ACADEMY OF ECONOMIC STUDIES

DOCTORAL SCHOOL OF FINANCE –DOFIN

CREDIT SCORING MODELLING :

A MICRO –MACRO APPROACH

MSc Student:SĂNDICĂ ANA-MARIA

Supervisor:Phd.MOISA ALTAR

Bucharest

2010

In order to reduce its capital requirement, banks use different credit risk models that are able to detect de difference between defaulter and a non-defaulter customer. In this paper I aim to make a comparison between these models and more to see which ones improve most when a macroeconomic variables is also introduce. What I would like to evidence in this paper is that more important than a particular model is the variables selection and the choice of a loss function that have to be minimized in order to treat the tradeoff between the profit considerations and best classification of customers.

Contents

1.Introduction

Banks and financial institutions play an important role in the economy as providers of credit. Beside government supervision and other regulatory conditions, capital requirements limit risks for depositors, and reduce insolvency and systemic risks. Unnecessary capital requirements restrain credit provision needlessly, whereas inadequate capital requirements may lead to undesirable levels of systemic risk

In December 2009, the Basel Committee on Banking Supervision has issued for consultation a package of proposals to strengthen global capital and liquidity regulations with the goal of promoting a more resilient banking sector.The Committee proposed a series of measures to promote the buildup of capital buffers in good times that can be drawn upon in periods of stress. A countercyclical capital framework will contribute to a more stable banking system, which will help attenuating, instead of amplifying, economic and financial shocks. In addition, the Committee suggested a forward looking provisioning based on expected losses, which captures actual losses more transparently and is also less pro-cyclical than the current "incurred loss"[1] provisioning model. There are many ways in which this can be done: dynamic provisioning, capital requirements change over time, capital requirements to reflect the expansion of credit and asset prices, setting a ceiling on the rate of lever.

Hugo Banziger[2] proposes mitigation measures pro-cyclicality, calibrating models to quantify risk based on extreme events, avoiding "disaster myopia”. Andrew G Haldane, Executive Director for Financial Stability Bank of England explained in his paper[3] that “disaster myopia refers to the propensity to underestimate the probability of adverse outcomes, in particular small probability events from the distant past. Economic agents have a tendency to base decision rules around rough heuristics (rules of thumb). The longer the period since an event occurred, the lower the subjective probability attached to it by agents (the “availability heuristic”) and below a certain bound, this subjective probability will effectively be set **to** zero (the “threshold heuristic”).Considering the fact that the financial system is composed largely of banks and financial institution, whose main activity is granting credits by taking into consideration a top-down approach from a macro-prudential analysis the convergence tends to a micro-prudential analysis.

---

[1] Strengthening the resilience of the banking sector - consultative document,December 2009.
[2] „Reform of the global financial architecture: a new social contract between society and finance”, Financial Stability Review, 2009, Chief Risk Officer and Member of the Management Board, Deutsche Bank
[3] “Why banks failed the stress test”, February 2009.

Christian Noyer[4] explains that it is necessary to complement the micro-prudential supervision of the macro-prudential, given the systemic importance and links between institutions, markets, instruments and how they evolve and lead to increased risk associated with the entire financial system.

Despite many innovations in banking, credit risk is typically the most significant source of risk and the largest source of credit risk is represented by loans; however, it also takes the form of positions in corporate bonds or transactions on over-the-counter markets, which involve the risk of default of the counterparty. Measuring credit risk involves estimation of a number of different parameters such as the likelihood of default on each instrument both on average and under extreme conditions; the extent of the losses in the event of default (or loss given default), which may involve estimating the value of collateral; and the likelihood that other counterparties will default at the same time. There are two general approaches to system-wide stress tests for credit risk, there are approaches based on loan performance data and there are approaches based on data on borrowers (financial leverage, interest coverage).

An important development in risk analysis introduced by the Basel II reforms is the consideration of changes in the quality of bank portfolios as a function of the business cycle and reflect capital requirements as a function of the credit quality of the borrower where credit quality is approximated by a rating, which may be public or internal to the bank.

Recent financial crises have highlighted the importance of macroeconomic analysis of the banking sector and its interactions with financial stability, which goes beyond the supervision of individual financial institutions by supervisory authorities and the macroeconomic analysis performed by central banks as part of the implementation of monetary policy. In this respect banks must take into consideration the financial stability and solvency of the entire financial system as a unit to the system.

In order to reduce its capital requirement, banks use different credit risk models that are able to detect de difference between a good and a bad customer. In this paper I want to make a comparison between these models and more to see which ones improve most when a macroeconomic variables is also introduced.

[4] Governor of Bank of France since 2003 and since March 2010 became the Chairman of Bank for International Settlements.

The paper is organized as follows. chapter 2 provides a review of literature on credit scoring models. chapter 3 describes the methodology ,data input, validation. Chapter 4 relates and analyses the results in a comparison approach and also a stress-testing scenario to capture the wage decreasing announced by the Finance Ministry at FMI's pressure. In Section 5 are presented the conclusion of this paper.

2.Literature Review.

In 1909 John M. Moody publishes first credit rating grades for publicly traded bonds and John Knowles Fitch founded the Fitch Publishing Company in 1913 in New York. David Durand is the pioneer of credit scoring when he in 1941 applied discriminant analysis proposed by Fisher (1936) to classifying prospective borrowers. In his paper published by the National Bureau of Economic Research he examined about 7200 reports on good and bad installment loans granted to 37 firms.

After World War II broke out, many finance lacked the experts to perform the work of credit analysis as many experienced people in the field joined the war. Those companies then asked experienced experts to put down their knowledge in credit assessment in the form of guidelines to help the relatively inexperienced make lending decision. The statisticians that designed the scorecard in the early days hoped to model after the practice of insurance companies who scored applicants based on age and gender to determine the premium. They reckoned that if banks could also have a scorecard for loan applicants as basis for making lending decision, it would help save the loan processing time and accomplish the objective of risk management.

In the 1950s, attempts had been made to merge automated credit decision making with statistical techniques to develop models that would help the making of credit decisions. But due to the deficiency of powerful computing tools, those models were substantially limited in sample size and model design. In 1963 Myers and Forgy compared discrimination analysis with regression in credit scoring application .In 1960 ,Altman introduced variables in a multivariate discriminant analysis and obtained a function depending on some financial ratios.

In 1988 ,Dutta & Shekhar were the first that developed neural networks model for corporate bond ratings and their results showed that this technique performed better in predicting bond rating from a given set o financial ratio. The advantages of this technique has been exploited in many researches such as the fact that non-numeric variables could be part of the model since there are no linearity constraints (Coats&Fant 1993).The most problem related to neural networks is that does not reveal the significance of each of the variables in the final, the derived weights could not be interpreted. In 1997, Hand and Henley made a comparison among logistic regression ,neural networks and other techniques and in their paper also present the Information Value criterion of selection variables.

The neural networks techniques dominates the literature on business failure in the second half of the 1990s and the main studies published are on corporate level due to data availability. West(2000) investigates the credit scoring accuracy of five neural network models and compared them with other techniques such as logistic regression, decision trees etc and the results demonstrate that although neural networks have better results logistic regression is a good alternative to them. In his paper he treats also the loss function and the same problem was evaluated by Liu(2002) ,when he focused on five techniques and one of the most accurate model was a multilayer perceptron. Komorád (2002) investigated credit scoring prediction accuracy and performance on a data set from a French bank. The credit score prediction performances of the following models were compared: logistic regression, multi-layer perceptron (MLP) neural network and radial basis neural networks were compared. The results obtained indicated that the methods, namely the logistic regression, multi-layer perceptron (MLP) and radial basis function (RBF) neural networks give very similar results, however the traditional logit model seems to perform marginally better. Baesens(2003) examines different credit scoring techniques and as a new approach he combined neural networks in a survival analysis function.

Roszbach(2003) evaluated loan applicants with a bivariate Tobit model with a variable censoring threshold considering that banks should take into account not only the status of default or not defaulted but the moment of this event.Lai, Yu, Wang and Zhou (2006a) indicated that a propagation neural network (BNN) with an identity transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous function arbitrarily well given a sufficient amount of middle-layer unit.

Bellotti and Crook (2007)show that survival analysis is competitive for prediction of default in comparison with logistic regression and also they included macroeconomic variables and a cost decision matrix. In a review of consumer credit risk models

,Crook,Edelman and Thomas (2007) discussed the difficulties in setting a cut-off and the concern about strategy curve. Malik and Thomas(2008) incorporated both consumer specific ratings and macroeconomic factors in the framework of Cox proportional hazard model.

A comparison between logistic regression and a classification tree was developed by Kocenda and Vojtek (2009) and their research conducted to the idea that although socio-demographic variables are important for the model but behavioural variables should be incorporated for managing the portfolio .Rommer(2005)come to idea that there is no major difference between logit and probit regression models. Rauhmeier(2006) analyzed the validation process for probabilities of default and includes also the concept of “rolling window 12 months “ and in 2010,Sabato also presents the importance of the model’s validation and how back testing is the essential part of this process.

3.Methodology

3.1.Comparison of credit scoring models

3.1.1.Discriminant analysis.

In 1936 Fischer introduced the linear discriminant function with the purpose to find a combination of variables that best separated two groups whose characteristics were available and in his work the groups were different subspecies of a plant for example.

In credit scoring the two groups are those classified by the lender as non-defaulter and defaulter and the characteristics are the application form details.

Let $Y = w_1X_1 + w_2X_2 + \cdots + w_pX_p$ be any linear combination of the characteristics $X = X_1 + X_2 + \cdots + X_p$.

Fisher recommended that if the two groups have a common sample variance then a sensible measure of separation is

$$M = \frac{\text{distance between sample means of two groups}}{(\text{sample variance of each group})^{\frac{1}{2}}} \quad (1)$$

For the goods and bads assume sample means, m_G respectively m_B and S is the common sample variance. If $Y = w_1X_1 + w_2X_2 + \cdots + w_pX_p$ then the corresponding separating distance M would be:

$$M = w^T \cdot \frac{m_G - m_B}{(w^T \cdot S \cdot w)^{\frac{1}{2}}} \quad (2)$$

Differentiating this with respect to w and setting derivative equal to zero the value of M is maximized when

$$w^T \propto (S^{-1}(m_G - m_B)^T) \quad (3)$$

3.1.2 Logistic regression.

In 1798, Malthus claimed that human population will increase in geometric progressions until 1845 when Pierre Francois Verhulst studied (1845) the population growth and used the logistic function. In credit scoring the first academic work was published by Wiginton in 1980 and the results were not very good.

If p_i is the probability that applicant *i* has defaulted, the purpose is to find w^* that best approximate

$$p_i = w_0 + x_{i1}w_1 + x_{i2}w_2 + \cdots + x_{ip}w_p \quad (4)$$

As it can be noticed in the equation (11) the right hand side could take any value from $-\infty$ to $+\infty$ but the left hand side is a probability and so should take only values between 0 and 1.The purpose was to find a function of p_i which could take values between 0 and 1 and one such function is the log of probability odds.

The linear combination of the characteristic variables is:

$$\log\left(\frac{p_i}{1-p_i}\right) = w_0 + w_1x_1 + w_2x_2 + \cdots w_px_p = w \cdot x^T \tag{5}$$

Taking exponential on both sides of (14) leads to the equation:

$$p_i = \frac{e^{w \cdot x}}{1 + e^{w \cdot x}} \tag{6}$$

Dividing by $e^{w \cdot x}$,the equation (15) becomes:

$$p_i = \frac{1}{1 + e^{-w \cdot x}} \tag{7}$$

Considering the encoding of good client, 0 and bad client 1, the probability of a customer to be bad is given by the following formula:

$$P(Y = 1|X) = \frac{1}{1 + e^{-wx}} \tag{8}$$

The probability of a client to be good is *1-probability of being bad*, thus the result is:

$$P(Y = 0|X) = \frac{e^{-wx}}{1 + e^{-wx}} \tag{9}$$

The probability of observing either class is given by the probability function of the Bernoulli distribution:

$$p(Y|X) = P(Y = 1|X)^{y}(1 - P(Y = 1|X))^{1-y} \quad (10)$$

The method used to calculate the coefficients *w* is the maximum likelihood approach and not ordinary least-squares. Considering the fact that the observations are drawn independently the joint probability function is:

$$\prod_{i=1}^{N} P(y_i = 1|x_i)^{y_i}(1 - P(y_i = 1|x_i)^{1-y_i} \quad (11)$$

The log likelihood function then becomes:

$$LL = \sum_{i=1}^{N} y_i \log(P(y_i = 1|x_i)) + (1 - y_i)\log(1 - P(y_i = 1|x_i)) \quad (12)$$

This leads to an iterative Newton-Raphson method to solve the equation that arises. Although theoretically logistic regression is optimal for a much wider class of distributions than linear regression, comparing these two types of regression, the results show that they are similar until either p becomes close to zero or close to 1.

3.1.3.Probit Regression

In 1934, Chester Bliss introduced a probit model in his paper[5] where suggested to transform a percentage into a probability unit (or probit).

Grablowsky and Talley in 1981 used for the first time the probit function in credit scoring. In probit analysis if N(x) is the cumulative normal distribution function so that:

[5] Bliss Cl. (1934)-“The methods of probits”-Science 79(2037):38-39.

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy \qquad (13)$$

Then the purpose is to estimate $N^{-1}(p_i)$ as a linear function of the characteristics of the applicant so:

$$N^{-1}(p_i) = w_0 + w_1 x_{1i} + w_2 x_{2i} + \cdots w_p x_{pi} = w \cdot x_i^T \qquad (14)$$

Again, p_i takes only values between 0 and 1, $N^{-1}(p_i)$ takes values between $-\infty$ to $+\infty$. Considering:

$$W = w_0 + w_1 X_1 + w_2 X_2 + \cdots w_p X_p = w \cdot x_i^T \qquad (15)$$

is a measure of goodness of an applicant and the fact that the applicant is defaulter or not depends on whether the value of W is greater or less than a cut-off level C. Supposing that C is a variable with standard normal distribution using maximum likelihood estimation w ,the vector of weights, could be estimated.

Consider the probability of a client to be defaulter (bad) as:

$$P(Y = 1|X) = N(w \cdot x_i^T) \qquad (16)$$

In order to calculate the log-likelihood function, the joint probability function is given by this formula:

$$\prod_{t=1}^{N} P(y_t = 1|x_t)^{y_t} (1 - P(y_t = 1|x_t)^{1-y_t} \qquad (17)$$

The logarithm function transforms the product into following sums:

$$LL = \sum_{i=1}^{N} y_i \log\left(P(y_i = 1|x_i)\right) + (1 - y_i)\log\left(1 - P(y_i = 1|x_i)\right) \quad (18)$$

In 2006 Bishop found that the results from probit regression tend to be similar to those of logistic regression.

3.1.4 Tobit Regression

In 1958 James Tobin proposed the Tobit Model in order to describe the relationship between a non-negative dependent variable and an independent vector, this assuming that can estimate p_i by:

$$p_i = \max(w \cdot x_i^T, 0) = \max\left(w_0 + w_1 x_{1i} + w_2 x_{2i} + \cdots w_p x_{pi}\right) \quad (19)$$

One issue is that the right-hand side should be positive and although the tobit transformation deals with negative probabilities, the estimated probabilities will not be greater than 1. A more symmetrical model would be:

$$p_i = \mathrm{Min}\left\{1, \max\{w \cdot x_i^T, 0\}\right\}, \quad (20)$$

3.1.5 Nearest-neighbor approach.

The nearest-neighbor method is a standard non-parametric approach to the classification problem first suggested by Fix and Hodges in 1952. In credit scoring the first approach was made by Barcun and Chatterjee in 1970 and later by Henley and Hand in 1996.

The main idea is to choose a metric on the space and then with a sample of past applicants as a representative standard, a new applicant is classified as good or bad depending on the proportions of defaulter and non-defaulters among the k nearest applicants from the representative sample—the new applicant's nearest neighbours. A neighbour is deemed nearest if it has the smallest distance, in the Euclidian[6] sense, in the input space.

The three parameters needed to run this approach are: the metric, how many applicants k constitute the set of nearest neighbours, and what proportion of these should be good for the applicant to be classified as non-defaulter. In 1984 Fukanaga and Flick introduced a general metric of the form:

$$d(x_1,x_2) = \sqrt{(x_1 - x_2)A(x_1)(x_1 - x_2)^T} \qquad (21)$$

Where A(x) is symmetric positive definite matrix and it is called local metric if it depends on x and global metric if it is independent of x.

In 1996, authors Henley and Hand suggested a metric of the form: $A = \pi r^2$

$$d(x_1,x_2) = \sqrt{((x_1 - x_2)^T (I + Dw \cdot w^T)(x_1 - x_2)} \, , \qquad (22)$$

where I is the identity matrix. In their working paper the values for D is between 1,4 and 1,8.

3.1.6 Linear Programming.

This is a technique that comes from the field of resource allocation problems and the original research in this area occurred during 1930's with studies on game theory (Morgenstern and von Neumann) and input-output models (Leontief).

In 1965, Mangasarian was the first to recognize that linear programming could be used in classification problems where there are two groups and there is a separating hyper plane. To find the weights $(w_1, w_2, \dots, w_p)$ that minimize the sum of the absolute values of these deviations (MSD) one has to solve the following linear program:

Minimize $\sum_i e_i^2$ subject to:

---

[6] Eucledian distance, $d(x_1,x_2) = \sqrt{(x_1 - x_2)(x_1 - x_2)^T}$

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + e_t \tag{23}$$

Hardy Jr. and Adrian Jr. (1985) presented an example to show how linear programming can be used to construct a credit scoring model and Vladimir et al. (2002) constructed a quadratic programming model which incorporated experts' judgment for credit risk evaluation. The review papers of Nath, Jackson and Jones (1992) compared the linear programming and regression approaches to classification on several data and their results suggest that the linear programming approach does not classify quite as well as the statistical methods.

3.1.7 Classification Trees.

The main idea is to split the set of application answers into different sets and then identify if these sets are good or bad depending on the majority in that set. In credit scoring the idea was developed by Makowski (1985) and Coffman (1986) .

The set of application data A is first split into two subsets and each of these sets is then again split into two in order to produce even more homogeneous subsets, then the process is repeated, from this coming the approach name of recursive partitioning. The process stops when the subsets meet the requirements to be terminal nodes of the tree. Each terminal node is then classified as a member of $\mathbf{A}_G$ or $\mathbf{A}_B$.

The decisions imply three procedures:

- What rule to use to split the sets into two – the splitting rule;
- How to decide that a set is a terminal node – the stopping rule;
- How to assign terminal nodes into good and bad categories-the assignment rule.

According to Thomas et al. (2002), Breiman and Friedman each independently came up with the idea of using analytical tools to determine the rule set in 1973 and after one year a procedure for deriving decision trees (Classification and Regression Trees) and their concept was first applied to credit scoring by Makowski and Coffman in 1985 and 1986 respectively.

3.1.8 Neural Networks

Early neural model-based approach dates back to 1943, once the first appearance of the neuron model, proposed model of neurophysiology W.S McCulloch and mathematician W. Pitts. Particular interest to the neuron model was observed after the first appearance of works in mathematical modeling of learning processes. A first occurrence of this kind took place in 1947, and is represented by the model of learning of D.O. Hebb, who opened unsuspected directions in neural calculations. Another important step on the road neural development approach was made in 1957, with the appearance of Frank Rosenblatt's work, dedicated to a simplified neural model probabilistic nature, known as the *perceptron*. Fundamental element of any neural network is an artificial neuron. Neurons that are part of neural networks, have different functions, they are specialized in performing certain types of activities. From this viewpoint, a neural network contains three basic types of neurons:

• input units, acquiring the input variables values or standard values of input variables, this means that the input neurons have no own computer functionality itself, but an interface role, the input neurons form the so-called input layer or the input;

• Neurons intermediaries are brain cells are located between the input layer and output layer having a function purely computer;

• output neurons, which calculates predicted values by neural network and comparing these values with specific target values or reference values, depending on the outcome comparisons, weights or connections are updated.

Each elementary unit of a neural network, i.e. each neuron has one or more an internal state and an exit. Functionality of a neuron consists in that it produces a single output, represented by a single numeric value, depending on the nature or status of such units, determined based on state information that the neuron input.

Each value of $X_1, X_2, \ldots, X_p$ is a variable and the weights, also known as synaptic weights[7] are written in the order (k, p) where k[8] indicates the neuron to which the weight applies and p indicates the variable.

$$u_k = w_{k0}x_0 + w_{k1}x_1 + \cdots + w_{kp}x_p = \sum_{q=0}^{p} w_{kq}x_{p'} \tag{24}$$

$$y_k = F(u_k) \tag{25}$$

The u_k value is then transformed using an activation function known as transfer function. Various alternative activation functions have been used:

- Threshold Function

$$\begin{cases} F(u) = 1, if\ u \geq 0, \\ F(u) = 0, if\ u < 0 \end{cases} \tag{26}$$

- Logistic Function:

$$F(u) = \frac{1}{1 + e^{-au}} \tag{27}$$

- Hyperbolic tangent :

$$F(u) = 1 - \frac{2}{1 + e^{2u}} \tag{28}$$

---

[7] If the sign is positive then the weights are known as excitory because they would increase the corresponding variable and if is negative they would reduce the value of u_k for positive variables are known as inhibitory.

[8] If the architecture is a single layer neuron then k is 1

In order to apply neural network technique the problem of specifying the weights that are used in the architecture built and this task is accomplished by the learning algorithm which trains the network and iteratively modifies those weights until a condition is satisfied ,especially when the error between the desired output and the one produced by the model is minimal.

There are three typologies of learning mechanism for neural networks: supervised, unsupervised and reinforced learning. The training set is used in order to offer the desired output and in this manner to adjust the weights. In comparison with this the second mechanism, unsupervised learning is using a set without the desired output and the weights are adjusted based on self-organizing. The reinforced learning mechanism assumes that the best method to adjust the weights is to introduce prizes and penalties as a function of network response.

A multilayer perceptron is composed of an input layer of signals, an output layer and a number of layers of neurons between, called hidden layers. The weights applied in the input neurons may differ from the weights applied in hidden layers. A three layer network is shown bellow.

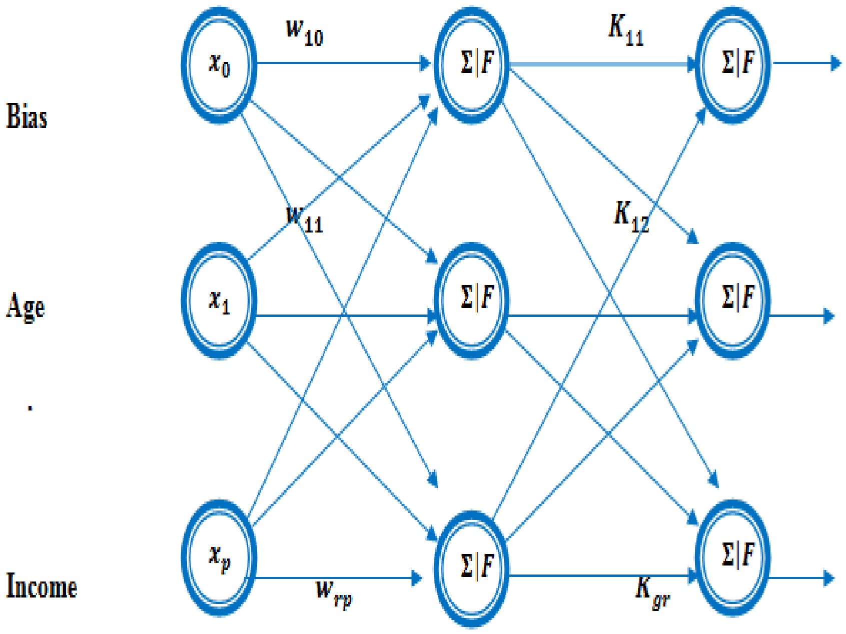


Figure 1-Multilayer perceptron

Input layer *p* inputs   Hidden layer *r* neurons   Output layer *s* neurons

q=1,... ,p   k=1,.., r   v=1,..s

$$y_k = F_1\left(\sum_{q=0}^{p} w_{kq} x_q\right) \quad \textbf{(29)}$$

Where the subscript 1 in equation (29) indicates the fact that it is the first layer and y_k are the outputs from the first hidden layer and the output of one layer is the input for the following layer; the relation became:

$$z_v = F_2\left(\sum_{k=1}^{r} K_{vk} y_k\right) = F_2\left(\sum_{k=1}^{r} K_{vk}\, F_1\left(\left(\sum_{q=0}^{p} w_{kq} x_q\right)\right)\right) \quad \textbf{(30)}$$

Where z_v is the output of neuron v in the output layer, v=1...s, F2 is the transfer function the output layer and the weight applied to the y_k layer is K_{vk}.

The method for calculating these weights is also known as training process, and most frequently method is the *back-propagation algorithm*, that looks for the minimum error function in weight space using the method of *gradient descent*. The solution of the learning problem is the combination of weights which minimizes the error function.

First, all weights are equal to some randomly chosen numbers and a training pair is selected, the forward pass is ending when z_v is calculated. The backward pass consists of distributing the error between known value o_v and calculated one, z_v, through the network proportionally with the contribution made by each weight. After that, a second pair is selected and both forward and back pass are calculated this process is known as epoch and the repeated process ends up when a stopping criterion has been fulfilled.

Defining the error, $e_v(t)$ as

$$e_v(t) = o_v(t) - y_v(t) \quad \textbf{(31)}$$

Where $e_v(t)$ is the observed outcome for case t in neuron v and $y_v(t)$ is the predicted outcome. The purpose is to choose a vector of weights that minimizes the average value over all training cases of:

$$E(t) = \frac{1}{2}\sum_{v=1}^{s} e_v^2(t) \tag{32}$$

where s, is the number of neurons in the output layer.For any neuron *v* in any layer *c* the relations could be written as follows:

$$u_v^{[c]} = \sum_{k=0}^{r} w_{vk} y_k^{[c-1]} \tag{33}$$

$$y_v^{[c]} = F(u_v^{[c]}) \tag{34}$$

Writing the partial derivative of E(t) with respect to weight $w_{vk}(t)$ and splitting into a chain rule:

$$\frac{\partial E(t)}{\partial w_{vk}(t)} = \frac{\partial E(t)}{\partial e_v(t)} \cdot \frac{\partial e_v(t)}{\partial y_v(t)} \cdot \frac{\partial y_v(t)}{\partial u_v(t)} \cdot \frac{\partial u_v(t)}{\partial w_{vk}(t)} \tag{35}$$

From equation (32):

$$\frac{\partial E(t)}{\partial e_v(t)} = e_v(t) \tag{36}$$

From equation (31):

$$\frac{\partial e_v(t)}{\partial y_v(t)} = -1 \tag{37}$$

From equation (34)

$$\frac{\partial y_v(t)}{\partial u_v(t)} = F'(u_v(t)) \tag{38}$$

From equation (33)

$$\frac{\partial u_v(t)}{\partial w_{vk}(t)} = y_k(t) \tag{39}$$

Substituting equations (36)-(39) in equation (35) the result is:

$$\frac{\partial E(t)}{\partial w_{vk}(t)} = -e_v(t)F'(u_v(t))y_k(t) \tag{40}$$

Between forward pass and backward pass is therefore:

$$\Delta w_{vk}(t) = -\eta \frac{\partial E(t)}{\partial w_{vk}(t)} = \eta\delta_v(t)y_k(t) \tag{41}$$

Where

$$\delta_v(t) = e_v(t)F'(u_v(t)) \tag{42}$$

η=training rate coefficient.

Smaller values for this training rate coefficient improve accuracy but extend the training time. The equation (51) is known as "Delta Rule" and was developed by Widrow and Hoff. It is one of the most commonly used learning rules. For a given input vector, the output vector is compared to the correct answer. If the difference is zero, no learning takes place; otherwise, the weights are adjusted to reduce this difference.

If the neuron v is in the output layer then the value $e_v(t)$ is directly observable but if it is in the hidden layer $o_v(t)$ it is not observable and in this case the formula for $\delta_v(t)$ is calculated different. In general this might be done by this formula:

$$\delta_k^{[c-1]} = F_k^{[c-1]'} \sum_{v=1}^{s} \delta_v^{[c]} w_{vk}^{[c]} \tag{43}$$

From (41) and (43) the change in weight becomes:

$$\Delta w_{vk}(t) = \eta \delta_v^{[c]} y_k^{[c-1]} \qquad (44)$$

For a giving training set the weights in the network are the only parameters that can be modified to make the quadratic error E as low as possible. This can be minimized by using an iterative process of gradient descent for which the gradient is:

$$\Delta E = \left(\frac{\partial E(1)}{\partial w_{vk}(1)}, \frac{\partial E(2)}{\partial w_{vk}(2)}, \dots, \frac{\partial E(t)}{\partial w_{vk}(t)}\right) \qquad (45)$$

The whole learning problem has now been reduced to the questions of calculating the gradient of a network function with respect to its weights, minim of the error function, where $\Delta E = 0$.

The main advantages have to be found in their learning capabilities and the fact that the derived model does not make any assumption on the relations among input variables and as an important drawback is that the development of neural networks requires quite a lot of expertise .

3.1.9.Survival Analysis

In 1992 Narain proposed survival analysis as a technique to be used in credit scoring and a comparison among basic survival analysis and logistic regression was developed by Banasik , Crook and Thomas[9] in 1999.

Let T be the time until a loan defaults then :

- Survival function:

$$S(t) = Prob(T \geq t), \text{where } F(t) = 1 - S(t) \text{ is the distribution function} \qquad (46)$$

- Density function f(t) ,where

[9] Banasik,Crook and Thomas (1999), Not if but when borrowers default,J. *Oper. Res. Soc.*, 50, 1185-1190.

$$Prob\{t \leq T \leq t + \delta t\} = f(t)\delta t \qquad (47)$$

- Hazard function :

$$H(t) = \frac{f(t)}{S(t)}, \text{and } Prob\{t \leq T \leq t + \delta t | T \geq t\} = h(t)\delta t \qquad (48)$$

In survival analysis two models have been proposed to explain the failure behavior of a customer: proportional hazard models and accelerated life models. Considering $X = (x_1, \dots, x_p)$ are the application (explanatory) characteristics the the accelerated life model assumes that:

$$S(t) = S_0(e^{wx} t) \qquad (49)$$

The proportional hazard assumes that:

$$h(t) = e^{wx} h_0(t) \qquad (50)$$

If an assumption is made by considering that $h_0(t)$ belong to a particular family of distributions then we deal with the parametric approach. In Cox (1972)[10] pointed out that in proportional hazard the vector of weights w could be estimated without knowing the baseline function.

3.2.Validation of Rating Models

The requirements of the IRB approach is that "the institution shall have a cycle of model validation that includes monitoring of model performance and stability "[11]This process includes a quantitative and a qualitative validation. The first part assumes a back testing and a

---

[10] D. R. Cox (1972), Regression models and life-tables (with discussion), *J. Roy. Statist. Soc.Ser. B,* 74, 187-220.

[11] Committee of European Banking Supervisors (CEBS) (2005) Guidelines on the implementation, validation and assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) Approaches.

benchmark analysis and for qualitative analysis the use test and data quality are the main components.

For statistical models the quantitative validation is very important and it is build up by the following criterions[12]:Discriminatory power, Calibration and Stability

When a model is used to determine the probability of default of a customer the main important aspect is to check if the model maintains the discriminatory power and it is better to use it instead of a random split of the customers.

The basic idea is that low probabilities of default should be mapped to those that didn't default and vice versa higher probabilities of default should correspond to defaulted client.

In order to see this concentration of probabilities of default, Cumulative Accuracy Profile Curve is plotting on X the cumulative frequencies of all cases and on y axis is the cumulative frequency of bad cases .

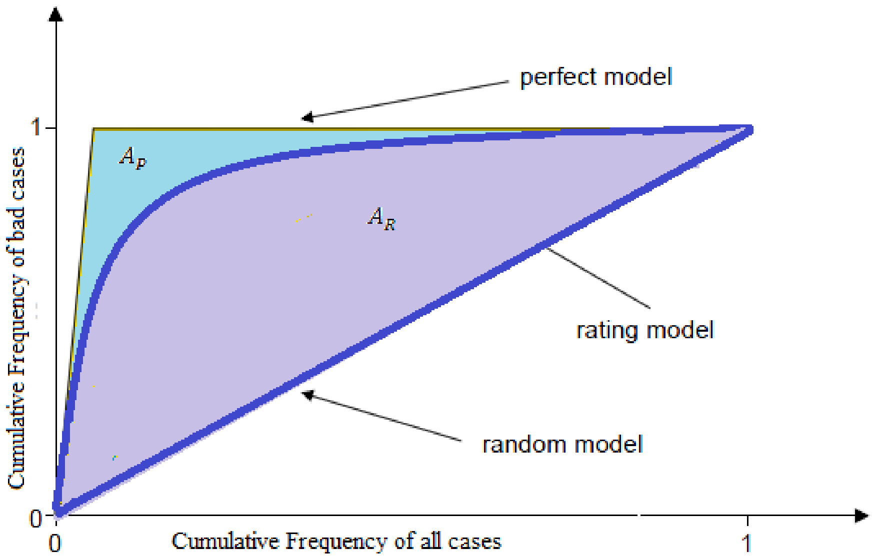


Figure 2-Cumulative Accuracy Profile

For a random model having no discriminative power the fraction x of all debtors with lowest rating scores will contain x percent of all defaulters .The rating model is between this model and the perfect one ,the one that will assign the lowest scores and implicit the higher

[12] DEUTSCHE BUNDESBANK, Monthly Report for September 2003, Approaches to the validation of internal rating systems.

probabilities of default to the defaulters. The quality of a rating system is measured using Accuracy Ratio (AR):

$$AR = \frac{A_R}{A_P} \tag{51}$$

And the closer value of AR to one the better the rating model is.

The confusion matrix offer a convenient way to compare the frequencies of actual versus predicted status for the given model applied. Considering that a default client also named as bad client have a status of ”1” and for a non-defaulter client (good client) the status is “0”.

		Actual	
		Defaulter	Non-Defaulter
Predicted	Defaulter	A	C
	Non-Defaulter	B	D

Table 1-Confusion Matrix

$$Type\ Error\ I - \frac{B}{B + A} \tag{52}$$

$$Type\ Error\ II = \frac{D}{C + D} \tag{53}$$

The Error of type I or α is also named the credit risk rate because is the rate of defaulters that are categorized as non-defaulters from the model ,this is usually when the accepting rate is very high and the proportion of clients accepted for receiving a loan is higher. Bank institutions should manage this accepting rate in order to reduce this misclassification rate .Also the Error of type II is a ratio of mismatch the category between the clients. Also know as commercial risk or β this error is happening when a non-defaulter is rejected because the model is considering his as being defaulter. This leads to a loss in the bank’s profit because the client rejected is seen as a potential cash flow asset. Also when a bank has an error of type II constantly higher during time, then its share of market is decreasing.

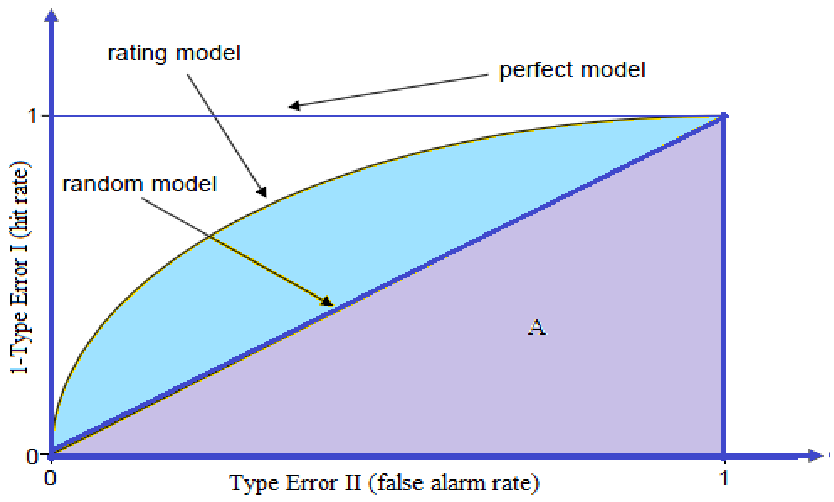


Figure 3-ROC Curve

By comparing ROC[13] curves one can study the difference in the classification accuracy between two classifiers ,for a new model used it is better to have for a given Type I error Rate a smaller Type II Error rate .

Defining hit rate, HR(C) as

$$HR(C) = \frac{H(C)}{N_D} \tag{54}$$

Where H(C) is the number of defaulters predicted correctly with the cut-off value C and N_D is the total number of defaulters in the sample .This could be expressed as the fraction of defaulters that was classified correctly given cut-off value C. The false alarm FAR(C) is:

$$FAR(C) = \frac{F(C)}{N_{ND}} \tag{55}$$

,where F(C) is the number of false alarms, also defined as the number of non-defaulters that were classified incorrectly as defaulters by using the same cut-off C. N_{ND} , is the number of non-defaulters in the sample. For all cut-off values C that are contained in the range of the

[13] Receiver Operating Characteristic was developed in 1940 to measure radar operator's ability to distinguish between a true signal and a noise.

rating scores the quartiles HR(C) and FAR(C) are calculated and plotted one versus other ,the result is ROC curve.

In order to analyze the performance of a model the area under curve need to be calculated, the relation is positive the larger the area the better the model. Denote this area by A this could be calculating by using this formula:

$$\int_0^1 HR(FAR)d(FAR) \tag{56}$$

The perfect model has an area equal to 1 and for a random model the value of A is 0.5.Using this area under the curve, another indicator is calculated, Gini[14] Coefficient, also known as Accuracy Ratio[15]. Vilfredo Pareto declared that income inequality would reduce in richer societies after in 1896 he noted how 80 percent of the land in Italy was owned by 20 percent of the population and this ratio also applied to land ownership and income in other countries too. In 1905 the American mathematician Max Otto Lorenz(1876-1959), develop the Lorenz Curve in order to display the income inequalities within society. In 1910, Corrado Gini proved that Pareto's statement is wrong by comparing income inequalities between countries using his coefficient.

Area under the curve(AUROC) and Accuracy Ratio are connected by means of the linear transformation and this fact was proven by Engelmann[16] in his paper.

$$AR = 2A - 1 \tag{57}$$

Pietra Index can be defined as the maximum area a triangle can obtain that is inscribed between the ROC curve and the diagonal of the unit square:

---

[14] In 1920, Gini founded the journal *Metron* and in 1923, he moved to the University of Rome, where he later became a professor, founded a sociology course, set up the School of Statistics (1928), and founded the Faculty of Statistical, Demographic, and Actuarial Sciences (1936). In 1926, he became president of the Central Institute of Statistics.

[15] The calculation of Accuracy Ratio it could me made either using Cumulative Accuracy Profile or deducted from Area under the Curve (AUROC) used for ROC Curve.

[16]Bernd Engelmann- Measures of a Rating's Discriminative Power- Applications and Limitations

$$Pietra\ Index = \frac{\sqrt{2}}{4}\max_{c}|HR(C) - FAR(C)| \tag{58}$$

Interpreting the Pietra Index as the maximum difference between the cumulative frequency distribution for the score values of goes and bad clients then the Kolmogorov Smirnov[17] test could be applied when the null hypothesis is that the score distributions are identical and could be rejected at level q if Pietra index equals or exceeds he following value:

$$D = \frac{D_q}{\sqrt{N \cdot p(1-p)}} \tag{59}$$

Where N is the number of cases in the sample examines and p refers to the observes default rate .If the Pietra Index is greater or equal to D then significant difference between those two distribution exists.

Information Entropy is a summary measure of the uncertainty that a probability distribution represents. This concept has its origin in the files of Statistical Mechanics and Information Theory[18]

Defining Information Entropy H(p) of an event with probability p as :

$$H(p) = -(plog(p) + (1-p)\,log(1-p)), \tag{60}$$

It can be observed that the information entropy takes its maximum at =1/2 the stat at which the uncertainty is maxim. If p is zero then the event will occur with certainty and thus not reveal any information. Consider the event of default as being D and the complementary event that does not default as $\bar{D}$,the information entropy *H* could be apply to $P(D|S)$,the conditional probability of default given the rating score *S*:

$$H\big(P(D|S)\big) = -(\,P(D|S)logP(D|S) + P(\bar{D}|S)logP(\bar{D}|S)) \tag{61}$$

[17] The test is named after the mathematician Andrei Nikolaevich Komogorov(1903-1985) ,who in 1933 published “Foundations of the Calculus of Probabilities”, a definitive work on probability theory.

[18] Shannon C and Weaver W-The Mathematical Theory of Communication-University of Illinois Press, Urbana ,1949

The expected value of (61) it is calculated and can be written as follows:

$$H_s = -E[P(D|S)logP(D|S) + P(\bar{D}|S)logP(\bar{D}|S)] \quad (62)$$

The difference between information entropy and conditional Entropy should be larger in order to have an information gain by application of the rating scores .This difference is also known as Kullback –Leibler Distance and it was introduced in 1951 by Solomon Kullback and Richard Leibler[19].This is a non-symmetric measure of the difference between two probability distributions.

$$Kullback - Leibler\ Distance = H(p) - H_s \quad (63)$$

In order to have a commen scale for any underlying population a nother measure is used and this is made by standardizing the Kullback –Leibler Distance:

$$CIER - \frac{H(p) - H_s}{H(p)} \quad (64)$$

This is named Conditional Information Entropy Ratio and compares the amount uncertainty there is about default in case where no model is applied to the amount of uncertainty left over after a model is introduced .If the model have no predictive power then CIER is zero and otherwise the perfect model has a ratio of 1.

If the CIER measures the gain information that is reached by using a rating model instead of other rating model ,Information Value measures the difference between the score defaulter distribution and the non-defaulter score distribution.

$$IV = E\left[log\frac{f_D(S)}{f_{\bar{D}}(S)}\middle|D\right] + E\left[log\frac{f_{\bar{D}}(S)}{f_D(S)}\middle|\bar{D}\right] \quad (65)$$

Considering the $f_D$the density of score distribution for defaulters and $f_{\bar{D}}$,for non-defaulters then (66) is defined as the sum of the relative entropy of non-defaulter distribution with respect to the defaulter distribution and the defaulter distribution with respect of non-

[19]Kullback ,S;Leibler,R.A(1951)"On Information and Sufficiency”, Annals of Mathematical Statistics

defaulter distribution[20].Higher values indicates a rating system with higher power of discrimination.

Brier[21] Score, is a measure was proposed by Brier in 1951 and the formula is:

$$BS = \frac{1}{n}\sum_{n=1}^{n}(p_n{}^{forecast} - \theta_j)^2 \tag{66}$$

Where,

$$\theta_j = \begin{cases} 1, & if\ obligor\ j\ defaults \\ 0, & otherwise \end{cases} \tag{67}$$

Hosmer Lemeshow[22] Test assumes that being $p_0^{forecast}, \dots, p_k^{forecast}$,the forecasted default probabilities of debtors the statistic is defined as follows:

$$T_k = \sum_{c=0}^{k} \frac{(N_c p_c^{forecast} - d_c)^2}{N_c p_c^{forecast}(1 - p_c^{forecast})} \tag{68}$$

This follows a χ^2 distribution with k-2 degree of freedom and this is available when this test is used in model finding in “in sample” analysis but when it is used for back testing this distribution is with k degree of freedom.

Normally the predicted default probability of each borrower is individually calculates and since Hosmer Lemeshow Chi Square Test requires averaging the predicted probability of defaults some bias might arise in the calculation. In order to avoid this problem Spiegelhalter in 1986 introduced a further generalization also known as Spiegelhalter[23] Test.

---

20 Basel Committee on Banking Supervision (BCBS) (2005b) Studies on the Validation of Internal Rating Systems (revised). Working Paper No. 14.

21 Brier, G. W., Monthly -"Verification of forecasts expressed in terms of probability". *Monthly weather review*

22 Hosmer, D. and Lemeshow, S. (2000), Applied logistic regression, Wiley series in Probability and Statistics.

23 Spiegelhalter, D. (1986), Probabilistic prediction in patient management and clinical trails,*Statistics in Medicine*, Vol. 5, pp. 421-433.

The test is based on Brier Score[24] ,eq.(66) and the null hypothesis is that the observed default rate is equal with the forecasted default rate and:

$$E[BS] = \frac{1}{N}\sum_{i=1}^{N} p_i^{forecast} \cdot (1 - p_i^{forecast}) \tag{69}$$

$$Var[BS] = \frac{1}{N^2}\sum_{i=1}^{N} (1 - 2p_i^{forecast})^2 \cdot p_i^{forecast} \cdot (1 - p_i^{forecast}) \tag{70}$$

Under the null hypothesis and the assumption that the defaults are independent and using Central Limit Theorem the statistic is:

$$Z = \frac{BS - E[BS]}{\sqrt{Var[BS]}} \tag{71}$$

,which follows a standard normal distribution.

Kuipers Score [25] is measuring the distance between the hit rate and false alarm rate and a model that discriminates between defaulters and non-defaulters has a value of this score of 1.

Granger and Pesaran(2000) show that the Pesaran –Timmermann ,having the null hypothesis assuming that the distribution of the forecasted and realized probabilities of default are independently and statistic can be expressed as:

---

[24]Brier Score is also knwon as Mean Square Error(MSE)

[25] Was originally proposed by Peirce (1884),also knwon as Hannsen-Kuipers test.

$$GP = \frac{\sqrt{N}KS}{\sqrt{\frac{p_f(1-p_f)}{p_o(1-p_o)}}} \sim N(0,1) \tag{72}$$

Most of test assumes independence of defaults and the existence of default correlation within a portfolio has the effect of reinforcing the volatility of default rate but "From a conservative risk management point of view, assuming independence of defaults is acceptable, as this approach will overestimate the significance of deviations in the realised default rate from the forecast rate." [26]Huschens and Stahl (2005)[27] show evidence that, for a well diversified German retail portfolio, asset correlations are in the range between 0% and 5%, which implies even smaller default correlations.

Taking into consideration the situation before 2008 and the creditworthiness of the companies that defaulted ,rating agencies might require now a higher capital buffer to attain the same credit rating as compared to the situation before 2008[28].

The goodness of fit tests are important for financial institutions when they are trying to use the model that is more suitable for its credit portfolio. Recently studies[29] showed that Hosmer-Lemeshow is too conservative."

3.3 Data Input

In a banking institution, the primary role of capital in addition to transfer of ownership is to act as a buffer for unexpected losses absorption, protect depositors and ensure the confidence of investors and rating agencies. In contrast, regulated capital (Regulatory Capital) refers to minimum capital requirements that banks are obliged to hold under the regulation of surveillance. While economic capital is to act as a buffer against all risks which

---

[26]François Coppens,Fernando González and Gerhard Winkler –The performance of Credit Rating Systems in the assesment of collateral used in eurosystem monetary ploicy operations ,European Central Bank,Occasional Paper Series, Nr 65.July 2007

[27] Huschens, S. and Stahl, G. (2005), A general framework for IRBS backtesting, *Bankarchiv,Zeitschrift für das gesamte Bank und Börsenwesen*, 53, pp. 241-248.

[28] Standard&Poor's has downgraded several financial companies on December 2008 such as Barclays Bank. Deutsche Bank ,Royal Bank of Scotland, Credit Suisse

[29] Andreas Blochlinger and Markus Leippold-„ New Goodness-of-Fit Test for Event Forecasting and Its Application to Credit Default Models"

may compromise the solvency of the bank, the economic capital for lending activity (Economic Credit Capital-ECC) is a guarantee against credit risks, such as bankruptcy counterparty rating of its deterioration, the development's credit spreads. Economic capital is used only to cover unexpected losses to a degree of confidence; expected losses are covered by reserves established for this purpose.

Therefore, in practice, economic capital is estimated as the difference between capital appropriately chosen by confidence interval and estimated expected loss. The main reason for expected losses low levels is that they are already incorporated in price credit risk product (in the spread of interest).

For ratings-based approach based on internal generation, only probability of default is calculated by the bank, the remaining components of risk being provided by the Steering Committee Basel banking institution or by national supervisors. If the IRB advanced approach is used, all four components of risk are calculated by the bank.

Measuring and monitoring the default rates is important form different several points of view. Based on past defaulted data expectations of future delinquency is one of the components that in general explains the level of bank spreads .The part of monitoring of default rate time series connect this with business cycles (Bangla et al,2002) and leads to construct anti cyclical regulations dealing with bank provision or capital (Jimenez and Saurina,2006).And all of this process have as a central part ,the estimation of these probabilities of default which is regulated by the Basel II .Finally the National Banks has the task of monitoring these default rates in order to maintain the financial stability as a supervisory authority.

The first step in a credit scoring model development is to define the default event .In the Basel II Capital Accord, the Basel Committee of Banking Supervision gave a reference definition of the default event And announced that banks should use this regulatory reference definition to estimate their model internal rating based .According to this a default is considered to have occurred with regard to particular obligor when either or both of the two following events taken place:

- The bank considers that the obligor is unlikely to pay
- The obligor is past due more than 90 days on any material credit obligation to the banking group.

Time horizon refers to period over which the default probability is estimated and also as a recommendation the time period is usually one year.

For this research I used the same definition of default as the one recommended by the Basel II Capital Accord (90days default) and as an observation period a rolling window of 1-year.

In order to determine the probability of default I chose to compare different credit scoring techniques on a client portfolio from a bank from Romania.

Taking into consideration that the data sample used contains customers with approval date of the credit between 2006 and 2008 „an observation period have been created for each one. For example if the client has been approved on January 2006 then for one year it have been observed to see it he meets the definition of default if this thing happened then a status of 1 have been recorded, otherwise a status of non-defaulter,0.In this way each client have the same time period of observation and the status represents the same thing over time ,90 days default plus a material threshold (100 euro overdue amount).This threshold is considered in order to avoid to have defaulter with small overdue amount above this value it has been considered that they are relevant for the exposure of default of the bank.

The available variables are split into two different categories: socio-demographical variables and financial information such as “Monthly Income” or “Financial Expenses”. These have proven to be of great importance in defining the profile of a default person. For instance, “Education” represents valuable information whereas persons with a higher degree of education tend to be more responsible. “Industry” is also very relevant especially during times like these affected by financial crisis when some fields (i.e. real estate, commerce, constructions etc) have reached an unemployment rate higher than others. “Marital status” and “Sex” have also shown significance in the rating process. For instance, married men are considered to be better payers than single ones who tend to be less responsible. Financial variables have considerable predictive power. They reveal the capacity of paying monthly instalments taking into consideration the wages of the applicants and their monthly expenses too. A full view on the variables used in this paper is available in the **Error! Reference source not found.**.

The data base consists of 33,321 observations representing private individuals that have been granted a loan between January 2006 and December 2008. Each client has been observed during the first year after the credit approval. Those having more than 90 days past due during observation period have been marked correspondingly as defaulters and have been encoded with *1,* whereas the others coincide with registrations having "Good_bad" *0 (non-defaulters).* So the ratio of default clients reaches the level of 14.81% on our database.

Most of the clients included in this PD estimation process are represented by males (70.12%) having an average age of 36 years. Of all the applicants 69.38% are married and 61.02% have graduated a university. The available data reveals that as industry of operating, public services has significant frequency (33.5%) among clients in our database. Unfortunately, this is one of the most affected fields in Romania as a consequence of the measures taken to confront the effects of the actual financial crisis.

More than half of the granted loans (65.43%) are mortgage loans and in what regards the currency, 46.84% of all approved credits are in CHF, mainly because of the low interest rate. Of all the 33,321 clients, 71.5% have never had any previous relationship with the bank and 13.42% have been clients for less than one year at the moment of approval. The collateral is also an important variable but this information wasn't available and considering the fact that studies[30] showed that loans having collateral leads to lower probabilities of default the fact that this variable is not using it seen as a measure of a conservatism.

The variable "Repayment" is very important especially in the case of those clients that before disposing of this loan have had another consumer credit. Out of these, 16.7% have required warnings in some cases and not surprisingly, most of them (79.5%) have defaulted with this loan too.

A simple statistical analysis for the numeric variables (age, term, income, monthly expenses, interest rate, loan value in RON, payment in RON and IMV1) is available in the Appendix 2

In order to get the best performance from a model the model or some parameters should be tuned .To do this three sample are selected from the available cases :one for building the model, one for choosing the optimal structure and parameters and one for testing the final model. The larger the train data the better the classifier and on the other hand the larger the test data the most accurate is the error rate estimation ,and this is seen as a trade-off

[30] Da Silva,Marins J,Da Neves,Brito G-„The influence of Collateral on Capital Requirements in The Brazilian Financial System: an approach through historical average and logistic regression on probability of default " ,Working Paper 187,June 2009, National Bank of Brazil

between these two requirements. For this research I used a split of 70% for the training sample, 20% for validation sample and 10% for the test sample.

3.4 Variable Selection

Selection of the variables is a very important process considering the fact that hose variables represents the base of model that it is developed .Having a lot of variables regarding the situation of a customer it is necessary to see which are relevant related to explained variable ,the good/bad status of the client.

Hand and Henley[31] in 1997 detailed the pressures on the number of the variables that need to be included in the model and they mentioned three commonly methods used in credit scoring :expert judgment ,stepwise selection and Information Value.

The forward selection first estimates parameters for effects forced into the model ,these effects are the intercept and the first n variables (n by default is zero).After this, the chi-square statistic for each effect not included in the model and verify which one is the largest. At this point the "selection entry " criterion interferes because this value could be set at different levels. If the chi-square values is significant at the selection level then the corresponding effect is added in the model .Once an effect is added to the model is never removed from the model.

The method of selection backward is starting with all variables in the model and after the Wald statistic is calculated then the effect that doesn't meet the significant level from the "selection stay " is removed .Once an effect is removed from the model is never added back.

The stepwise selection is a combination of the two procedures described above and it is starting with a forward selection and then continues with a backward selection in this way a variable could enter and could be removed from the model several times until no further effect can be added to the model or if the effect just enter into the model is the only effect removed in the subsequent backward elimination.

[31] W.E. Henley, D.J. Hand (1997), „Statistical Classification in Customer Credit Scoring",Journal of the Royal Statistical Society. Series A, Vol. 160, Issue 3

The second method I used for selection the variables is Information Value Criterion which calculates how much gain is provided from each variable. The concept is based on calculating the Weights on Evidence (WOE) on each category :

$$WOE_c - \ln\left(\frac{\%Non - defaulters}{\%Defaulters}\right) \tag{73}$$

Where %Defaulters represents the proportion of defaulter from the category calculated over the all clients from that category ,analogue is made for %Non Defaulters.

Information Value per category is calculating based on this formula:

$$IV(c) = (\%Non - defaulters - \%Defaulters) * WOE_c \tag{74}$$

The Information Value of the variable is the sum of Information Value per category (if the variable is categorical then the possible characteristics of that variable are selected ,if the variable is continuous then the categories are made by splitting into several homogenous groups).

$$Information\ Value = \sum_{i=1}^{k} IV(k) \tag{75}$$

,where k is the numbers of category per variable.

According to Hand and Heley (1997) if the value of this indicator is zero then the variable shouldn't be included in the model and as a threshold they recommend 0.1 ,from this value the variable could be entered in the model. Kočenda and Vojtek (2009) [32]analyzed a comparison among models including different variables and even they mention that in banking practice the threshold used is 0.2 they also used 0.1 in selection the variables to enter in the model.

Information Value			
Variable	2006	2007	2008
AGE	0.39398	0.47938	0.44900
BANK_R	0.24589	0.00696	0.05127
CCY	0.01337	0.02158	0.00689
COUNTY_ID	0.00014	0.00124	0.01049

[32] Evžen Kočenda, Martin Vojtek - "Default Predictors and Credit Scoring Models for Retail Banking", CESIFO Working paper, Category 12, December 2009

EDUCATION	1.06506	0.22236	0.20623
EXPENSES	0.78089	0.62239	0.33262
INCOME	0.87698	0.27902	0.13908
INDUSTRY	0.39440	0.49011	0.16557
INTEREST_RATE	0.31112	0.16148	0.12133
LOAN_VALUE	0.67563	0.26445	0.25619
MARITAL_STATUS	0.52518	0.33669	0.52125
PAYMENT	0.59730	0.31969	0.11234
PHONE_ID	0.03745	0.00665	0.04046
PRODUCT_ID	0.13533	0.17437	0.09027
PROFESSION	0.39685	0.07986	0.01145
REPAYMENT	1.18685	1.49617	1.15581
RESIDENCE	0.87919	0.37306	0.72286
SENIORITY	0.17727	0.66712	0.45028
SEX	0.00116	0.00792	0.00299
TERM	0.44065	0.18200	0.26365

*The red colour is for values < 0.1 ,yellow is for values between 0.1 and 0.2 and green otherwise

Table 2-Information Value Results

As it can be observed some variables are not significant in any of the samples analyzed, such as Sex, County_ID, Currency and Phone ID. Other variables such as Profession or Relation with Bank , lost the informational value during time.

Each sample analysis involves a number of different techniques and for each sample, I decided to determine the default probabilities by three techniques: logistic regression, probit regression and neural networks.

Each of these three techniques has two features, thus for first two techniques I have used both variable selection method using stepwise method(Logit/Probit 1) and Information Value criteria(Logit/probit 2). When apply neural networks it is very important to choose its architecture. Studies[33] showed that 3 neurons are the most commonly used and which give the best results. Also activation function used logistic function and for comparison I have decided to use also hyperbolic tangent function.

3.5 Macroeconomic Variables in Credit Scoring

[33] Biancotti,DÁurizio and Polcini(2007)-"A neural network architecture for data editing in the Bank of Italy's business surveys", Bank of Italy

With the advent of the Basel II banking regulation it is just not enough to correctly rank customers according to their default risk but also to have an accurate probability of default for each client as these predicted values are used to determine the minimum capital requirement for the portfolio of the retail sector.

In order to incorporate the changes in economic conditions and to observe the modifications of the quality of the portfolio, variables that catch up the macroeconomic vulnerabilities have been introduced in model.

After numerous empirical analysis found that a great crises can be divided in three categories: banking, debt and foreign currency. But this is a robust classification, as events have shown that there isn't a pure type of crisis. Chang and Velasco (1998,1999,2004) show that a banking crisis may turn meet expenses, they called the "twin crisis". In 1996, Frankel and Rose define currency crisis as that situation where the exchange rate recorded a nominal depreciation of at least 25% over a year and its dynamic impairment progresses at least 10 percentage points in the same period of time.

Therefore based on empirical analysis of the crisis has appeared different defining kinds of crisis, based on which some indices have been developed to detect such events. In 1994, Eichengreen[34], Rose and Wyplosz formulated based on empirical analysis carried out on crisis in 22 countries between 1967 to 1992, an index of speculative pressure quantification.

In 1999 Herrera and Garcia[35] proposed a different approach for defining speculative pressure. This index assumes that when the modification of exchange rate and interest rate is over the modification of currency reserves then a speculative pressure exist:

$$ISP_{HG} = \Delta\%ER + \Delta\%IR - \Delta\%Res \tag{76}$$

,where ΔER is the exchange rate variation, ΔIR is the interest rate variation and ΔRes is the currency reserve variation .

---

[34] Eichengreen, Barry, Andrew K. Rose şi Charles Wyplosz, 1994, „Speculative Attacks on Pegged Exchange Rates: An Empirical Exploration with Special Reference to the European Monetary System", NBER WP 4898

[35] Herrera, Santiago şi Conrado Garcia, 1999, „A user's Guide to an Early Warning System of Macroeconomic Vulnerability for Lac Countries", XVII Latin American Meeting for Econometric Society

Also in 1999 the same authors proposed a macroeconomic vulnerability indicator that could be able to detect and extract signal for a currency crisis:

$$IMV = \frac{M2}{Res} + \Delta\%CN + REER + I \qquad (77)$$

,where

$\frac{M2}{Res}$ - M2 reported on currency reserves

$\Delta\%CN$-non-government credit

$REER$-real effective exchange rate

I=inflation

Carling et al (2002) estimate a duration model to explain the survival time to default for borrowers in the business loan portfolio of a major Swedish bank over the period 1994-2000 and as a significant variables the obtain output gap and the yield curve. Virolainen(2004),using Finnish data over seven years starting with 1986 finds a significant relationship between corporate default rates and macroeconomic factors including GDP ,interest rate and corporate indebtedness.

Based on these studies a macroeconomic vulnerability indicator for debt pressure for population segment could be calculating as follows:

$$IMV_G = \Delta\%UR + \Delta\%IR - \Delta\%IPI + \Delta ER - \Delta\%BET + \Delta\%CPI \qquad (78)$$

Where

UR –unemployment rate

IR-reference interest rate

IPI=index of industrial production

CS=exchange rate

BET=Stock Market Index

CPI=consumer price index.

This indicator is calculated on monthly data from January 2006-December 2009 and captures the macroeconomic pressures in several sectors. To find a composition as homogeneous as this indicator and how better to capture the evolution rate arrears population segment I made a comparison between various scenarios to conclude that the weights used are those which includes as much information about the situation economic.

$$IMV_{client} = IMV_{G} \cdot DTI \cdot Spread \tag{79}$$

Where

$$DTI = \frac{Monthly\ Payment}{Income - Expenses} \tag{80}$$

$$Spread = Interest\ Rate - Benchmark\ Rate \tag{81}$$

Macroeconomic vulnerability index is transmitted to each client differently because each has a different capacity to respond to such pressure. Because the loan was granted after an assessment of the extent or debt, this indicator is the multiplier effect of this pressure. Higher degree of indebtedness leads to a lower repayment capacity and this implies a high probability of default.

If this effect is added to this first general economic pressure given then the probability of default increases. This is amplified even more if the interest rate at which the client took the credit is higher compared to a benchmark. The higher the spread is then the its capacity of repayment decreases and this scenario where the debt to income and the pressure is high at the macroeconomic level it is just the worst case possible for the bank.

4.Empirical Results

4.1 Comparison of the models in a multiyear analysis.

In 2006 ,at national level[36] the overdue ratio for retail clients has decreased reaching the level of 0.37% in December. However, the number of bad payers among private

[36] Financial Stability Report-2006-National Bank of Romania ,www.bnro.ro

individuals has increased in 2006 in comparison with the previous year, exceeding 300,000 people and the gross increase in the in-balance debt has also increased by 49%. Many of the high value loans have been granted to people with ages between 30 and 40 years, mainly because their monthly income is higher than in the case of people with other ages.

At the end of the following year (in 2007) the overdue ratio has surpassed 0.5%, reaching 0.59% in February 2008. An alarming fact is the increase by 130% in the amount of arrears in February 2007- February 2008, significantly higher than the increase in the amount of loans granted to private individuals.

By February 2009, the overdue ratio has doubled, becoming 1.42%. During the same time period, the overdue amount has tripled and the number of people with loans greater than 20,000 RON and with arrears increased significantly, by 87%. This overdue comes mainly from people with monthly incomes under 1,500 RON (80% of the total no. of arrears). Thus, a decrease in the level of wages would have a significant impact on the capacity of paying of individuals

Logistic Regression –First Method

The selected model, based on the CHOOSE=AIC criterion,for the 2006 sample, is the model trained in Step 11. Null hypothesis, that is why all the parameters are null is rejected because the value of the test indicate a p-value less than 0.001(Table 3).

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates	Chi-Square		
1234.699	578.159	656.5398	32	<.0001

Table 3-2006-LR Test Logistic Regression(1)

For 2006 sample results indicate that financial variable such as I*ncome, Expenses* or *monthly rate* are significant at 1%. Estimated *Income* coefficient is -0.00237, which indicates that if this increases the probability of default decreases. For variable *Expenses*, the coefficient is 0.0061 an is positive due to direct relationship between it and the probability of default. A positive coefficient is estimated also for *interest rate* variable which is explained by the fact that when interest rate increases then the capacity of repay is decreasing and this leads to a higher probability of default. The socio-demographic variables and the other results are detailed in Appendix 4

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-3.0700	1.6965	3.2700	0.0704		0.0460
Expenses	1	0.0066	0.0008	66.9300	<.0001	1.4642	1.0070
Income	1	-0.0024	0.0004	36.4400	<.0001	-2.2838	0.9980
Interest_rate	1	0.1465	0.0568	6.6500	0.0099	0.1648	1.1580
Loan_Value	1	0.0000	0.0000	6.3200	0.0119	0.3965	1.0000
Payment	1	0.0030	0.0006	24.4200	<.0001	0.8307	1.0030

Table 4-2006-Logistic Regression Output(1)

The selected mode for the 2007 sample is the model trained in Step 12 and it consists of the following effects: Intercept Age Bank_r Education ,Expenses, Income, Industry, loan value, Marital Status, Profession, Repayment, Residence and Seniority.

Parameters estimated for samples of 2007 shows that the variable *Age* is negatively correlated with default probability , having a coefficient of -0.0299. Variables *Interest rate* and *Payment* are excluded from the model because do not comply with the stay value from the stepwise selection. Also variable *Loan Value* is significant at 1% and the coefficient is positive because if the loan increases then a higher leads to a greater indebtedness. For the other details the results are in Appendix 16

For the 2008 sample in the step 15 the model has been chosen based on AIC criterion and the following effects have been entered in the model: Intercept Age CCY Education Expenses Income Industry Interest rate loan value Marital Status Payment Product_id Profession Repayment Residence Seniority.

Variable *Age* is significant for 2008 sample, having a negative factor as well as income with a coefficient of -0.00047. Positively correlated with the probability of default are *Interest rate* , *Monthly rate* and *Expenses* each of them having negative coefficients, the rest of the parameters are detailed in Appendix 28

For 2006 sample according to Information Value criterion variables introduced in the model are: EDUCATION,EXPENSES,INCOME, NDUSTRY,INTEREST_RATE,LOAN_VALUE_RON, MARITAL_STATUS,PAYMENT_RON,PRODUCT_ID,PROFESSION,REPAYMENT,RESIDENCE,TERM and SENIORITY.

The null hypothesis of the Likelihood Ratio test[37] is rejected this meaning that none of the parameters are equal to zero.

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	-8.5738	2.1930	15.2900	<.0001		0.0000
Age	1	-0.0240	0.0190	1.5800	0.2082	-0.1210	0.9760
Expenses	1	0.0062	0.0008	60.6900	<.0001	1.3770	1.0060
Income	1	-0.0022	0.0004	31.0400	<.0001	-2.1109	0.9980
Interest_rate	1	0.3224	0.0911	12.5400	0.0004	0.3629	1.3810
Loan_Value	1	0.0000	0.0000	0.1200	0.7344	-0.0986	1.0000
Payment	1	0.0039	0.0010	13.6700	0.0002	1.0640	1.0040
Term	1	0.0001	0.0001	1.6300	0.2011	0.1649	1.0000

Table 5-2006-Logistic Regression Output(2)

For the second method of variable selection variables *Term, Age* and *Loan Value* did not meet the conditions of being significant ,not even at 10% confidence level. *Payment ,Interest Rate* and *Expenses* have a positive coefficient estimated and this indicates that for every increase in this variable increases the probability of default of the borrower(Table 2)

For 2007 Sample considering the Information Value Criterion (Table 2) I introduced in the model the variables with IV larger than 0.1 and the results obtained indicates that for instance, comparing with 2006, variable *Age* is significant and negative correlated with the default probability of a client considering the fact that if a person is getting old then its income should increase and be more responsible and the results are shown in his capacity of repayment of the credit. In the same manner as for the 2006 sample the variables regarding his capacity of repayment are significant (*Expenses, Income* ,and *Payment*).What is very interesting is that variable *Loan Value* is also not significant but the *Interest Rate* variable is only at 5% level of confidence(Appendix 18-2007-Logistic Regression Output(2)Appendix 18)

[37] More details are shown in Annex:

After applying the Information Value criterion(Table 2) the results for 2008 sample indicates that variable Term is indicative for finding the probability of default and related coefficient of -0.00009 explained by the fact that increasing of term the monthly repayment decreases and its ability to pay increases. Also long term loans are mortgages that have a lower risk compared to the consumer due to collateralized process(Appendix 30)

Probit Regression-First Method

The selected model for 2006 sample, based on the CHOOSE=AIC criterion, is the model trained in Step 12 and it consists of the following effects:Intercept, Age, Education, Expenses, Income, Industry .Interest_rate, Marital_Status, Payment_ron, Profession, Repayment,Residence,Term.

Compared with logistic regression variable *Term* is significant at a confidence level of 5%. Probability of default of the customer is directly proportional with the variables: *Expenses*, *Interest Rate* and *Payment* and negative correlated to *Income* (Appendix 9)

For 2007 sample the results obtained indicate that *Loan Value* is significant at 10% confidence level and the negative coefficient evidence the fact that the higher value of the loan is specific to mortgage loans and due to lower interest rates on long term decreases the probability of default also the p-value of 0.0185 for the term coefficient point out he significance at 5% level of the variable *Term*.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-4.2889	21.0212	0.0400	0.8383	
Age	1	-0.0159	0.0039	16.7800	<.0001	-0.1397
Expenses	1	0.0009	0.0001	284.8100	<.0001	0.6396
Income	1	-0.0004	0.0000	193.1500	<.0001	-1.1973
Interest_rate	1	0.1643	0.0420	15.3200	<.0001	0.3042
Loan Value	1	-0.0000	0.0000	3.6100	0.0573	-0.1398
Payment	1	0.0008	0.0001	53.9200	<.0001	0.5929
Term	1	0.0000	0.0000	5.5500	0.0185	0.1113

Table 6-2007-Probit Regression Estimates(1)

Interest Rate with a estimated coefficient of 0.1963 is significant at 1% confidence level the same as *Income* ,*Expenses* and *Loan Value* for 2008 sample, and at 5 % confidence level are significant *Age* and *Payment* (Appendix).The stepwise selection in comparison with 2007 sample excluded from the model the variable *Term* ,for the rest of the variables that entered in the model more details are presented in Appendix 33

Probit regression 2nd Method

According to Information Value criterion the variables introduced in the model are the same used for Logistic Regression for the 2006 sample. The results indicates that the variable *Loan Value* is not significant for the model and also *Term* variable is only at 10% confidence level significant. In the same frame the output explains the economic relation between Income and Expenses and the probability that the client defaults Appendix 11

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-1.3171	0.4531	8.4500	0.0037	
Age	1	-0.0155	0.0038	16.7100	<.0001	-0.1357
Expenses	1	0.0009	0.0001	281.2100	<.0001	0.6277
Income	1	-0.0005	0.0000	202.7600	<.0001	-1.2179
Interest_rate	1	0.0421	0.0224	3.5200	0.0607	0.0780
Loan Value	1	0.0000	0.0000	0.0100	0.9248	-0.0118
Payment	1	0.0006	0.0002	14.0200	0.0002	0.4743
Term	1	0.0000	0.0000	1.4300	0.2312	0.0608

Table 7-2007-Probit Regression Estimates(2)

In the 2007 sample model the variables were the same as for the 2006 with the xception of Profession variable. The results indicates that neither Term or Loan Value are significant at all and for Interest Rate the confidence level is 10%-Table 7

The fact that the Information Value for the variable Term is increasing in 2008 in comparison with 2007 it is revealed also by the significance at 5% confidence level the opposite process happened with the Payment variable that the decreasing of Information Value makes the level of confidence to be 5% in 2008 versus 2007.

Neural Networks

West(2000) in his paper made a comparison between neural networks and other techniques and as an activation function he uses hyperbolic tangent .Bart(2002) recommends using the logistic function as the activation function.

In order to see the difference between them I compared the results and although the Misclassification Rate, for the first activation function, is 0.0209 for test sample the error increases at 0.04 the number of wrong classified clients is 20.Logistic function used in the second architecture improves the efficiency of the model considering the value of information criterion AIC. Regarding test sample misclassification error is smaller than the first model and the number of wrong classified customers is decreasing with 3.For the 2007 sample the results of AIC and BIC pointed out the neural network using logistic function as fitting better the data. From the error of misclassification point of view both validation and test error are smaller for this architecture and the number of wrong classified is decreasing with 5 on the test sample.

As it can be observed in Table 8 the informational criterion AIC and BIC indicate that the model that fits better is the one with the logistic type as the activation function. Although the number of wrong classification is greater in case of this architecture the Average Error Function is lower and for Validation Sample the Misclassification Rate is lower for this model than for the one with the activation function with hyperbolic tangent.

Neural Networks	Tanh	Logistic
Train: Akaike's Information Criterion	4121.68000	3953.43000
Train: Schwarz's Bayesian Criterion	4992.89000	4824.64000
Train: Average Error Function	0.19598	0.18748
Train: Error Function	3879.68000	3711.43000
Train: Misclassification Rate	0.07234	0.07254
Train: Number of Wrong Classifications	716.00000	718.00000
Valid: Average Error Function	0.21115	0.20798
Valid: Error Function	1194.24000	1176.34000
Valid: Mean Squared Error	0.05944	0.05800
Valid: Misclassification Rate	0.07284	0.07178
Valid: Number of Wrong Classifications	206	203
Test: Average Error Function	0.20589	0.19242
Test: Error Function	582.25500	544.16500
Test: Mean of Squared Error	0.05774	0.05252
Test: Misclassification Rate	0.07497	0.07143
Test: Number of Wrong Classifications	106	101

Table 8-2007-Neural Networks Results

Goodness of Fit

In order to evaluate these models I have performed tests that lead to a model that maps the best outcomes from actual data. Considering that the sample is split tests were performed on all three samples but the final decision was taken based on the test results.

Results "in -sample" are the one from the training part, where the parameters were estimated ,and AUROC test ,for 2006,indicates that the neural network model is closest to the perfect model. In terms of prediction error (Brier score) all same type of model is the better one. Considering that one of the main reasons for achieving these models is the detection of the defaulters, this rate have the highest value for the model that uses neural networks with logistic activation function type.(Appendix 13)

The results of the test sample shows that the model discriminates best non-defaulters from defaulters customers, through the KS test, is the model of neural networks with logistic activation function. Regarding the Brier Score, minimum error is for the same model but which the stepwise selection logistic regression error is smaller (0.0334) than for neural network having as activation function the hyperbolic tangent.

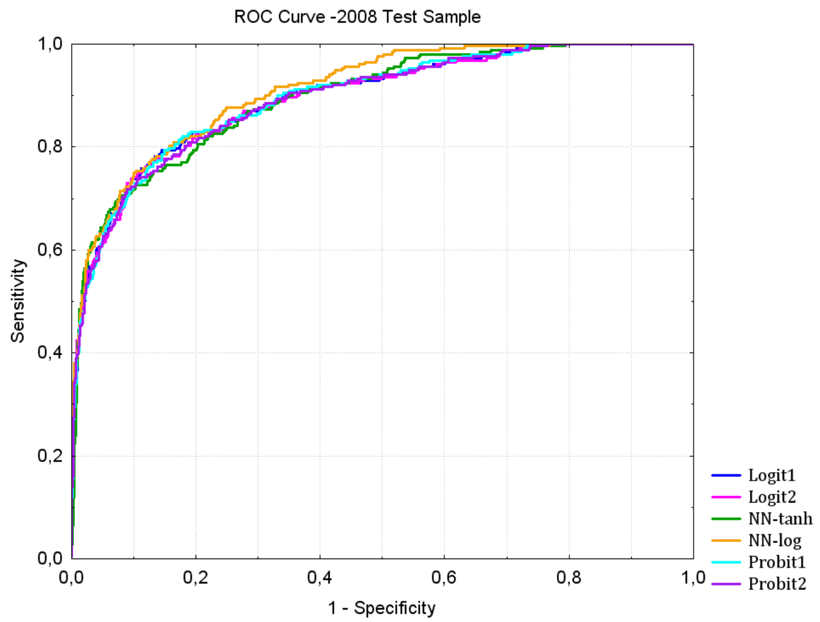


Figure 4-ROC Curve 2008 Sample

In the graph above are plotted ROC Curves for the six models analyzed for 2008 Test Sample and as it can be observed and sustained by ,AUROC indicator, the second architecture of neural network is the mist suitable for this data.

Gini coefficient, or Accuracy Ratio has a value of 0.858 on test data from 2007 for the logistics function neural network, this model all the other validation criteria. What differs from 2006 is that the detection of bad customers in the stepwise probit regression is better than on any type of logistic regression.

For the test sample from the 2008 indicators pointed out that the model that uses a neural network with the logistic function is better than the rest of the models in terms of defaulters detection, discrimination between them and non-defaulters (KS = 0.6499) and in what concerns the prediction error on each client, Brier Score, achieve minimum to this model (0.0754).

To be noted that the defaulters detection accuracy is 0.5952 when logistic regression with the selection criterion variables Information Value is used and this value is equal to the neural network model using hyperbolic tangential function.

Out of sample and out-of time estimation.

To validate a model it must meet certain minimum conditions in terms of error on the test samples (out of sample) and the scale of time. Most variables, the socio-demographic changes during the one year (the period of estimated probabilities of default) but this change is a slower than that in case of financial variables (income, expenses).

In order to observe this error on test sample out of time of the model estimated I apply the best model from 2006 sample on 2007 test data and the best model from 2007 I tested on 2008 data .Like I presented in the Goodness of Fit Test Section for the each sample in part the model that has the most accurate results is the Neural Network using logistic function for activating the nodes.

		Confusion Matrix				Goodness of Fit						
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Misclass Rate	KS	AUROC	AR	Brier Score
NN2_2008	test	1199	95	157	50	0.6230	0.9600	0.0966	0.6499	0.9104	0.8208	0.0754
NN2_07_08	test	1230	162	90	19	0.3571	0.9848	0.1206	0.5679	0.8550	0.7100	0.0978
NN2_2007	test	1203	74	110	27	0.5978	0.9780	0.0714	0.7415	0.9290	0.8580	0.0525
NN2_06_07	test	1180	134	50	50	0.2717	0.9593	0.1301	0.4618	0.7945	0.5891	0.1096

Table 9-Out of time /sample Results

As it can be observed in Table 9,Area under the Curve, for the out-of time estimation for 2007 is 0.7945 and considering that the same indicator for 2006 sample for test data was 0.9491 and the misclassification rate is 13%.

For the other out-of time analysis applied on 2008 data the results are more closely ,for instance Accuracy Ratio is 0.71 and for 2008 data the value is 0.8208 .Considering the fact that applying the same model in the same period 2007 the area under the curve is improving with only 9% relative difference .

This approach is a recommendation from the Basel II Validation Guide and the fact that a model applied on a different test data on a different scale of time it only sustains its robustness and the fact that the models accuracy are higher is not only due to data variables and connection and also its model itself.

4.2 Portfolio analysis.

In the second logistic regression variable portfolio *Interest Rate* is not significant but in this included in the *Term* variable model and has a coefficient (-0.0006) p-value less than 0001, explaining the relationship between long-term loans, generally mortgages and default probability. Variable *loan value* is significant in both logistic regression and 5% *Income* and *Expenses* variables have coefficients very close as values for the two models.

For the probit regression the results indicated also that Interest Rate for the second regression, the one with variable selection based on Information Value ,is not significant at any level and also the variable term is included in the model. For the stepwise probit regression all financial variables are significant at 1% confidence level(Appendix 46)

Neural Networks	Tanh	Logistic
Train: Akaike's Information Criterion	10510.42000	10046.78000
Train: Schwarz's Bayesian Criterion	11678.73000	11215.09000
Train: Average Error Function	0.21909	0.20915
Train: Error Function	10220.42000	9756.78000
Train: Misclassification Rate	0.07893	0.07631
Train: Number of Wrong Classifications	1841	1780
Valid: Average Error Function	0.22910	0.22547
Valid: Error Function	3053.48000	3005.10000
Valid: Mean Squared Error	0.06530	0.06478

Valid: Misclassification Rate	0.08178	0.08373
Valid: Number of Wrong Classifications	545	558
Test: Average Error Function	0.23237	0.23123
Test: Error Function	1548.52000	1540.89000
Test: Mean of Squared Error	0.06500	0.06497
Test: Misclassification Rate	0.07743	0.08103
Test: Number of Wrong Classifications	258	270

Table 10-Portfolio-Neural Network

The misclassification Rate for test data ,7.7743% for neural network with hyperbolic tangent function is lower comparing with the other neural network although on validation sample the situation is inverted. The model that best explain the data ,having the lowest AIC value is the neural network with logistic function.

Goodness of Fit Tests

Tests "in sample" portfolio confirms the second ANN architecture has the best results. The value of the KS distance of 0.6815 for the first neural network is ranked as the second model as validation. In connection with the prediction accuracy, Brier Score, reaches its peak on Probit regression, with Information Value as the criterion for selection the variables and for this model, Gini coefficient, chive its minimum of 0.8187,compared for example with stepwise logistic regression, 0.8229.

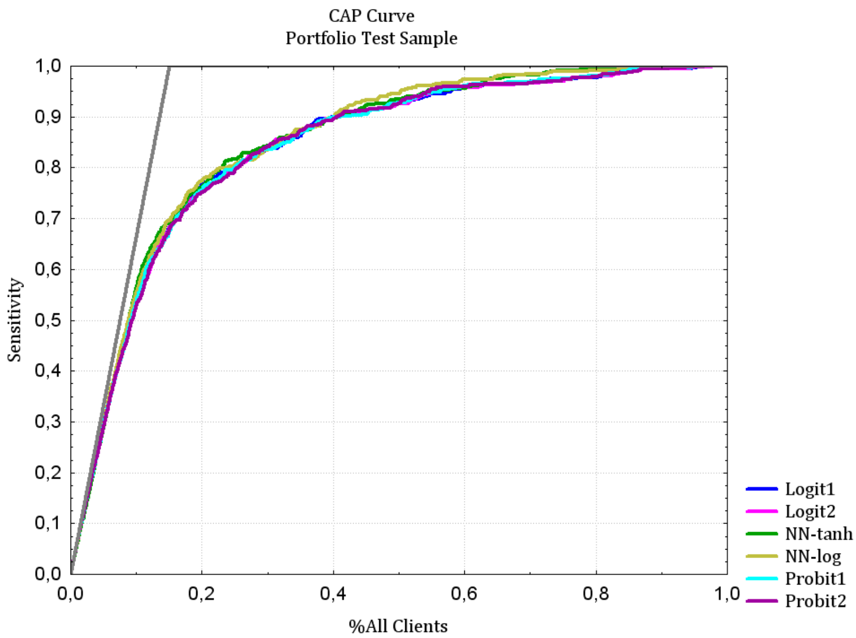


Figure 5-Portfolio-Test Sample-CAP Curve

The fact that the results of the out-of samples indicate the same pattern as at best, can only confirm the consistency of analysis. However in this case, detection accuracy of the defaulters is higher for neural network model with the hyperbolic tangent function, which is confirmed by results presented above, when I mentioned the increasing number of wrong classified customers if it is using the logistics function for neural network.

Cumulative Accuracy Profile graph for wich i have calculated the area under this curve ,this being the Accuracy Ratio points aut that the two curves og=f the that neural networks are well above the curves of the other models, especially over probit regression who has the smallest value of this indicator 0.8061 compared with the second neural network, 0.8314 and compared to the perfect model that reaches the value of 1.

4.3 Portfolio with macroeconomic variable.

The implementation of this macroeconomic variable is made on the same samples of portfolio and the model is estimated on training sample with regarding of validation test on the test sample.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-3.6365	1.4412	6.3700	0.0116	
Age	1	-0.0184	0.0044	17.8900	<.0001	-0.0914
Expenses	1	0.0018	0.0001	552.9500	<.0001	0.6611
Income	1	-0.0007	0.0000	387.6800	<.0001	-1.3555
Interest_rate	1	0.1146	0.0279	16.8800	<.0001	0.1417
Loan Value	1	0.0000	0.0000	27.7500	<.0001	-0.3528
Payment	1	0.0023	0.0001	250.6900	<.0001	1.2525
IMV_customer	1	5.1807	0.3196	262.7200	<.0001	0.2641

Table 11-Portfolio Macro Stepwise Logistic Regression

The model with macroeconomic variable included leads to an improvement of the significance of some variables such as *loan value* and *Interest Rate*. Also another deduction is that *Income* and *Expenses* have smaller values of coefficients in this model a part of their importance being transfer to the new variable *IMV_customer* also significant at 1% confidence level. For the stepwise probit regression the financial variables are all significant at 1% and in both models the *Term* variable is not selected to be part of the estimation.

Neural Networks	Tanh	Logistic

Train: Akaike's Information Criterion	9381.67000	9337.05000
Train: Schwarz's Bayesian Criterion	10574.14000	10529.53000
Train: Average Error Function	0.19476	0.19381
Train: Error Function	9085.67000	9041.05000
Train: Misclassification Rate	0.07374	0.07117
Train: Number of Wrong Classifications	1720.00000	1660.00000
Valid: Average Error Function	0.21533	0.21072
Valid: Error Function	2869.89000	2808.47000
Valid: Mean Squared Error	0.06174	0.06047
Valid: Misclassification Rate	0.08013	0.07773
Valid: Number of Wrong Classifications	534	518
Test: Average Error Function	0.21861	0.20920
Test: Error Function	1456.78000	1394.12000
Test: Mean of Squared Error	0.06138	0.05939
Test: Misclassification Rate	0.07713	0.07593
Test: Number of Wrong Classifications	257	253

Table 12-Portfolio Macro -Neural Network

Difference in the number of training cases wrong classified is 60 in favour of neural network that uses logistic function as activation function and also for the validation sample the detection error is smaller for this model. Among 3332 of clients 253 were wrong classified on test sample for the model mentioned above and considering the fact that for the portfolio analysis the situation was reversed and for the first method the number was 258 and the for the second method the same number was 270 ,it can be observed the major improvement brought by the macroeconomic variable.

Goodness of Fit Tests:

Testing "out of sample" results for portfolio with macroeconomic variables incorporated maintain the idea of the performance for neural networks models over the rest models. In this manner to see the improved results for these models compared with the portfolio, I made an analysis showing relative changes of each indicator in order to move from the global analysis of the overall portfolio, with and without this new variable, to each indicator and the improvement in every technique I used for modelling.(Appendix 63)

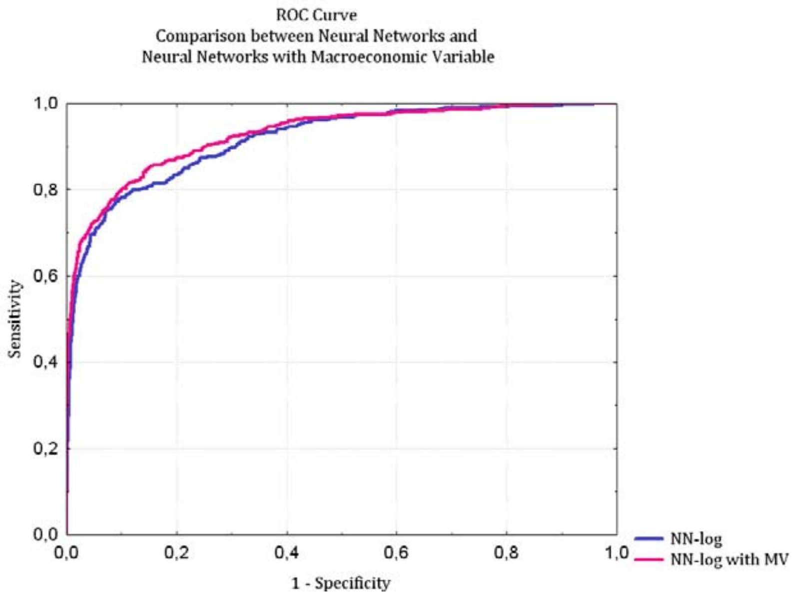


Figure 6-Comparison between Neural Networks models

Detection accuracy of defaulters customers increases on average with 5.85% for probit regressions and with 3% for logistic regressions, neural networks instead recorded an increase of only 1.13%. Regarding the prediction error for each customer, on average decreases 4% for the two types of regression and the improvement gave by the neural networks is 7.5% .Accuracy Ratio improves with 1.73% for the neural network with hyperbolic tangent function while logistic regressions bring increase in average equal to 1%.

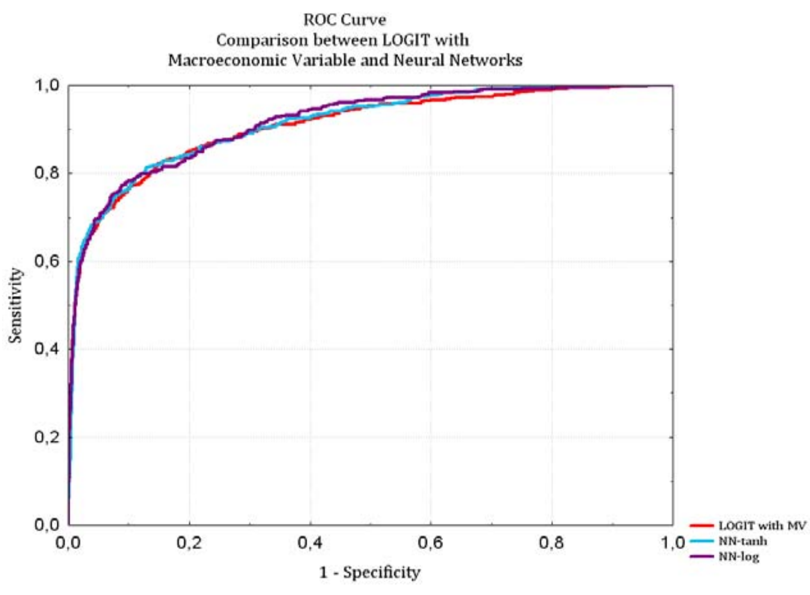


Figure 7-Comparison among logistic regression and Neural Networks

As it can be seen in the graph above the regression logistic with macroeconomic variable incorporated is comparable with neural networks and on some fractions the report between 1-error type 1(hit rate) and error type II (false alarm rate) is greater than for the neural networks using hyperbolic tangent function.

The Spiegelhalter Test indicates that, by accepting the null hypothesis on both portfolios with and without the macroeconomic variable, the observed default rates are close to the estimated probabilities of default(Appendix 68)

Dynamic Cut off portfolio with macroeconomic variables

When using the cut-off or 0.5 in order to classify customers a disequilibrium could increase costs for defaulter detection. To solve this problem a new cut-off should be used and in order to find the theoretical cut-off the intersection of the corrected rate for discriminating the non defaulters and defaulters is offering the desired result. When this process is made the differences between Sensitivity and Specificity should be zero because their point of intersection is the new value of the cut-off.

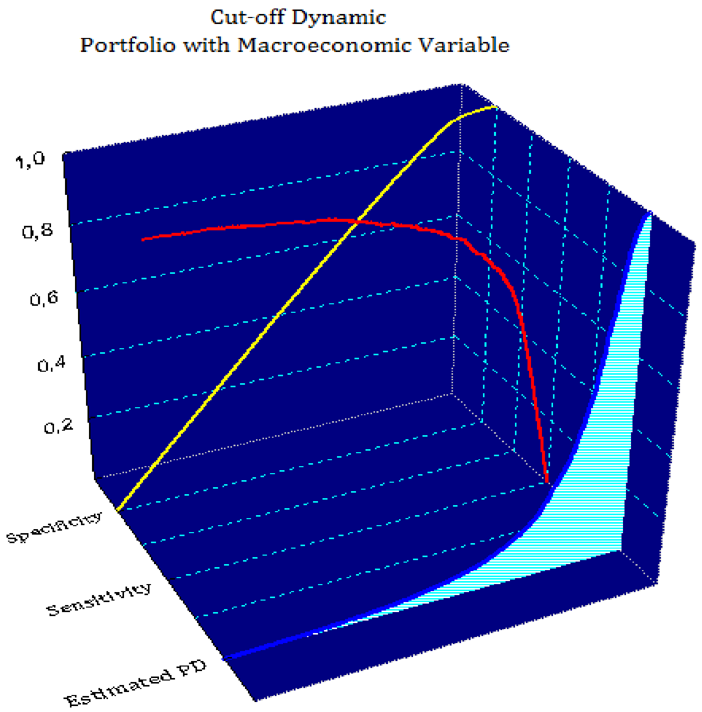


Figure 8-Portfolio Macro Cut-Off Dynamic

The results of the portfolio with macroeconomic variable incorporated are recalculated with the new values of cut-off. The value of cut-off is determined on the training sample and applied then on the default probabilities obtained from the modes described in this paper. Although the optimal threshold is then applied on the test sample and this could affect the performance of the models but the predictions made on test samples are completely independent on the training samples which will be also seen as in practice where credit decision and management is involved in setting this cut-off. Many factors are involved when setting this threshold but the one that affect the portfolio of a bank is the fact that a lower value of this cut off will be translated as an acceptance rate higher which it turn into profitability for the bank if the clients accepted wouldn't default in a larger proportion.

This issue of acceptance rate is a trade-off between the higher acceptance rate as profit generator and lower acceptance rate as loss in market share is well weighted in a bank strategic decisions .The results on test sample indicates the fact that for the new cut-off the detection of defaulter clients has been increased while the detection on non-defaulter clients has been decreased. Even if on training sample the difference between those two is on average equal to 0.001 the same average on the test results is -0.0079 meaning that using the cut-off from training sample to test sample isn't a mismeasurement.

4.3 Misclassification Cost

The model that minimizes the expected future loss is an optimal model of classification and considering the fact that there are two classes of customers the future loss depends on the two types of misclassification errors.

$$Expected = P_B Cost_{Type\,I} Error\ Rate\ Type\ I + P_G Cost_{TypeI\,I} Error\ Rate\ Type\ II$$

Where P_B and P_G are the population percentage of defaulters and non-defaulters clients and $Cost_{Type\,I}$ is the cost of error type I respectively cost of error type II. The choice of these two costs has a major impact on the evaluation of the model and the factor that affect the costs are difficult to be quantified. The Error of Type I is the cost of granting a loan to a customer that is defaulter and the Error of Type II is the opportunity profit of rejecting a non-defaulter client considered as bad. In this way for the first type of error the costs are related to loss of principal money and other costs that interfere in the process of recovery. For the type II error the lost of interest paid by de client and the profit obtained from his loan is the virtual earn missed by the bank.

Having clarified this idea is understandable that potential loss is more expensive than lost profit, in this way cost values should be differentiate. In this paper I took into consideration a proportion of the cost of 5.10 and 15 this being the multiplier for Type I error compared with type II error.These selected values for cost are sustained by bank policy whose portfolio I used in this research,and this was calculated by incorportaing the cost of risk.

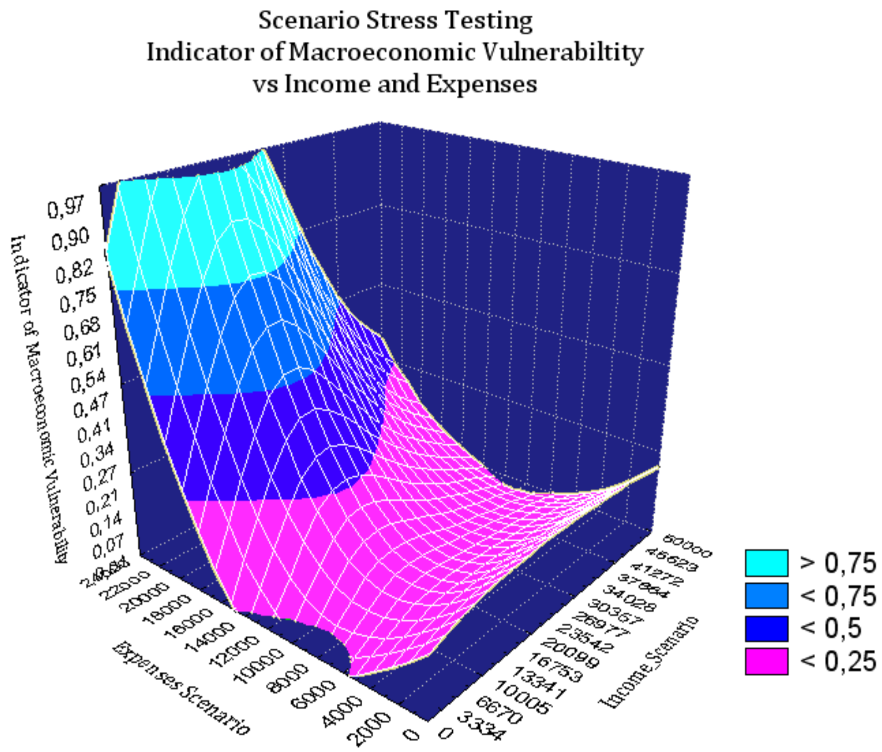
The main idea of this analysis is to see the impact of misclassification cost on the results of the models analyzed. First I explored the results with cost of 5 and by comparison with the initials models for portfolio ,without macroeconomic variable ,on test sample ,the defaulter accuracy ratio has improved most on Probit Regression with Information Value and on the second place is the logistic regression with the same selection type of variables. Even the first type of probit and logistic regression recorded a higher increase on this indicator than other neural network model.Cost of 10 improves probit regression and reduce the error of prediction with 3.3% and neural networks on the same indicator has been improved with 6.8%.

The same thing happened when a cost of 15 has been used ,but the improvement of the stepwise logistic regression on misclassification error ratio is 5.31% and for neural network is 5.55 % but if the latter value is the highest for neural network ,for regressions the higher improvement is recorded on the second type of probit regression with a 8.44% relative modification.

After all three types of scores gave been incorporated into the model analyzed I compared them through *Kuipers Score* and *Granger-Pesaran Test.*As it can be noticed the hypothesis null of classification failure could be rejected at one-percent level of significance for all model for all three types of costs. *Kuipers Score* is the difference between hit rate and false alarm rate and the grater the difference the better the classification between defaulters and non-defaulters is made. The models that have la higher score are neural networks and from regressions class the stepwise logistic is the one that discriminate better.

4.4 Stress Testing

For credit risk modeling the stress testing is based not only on scenario tests but on sensitivity tests and according to Basel II the objective is not to require banks to consider worst-case scenarios but to capture the different behaviors and mixtures of simulations in order to create a real scenario possibility. Studies on credit risk stress testing come with three or four scenarios ,one is the baseline and the other are related to decreasing of GDP ,a rise of real interest rate and a reduction in real property prices by different values.



Considering the related to the loan agreement signed with the IMF the reduction plan is to cutt wages with the following percentages 25% and those Cuts will come into effect starting June 1st.

The scenario that I considered it is based on stressing the income of customers having „Public Service" as industry .The assumptions are that the their income are decreasing with 25% and for the rest of the portfolio this variables remain the same. Regarding expenses I proposed to capture a raise in inflation, that will be translated to an increase in the level of expenses and considering that the target inflation is 3.5% plus 1% error band I stressed the values of the variable Expenses with an increase with 4.5%

After recalculating the probabilities of default the results concluded that this scenario impact the losses with an increase of 0.25% on the entire portfolio and only on the public employers the impact is 1% on the average probability of default on both models (stepwise logistic regression and logistic neural networks).On the graph below it can be observed the differences between the estimated probability of default on original data and on stressed data , on the selected portfolio public employers, on test sample .

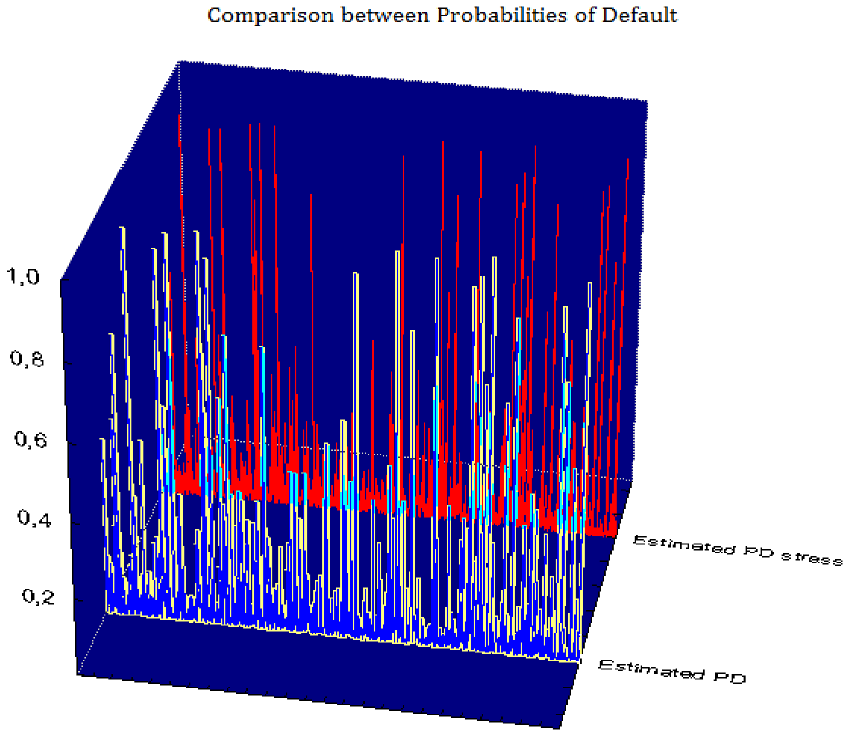


Figure 9-Portfolio Macro Stress Testing

5.Conclusions

In this paper I have highlighted both the comparison of several models of credit scoring and improvement necessities that have to be done for such models. The initial idea I had was a transposition of several models on both scale of time and on the unit, meaning model estimation and data analysis from different years and on different customers. The conclusion of applying these models on different years was that, certain financial variables, like *Income* or E*xpenses* are significant regardless of the chosen time axis. What is important to emphasize here is that although socio-demographic variables, during one year, tend to have a rate of change smaller than for financial variables and there exist some connections such as variable Seniority, age of work at previous job, was included in almost all models and the explanation being that this variable is the bridge between financial and non-financial variables. If the seniority is higher then there are two options, have an adequately income and wants stability, or there is a higher probability that in the next period will want to change their workplace but automatically income will be raised or remains the same which is negatively correlated with probability of default given. This analysis captures both the evolution of the three periods of varying models importance and their importance in itself and an important aspect is that usually a multi-year analysis incorporates behavioral variables that are meant to hold a more correct image of the client's position just like a compass, so the bank will be able to act in time. Finding that the results on the test samples for out-of sample and out-of-time are quite robust I grouped data at a level of portfolio in order to capture the relationships with the economic environment of the period between 2007 and 2009 in Romania.

Portfolio results supported findings from different years, so that neural networks have a higher accuracy than the regressions. In order to surprise the whole picture and the framework of the portfolio I analyzed in this paper, I proposed a new approach of determining the level of customer macroeconomic impacts. So going from one macro vulnerability indicator proposed in the literature by Herrera and Garcia I calculated an indicator that its designation is to capture the capacity of repayment of a customer and to also to see its future problems with this issue. First I made this indicator at global level on monthly data and the impact will be performed by reporting this to its degree of indebtedness and the spread for the interest rate that has taken credit.

This indicator of macroeconomic vulnerability could be part of a development model for credit risk based on a scorecard where the capacity of a client would be aligned to the period he is taking the loan because if the period is under pressure a small deviation of his behavior will be amplified by the macroeconomic conditions and he will be overdue with his monthly payments and in the end classified as default client.

What is interesting is that once I re-estimated the parameters , results showed that models like logistic regressions have accuracy as high as one of the neural network architectures. To study the impact of each model I computed an improvement ratio to detect which technique is getting improved related to the inclusion of this new variable because some models had high accuracy before, like neural networks. Considering the detection accuracy of default clients the regressions techniques have a much greater improvement than any other model. This detection is very important for bad customers due to their expensiveness in comparison with the non-detection of good clients and in order to explore this area I included in the models a loss function depending on the two types of costs. I analyzed three types of proportions between those costs and the results indicated that cost improvements of a logistic regression or probit type are comparable or even higher than an improvement of the neural networks.

All analysis have been sustained by the minimal error of detection between default realized rate and default predicted rate and by statistics test that confirmed that models have no major differences between distributions of the two probabilities of default.
What I wanted to evidence in this paper is that more important than a particular model is the variable selection and choice of loss function that need to be minimized in order to treat the tradeoff between the profit considerations and best classification of customers.

For the further research I would like to incorporate both behavioral and macroeconomic variable in a survival analysis to detect not only if the customer defaults but when this event happens in order to help a bank to have enough capital when a part of the portfolio is translated from a rating class to other and you don't know when this migration ends because in the end you know it will be a default client.

6.References

Andersen, H(2008),"Failure prediction of Norwegian banks: A logit approach" , Financial Market Department, Norges Bank.

Baesens,B. (2003), “Developing intelligent systems for credit scoring using machine learning techniques” Ph.D. thesis, K.U.Leuven

Balcaen, S. and H. Ooghe, (2004). Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods? Working Paper 2004/249, Faculteit Economie en Bedrijfskunde

Balzarotti V, M. Gutiérrez and G.V. Vallés (2006),"Credit Scoring models with truncated samples and their validation "NationalBank of Argentina

Banasik, J,J.N. Crook, and L.C. Thomas.(1999),”Not if but when will borrowers default” Journal of the Operational Research Society

Banasik, J.N. Crook, and L.C. Thomas (2001),” Sample selection bias in credit scoring models” In Proceedings of the Seventh Conference on Credit Scoring and Credit Control (CSCCVII'2001), Edinburgh, Scotland

Banasik, J., Crook, J., Thomas, L. (2003), "Sample selection bias in credit scoring models", Journal of the Operational Research Society, Vol. 54 No.8

Biancotti C, L. D’Aurizio and R. Polcini (2007),"A neural network architecture for data editing in the Bank of Italy’s business surveys" National Bank of Italy

Bishop, C. (1995), “Neural Networks for Pattern Recognition”, Clarendon Press, Oxford

BCBS, Basel Committee on Banking Supervision, 1988. *International Convergence of Capital Measurement and Capital Standards.* Bank for International Settlements,July 1988

BCBS, Basel Committee on Banking Supervision, 2003. *Press releas: Significant Progress on Major Issues* . Bank for International Settlements, 11 October 2004

BCBS, Basel Committee on Banking Supervision, 2004. *International Convergence of Capital Measurement and Capital Standards. A Revised Framework.* Bank for International Settlements, June 2004

BCBS, Basel Committee on Banking Supervision, 2005. *Studies on the Validation of Internal Rating Systems.* Working paper no. 14, Basel Committee on Banking Supervision, February 2005

Crook,J John Banasik,(2004)," Does reject inference really improve the performance of application scoring models? "Journal of Banking & Finance

Engelmann B and R Rauhmeier(2006),"The Basel II Risk Parameters",Springer

Härdle W,2007,"Estimating probabilities of default with support vector machines ", Deutsche Bank

Kočenda E, Martin Vojtek (2009), "Default Predictors and Credit Scoring Models for Retail Banking", Working paper, Prague

Komorád K.(2002), "On Credit Scoring Estimation" Master's Thesis, Institute for Statistics and Econometrics, Humboldt University, Berlin

Ming-Chun Tsai, Shu-Ping Lin, Ching-Chan Cheng, Yen-Ping Lin (2009), "The consumer loan default predicting model – An application of DEA–DA and neural network", Expert Systems with Applications 36

Oestereichische Nationalbank, 2004. Guidelines on Credit Risk Management, Rating Models and Validation

Rommer, A. D., 2005b. Testing the Assumptions of Credit-scoring Models. WP no.28, Danmarks NationalbankŘezáč,M,F. Řezáč and Masaryk(2008),"Measuring the Quality of Credit Scoring Models", University, Czech Republic

Roszbach(2003)," Bank Lending Policy,Credit Scoring and the Survival of Loans" , SVERIGES RIKSBANK

Silva da R.,„A.C.Magalhães da Silva(2009)-" The Influence of Collateral on Capital Requirements in the Brazilian Financial System: an approach through historical average and logistic regression on probability of default",National Bank of Brazilia

Siddiqi N(2006),"Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring"

Thomas, D.B. Edelman si J.N. Crook (2002), „Credit Scoring and Its Applications",SIAM

West,D.(2000)," Neural network credit scoring models". Computers and Operations Research

7.APPENDIX

Variable	Characteristics	No.	%
Credit_nr	Unique no.	33,321	
Repayment	so far no credit customer	24,001	72.03%
	Yes (In some cases warnings required)	1,556	4.67%
	Yes (never warned or deferred)	7,764	23.30%
Sex	Female	9,955	29.88%
	Male	23,366	70.12%
Age	Continuous variable		
Marital_Status	Divorce	1,111	3.33%
	Married	23,118	69.38%
	Single	8,787	26.37%
	Widowed	305	0.92%
Education	High School	12,536	37.62%
	Primary school	451	1.35%
	University	20,334	61.02%
Profession	Employee	31,031	93.13%
	Own Empl	1,259	3.78%
	Unemploy	511	1.53%
	Worker	520	1.56%
Seniority	< 6 Months	2,116	6.35%
	> 10 Years	9,570	28.72%
	0,5 - 1 Year	3,437	10.31%
	1 - 2 Years	6,953	20.87%
	3 - 5 Years	5,951	17.86%
	6 - 10 Years	5,294	15.89%
Industry	Agriculture	121	0.36%
	Bank and Financial Services	1,990	5.97%
	Construction	1,068	3.21%
	Electronics / Pharmaceutical / Optics	700	2.10%
	Food	259	0.78%
	Gastronomie	127	0.38%
	Leather / Textile / Clothing	210	0.63%
	Other	14,799	44.41%
	Plastic / rubber / asbestos	71	0.21%
	Public Service	11,161	33.50%
	Retail	765	2.30%
	Steel/metal processing	926	2.78%
	Stones / earth / gas / ceramic	730	2.19%
	Wholesale	207	0.62%
	Wood	187	0.56%
Residence	Job apartment	117	0.35%
	Own house	19,622	58.89%

	Rent	1,300	3.90%
	With parents	12,282	36.86%
Income	Continuous variable		
Expenses	Continuous variable		
Good_bad	0 (non-defaulter)	28,386	85.19%
	1 (defaulter)	4,935	14.81%
Bank_r	0 (not a client before)	23,826	71.50%
	< 1 Year	4,471	13.42%
	> 10 Year	52	0.16%
	1 -2 Year	1,124	3.37%
	1 Year	1,766	5.30%
	2 Years	1,097	3.29%
	2-3 Year	410	1.23%
	3-5 Years	535	1.61%
	6-10 Year	40	0.12%
Term (days)	Continuous variable		
CCY	CHF	15,609	46.84%
	EUR	10,789	32.38%
	RON	6,923	20.78%
loan_value_ron	Continuous variable		
Interest_rate	Continuous variable		
Payment_ron	Continuous variable		
Product_id	CAR	3,752	11.26%
	CONSUMER	7,766	23.31%
	MORTGAGE	21,803	65.43%
Phone_id	Fix	14,081	42.26%
	Mobile	18,095	54.31%
	no information	1,145	3.44%
County_ID	0 (Other than Bucharest)	27,539	82.65%
	1 (Bucharest)	5,782	17.35%
IMV	Continuous variable		

Appendix 1-Data Description

	Min	Max	Mean	Stdev
Age	18	68	36	9
Income	501.00	89265.00	3670.02	3292.19
Expenses	20.00	20763.00	387.74	692.96
Term	149.00	13967.00	5872.12	2923.40
loan_value_ron	1200.00	2059675.00	123478.81	143863.88
Interest_rate	3.95	19.75	5.83	2.23
Payment_ron	40.00	17325.00	925.63	979.53
IMV	0.01	0.71	0.16	0.09

Appendix 2-Decriptive Statistics

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	267.4513		<.0001
2	Education		2	2	135.4064		<.0001
3	Expenses		1	3	128.4606		<.0001
4	Residence		3	4	96.2262		<.0001
5	Profession		3	5	38.8362		<.0001
6	Industry		14	6	55.5151		<.0001
7	Marital_Status		3	7	26.8423		<.0001
8	Interest_rate		1	8	11.5598		0.0007
9	Income		1	9	6.4404		0.0112
10	Payment_ron		1	10	94.0618		<.0001
11	loan_value		1	11	6.3744		0.0116
12	Product_id		2	12	3.9878		0.1362
13	Age		1	13	1.6344		0.2011

Appendix 3-2006-Stepwise Seletion Logistic Regression

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-3.0700	1.6965	3.27	0.0704		0.046	-6.6080	0.0852
Education	High School	1	2.1242	0.4035	27.72	<.0001		8.367	1.2910	2.8704
Education	Primary school	1	2.9904	0.6013	24.73	<.0001		19.894	1.7693	4.1585
Education	University	0	0	.	.	.	.	.	1.7693	4.1585
Expenses		1	0.00661	0.000808	66.93	<.0001	1.4642	1.007	0.00508	0.00823
Income		1	-0.00237	0.000393	36.44	<.0001	-2.2838	0.998	-0.00312	-0.00159
Industry	Agriculture	1	-8.4056	277.4	0.00	0.9758		0.000	-1197.5	1178.2
Industry	Bank and Financial Services	1	-0.1805	1.3904	0.02	0.8967		0.835	-2.8162	2.6166
Industry	Construction	1	1.3808	1.2056	1.31	0.2521		3.978	-0.9214	3.8179
Industry	Electronics / Pharmaceutical / O	1	-11.7791	112.5	0.01	0.9166		0.000	-253.1	229.1
Industry	Food	1	2.2275	1.2260	3.30	0.0692		9.277	-0.0371	4.8024
Industry	Gastronomie	1	-8.4999	307.1	0.00	0.9779		0.000	-1288.5	1268.5
Industry	Leather / Textile / Clothing	1	0.5076	1.2887	0.16	0.6937		1.661	-1.9866	3.1107
Industry	Other	1	0.0563	1.0946	0.00	0.9589		1.058	-1.9908	2.3236
Industry	Plastic / rubber / asbestos	1	-0.5594	2.2673	0.06	0.8051		0.572	-5.5384	4.1743
Industry	Public Service	1	0.4324	1.1021	0.15	0.6948		1.541	-1.6255	2.7202
Industry	Retail	1	2.4074	1.2378	3.78	0.0518		11.105	0.1167	4.9871
Industry	Steel/metal processing	1	0.4711	1.3257	0.13	0.7223		1.602	-2.0253	3.1915

Industry	Stones / earth / gas / ceramic	1	-0.6281	1.4762	0.18	0.6705		0.534	-3.4451	2.4301
Industry	Wholesale	1	2.4505	1.1868	4.26	0.0389		11.594	0.1927	4.8692
Industry	Wood	0	0	.	.	.	.	.	0.1927	4.8692
Interest_rate		1	0.1465	0.0568	6.65	0.0099	0.1648	1.158	0.0777	0.3862
loan_value		1	0.000023	9.301E-6	6.32	0.0119	0.3965	1.000	-3.49E-6	0.000037
Marital_Status	Divorce	1	-0.6702	1.0052	0.44	0.5050		0.512	-2.6020	1.3195
Marital_Status	Married	1	-1.0003	0.9174	1.19	0.2756		0.368	-2.7928	0.7889
Marital_Status	Single	1	-0.0577	0.9406	0.00	0.9511		0.944	-1.8780	1.7979
Marital_Status	Widowed	0	0	.	.	.	.	.	-1.8780	1.7979
Payment_ron		1	0.00301	0.000609	24.42	<.0001	0.8307	1.003	0.00194	0.00438
Profession	Employee	1	-1.7780	0.3615	24.19	<.0001		0.169	-2.0336	0.3766
Profession	Own Empl	1	-2.1146	1.2586	2.82	0.0930		0.121	-1.6446	0.0588
Profession	Unemploy	1	2.3898	1.1496	4.32	0.0376		10.911	-2.4808	-1.0576
Profession	Worker	0	0	.	.	.	.	.	-2.4808	-1.0576
Repayment	Yes (In some cases warnings requ	1	2.5853	0.5657	20.88	<.0001		13.268	-4.6655	0.3111
Repayment	Yes (never warned or deferred)	1	-1.3447	0.4341	9.60	0.0020		0.261	0.1350	4.8735
Repayment	so far no credit customer	0	0	.	.	.	.	.	0.1350	4.8735
Residence	Job apartment	1	-10.0146	139.7	0.01	0.9429		0.000	1.4278	3.6824
Residence	Own house	1	-0.7979	0.2945	7.34	0.0067		0.450	-2.2972	-0.5743
Residence	Rent	1	1.0000	0.3461	8.35	0.0039		2.718	-337.9	317.4
Residence	With parents	0	0	.	.	.	.	.	-337.9	317.4

Appendix 4-2006-Logistic Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
1234.699	569.680	665.0185	47	<.0001

Appendix 5-2006-LR test Logistic Regression(2)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-8.5738	2.1930	15.29	<.0001		0.000	-12.8719	-4.2757
Age		1	-0.0240	0.0190	1.58	0.2082	-0.1210	0.976	-0.0613	0.0134
Bank_r	0	1	5.3413	.	.	.	.	208.788	-0.0613	0.0134
Bank_r	1 Year	1	5.0651	1.5712	10.39	0.0013		158.390	1.9856	8.1446
Bank_r	2 Years	1	6.5273	1.5765	17.14	<.0001		683.564	3.4374	9.6172
Bank_r	3-5 Years	1	-1.1695	146.8	0.00	0.9936		0.311	-288.9	286.5
Bank_r	6-10 Year	1	-4.8584	1419.3	0.00	0.9973		0.008	-2786.7	2777.0
Bank_r	< 1 Year	1	5.9566	1.3532	19.38	<.0001		386.282	3.3043	8.6088
Bank_r	> 10 Year	0	0	.	.	.	.	.	3.3043	8.6088
Education	High School	1	2.0417	0.4039	25.56	<.0001		7.703	1.2501	2.8332
Education	Primary school	1	2.9331	0.6048	23.52	<.0001		18.786	1.7477	4.1186
Education	University	0	0	.	.	.	.	.	1.7477	4.1186
Expenses		1	0.00622	0.000798	60.69	<.0001	1.3770	1.006	0.00466	0.00779
Income		1	-0.00219	0.000393	31.04	<.0001	-2.1109	0.998	-0.00296	-0.00142
Industry	Agriculture	1	-6.8117	162.6	0.00	0.9666		0.001	-325.5	311.9
Industry	Bank and Financial Services	1	-0.0717	1.4358	0.00	0.9602		0.931	-2.8859	2.7425
Industry	Construction	1	1.5057	1.2188	1.53	0.2167		4.507	-0.8831	3.8944
Industry	Electronics / Pharmaceutical / O	1	-13.6279	278.0	0.00	0.9609		0.000	-558.4	531.1
Industry	Food	1	2.3621	1.2435	3.61	0.0575		10.613	-0.0752	4.7994
Industry	Gastronomie	1	-7.6245	198.5	0.00	0.9694		0.000	-396.6	381.3
Industry	Leather / Textile / Clothing	1	0.5295	1.3076	0.16	0.6855		1.698	-2.0334	3.0924
Industry	Other	1	0.1481	1.1056	0.02	0.8935		1.160	-2.0188	2.3149
Industry	Plastic / rubber / asbestos	1	-1.0291	2.8359	0.13	0.7167		0.357	-6.5874	4.5291
Industry	Public Service	1	0.5300	1.1142	0.23	0.6343		1.699	-1.6537	2.7137
Industry	Retail	1	2.5229	1.2474	4.09	0.0431		12.465	0.0780	4.9678
Industry	Steel/metal processing	1	0.5168	1.3372	0.15	0.6991		1.677	-2.1040	3.1377
Industry	Stones / earth / gas / ceramic	1	-0.5112	1.5077	0.11	0.7345		0.600	-3.4662	2.4437
Industry	Wholesale	1	2.5843	1.2021	4.62	0.0316		13.254	0.2282	4.9404
Industry	Wood	0	0	.	.	.	.	.	0.2282	4.9404
Interest_rate		1	0.3224	0.0911	12.54	0.0004	0.3629	1.381	0.1440	0.5009
loan_value_ron		1	-2.43E-6	7.172E-6	0.12	0.7344	-0.0986	1.000	-0.00002	0.000012
Marital_Status	Divorce	1	-0.8072	1.0136	0.63	0.4258		0.446	-2.7939	1.1795
Marital_Status	Married	1	-1.2851	0.9455	1.85	0.1741		0.277	-3.1382	0.5679
Marital_Status	Single	1	-0.4274	1.0102	0.18	0.6722		0.652	-2.4073	1.5525
Marital_Status	Widowed	0	0	.	.	.	.	.	-2.4073	1.5525
Payment_ron		1	0.00385	0.00104	13.67	0.0002	1.0640	1.004	0.00181	0.00590
Profession	Employee	1	-1.8099	0.3666	24.37	<.0001		0.164	-2.5285	-1.0914
Profession	Own Empl	1	-2.1142	1.3310	2.52	0.1122		0.121	-4.7230	0.4945
Profession	Unemploy	1	2.7179	1.3034	4.35	0.0370		15.148	0.1633	5.2725
Profession	Worker	0	0	.	.	.	.	.	0.1633	5.2725
Repayment	Yes (In some cases warnings requ	1	2.2087	1.4773	2.24	0.1349		9.103	-0.6868	5.1041
Repayment	Yes (never warned or deferred)	1	-1.9220	1.4262	1.82	0.1778		0.146	-4.7173	0.8734
Repayment	so far no credit customer	0	0	.	.	.	.	.	-4.7173	0.8734
Residence	Job apartment	1	-9.1597	100.0	0.01	0.9270		0.000	-205.2	186.9
Residence	Own house	1	-0.7169	0.3200	5.02	0.0251		0.488	-1.3440	-0.0898
Residence	Rent	1	1.0935	0.3515	9.68	0.0019		2.985	0.4045	1.7824
Residence	With parents	0	0	.	.	.	.	.	0.4045	1.7824
Term		1	0.000091	0.000071	1.63	0.2011	0.1649	1.000	-0.00005	0.000230
Product_id	CAR	1	-0.5216	0.6847	0.58	0.4462		0.594	-1.8636	0.8204
Product_id	CONSUMER	1	-0.8303	0.4402	3.56	0.0593		0.436	-1.6930	0.0325
Product_id	MORTGAGE	0	0	.	.	.	.	.	-1.6930	0.0325
Seniority	0.5 - 1 Year	1	-0.3220	0.4767	0.46	0.4993		0.725	-1.2564	0.6123
Seniority	1 - 2 Years	1	-0.1493	0.3878	0.15	0.7002		0.861	-0.9093	0.6107
Seniority	3 - 5 Years	1	-0.0898	0.3841	0.05	0.8152		0.914	-0.8426	0.6631
Seniority	6 - 10 Years	1	-0.0318	0.4611	0.00	0.9450		0.969	-0.9356	0.8721
Seniority	< 6 Months	1	-0.0113	0.4871	0.00	0.9815		0.989	-0.9659	0.9434
Seniority	> 10 Years	0	0	.	.	.	.	.	-0.9659	0.9434

Appendix 6-2006-Logistic Regression Output(2)

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	162.7283		<.0001
2	Education		2	2	126.5888		<.0001
3	Expenses		1	3	90.2411		<.0001
4	Residence		3	4	81.5458		<.0001
5	Profession		3	5	34.1309		<.0001
6	Industry		14	6	50.1257		<.0001
7	Marital_Status		3	7	25.7411		<.0001
8	Interest_rate		1	8	12.2790		0.0005
9	Age		1	9	5.9322		0.0149
10	Income		1	10	5.2365		0.0221
11	Payment_ron		1	11	49.9670		<.0001
12	Term		1	12	4.7754		0.0289
13	County_ID		1	13	1.8726		0.1712
14		County_ID	1	12		1.8672	0.1718

Appendix 7-2006-Stepwise Selection Probit Regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood				
Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
1234.699	594.155	640.5439	33	<.0001

Appendix 8-2006-LR test Probit Regression (1)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-1.8746	1.0535	3.17	0.0752		-3.9394	0.1902
Age		1	-0.0125	0.00869	2.08	0.1489	-0.1148	-0.0296	0.00449
Education	High School	1	0.9520	0.1713	30.88	<.0001		0.6162	1.2877
Education	Primary school	1	1.4956	0.2885	26.87	<.0001		0.9302	2.0611
Education	University	0	0	.	.	.	.	0.9302	2.0611
Expenses		1	0.00228	0.000275	68.49	<.0001	0.9139	0.00174	0.00281
Income		1	-0.00094	0.000162	34.04	<.0001	-1.6487	-0.00126	-0.00063
Industry	Agriculture	1	-2.3191	107.9	0.00	0.9829		-213.8	209.1
Industry	Bank and Financial Services	1	-0.0530	0.6980	0.01	0.9395		-1.4210	1.3150
Industry	Construction	1	0.8592	0.6523	1.74	0.1878		-0.4192	2.1376
Industry	Electronics / Pharmaceutical / O	1	-4.5939	22.6045	0.04	0.8390		-48.8979	39.7100
Industry	Food	1	1.3406	0.6565	4.17	0.0411		0.0539	2.6273
Industry	Gastronomie	1	-2.4192	122.6	0.00	0.9843		-242.6	237.8
Industry	Leather / Textile / Clothing	1	0.3554	0.6967	0.26	0.6099		-1.0101	1.7209
Industry	Other	1	0.1859	0.5943	0.10	0.7545		-0.9789	1.3506
Industry	Plastic / rubber / asbestos	1	-0.3543	1.3114	0.07	0.7870		-2.9246	2.2160
Industry	Public Service	1	0.3770	0.5984	0.40	0.5287		-0.7958	1.5497
Industry	Retail	1	1.3435	0.6625	4.11	0.0426		0.0451	2.6419
Industry	Steel/metal processing	1	0.3035	0.6958	0.19	0.6627		-1.0603	1.6672
Industry	Stones / earth / gas / ceramic	1	-0.0998	0.7726	0.02	0.8973		-1.6141	1.4145
Industry	Wholesale	1	1.6174	0.6462	6.27	0.0123		0.3509	2.8839
Industry	Wood	0	0	.	.	.	.	0.3509	2.8839
Interest_rate		1	0.1353	0.0391	11.99	0.0005	0.2761	0.0587	0.2118
Marital_Status	Divorce	1	-0.4045	0.5416	0.56	0.4551		-1.4660	0.6569
Marital_Status	Married	1	-0.6975	0.5066	1.90	0.1685		-1.6904	0.2954
Marital_Status	Single	1	-0.2507	0.5341	0.22	0.6387		-1.2975	0.7960
Marital_Status	Widowed	0	0	.	.	.	.	-1.2975	0.7960
Payment_ron		1	0.00157	0.000256	37.42	<.0001	0.7847	0.00106	0.00207
Profession	Employee	1	-0.9624	0.1955	24.23	<.0001		-1.3456	-0.5791
Profession	Own Empl	1	-0.9567	0.5197	3.39	0.0656		-1.9753	0.0618
Profession	Unemploy	1	1.4602	0.6182	5.58	0.0182		0.2485	2.6719
Profession	Worker	0	0	.	.	.	.	0.2485	2.6719
Repayment	Yes (In some cases warnings requ	1	1.4920	0.2570	33.70	<.0001		0.9883	1.9958
Repayment	Yes (never warned or deferred)	1	-0.5834	0.1936	9.08	0.0026		-0.9629	-0.2039
Repayment	so far no credit customer	0	0	.	.	.	.	-0.9629	-0.2039
Residence	Job apartment	1	-3.5913	52.4325	0.00	0.9454		-106.4	99.1745
Residence	Own house	1	-0.2762	0.1515	3.32	0.0683		-0.5732	0.0208
Residence	Rent	1	0.6283	0.1785	12.39	0.0004		0.2785	0.9782
Residence	With parents	0	0	.	.	.	.	0.2785	0.9782
Term		1	0.000060	0.000027	4.75	0.0292	0.1969	6.046E-6	0.000114

Appendix 9-2006-Probit Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
1234.699	586.732	647.9671	47	<.0001

Appendix 10-2006-LR Test -Probit Regression (2)

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-2.9645	36.8140	0.01	0.9358		-75.1186	69.1896
Age		1	-0.0144	0.00944	2.33	0.1270	-0.1318	-0.0329	0.00409
Bank_r	0	1	1.1582	36.7973	0.00	0.9749		-70.9632	73.2795
Bank_r	1 Year	1	1.0976	36.7915	0.00	0.9762		-71.0124	73.2077
Bank_r	2 Years	1	1.8099	36.7919	0.00	0.9608		-70.3008	73.9206
Bank_r	3-5 Years	1	-0.4051	48.1250	0.00	0.9933		-94.7284	93.9183
Bank_r	6-10 Year	1	-2.2822	1283.8	0.00	0.9986		-2518.6	2514.0
Bank_r	< 1 Year	1	1.4349	36.7910	0.00	0.9689		-70.6740	73.5439
Bank_r	> 10 Year	0	0	.	.	.	.	-70.6740	73.5439
Education	High School	1	0.9398	0.1736	29.31	<.0001		0.5996	1.2801
Education	Primary school	1	1.4558	0.2926	24.75	<.0001		0.8823	2.0294
Education	University	0	0	.	.	.	.	0.8823	2.0294
Expenses		1	0.00224	0.000283	62.38	<.0001	0.8982	0.00168	0.00279
Income		1	-0.00095	0.000167	32.28	<.0001	-1.6604	-0.00128	-0.00062
Industry	Agriculture	1	-1.7448	36.6706	0.00	0.9620		-73.6180	70.1283
Industry	Bank and Financial Services	1	-0.0411	0.7076	0.00	0.9537		-1.4279	1.3457
Industry	Construction	1	0.8659	0.6522	1.76	0.1842		-0.4122	2.1441
Industry	Electronics / Pharmaceutical / O	1	-4.9626	53.9208	0.01	0.9267		-110.6	100.7
Industry	Food	1	1.3936	0.6583	4.48	0.0343		0.1033	2.6839
Industry	Gastronomie	1	-2.0245	42.2283	0.00	0.9618		-84.7905	80.7414
Industry	Leather / Textile / Clothing	1	0.3553	0.6970	0.26	0.6102		-1.0108	1.7213
Industry	Other	1	0.1782	0.5927	0.09	0.7636		-0.9834	1.3399
Industry	Plastic / rubber / asbestos	1	-0.3963	1.3940	0.08	0.7762		-3.1284	2.3359
Industry	Public Service	1	0.3668	0.5975	0.38	0.5393		-0.8044	1.5380
Industry	Retail	1	1.4395	0.6624	4.72	0.0298		0.1412	2.7378
Industry	Steel/metal processing	1	0.3073	0.6977	0.19	0.6596		-1.0602	1.6748
Industry	Stones / earth / gas / ceramic	1	-0.1986	0.7932	0.06	0.8023		-1.7533	1.3561
Industry	Wholesale	1	1.5950	0.6456	6.10	0.0135		0.3295	2.8604
Industry	Wood	0	0	.	.	.	.	0.3295	2.8604
Interest_rate		1	0.1706	0.0454	14.10	0.0002	0.3482	0.0816	0.2597
loan_value_ron		1	-1.98E-6	3.51E-6	0.32	0.5731	-0.1454	-8.86E-6	4.902E-6
Marital_Status	Divorce	1	-0.3874	0.5421	0.51	0.4748		-1.4499	0.6750
Marital_Status	Married	1	-0.7087	0.5067	1.96	0.1620		-1.7018	0.2845
Marital_Status	Single	1	-0.2279	0.5348	0.18	0.6700		-1.2761	0.8203

Marital_Status	Widowed	0	0	.	.	.	.	-1.2761	0.8203
Payment_ron		1	0.00179	0.000443	16.36	<.0001	0.8982	0.000924	0.00266
Product_id	CAR	1	-0.1074	0.3278	0.11	0.7431		-0.7498	0.5350
Product_id	CONSUMER	1	-0.3363	0.2066	2.65	0.1036		-0.7413	0.0686
Product_id	MORTGAGE	0	0	.	.	.	.	-0.7413	0.0686
Profession	Employee	1	-0.9626	0.1974	23.78	<.0001		-1.3495	-0.5757
Profession	Own Empl	1	-0.9868	0.5405	3.33	0.0679		-2.0462	0.0725
Profession	Unemploy	1	1.3617	0.6578	4.29	0.0384		0.0724	2.6511
Profession	Worker	0	0	.	.	.	.	0.0724	2.6511
Repayment	Yes (In some cases warnings requ	1	1.3666	0.7331	3.47	0.0623		-0.0703	2.8035
Repayment	Yes (never warned or deferred)	1	-0.8815	0.7132	1.53	0.2165		-2.2793	0.5163
Repayment	so far no credit customer	0	0	.	.	.	.	-2.2793	0.5163
Residence	Job apartment	1	-3.2278	24.6305	0.02	0.8957		-51.5027	45.0471
Residence	Own house	1	-0.3066	0.1553	3.90	0.0483		-0.6109	-0.00226
Residence	Rent	1	0.6547	0.1814	13.02	0.0003		0.2991	1.0104
Residence	With parents	0	0	.	.	.	.	0.2991	1.0104
Seniority	0.5 - 1 Year	1	-0.2010	0.2446	0.68	0.4111		-0.6804	0.2784
Seniority	1 - 2 Years	1	-0.0968	0.1930	0.25	0.6159		-0.4751	0.2814
Seniority	3 - 5 Years	1	-0.0521	0.1922	0.07	0.7864		-0.4288	0.3246
Seniority	6 - 10 Years	1	0.0269	0.2210	0.01	0.9032		-0.4062	0.4600
Seniority	< 6 Months	1	0.00222	0.2554	0.00	0.9931		-0.4985	0.5029
Seniority	> 10 Years	0	0	.	.	.	.	-0.4985	0.5029
Term		1	0.000063	0.000035	3.29	0.0697	0.2070	-5.07E-6	0.000131

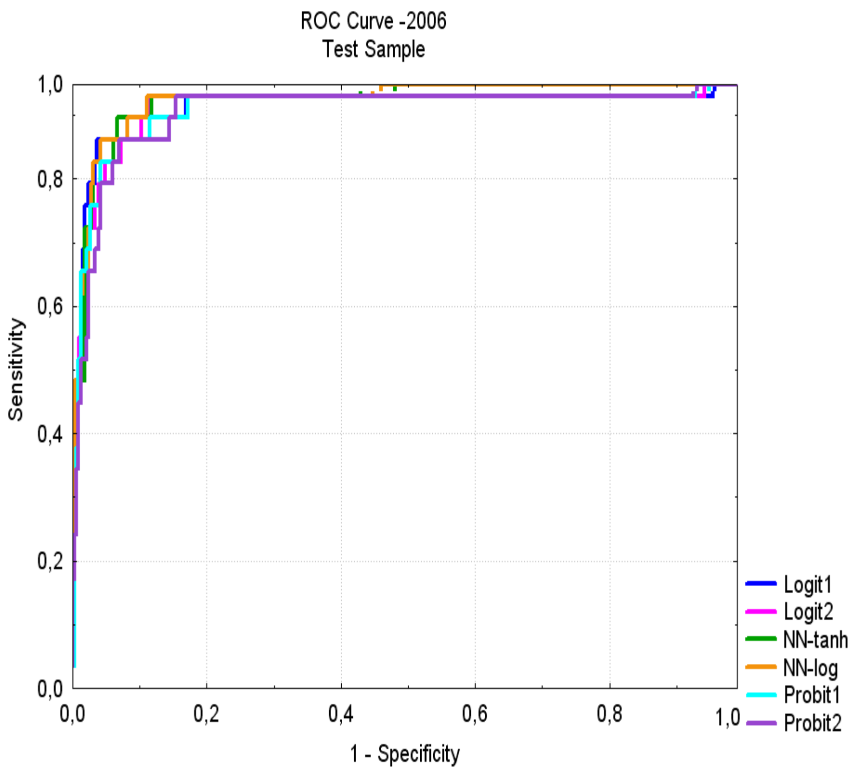
Appendix 11-2006-Probit Regression Output(2)

Optimization Results			
Parameter Estimates			
N	Parameter	Estimate	Gradient Objective Function
1	Age_H11	-0.149128	-0.000392
2	Expenses_H11	-2.972442	0.000212
3	Income_H11	6.374286	0.000110
4	Interest_rate_H11	0.325058	-0.000489
5	loan_value_ron_H11	0.090121	0.000015284
6	Payment_ron_H11	-2.399424	-0.000111
7	Term_H11	0.472027	0.000545
8	Age_H12	-0.887930	0.000046461
9	Expenses_H12	0.433675	-0.000256
10	Income_H12	0.184811	-0.000002871
11	Interest_rate_H12	-1.609689	0.000002559
12	loan_value_ron_H12	-1.565278	0.000091536
13	Payment_ron_H12	-7.212142	0.000162
14	Term_H12	-1.063442	-0.000043000
15	Age_H13	-0.774639	-0.000653
16	Expenses_H13	6.258555	-0.000400
17	Income_H13	-6.692681	-0.000351
18	Interest_rate_H13	0.997329	-0.000007847
19	loan_value_ron_H13	-0.772026	-0.000294
20	Payment_ron_H13	5.621034	-0.000443
21	Term_H13	0.574012	-0.000114
22	Bank_r0_H11	0.230963	-0.000400
23	Bank_r1Year_H11	1.101856	-0.000148
24	Bank_r2Years_H11	0.476262	-0.000099082
25	Bank_r35Years_H11	-0.188375	0.000000220
26	Bank_r610Year_H11	0.212061	0.000000220
27	Bank_r_1Year_H11	-0.290332	0.000207
28	EducationHigh School_H11	0.466297	-0.000077822
29	EducationPrimaryschool_H11	-0.948660	0.000133
30	IndustryAgriculture_H11	0.132617	0.000186
31	IndustryBankandFinancial Services	-0.967544	0.000120
32	IndustryConstruction_H11	-0.120833	-0.000011663
33	IndustryElectronics_Pharmaceutic	2.214396	0.000184
34	IndustryFood_H11	-0.580301	0.000208
35	IndustryGastronomie_H11	0.335866	0.000186
36	IndustryLeather_Textile_Clothing	-0.577288	0.000220
37	IndustryOther_H11	0.291394	0.000605
38	IndustryPlastic_rubber_asbestos_	-0.082741	0.000150
39	IndustryPublic Service_H11	0.537509	-0.000189
40	IndustryRetail_H11	-0.408738	0.000207
41	Industry Steel_metalprocessing_H1	-0.569911	0.000318
42	Industry Stones_earth_gas_ceramic	0.757945	-0.000066630
43	IndustryWholesale_H11	-1.375284	0.000232
44	Marital_ StatusDivorce_H11	-0.159386	-0.000130
45	Marital_ StatusMarried_H11	0.164749	-0.000236
46	Marital_ Status Single_H11	-0.264425	-0.000061399
47	Product_IdCAR_H11	-0.489539	-0.000124
48	Product_IdCONSUMER_H11	0.642238	-0.000371
49	ProfessionEmployee_H11	0.549478	-0.000934
50	ProfessionOwnEmpl_H11	-1.190398	-0.000313
51	ProfessionUnemploy_H11	-0.143828	-0.000559
52	RepaymentYes_Insomecaseswarnings	0.701186	0.000308
53	RepaymentYes_neverwarnedordeferr	-0.312221	0.000663
55	ResidenceOwnhouse_H11	-0.252355	0.000000691
56	ResidenceRent_H11	-0.982771	0.000478
57	Seniority0_51Year_H11	-0.243044	-0.000276
58	Seniority12Years_H11	0.567114	0.000307
59	Seniority35Years_H11	-0.262152	0.000262
60	Seniority610Years_H11	-0.237971	-0.000040345
61	Seniority_6Months_H11	-0.193911	0.000088238
62	Bank_r0_H12	-0.493588	-0.000195
63	Bank_r1Year_H12	-1.702014	-0.000114
64	Bank_r2Years_H12	1.638833	-0.000023457
65	Bank_r35Years_H12	0.064597	-6.204581E-8
66	Bank_r610Year_H12	0.022669	-1.630999E-9
67	Bank_r_1Year_H12	-0.230207	0.000049108
68	EducationHigh School_H12	-1.189517	-0.000068817
69	EducationPrimaryschool_H12	4.903429	-0.000061480
70	IndustryAgriculture_H12	0.103753	0.000004160
71	_DUP	4.052332	0.000003679
72	IndustryConstruction_H12	-2.715685	0.000050360
73	_DUP1	0.718415	-0.000003943
74	IndustryFood_H12	1.564069	-0.000122
75	IndustryGastronomie_H12	0.203000	0.000004160
76	_DUP2	0.730068	-0.000007121
77	IndustryOther_H12	2.075712	-0.000116
78	_DUP3	-0.739614	0.000004257
79	IndustryPublic Service_H12	-1.442686	-0.000168
80	IndustryRetail_H12	-4.594361	0.000043308
81	_DUP4	2.386484	0.000008677
82	_DUP5	0.754263	0.000028257
83	IndustryWholesale_H12	-3.096883	0.000047568

Appendix 12-2006-Neural Networks Output

Technique	Sample	Confusion Matrix				Goodness of fit Tests						
		TN	FN	TP	FP	Sensitivity	Specificity	Misclassification Rate	KS	AUROC	AR	Brier Score
Logit1	validation	763	27	33	11	0.5500	0.9858	0.0456	0.8208	0.9422	0.8843	0.0350
Logit1	test	387	17	12	2	0.4138	0.9949	0.0455	0.8261	0.9153	0.8307	0.0334
Logit1	training	2747	77	82	14	0.5157	0.9949	0.0312	0.7990	0.9641	0.9282	0.0252
Logit 2	validation	765	29	31	9	0.5167	0.9884	0.0456	0.8258	0.9438	0.8875	0.0347
Logit 2	test	387	18	11	2	0.3793	0.9949	0.0478	0.8154	0.9156	0.8312	0.0346
Logit 2	training	2747	77	82	14	0.5157	0.9949	0.0312	0.7990	0.9641	0.9282	0.0252
NN1	test	384	15	14	5	0.4828	0.9871	0.0478	0.8297	0.9520	0.9041	0.0342
NN1	training	2749	49	110	12	0.6918	0.9957	0.0209	0.8493	0.9826	0.9651	0.0170
NN1	validation	764	17	43	10	0.7167	0.9871	0.0324	0.8067	0.9669	0.9339	0.0294
NN2	training	2749	40	119	12	0.7484	0.9957	0.0178	0.9114	0.9901	0.9801	0.0147
NN2	validation	767	14	46	7	0.7667	0.9910	0.0252	0.8611	0.9745	0.9491	0.0236
NN2	test	387	15	14	2	0.4828	0.9949	0.0407	0.8209	0.9530	0.9060	0.0315
Probit 1	validation	764	35	25	10	0.4167	0.9871	0.0540	0.8158	0.9414	0.8829	0.0382
Probit 1	test	387	20	9	2	0.3103	0.9949	0.0526	0.7927	0.9136	0.8271	0.0362
Probit 1	training	2747	86	73	14	0.4591	0.9949	0.0342	0.7856	0.9597	0.9194	0.0281
Probit2	validation	766	33	27	8	0.4500	0.9897	0.0492	0.7975	0.9401	0.8803	0.0377
Probit2	test	387	21	8	2	0.2759	0.9949	0.0550	0.7927	0.9105	0.8209	0.0375
Probit2	training	2749	89	70	12	0.4403	0.9957	0.0346	0.7911	0.9608	0.9216	0.0279

Appendix 13-2006-Goodness of fit results



Appendix 14-2006-ROC Curve Test Sample

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	2448.4263		<.0001
2	Seniority		5	2	653.6987		<.0001
3	Education		2	3	308.4385		<.0001
4	Residence		3	4	176.5847		<.0001
5	Industry		14	5	140.2065		<.0001
6	Income		1	6	70.2653		<.0001
7	Expenses		1	7	278.4038		<.0001
8	loan_value_ron		1	8	98.9083		<.0001
9	Profession		3	9	57.7278		<.0001
10	Marital_Status		3	10	36.7092		<.0001
11	Age		1	11	17.0333		<.0001
12	Bank_r		6	12	27.0577		0.0001
13		Bank_r	6	11		8.4564	0.2065

Appendix 15-2007-Stepwise Selection Logistic Regression

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-5.6747	25.3538	0.05	0.8229		0.003	-1.6223	1.6455
Age		1	-0.0299	0.00738	16.48	<.0001	-0.1447	0.970	-0.0445	-0.0158
Bank_r	0	1	5.7149	25.3397	0.05	0.8216		303.339	0.5121	0.8535
Bank_r	1 -2 Years	1	5.0970	25.3392	0.04	0.8406		163.531	1.5081	2.6069
Bank_r	2-3 Years	1	4.7156	25.3408	0.03	0.8524		111.674	0.00180	0.00231
Bank_r	3-5 Years	1	5.2438	25.3434	0.04	0.8361		189.384	-0.00112	-0.00086
Bank_r	6-10 Years	1	-8.5930	53.2125	0.03	0.8717		0.000	0.4115	2.9697
Bank_r	< 1 Year	1	5.0480	25.3388	0.04	0.8421		155.714	-0.4847	1.4224
Bank_r	> 10 Years	0	0	.	.	.	.	.	-0.4847	1.4224
Education	High School	1	0.6719	0.0876	58.86	<.0001		1.958	0.4439	2.3531
Education	Primary scho	1	2.0739	0.2822	54.02	<.0001		7.956	-0.3159	1.6928
Education	University	0	0	.	.	.	.	.	-0.3159	1.6928
Expenses		1	0.00214	0.000131	266.34	<.0001	0.8594	1.002	0.1195	2.2494
Income		1	-0.00101	0.000067	224.58	<.0001	-1.4999	0.999	1.0985	3.4084
Industry	Agriculture	1	1.7046	0.6577	6.72	0.0096		5.499	0.1030	2.2177
Industry	Bank and Financial Services	1	0.6008	0.4912	1.50	0.2213		1.824	-0.4801	1.3373
Industry	Construction	1	1.3983	0.4899	8.15	0.0043		4.048	-0.4548	2.8354
Industry	Electronics / Pharmaceutical / O	1	0.6893	0.5156	1.79	0.1812		1.992	-0.9374	0.8780
Industry	Food	1	1.1684	0.5471	4.56	0.0327		3.217	-0.4700	1.4866
Industry	Gastronomie	1	2.2583	0.5916	14.57	0.0001		9.567	-1.2555	0.8007

Industry	Leather / Textile / Clothing	1	1.1680	0.5419	4.65	0.0311		3.215	-0.9517	1.2552
Industry	Other	1	0.4696	0.4664	1.01	0.3140		1.599	-0.7909	1.5410
Industry	Plastic / rubber / asbestos	1	1.1999	0.8409	2.04	0.1536		3.320	6.566E-6	0.000010
Industry	Public Service	1	-0.0255	0.4656	0.00	0.9564		0.975	-0.1238	2.1448
Industry	Retail	1	0.5010	0.5019	1.00	0.3182		1.650	-1.2863	0.9026
Industry	Steel/metal processing	1	-0.2411	0.5284	0.21	0.6482		0.786	-0.9726	1.2412
Industry	Stones / earth / gas / ceramic	1	0.1693	0.5650	0.09	0.7644		1.184	-1.8445	-1.0671
Industry	Wholesale	1	0.3598	0.5982	0.36	0.5475		1.433	-2.2152	-0.8862
Industry	Wood	0	0	.	.	.	.	.	-2.2152	-0.8862
loan_value_ron		1	8.737E-6	9.181E-7	90.56	<.0001	0.6016	1.000	-2.5202	0.9158
Marital_Status	Divorced	1	0.9796	0.5776	2.88	0.0899		2.663	4.1608	4.8675
Marital_Status	Married	1	-0.1984	0.5568	0.13	0.7216		0.820	-1.3127	-0.6737
Marital_Status	Single	1	0.1002	0.5637	0.03	0.8590		1.105	-0.9845	1.3232
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.9845	1.3232
Profession	Employee	1	-1.4854	0.1994	55.48	<.0001		0.226	-1.1172	-0.7024
Profession	Own Empl	1	-1.5863	0.3420	21.51	<.0001		0.205	-0.3300	0.4104
Profession	Unemploy	1	-0.9412	0.9146	1.06	0.3034		0.390	1.1431	1.8047
Profession	Worker	0	0	.	.	.	.	.	1.1431	1.8047
Repayment	Yes (in some cases warnings requ	1	5.1015	0.2800	331.84	<.0001		164.267	0.6406	1.2741
Repayment	Yes (never warned or deferred)	1	-0.3873	0.2806	1.91	0.1675		0.679	0.3870	1.0301
Repayment	so far no credit customer	0	0	.	.	.	.	.	0.3870	1.0301
Residence	Job apartment	1	0.1629	0.5904	0.08	0.7826		1.177	-0.2927	0.4521
Residence	Own house	1	-0.9344	0.1064	77.10	<.0001		0.393	1.2072	1.9184
Residence	Rent	1	0.0424	0.1898	0.05	0.8234		1.043	-0.3296	0.4143
Residence	With parents	0	0	.	.	.	.	.	-0.3296	0.4143
Seniority	0.5 - 1 Year	1	1.4716	0.1700	74.95	<.0001		4.356	1.1384	1.8048
Seniority	1 - 2 Years	1	0.9540	0.1628	34.35	<.0001		2.596	0.6350	1.2730
Seniority	3 - 5 Years	1	0.7004	0.1652	17.97	<.0001		2.015	0.3766	1.0242
Seniority	6 - 10 Years	1	0.1402	0.1901	0.54	0.4607		1.151	-0.2324	0.5128
Seniority	< 6 Months	1	1.5920	0.1828	75.88	<.0001		4.914	1.2338	1.9502
Seniority	> 10 Years	0	0	.	.	.	.	.	1.2338	1.9502

Appendix 16-2007-Logistic Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
7908.011	4392.604	3515.4068	38	<.0001

Appendix 17-2007-LR Test Logistic Regression(2)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-2.0065	0.8618	5.42	0.0199		0.134	-3.6956	-0.3175
Age		1	-0.0315	0.00741	18.02	<.0001	-0.1520	0.969	-0.0460	-0.0169
Education	High School	1	0.7822	0.0858	83.07	<.0001		2.186	0.6140	0.9504
Education	Primary scho	1	2.0747	0.2827	53.87	<.0001		7.962	1.5207	2.6287
Education	University	0	0	.	.	.	.	.	1.5207	2.6287
Expenses		1	0.00220	0.000139	248.80	<.0001	0.8805	1.002	0.00192	0.00247
Income		1	-0.00108	0.000075	205.39	<.0001	-1.6045	0.999	-0.00123	-0.00093
Industry	Agriculture	1	1.6919	0.6502	6.77	0.0093		5.430	0.4175	2.9664
Industry	Bank and Financial Services	1	0.4479	0.4860	0.85	0.3566		1.565	-0.5045	1.4004
Industry	Construction	1	1.4191	0.4852	8.55	0.0034		4.133	0.4681	2.3700
Industry	Electronics / Pharmaceutical / O	1	0.6929	0.5127	1.83	0.1765		1.999	-0.3119	1.6977
Industry	Food	1	1.1281	0.5449	4.29	0.0384		3.090	0.0602	2.1961
Industry	Gastronomie	1	2.1807	0.5927	13.54	0.0002		8.853	1.0190	3.3425
Industry	Leather / Textile / Clothing	1	1.1753	0.5381	4.77	0.0290		3.239	0.1206	2.2300
Industry	Other	1	0.4774	0.4628	1.06	0.3023		1.612	-0.4297	1.3845
Industry	Plastic / rubber / asbestos	1	0.9976	0.8533	1.37	0.2423		2.712	-0.6748	2.6700
Industry	Public Service	1	0.0107	0.4627	0.00	0.9816		1.011	-0.8962	0.9175
Industry	Retail	1	0.4770	0.4994	0.91	0.3395		1.611	-0.5018	1.4558
Industry	Steel/metal processing	1	-0.2073	0.5247	0.16	0.6927		0.813	-1.2357	0.8210
Industry	Stones / earth / gas / ceramic	1	0.1184	0.5658	0.04	0.8342		1.126	-0.9904	1.2273
Industry	Wholesale	1	0.3517	0.5955	0.35	0.5548		1.422	-0.8154	1.5188
Industry	Wood	0	0	.	.	.	.	.	-0.8154	1.5188
Interest_rate		1	0.0900	0.0421	4.58	0.0323	0.0919	1.094	0.00759	0.1725
loan_value_ron		1	1.866E-6	2E-6	0.87	0.3510	0.1285	1.000	-2.05E-6	5.786E-6
Marital_Status	Divorced	1	1.0346	0.5766	3.22	0.0727		2.814	-0.0954	2.1646
Marital_Status	Married	1	-0.1787	0.5574	0.10	0.7484		0.836	-1.2712	0.9137
Marital_Status	Single	1	0.1212	0.5636	0.05	0.8298		1.129	-0.9835	1.2259
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.9835	1.2259
Payment_ron		1	0.00128	0.000307	17.30	<.0001	0.5370	1.001	0.000676	0.00188
Product_id	CAR	1	-0.5736	0.2101	7.45	0.0063		0.563	-0.9854	-0.1618
Product_id	CONSUMER	1	-0.4155	0.3149	1.74	0.1870		0.660	-1.0326	0.2017
Product_id	MORTGAGE	0	0	.	.	.	.	.	-1.0326	0.2017
Repayment	Yes (In some cases warnings requ	1	4.4262	0.1798	605.87	<.0001		83.613	4.0738	4.7786
Repayment	Yes (never warned or deferred)	1	-1.0627	0.1627	42.65	<.0001		0.346	-1.3817	-0.7438
Repayment	so far no credit customer	0	0	.	.	.	.	.	-1.3817	-0.7438
Residence	Job apartment	1	0.2039	0.5719	0.13	0.7214		1.226	-0.9170	1.3248
Residence	Own house	1	-0.8939	0.1065	70.51	<.0001		0.409	-1.1025	-0.6852
Residence	Rent	1	0.0224	0.1895	0.01	0.9061		1.023	-0.3491	0.3938
Residence	With parents	0	0	.	.	.	.	.	-0.3491	0.3938
Seniority	0.5 - 1 Year	1	1.4647	0.1683	75.77	<.0001		4.326	1.1349	1.7945
Seniority	1 - 2 Years	1	0.9728	0.1605	36.72	<.0001		2.645	0.6581	1.2874
Seniority	3 - 5 Years	1	0.7326	0.1627	20.27	<.0001		2.081	0.4137	1.0516
Seniority	6 - 10 Years	1	0.1156	0.1879	0.38	0.5385		1.123	-0.2526	0.4838
Seniority	< 6 Months	1	1.5321	0.1816	71.19	<.0001		4.628	1.1762	1.8880
Seniority	> 10 Years	0	0	.	.	.	.	.	1.1762	1.8880

Appendix 18-2007-Logistic Regression Output(2)

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	1548.5035		<.0001
2	Seniority		5	2	623.1347		<.0001
3	Education		2	3	273.0330		<.0001
4	Residence		3	4	167.2741		<.0001
5	Industry		14	5	125.1009		<.0001
6	Payment_ron		1	6	65.3441		<.0001
7	Expenses		1	7	73.9997		<.0001
8	Income		1	8	220.7070		<.0001
9	Profession		3	9	48.3775		<.0001
10	Product_id		2	10	37.3677		<.0001
11	Marital_Status		3	11	37.7538		<.0001
12	Age		1	12	21.4078		<.0001
13	Bank_r		6	13	23.4554		0.0007
14	Interest_rate		1	14	5.0591		0.0245
15	CCY		2	15	25.8455		<.0001
16	Term		1	16	3.0824		0.0791
17	loan_value		1	17	3.5902		0.0581
18	County_ID		1	18	1.5810		0.2086
19		County_ID	1	17		1.6005	0.2058

Appendix 19-2007-Stepwise Selection-Probit Regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio		
Intercept Only	Intercept & Covariates	Chi-Square	DF	Pr > ChiSq
7908.011	4366.248	3541.7622	49	<.0001

Appendix 20-2007-LR Test-Probit Regression(1)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-4.2889	21.0212	0.04	0.8383		-45.4896	36.9118
Age		1	-0.0159	0.00389	16.78	<.0001	-0.1397	-0.0236	-0.00831
Bank_r	0	1	2.8219	21.0141	0.02	0.8932		-38.3650	44.0087
Bank_r	1 -2 Years	1	2.4822	21.0139	0.01	0.9060		-38.7043	43.6688
Bank_r	2-3 Years	1	2.1778	21.0145	0.01	0.9175		-39.0098	43.3654
Bank_r	3-5 Years	1	2.6453	21.0150	0.02	0.8998		-38.5434	43.8339
Bank_r	6-10 Years	1	-4.1145	90.2626	0.00	0.9636		-181.0	172.8
Bank_r	< 1 Year	1	2.4689	21.0138	0.01	0.9065		-38.7174	43.6552
Bank_r	> 10 Years	0	0	.	.	.	.	-38.7174	43.6552
CCY	CHF	1	0.4503	0.1448	9.67	0.0019		0.1664	0.7342
CCY	EUR	1	0.0124	0.1170	0.01	0.9153		-0.2169	0.2418
CCY	RON	0	0	.	.	.	.	-0.2169	0.2418
Education	High School	1	0.3260	0.0465	49.22	<.0001		0.2349	0.4171
Education	Primary scho	1	1.1836	0.1623	53.21	<.0001		0.8656	1.5016
Education	University	0	0	.	.	.	.	0.8656	1.5016
Expenses		1	0.000879	0.000052	284.81	<.0001	0.6396	0.000777	0.000981
Income		1	-0.00044	0.000032	193.15	<.0001	-1.1973	-0.00051	-0.00038
Industry	Agriculture	1	0.6882	0.3586	3.68	0.0550		-0.0146	1.3911
Industry	Bank and Financial Services	1	0.3743	0.2629	2.03	0.1546		-0.1411	0.8896
Industry	Construction	1	0.7481	0.2647	7.99	0.0047		0.2294	1.2668
Industry	Electronics / Pharmaceutical / O	1	0.3450	0.2799	1.52	0.2177		-0.2036	0.8937
Industry	Food	1	0.6418	0.2981	4.63	0.0313		0.0575	1.2261

Industry	Gastronomie	1	1.2787	0.3293	15.08	0.0001		0.6333	1.9241
Industry	Leather / Textile / Clothing	1	0.6576	0.2981	4.87	0.0274		0.0734	1.2419
Industry	Other	1	0.2767	0.2515	1.21	0.2712		-0.2162	0.7695
Industry	Plastic / rubber / asbestos	1	0.4334	0.4949	0.77	0.3812		-0.5366	1.4034
Industry	Public Service	1	0.0179	0.2505	0.01	0.9431		-0.4731	0.5088
Industry	Retail	1	0.2405	0.2711	0.79	0.3749		-0.2908	0.7718
Industry	Steel/metal processing	1	-0.1942	0.2871	0.46	0.4987		-0.7568	0.3685
Industry	Stones / earth / gas / ceramic	1	0.0598	0.3007	0.04	0.8423		-0.5296	0.6493
Industry	Wholesale	1	0.1686	0.3261	0.27	0.6052		-0.4705	0.8076
Industry	Wood	0	0	.	.	.	.	-0.4705	0.8076
Interest_rate		1	0.1643	0.0420	15.32	<.0001	0.3042	0.0820	0.2466
loan_value		1	-2.64E-6	1.389E-6	3.61	0.0573	-0.1398	-5.36E-6	8.225E-8
Marital_Status	Divorced	1	0.5984	0.3083	3.77	0.0523		-0.00583	1.2026
Marital_Status	Married	1	-0.0126	0.2969	0.00	0.9662		-0.5944	0.5693
Marital_Status	Single	1	0.1423	0.3005	0.22	0.6360		-0.4467	0.7313
Marital_Status	Widowed	0	0	.	.	.	.	-0.4467	0.7313
Payment_ron		1	0.000779	0.000106	53.92	<.0001	0.5929	0.000571	0.000986
Product_id	CAR	1	-0.1173	0.1270	0.85	0.3556		-0.3663	0.1316
Product_id	CONSUMER	1	-0.4393	0.2180	4.06	0.0439		-0.8665	-0.0121
Product_id	MORTGAGE	0	0	.	.	.	.	-0.8665	-0.0121
Profession	Employee	1	-0.8212	0.1159	50.22	<.0001		-1.0483	-0.5941
Profession	Own Empl	1	-1.0415	0.1855	31.52	<.0001		-1.4051	-0.6779
Profession	Unemploy	1	-0.3298	0.4003	0.68	0.4100		-1.1145	0.4548
Profession	Worker	0	0	.	.	.	.	-1.1145	0.4548
Repayment	Yes (In some cases warnings requ	1	2.6846	0.1356	392.01	<.0001		2.4188	2.9503
Repayment	Yes (never warned or deferred)	1	-0.1797	0.1378	1.70	0.1922		-0.4499	0.0904
Repayment	so far no credit customer	0	0	.	.	.	.	-0.4499	0.0904
Residence	Job apartment	1	0.00747	0.3259	0.00	0.9817		-0.6313	0.6462
Residence	Own house	1	-0.4633	0.0539	73.89	<.0001		-0.5689	-0.3577
Residence	Rent	1	0.0239	0.1068	0.05	0.8230		-0.1854	0.2332
Residence	With parents	0	0	.	.	.	.	-0.1854	0.2332
Seniority	0.5 - 1 Year	1	0.7434	0.0870	73.00	<.0001		0.5729	0.9139
Seniority	1 - 2 Years	1	0.4451	0.0809	30.23	<.0001		0.2864	0.6037
Seniority	3 - 5 Years	1	0.3386	0.0810	17.49	<.0001		0.1799	0.4973
Seniority	6 - 10 Years	1	0.0670	0.0905	0.55	0.4594		-0.1104	0.2443
Seniority	< 6 Months	1	0.8333	0.0949	77.07	<.0001		0.6473	1.0194
Seniority	> 10 Years	0	0	.	.	.	.	0.6473	1.0194
Term		1	0.000035	0.000015	5.55	0.0185	0.1113	5.803E-6	0.000063

Appendix 21-2007-Probit Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
7908.011	4474.680	3433.3311	38	<.0001

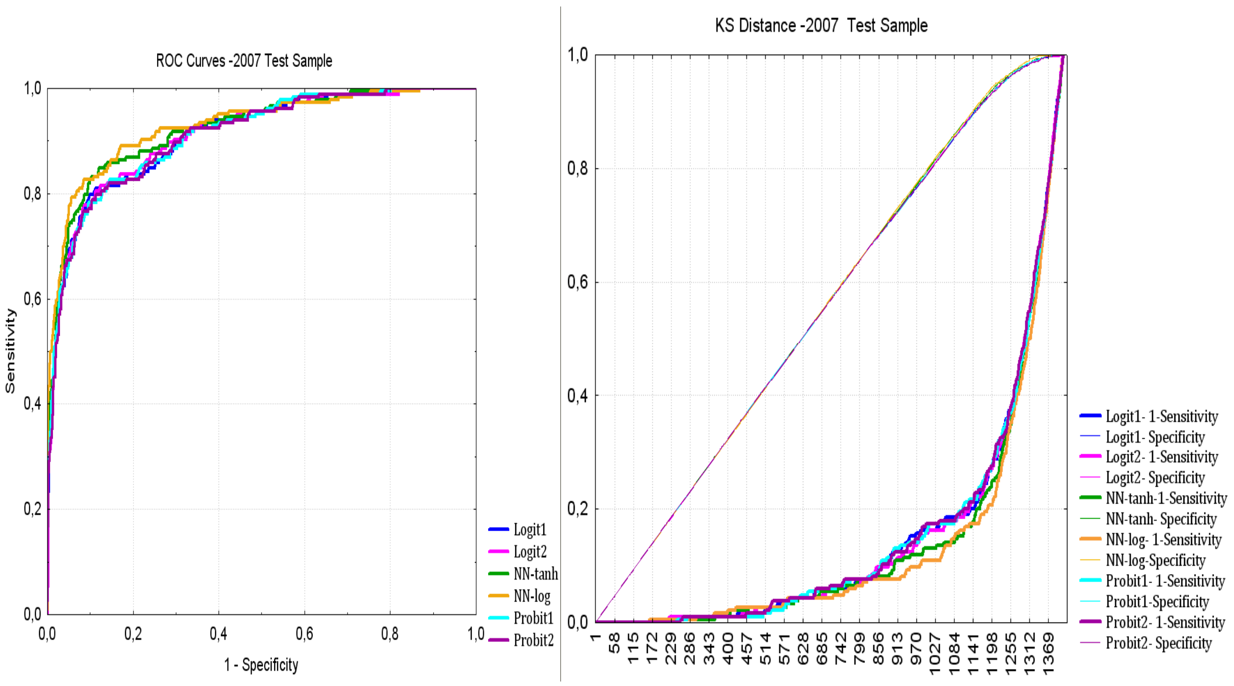
Appendix 22-2007-LR Test Probit Regression (2)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-1.3171	0.4531	8.45	0.0037		-2.2051	-0.4290
Age		1	-0.0155	0.00379	16.71	<.0001	-0.1357	-0.0229	-0.00806
Education	High School	1	0.3966	0.0449	77.98	<.0001		0.3086	0.4846
Education	Primary scho	1	1.2126	0.1585	58.53	<.0001		0.9019	1.5232
Education	University	0	0	.	.	.	.	0.9019	1.5232
Expenses		1	0.000863	0.000051	281.21	<.0001	0.6277	0.000762	0.000964
Income		1	-0.00045	0.000032	202.76	<.0001	-1.2179	-0.00051	-0.00039
Industry	Agriculture	1	0.7982	0.3517	5.15	0.0232		0.1089	1.4874
Industry	Bank and Financial Services	1	0.2755	0.2587	1.13	0.2870		-0.2317	0.7826
Industry	Construction	1	0.7394	0.2609	8.03	0.0046		0.2281	1.2507
Industry	Electronics / Pharmaceutical / O	1	0.3359	0.2765	1.48	0.2245		-0.2061	0.8779
Industry	Food	1	0.6571	0.2942	4.99	0.0255		0.0804	1.2337
Industry	Gastronomie	1	1.2375	0.3279	14.24	0.0002		0.5948	1.8803
Industry	Leather / Textile / Clothing	1	0.6433	0.2946	4.77	0.0290		0.0659	1.2206
Industry	Other	1	0.2559	0.2483	1.06	0.3026		-0.2307	0.7425
Industry	Plastic / rubber / asbestos	1	0.4716	0.4810	0.96	0.3269		-0.4712	1.4144
Industry	Public Service	1	0.0205	0.2476	0.01	0.9339		-0.4648	0.5059
Industry	Retail	1	0.2278	0.2678	0.72	0.3949		-0.2971	0.7527
Industry	Steel/metal processing	1	-0.1697	0.2823	0.36	0.5478		-0.7231	0.3837
Industry	Stones / earth / gas / ceramic	1	0.0857	0.2947	0.08	0.7713		-0.4920	0.6634
Industry	Wholesale	1	0.1798	0.3226	0.31	0.5773		-0.4525	0.8122
Industry	Wood	0	0	.	.	.	.	-0.4525	0.8122
Interest_rate		1	0.0421	0.0224	3.52	0.0607	0.0780	-0.00189	0.0861
loan_value_ron		1	-9.46E-8	1.002E-6	0.01	0.9248	-0.0118	-2.06E-6	1.869E-6
Marital_Status	Divorced	1	0.6414	0.3075	4.35	0.0370		0.0387	1.2440
Marital_Status	Married	1	0.0121	0.2964	0.00	0.9674		-0.5688	0.5931
Marital_Status	Single	1	0.1790	0.2997	0.36	0.5504		-0.4084	0.7664
Marital_Status	Widowed	0	0	.	.	.	.	-0.4084	0.7664
Payment_ron		1	0.000623	0.000166	14.02	0.0002	0.4743	0.000297	0.000949
Product_id	CAR	1	-0.3311	0.1094	9.16	0.0025		-0.5454	-0.1167
Product_id	CONSUMER	1	-0.1403	0.1667	0.71	0.4002		-0.4670	0.1865
Product_id	MORTGAGE	0	0	.	.	.	.	-0.4670	0.1865
Repayment	Yes (In some cases warnings requ	1	2.3675	0.0871	738.65	<.0001		2.1967	2.5382
Repayment	Yes (never warned or deferred)	1	-0.5219	0.0765	46.52	<.0001		-0.6719	-0.3720
Repayment	so far no credit customer	0	0	.	.	.	.	-0.6719	-0.3720
Residence	Job apartment	1	0.0286	0.3172	0.01	0.9282		-0.5931	0.6502
Residence	Own house	1	-0.4439	0.0530	70.11	<.0001		-0.5478	-0.3400
Residence	Rent	1	0.00779	0.1055	0.01	0.9412		-0.1990	0.2146
Residence	With parents	0	0	.	.	.	.	-0.1990	0.2146
Seniority	0.5 - 1 Year	1	0.7423	0.0855	75.44	<.0001		0.5748	0.9098
Seniority	1 - 2 Years	1	0.4533	0.0793	32.71	<.0001		0.2979	0.6086
Seniority	3 - 5 Years	1	0.3529	0.0794	19.75	<.0001		0.1972	0.5085
Seniority	6 - 10 Years	1	0.0672	0.0879	0.59	0.4443		-0.1050	0.2394
Seniority	< 6 Months	1	0.7921	0.0934	71.97	<.0001		0.6091	0.9751
Seniority	> 10 Years	0	0	.	.	.	.	0.6091	0.9751
Term		1	0.000019	0.000016	1.43	0.2312	0.0608	-0.00001	0.000050

Appendix 23-2007-Probit Regression Output (2)

Model	Sample	Confusion Matrix				Goodness of Fit						
		TN	FN	TP	FP	Sensitivity	Specificity	Misclass Rate	KS	AUROC	AR	Brier Score
Logit1	training	8426	670	686	116	0.5059	0.9864	0.0794	0.6897	0.9187	0.8374	0.0611
Logit1	validation	2395	199	200	34	0.5013	0.9860	0.0824	0.6813	0.9071	0.8143	0.0642
Logit1	test	1207	92	92	23	0.5000	0.9813	0.0813	0.7005	0.9110	0.8220	0.0621
Logit 2	training	8420	684	672	122	0.4956	0.9857	0.0814	0.6864	0.9177	0.8354	0.0622
Logit 2	validation	2388	196	203	41	0.5088	0.9831	0.0838	0.6773	0.9066	0.8131	0.0645
Logit 2	test	1205	93	91	25	0.4946	0.9797	0.0835	0.6942	0.9125	0.8249	0.0623
Probit 1	training	8429	704	652	113	0.4808	0.9868	0.0825	0.6881	0.9181	0.8361	0.0625
Probit 1	validation	2394	208	191	35	0.4787	0.9856	0.0859	0.6743	0.9073	0.8146	0.0653
Probit 1	test	1209	90	94	21	0.5109	0.9829	0.0785	0.6875	0.9108	0.8217	0.0621
Probit2	training	8434	739	617	108	0.4550	0.9874	0.0856	0.6794	0.9145	0.8289	0.0643
Probit2	validation	2395	213	186	34	0.4662	0.9860	0.0873	0.6743	0.9073	0.8146	0.0656
Probit2	test	1209	101	83	21	0.4511	0.9829	0.0863	0.6851	0.9090	0.8180	0.0641
NN1	training	8368	542	814	174	0.6003	0.9796	0.0723	0.7146	0.9340	0.8681	0.0564
NN1	validation	2382	159	240	47	0.6015	0.9807	0.0728	0.6975	0.9207	0.8414	0.0594
NN1	test	1201	77	107	29	0.5815	0.9764	0.0750	0.7291	0.9222	0.8443	0.0577
NN2	training	8365	541	815	177	0.6010	0.9793	0.0725	0.7313	0.9404	0.8808	0.0544
NN2	validation	2381	155	244	48	0.6115	0.9802	0.0718	0.6858	0.9228	0.8457	0.0580
NN2	test	1203	74	110	27	0.5978	0.9780	0.0714	0.7415	0.9290	0.8580	0.0525

Appendix 24-2007-Goodness of Fit Results



Appendix 25-2007 ROC Curve and KS Distance

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	1652.3664		<.0001
2	Residence		3	2	1059.4059		<.0001
3	Education		2	3	386.1939		<.0001
4	Marital_Status		3	4	408.3472		<.0001
5	Seniority		5	5	219.6831		<.0001
6	Product_id		2	6	81.7989		<.0001
7	Interest_rate		1	7	103.1738		<.0001
8	Payment_ron		1	8	63.3444		<.0001
9	Income		1	9	107.2693		<.0001
10	Expenses		1	10	280.8578		<.0001
11	Industry		14	11	50.1239		<.0001
12	CCY		2	12	15.2085		0.0005
13	loan_value_ron		1	13	11.0652		0.0009
14	Age		1	14	8.2165		0.0042
15	Profession		3	15	8.1472		0.0431
16	Bank_r		6	16	11.3471		0.0782
17	County_ID		1	17	1.5676		0.2106
18		County_ID	1	16		1.5701	0.2102

Appendix 26-2008-Stepwise Selection Logistic regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
10105.006	5787.769	4317.2374	42	<.0001

Appendix 27-2008-LR Test Logistic Regression (1)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-4.0368	0.8691	21.57	<.0001		0.018	-8.8491	-3.0758
Age		1	-0.0188	0.00579	10.57	0.0012	-0.0955	0.981	-0.0301	-0.00731
CCY	CHF	1	1.2239	0.2890	17.94	<.0001		3.401	-0.4184	4.2068
CCY	EUR	1	0.8069	0.2078	15.08	0.0001		2.241	0.0187	4.6546
CCY	RON	0	0	.	.	.	.	.	0.0187	4.6546
Education	High School	1	1.2397	0.0734	285.07	<.0001		3.455	-0.7574	3.9580
Education	Primary scho	1	1.5759	0.2353	44.86	<.0001		4.835	0.0413	4.8793
Education	University	0	0	.	.	.	.	.	0.0413	4.8793
Expenses		1	0.00161	0.000120	180.48	<.0001	0.5521	1.002	-1.5597	5.6078
Income		1	-0.00047	0.000029	269.34	<.0001	-1.0006	1.000	0.0146	4.6070
Industry	Agriculture	1	0.5518	0.8420	0.43	0.5123		1.736	0.6621	1.7969
Industry	Bank and Financial Services	1	-0.1428	0.4957	0.08	0.7733		0.867	0.4020	1.2183
Industry	Construction	1	0.3226	0.5028	0.41	0.5211		1.381	1.1022	1.3908
Industry	Electronics / Pharmaceutical / O	1	-0.4563	0.5547	0.68	0.4108		0.634	1.1107	2.0371
Industry	Food	1	-1.2235	0.7023	3.04	0.0815		0.294	0.00138	0.00185
Industry	Gastronomie	1	0.3946	0.6782	0.34	0.5607		1.484	-0.00053	-0.00041
Industry	Leather / Textile / Clothing	1	0.3437	0.6881	0.25	0.6174		1.410	-1.0909	2.2001
Industry	Other	1	0.0718	0.4737	0.02	0.8796		1.074	-1.2184	0.7368
Industry	Plastic / rubber / asbestos	1	-1.0384	1.2361	0.71	0.4009		0.354	-0.6308	1.3387
Industry	Public Service	1	-0.3850	0.4799	0.64	0.4224		0.680	-1.5334	0.6406
Industry	Retail	1	-0.5458	0.5358	1.04	0.3084		0.579	-2.5975	0.1519
Industry	Steel/metal processing	1	-1.1308	0.5753	3.86	0.0494		0.323	-0.9048	1.7487
Industry	Stones / earth / gas / ceramic	1	-0.1783	0.5626	0.10	0.7513		0.837	-0.9916	1.7070

Industry	Wholesale	1	-0.1983	0.6459	0.09	0.7588		0.820	-0.8508	1.0054
Industry	Wood	0	0	.	.	.	.	.	-0.8508	1.0054
Interest_rate		1	0.3604	0.0479	56.64	<.0001	0.4635	1.434	-3.4625	1.3591
loan_value_ron		1	4.368E-6	1.352E-6	10.44	0.0012	0.4004	1.000	-1.3210	0.5593
Marital_Status	Divorced	1	1.1467	0.3539	10.50	0.0012		3.148	-1.5995	0.5005
Marital_Status	Married	1	-0.8200	0.3337	6.04	0.0140		0.440	-2.2388	0.0138
Marital_Status	Single	1	-0.0631	0.3433	0.03	0.8541		0.939	-1.2760	0.9301
Marital_Status	Widowed	0	0	.	.	.	.	.	-1.2760	0.9301
Payment_ron		1	0.000669	0.000189	12.59	0.0004	0.4386	1.001	-1.4448	1.0828
Product_id	CAR	1	-0.3647	0.2018	3.27	0.0707		0.694	0.2656	0.4537
Product_id	CONSUMER	1	-0.7455	0.1238	36.24	<.0001		0.474	1.678E-6	6.987E-6
Product_id	MORTGAGE	0	0	.	.	.	.	.	1.678E-6	6.987E-6
Profession	Employee	1	0.4401	0.3369	1.71	0.1914		1.553	0.4565	1.8433
Profession	Own Employee	1	0.3367	0.3762	0.80	0.3707		1.400	-1.4668	-0.1593
Profession	Unemployed	1	1.2653	0.4769	7.04	0.0080		3.544	-0.7274	0.6182
Profession	Worker	0	0	.	.	.	.	.	-0.7274	0.6182
Repayment	Yes (In some cases warnings requ	1	4.6546	0.2153	467.36	<.0001		105.071	0.000310	0.00105
Repayment	Yes (never warned or deferred)	1	-2.1088	0.1861	128.44	<.0001		0.121	-0.7523	0.0399
Repayment	so far no credit customer	0	0	.	.	.	.	.	-0.7523	0.0399
Residence	Job apartment	1	-0.0923	0.5379	0.03	0.8637		0.912	-0.9839	-0.4978
Residence	Own house	1	-1.3874	0.0882	247.16	<.0001		0.250	-0.2141	1.1063
Residence	Rent	1	1.2059	0.1326	82.74	<.0001		3.340	-0.3589	1.1171
Residence	With parents	0	0	.	.	.	.	.	-0.3589	1.1171
Seniority	0.5 - 1 Year	1	0.9633	0.1328	52.59	<.0001		2.620	0.3284	2.1980
Seniority	1 - 2 Years	1	0.6825	0.1215	31.54	<.0001		1.979	3.8034	4.9771
Seniority	3 - 5 Years	1	0.3167	0.1297	5.96	0.0147		1.373	-2.9398	-1.8638

Appendix 28-2008-Logistic Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
10105.006	5808.379	4296.6269	38	<.0001

Appendix 29-2008-LR Test Logistic Regression(2)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-0.9437	0.7148	1.74	0.1868		0.389	-2.3448	0.4573
Age		1	-0.0165	0.00560	8.71	0.0032	-0.0839	0.984	-0.0275	-0.00555
Education	High School	1	1.2196	0.0729	279.58	<.0001		3.386	1.0767	1.3626
Education	Primary scho	1	1.5908	0.2321	46.97	<.0001		4.908	1.1359	2.0457
Education	University	0	0	.	.	.	.	.	1.1359	2.0457
Expenses		1	0.00161	0.000120	181.71	<.0001	0.5534	1.002	0.00138	0.00185
Income		1	-0.00047	0.000028	272.10	<.0001	-1.0001	1.000	-0.00052	-0.00041
Industry	Agriculture	1	0.9420	0.8142	1.34	0.2473		2.565	-0.6537	2.5377
Industry	Bank and Financial Services	1	-0.1868	0.4957	0.14	0.7063		0.830	-1.1585	0.7848
Industry	Construction	1	0.3338	0.5026	0.44	0.5067		1.396	-0.6514	1.3189
Industry	Electronics / Pharmaceutical / O	1	-0.2909	0.5502	0.28	0.5970		0.748	-1.3693	0.7875
Industry	Food	1	-1.1921	0.7021	2.88	0.0895		0.304	-2.5682	0.1840
Industry	Gastronomie	1	0.4674	0.6722	0.48	0.4868		1.596	-0.8500	1.7849
Industry	Leather / Textile / Clothing	1	0.3068	0.6889	0.20	0.6560		1.359	-1.0433	1.6570
Industry	Other	1	0.1010	0.4736	0.05	0.8311		1.106	-0.8273	1.0293
Industry	Plastic / rubber / asbestos	1	-1.0052	1.2336	0.66	0.4151		0.366	-3.4230	1.4125
Industry	Public Service	1	-0.3638	0.4796	0.58	0.4482		0.695	-1.3038	0.5762
Industry	Retail	1	-0.5073	0.5351	0.90	0.3432		0.602	-1.5561	0.5416
Industry	Steel/metal processing	1	-0.8558	0.5637	2.31	0.1289		0.425	-1.9606	0.2489
Industry	Stones / earth / gas / ceramic	1	0.2191	0.5351	0.17	0.6822		1.245	-0.8297	1.2680
Industry	Wholesale	1	-0.1000	0.6371	0.02	0.8753		0.905	-1.3486	1.1486
Industry	Wood	0	0	.	.	.	.	.	-1.3486	1.1486
Interest_rate		1	0.1204	0.0371	10.50	0.0012	0.1548	1.128	0.0476	0.1932
loan_value_ron		1	4.416E-6	1.354E-6	10.63	0.0011	0.4047	1.000	1.761E-6	7.07E-6
Marital_Status	Divorced	1	1.0495	0.3496	9.01	0.0027		2.856	0.3643	1.7347
Marital_Status	Married	1	-0.9192	0.3288	7.81	0.0052		0.399	-1.5636	-0.2747
Marital_Status	Single	1	-0.1567	0.3390	0.21	0.6440		0.855	-0.8212	0.5078
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.8212	0.5078
Payment_ron		1	0.000665	0.000189	12.41	0.0004	0.4357	1.001	0.000295	0.00103
Product_id	CAR	1	-1.0726	0.1386	59.89	<.0001		0.342	-1.3442	-0.8009
Product_id	CONSUMER	1	-0.4077	0.1489	7.49	0.0062		0.665	-0.6996	-0.1158
Product_id	MORTGAGE	0	0	.	.	.	.	.	-0.6996	-0.1158
Repayment	Yes (In some cases warnings requ	1	4.6315	0.2144	466.77	<.0001		102.669	4.2113	5.0517
Repayment	Yes (never warned or deferred)	1	-2.1039	0.1861	127.84	<.0001		0.122	-2.4686	-1.7392
Repayment	so far no credit customer	0	0	.	.	.	.	.	-2.4686	-1.7392
Residence	Job apartment	1	-0.0708	0.5363	0.02	0.8949		0.932	-1.1219	0.9802
Residence	Own house	1	-1.3910	0.0878	251.01	<.0001		0.249	-1.5631	-1.2189
Residence	Rent	1	1.2041	0.1321	83.12	<.0001		3.334	0.9452	1.4629
Residence	With parents	0	0	.	.	.	.	.	0.9452	1.4629
Seniority	0.5 - 1 Year	1	0.9702	0.1320	54.05	<.0001		2.638	0.7115	1.2288
Seniority	1 - 2 Years	1	0.6929	0.1203	33.17	<.0001		2.000	0.4571	0.9287
Seniority	3 - 5 Years	1	0.3225	0.1290	6.25	0.0124		1.381	0.0697	0.5753
Seniority	6 - 10 Years	1	0.1677	0.1360	1.52	0.2174		1.183	-0.0988	0.4343
Seniority	< 6 Months	1	1.2243	0.1455	70.77	<.0001		3.402	0.9390	1.5095
Seniority	> 10 Years	0	0	.	.	.	.	.	0.9390	1.5095
Term		1	-0.00009	0.000035	6.69	0.0097	-0.1089	1.000	-0.00016	-0.00002

Appendix 30-2008-Logistic Regression Output(2)

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	1267.2824		<.0001
2	Residence		3	2	1014.8056		<.0001
3	Marital_Status		3	3	359.5553		<.0001
4	Education		2	4	375.0903		<.0001
5	Seniority		5	5	213.8530		<.0001
6	Product_id		2	6	86.8216		<.0001
7	Interest_rate		1	7	99.5847		<.0001
8	Payment_ron		1	8	61.0656		<.0001
9	Income		1	9	98.7456		<.0001
10	Expenses		1	10	204.0061		<.0001
11	Industry		14	11	46.2683		<.0001
12	CCY		2	12	13.2293		0.0013
13	loan_value_ron		1	13	13.8412		0.0002
14	Age		1	14	4.3547		0.0369
15	Profession		3	15	7.4778		0.0581
16	loan_value		1	16	1.9211		0.1657
17		loan_value	1	15		2.0084	0.1564

Appendix 31-2008-Stepwise Probit Regression (1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood				
Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
10105.006	5908.300	4196.7060	43	<.0001

Appendix 32-2008-LR Test Probit Regression(1)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-2.3134	0.4657	24.68	<.0001		-3.2261	-1.4007
Age		1	-0.00724	0.00302	5.75	0.0165	-0.0666	-0.0132	-0.00132
CCY	CHF	1	0.6175	0.1538	16.12	<.0001		0.3161	0.9190
CCY	EUR	1	0.4244	0.1117	14.45	0.0001		0.2056	0.6433
CCY	RON	0	0	.	.	.	.	0.2056	0.6433
Education	High School	1	0.6514	0.0395	272.50	<.0001		0.5741	0.7288
Education	Primary scho	1	0.8217	0.1304	39.72	<.0001		0.5662	1.0772
Education	University	0	0	.	.	.	.	0.5662	1.0772
Expenses		1	0.000629	0.000045	198.37	<.0001	0.3912	0.000542	0.000717
Income		1	-0.00020	0.000012	269.67	<.0001	-0.7827	-0.00023	-0.00018
Industry	Agriculture	1	0.2573	0.4489	0.33	0.5666		-0.6226	1.1372
Industry	Bank and Financial Services	1	-0.1687	0.2591	0.42	0.5151		-0.6766	0.3393
Industry	Construction	1	0.0821	0.2641	0.10	0.7561		-0.4356	0.5997
Industry	Electronics / Pharmaceutical / O	1	-0.3475	0.2902	1.43	0.2310		-0.9162	0.2212
Industry	Food	1	-0.8021	0.3909	4.21	0.0402		-1.5682	-0.0360
Industry	Gastronomie	1	0.1165	0.3734	0.10	0.7550		-0.6154	0.8484
Industry	Leather / Textile / Clothing	1	0.0241	0.3828	0.00	0.9499		-0.7263	0.7744
Industry	Other	1	-0.0469	0.2470	0.04	0.8496		-0.5310	0.4373
Industry	Plastic / rubber / asbestos	1	-0.6981	0.6890	1.03	0.3109		-2.0485	0.6523
Industry	Public Service	1	-0.2779	0.2501	1.24	0.2664		-0.7680	0.2122

Industry	Retail	1	-0.3702	0.2816	1.73	0.1886		-0.9222	0.1817
Industry	Steel/metal processing	1	-0.6068	0.2961	4.20	0.0404		-1.1872	-0.0265
Industry	Stones / earth / gas / ceramic	1	-0.2372	0.2959	0.64	0.4227		-0.8172	0.3427
Industry	Wholesale	1	-0.1298	0.3499	0.14	0.7107		-0.8156	0.5560
Industry	Wood	0	0	.	.	.	.	-0.8156	0.5560
Interest_rate		1	0.1963	0.0255	59.42	<.0001	0.4580	0.1464	0.2463
loan_value_ron		1	2.516E-6	6.906E-7	13.28	0.0003	0.4183	1.163E-6	3.87E-6
Marital_Status	Divorced	1	0.6862	0.1965	12.19	0.0005		0.3010	1.0714
Marital_Status	Married	1	-0.3729	0.1839	4.11	0.0426		-0.7333	-0.0125
Marital_Status	Single	1	0.0828	0.1893	0.19	0.6616		-0.2881	0.4538
Marital_Status	Widowed	0	0	.	.	.	.	-0.2881	0.4538
Payment_ron		1	0.000219	0.000093	5.50	0.0190	0.2600	0.000036	0.000401
Product_id	CAR	1	-0.2931	0.1083	7.33	0.0068		-0.5053	-0.0809
Product_id	CONSUMER	1	-0.4165	0.0669	38.80	<.0001		-0.5476	-0.2855
Product_id	MORTGAGE	0	0	.	.	.	.	-0.5476	-0.2855
Profession	Employee	1	0.2894	0.1934	2.24	0.1346		-0.0897	0.6686
Profession	Own Employee	1	0.2108	0.2120	0.99	0.3199		-0.2046	0.6263
Profession	Unemployed	1	0.6588	0.2627	6.29	0.0121		0.1440	1.1737
Profession	Worker	0	0	.	.	.	.	0.1440	1.1737
Repayment	Yes (In some cases warnings requ	1	2.3159	0.1007	528.61	<.0001		2.1185	2.5133
Repayment	Yes (never warned or deferred)	1	-1.1252	0.0880	163.47	<.0001		-1.2977	-0.9527
Repayment	so far no credit customer	0	0	.	.	.	.	-1.2977	-0.9527
Residence	Job apartment	1	-0.0552	0.2899	0.04	0.8490		-0.6234	0.5130
Residence	Own house	1	-0.7223	0.0467	239.11	<.0001		-0.8138	-0.6307
Residence	Rent	1	0.6919	0.0757	83.58	<.0001		0.5436	0.8403
Residence	With parents	0	0	.	.	.	.	0.5436	0.8403
Seniority	0.5 - 1 Year	1	0.5297	0.0708	55.94	<.0001		0.3909	0.6685
Seniority	1 - 2 Years	1	0.3581	0.0632	32.06	<.0001		0.2341	0.4820
Seniority	3 - 5 Years	1	0.1602	0.0669	5.73	0.0167		0.0290	0.2914
Seniority	6 - 10 Years	1	0.0592	0.0712	0.69	0.4054		-0.0803	0.1987
Seniority	< 6 Months	1	0.6865	0.0792	75.16	<.0001		0.5313	0.8417
Seniority	> 10 Years	0	0	.	.	.	.	0.5313	0.8417

Appendix 33-2008-Probit Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
10105.006	5929.508	4175.4976	38	<.0001

Appendix 34-2008-LR Test Probit Regression(2)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-0.6833	0.3790	3.25	0.0714		-1.4261	0.0595
Age		1	-0.00640	0.00293	4.79	0.0287	-0.0589	-0.0121	-0.00067
Education	High School	1	0.6392	0.0392	265.59	<.0001		0.5623	0.7161
Education	Primary scho	1	0.8257	0.1288	41.07	<.0001		0.5732	1.0782
Education	University	0	0	.	.	.	.	0.5732	1.0782
Expenses		1	0.000629	0.000045	199.41	<.0001	0.3911	0.000542	0.000716
Income		1	-0.00020	0.000012	273.45	<.0001	-0.7827	-0.00023	-0.00018
Industry	Agriculture	1	0.4189	0.4385	0.91	0.3394		-0.4406	1.2783
Industry	Bank and Financial Services	1	-0.1996	0.2583	0.60	0.4397		-0.7057	0.3066
Industry	Construction	1	0.0845	0.2639	0.10	0.7489		-0.4328	0.6018
Industry	Electronics / Pharmaceutical / O	1	-0.2691	0.2881	0.87	0.3502		-0.8338	0.2955
Industry	Food	1	-0.7915	0.3903	4.11	0.0426		-1.5564	-0.0265
Industry	Gastronomie	1	0.1496	0.3711	0.16	0.6869		-0.5779	0.8770
Industry	Leather / Textile / Clothing	1	0.00370	0.3828	0.00	0.9923		-0.7466	0.7540
Industry	Other	1	-0.0342	0.2468	0.02	0.8897		-0.5179	0.4494
Industry	Plastic / rubber / asbestos	1	-0.6846	0.6877	0.99	0.3195		-2.0325	0.6633
Industry	Public Service	1	-0.2719	0.2498	1.18	0.2764		-0.7614	0.2177
Industry	Retail	1	-0.3572	0.2812	1.61	0.2041		-0.9084	0.1941
Industry	Steel/metal processing	1	-0.4721	0.2876	2.70	0.1006		-1.0357	0.0914
Industry	Stones / earth / gas / ceramic	1	-0.0562	0.2815	0.04	0.8417		-0.6080	0.4955
Industry	Wholesale	1	-0.0941	0.3462	0.07	0.7858		-0.7727	0.5846
Industry	Wood	0	0	.	.	.	.	-0.7727	0.5846
Interest_rate		1	0.0762	0.0202	14.30	0.0002	0.1779	0.0367	0.1158
loan_value_ron		1	2.511E-6	6.915E-7	13.19	0.0003	0.4175	1.156E-6	3.867E-6
Marital_Status	Divorced	1	0.6440	0.1941	11.01	0.0009		0.2636	1.0243
Marital_Status	Married	1	-0.4146	0.1811	5.24	0.0221		-0.7696	-0.0596
Marital_Status	Single	1	0.0427	0.1868	0.05	0.8193		-0.3235	0.4089
Marital_Status	Widowed	0	0	.	.	.	.	-0.3235	0.4089
Payment_ron		1	0.000220	0.000094	5.54	0.0185	0.2618	0.000037	0.000404
Product_id	CAR	1	-0.6524	0.0753	75.06	<.0001		-0.8000	-0.5048
Product_id	CONSUMER	1	-0.2475	0.0815	9.23	0.0024		-0.4073	-0.0878
Product_id	MORTGAGE	0	0	.	.	.	.	-0.4073	-0.0878
Repayment	Yes (In some cases warnings requ	1	2.3098	0.1004	528.92	<.0001		2.1129	2.5066
Repayment	Yes (never warned or deferred)	1	-1.1144	0.0876	162.02	<.0001		-1.2860	-0.9428
Repayment	so far no credit customer	0	0	.	.	.	.	-1.2860	-0.9428
Residence	Job apartment	1	-0.0281	0.2860	0.01	0.9217		-0.5886	0.5323
Residence	Own house	1	-0.7253	0.0465	243.16	<.0001		-0.8165	-0.6341
Residence	Rent	1	0.6957	0.0755	84.93	<.0001		0.5478	0.8437
Residence	With parents	0	0	.	.	.	.	0.5478	0.8437
Seniority	0.5 - 1 Year	1	0.5313	0.0705	56.78	<.0001		0.3931	0.6695
Seniority	1 - 2 Years	1	0.3611	0.0626	33.28	<.0001		0.2384	0.4838
Seniority	3 - 5 Years	1	0.1609	0.0666	5.84	0.0156		0.0304	0.2914
Seniority	6 - 10 Years	1	0.0800	0.0697	1.32	0.2508		-0.0565	0.2165
Seniority	< 6 Months	1	0.6869	0.0791	75.50	<.0001		0.5319	0.8418
Seniority	> 10 Years	0	0	.	.	.	.	0.5319	0.8418
Term		1	-0.00005	0.000019	5.66	0.0173	-0.0985	-0.00008	-7.96E-6

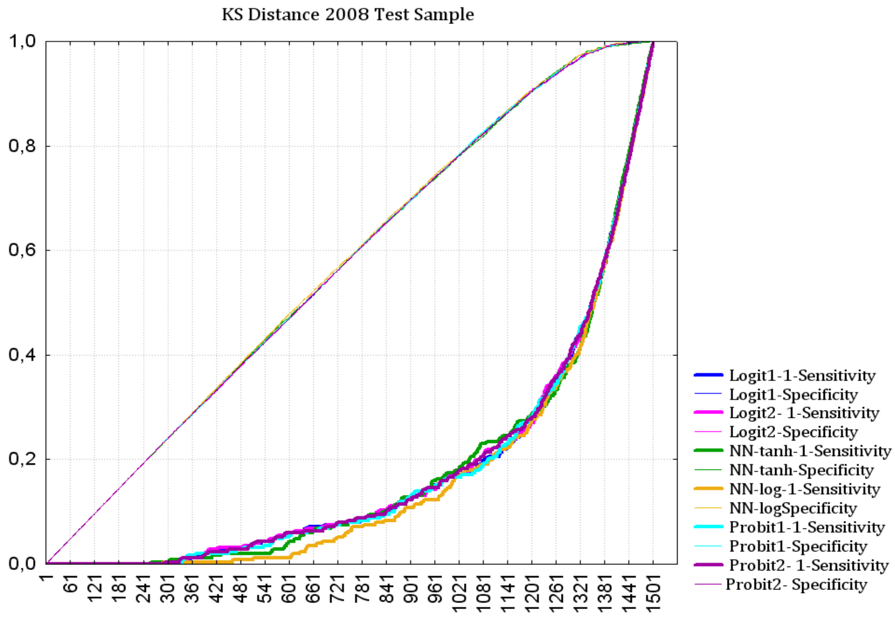
Appendix 35-2008-Probit Regression Output(2)

Neural Networks	Tahn	Logistic
Train: Akaike's Information Criterion	5104.87000	5031.93000
Train: Schwarz's Bayesian Criterion	5983.29000	5910.36000
Train: Average Error Function	0.23143	0.22796
Train: Error Function	4862.87000	4789.93000
Train: Misclassification Rate	0.08576	0.08443
Train: Number of Wrong Classifications	901.00000	887.00000
Valid: Average Error Function	0.24769	0.23676
Valid: Error Function	1487.12000	1421.52000
Valid: Mean Squared Error	0.06901	0.06614
Valid: Misclassification Rate	0.08394	0.07961
Valid: Number of Wrong Classifications	252	239
Test: Average Error Function	0.26263	0.24935
Test: Error Function	788.40600	748.55000
Test: Mean of Squared Error	0.07399	0.07159
Test: Misclassification Rate	0.08861	0.08994
Test: Number of Wrong Classifications	133	135

Appendix 36-2008-Neural Netoworks Results

		Confusion Matrix				Goodness of Fit						
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Misclass Rate	KS	AUROC	AR	Brier Score
Logit1	training	8287	797	1161	261	0.5930	0.9695	0.1007	0.6666	0.9072	0.8143	0.0789
Logit1	validation	2396	228	310	68	0.5762	0.9724	0.0986	0.6537	0.9002	0.8004	0.0787
Logit1	test	1200	104	148	49	0.5873	0.9608	0.1019	0.6480	0.8956	0.7913	0.0785
Logit 2	training	8284	812	1146	264	0.5853	0.9691	0.1024	0.6674	0.9065	0.8130	0.0793
Logit 2	validation	2394	236	302	70	0.5613	0.9716	0.1019	0.6436	0.9000	0.8000	0.0791
Logit 2	test	1196	102	150	53	0.5952	0.9576	0.1033	0.6427	0.8936	0.7872	0.0790
Probit 1	training	8297	856	1102	251	0.5628	0.9706	0.1054	0.6631	0.9050	0.8100	0.0816
Probit 1	validation	2399	243	295	65	0.5483	0.9736	0.1026	0.6351	0.8975	0.7950	0.0814
Probit 1	test	1204	109	143	45	0.5675	0.9640	0.1026	0.6439	0.8961	0.7921	0.0794
Probit2	training	8292	860	1098	256	0.5608	0.9701	0.1062	0.6636	0.9043	0.8086	0.0820
Probit2	validation	2398	246	292	66	0.5428	0.9732	0.1039	0.6314	0.8975	0.7949	0.0816
Probit2	test	1204	109	143	45	0.5675	0.9640	0.1026	0.6316	0.8935	0.7870	0.0798
NN1	training	8349	743	1215	199	0.6205	0.9767	0.0897	0.6891	0.9190	0.8379	0.0714
NN1	validation	2401	207	331	63	0.6152	0.9744	0.0899	0.6757	0.9024	0.8048	0.0741
NN1	test	1212	102	150	37	0.5952	0.9704	0.0926	0.6279	0.8957	0.7914	0.0763
NN2	training	8276	642	1316	272	0.6721	0.9682	0.0870	0.7165	0.9347	0.8694	0.0666
NN2	validation	2374	178	360	90	0.6691	0.9635	0.0893	0.7063	0.9175	0.8350	0.0704
NN2	test	1199	95	157	50	0.6230	0.9600	0.0966	0.6499	0.9104	0.8208	0.0754

Appendix 37-2008-Goodness of Fit Results



Appendix 38-2008-KS Distance Test Sample

Summary of Stepwise Selection						
Step	Effect Entered	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Age	1	1	796.4398		<.0001
2	Bank_r	8	2	59.0008		<.0001
3	CCY	2	3	18.1093		0.0001
4	Education	2	4	684.7794		<.0001
5	Expenses	1	5	1214.4073		<.0001
6	Income	1	6	57.7857		<.0001
7	Industry	14	7	607.3997		<.0001
8	Interest_rate	1	8	198.3950		<.0001
9	loan_value_ron	1	9	1027.7474		<.0001
10	Marital_Status	3	10	488.7051		<.0001
11	Payment_ron	1	11	697.9895		<.0001
12	Product_id	2	12	13.7783		0.0010
13	Repayment	2	13	2242.6978		<.0001
14	Residence	3	14	503.7747		<.0001
15	Seniority	5	15	218.3690		<.0001

Appendix 39-Portfolio-Stepwise Selection Logistic Regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood				
Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
19352.260	10942.483	8409.7771	47	<.0001

Appendix 40-Portfolio-LR Test Logistic Regression(1)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-3.2254	1.3597	5.63	0.0177		0.040	-5.8903	-0.5605
Age		1	-0.0150	0.00418	12.94	0.0003	-0.0746	0.985	-0.0232	-0.00684
Bank_r	0	1	1.0266	1.2486	0.68	0.4109		2.792	-1.4205	3.4738
Bank_r	1 -2 Year	1	0.2552	1.2624	0.04	0.8398		1.291	-2.2192	2.7295
Bank_r	1 Year	1	0.8626	1.2562	0.47	0.4923		2.369	-1.5995	3.3246
Bank_r	2 Years	1	1.0393	1.2673	0.67	0.4122		2.827	-1.4447	3.5232
Bank_r	2-3 Year	1	-0.0465	1.2960	0.00	0.9714		0.955	-2.5866	2.4935
Bank_r	3-5 Years	1	1.3350	1.2805	1.09	0.2971		3.800	-1.1746	3.8447
Bank_r	6-10 Year	1	-2.6111	1.8806	1.93	0.1650		0.073	-6.2970	1.0749
Bank_r	< 1 Year	1	0.9940	1.2469	0.64	0.4253		2.702	-1.4498	3.4379
Bank_r	> 10 Year	0	0	.	.	.	.	.	-1.4498	3.4379
CCY	CHF	1	0.4326	0.1468	8.69	0.0032		1.541	0.1450	0.7202
CCY	EUR	1	-0.3520	0.1004	12.29	0.0005		0.703	-0.5487	-0.1552
CCY	RON	0	0	.	.	.	.	.	-0.5487	-0.1552
Education	High School	1	1.0177	0.0537	359.41	<.0001		2.767	0.9125	1.1229
Education	Primary school	1	1.7605	0.1681	109.63	<.0001		5.815	1.4310	2.0901
Education	University	0	0	.	.	.	.	.	1.4310	2.0901
Expenses		1	0.00201	0.000074	744.30	<.0001	0.7560	1.002	0.00187	0.00215
Income		1	-0.00099	0.000038	684.55	<.0001	-1.8114	0.999	-0.00107	-0.00092
Industry	Agriculture	1	1.5829	0.4746	11.12	0.0009		4.869	0.6527	2.5132
Industry	Bank and Financial Services	1	0.2725	0.3552	0.59	0.4429		1.313	-0.4236	0.9686
Industry	Construction	1	0.9405	0.3573	6.93	0.0085		2.561	0.2402	1.6408
Industry	Electronics / Pharmaceutical / O	1	0.3816	0.3774	1.02	0.3120		1.465	-0.3582	1.1213
Industry	Food	1	0.6276	0.3975	2.49	0.1144		1.873	-0.1515	1.4067
Industry	Gastronomie	1	1.7356	0.4513	14.79	0.0001		5.672	0.8511	2.6201
Industry	Leather / Textile / Clothing	1	0.5501	0.4212	1.71	0.1915		1.734	-0.2753	1.3756
Industry	Other	1	0.5193	0.3378	2.36	0.1242		1.681	-0.1427	1.1814
Industry	Plastic / rubber / asbestos	1	0.1181	0.6108	0.04	0.8467		1.125	-1.0790	1.3151
Industry	Public Service	1	-0.1999	0.3409	0.34	0.5576		0.819	-0.8680	0.4682
Industry	Retail	1	0.2600	0.3673	0.50	0.4791		1.297	-0.4600	0.9800
Industry	Steel/metal processing	1	-0.1760	0.3843	0.21	0.6469		0.839	-0.9291	0.5771
Industry	Stones / earth / gas / ceramic	1	0.2417	0.3961	0.37	0.5418		1.273	-0.5347	1.0181
Industry	Wholesale	1	0.1071	0.4219	0.06	0.7996		1.113	-0.7199	0.9341
Industry	Wood	0	0	.	.	.	.	.	-0.7199	0.9341
Interest_rate		1	0.0697	0.0272	6.58	0.0103	0.0861	1.072	0.0164	0.1229
loan_value_ron		1	-5.36E-6	8.549E-7	39.31	<.0001	-0.4243	1.000	-7.04E-6	-3.68E-6
Marital_Status	Divorce	1	0.8999	0.2722	10.93	0.0009		2.459	0.3665	1.4334
Marital_Status	Married	1	-0.6750	0.2572	6.89	0.0087		0.509	-1.1791	-0.1708
Marital_Status	Single	1	-0.1958	0.2641	0.55	0.4584		0.822	-0.7134	0.3218
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.7134	0.3218
Payment_ron		1	0.00317	0.000143	490.89	<.0001	1.7160	1.003	0.00289	0.00345
Product_Id	CAR	1	-0.7008	0.1290	29.51	<.0001		0.496	-0.9537	-0.4480
Product_Id	CONSUMER	1	-0.0391	0.0904	0.19	0.6652		0.962	-0.2164	0.1381
Product_Id	MORTGAGE	0	0	.	.	.	.	.	-0.2164	0.1381
Repayment	Yes (In some cases warnings requ	1	4.3002	0.1822	557.16	<.0001		73.715	3.9431	4.6573
Repayment	Yes (never warned or deferred)	1	-1.4962	0.1815	67.97	<.0001		0.224	-1.8519	-1.1405
Repayment	so far no credit customer	0	0	.	.	.	.	.	-1.8519	-1.1405
Residence	Job apartment	1	-0.1781	0.3886	0.21	0.6467		0.837	-0.9398	0.5836
Residence	Own house	1	-1.1310	0.0660	293.91	<.0001		0.323	-1.2603	-1.0017
Residence	Rent	1	0.8856	0.0965	84.20	<.0001		2.424	0.6964	1.0747
Residence	With parents	0	0	.	.	.	.	.	0.6964	1.0747
Seniority	0,5 - 1 Year	1	1.1771	0.0987	142.23	<.0001		3.245	0.9836	1.3705
Seniority	1 - 2 Years	1	0.8162	0.0922	78.41	<.0001		2.262	0.6355	0.9968
Seniority	3 - 5 Years	1	0.4894	0.0961	25.92	<.0001		1.631	0.3010	0.6778
Seniority	6 - 10 Years	1	0.2442	0.1045	5.46	0.0194		1.277	0.0395	0.4490
Seniority	< 6 Months	1	1.2086	0.1099	121.01	<.0001		3.349	0.9933	1.4239
Seniority	> 10 Years	0	0	.	.	.	.	.	0.9933	1.4239

Appendix 41-Portfolio -Logistic Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
19352.260	11041.857	8310.4027	46	<.0001

Appendix 42-Portfolio –LR Test Logistic regression(2)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-1.8361	1.3083	1.97	0.1605		0.159	-4.4003	0.7282
Age		1	-0.0152	0.00415	13.46	0.0002	-0.0757	0.985	-0.0234	-0.00710
Bank_r	0	1	0.8612	1.2099	0.51	0.4766		2.366	-1.5102	3.2327
Bank_r	1 -2 Year	1	0.0890	1.2239	0.01	0.9420		1.093	-2.3098	2.4879
Bank_r	1 Year	1	0.6732	1.2179	0.31	0.5804		1.960	-1.7138	3.0601
Bank_r	2 Years	1	0.8344	1.2298	0.46	0.4974		2.304	-1.5755	3.2444
Bank_r	2-3 Year	1	-0.2352	1.2572	0.03	0.8516		0.790	-2.6992	2.2289
Bank_r	3-5 Years	1	1.1554	1.2417	0.87	0.3521		3.175	-1.2783	3.5892
Bank_r	6-10 Year	1	-2.9076	1.8341	2.51	0.1129		0.055	-6.5023	0.6871
Bank_r	< 1 Year	1	0.7977	1.2082	0.44	0.5091		2.220	-1.5703	3.1658
Bank_r	> 10 Year	0	0	.	.	.	.	.	-1.5703	3.1658
Education	High School	1	1.0199	0.0535	363.26	<.0001		2.773	0.9150	1.1248
Education	Primary school	1	1.7579	0.1666	111.37	<.0001		5.800	1.4314	2.0844
Education	University	0	0	.	.	.	.	.	1.4314	2.0844
Expenses		1	0.00200	0.000073	748.71	<.0001	0.7522	1.002	0.00186	0.00214
Income		1	-0.00099	0.000038	694.92	<.0001	-1.8147	0.999	-0.00107	-0.00092
Industry	Agriculture	1	1.5270	0.4718	10.48	0.0012		4.604	0.6023	2.4517
Industry	Bank and Financial Services	1	0.2045	0.3517	0.34	0.5610		1.227	-0.4848	0.8938
Industry	Construction	1	0.8899	0.3537	6.33	0.0119		2.435	0.1968	1.5831
Industry	Electronics / Pharmaceutical / O	1	0.3239	0.3738	0.75	0.3861		1.383	-0.4086	1.0565
Industry	Food	1	0.6126	0.3935	2.42	0.1195		1.845	-0.1587	1.3838
Industry	Gastronomie	1	1.6589	0.4477	13.73	0.0002		5.254	0.7814	2.5365
Industry	Leather / Textile / Clothing	1	0.4911	0.4175	1.38	0.2395		1.634	-0.3272	1.3094
Industry	Public Service	1	-0.2363	0.3373	0.49	0.4835		0.790	-0.8973	0.4247
Industry	Retail	1	0.2236	0.3640	0.38	0.5390		1.251	-0.4898	0.9370
Industry	Steel/metal processing	1	-0.1795	0.3805	0.22	0.6370		0.836	-0.9253	0.5662
Industry	Stones / earth / gas / ceramic	1	0.1887	0.3923	0.23	0.6306		1.208	-0.5802	0.9575
Industry	Wholesale	1	0.0218	0.4196	0.00	0.9586		1.022	-0.8006	0.8442
Industry	Wood	0	0	.	.	.	.	.	-0.8006	0.8442
Interest_rate		1	-0.0287	0.0176	2.66	0.1027	-0.0354	0.972	-0.0631	0.00576
loan_value_ron		1	-2.49E-6	1.046E-6	5.66	0.0173	-0.1970	1.000	-4.54E-6	-4.39E-7
Marital_Status	Divorce	1	0.8692	0.2714	10.26	0.0014		2.385	0.3373	1.4011
Marital_Status	Married	1	-0.7012	0.2565	7.47	0.0063		0.496	-1.2040	-0.1985
Marital_Status	Single	1	-0.2431	0.2634	0.85	0.3560		0.784	-0.7592	0.2731
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.7592	0.2731
Payment_ron		1	0.00275	0.000166	276.42	<.0001	1.4906	1.003	0.00243	0.00308
Product_Id	CAR	1	-0.9645	0.1026	88.32	<.0001		0.381	-1.1656	-0.7633
Product_Id	CONSUMER	1	-0.0951	0.0864	1.21	0.2714		0.909	-0.2645	0.0743
Product_Id	MORTGAGE	0	0	.	.	.	.	.	-0.2645	0.0743
Repayment	Yes (In some cases warnings requ	1	4.2704	0.1801	562.35	<.0001		71.554	3.9175	4.6234
Repayment	Yes (never warned or deferred)	1	-1.5171	0.1789	71.91	<.0001		0.219	-1.8677	-1.1664
Repayment	so far no credit customer	0	0	.	.	.	.	.	-1.8677	-1.1664
Residence	Job apartment	1	-0.2159	0.3894	0.31	0.5793		0.806	-0.9790	0.5473
Residence	Own house	1	-1.1288	0.0657	294.97	<.0001		0.323	-1.2576	-1.0000
Residence	Rent	1	0.8434	0.0960	77.11	<.0001		2.324	0.6552	1.0316
Residence	With parents	0	0	.	.	.	.	.	0.6552	1.0316
Seniority	0,5 - 1 Year	1	1.1530	0.0981	138.06	<.0001		3.168	0.9607	1.3453
Seniority	1 - 2 Years	1	0.8001	0.0917	76.12	<.0001		2.226	0.6204	0.9798
Seniority	3 - 5 Years	1	0.4630	0.0956	23.44	<.0001		1.589	0.2755	0.6504
Seniority	6 - 10 Years	1	0.2390	0.1039	5.29	0.0215		1.270	0.0353	0.4427
Seniority	< 6 Months	1	1.1795	0.1093	116.45	<.0001		3.253	0.9652	1.3937
Seniority	> 10 Years	0	0	.	.	.	.	.	0.9652	1.3937
Term		1	-0.00006	0.000015	16.08	<.0001	-0.0959	1.000	-0.00009	-0.00003

Appendix 43-Logistic Regression Output(2) Portfolio

Summary of Stepwise Selection						
	Effect					
Step	Entered	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Age	1	1	788.5915		<.0001
2	Bank_r	8	2	60.0088		<.0001
3	CCY	2	3	16.4233		0.0003
4	Education	2	4	641.6762		<.0001
5	Expenses	1	5	804.9346		<.0001
6	Income	1	6	52.6949		<.0001
7	Industry	14	7	611.0073		<.0001
8	Interest_rate	1	8	201.5713		<.0001
9	loan_value_ron	1	9	809.5452		<.0001
10	Marital_Status	3	10	463.9509		<.0001
11	Payment_ron	1	11	618.6870		<.0001
12	Product_Id	2	12	16.3055		0.0003
13	Repayment	2	13	1913.7939		<.0001
14	Residence	3	14	483.0227		<.0001
15	Seniority	5	15	219.1454		<.0001

Appendix 44-Portfolio-Stepwise Selection Probit

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
19352.260	11163.676	8188.5836	47	<.0001

Appendix 45-Portfolio LR Test Probit Regression(1)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-1.9537	0.7119	7.53	0.0061		-3.3490	-0.5584
Age		1	-0.00726	0.00216	11.30	0.0008	-0.0654	-0.0115	-0.00303
Bank_r	0	1	0.5546	0.6506	0.73	0.3940		-0.7206	1.8299
Bank_r	1 -2 Year	1	0.1680	0.6570	0.07	0.7982		-1.1197	1.4557
Bank_r	1 Year	1	0.4439	0.6543	0.46	0.4976		-0.8386	1.7263
Bank_r	2 Years	1	0.5569	0.6589	0.71	0.3980		-0.7345	1.8484
Bank_r	2-3 Year	1	-0.0653	0.6745	0.01	0.9229		-1.3872	1.2566
Bank_r	3-5 Years	1	0.7018	0.6660	1.11	0.2920		-0.6036	2.0072
Bank_r	6-10 Year	1	-1.1544	1.0785	1.15	0.2844		-3.2682	0.9594
Bank_r	< 1 Year	1	0.4940	0.6502	0.58	0.4474		-0.7803	1.7683
Bank_r	> 10 Year	0	0	.	.	.	.	-0.7803	1.7683
CCY	CHF	1	0.2380	0.0777	9.38	0.0022		0.0857	0.3903
CCY	EUR	1	-0.1664	0.0538	9.56	0.0020		-0.2719	-0.0609
CCY	RON	0	0	.	.	.	.	-0.2719	-0.0609
Education	High School	1	0.5189	0.0284	334.71	<.0001		0.4633	0.5745
Education	Primary school	1	0.9335	0.0928	101.12	<.0001		0.7515	1.1154
Education	University	0	0	.	.	.	.	0.7515	1.1154
Expenses		1	0.000876	0.000031	808.89	<.0001	0.5976	0.000815	0.000936
Income		1	-0.00044	0.000017	718.17	<.0001	-1.4683	-0.00048	-0.00041
Industry	Agriculture	1	0.8765	0.2626	11.14	0.0008		0.3617	1.3912
Industry	Bank and Financial Services	1	0.2102	0.1950	1.16	0.2812		-0.1721	0.5924
Industry	Construction	1	0.5176	0.1975	6.87	0.0088		0.1305	0.9046
Industry	Electronics / Pharmaceutical / O	1	0.2230	0.2078	1.15	0.2830		-0.1842	0.6302

Industry	Gastronomie	1	1.0113	0.2569	15.49	<.0001		0.5077	1.5149
Industry	Leather / Textile / Clothing	1	0.3501	0.2311	2.30	0.1297		-0.1028	0.8031
Industry	Other	1	0.3158	0.1870	2.85	0.0913		-0.0508	0.6824
Industry	Plastic / rubber / asbestos	1	0.0772	0.3384	0.05	0.8196		-0.5860	0.7404
Industry	Public Service	1	-0.0568	0.1883	0.09	0.7631		-0.4258	0.3123
Industry	Retail	1	0.1537	0.2033	0.57	0.4497		-0.2448	0.5522
Industry	Steel/metal processing	1	-0.0457	0.2086	0.05	0.8264		-0.4546	0.3632
Industry	Stones / earth / gas / ceramic	1	0.1384	0.2137	0.42	0.5172		-0.2804	0.5572
Industry	Wholesale	1	0.1789	0.2326	0.59	0.4417		-0.2769	0.6347
Industry	Wood	0	0	.	.	.	.	-0.2769	0.6347
Interest_rate		1	0.0415	0.0144	8.32	0.0039	0.0931	0.0133	0.0697
loan_value_ron		1	-2.71E-6	4.279E-7	40.11	<.0001	-0.3891	-3.55E-6	-1.87E-6
Marital_Status	Divorce	1	0.5435	0.1476	13.56	0.0002		0.2542	0.8328
Marital_Status	Married	1	-0.2907	0.1385	4.41	0.0358		-0.5621	-0.0193
Marital_Status	Single	1	-0.0217	0.1423	0.02	0.8787		-0.3006	0.2572
Marital_Status	Widowed	0	0	.	.	.	.	-0.3006	0.2572
Payment_ron		1	0.00148	0.000068	473.58	<.0001	1.4534	0.00135	0.00161
Product_id	CAR	1	-0.4018	0.0676	35.38	<.0001		-0.5343	-0.2694
Product_id	CONSUMER	1	-0.0179	0.0476	0.14	0.7064		-0.1113	0.0754
Product_id	MORTGAGE	0	0	.	.	.	.	-0.1113	0.0754
Repayment	Yes (In some cases warnings requ	1	2.2593	0.0888	647.78	<.0001		2.0853	2.4333
Repayment	Yes (never warned or deferred)	1	-0.7350	0.0849	74.89	<.0001		-0.9015	-0.5685
Repayment	so far no credit customer	0	0	.	.	.	.	-0.9015	-0.5685
Residence	Job apartment	1	-0.1200	0.2055	0.34	0.5593		-0.5228	0.2828
Residence	Own house	1	-0.5710	0.0339	283.82	<.0001		-0.6374	-0.5046
Residence	Rent	1	0.5038	0.0544	85.63	<.0001		0.3971	0.6106
Residence	With parents	0	0	.	.	.	.	0.3971	0.6106
Seniority	0,5 - 1 Year	1	0.6230	0.0514	146.87	<.0001		0.5222	0.7237
Seniority	1 - 2 Years	1	0.4137	0.0468	78.10	<.0001		0.3220	0.5055
Seniority	3 - 5 Years	1	0.2354	0.0487	23.36	<.0001		0.1399	0.3308
Seniority	6 - 10 Years	1	0.1221	0.0517	5.59	0.0181		0.0209	0.2234
Seniority	< 6 Months	1	0.6411	0.0584	120.47	<.0001		0.5266	0.7555
Seniority	> 10 Years	0	0	.	.	.	.	0.5266	0.7555

Appendix 46-Portfolio-Probit Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
19352.260	11253.997	8098.2628	46	<.0001

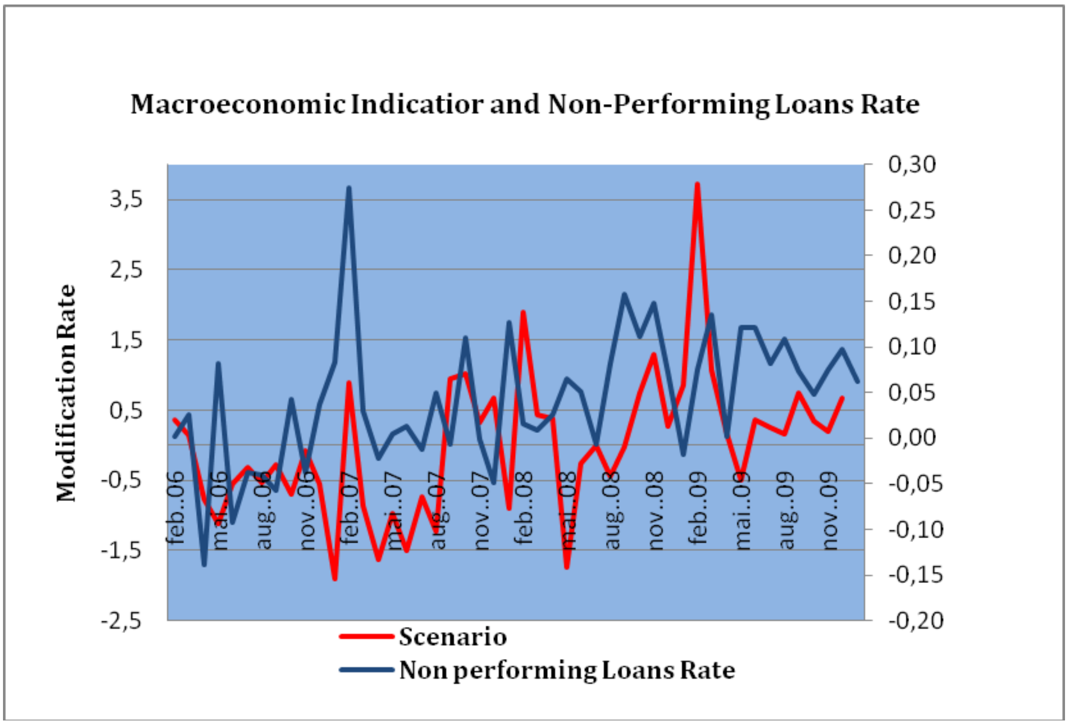
Appendix 47-Portfolio-LR Test Probit Regression(2)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-1.2039	0.6881	3.06	0.0802		-2.5526	0.1449
Age		1	-0.00747	0.00215	12.03	0.0005	-0.0673	-0.0117	-0.00325
Bank_r	0	1	0.4752	0.6341	0.56	0.4536		-0.7676	1.7180
Bank_r	1 -2 Year	1	0.1073	0.6406	0.03	0.8669		-1.1483	1.3629
Bank_r	1 Year	1	0.3548	0.6379	0.31	0.5781		-0.8955	1.6050
Bank_r	2 Years	1	0.4754	0.6427	0.55	0.4595		-0.7843	1.7350
Bank_r	2-3 Year	1	-0.1314	0.6585	0.04	0.8418		-1.4220	1.1591
Bank_r	3-5 Years	1	0.6297	0.6494	0.94	0.3322		-0.6432	1.9026
Bank_r	6-10 Year	1	-1.3040	1.0507	1.54	0.2146		-3.3634	0.7554
Bank_r	< 1 Year	1	0.4085	0.6336	0.42	0.5191		-0.8334	1.6504
Bank_r	> 10 Year	0	0	.	.	.	.	-0.8334	1.6504
Education	High School	1	0.5224	0.0283	340.74	<.0001		0.4669	0.5779
Education	Primary school	1	0.9289	0.0925	100.84	<.0001		0.7476	1.1102
Education	University	0	0	.	.	.	.	0.7476	1.1102
Expenses		1	0.000877	0.000031	818.61	<.0001	0.5987	0.000817	0.000937
Income		1	-0.00045	0.000016	734.89	<.0001	-1.4829	-0.00048	-0.00041
Industry	Agriculture	1	0.8478	0.2610	10.55	0.0012		0.3363	1.3593
Industry	Bank and Financial Services	1	0.1716	0.1928	0.79	0.3735		-0.2063	0.5495
Industry	Construction	1	0.4907	0.1953	6.31	0.0120		0.1080	0.8735
Industry	Electronics / Pharmaceutical / O	1	0.1961	0.2055	0.91	0.3399		-0.2066	0.5988
Industry	Food	1	0.3776	0.2183	2.99	0.0837		-0.0503	0.8055
Industry	Gastronomie	1	0.9734	0.2549	14.58	0.0001		0.4737	1.4731
Industry	Leather / Textile / Clothing	1	0.3171	0.2290	1.92	0.1661		-0.1317	0.7660
Industry	Other	1	0.2953	0.1847	2.55	0.1099		-0.0668	0.6574
Industry	Plastic / rubber / asbestos	1	0.0710	0.3338	0.05	0.8316		-0.5833	0.7252
Industry	Public Service	1	-0.0732	0.1860	0.15	0.6939		-0.4377	0.2914
Industry	Retail	1	0.1358	0.2012	0.46	0.4997		-0.2586	0.5302
Industry	Steel/metal processing	1	-0.0441	0.2061	0.05	0.8306		-0.4481	0.3599
Industry	Stones / earth / gas / ceramic	1	0.1213	0.2113	0.33	0.5658		-0.2928	0.5355
Industry	Wholesale	1	0.1377	0.2309	0.36	0.5509		-0.3148	0.5902
Industry	Wood	0	0	.	.	.	.	-0.3148	0.5902
Interest_rate		1	-0.0142	0.00938	2.29	0.1298	-0.0319	-0.0326	0.00417
loan_value_ron		1	-1.38E-6	5.187E-7	7.13	0.0076	-0.1988	-2.4E-6	-3.68E-7
Marital_Status	Divorce	1	0.5246	0.1469	12.74	0.0004		0.2366	0.8126
Marital_Status	Married	1	-0.3075	0.1378	4.98	0.0256		-0.5775	-0.0374
Marital_Status	Single	1	-0.0494	0.1416	0.12	0.7272		-0.3269	0.2281
Marital_Status	Widowed	0	0	.	.	.	.	-0.3269	0.2281
Payment_ron		1	0.00130	0.000080	264.94	<.0001	1.2728	0.00114	0.00145
Product_id	CAR	1	-0.5454	0.0543	100.95	<.0001		-0.6518	-0.4390
Product_id	CONSUMER	1	-0.0294	0.0457	0.41	0.5197		-0.1189	0.0601
Product_id	MORTGAGE	0	0	.	.	.	.	-0.1189	0.0601
Repayment	Yes (In some cases warnings requ	1	2.2424	0.0882	646.08	<.0001		2.0695	2.4153
Repayment	Yes (never warned or deferred)	1	-0.7501	0.0846	78.68	<.0001		-0.9159	-0.5844
Repayment	so far no credit customer	0	0	.	.	.	.	-0.9159	-0.5844
Residence	Job apartment	1	-0.1414	0.2055	0.47	0.4914		-0.5441	0.2613
Residence	Own house	1	-0.5704	0.0338	285.13	<.0001		-0.6366	-0.5042
Residence	Rent	1	0.4846	0.0542	79.91	<.0001		0.3783	0.5908
Residence	With parents	0	0	.	.	.	.	0.3783	0.5908
Seniority	0,5 - 1 Year	1	0.6108	0.0512	142.38	<.0001		0.5105	0.7112
Seniority	1 - 2 Years	1	0.4059	0.0466	75.76	<.0001		0.3145	0.4973
Seniority	3 - 5 Years	1	0.2241	0.0485	21.33	<.0001		0.1290	0.3191
Seniority	6 - 10 Years	1	0.1198	0.0515	5.42	0.0199		0.0189	0.2207
Seniority	< 6 Months	1	0.6260	0.0582	115.79	<.0001		0.5119	0.7400
Seniority	> 10 Years	0	0	.	.	.	.	0.5119	0.7400
Term		1	-0.00003	7.865E-6	15.29	<.0001	-0.0898	-0.00005	-0.00002

Appendix 48-Portfolio-Probit Regression Output(2)

		Confusion Matrix				Goodness of Fit						
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Misclass	KS	AUROC	AR	Brier Score
Logit1	training	19534	1580	1814	397	0.5345	0.9801	0.0848	0.6674	0.9115	0.8229	0.0665
Logit1	validation	5527	484	538	115	0.5264	0.9796	0.0899	0.6730	0.9123	0.8245	0.0695
Logit1	test	2755	224	295	58	0.5684	0.9794	0.0846	0.6739	0.9042	0.8084	0.0694
Logit 2	training	19522	1598	1796	409	0.5292	0.9795	0.0860	0.6649	0.9100	0.8200	0.0674
Logit 2	validation	5511	487	535	131	0.5235	0.9768	0.0927	0.6693	0.9102	0.8204	0.0705
Logit 2	test	2751	229	290	62	0.5588	0.9780	0.0873	0.6608	0.9034	0.8067	0.0704
Probit 1	training	19612	1693	1701	319	0.5012	0.9840	0.0863	0.6640	0.9108	0.8216	0.0683
Probit 1	validation	5539	519	503	103	0.4922	0.9817	0.0933	0.6710	0.9114	0.8228	0.0714
Probit 1	test	2763	249	270	50	0.5202	0.9822	0.0897	0.6658	0.9038	0.8075	0.0712
Probit2	training	19575	1738	1656	356	0.4879	0.9821	0.0898	0.6636	0.9094	0.8187	0.0690
Probit2	validation	5533	535	487	109	0.4765	0.9807	0.0966	0.6661	0.9095	0.8190	0.0721
Probit2	test	2758	253	266	55	0.5125	0.9804	0.0924	0.6569	0.9030	0.8061	0.0720
NN1	training	19472	1382	2012	459	0.5928	0.9770	0.0789	0.6815	0.9205	0.8410	0.0624
NN1	validation	5495	398	624	147	0.6106	0.9739	0.0818	0.6794	0.9159	0.8318	0.0653
NN1	test	2742	187	332	71	0.6397	0.9748	0.0774	0.6844	0.9129	0.8259	0.0650
NN2	training	19518	1367	2027	413	0.5972	0.9793	0.0763	0.6879	0.9273	0.8546	0.0600
NN2	validation	5498	414	608	144	0.5949	0.9745	0.0837	0.6958	0.9193	0.8386	0.0648
NN2	test	2740	197	322	73	0.6204	0.9740	0.0810	0.6851	0.9157	0.8314	0.0650

Appendix 49-Portfolio -Goodness of fit Test



Appendix 50-Scenario Comparison

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	4228.1679		<.0001
2	Residence		3	2	1520.1693		<.0001
3	IMV_client2		1	3	1074.1177		<.0001
4	Education		2	4	778.9710		<.0001
5	Seniority		5	5	534.8750		<.0001
6	Marital_Status		3	6	329.1663		<.0001
7	Expenses		1	7	127.5968		<.0001
8	Income		1	8	175.0869		<.0001
9	Payment_ron		1	9	386.0895		<.0001
10	Industry		14	10	179.8572		<.0001
11	Profession		3	11	52.3166		<.0001
12	Interest_rate		1	12	41.5603		<.0001
13	CCY		2	13	105.1623		<.0001
14	loan_value_ron		1	14	21.7220		<.0001
15	Age		1	15	19.1818		<.0001
16	Product_id		2	16	13.2767		0.0013
17	Bank_r		8	17	23.7786		0.0025
18	Phone_id		2	18	3.3647		0.1859
19		Phone_id	2	17		3.3560	0.1867

Appendix 51-Portfolio Macro Stepwise Selection Logistic Regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood				
Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
19352.260	10622.993	8729.2664	51	<.0001

Appendix 52-Portfolio Macro LR Test Logistic Regression(1)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-3.6365	1.4412	6.37	0.0116		0.026	-6.4613	-0.8118
Age		1	-0.0184	0.00435	17.89	<.0001	-0.0914	0.982	-0.0269	-0.00986
Bank_r	0	1	1.2368	1.3198	0.88	0.3487		3.445	-1.3498	3.8235
Bank_r	1 -2 Year	1	0.3852	1.3324	0.08	0.7725		1.470	-2.2263	2.9966
Bank_r	1 Year	1	0.8984	1.3276	0.46	0.4986		2.456	-1.7037	3.5005
Bank_r	2 Years	1	0.9899	1.3396	0.55	0.4599		2.691	-1.6357	3.6155
Bank_r	2-3 Year	1	0.1228	1.3647	0.01	0.9283		1.131	-2.5520	2.7975
Bank_r	3-5 Years	1	1.4090	1.3495	1.09	0.2964		4.092	-1.2359	4.0540
Bank_r	6-10 Year	1	-3.0872	1.9253	2.57	0.1088		0.046	-6.8608	0.6864
Bank_r	< 1 Year	1	1.0789	1.3186	0.67	0.4132		2.942	-1.5056	3.6634
Bank_r	> 10 Year	0	0	.	.	.	.	.	-1.5056	3.6634
CCY	CHF	1	0.4109	0.1503	7.48	0.0063		1.508	0.1164	0.7054
CCY	EUR	1	-0.3018	0.1029	8.60	0.0034		0.739	-0.5035	-0.1001
CCY	RON	0	0	.	.	.	.	.	-0.5035	-0.1001
Education	High School	1	1.0232	0.0554	340.99	<.0001		2.782	0.9146	1.1318
Education	Primary school	1	1.7823	0.1713	108.29	<.0001		5.944	1.4466	2.1180
Education	University	0	0	.	.	.	.	.	1.4466	2.1180
Expenses		1	0.00176	0.000075	552.95	<.0001	0.6611	1.002	0.00161	0.00190
Income		1	-0.00074	0.000038	387.68	<.0001	-1.3555	0.999	-0.00082	-0.00067
Industry	Agriculture	1	1.1456	0.4939	5.38	0.0204		3.144	0.1777	2.1136
Industry	Bank and Financial Services	1	0.1522	0.3625	0.18	0.6745		1.164	-0.5583	0.8628
Industry	Construction	1	0.8826	0.3649	5.85	0.0156		2.417	0.1675	1.5977
Industry	Electronics / Pharmaceutical / O	1	0.1552	0.3860	0.16	0.6876		1.168	-0.6013	0.9118
Industry	Food	1	0.6426	0.4044	2.53	0.1120		1.901	-0.1499	1.4352
Industry	Gastronomie	1	1.7262	0.4641	13.83	0.0002		5.620	0.8166	2.6359
Industry	Leather / Textile / Clothing	1	0.5483	0.4317	1.61	0.2041		1.730	-0.2979	1.3945
Industry	Other	1	0.3430	0.3450	0.99	0.3201		1.409	-0.3331	1.0191
Industry	Plastic / rubber / asbestos	1	0.1202	0.6306	0.04	0.8488		1.128	-1.1158	1.3562
Industry	Public Service	1	-0.2823	0.3480	0.66	0.4172		0.754	-0.9644	0.3997
Industry	Retail	1	0.0880	0.3749	0.06	0.8143		1.092	-0.6467	0.8228
Industry	Steel/metal processing	1	-0.4362	0.3943	1.22	0.2685		0.646	-1.2090	0.3365
Industry	Stones / earth / gas / ceramic	1	-0.1864	0.4139	0.20	0.6525		0.830	-0.9977	0.6249
Industry	Wholesale	1	0.00433	0.4327	0.00	0.9920		1.004	-0.8438	0.8524
Industry	Wood	0	0	.	.	.	.	.	-0.8438	0.8524
Interest_rate		1	0.1146	0.0279	16.88	<.0001	0.1417	1.121	0.0599	0.1693
loan_value_ron		1	-4.46E-6	8.462E-7	27.75	<.0001	-0.3528	1.000	-6.12E-6	-2.8E-6
Marital_Status	Divorce	1	0.9881	0.2801	12.44	0.0004		2.686	0.4390	1.5371
Marital_Status	Married	1	-0.5761	0.2646	4.74	0.0295		0.562	-1.0947	-0.0575
Marital_Status	Single	1	-0.1376	0.2713	0.26	0.6120		0.871	-0.6694	0.3942
Marital_Status	Widowed	0	0	.	.	.	.	.	-0.6694	0.3942
Payment_ron		1	0.00231	0.000146	250.69	<.0001	1.2525	1.002	0.00203	0.00260
Product_id	CAR	1	-0.4709	0.1318	12.76	0.0004		0.624	-0.7293	-0.2125
Product_id	CONSUMER	1	-0.0664	0.0914	0.53	0.4676		0.936	-0.2456	0.1128
Product_id	MORTGAGE	0	0	.	.	.	.	.	-0.2456	0.1128
Profession	Employee	1	-0.9152	0.1461	39.26	<.0001		0.400	-1.2015	-0.6289
Profession	Own Empl	1	-0.9986	0.2069	23.29	<.0001		0.368	-1.4041	-0.5930
Profession	Unemploy	1	0.2601	0.3049	0.73	0.3936		1.297	-0.3375	0.8577
Profession	Worker	0	0	.	.	.	.	.	-0.3375	0.8577
Repayment	Yes (In some cases warnings requ	1	4.5481	0.1808	632.89	<.0001		94.453	4.1938	4.9024
Repayment	Yes (never warned or deferred)	1	-1.3006	0.1836	50.17	<.0001		0.272	-1.6605	-0.9407
Repayment	so far no credit customer	0	0	.	.	.	.	.	-1.6605	-0.9407
Residence	Job apartment	1	-0.1459	0.3895	0.14	0.7080		0.864	-0.9094	0.6176
Residence	Own house	1	-1.1237	0.0669	282.00	<.0001		0.325	-1.2549	-0.9926
Residence	Rent	1	0.8454	0.0989	73.10	<.0001		2.329	0.6516	1.0392
Residence	With parents	0	0	.	.	.	.	.	0.6516	1.0392
Seniority	0,5 - 1 Year	1	1.1516	0.1006	131.15	<.0001		3.163	0.9545	1.3486
Seniority	1 - 2 Years	1	0.7770	0.0938	68.56	<.0001		2.175	0.5931	0.9609
Seniority	3 - 5 Years	1	0.4652	0.0976	22.74	<.0001		1.592	0.2740	0.6564
Seniority	6 - 10 Years	1	0.1413	0.1071	1.74	0.1872		1.152	-0.0687	0.3512
Seniority	< 6 Months	1	1.1907	0.1120	113.02	<.0001		3.289	0.9712	1.4102
Seniority	> 10 Years	0	0	.	.	.	.	.	0.9712	1.4102
IMV_client2		1	5.1807	0.3196	262.72	<.0001	0.2641	177.810	4.5543	5.8072

Appendix 53-Portfolio Macro Logistic Regression Output(1)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
19352.260	10758.365	8593.8942	47	<.0001

Appendix 54-Portfolio Macro LR Test Logistic regression(2)

Analysis of Maximum Likelihood Estimates										
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)	95% Confidence Limits	
Intercept		1	-3.4977	1.3938	6.30	0.0121		0.030	-6.2295	-0.7658
Age		1	-0.0148	0.00422	12.30	0.0005	-0.0735	0.985	-0.0231	-0.00652
Bank_r	0	1	1.1096	1.2936	0.74	0.3910		3.033	-1.4258	3.6449
Bank_r	1 -2 Year	1	0.2647	1.3064	0.04	0.8394		1.303	-2.2958	2.8252
Bank_r	1 Year	1	0.7807	1.3021	0.36	0.5488		2.183	-1.7714	3.3327
Bank_r	2 Years	1	0.8701	1.3144	0.44	0.5080		2.387	-1.7060	3.4461
Bank_r	2-3 Year	1	-0.0260	1.3385	0.00	0.9845		0.974	-2.6495	2.5974
Bank_r	3-5 Years	1	1.2913	1.3235	0.95	0.3292		3.638	-1.3028	3.8854
Bank_r	6-10 Year	1	-3.4454	1.8992	3.29	0.0697		0.032	-7.1677	0.2769
Bank_r	< 1 Year	1	0.9531	1.2929	0.54	0.4610		2.594	-1.5808	3.4870
Bank_r	> 10 Year	0	0	.	.	.	.	.	-1.5808	3.4870
Education	High School	1	1.0782	0.0547	388.94	<.0001		2.939	0.9710	1.1853
Education	Primary school	1	1.8196	0.1687	116.39	<.0001		6.170	1.4891	2.1502
Education	University	0	0	.	.	.	.	.	1.4891	2.1502
Expenses		1	0.00175	0.000074	558.23	<.0001	0.6584	1.002	0.00160	0.00190
IMV_client2		1	5.2456	0.3172	273.46	<.0001	0.2674	189.728	4.6239	5.8673
Income		1	-0.00075	0.000038	393.64	<.0001	-1.3626	0.999	-0.00082	-0.00067
Industry	Agriculture	1	1.4556	0.4824	9.11	0.0025		4.287	0.5101	2.4011
Industry	Bank and Financial Services	1	0.1000	0.3585	0.08	0.7803		1.105	-0.6027	0.8027
Industry	Construction	1	0.8565	0.3611	5.63	0.0177		2.355	0.1488	1.5643
Industry	Electronics / Pharmaceutical / O	1	0.1995	0.3814	0.27	0.6010		1.221	-0.5481	0.9471
Industry	Food	1	0.6399	0.4008	2.55	0.1104		1.896	-0.1457	1.4256
Industry	Gastronomie	1	1.6804	0.4594	13.38	0.0003		5.368	0.7801	2.5807
Industry	Public service	1	-0.2815	0.3444	0.67	0.4136		0.755	-0.9565	0.3935
Industry	Retail	1	0.0858	0.3714	0.05	0.8174		1.090	-0.6421	0.8136
Industry	Steel/metal processing	1	-0.2566	0.3880	0.44	0.5084		0.774	-1.0171	0.5039
Industry	Stones / earth / gas / ceramic	1	0.1058	0.3994	0.07	0.7911		1.112	-0.6771	0.8887
Industry	Wholesale	1	-0.0113	0.4288	0.00	0.9789		0.989	-0.8518	0.8292
Industry	Wood	0	0	.	.	.	.	.	-0.8518	0.8292
Interest_rate		1	0.0298	0.0180	2.73	0.0983	0.0369	1.030	-0.00554	0.0652
loan_value_ron		1	-2.26E-6	1.038E-6	4.74	0.0294	-0.1789	1.000	-4.3E-6	-2.26E-7
Marital_status	Divorce	1	0.8606	0.2764	9.70	0.0018		2.365	0.3189	1.4022
Marital_status	Married	1	-0.6923	0.2608	7.04	0.0079		0.500	-1.2035	-0.1811
Marital_status	Single	1	-0.2590	0.2679	0.93	0.3337		0.772	-0.7841	0.2662
Marital_status	Widowed	0	0	.	.	.	.	.	-0.7841	0.2662
Payment_ron		1	0.00198	0.000168	139.54	<.0001	1.0727	1.002	0.00165	0.00231
Product_id	CAR	1	-0.6784	0.1052	41.56	<.0001		0.507	-0.8847	-0.4722
Product_id	CONSUMER	1	-0.0987	0.0869	1.29	0.2559		0.906	-0.2691	0.0716
Product_id	MORTGAGE	0	0	.	.	.	.	.	-0.2691	0.0716
Repayment	Yes (in some cases warnings requ	1	4.4837	0.1782	633.15	<.0001		88.564	4.1345	4.8330
Repayment	Yes (never warned or deferred)	1	-1.3424	0.1811	54.93	<.0001		0.261	-1.6974	-0.9874
Repayment	so far no credit customer	0	0	.	.	.	.	.	-1.6974	-0.9874
Residence	Job apartment	1	-0.2138	0.3903	0.30	0.5838		0.807	-0.9788	0.5511
Residence	Own house	1	-1.1147	0.0664	281.48	<.0001		0.328	-1.2449	-0.9845
Residence	Rent	1	0.8041	0.0980	67.35	<.0001		2.235	0.6120	0.9961
Residence	With parents	0	0	.	.	.	.	.	0.6120	0.9961
Seniority	0,5 - 1 Year	1	1.1479	0.0995	133.01	<.0001		3.152	0.9528	1.3430
Seniority	1 - 2 Years	1	0.7880	0.0926	72.37	<.0001		2.199	0.6065	0.9696
Seniority	3 - 5 Years	1	0.4619	0.0964	22.94	<.0001		1.587	0.2729	0.6509
Seniority	6 - 10 Years	1	0.2041	0.1050	3.78	0.0519		1.226	-0.00164	0.4098
Seniority	< 6 Months	1	1.1825	0.1110	113.54	<.0001		3.263	0.9650	1.4001
Seniority	> 10 Years	0	0	.	.	.	.	.	0.9650	1.4001
Term		1	-0.00004	0.000015	8.11	0.0044	-0.0692	1.000	-0.00007	-0.00001

Appendix 55-Portfolio Macro Logistic Regression Output(2)

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Repayment		2	1	2891.5042		<.0001
2	Residence		3	2	1475.7909		<.0001
3	IMV_client2		1	3	975.2489		<.0001
4	Education		2	4	708.6659		<.0001
5	Seniority		5	5	510.4239		<.0001
6	Marital_Status		3	6	310.6490		<.0001
7	Expenses		1	7	158.1529		<.0001
8	Income		1	8	167.3244		<.0001
9	Payment_ron		1	9	306.1311		<.0001
10	Industry		14	10	170.1825		<.0001
11	Interest_rate		1	11	44.8925		<.0001
12	CCY		2	12	106.5234		<.0001
13	Profession		3	13	46.4379		<.0001
14	loan_value_ron		1	14	22.8916		<.0001
15	Age		1	15	17.2471		<.0001
16	Product_id		2	16	16.2620		0.0003
17	Bank_r		8	17	25.6863		0.0012
18	County_ID		1	18	1.5571		0.2121
19		County_ID	1	17		1.5539	0.2126

Appendix 56-Portfolio Macro Stepwise Selection Probit Regression

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood				
Intercept Only	Intercept & Covariates	Likelihood Ratio Chi-Square	DF	Pr > ChiSq
19352.260	10832.701	8519.5583	51	<.0001

Appendix 57-Portfolio Macro-LR Test Probit Regression(1)

Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-2.1725	0.7504	8.38	0.0038		-3.6434	-0.7017
Age		1	-0.00888	0.00225	15.59	<.0001	-0.0801	-0.0133	-0.00447
Bank_r	0	1	0.6580	0.6817	0.93	0.3345		-0.6782	1.9941
Bank_r	1 -2 Year	1	0.2342	0.6880	0.12	0.7336		-1.1142	1.5825
Bank_r	1 Year	1	0.4625	0.6857	0.45	0.5000		-0.8814	1.8064
Bank_r	2 Years	1	0.5525	0.6906	0.64	0.4237		-0.8010	1.9061
Bank_r	2-3 Year	1	0.00390	0.7042	0.00	0.9956		-1.3762	1.3841
Bank_r	3-5 Years	1	0.7490	0.6964	1.16	0.2822		-0.6160	2.1139
Bank_r	6-10 Year	1	-1.4159	1.1096	1.63	0.2019		-3.5907	0.7589
Bank_r	< 1 Year	1	0.5486	0.6815	0.65	0.4208		-0.7871	1.8844
Bank_r	> 10 Year	0	0	.	.	.	.	-0.7871	1.8844
CCY	CHF	1	0.2320	0.0795	8.52	0.0035		0.0762	0.3878
CCY	EUR	1	-0.1453	0.0551	6.96	0.0084		-0.2533	-0.0373
CCY	RON	0	0	.	.	.	.	-0.2533	-0.0373
Education	High School	1	0.5155	0.0292	312.67	<.0001		0.4584	0.5727
Education	Primary school	1	0.9363	0.0941	98.97	<.0001		0.7518	1.1208
Education	University	0	0	.	.	.	.	0.7518	1.1208
Expenses		1	0.000745	0.000031	564.52	<.0001	0.5084	0.000684	0.000806
Income		1	-0.00033	0.000017	369.32	<.0001	-1.0864	-0.00036	-0.00029
Industry	Agriculture	1	0.6316	0.2702	5.47	0.0194		0.1021	1.1611
Industry	Bank and Financial Services	1	0.1396	0.1989	0.49	0.4829		-0.2503	0.5295
Industry	Construction	1	0.4978	0.2010	6.13	0.0133		0.1038	0.8919
Industry	Electronics / Pharmaceutical / O	1	0.1109	0.2118	0.27	0.6005		-0.3042	0.5260
Industry	Food	1	0.4011	0.2239	3.21	0.0732		-0.0377	0.8399
Industry	Gastronomie	1	0.9982	0.2618	14.54	0.0001		0.4851	1.5113
Industry	Leather / Textile / Clothing	1	0.3453	0.2358	2.14	0.1432		-0.1169	0.8075
Industry	Other	1	0.2203	0.1906	1.34	0.2478		-0.1533	0.5940
Industry	Plastic / rubber / asbestos	1	0.0948	0.3441	0.08	0.7829		-0.5796	0.7692
Industry	Public Service	1	-0.0960	0.1919	0.25	0.6169		-0.4720	0.2801
Industry	Retail	1	0.0640	0.2070	0.10	0.7572		-0.3417	0.4697
Industry	Steel/metal processing	1	-0.1884	0.2140	0.78	0.3787		-0.6079	0.2311
Industry	Stones / earth / gas / ceramic	1	-0.0745	0.2225	0.11	0.7376		-0.5106	0.3615
Industry	Wholesale	1	0.1440	0.2367	0.37	0.5429		-0.3199	0.6080
Industry	Wood	0	0	.	.	.	.	-0.3199	0.6080
Interest_rate		1	0.0636	0.0148	18.58	<.0001	0.1428	0.0347	0.0926
loan_value_ron		1	-2.41E-6	4.32E-7	31.25	<.0001	-0.3467	-3.26E-6	-1.57E-6
Marital_Status	Divorce	1	0.6155	0.1543	15.91	<.0001		0.3131	0.9179
Marital_Status	Married	1	-0.2104	0.1451	2.10	0.1470		-0.4948	0.0739
Marital_Status	Single	1	0.0310	0.1487	0.04	0.8350		-0.2605	0.3224
Marital_Status	Widowed	0	0	.	.	.	.	-0.2605	0.3224
Payment_ron		1	0.00108	0.000072	227.57	<.0001	1.0657	0.000944	0.00123
Product_id	CAR	1	-0.2789	0.0691	16.31	<.0001		-0.4143	-0.1436
Product_id	CONSUMER	1	-0.0311	0.0483	0.42	0.5192		-0.1257	0.0635
Product_id	MORTGAGE	0	0	.	.	.	.	-0.1257	0.0635
Profession	Employee	1	-0.4915	0.0834	34.77	<.0001		-0.6549	-0.3281
Profession	Own Empl	1	-0.5217	0.1117	21.81	<.0001		-0.7407	-0.3028
Profession	Unemploy	1	0.1197	0.1606	0.56	0.4562		-0.1951	0.4344
Profession	Worker	0	0	.	.	.	.	-0.1951	0.4344

Repayment	Yes (In some cases warnings requ	1	2.4069	0.0888	734.45	<.0001		2.2328	2.5809
Repayment	Yes (never warned or deferred)	1	-0.6482	0.0861	56.71	<.0001		-0.8169	-0.4795
Repayment	so far no credit customer	0	0	.	.	.	.	-0.8169	-0.4795
Residence	Job apartment	1	-0.1021	0.2068	0.24	0.6214		-0.5075	0.3032
Residence	Own house	1	-0.5676	0.0344	271.77	<.0001		-0.6351	-0.5001
Residence	Rent	1	0.4850	0.0553	76.98	<.0001		0.3766	0.5933
Residence	With parents	0	0	.	.	.	.	0.3766	0.5933
Seniority	0,5 - 1 Year	1	0.6092	0.0523	135.61	<.0001		0.5067	0.7117
Seniority	1 - 2 Years	1	0.3957	0.0477	68.74	<.0001		0.3022	0.4892
Seniority	3 - 5 Years	1	0.2243	0.0495	20.52	<.0001		0.1272	0.3213
Seniority	6 - 10 Years	1	0.0702	0.0534	1.73	0.1887		-0.0345	0.1748
Seniority	< 6 Months	1	0.6271	0.0595	111.14	<.0001		0.5105	0.7437
Seniority	> 10 Years	0	0	.	.	.	.	0.5105	0.7437
IMV_client2		1	2.8377	0.1702	278.12	<.0001	0.2624	2.5042	3.1712

Appendix 58-Portfolio Macro-Probit Regression Output(2)

Likelihood Ratio Test for Global Null Hypothesis: BETA=0				
-2 Log Likelihood		Likelihood Ratio Chi-Square	DF	Pr > ChiSq
Intercept Only	Intercept & Covariates			
19352.260	10964.259	8388.0004	47	<.0001

Appendix 59-Portfolio Macro –LR Test Probit Regression(2)

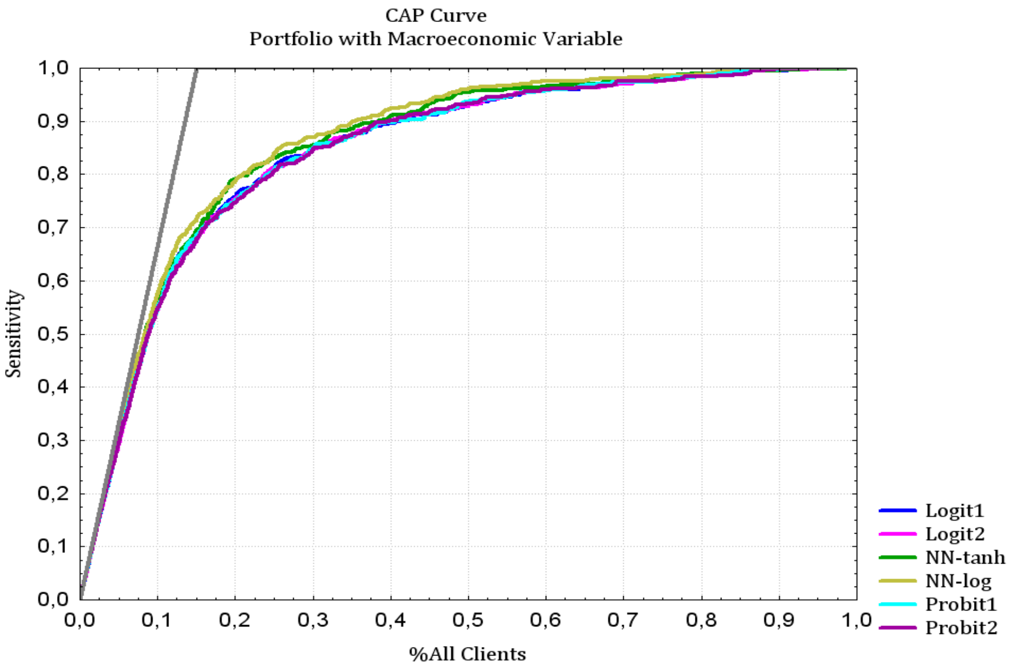
Analysis of Maximum Likelihood Estimates									
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	95% Confidence Limits	
Intercept		1	-2.0873	0.7300	8.18	0.0042		-3.5180	-0.6566
Age		1	-0.00684	0.00218	9.79	0.0018	-0.0616	-0.0111	-0.00255
Bank_r	0	1	0.5889	0.6745	0.76	0.3826		-0.7331	1.9108
Bank_r	1 -2 Year	1	0.1828	0.6809	0.07	0.7883		-1.1518	1.5174
Bank_r	1 Year	1	0.4045	0.6787	0.36	0.5512		-0.9257	1.7346
Bank_r	2 Years	1	0.5066	0.6837	0.55	0.4587		-0.8334	1.8465
Bank_r	2-3 Year	1	-0.0655	0.6973	0.01	0.9251		-1.4323	1.3013
Bank_r	3-5 Years	1	0.6962	0.6893	1.02	0.3124		-0.6547	2.0472
Bank_r	6-10 Year	1	-1.5989	1.0965	2.13	0.1448		-3.7480	0.5502
Bank_r	< 1 Year	1	0.4871	0.6745	0.52	0.4702		-0.8349	1.8090
Bank_r	> 10 Year	0	0	.	.	.	.	-0.8349	1.8090
Education	High School	1	0.5457	0.0287	360.57	<.0001		0.4893	0.6020
Education	Primary school	1	0.9533	0.0933	104.48	<.0001		0.7705	1.1362
Education	University	0	0	.	.	.	.	0.7705	1.1362
Expenses		1	0.000749	0.000031	573.46	<.0001	0.5110	0.000688	0.000810
IMV_client2		1	2.8475	0.1693	282.97	<.0001	0.2633	2.5158	3.1793
Income		1	-0.00033	0.000017	382.87	<.0001	-1.1077	-0.00037	-0.00030
Industry	Agriculture	1	0.8010	0.2653	9.12	0.0025		0.2811	1.3210
Industry	Bank and Financial Services	1	0.1116	0.1964	0.32	0.5701		-0.2734	0.4966
Industry	Construction	1	0.4806	0.1989	5.84	0.0157		0.0908	0.8704
Industry	Electronics / Pharmaceutical / O	1	0.1343	0.2092	0.41	0.5208		-0.2757	0.5444
Industry	Food	1	0.3985	0.2216	3.23	0.0721		-0.0358	0.8329
Industry	Gastronomie	1	0.9729	0.2597	14.04	0.0002		0.4639	1.4818

Industry	Leather / Textile / Clothing	1	0.3184	0.2336	1.86	0.1729		-0.1395	0.7763
Industry	Other	1	0.2217	0.1885	1.38	0.2394		-0.1476	0.5911
Industry	Plastic / rubber / asbestos	1	0.1047	0.3364	0.10	0.7555		-0.5545	0.7640
Industry	Public Service	1	-0.0962	0.1897	0.26	0.6119		-0.4680	0.2755
Industry	Retail	1	0.0598	0.2049	0.09	0.7704		-0.3419	0.4615
Industry	Steel/metal processing	1	-0.0811	0.2099	0.15	0.6991		-0.4925	0.3302
Industry	Stones / earth / gas / ceramic	1	0.0922	0.2149	0.18	0.6679		-0.3290	0.5134
Industry	Wholesale	1	0.1293	0.2348	0.30	0.5818		-0.3310	0.5896
Industry	Wood	0	0	.	.	.	.	-0.3310	0.5896
Interest_rate		1	0.0157	0.00963	2.65	0.1034	0.0352	-0.00319	0.0346
loan_value_ron		1	-1.44E-6	5.27E-7	7.49	0.0062	-0.2070	-2.47E-6	-4.09E-7
Marital_Status	Divorce	1	0.5380	0.1505	12.77	0.0004		0.2429	0.8330
Marital_Status	Married	1	-0.2814	0.1412	3.97	0.0463		-0.5582	-0.00456
Marital_Status	Single	1	-0.0401	0.1451	0.08	0.7823		-0.3244	0.2442
Marital_Status	Widowed	0	0	.	.	.	.	-0.3244	0.2442
Payment_ron		1	0.000946	0.000083	130.31	<.0001	0.9290	0.000783	0.00111
Product_id	CAR	1	-0.3950	0.0555	50.57	<.0001		-0.5039	-0.2861
Product_id	CONSUMER	1	-0.0347	0.0461	0.57	0.4519		-0.1249	0.0556
Product_id	MORTGAGE	0	0	.	.	.	.	-0.1249	0.0556
Repayment	Yes (In some cases warnings requ	1	2.3736	0.0880	726.90	<.0001		2.2011	2.5462
Repayment	Yes (never warned or deferred)	1	-0.6752	0.0855	62.35	<.0001		-0.8428	-0.5076
Repayment	so far no credit customer	0	0	.	.	.	.	-0.8428	-0.5076
Residence	Job apartment	1	-0.1360	0.2071	0.43	0.5114		-0.5419	0.2699
Residence	Own house	1	-0.5661	0.0342	274.65	<.0001		-0.6330	-0.4991
Residence	Rent	1	0.4648	0.0548	71.86	<.0001		0.3573	0.5723
Residence	With parents	0	0	.	.	.	.	0.3573	0.5723
Seniority	0,5 - 1 Year	1	0.6118	0.0519	139.01	<.0001		0.5101	0.7135
Seniority	1 - 2 Years	1	0.4045	0.0472	73.52	<.0001		0.3121	0.4970
Seniority	3 - 5 Years	1	0.2259	0.0490	21.22	<.0001		0.1298	0.3220
Seniority	6 - 10 Years	1	0.1076	0.0521	4.27	0.0388		0.00556	0.2097
Seniority	< 6 Months	1	0.6262	0.0590	112.51	<.0001		0.5105	0.7419
Seniority	> 10 Years	0	0	.	.	.	.	0.5105	0.7419
Term		1	-0.00002	8.017E-6	6.21	0.0127	-0.0583	-0.00004	-4.27E-6

Appendix 60-Portfolio Macro -Probit Regression Output (2)

		Confusion Matrix				Goodness of fit						
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Misclass	KS	AUROC	AR	Brier Score
Logit1	training	19546	1526	1868	385	0.55	0.981	0.082	0.676	0.916	0.832	0.064
Logit1	validation	5510	464	558	132	0.546	0.977	0.089	0.669	0.914	0.828	0.068
Logit1	test	2757	214	305	56	0.588	0.98	0.081	0.674	0.908	0.817	0.067
Logit 2	training	19547	1528	1866	384	0.55	0.981	0.082	0.671	0.914	0.827	0.065
Logit 2	validation	5512	472	550	130	0.538	0.977	0.09	0.664	0.911	0.822	0.069
Logit 2	test	2757	221	298	56	0.574	0.98	0.083	0.67	0.907	0.814	0.068
Probit 1	training	19593	1620	1774	338	0.523	0.983	0.084	0.674	0.915	0.83	0.066
Probit 1	validation	5533	492	530	109	0.519	0.981	0.09	0.668	0.913	0.825	0.069
Probit 1	test	2764	235	284	49	0.547	0.983	0.085	0.667	0.908	0.816	0.068
Probit2	training	19594	1639	1755	337	0.517	0.983	0.085	0.669	0.912	0.825	0.067
Probit2	validation	5529	501	521	113	0.51	0.98	0.092	0.662	0.91	0.819	0.071
Probit2	test	2765	236	283	48	0.545	0.983	0.085	0.666	0.907	0.814	0.069
NN1	training	19537	1296	2098	394	0.618	0.98	0.072	0.709	0.933	0.867	0.056
NN1	validation	5500	394	628	142	0.614	0.975	0.08	0.706	0.923	0.846	0.062
NN1	test	2747	196	323	66	0.622	0.977	0.079	0.704	0.92	0.84	0.061
NN2	training	19557	1294	2100	374	0.619	0.981	0.072	0.714	0.936	0.872	0.055
NN2	validation	5502	380	642	140	0.628	0.975	0.078	0.7	0.927	0.855	0.061
NN2	test	2754	181	338	59	0.651	0.979	0.072	0.704	0.928	0.856	0.059

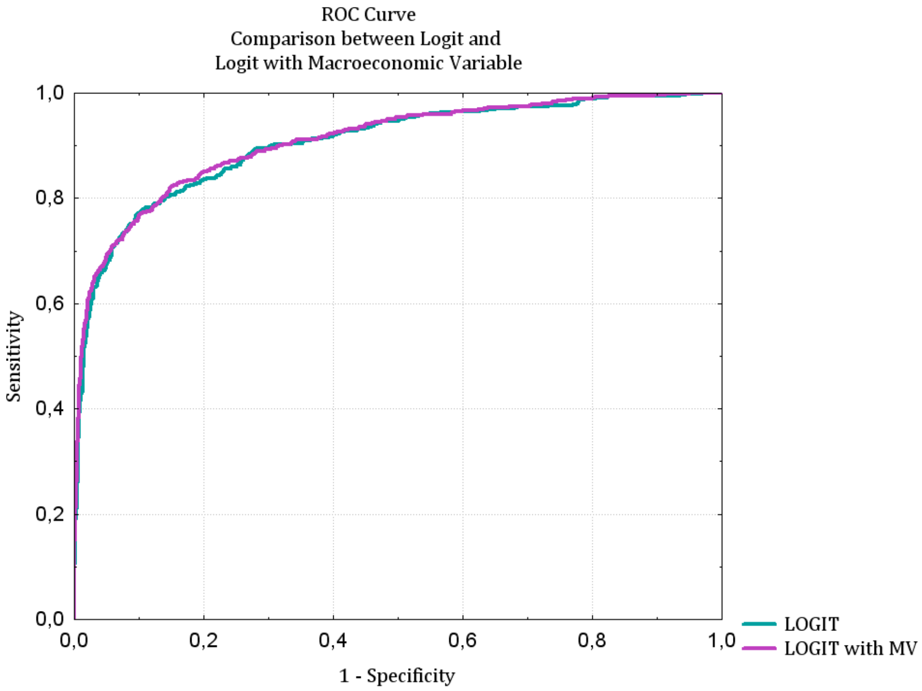
Appendix 61-Portfolio Macro Goodness of Fit Results



Appendix 62-Portfolio Macro CAP Curve

		Confusion Matrix				Goodness of Fit						
Model	**Sample**	**TN**	**FN**	**TP**	**FP**	**Sensitivity**	**Specificity**	**Misclass**	**KS**	**AUROC**	**AR**	**Brier Score**
Logit 2	test	0.22%	-3.49%	2.76%	-9.68%	2.76%	0.22%	-4.81%	1.40%	0.43%	0.96%	-3.72%
Logit 2	training	0.13%	-4.38%	3.90%	-6.11%	3.90%	0.13%	-4.73%	0.85%	0.39%	0.86%	-3.23%
Logit 2	validation	0.02%	-3.08%	2.80%	-0.76%	2.80%	0.02%	-2.59%	-0.74%	0.07%	0.15%	-1.89%
Logit1	test	0.07%	-4.46%	3.39%	-3.45%	3.39%	0.07%	-4.26%	0.03%	0.47%	1.04%	-4.01%
Logit1	training	0.06%	-3.42%	2.98%	-3.02%	2.98%	0.06%	-3.34%	1.30%	0.48%	1.07%	-3.47%
Logit1	validation	-0.31%	-4.13%	3.72%	14.78%	3.72%	-0.31%	-0.50%	-0.59%	0.19%	0.43%	-2.32%
NN1	test	0.18%	4.81%	-2.71%	-7.04%	-2.71%	0.18%	1.55%	2.88%	0.78%	1.73%	-5.39%
NN1	training	0.33%	-6.22%	4.27%	-14.16%	4.27%	0.33%	-8.20%	4.05%	1.40%	3.08%	-9.63%
NN1	validation	0.09%	-1.01%	0.64%	-3.40%	0.64%	0.09%	-1.65%	3.89%	0.80%	1.76%	-4.29%
NN2	test	0.51%	-8.12%	4.97%	-19.18%	4.97%	0.51%	-11.11%	2.72%	1.32%	2.91%	-9.63%
NN2	training	0.20%	-5.34%	3.60%	-9.44%	3.60%	0.20%	-6.29%	3.86%	0.94%	2.04%	-8.08%
NN2	validation	0.07%	-8.21%	5.59%	-2.78%	5.59%	0.07%	-6.81%	0.56%	0.87%	1.91%	-5.89%
Probit 1	test	0.04%	-5.62%	5.19%	-2.00%	5.19%	0.04%	-5.02%	0.21%	0.46%	1.03%	-4.22%
Probit 1	training	-0.10%	-4.31%	4.29%	5.96%	4.29%	-0.10%	-2.68%	1.47%	0.45%	0.99%	-3.49%
Probit 1	validation	-0.11%	-5.20%	5.37%	5.83%	5.37%	-0.11%	-3.38%	-0.40%	0.14%	0.30%	-2.64%
Probit2	test	0.25%	-6.72%	6.39%	-12.73%	6.39%	0.25%	-7.79%	1.44%	0.43%	0.96%	-3.81%
Probit2	training	0.10%	-5.70%	5.98%	-5.34%	5.98%	0.10%	-5.64%	0.82%	0.34%	0.77%	-3.20%
Probit2	validation	-0.07%	-6.36%	6.98%	3.67%	6.98%	-0.07%	-4.66%	-0.68%	0.01%	0.03%	-2.14%

Appendix 63-Macroeconomic improvement on portfolio models



Appendix 64-Portfolio Macro -Comparison Models

Model	Cut-off
Logit1	0.14346
Logit 2	0.14364
Probit 1	0.16261
Probit2	0.16093
NN1	0.12341
NN2	0.12821

Appendix 65-Portfolio Macro -Cut-off Values

Cut-off dynamic		Confusion Matrix				Goodness of Fit						
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Misclass Rate	KS	AUROC	AR	Brier Score
Logit1	training	16600	567	2827	3331	0.8329	0.8329	0.1671	0.6761	0.9159	0.8317	0.0642
Logit1	validation	4666	166	856	976	0.8376	0.8270	0.1714	0.6690	0.9140	0.8280	0.0679
Logit1	test	2350	88	431	463	0.8304	0.8354	0.1654	0.6741	0.9084	0.8168	0.0666
Logit 2	training	16559	574	2820	3372	0.8309	0.8308	0.1692	0.6705	0.9135	0.8270	0.0652
Logit 2	validation	4646	169	853	996	0.8346	0.8235	0.1748	0.6643	0.9108	0.8217	0.0691
Logit 2	test	2327	91	428	486	0.8247	0.8272	0.1732	0.6700	0.9072	0.8144	0.0677
Probit 1	training	16606	566	2828	3325	0.8332	0.8332	0.1668	0.6738	0.9149	0.8298	0.0659
Probit 1	validation	4672	166	856	970	0.8376	0.8281	0.1705	0.6683	0.9127	0.8253	0.0695
Probit 1	test	2349	92	427	464	0.8227	0.8351	0.1669	0.6671	0.9079	0.8159	0.0682
Probit2	training	16518	581	2813	3413	0.8288	0.8288	0.1712	0.6691	0.9125	0.8250	0.0668
Probit2	validation	4644	170	852	998	0.8337	0.8231	0.1753	0.6616	0.9096	0.8192	0.0706

Probit2	test	2328	93	426	485	0.8208	0.8276	0.1735	0.6664	0.9069	0.8138	0.0693
NN1	training	16993	500	2894	2938	0.8527	0.8526	0.1474	0.7092	0.9335	0.8669	0.0564
NN1	validation	4772	153	869	870	0.8503	0.8458	0.1535	0.7058	0.9232	0.8464	0.0625
NN1	Test	2391	85	434	422	0.8362	0.8500	0.1522	0.7042	0.9201	0.8401	0.0615
NN2	Trening	17019	495	2899	2912	0.8542	0.8539	0.1461	0.7144	0.9361	0.8721	0.0552
NN2	validation	4797	166	856	845	0.8376	0.8502	0.1517	0.6997	0.9273	0.8547	0.0610
NN2	Test	2404	79	440	409	0.8478	0.8546	0.1465	0.7037	0.9278	0.8556	0.0587

=5

Appendix 66-Portfolio Macro -Dynamic Cut-off

odel	ample	N	N	P	P	Sensitivity	Specificity	Miscalss Rate	KS	AUROC	AR	Brier Score
ogit1	raining	0.07%	0.95%	.83%	.27%	.83%	0.07%	0.10%	.00%	.00%	.00%	0.01%
ogit1	alidation	0.05%	1.24%	.12%	.61%	.12%	0.05%	0.50%	.00%	.00%	.00%	0.04%
ogit1	est	.11%	5.36%	.07%	5.17%	.07%	.11%	5.32%	0.02%	.48%	.06%	4.07%
ogit 2	raining	.09%	5.38%	.79%	4.16%	.79%	.09%	5.13%	.85%	.39%	.86%	3.24%
ogit 2	alidation	0.11%	3.29%	.99%	.58%	.99%	0.11%	1.62%	0.74%	.07%	.15%	1.89%
ogit 2	est	.15%	5.68%	.48%	6.45%	.48%	.15%	5.84%	.40%	.43%	.96%	3.70%
robit 1	raining	0.06%	1.54%	.53%	.76%	.53%	0.06%	0.70%	.00%	.00%	.00%	0.03%
robit 1	alidation	0.05%	0.77%	.80%	.91%	.80%	0.05%	0.16%	.00%	.00%	.00%	0.05%
robit 1	est	0.04%	8.43%	.78%	.00%	.78%	0.04%	6.69%	.21%	.47%	.05%	4.28%
robit2	raining	.04%	7.02%	.37%	1.97%	.37%	.04%	6.16%	.82%	.34%	.77%	3.21%
robit2	alidation	0.09%	6.92%	.60%	.59%	.60%	0.09%	4.97%	0.68%	.01%	.03%	2.15%
robit2	est	.25%	7.51%	.14%	12.73%	.14%	.25%	8.44%	.44%	.43%	.96%	3.80%
N1	raining	0.04%	10.13%	.96%	.53%	.96%	0.04%	7.22%	.26%	.87%	.09%	10.08%
N1	alidation	0.20%	5.28%	.37%	.48%	.37%	0.20%	1.83%	.67%	.63%	.58%	6.30%
N1	est	0.04%	0.53%	.30%	.41%	.30%	0.04%	.00%	.42%	.51%	.34%	5.02%
N2	raining	.23%	5.85%	.95%	10.65%	.95%	.23%	6.97%	.28%	.96%	.09%	7.70%
N2	alidation	.09%	7.97%	.43%	3.47%	.43%	.09%	6.81%	.89%	.95%	.08%	6.61%
N2	est	.47%	2.03%	.24%	17.81%	.24%	.47%	6.30%	.07%	.51%	.33%	8.58%

=10

odel	ample	N	N	P	P	Sensitivity	Specificity	Miscalss Rate	KS	AUROC	AR	Brier Score
ogit1	raining	.11%	.59%	2.26%	5.54%	2.26%	.11%	.96%	1.90%	0.46%	1.02%	.96%
ogit1	alidation	.00%	.31%	2.97%	.00%	2.97%	.00%	.67%	1.34%	0.32%	0.70%	.02%
ogit1	est	0.18%	.46%	3.39%	.62%	3.39%	0.18%	.32%	1.03%	0.52%	1.16%	.36%
ogit 2	raining	.09%	5.38%	.79%	4.16%	.79%	.09%	5.13%	.85%	.39%	.86%	3.24%
ogit 2	alidation	0.11%	3.29%	.99%	.58%	.99%	0.11%	1.62%	0.74%	.07%	.15%	1.89%
ogit 2	est	.15%	5.68%	.48%	6.45%	.48%	.15%	5.84%	.40%	.43%	.96%	3.70%
robit 1	raining	0.06%	1.54%	.53%	.76%	.53%	0.06%	0.70%	.00%	.00%	.00%	0.03%
robit 1	alidation	0.05%	0.77%	.80%	.91%	.80%	0.05%	0.16%	.00%	.00%	.00%	0.05%

robit 1	est	.14%	6.02%	.56%	8.00%	.56%	.14%	6.35%	0.26%	.31%	.70%	2.69%
robit2	raining	.04%	7.02%	.37%	1.97%	.37%	.04%	6.16%	.82%	.34%	.77%	3.21%
robit2	alidation	0.09%	6.92%	.60%	.59%	.60%	0.09%	4.97%	0.68%	.01%	.03%	2.15%
robit2	est	.25%	7.51%	.14%	12.73%	.14%	.25%	8.44%	.44%	.43%	.96%	3.80%
N1	raining	0.05%	10.13%	.96%	.96%	.96%	0.05%	7.12%	.26%	.86%	.07%	10.09%
N1	alidation	0.20%	5.28%	.37%	.48%	.37%	0.20%	1.83%	.79%	.62%	.57%	6.26%
N1	est	0.07%	0.53%	.30%	.82%	.30%	0.07%	.39%	.99%	.50%	.31%	5.00%
N2	raining	.15%	6.58%	.44%	7.26%	.44%	.15%	6.74%	.45%	.96%	.09%	7.56%
N2	alidation	0.02%	7.73%	.26%	.69%	.26%	0.02%	5.56%	.35%	.96%	.10%	6.49%
N2	est	.33%	3.05%	.86%	12.33%	.86%	.33%	5.56%	.85%	.52%	.35%	8.67%

=15

odel	ample	N	N	P	P	Sensitivity	Specificity	Miscalss Rate	KS	AUROC	AR	Brier Score
ogit1	raining	0.07%	0.95%	.83%	.27%	.83%	0.07%	0.10%	.00%	.00%	.00%	0.01%
ogit1	alidation	0.05%	1.24%	.12%	.61%	.12%	0.05%	0.50%	.00%	.00%	.00%	0.04%
ogit1	est	.11%	5.36%	.07%	5.17%	.07%	.11%	5.32%	0.02%	.48%	.06%	4.07%
ogit 2	raining	.09%	5.38%	.79%	4.16%	.79%	.09%	5.13%	.85%	.39%	.86%	3.24%
ogit 2	alidation	0.11%	3.29%	.99%	.58%	.99%	0.11%	1.62%	0.74%	.07%	.15%	1.89%
ogit 2	est	.15%	5.68%	.48%	6.45%	.48%	.15%	5.84%	.40%	.43%	.96%	3.70%
robit 1	raining	0.06%	1.54%	.53%	.76%	.53%	0.06%	0.70%	.00%	.00%	.00%	0.03%
robit 1	alidation	0.05%	0.77%	.80%	.91%	.80%	0.05%	0.16%	.00%	.00%	.00%	0.05%
robit 1	est	0.04%	8.43%	.78%	.00%	.78%	0.04%	6.69%	.21%	.47%	.05%	4.28%
robit2	raining	.04%	7.02%	.37%	1.97%	.37%	.04%	6.16%	.82%	.34%	.77%	3.21%
robit2	alidation	0.09%	6.92%	.60%	.59%	.60%	0.09%	4.97%	0.68%	.01%	.03%	2.15%
robit2	est	.25%	7.51%	.14%	12.73%	.14%	.25%	8.44%	.44%	.43%	.96%	3.80%
N1	raining	0.02%	10.35%	.11%	.65%	.11%	0.02%	7.60%	.26%	.88%	.12%	10.16%
N1	alidation	0.16%	4.27%	.72%	.12%	.72%	0.16%	1.47%	.58%	.65%	.64%	6.39%
N1	est	.07%	1.60%	.90%	2.82%	.90%	.07%	1.94%	.81%	.61%	.55%	5.35%
N2	raining	.15%	6.58%	.44%	7.26%	.44%	.15%	6.74%	.45%	.96%	.09%	7.56%
N2	alidation	0.02%	7.73%	.26%	.69%	.26%	0.02%	5.56%	.35%	.96%	.10%	6.49%
N2	est	.33%	3.05%	.86%	12.33%	.86%	.33%	5.56%	.85%	.52%	.35%	8.67%

Appendix 67-Cost Comparison -Model Improvement vs. Portfolio Results

Spiegelhalter Test	LOGIT1	LOGIT2	PROBIT1	PROBIT2	NN1	NN2
Port	0.7605	0.7330	0.1932	0.1773	0.8992	0.2235
Macro	0.7852	0.6458	0.1626	0.1133	0.2602	0.5743

Appendix 68-Spiegelhalter Test

Cost comparison			Confusion Matrix				Goodness of Fit			
Model	Sample	TN	FN	TP	FP	Sensitivity	Specificity	Kuipers Score	Granger Pesaran	p-value
Logit1	test	2758	212	307	55	0.5915	0.9804	0.5720	38.47	0.000
Logit 2	test	2755	216	303	58	0.5838	0.9794	0.5632	37.93	0.000
Probit 1	test	2762	228	291	51	0.5607	0.9819	0.5426	37.42	0.000
Probit2	test	2765	234	285	48	0.5491	0.9829	0.5321	37.13	0.000
NN1	test	2741	186	333	72	0.6416	0.9744	0.6160	39.46	0.000
NN2	test	2753	193	326	60	0.6281	0.9787	0.6068	39.69	0.000
Logit1	test	2750	234	285	63	0.5491	0.9776	0.5267	36.05	0.000
Logit 2	test	2755	216	303	58	0.5838	0.9794	0.5632	37.93	0.000
Probit 1	test	2767	234	285	46	0.5491	0.9836	0.5328	37.28	0.000
Probit2	test	2765	234	285	48	0.5491	0.9829	0.5321	37.13	0.000
NN1	test	2740	186	333	73	0.6416	0.9740	0.6157	39.40	0.000
NN2	test	2749	191	328	64	0.6320	0.9772	0.6092	39.58	0.000
Logit1	test	2758	212	307	55	0.5915	0.9804	0.5720	38.47	0.000
Logit 2	test	2755	216	303	58	0.5838	0.9794	0.5632	37.93	0.000
Probit 1	test	2762	228	291	51	0.5607	0.9819	0.5426	37.42	0.000
Probit2	test	2765	234	285	48	0.5491	0.9829	0.5321	37.13	0.000
NN1	test	2744	184	335	69	0.6455	0.9755	0.6209	39.82	0.000
NN2	test	2749	191	328	64	0.6320	0.9772	0.6092	39.58	0.000

Appendix 69-Cost Comparison Tests