Introduction, functionality and improvement of individual creditworthiness assessment in Central and Eastern Europe: the Bulgarian case

By
Simeonka Stoykova Manova

Dissertation

XVII Cycle of the
Ph.D. in Business Administration and Management
Bocconi University, Milan

Supervisor: Prof. Andrea Sironi
Other committee members: Prof. Barbara Alemanni
Prof. Giampaolo Gabbi

Introduction, functionality and improvement of individual creditworthiness assessment in Central and Eastern Europe: the Bulgarian case*

By

Simeonka Stoykova Manova**

* Special thanks to Ivan Cavinato without whom this research would have not been possible.
** E-mail: simeonkamanova@yahoo.com

На моето семейство за безграничната му любов и подкрепа по време на обучението ми в чужбина

(To my family for their infinite love and support during my stay abroad)

Introduction, functionality and improvement of individual creditworthiness assessment in Central and Eastern Europe: the Bulgarian case

Abstract: The starting point of this study is the role of foreign bank institutions in the introduction, application and development of individual creditworthiness assessment (scoring models and policy rules) and individual loan market into Central and Eastern Europe (CEE) banking systems. Its main objective is to show how a generic individual credit granting model is tested and a new statistical one is built up in markets where many obstacles (e.g. lack of long time history and clean data) make difficult, but not impossible, the development of real statistical credit scoring models. For this reason, I firstly outline the main problems (unavailable historical data, no or unreliable Credit Bureau, etc.) and peculiarities (post-communist environment, lack of retail banking market, legal differences) that one of the biggest Western European (WE) investors in the CEE banking sector industry encountered and had to take into consideration while implementing individual credit granting systems into all of its CEE affiliates. I further describe the WE and the CEE banks underwriting processes (with particular focus on the Bulgarian case) and explain why at present it is not possible to adopt the entire WE practice in CEE countries. The individual creditworthiness assessments of the WE bank in question and its CEE affiliates consist of two parts: credit scoring model and policy rules application. The credit-scoring model is developed on a sample of clients' data and includes *statistical scorecards* (WE) and *generic/judgemental scorecards* and *debt ratio* (CEE). Policy rules concern the credit policy of the bank and address applicant's age, minimum monthly income, adversely classified clients, loan-to-value ratio (LTV), etc. The review of the actual individual granting model and the development of new sophisticated (statistical) one are of prior importance for CEE where individual loans boom, resulting from stable or increasing GDP growth, is expected to lead to enlarged non-performing loans in the mid-term. I use 8274 consumer and mortgage loans from the Bulgarian bank's portfolio data in order to test the individual granting process that was firstly implemented in this CEE affiliate of the WE bank. In particular, I back test the quality and the stability of the existing scoring model (i.e. the discriminatory power of each variable it comprises) through Cumulative Accuracy Profiles (CAP) validation technique, the Gini coefficient, and the stability index respectively. I then examine the effects of the policy rules on banks' final decision and the overlapping area between rules and models. Subsequently, the insufficient level of the Gini coefficients for most of the variables (below 0.10), for the entire generic scorecards (0.32), and for the debt ratio (-0.08) leads to the development of new statistical version of the scoring model (scorecards and debt ratio). The new granting system is based on both practical experience and theoretical literature findings and recommendations. I run univariate and multivariate logit regressions and review the policy rules. I assess the goodness-of-fit of the new models through Wald test, Likelihood-ratio test, Hosmer-Lemeshow test, and leave-one-out cross validation methodology. Even if with significantly superior discriminatory power (0.46 for the new statistical scorecards and –18.16% for the new debt ratio), the new credit scoring model can be further improved after the development of Credit Bureau (CB) register and CB score. CB investigation has become a reliable source for fraud or negative records of individual credits applicants in the USA and WE, but its foundation, utilization, and efficiency have been just tackled in most of the CEE countries. I further discuss what implications the newly created individual creditworthiness assessment (which is quite similar for all CEE members of this WE bank, even if not subject to discussion in this study) could have on gaining competitive advantage on the very intense and severe CEE retail banking market. Finally, I suggest further improvements of the individual scoring model through the introduction of CB scoring and larger local data samples.

Key words: individuals, creditworthiness assessment, credit granting, credit underwriting, back-testing and credit scoring models development

Table of Contents

Introduction

Individual credit scoring has received little attention from credit risk management scholars even if the subject has been crucial for the considerable growth in individual loans in the recent years. Almost all of the academic works on credit risk management deal with corporate banking problems: *corporate borrowers' probability of default (PD) estimation* (internal/external rating systems, Altman Z-score model, KMV and Black-Scholes-Merton's models, etc.), *unexpected loss prediction, recovery rates and default rates (DR) correlation, Basel II capital requirements.*

Only a small part of the research has been concentrated on individual loans[1]. However, it has been mainly focused on the impact of different variables - predictors of default - on loan pricing, default risk, or repayment capability of the applicant, but it does not investigate what should be the weight and the significance of each of these variables in one entire model. Regarding mortgage loans (MG), researchers explore correlations between observable personal characteristics and individual loans spreads (Chiang, Chow, and Liu, 2002), minority discrimination and DR (Boyd, 1997), investment characteristics and income level (Goldberg and Harding, 2003), etc. Concerning the credit cards market, dependencies between personal delinquency and default costs (Gross and Souseles, 2003), consumer search costs and credit cards rates (Berlin and Mester, 2004), were examined. Finally, some scholars explored bank lending policy and credit scoring (Jacobson and Roszbach, 2003) or relationships and rationing (Chakravarty and Scott, 1998) in consumer loans (CL) market.

Researchers themselves argued that "historically, credit analysis had experienced little progress due to lack of data" (Falkenstein, 2002). It is fascinating, but understandable how a large part of the empirical literature regarding credit scoring techniques is developed by practitioners in rating agencies (e.g. Moody's - Stein, 2002; Crosbie and Bohn, 2002) or in banks (e.g. Deutsche Bank AG - Fritz, Popken,

[1] Individual loans (loans to private individuals) include any credit asked by a citizen or a physical person that is addressed to personal needs and expenses. E.g. credit cards, consumer loans, mortgages, overdrafts, and special purposes loans.

and Wagner, 2002). It is fascinating because many research areas are invaded and completely conquered by university scholars. This, however, is not the case of credit scoring models and rating systems. Academic studies prior to 1990 normally used less than 100 defaults in a sample, which is not enough for model development (Falkenstein, 2002). Hence, professionals' penetration in a research subject where statistical techniques and data availability are of major importance is justified. A hybrid situation of collaboration between academicians and practitioners can be both reasonable and fertile as banking research department members have access to data, while scholars have broad theoretical knowledge. An example of such a mutual project is the year-2003 paper of Engelmann and Tasche (Deutsche Bank AG), and Hayden (University of Vienna) on back testing of rating systems.

Still, no studies on the whole individual creditworthiness assessment (which includes both scoring models and policy rules) exist as publicly disclosed financial institutions' (FI) information is very limited. My research aims at completing the gap in the scientific literature through the analysis of the introduction and functionality of the entire individual credit granting model (scoring models and policy rules) into Central and Eastern Europe (CEE) banking systems (the Bulgarian case) with the help of WE bank institutions. The access to confidential data of the WE bank in question allowed the realization of this project, which gives a significant basis for the alignment between theoretical literature and financial institutions (FI) practical experience.

Regarding loan type, a prevalence of research work on MG rather than on CL[2] exists. However, nobody (with rare exceptions like Chakravarty and Scott, 1998) tried to

---

[2] Mortgage loans give conditional ownership of a real estate, secured by the real estate being financed. The borrower gives the lender a mortgage for the right to use the property while the mortgage is in effect, and agrees to make regular monthly instalments of principal and interests. The lien is removed when the debt is fully paid. It is a long-term debt with a tenor between 10-25 years. Consumer loans are loans extended to individuals for purposes other than home mortgages, i.e. for financing the purchase of movable properties (e.g. automobile, household appliance and furniture, technology, events/celebrations, etc.). The borrower agrees to make regular monthly instalments of principal and interests with tenor of 1 to 5-6 years. Consumer loans may be secured by personal guarantee, cash collateral, pledge, or real estate.
The repayment scheme for both types of loan can be either constant instalment repayment (annuity) or constant principal repayment.

analyse simultaneously different types of individual credits. My paper examines both CL and MG loans at their first introduction into CEE affiliates of the WE bank in question with a particular focus on its Bulgarian member.

On the other hand, all of the existing literature examines either samples of only one lending institution (Jacobson and Roszbach, 2003; Boyd, 1997) or samples comprising data of various FI, but operating on the same domestic market (usually the USA one). My study is a mixture of the two-abovementioned possibilities in the sense that it explores one CEE bank, situated in Bulgaria, that operates on its national individual lending market (the Bulgarian one), but is acquired and governed by a Western European (WE) FI that operates on another country market. Of course, the Bulgarian bank is autonomous in most of its actions, but WE bank expertise and experience on the domestic market are crucial. In the mean time, I also look at CEE banking markets similarities and differences that come out from close past economic and political history of the countries in the region and present affiliation of some of the largest local banks to the WE bank. Consequently, in contrast to previous literature that analyses mainly US data samples, with the exception of some evidences from Swedish CL loans (Jacobson and Roszbach, 2003) or Hong Kong MG markets (Chiang et al., 2002), my paper elaborates individual credits data sets from CEE that has become lately one of the most attractive markets for foreign investors.

Another stream of the retail banking research work, though very limited, tends to explore the superiority of artificial intelligence models (like neural networks) to traditional scoring techniques. For instance, Yang, Platt, and Platt (1999) illustrated that the back-propagation model obtained the highest classification accuracy when compared to another neural network model and discriminant analysis. Despite all its advantages in identifying bad customers, (Malhotra and Malhotra, 2003; Yang, Platt, and Platt, 1999), in practice almost nobody introduces neural networks, because of their complexity (Mester, 1997). Actually, FI prefer traditional statistical scoring models.

My study does not only show how an individual creditworthiness assessment is tested and built up, but how it is improved and applied in markets where many obstacles (e.g. lack of long time history and clean data) make difficult, but not impossible, the development of real statistical credit scoring models. The results of this study can be generalized for all of the CEE banks acquired by the WE bank in question as far as the initial versions of the individual creditworthiness assessments introduced in all of them are pretty similar (also economic and banking environment is quite alike in all of the CEE countries) and are undergoing the same changes and unifications with the WE bank model. The only minor difference consists in the presence of Credit Bureau (CB) in some of the CEE countries.[3] However, CB databases are still incomplete and unreliable source of information as some period of transition and testing is needed. Therefore, even if I have to take the example of Bulgaria where no CB has been developed yet, the results can be generalized for all CEE banks.

In the first paragraph, I offer a brief explanation of the nature of individual scoring models and back testing putting together practical (the experience of the WE bank in consideration on its domestic market and in CEE) and theoretical (review of the existing literature) views; I then depict the particularities of the individual loans market and list the problems with data quality and information availability in CEE financial system. In the second paragraph, I give details about the WE and the CEE banks underwriting processes (focusing on the Bulgarian affiliate) and explain why at present it is not possible to adopt the entire WE bank practice in CEE countries. In the third paragraph, I describe the dataset and the sample under consideration, while in the fourth paragraph I depict the methodology used for the quality and the stability back testing of the scoring model currently applied in the Bulgarian affiliate of the WE bank. I then examine the effects of the policy rules on banks' final decision and check the overlapping area between rules and models. The results at this stage of the analysis showed, as expected, that the current generic credit scoring model does not have significant discriminatory power and a new statistical one can be built up.

[3] Credit Bureau (CB) investigation has become a reliable source for fraud or negative records of individual credits applicants in the USA and WE, but it is only now tackled in CEE (Thomas, L., Edelman, D., and Crook, J., 2002). For example, four of the CEE countries under discussion (Bulgaria, Croatia, Slovakia, and Romania) do not have CB, while the other three (Poland, Czech Republic, and Turkey) do have it.

Therefore, in the fifth paragraph, I explain the methodology used for the development of the new statistical scoring model, which is followed by a review of the policy rules and summary of the last results. Even if with significantly superior discriminatory, the new creditworthiness assessment can be further improved after the development of Credit Bureau (CB) register and CB score. Finally, I discuss the implications of the empirical results and outline future improvements of individual creditworthiness assessment in CEE.

1. Theoretical background and literature review

In this section I first provide some theoretical definitions and explanations followed by a review of the related literature; I then give a brief overview of CEE banking system and its evolution under foreign credit institutions penetration. Finally, I discuss the main problems and peculiarities of individual loans market in CEE.

1.1 Theoretical background and main definitions

The main concern of each FI is to assess the creditworthiness of its potential borrowers no matter whether they belong to the corporate or retail segment[4]. In order to do that, banks either develop their own internal models for evaluation or use the ones already created by researchers, rating agencies and other centralized institutions (e.g. Credit Bureau - CB). The main objective of all models present in the theoretical and empirical literature is to discriminate relatively good from relatively bad clients. According to the theoretical literature, scoring models can be divided into non-statistical/judgemental, statistical, and artificial intelligence models. In the past, when not enough significant statistical information for model development was available, banks used to rely on **non-statistical/judgemental** methods. Concretely, the credit granting decision was based on subjective human assessment (Hand and Hanley, 1997). With the increased quantity and quality of historical default data **statistical credit scoring models** moved aside the less accurate and efficient

[4] For the sake of simplicity, under corporate segment, I understand loans to large and medium companies (turnover >500 000 Euro) plus to small and medium enterprises (SME) (turnover between 250 000 and 500 000 Euro), while under retail segment I intend individual loans.

judgemental assessments (Rosenberg and Gleit, 1994; Reichert, Cho, and Wagner, 1983; Chandler and Coffman, 1979; Hsia, 1978). Statistical credit scoring models can have one of the four methodological forms: (1) *discriminant analysis* models, (2) *linear probability* model, (3) *logit* model and (4) *probit* model. All of the above mentioned techniques are aimed at aggregating data contained in different indicators through statistical analysis that assign weights to various financial ratios or personal characteristics depending on their discriminant power. The main feature of *discriminant analysis* (Fisher, 1936; Altman, 1968) models is the segregation of the borrowers in two classes: one characterized by low level of default risk and one characterized by high level of default risk. The remaining three models differ among one another on the assumed relationship between the PD (the dependent variable) and borrowers' characteristics (independent variables). Thus, the *linear probabilistic* model hypothesizes a linear relationship, but as the problem of heteroscedacity occurs, *logit* and *probit* regressions have been suggested. The only difference between the last two techniques consists in the assumption that the dependent variable (PD) represents cumulative function of the logistic distribution in the first case and cumulative normal distribution in the second case. Still according to Thomas, Edelman, and Crook (2002) logistic regression is:

> "the most successful scorecards development method by both academicians and practitioners as it has less restrictive assumptions to guarantee its optimality".

The last group of credit scoring methods is represented by **artificial intelligence** models that embody a heterogeneous set of methodologies for solving problems based on mathematical logic. Being quite complex, these techniques are rarely implemented in practice and the research work on them is very limited with the exception of the neural networks model (e.g. Malhotra and Malhotra, 2003; Saunders and Allen, 2002; Yang, Platt, and Platt, 1999). Being less complicated and easily managed, statistical models are the ones prevailing in practice. According to Mester (1997), 97% of banks in the U.S. (no research generalized for the whole Europe exists because of data availability constraints) use statistical models to approve credit card applicants, while 70% of the banks use it for the SME evaluation.

In reality, however, individual creditworthiness assessment consists of both: a) statistical credit scoring models (usually logit or probit) and b) policy rules application (which is also the WE bank case).

Statistical credit scoring models build distinct scorecards on a sample of clients' data. Every scorecard consists of various variables/characteristics each with closed number of answers/attributes. Each attribute is assigned with a score (point) that indicates borrower's degree of creditworthiness. Usually, a higher score corresponds to lower PD. The scorecard for individual clients is based on six main categories of variables: *socio-demographical characteristics* - stability of the consumer (e.g. Time at address), *his /her resources* (e.g. Residential status, employment), *possible outgoings* (number of family members); *income and financial sophistication characteristics* (e.g. Family income level, total debt to income ratio, number of credit cards held, existing liabilities, current account); *CB characteristics* (e.g. Internal and external database sources, Credit Bureau score), and *product characteristics* (e.g. Purpose and tenor of the loan, own contribution). No subjective evaluations of credit officers are present. However, it is not enough to calculate the score of each applicant. A threshold score C dividing potential survivals from potential defaulters must be set. Therefore, every FI introduces a predefined cut-off value C that reflects its credit policy and market share objectives. Each applicant with a score higher than C will be classified as a non-defaulter and each applicant with a score lower than C as potential defaulter. Four decision outcomes are possible. If the score is above C and the obligor does not default or the score is below C and the debtor defaults than the decision was correct. Otherwise, if the score is above C and the debtor defaults, the decision was incorrect and the Type I error is present. Correspondingly, if the score is below C and the debtor does not default than the decision was wrong again and the Type II error occurs, i.e. a non-defaulter was wrongly classified as a defaulter. Banks try to minimize both types of errors, as in the first case they face incurred debt and in the second case they suffer loss of potential profit. They construct a cost function (whenever enough data about the costs are available) whose minimum is taken as the cut-off value C.

Policy rules as part of individual creditworthiness assessment concern minimum age, minimum monthly income, loan-to-value ratio (LTV), already adversely classified clients, etc. These rules are based on the internal credit policy of each FI and can vary with time, bank's strategy, and local environment. Two kinds of policy rules exist in practice: headquarter and black rules. The first one sends an application for a better examination to headquarters - HQ - (where subjective evaluations of credit analysts are introduced), while the second leads to automatic rejection of the application. Despite mentioning that policy rules can influence the final decision of the credit analyst, empirical research does not analyse openly their role in the individual creditworthiness assessment. No explicit examination of banks credit policy and its impact on credit scoring models is available as policy rules are subject to private FI information.

1.2 Literature review

In 2001 Jacobson and Roszbach stated that:

> "Individual credit has come to play an increasingly important role, both as an instrument in the financial planning of households and as an asset on the balance sheet of FI. Since a creditor cannot observe borrowers' probabilities of default, credit-scoring models – by enabling a lending institution to rank potential customers according to their default risk – can improve the allocation of resources, from a second best towards first best equilibrium".

No study explaining the whole structure and functionality of individual creditworthiness assessment (scoring models and policy rules) exists. What normally happens is that scholars examine the impact of different variables - predictors of default - on loan pricing, default risk, or repayment capability of the applicant, but they do not investigate what should be the weight and the significance of each of these variables in one entire model. Furthermore, because of publicly undisclosed FI information, academicians could not analyse FI's policy rules or entire underwriting processes. Jacobson and Roszbach's 2001 study is a rare example of a paper that tries to analyse bank's policy, but as their information about bank's strategy was very limited they could draw conclusion only on the basis of the results from their simulations, which does not corresponds necessary to reality. They argued that lending institutions neither minimized credit risk nor maximized the rate of return, but practice a simple decision-rule scheme. According to their results, bank's behaviour is either a symptom of an inefficient lending policy or the result of some other type of optimising behaviour, as the provided loans were not consistent with the default-risk minimization theory. Having available all the data for the creditworthiness assessment of the WE bank and its CEE affiliates, I argue that my results regarding policy rules and bank's strategy will be directly related to practice. Furthermore, my study incorporates both policy rules and credit scoring model in an entire individual underwriting process.

Another specificity of the existing studies is the prevalence of MG loans analysis where long loan tenor (large time series) and high loan amounts attract academicians' interest. In their literature review of 1992 on MG loans, Quercia and Stegman identified three groups of default studies: first-, second- and third-generation models depending on the perspective from which MG risk is analysed and on the measure of default risk. The first-generation models applied mostly regressions, the second-generation ones - both regressions and multinomial logit, whereas in the third-generation models prevailed proportional hazard models[5]. Quercia and Stegman argued that academicians generally agree on the impact of loan characteristics (home equity and LTV) on default, while what is unknown is the effect of transaction costs and borrowers related factors that come out to have even opposite effects according to different scholars (Kau, Keenan, and Kim, 1991 vs. Quigley and Van Order 1992).

As most of the literature on individual loans examines scoring models and variables correlations, various predictors of default were found. Thus, Boyd (1997) linked default decision in the student loans market to borrowing expectations in the MG home loan market. She stated that students that default on an educational loan face

[5] The proportional hazard model is suited to examine default risks in loan pools and loans prepayment behaviour. The main elements of the model are: the hazard (the default event) and the hazard rate (the probability that this event will occur in a particular period given that it did not occur at the beginning of the period). The model is used for the estimation of the PD in the first year, second year, etc. of the loan existence. For further details see Quercia and Stegman (1992).

poor credit rating and consequently increased cost of borrowing in the future. Therefore,

> "given that the primary cost associated with defaulting on a student loan is the potential restriction from other credit markets or an increased cost of borrowing, for those students who foresee themselves not being able to borrow for another loan such as a mortgage, the rational decision is to not repay their guaranteed student loan".

Borrower's characteristics influencing the default decision came out to be the ability to repay[6], type of education (perceived type of degree), parental income. Expectations of future credit market came out to be significant too. Finally, it was argued that applicants who are older, married, and have children have significantly higher probabilities of homeownership. Despite Boyd's study (1997) where a comparison between black and white students has been made, scholars and professionals tend to exclude gender and race from credit scoring models, while characteristics like age, marital status, and number of family members are always present in the scorecards. Through the method of hedonic indexes Lacour-Little and Malpezzi (2003) examined the effect of appraisal quality[7] on MG loan performance using data from the high volatility market of Alaska in the 1980s. Their results showed that over-appraisal is significantly related to default, whereas under-appraisal has no effect. Besides, 40% of defaulting borrowers attributed their default to the trigger events of job-loss and divorce, where job-loss was more often cited event. Hence, employment and marital status are two other variables that have been reasonably included in the individual scoring models that are explored in this paper. Lacour-Little and Malpezzi (2003) used also housing payment to income ratio and total debt to income ratio (used also in the WE institution and its CEE banks) for assessing creditworthiness. Briefly, while some studies focus on borrower's

[6] The ability to repay is a result of the comparison between the debt ratio and the relative thresholds (cut-off points) defined by each FI. Usually, the ability to repay is a dummy variable that takes a value of 0 if the debt ratio exceeds the corresponding threshold (i.e. the applicant is not able to bear the burden of its liabilities), and a value of 1 if the debt ratio is below this threshold (i.e. the applicant is able to repay the new loan).

[7] According to the MG literature terminology the appraisal value of a real estate property is defined as its expert evaluation. In practice, whenever an individual applies for a MG in a certain FI, he has to declare not only the market value (the contract/purchase price) of the real estate, but also its expert evaluation. This evaluation is estimated by a specialist that belongs to a particular list of licensed real estate experts approved by the FI at issue.

characteristics, others focus on the difference between property and loan values as determinants of default and losses, and thirds focus on cash flows as predictors of default[8]. For example, Goldberg and Harding (2003) examined the relationship between level of income, default and refinancing rate through a multinomial logit model. With the help of Monte Carlo simulation of future cash flows, they found that low and medium income portfolio exhibits a significantly higher DR but a significantly lower refinancing rate when interest rates fall. Some of the correlations that they verified are important for my study and for the redesign of individual scoring models as far as income characteristics are concerned. Firstly, Goldberg et al. (2003) demonstrated that high-income individuals generally have greater wealth and cash reserves to help weather income disruptions that might cause a low-income, low-wealth borrower to default. Second, they argued that some lenders apply stricter underwriting standards to refinance loans and so low-income borrowers may face underwriting constraints limiting their ability to refinance. And third, it was shown that high-income borrowers have higher rates of mobility. Their results confirmed the existing strong theoretical arguments that the level of income influences borrower behaviour. Meanwhile, Moss and Johnson (1998) argued that lower income households have gained increased access to credit over time. Actually, this is one of the main concerns for the CEE affiliates of the WE bank in consideration. A way of granting money to "poorer" households without increasing the default risk is one of the problems that the individuals' creditworthiness assessment tries to solve. However, Vandell and Thibodeau (1985) and Cunningham and Capone (1991) argued that sources (e.g. self-employed, retired, from child support or alimony) and types of income (e.g. income from salary, capital and investments, commissions) together with the length of employment at one's current job have a direct impact on default rather than income per se. As their findings come out to be significant also in practice, I will consider them for the new model development. Moreover, their results indicated that the payment to income and the after-tax non-housing debt to income ratio[9], together with LTV, are significant predictors of default.

[8] In a way the WE bank granting process incorporates all of them in its creditworthiness assessment.
[9] These two ratios are always considered together in the mortgage scoring models literature. Because of that, usually the debt to income ratio (debt ratio) is equal to (New Monthly liabilities + Existing Monthly liabilities + Monthly Cost of Living)/ (Net Monthly Income), where Cost of living is net of

In 1978 Vandell examined how the impact on default of some variables varies when alternative MG instruments are applied. He found that households with low or negative equity in their residence or with high payment burden would experience higher default risk. The same was demonstrated to be true for applicants with low income, loans with higher term to maturity and high contract interest rates. In 1993 still Vandell summarized most of the variables used in the creditworthiness assessment for MG loans, namely: borrower (payment to income, the after-tax non-housing debt to income ratio, potential – e.g. period of employment, number of liquid assets), loan (LTV; purpose – e.g. for building, reconstruction), and property (type of property; condition of the real estate, marketability – e.g. age of the real estate, number of owners) characteristics. However, further review of the literature reveals differences between theory and practice: economic models focus on the role of equity, while many lenders focus more on ability to pay (Sandor and Sosin, 1975; Vandell, 1978; Jacobson and Roszbach, 2001). Even if I analyse MG risk from the perspective of the lender (the WE bank and its CEE affiliates) the credit-scoring model that I develop afterwards considers both ability to repay and LTV (measure of home equity).

Most of the existing literature on individual loans (apart from MG loans) is oriented towards non-instalment or revolving credits and less often towards CL loans (instalment loans)[10]. Therefore the CL loan literature is very limited and deals with questions like bank strategies for loan-pricing (Lown and Peristiani, 1996) and contract conditions (Jacobson and Roszbach, 2001). Most of the significant borrowers' characteristics and credit scoring techniques are the same as in the MG

housing expenditures; thus, the nominator of the ratio includes only debt and consumption expenditures. Vandell (1978) argued that the payment to income ratio is used also for consumer loans. In the WE bank model the debt ratio is equal to, (New Monthly liabilities + Existing Monthly liabilities)/ (Net Monthly Income- Monthly Cost of Living), while in the CEE banks it is equal to, (New Monthly liabilities)/(Net Monthly Income - Monthly Cost of Living – Existing Monthly liabilities) and is used for both MG and CL loans.

[10] Monthly instalment loans are the ones that request regular pre-defined monthly payments – CL loans, MG loans, special purposes loans, cash loans. Non-instalment or revolving loans are the ones that do not have any fixed prepayment scheme, but only a predetermined maturity (overdrafts) or repayment amount depending on the total expenditure (credit cards). Usually, this amount consists in paying a percentage of the total outstanding of the credit account every month. Non-instalment loans consist mostly of credit cards, overdrafts and rarely of special purpose loans.

literature, even if loan characteristics depend on the specificity of the bank product under examination. Thus, Jacobson and Roszbach (2001) designed a bivariate probit model that consisted of two simultaneous equations (one for binary decision to provide a loan or not, and another for binary outcome – default or proper payment) in order to define the portfolio credit risk of a Swedish FI. They demonstrated that income, house ownership, type of employment (self-employed), type of income (from capital), number of collateral-free loans already outstanding and number of guarantors have positive impact on loan granting, while product limit, used amount, and no collateral-free loans already outstanding affect negatively a bank's decision. Borrowers' characteristics like age, marital status (divorce), and number of requests in Credit Bureau were included in the model. Such variables are also considered in the WE bank and CEE affiliates scorecards (apart from "number of requests in Credit Bureau" that is conditional on the existence of a Credit Bureau in the countries). Analysing credit card delinquency and personal bankruptcy within the stability of credit risk models, Gross and Souseles (2003) found important relationships between default risk and scorecard results. They estimated a duration model for consumer default and assessed the relative importance of different variables in predicting default. Future DR was confirmed to increase with the increase in credit supply. They also observed that people had become more willing to default over time. Thus, a change in the relationship between default and the variables that lenders typically use to predict it, such as debt levels, was verified (also in CEE where the economic situation is still volatile, most of the banks run the risk of facing more and more customers with increased debt ratio). Moreover, applicants with higher scores have been demonstrated to be much less likely to go bankrupt, while greater unemployment, weaker house prices, and lack of health insurance were associated with more bankruptcy. Such results underline the importance of not only developing and improving individual scoring models (scorecards and debt ratio), but also exploring the move in the correlations among predictors of default over time.

The last stream of the literature explores the superiority of artificial intelligence models (like neural networks) to traditional scoring techniques. Usually, these kinds of models are tested and applied on CL loans datasets (Malhotra and Malhotra, 2003;

Hand and Hanley, 1996). For instance, Yang, Platt, and Platt (1999) illustrated that the back-propagation model obtained the highest classification accuracy when compared to another neural network model and discriminant analysis. Hand and Hanley (1996) looked at the application of the k-nearest-neighbour (k-NN) method in CL credit scoring. They proposed an adjusted version of the Euclidean distance metric and compared its performance to the traditional credit scoring techniques (linear and logistic regressions and decision trees). The k-NN performed well, achieving the lowest expected bad risk rate, but the results were confirmed to be valid for populations with lower proportions of bad risks in the whole population. Malhotra and Malhotra (2003) compared the performance of multiple discriminant analysis to neural networks (back-propagation model) in identifying potential loan defaulters. Applied to seven different random cross-sections of the data sample, a paired t-test showed that there is no significant difference between the average predictive performance of artificial neural systems and discriminant analysis models in analysing "good" loans, while the difference is significant for problematic loans. However, they argued that the use of credit scoring technique depends on the complexity of the institution, and the size and the type of the loan. Even if in 1996 Hand and Hanley demonstrated that k-NN is superior to some of the traditional credit scoring techniques, in 1997 in their literature review of the classification methods in credit scoring, they argued that this superiority is not very significant

> "in the context of population drift and looseness of the bad and good risk class definition. When one factors in the cost of changing the scoring system, and the likely future life of any system that one does install, one questions whether the differences are of any practical value".

Therefore, despite evidence about improved classification accuracy under artificial intelligence models, FI still explore their advantages to traditional scoring models especially as far their individual loans portfolio is concerned (Hand and Hanley, 1996).

As academicians experience difficulties in accessing data on MG and CL credits with numerous individual characteristics, as well as loan information, empirical research relating MG and consumer borrower/loan characteristics within an entire credit underwriting process (scoring models and policy rules together) is scant and more precisely does not exist. Scholars usually analyze the impact of different variables on default risk, but they do not investigate what should be the weight and the significance of each of these variables in one whole model. The most significant predictors of default in the MG literature can be separated into two main categories:

- Demographic characteristics - type of education (perceived type of degree), age, marital status, number of children (Boyd, 1997), employment, (Lacour-Little and Malpezzi, 2003), length of employment at one's current job (Vandell and Thibodeau, 1985; Cunningham and Capone, 1991), etc.;
- Financial and income characteristics - ability to repay, parental income (Boyd, 1997), housing payment to income ratio and total debt to income ratio, payment to income and the after-tax non-housing debt to income ratio, LTV, sources (e.g. self-employed, retired, from child support or alimony) and types of income (e.g. income from salary, capital and investments, commissions) (Lacour-Little and Malpezzi, 2003; Vandell and Thibodeau, 1985; Cunningham and Capone, 1991).

According to the limited CL research the most relevant variables for the occurrence of default are age, marital status (divorce), and number of requests in Credit Bureau, income, house ownership, type of employment (self-employed), type of income (from capital), number of collateral-free loans already outstanding, number of guarantors, product limit, and used amount (Jacobson and Roszbach, 2001).

In effect, most of these variables have also been taken into consideration in the existing scoring models in the WE institution and its CEE affiliates. However, my study aims at completing the credit risk management literature through examination not only of the scoring model, but of the whole individual creditworthiness assessment (scoring model and policy rules) in the CEE banks of the WE bank under examination. In fact, my objective is to back-test the current generic individual credit granting model in CEE and to develop a new statistical one, based on practical experience and theoretical literature findings and recommendations. As far as the structure of the CEE banks underwriting processes depends also on the local reality, I first give a short description of the banking system in the region and legal environment of concretely in Bulgaria, followed by the presentation of the existing credit granting processes in the WE bank and its CEE affiliates with a particular focus on its Bulgarian member.

1.3 Individual banking and legal environment in CEE

1.3.1 Banking environment

The largest Austrian, Italian, German, French and English banks[11] are in the process of competition and expansion on CEE markets as demand for financial services in the region is growing faster than in WE. The euro zone economy grew by 1.3 percent in the first quarter of year 2004 compared to 3.2 percent in Croatia, Slovakia and Bulgaria, and 6.1 percent in Romania. The overall upraise in the economies of the ex-communist countries that step by step are doing their way through transition towards EU stability has been appreciated also by the latest enlargement of the community. Direct foreign investments together with GDP growth and improvement of the main macroeconomic indicators gave positive signals to foreign banks that from 1996 on started to penetrate the regional banking system (see Table 1).

Table 1

Country	Total banking assets belonging to foreign institutions	Key international players
Bulgaria	About 80% as of December 2004	German HVB Bank, Italian UCI, German Raiffeisenbank, and French Société Générale
Croatia	91% as of September 2003	Italian UCI
Czech Republic	87% as of December 2003	Belgium KBC Bank, French Société Générale, and Austrian Erste Bank
Poland	63% as of December 2004	German HVB Bank, Italian UCI, and German Allianz insurance company consortium, American Citigroup, and German Deutsche Bank
Romania	58% as of December 2003	French Société Générale, Dutch ING Bank, German Raiffeisenbank and Romanian-American Enterprise Fund consortium, and American Citibank
Slovakia	85% as of May 2003	German HVB Bank, Italian UCI and Banca Intesa, German Raiffeisenbank, and Austrian Erste Bank

Source: CEE National Banks reports

[11] I intend largest banks by total assets.

Thus, in Czech Republic foreign entities controlled 90% of the total assets of the sector as of 31 December 2001, which is 24 percentage points more than a year earlier. Taking into account banks owned indirectly by international institutions through subsidiaries operating in the country, the figure increases to 94.2% (Czech National Bank). At the end of December 2004, the corresponding numbers are 87% (3 points more compared to 2002) and 95.7% (2.6 points more in year-on-year terms). The new shareholders of the large Czech banks are foreign entities based in Belgium (KBC Bank), France (Société Générale), and Austria (Erste Bank). In Bulgaria foreign institutions control more than 80% of total banking sector assets as of December 2004 (Bulgarian National Bank; Barisitz, January 2001). With the privatisation of Bulgarian third largest bank - DSK Bank - the market structure, in terms of key international players, is perceived to be largely fixed (HVB Bank, UCI, Raiffeisenbank, and Société Générale). In Poland foreign banks controlled 70% of total banking system assets as of 1999 year-end, but this figure underestimates the presence of foreign strategic investors which, although minority shareholders, exercise control of banks (Damutz and Gabbai, September 2001). The corresponding number as December 2004 is 63%. Such decrease is explained by the changes in total assets structure in the banking system. Main foreign competitors on the Polish banking market are found out to be German HVB Bank, Italian UCI and German Allianz insurance company consortium, City Bank, and Deutsche Bank. Foreign banks play a dominant role also in Slovakia, where their market share increased between 1999 and 2002 to 85%. They already hold 90% of total banking assets (Gardò, May 2003), where key international players are Erste Bank, Banca Intesa, Raiffeisenbank, UCI, and HVB Bank. The situation in Croatia is quite similar. Foreign institutions control 90.8% of total banking sector assets as of September 2003 (National Bank of Croatia). In 2004, 39 banks are operating in Croatia, of which 19 of foreign ownership. By end-2000 (before the privatisation of Banca Agricola), foreign-owned banks were estimated to account for over 53% of total Romanian banking assets. Six out of the ten largest banks (in terms of assets) were foreign-owned; three of these six were branches of foreign credit institutions (Barisitz, January 2001). According to the Romanian National Bank 2003 annual report, during the last three years market shares held by domestic banks contracted in favour of foreign banks that hold 58.3% of the total assets in Romanian credit institutions at end-2003. Key foreign investors in the Romanian banking sector are Société Générale, ING Bank, German Raiffeisenbank and Romanian-American Enterprise Fund consortium, Citibank Overseas Investment. Still, according to one of the dominant players in the banking system – Raiffeisenbank – CEE has high growth potential and remains a region that is still largely under-banked.

International FI is looking for exporting expertise they have built up domestically in their new markets. However, the unavailability of strong or any individual loans market (e.g. in some of the CEE banks there were no credit cards until 1999) impeded and slowed down the introduction of retail banking products and individuals' creditworthiness assessment. According to the "Household Wealth in the New Europe Countries" report made by the WE bank in co-operation with Pioneer Global Asset Management and PKO SA, people in CEE have 22-times less financial savings than those living in the "old" EU member states. Moreover, they borrow 40 times less than individuals in WE. People's mistrust and lack of market economy reasoning are among the drawbacks for the development and expansion of retail products. In such an environment, major WE banks and their affiliates enter into severe competition for larger shares on the newly created individual loans market. In 2003 most of the banks spread their credit expansion on all fronts - loans to corporate clients, to small and medium enterprises, and to private individuals. However, their main target still remains crediting the population. In the last year and a half a strong increase in the individual credit demand has been recorded. The credit expansion resulting from long-delayed consumption over the years, aims at catching up after years of delayed crediting. A real example is the Bulgarian case (see Table 2 below) were the aggregate amount of CL loans, extended by the largest credit institutions (assets above 500mln BGN), has increased several-fold: 8 times in the case of Hebrosbank, 6 times in the case of Biochim, 5 times in the case of Raiffeisenbank (Bulgaria). Bulgarian National Bank (BNB) and IMF agree that there is a boom in retail credits (the year-on-year – yoy - growth is 66.8% in 2004 vs. 80.7% in 2003) and decide measures for its restriction:

"Following Fund technical assistance on alternative responses to the credit boom, the BNB announced in February 2005 that, beginning with the second quarter of 2005, banks whose credit portfolio expands by more than 6 percent per calendar quarter will be subject to an unremunerated deposit requirement of twice the excess credit expansion, unless the ratio of their credits (including risk-weighted off-balance sheet items) minus capital to deposits, other than those by other financial institutions, is below 60 percent. Staff supported this measure - particularly as a counterweight to fiscal easing—but questioned two elements in its design: first, the credit-to-deposit criterion exempted 14 out of the 35 banks (including the largest one); and second, given that the measure had been announced over one month before coming into force, banks could boost lending in March so as to create a larger base from which to calculate growth rates[12]. Taking into account those loopholes, staff estimates that the credit expansion slows to 30 percent in 2005 (equivalent to a reduction of net credit. expansion to 10 percent of GDP from 12 percent in 2004), but some of this slowdown is likely to be replaced by credit from non-banks (especially leasing companies) and direct foreign borrowing. Partly for this reason, the authorities plan to enhance the monitoring of leasing companies."[13]

Table 2[14]
It shows the actual values for Bulgarian main macroeconomic indicators, credit growth, and Non-performing loans (NPL) according to International Monetary Fund (IMF) and World Bank (WB).

BULGARIA	1998	1999	2000	2001	2002	2003	2004
GDP growth (% real)	4,0	2,3	5,4	4,1	4,9	4,5	5,6
Inflation (CPI) yoy, avg	18,7	2,6	10,3	7,5	5,8	2,3	6,1
Unemployment rate %	12,5	14,1	18,7	17,5	17,4	14,3	12,7
Total Loans (FX and local currency) - yoy in %	n.a.	n.a.	n.a.	32,1	44,0	48,3	48,7
Corporate	n.a.	n.a.	n.a.	42,1	42,1	38,0	33,0
Retail	n.a.	n.a.	n.a.	44,1	44,1	80,7	66,8
NPL ratio (NPL/Total Loans) %	n.a.	n.a.	n.a.	13,1	8,6	7,3	7,0

Source: IMF and WB Annual Reports

It is a well-known fact that economic cycles have impact on companies' financial results and rating migration in corporate lending (Falkenstein, 2002). Consequently, such firms' performance volatility influences households' budget constraint and propensity to consume. Normally, GDP growth is translated into improved credit

[12] This has indeed happened, but the BNB has announced that it will exclude from the base temporary credits extended in late March to large corporations.
[13] IMF Country report No.05/ 169. May 2005. "Bulgaria: First Review under the Stand-By Arrangement and Request for Waiver of Performance Criteria - Staff Report; Staff statement; Press Release on the Executive Board Discussion; and Statement by the Executive Director for Bulgaria".
[14] It is difficult to find data about total loans and NPL ratio (the "n.a." fields in Table2) in CEE as the methodology of calculation was different before 2001 and even if some numbers are available they are not comparable to the ones presented in the table as they include different information and indications (See also paragraph 1.4).

activity through increased total loans and decreased non-performing loans (NPL) in the short-term (see Table 2). However, individual loans boom in CEE resulting from stable or increasing GDP growth is expected to lead to increased NPL in the mid-term (IMF Country report No.05/ 169. May 2005 on Bulgaria). Such a dynamic environment characterized by strong competition, elevated demand, and high retail loans growth may translate into weakening of loan eligibility requirements and eventually may cause selection problems and loans quality deterioration. For that reason, banks are obliged to revise, test and improve their individual granting models.

1.3.2 Legal environment (WE vs. Bulgaria)

As in any economic activity, the legal environment is essential also for credit granting. However, regarding individual loans particular attention should be paid to the Family code (family property regulation), and the types of guarantees provided by the law (personal guarantee, pledge, promissory note, and MG). The latter is of prior importance for the individuals' credit risk management as the better secured the loan (by types of guarantees and family property), the higher the recovered amount and respectively the lower the loss given default will be.

Regarding the Family codes according to the legislation of the WE country in question a married couple can choose between community and separation of assets as family property regime. If and only if an applicant for CL or MG loan is married and is in regime of community of assets, his/her spouse is obliged to sign the contract as co-applicant. In some of the CEE countries the "separation of assets" concept is still under examination by the corresponding authorities and is not mentioned in the Family codes (e.g. Bulgaria, Turkey). In others (e.g. Slovakia and Czech Republic), it exists but is not quite popular yet. Hence, it is compulsory that both spouses sign the credit contract if CL and MG loan are under consideration. In case of overdrafts and credit cards only the spouse who needs this specific bank product signs the contract, as income does not take part in the family property. The same logic is applied also to personal guarantors, where if a person agrees to be a guarantor and is married with no separation of assets his/her spouse should also sign the credit contract as guarantor. These conditions are crucial in case of individual loan recovery as the process becomes more complicated if the ownership of the real estate or the movable is joint, but only one of the spouses have signed the credit contract.

Four are the main types of guarantees that usually accompany CL and MG loans: personal guarantee, promissory note, pledge, and mortgage. The law regulation of **personal guarantees** is quite similar in the examined WE country and Bulgaria. A guarantor is a person who assures the performance of another party's obligation by binding himself expressly and personally to the creditor. The guarantor is liable jointly and severally with the principal debtor. If a debtor has several guarantors for one and the same obligation, each one of them is liable for the whole obligation unless there is an arrangement for the division thereof. Where a debtor has several guarantors for one and the same obligation, a guarantor who has performed the obligation may claim from the other guarantors their due parts. The guarantor shall also assume the rights of the creditor against third parties that have provided a pledge or mortgage for the obligation, but only up to the amount to which he would have had an action against them if they had been guarantors. The main difference is the fact that in the WE country under examination a guarantor can raise an action against the debtor even before making any payment, while in Bulgaria the guarantor can do it if and only if he has satisfied the debt (if there is no special agreement between the guarantor and the debtor). The second type of guarantee mainly accompanying CL loans is the **promissory note**. In the WE country in consideration, if an injunction is issued on the basis of promissory notes, it may be declared immediately enforceable. If the defendant does not protest in 40 days period, the injunction becomes final and enforceable. The only difference for Bulgaria is the period that the defendant has available for filing any protests against the injunction. However, the promissory note is less preferred to the other types of guaranties because of its juridical nature. The third kind of guarantee that is popular for CL loans is the **pledge**. In the WE country in question the possession of the thing given in pledge passes from the debtor (or the third party) to the creditor, while in Bulgaria two types of pledge exist. One of them satisfies the WE country regulation, while the second one – the registered pledge - is constituted without submitting the pledged property. The last type of guarantee that can accompany individual loans, mostly MG loans, is the **mortgage**. Also in this case some differences between the WE country and the Bulgarian legislations exist. Thus, in the WE country under examination a mortgage may be imposed on immovable property of the debtor or a third party, while in Bulgaria the property must belong to the mortgagor at the time of signing the contract. Furthermore, a creditor whose claim is secured by a mortgage is entitled to be satisfied preferentially from the mortgaged property's price, whoever its owner might be.

The type and the treatment of individual loans guarantees is of prior importance for the CEE countries, as one of their product characteristics is strictly related to the compulsory personal guarantors, pledge or mortgages when CL loan is requested. The kind of guarantee depends also on the requested loan amount.

1.4 Main problems encountered by the WE bank on the individual loans market in CEE

Various problems (no data storing and data cleaning, no retail banking market) impeded the development and implementation of the first version of individual underwriting systems in five of the CEE affiliates of the WE bank under examination. For example, one of them was acquired in 1999, but it introduced outdoor[15] designed Application Processing System[16] for individuals (APS) in 2002 (before that the workflow was managed manually); other two did the same in June 2003, even if they have become part of the WE bank since 2000, while the last two developed indoor APS correspondingly in September 2003 and during 2002. No matter who created the APS, the WE institution tends to unify all CEE banks' individual credit granting processes to its own model. As there was no retail segment

[15] Outdoor means that the APS was developed and implemented by external consultants, while indoor means that internal experts developed and implemented the APS.

[16] APS manages automatically the whole creditworthiness assessment (credit scoring model, debt ratio, and policy rules). It generates the workflow from the application of the potential client to the final decision of the competent bank employee. The APS calculates the result of the scoring model and suggests final decisions unless headquarter rules are present. In case of headquarter rules the application is automatically sent to headquarters where the competent credit analyst takes the final decision that can be different than the one given by the system. In case of black rules the system automatically suggests rejection of the application. Actually, the APS should prevent the application from further processing when black rules are available, but as far as data collection for statistical analysis is useful, the loan request is rejected at the end when all relevant information is stored.

in the portfolio of some of the banks the introduction of individual underwriting process resulted only after the development of general business and credit strategies for customers' attraction. Other CEE banks had had individual credit for some time, but bad data storing and no data cleaning did not allow the creation of granting model based on good samples of real data. Thus, in both cases the creditworthiness assessment of individuals had to be designed on the basis of the WE expertise and experience, where scant local banks' data forced the WE bank to implement generic scorecards developed for emerging markets. These generic scorecards were created by external consultants on a sample that was not representative for each single country of the region. The inexistence of historical data (because of bad data storing or lack of retail segment in the portfolio) impeded the possibility to test the newly implemented scorecard. Another limitation of the local market is the unavailability of official data about Cost of Living[17] (CoL) by different income groups[18], which is essential for the calculation of the level of indebtedness. The only data that can be found on the websites of National Statistical Institutes in CEE is the average income and its corresponding CoL value per capita or per household. Rarely, one can find also the CoL relative to the minimum income. Therefore, the CoL variations with income and number of family members were a matter of subjective evaluations and discussions between employees of the WE and the CEE banks. The lack of National Central Bank Registers (NCBR) and Credit Bureaus where all the banks in each country keep track of name, current address, total exposure, number of loans, credit history, payment records (positive or negative – defaults or delays) for all their clients comes out to be a crucial drawback for the realization of objective creditworthiness assessment. CB and NCBR exist in Poland, Czech Republic and Turkey since end-2002 – 2003, while in Slovakia and Bulgaria CB should start functioning in 2005 (the only external database in Bulgaria is NCBR which was activated just in November 2004 so it is possible now to gather some official data about applicants' credits in the banking system). Even if introduced, CB database is

[17] Cost of Living (CoL) usually includes living expenses for house maintenance, food, beverages, transport, electricity, and clothes. In developed countries the consideration of CoL in the debt-ratio is under revision, while for emerging markets it is still an important factor for the family budget planning. For example, in Bulgaria CoL for an average household income is around 70% of its budget.
[18] The WE country national statistical institute divides households into different income buckets and gives official data for the CoL of each bucket depending on the number of family members.

not complete and fully reliable source of information in CEE. CB score is not calculated at the beginning (e.g. Czech Republic), while in Poland and Turkey it is, but because of limited number of banks participating in CB and short time history it has not been tested yet and as such is not reliable. As some period of transition and testing of the CB database is needed, such a powerful instrument for citizen's evaluation as the CB score cannot be still built-in into CEE banks' scoring models and the introduced individual granting systems could not be a mirror image of the one of the WE bank in question. In fact, the design of specific and precise credit rules regarding the negative information available in existing CB and NCBR is the unique, but partial solution to the problem.

As a result, after almost one year and a half of its implementation, the WE bank has to update and renew the originally applied model (see paragraph 2 below) for individual creditworthiness assessment. Now it has good local data (even if with short history) necessary for the improvement of the generic scorecards so as to make them representative of each country itself. I take the example of Bulgaria, but the individual creditworthiness assessments in all of the CEE affiliates of the WE bank under examination are very similar for the reasons that I listed above. The only minor difference consists in the presence of CB in some of the CEE countries, but its databases are still incomplete (see above) and only some generic policy rules regarding CB can be considered. Therefore, even if I have to take the example of Bulgaria where no CB has been developed yet, the results can be generalized for all CEE banks.

2. WE and CEE banks' creditworthiness assessment

The final goal of each individual creditworthiness assessment is to accurately predict which potential borrowers could be subject to future default and which not. The individual creditworthiness assessments of the WE bank and its CEE affiliates consist of both: a) statistical credit scoring models (usually logit or probit) and b) policy rules application and are processed by APS. If the APS does not include or consider certain information known to the branch or relationship manager, he/she can refuse the application under examination or send it to a competent credit analyst in HQ entitled to take the final decision.

2.1 The WE bank creditworthiness assessment

When an individual applies for a personal loan he/she fills in the application form where he/she has to supply all the data that can be useful for his/her evaluation. Second, he/she has to bring all the official documents: ID card, income/tax declaration, documents related to the collateral he/she is offering. Once the bank receives all the compulsory information, it makes a request to existing external databases (e.g. Central Register, CB) that store information about individuals' payment history in the WE banking system during the last five years. The local operator can see not only if the applicant has been previously granted a loan in any bank in the concerned WE country, but also the kind of loan, the credit exposure, and the payment history till the date of the new request. Furthermore, the score of the applicant calculated by CB is included in the scoring model together with all the characteristics that I mentioned in paragraph 1.1. Calculation of the total exposure defines the authority level (e.g. branch manager, senior credit analyst, head of credit underwriting) that is entitled to take the final decision. Different scorecards are developed on the basis of the product under request. Thus for a MG loan the scorecard will include certain demographic and product characteristics, while for a CL loan it will include other demographic and product variables. Once the score of the applicant is computed, the cut-off values are verified and black or white rating is assigned, where individuals with black rating should be rejected automatically and with white rating can be rejected, accepted or sent to HQ depending on the policy rules (see Table 3). The policy rules check for potential borrowers' age, income, level of indebtedness, type of collateral/guarantors, LTV, negative external and internal databases records, etc. Some of these rules lead to automatic rejection of the application and are named as "black rules". For instance, if applicant's income is lower than a minimum stabilized by the bank or he/she has negative records in CB or in the bank's internal databases, he/she is refused the loan. Other rules, on the contrary, request better examination of the case in HQ, where a more competent expert makes the final decision. In this case, I talk about "HQ rules". A good example of such a case is a person that applies for a CL loan where the tenor of the loan added to the person's age gives a value higher than the retirement age. Another important rule concerns the applicant's level of indebtedness that is equal to applicant's (existing monthly liabilities + new monthly liabilities)/ (monthly available income – CoL). Another HQ rule regards applicant's type of contract, i.e. if the source of his/her monthly income is a short-term contract the loan request must be sent to HQ. The policy rules are defined again according to the WE bank credit policy, target rejection rate and market share objectives.

Table 3

It illustrates the possible outcomes of the intersection between applicant's score and policy rules availability and its impact on the decision level. Whenever some policy rules are violated the column "Policy rules" contains "Yes" (correspondingly there can be some HQ, Black or both types of rules violated), while if all policy rules are satisfied it contains "No" (all Black and HQ rules are satisfied). The intersection of Rating and Policy rules can have Black, HQ faculty or Proposable results. According to the latter the application can be accepted / rejected at a Branch/Regional level or sent to HQ for a better examination.

Rating	Policy rules	Black rule	HQ rule	Result	Branch/ Regional	Headquarters (HQ)
White	Yes	Yes	Any	Black	Reject	//
		No	Yes	HQ faculty	Send to HQ	Accept/Reject
					Reject	//
	No	No	No	Proposable[19]	Accept/Reject	//
Black	Any	Any	Any	Black	Reject	//

Source: WE bank Credit Department

Finally, an individual can be refused a loan because of scarce rating or because of policy rules. As the quantity and the quality of the data in WE are very high, the scoring model has higher impact on rejection rate.

One important fact to mention is the impact of guarantors on the applicant's creditworthiness assessment. In the WE bank practice; an individual can decide whether to bring a guarantor or not. However, if he/she does it, checks on the guarantor in internal and external databases are made. Any negative information about the guarantor has a negative impact on the whole application. Another crucial detail is the relationship of the guarantor to the applicant. An internal research of the WE institution in consideration showed that in 80% of the cases guarantors that are very close relatives (parents, spouses, brothers, sisters) behave as co-applicants in

[19] Proposable means that the application has high rating and no black or HQ rules are abused.

case the applicant runs the risk of default. Hence, such guarantors are defined as "eligible" and are included in the debt ratio. As a result, they can increase the limit of the granted loan as the level of indebtedness (debt ratio) is computed for the whole application, and not only for the applicant.

2.2 CEE banks creditworthiness assessment (a Bulgarian bank as representative example)

As a result of previously mentioned legal and market differences, and insufficient and unreliable data, the underwriting process in the CEE affiliates of the WE bank in question underwent some modifications. The procedure regarding the collection of potential borrowers' data and the corresponding compulsory documents is quite similar. Of course, application forms are different and some information requested in the country of origin of the WE bank in consideration is not relevant for the CEE banks. One significant shortcoming of the CEE banks creditworthiness assessment is the unavailability of CB score (or CB at all) and any other external databases in some of the countries in the region (see paragraph 1.4). As a result, in Bulgaria the declaration of the applicant about his/her existing credits towards other banks and their proper repayment has to be considered as a unique source for the computation of the debt ratio (NCBR started operating 4 months ago). As it was not clear how reliable the generic scorecard was and no CB score was available, the debt ratio was shifted from the policy rules to the credit-scoring model. In such a way, the only source of the applicant's ability to repay (the debt ratio) compensated partially the generality of the scorecard. One inconsistency of the existing CEE banks' creditworthiness assessment, however, is the fact that the debt ratio was transferred to the scoring model, but rules regarding applicant's source of monthly income[20] were maintained among the policy rules. Significant differences regarding the development of the WE and CEE scorecards exist. The WE bank scorecards are statistical and include both borrowers' (demographic) and product characteristics,

[20] Because of inefficient data storing, differentiation among various types of income (e.g. from rent, own assets, additional activities) was not taken into consideration for the debt ratio calculation. Accordingly, any income proved by official documents and coming from long-term labour contract was taken into account. The problem is that, even if officially declared, some sources of income can be more risky than others, and as a consequence corresponding haircut levels have to be applied, while now all types of income are taken or discarded at 100%.

where also demographic variables in each scorecard vary with the type of loan. On the contrary, in the Bulgarian bank, one common demographic and two separate product scorecards (one for MG and one for CL loans) are present. Moreover, these scorecards are generic and do not consider local data (see paragraph 1.4). Additionally, their scores are defined by subjective personal evaluation and without any statistical calculation. One weakness of the existing creditworthiness assessment common to both the WE and Bulgaria is the way multiple applicants are treated. That is, if one application includes more than one applicant, the highest score among all the applicants' scores is considered for further evaluation, while it should be the score of the applicant with the highest income[21]. Finally, in the Bulgarian bank, a total score equal to the sum of maximum demographic score and the product score compared to pre-defined cut-off levels assigns green, yellow, red or black colour. Another main difference between WE and CEE credit scoring models (see above) consists in the effect of the debt ratio on the final rating. In CEE (more precisely in Bulgaria) the debt ratio is compared to a certain threshold of indebtedness (called maximum accepted ratio) that varies with income (applicants with higher income have higher threshold for the debt ratio) and the ability to repay[22] of the applicant is defined. Afterwards a matrix (see Table 4) between applicant's total score and ability to repay gives the final - black or white rating. The last deviation from the WE bank creditworthiness assessment is the formula used in the calculation of the debt ratio: new monthly liabilities/ (monthly available income – monthly CoL – existing monthly liabilities) and the fact that the it is calculated separately for applicants and guarantors where the ability to repay of the guarantors can have a positive impact on the final decision for credit granting.

[21] In the case of multiple applicants, the individual who will contribute mostly for the loan repayment is not the one with the highest demographic score, but the one with the highest income. Therefore, the regular performance or the default of the credit will depend mainly on the "richest" applicant".

[22] Flag ability to repay is a dummy variable that takes a value of 0 if the debt ratio exceeds the corresponding maximum accepted ratio, and a value of 1 if the debt ratio is below this maximum accepted ratio.

Table 4 [23]

Total Score \ Flag ability to repay	Flag ability to repay = 1	Flag ability to repay = 0
Green	White	Black
Yellow	White	Black
Red	White	Black
Black	Black	Black

The second part of the individual creditworthiness assessment includes verification of certain bank's policy rules (see examples in paragraph 3.1).[24] Some of these rules coincide with the WE bank ones, while others are adapted to local economic conditions and legal differences, but are compliant with Basel II and the WE bank practice. At present, there are no Black rules in the Bulgarian bank creditworthiness assessment. The main weakness in this part of the creditworthiness assessment is the lack of any policy rules regarding negative records in external databases (the NBCR is already operating). Furthermore, no rules regulating the treatment of applicants already rejected in the past have been designed. The weight of the rating and the policy rules for the set of rejection rate cut-off is based on the type of product (CL loan or MG loan) and on the reliability of the data. Thus, at the beginning of the improvement more influence will be given to the policy rules in the whole creditworthiness assessment and to the debt ratio in the scoring model, as there are not many defaults and CB score is not reliable or missing.

3. Data set and sample selection

The time span of the credit applications origination varies among the banks as the APS and the implementation of the individual creditworthiness assessment were introduced in different years. Therefore, I evaluate loans requests in the Bulgarian bank between August 2003 and September 2004. The number of observations used for the back testing is different from the one considered in the new scoring models

[23] This is a generalized table for mortgage and consumer loans, where no guarantors are available in the application. Guarantors are not included because if the applicant has black score or 0 ability to repay, guarantors cannot contribute for his acceptance.

[24] However, policy rules do not change application rating. That is, if an application has a white rating before policy rules verification, it should still have white rating even if it is accompanied by some HQ rules and is rejected.

development. Concretely, for the back testing sample I assess only applications that have been granted a loan. In the literature, this fact is known as the "rejection inference bias" (Thomas, L., Edelman, D., and Crook, J., 2002; Crook and Banasik, 2004). The best solution to the problem is to collect from the system (e.g. in other banks) information about applicants that the bank under examination rejected but its competitors accepted (Hand and Henley, 1997). For example, in countries with developed centralized registers such information can be obtained from CB. In such cases the rejection inference bias is overcome by defining the rejected clients as bad, indeterminate or good on the basis of their CB score. At present, such a solution is impossible for the Bulgarian bank scoring model analysis. Other techniques (extrapolation, bivariate probit, augmentation etc.) for dealing with the rejection inference bias exist, but their application does not improve significantly the results[25]. For this reason, during the back testing, I use only accepted (bad and good) applicants. As second best solution for minimizing the bias, during the phase of new model development, I will recalculate the new scores for all of the applicants (rejected and accepted). Finally, I will set cut-offs on the basis of the whole sample. Thus, for the back testing, I use only defaulted and non-defaulted clients, while for the improvement of the existing underwriting process I evaluate both accepted and rejected CL and MG loans applications.

3.1 Datasets

The main problem that I encountered while organizing the dataset was the fact that the information I needed for the analysis has been stored in two separate databases, where no unique key for matching was available. Thus, I had:

- **APS dataset** with 8552 applications processed in the period 1/08/2003 – 09/09/2004;
- **Payment history dataset** with 7141 credits' performance in the period 1/08/2003 – 31/08/2004.

[25] Crook, J., and Banasik, J., 2004. "Does Reject Inference Really Improve the Performance of Application Scoring Models", *Journal of Banking and Finance*, 28, 857-874. Reichert, Cho, and Wagner (1983 argued that "the inclusion of the group of rejected applicants appears to have little information that is useful in classifying marginal credit risks".

The **APS** dataset contains the *application form information*, while the **payment history dataset** - the *monthly payment performance* one.

Furthermore, the payment history information has been prepared in separate files (one for every of the 13 months), which I had to put in one entire database. I established some characteristics available in all these 13 files in order to obtain the real number of observations without running the risk of doubling some records. Therefore, I used multiple-key-match, where **personal number, product type, application date,** and **branch code** could represent a unique combination for each record. After the merge I obtained **7141** observations with payment performance.

I met the same problem while merging the APS and the payment history datasets, where I had to associate the performance of the clients in a univocal manner to the application form variables. I used the same combination of variables (**personal number, product type, application date,** and **branch code**) and obtained the final database **8552** with the corresponding performance for **7141** of them. After data cleaning, I obtained 8274 records, of which 7852 accepted and 422 rejected. Out of these 8274 applications, 7063 were granted a credit, as they signed a loan contract. Table 5 gives information about the distribution and the rejection rate of the whole sample by type of product and total/average loan amount application. Notice that, even if CL applications are more than two times MG applications, their rejection rate is lower than the one for MG, while the Bulgarian bank objective is exactly the opposite. This fact underlines the revision of the cut-offs and the whole creditworthiness assessment. Appendix A includes additional information about sample distribution and descriptive statistics for all the variables in the scoring model.

Table 5

Type of Loan	Total	Total Value of Applications in BGN	Average Value of Applications in BGN	Distribution in %	Rejection Rate
CL	5771	19,145,200	3,317	69.75%	4.99%
MG	2503	86,823,553	34,688	30.25%	5.35%
Total	**8274**	**105,968,753**			**5.10%**

3.2 Variables definition

The whole dataset (8274 applications between 1/08/2003 and 09/09/2004) consisted of about 215 variables divided into:

- socio-demographic and household budget information - 40;
- monthly payment performance history – 80;
- additional variables[26] - 95.

I use the first group of variables - 40 (see Table 6) for the back testing and the new scoring model development and the second group for the definition of bad customers.

Table 6

Variable	Label	Definition
Housing rights (discrete variable)	Score_house_rights1_2	The applicant can live in own house, on rent, with his/her parents, etc.
Education (discrete variable)	SC_level_of_education_old	Level of education – elementary school, high school, bachelor degree, etc.
Time at address (discrete variable)	Score_time_at_address1	The number of years the applicant lives on his/her last address
Time at present occupation (discrete variable)	SC_time_at_occupation_old	The number of years the applicant works on his/her current job
Marital status (discrete variable)	score_marit_status1_2	Whether the applicant is single, divorced, married, etc.
Type of contract (discrete variable)	SC_type_of_contract_old	Whether the applicant has long-term, short-term, seasonal or other contract
Profession (discrete variable)	SC_profession_old	Kind of occupation: doctor, architect, economist, driver, etc.
Salary drawn in the bank (discrete variable)	SC_salary_drawn_old	Whether the applicant has his/her salary is paid on account in the bank and for what period (1, 2, 3, etc. months)
Age (discrete variable)	SC_age_old	Applicant's age
Kind of employment (discrete variable)	SC_kind_of_employment_old	Whether the applicant is self-employed, employee, student, unemployed, etc.
Demographic score (discrete variable)	Demoscorefinal_old	The sum of the of the scores corresponding to the above demographic variables and a constant
Sex (discrete variable)	Sex	Whether the applicant is male or female
Kind of company (discrete variable)	CF1_Kind_of_Company	Whether applicant's employer company is stock company, limited or unlimited company, etc.
Position (discrete variable)	PC1_Position	Applicant's level in company's hierarchy: specialist, manager, director, etc.
Employer origin (discrete variable)	PC1_Employer_ownership	Whether the employer is domestic/foreign private/public company

[26] Variables that concern decision dates, contract date, branch code, decision level (branch or headquarters), decline reason, client's decision, client's total exposure, HQ rules, etc. I use these data for the analysis of the workflow and the application statuses.

Internal loan (MG) (dummy variable)	SC_internal_loan_MG	Takes value 1 if the applicant/guarantor is bank employee and 0 otherwise
Purpose (MG) (discrete variable)	SC_Destination_MG	Whether the loans will be used for repayment of another mortgage, building, reconstruction or purchase of a real estate
Kind of property (MG) (discrete variable)	SC_kind_of_property_MG	Whether the new property will be first house or villa
Own contribution (MG) (discrete variable)	SC_Own_contrib_MG	The percentage of the market value of the real estate that will be financed by the applicant himself/herself
Tenor (MG) (discrete variable)	SC_tenor_MG	The number of years, i.e. instalments of the loan
Currency (MG) (discrete variable)	SC_currency_MG	Whether the loan is in local or foreign (EUR, USD) currency
MG loan score (discrete variable)	MGscore	The sum of the scores of the mortgage variables and a constant
Total score 1 (discrete variable)	Demo_MGscore_old	The sum of the demographic and mortgage product score
Tenor (CL loan) (discrete variable)	SC_number_of_instalment_CS	The number of years, i.e. instalments of the loan
Internal loan (CL loan) (discrete variable)	SC_internal_loan_CL	Whether the loan is in local or foreign (EUR, USD) currency
CL loan score (discrete variable)	CLscore	The sum of the scores of the consumer loan variables and a constant
Total score 2 (discrete variable)	Demo_CLscore_old	The sum of the demographic and consumer loan product score
Destination (CL loan) (discrete variable)	CL1_DESTINATION	Whether the applicant has declared some specific loan purpose or not
Purpose (CL loan) (discrete variable)	CL1_PURPOSE	Whether the loan will be used for the purchase of a computer, furniture, car, etc.
Debt ratio OLD (continuous variable)	Couple_debt_ratio1_2	It is equal to (New Monthly liabilities)/(Net Monthly Income – Monthly Cost of Living – Existing Monthly liabilities
Maximum accepted ratio (discrete variable)	Max_accept_ratio_1_2	Threshold value for the old debt ratio that varies with income – it goes from 31 to 85
Ability to repay (dummy variable)	Ability_to_repay1_2flag	Takes value 1 if the applicant' level of indebtedness allows him/her to repay the new loan
Net monthly income (continuous variable)	Net_monthly_income1, Net_monthly_income2	Applicant's income net of taxes and wage deductions
Couple net monthly income (continuous variable)	Couple_modif_month_income 1_2	The sum of two applicants' incomes if they are a family
New monthly liabilities (continuous variable)	New_monthly_liabilities1	The sum that the applicant is going to pay every month if he/she receives the loan

Existing monthly liabilities (continuous variable)	Total_existing_liabilities1, Total_existing_liabilities2	The sum that the applicant is paying every month on his/her existing loans
Cost of living (continuous variable)	Couple_cost_of_living1_2	Applicant's living expenses for house maintenance, food, beverages, transport, electricity, clothes
Number of HQ rules (discrete variable)	Number_exc	Variable that shows how many HQ rules are present in the application

3.3 Default definition

Whenever a scoring model is tested a particular point in time defined as "observation point" is chosen. The period preceding this point is called an "outcome period", while the period following it, is named "performance period". The first period is the one needed for the loan to mature and to accumulate some payment history, while the second one is important for the classification of borrowers as good or bad on the basis of their status during that period (Thomas, L., Edelman, D., and Crook, J., 2002). On one hand, academicians suggest 12-months length for the performance period and 12-to-18-months for the outcome one. On the other hand, in these two or three years the stability of some variables in the scorecard can change as a result of a population drift[27], which can distort the results. Therefore, shorter time span (e.g. of six months for both periods, or of twelve and six months correspondingly) are suggested. As the time period of my data is relatively short (thirteen months), in order to have significant results, I apply different performance and outcome periods for the classification of bad and good borrowers. Because of increasing lending growth and underdeveloped underwriting techniques in Bulgaria (and the whole CEE), the default probability for an individual loan (mainly CL one) is quite high since the origination of the loan. Consequently, I could identify a bad client even after three months of his loan withdrawal. Hence, I do not consider any explicit performance period for bad loans classification, while the outcome period is equal to 13 months. The situation is different for good clients, where I do consider a 9-month performance period during which both types of loans (MG and CL) need to mature and to confirm themselves as good ones.

[27] A drift in the population is observed when the population characteristics change/evolve over time, so as the distributions change. For example, the number of divorced, younger people or self-employed has increased compared to the past.

The main problem comes out to be the small number of granted MG *(the Bulgarian bank used to be a corporate bank in the past)* with existence history for more than 9 months. Another problem is that a quantitative validation requires a sufficient number of loan defaults. As in the Bulgarian bank under examination the individual underwriting system together with the new APS has been implemented thirteen months ago, I run the risk not to have significant number of defaults (more than 100). Therefore, I adopt two definitions of default, and despite running the analysis with both of them, I take into consideration for the new model development the broader one. The first definition is the one proposed by the New Basel Capital Accord and includes all the loans that were "more than 90 days late in payment in capital and /or interest". This criterion is also in line with the definitions of distressed risk classes as defined by the majority of the Central Banks of CEE. According to this definition, among the applicants that have signed a contract I have:

- Bad records - credits with over 90 days of delay;
- Indeterminate - credits with past due days between 31 and 90 days;
- Few observations – credits that have not been already defined as bad or indeterminate and have null Exposure at the end of the 9th month of existence or exist for less than 9 months;
- Good - all credits not included in the above-mentioned categories and have always been classified as good.

The second and broader definition for bad customers has its base in the literature where Lacour-Little and Malpezzi (2003), for example, defined as bad any client that had experienced some negative event[28]. Therefore, among the granted credit applications I define as:

- Bad - credits with delay in payment for more than 60 days;
- Indeterminate - credits with delay in payment from 31 to 60 days;
- Few observations – credits that have not been already defined as bad or indeterminate and have null Exposure at the end of the 9th month of existence or exist for less than 9 months;

[28] The group of negative events includes any delay in payment (for whatever period), any downgrading of the client or any negative/bad information from internal and external databases (see paragraph 2.2 also).

- Good - all credits that are not included in the above-mentioned categories and have been always classified good.

Finally, whenever a client (bad or good) has more than two credits I took the one with the worst classification. Hence, according to the Basel II definition of bad clients I end up with 1506 observations (1423 good and 83 bad) while the broader definition leads to 1551 observations (1423 good and 128 bad)[29]; where the number of defaults in the MG sample does not allow any significant statistical analysis (I identify only 12 bad MG).

The final bad-good sample size has received significant attention in the literature. Some scholars suggest at least 100 bad and as many good as available, others opt for equal number of bad and good, thirds say that good should be no more than ten times bad (Thomas, Edelman, and Crook, 2002). These multiple points of view are expressed also in the empirical literature where bad-good ratios like 1:4 (Boyd, 1997), 1:8 (Vandell and Thibodeau, 1985); 1:20 (Roszbach, 1998), 2:1 (Gross and Souseles, 2001), etc can be found. Even Crook and Banasik (2004) ran a number of models where the bad-good ratio was between 1:5 and 1:9. In the light of this discussion, one thing is sure, namely that the number of bad can have impact on the significance of the results only in case it is less than 100 or it is too low compared to the number of good (e.g. 1:20). As there is no clear view on the bad-good ratio, for the sake of completeness, I performed the back testing analysis in two ways: a) with 128 bad and 1423 good and b) with ten sub-samples each consisting of 128 bad and 142 (143) good. However, the results came out to be very similar for each of the ten sub-samples, while the average result is exactly the one of the whole sample 1423 good vs. 128 bad (for further comments see paragraph 4.1). Therefore, for the back testing I consider both samples, while for the improvement of the scoring model - only the second one (where the number of bad is above 100).

[29] I will call as bad rate the one corresponding to the Basel II definition of default and bad rate1 the one corresponding to the broader definition.

3.4 Descriptive statistics

Table 7 contains descriptive statistics for the bad-good sample of 1551 records that is used for the back testing and the new model development (see Appendix A – Table 35 for the descriptive statistics of the whole sample of accepted and rejected applicants).

Table 7

Variables	N	Mean	Std. Dev	Minimum	Maximum
Score_house_rights1_2	1551	24,36	9,78	0	30
SC_age_old	1551	-0,97	9,96	-45	10
Score_marit_status1_2	1551	10,52	12,30	-15	20
SC_time_at_occupation_old	1551	0,23	13,75	-20	20
SC_salary_drawn_old	1551	7,72	14,52	0	35
SC_type_of_contract_old	1551	27,56	8,20	0	30
SC_level_of_education_old	1551	3,10	6,05	-15	10
SC_profession_old	1551	5,35	14,83	-25	30
SC_kind_of_employment_old	1551	-0,84	3,05	-30	0
Score_time_at_address1	1551	2,22	10,90	-15	10
Demoscorefinal_old	1551	879,26	42,43	755	995
SC_Destination_MG	230	16,39	11,43	0	25
SC_Own_contrib_MG	230	10,28	3,74	0	20
SC_kind_of_property_MG	230	17,28	11,57	0	25
SC_currency_MG	230	-1,96	7,42	-30	0
SC_tenor_MG	230	10,09	12,00	-20	35
SC_internal_loan_MG	1	25,00	.	25	25
MGscore	230	38,89	20,52	-20	80
Demo_MGscore_old	230	913,02	45,56	785	1060
SC_internal_loan_CL	14	0,00	0,00	0	0
SC_number_of_instalment_CS	1321	-8,63	14,00	-20	15
CLscore	1321	-28,63	14,00	-40	-5
Demo_CLscore_old	1321	851,53	45,81	720	985
Couple_debt_ratio1_2	1551	75,80	627,58	4	9999
Max_accept_ratio_1_2	1551	45,80	12,12	30	85
Ability_to_repay1_2flag	1551	0,85	0,35	0	1
Net_monthly_income1	1551	530,05	688,26	82	14906
Net_monthly_income2	927	305,19	1078,31	0	30371
Couple_modif_month_income1_2	1551	703,07	1107,81	0	31871
New_monthly_liabilities1	1551	139,63	227,67	24	4980
Total_existing_liabilities1	1551	6,83	46,76	-375	1310
Total_existing_liabilities2	1551	2,64	23,63	-375	477
Couple_cost_of_living1_2	1551	277,79	298,83	100	2050
Number_exc	1551	0,10	0,34	0	4

Descriptive statistics about Sex, Kind of company, Position, Employer origin, Destination (CL loan), and Purpose (CL loan) are not shown in the table as these variables have not been included in the existing scorecards, even if information about them has been stored. As far as no score was previously assigned to these characteristics, no quantitative measure for the calculation of mean and standard deviations is available.

According to the remaining descriptive statistics (Appendix A – Table 36, Table 37, and Table 38) and the present structure of the individual underwriting process only 0.08% applications received the worst black score and 7.67% obtained the best green one. The fact that 75% of the sample has yellow score suggests again that cut-offs are not efficient (usually the "middle area", which is neither bad nor good area, should be the smallest one) and four colours for the total score are not necessary (see Appendix A - Table 37). In the mean time, 12.48% couples of applicants (of which 15% with red, 12% with yellow, and 10% with green score) have 0 ability to repay. Thus, the numbers support the fact that the debt ratio was included in the scoring model in order to improve the results of the generic scorecard (see Appendix A - Table 38). Final black rating was assigned to 13.43% of the loan requests, of which 76% have been accepted. Such evidence raises doubts about the usefulness of the whole assessment. The descriptive statistics regarding the workflow of the Bulgarian bank individual creditworthiness assessment show that 18.6% of the applications (of which 48% CL and 52% MG loan requests) have been sent to HQ for a better examination (see Appendix A – Table 36). The overlapping area between policy rules and rating in the rejection region is 21.4%, which is the percentage of black rated requests that faced some HQ rules. Regarding the two products, even if the number of MG applications is lower than the CL ones, more MG have both black rating and exceptions, but the rejection rate is lower (see Appendix A – Table 38).

The empirical analysis of the creditworthiness assessment (scoring model and policy rules) consists mainly of two parts: a) back testing of the existing two CEE banks' individual underwriting processes and b) new credit granting models development. The first phase is described in paragraph four and aims at proving that the present scoring model do not function properly, while the second phase is presented in paragraph five and aims at developing statistical credit scoring models for CEE banks.

4. Back testing methodology and results

Once a scoring system is implemented, questions regarding its quality and stability over time arise. Basel II regulatory principles for validation have been designed to ensure a uniform framework for certification and monitoring of the scoring system in use. However, it is not only a matter of external regulations, but of banks' interest frequently to review their scoring systems. The quality and stability of a scoring system is assessed through its back testing. For verifying the quality of the scoring model, I compare the estimated to the observed individual DR. A scoring system has a high discriminatory power if the high scores turn out to contain only a small percentage of defaulters and a large percentage of non-defaulters and vice versa. For checking the stability of the model, I match the characteristics of the most recent population of applicants up to the characteristics of the development sample. A scoring model is said to be stable if the mean and the standard deviation of its variables are constant over time.

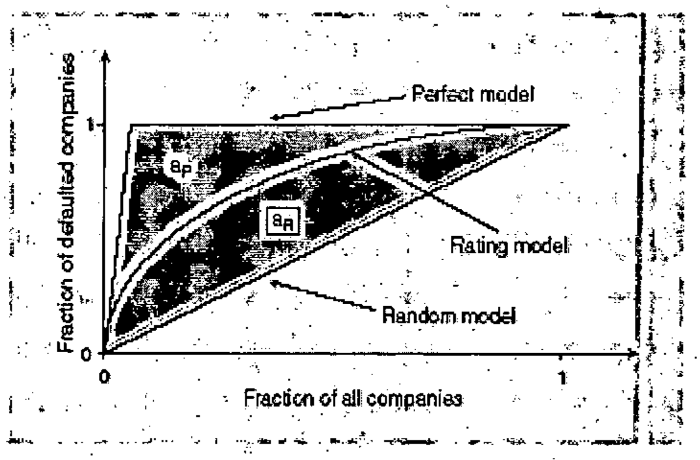
4.1 Scoring model quality back testing – methodology and empirical results

Contingency tables (confusion matrices) and classification errors are among the most basic and efficient tools for understanding the performance of a default prediction model (Hosmer and Lemeshow, 2000; Stein, 2002). Although they do not make any assumptions on the score distribution, they are strictly dependent on the cut-off definition, which makes them less attractive to FI. In fact, the most popular validation techniques of models' discriminatory power in practice are: CAP (Cumulative Accuracy Profiles), ROC (Receiving Operating Characteristic), and their corresponding Gini coefficient (Fritz, Popken, and Wagner, 2002; Engelmann, Hayden, and Tasche,. 2003). All of them provide information on the model performance at any cut-off level. CAP is the one that I use for the back testing of the CEE banks scoring models. Concretely, the CAP curve (Figure 1) is determined by plotting the cumulative percentage of all borrowers ("alarm rate") on the x-axis and the cumulative percentage of defaulted borrowers ("hit rate") on the y-axis. The steeper the slope at the beginning of the curve, the more accurate the model is. A perfect model will discriminate between bad and good clients at 100% and will assign the lowest scores to all defaulters. Its CAP curve would rise linearly before becoming horizontal. A random model will correspond to the 45-degree line and will have 0% discriminant power. Real scoring models are somewhere in the middle. CAP as validation technique aggregates the discriminatory power of a scoring model into a single number the so-called Gini coefficient. In this case the Gini coefficient is also labelled as Accuracy ratio (AR) and is defined as the ratio of the area a_r between the CAP of the scoring model (variable) being validated and the CAP of the random model, and the area a_p between the CAP of the perfect scoring model and the CAP of the random model (Engelmann, Hayden, and Tasche, 2003):

$$\text{Gini coefficient} = AR = a_r / a_p$$

Therefore, the Gini coefficient varies always between –1 and 1.

Figure 1: CAP curve



Source: Bernd Engelmann, Evelyn Hayden and Dirk Tasche (2003)

From a theoretical point of view, for a single characteristic, a Gini coefficient greater than 0.4-0.6 is considered high, and the variable should be taken forward into the variable selection process, while any characteristic with Gini coefficient lower than 0.2 is unlikely to be used in the multivariable analysis. Despite these rules, typically, scoring model quality measures depend on the underlying data. For instance, in case of only socio-demographic information a Gini coefficient of 0.3 is quite high, whereas a scoring system based on individual current account data (used in credit risk monitoring) may yield a Gini coefficient of 0.8 (Ong, 2002). Hence, considering my data specificity and sample peculiarity, I define any scorecard variable with Gini coefficient:

- higher or equal to 0.2 as variable with significant discriminatory power;
- between 0.10 and 0.20 is defined as variable with acceptable discriminatory power;
- lower or equal to 0.10 is defined as variable with insignificant discriminatory power.

The debt ratio variable is analysed following the same methodology, but apart from the other characteristics as it is the only continuous factor that takes part in the scoring model.

Before analysing the distribution of the variables in the current credit-scoring model, I select the characteristics of the person that had impact on the finale decision, i.e. I consider the applicant with the highest demographic score in the application (as implemented in the current underwriting process). The table below shows the two bad rates (according to the two definitions of problematic credits) by each variable from the scoring model (scorecard and debt ratio) and its attributes.

Table 8

DEMOGRAPHIC VARIABLES				
Housing Rights	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Exclusive ownership	1152	74,27%	5,27%	7,99%
Rent from state/private	50	3,22%	8,00%	8,00%
Provided by employer	26	1,68%	8,00%	11,54%
Other	323	20,83%	5,77%	8,98%
Total	**1551**	**100,00%**		
Age	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
low-20	9	0,58%	0,00%	0,00%
21-25	96	6,19%	7,61%	11,46%
26-30	208	13,41%	6,90%	9,13%
31-35	266	17,15%	6,92%	9,02%
36-40	288	18,57%	5,28%	6,60%
41-45	261	16,83%	4,00%	8,05%
46-50	236	15,22%	4,41%	8,05%
51-55	141	9,09%	4,41%	7,80%
56-60	39	2,51%	7,89%	10,26%
61+	7	0,45%	0,00%	0,00%
Total	**1551**	**100,00%**		
Marital Status	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Married (M)	930	59,96%	3,18%	4,95%
Single (S)	431	27,79%	8,83%	11,37%

Divorced (D)	142	9,16%	11,45%	18,31%
Separated (X)	10	0,64%	0,00%	20,00%
Widow/er (W)	38	2,45%	5,71%	13,16%
Total	**1551**	**100,00%**		
Time at address (years)	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
0-5	294	18,96%	7,32%	9,52%
6-10	236	15,22%	5,68%	8,47%
11+	1021	65,83%	4,95%	7,84%
Total	**1551**	**100,00%**		
Time at present occupation (years)	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
0-1	356	22,95%	10,76%	13,76%
2	185	11,93%	8,14%	14,59%
3-4	203	13,09%	3,48%	4,43%
5-6	164	10,57%	3,68%	4,27%
7-15	419	27,01%	3,19%	5,73%
16+	216	13,93%	2,38%	5,09%
Empty	8	0,52%	12,50%	12,50%
Total	**1551**	**100,00%**		
Salary drawn in the bank (months)	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
No salary payment	683	44,04%	6,50%	9,37%
> 1 month	342	22,05%	1,19%	2,92%
Empty	526	33,91%	7,09%	10,27%
Total	**1551**	**100,00%**		
Kind of employment	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Running own Business (B)	80	5,16%	8.97%	11,25%
Renter (C)	2	0,13%	0,00%	0,00%
Employee (E)	1420	91,55%	5,37%	8,17%
Farmer (F)	6	0,39%	0,00%	0,00%
Other (O)	2	0,13%	0,00%	0,00%
Retired (R)	8	0,52%	12,50%	12,50%
Unemployed (U)	0	0,00%	0,00%	0,00%
Student (T)	0	0,00%	0,00%	0,00%
Self Employed (S)	33	2,13%	3,13%	6,06%
Total	**1551**	**100,00%**		
Type of contract	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
No term contract (L)	1425	91,88%	5,35%	8,14%
Term contract (T)	0	0,00%	0,00%	0,00%
Seasonal (S)	0	0,00%	0,00%	0,00%
Other (O)	0	0,00%	0,00%	0,00%
Empty	126	8,12%	7,32%	9,52%
Total	**1551**	**100,00%**		
Level of education	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Elementary (E)	63	4,06%	4,92%	7,94%
High school degree (H)	484	31,21%	6,37%	8.88%

College (I)	86	5,54%	2,41%	5.81%
Master (M)	490	31,59%	5,04%	7,76%
High school no degree (S)	428	27,60%	5,78%	8,64%
Total	**1551**	**100,00%**		
Profession	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
agent/solicitor	15	0,97%	0,00%	0,00%
architect/designer	2	0,13%	0,00%	0,00%
Artist	3	0,19%	16,67%	33,33%
assistant/secretary	14	0,90%	8,33%	21,43%
journalist	4	0,26%	25,00%	25,00%
economist	135	8,70%	8,40%	11,11%
pharmacist	8	0,52%	0,00%	0,00%
hairdresser/beautician	6	0,39%	0,00%	0,00%
public officer	31	2,00%	6,65%	6,45%
computer scientist/ programmer	22	1,42%	4,76%	9,09%
engineer	59	3,80%	5,08%	5,08%
cashier/collector	21	1,35%	9,52%	9,52%
driver	80	5,16%	10,00%	10,00%
clerical	27	1,74%	12,00%	18,52%
accountant	74	4,77%	2,74%	4,05%
cooker/waiter	20	1,29%	15,79%	20,00%
doctor	34	2,19%	0,00%	0,00%
broker/investment adviser	3	0,19%	0,00%	0,00%
mariner/fishermen	5	0,32%	0,00%	0,00%
mechanic	65	4,19%	4,92%	10,77%
teacher	78	5,03%	1,32%	3,84%
nurse/midwife	31	2,00%	6,67%	9,68%
pilot/stewardess	2	0,13%	0,00%	0,00%
policeman	27	1,74%	7,41%	7,41%
politician	0	0,00%	0,00%	0,00%
doorman/watchman/receptionist	10	0,64%	10,00%	10,00%
office worker	32	2,06%	6,45%	9,38%
scientist	5	0,32%	0,00%	0,00%
tourism worker	7	0,45%	28,57%	28,57%
advertising worker	6	0,39%	16,67%	16,67%
bodyguard	4	0,26%	0,00%	0,00%
lawyer	16	1,03%	6,25%	6,25%
psychologist/sociologist	0	0,00%	0,00%	0,00%
editor	0	0,00%	0,00%	0,00%
non-qualified worker	30	1,93%	0,00%	0,00%
qualified worker	374	24,11%	4,44%	8.02%
farmer/gardener/raiser/grower	5	0,32%	0,00%	20.00%
salesman	31	2,00%	10,34%	16.13%
fireman	3	0,19%	0,00%	0,00%
technician	39	2,51%	5,13%	5.13%
civil servant	20	1,29%	0,00%	0,00%
professional soldier	15	0,97%	7,14%	13.33%

others	170	10,96%	4,32%	8,88%
empty	18	1,16%	5,56%	5,56%
Total	**1551**	**100,00%**		

According to the distribution and bad rate1 for each variable and its attributes according to the above table, I can say that:

- **housing rights variable** experience high concentration of applicants into the exclusive ownership group. Thus, even if the bad rate1 for this group is lower than the others, the evidence is not strong. Besides, there are quite many observations with answer "others". Hence, I will suggest some changes in the variable attributes (see paragraph 5.2.1) so as to improve the quality of the collected data and their corresponding distribution (e.g. I would add a "living with parents" attribute - economic conditions in CEE countries have significant impact on children independence; furthermore I will change the "provided by employer" category to "right to use", where one can live in a house provided also by friends, relatives, etc);
- **age variable** attributes (number of groups) are usually expected to follow a U-shaped bad rate. According to the current scorecard, people with age between 46-55 years should be the least risky group, followed by 56-60 and 41-45 years old. However, bad rate1 for 41-45 and 51-55 comes out to be higher than the one for 36-40. Consequently, attention should be paid to the minimum of the function, i.e. whether to expect the lowest bad rate1 at e.g. 46-50 or at 36-40;
- **marital status variable** performs best. The real bad rate1 follows the path expected by the scorecard. Namely, the most risky are divorced and separated applicants, followed by single and widowers;
- **time at address variable** has bad rate1 that follows the trend expected by the current scorecard, but it is not very reliable as more than 65% of the sample is concentrated in the category "11+". For the new model, I would rather split the last group into two categories "10-15" and "16+";
- **time at present occupation variable** presents some contradictory results. Thus, clients that have been working for 2 years at their current job have bad rate1 higher than the ones in the "0-1" group, while the opposite is expected. Surprisingly, bad rate1 for the "7-15" category is one of the highest, while it

should be the least risky group. The results suggest that some further analysis and redistribution of the groups should be made;

- **salary drawn in the bank variable** was not compulsory in the first edition of the APS, therefore about 33.9% of the applicants did not answer. Thus, it cannot be said whether the characteristic behaves according to the expectations or not. I suggest to change the categories of the variable as follows – 0, 1-3, and more than 3 months (4+);
- **kind of employment variable** suffers high concentration of observations in the "employee" (91.55%) attribute is explained both by the economic situation and by the fact that the system was processing only applicants with "long-term" contract. Apart from that, bad rate1 for "running own business" is higher than the one for "employee", but the results are not reliable because of highly concentrated distribution of this characteristic;
- **type of contract variable** does not have any discriminatory power as the majority of the sample has no term contract (91.88%). As different kinds of income have not been considered according to their riskiness (i.e. with different discount factor), loan requests were not processed if applicants did not have "long term contract". Thus, such a distribution is understandable and the variable cannot be judged;
- **level of education variable** has 5 attributes of which "College" and "Elementary" present the lowest distribution. The "master" attribute contains very high percentage of the population (31.59%), as till 1996 the only university degree was master; nowadays it is split into bachelor and master degrees (hence, the master/PhD can be separated from the Bachelor degree attribute). Furthermore, opposed to the expectations, "high school degree" has higher bad rate1 than "high school no degree" and "elementary school";
- **profession variable –** contains some professions that are rarely nominated and are even not present in the bad-good sample so bad rate1 is not very reliable. "Economists" and qualified workers include about 33% of the sample. "Teachers" have a relatively lower bad rate1 (3.84%), while "qualified workers" (374 observations) have a bad rate1 of 8.02%. Attention should be paid to the fairly high bad rate1 of the "economists" (11.11% with

135 observations) and the "others" group (8.88% with 170 records). Hence, I suggest the addition of some professions (such as "public notary" in the "lawyer" category, "artisan/merchant/craftsman", "sportsmen", and "insurer").

As I explained in paragraph 3.2, two separated product cards (one for MG and one for CL loans) exist. Table 9 summarizes all of them with their distributions and bad rates by attributes. In both scorecards the "internal loan" variable is not useful, as the bank decided not to process through APS applications of its employees and managers. Thus, this characteristic is unused in the model and I will not analyse it from now on.

Table 9

PRODUCT VARIABLES				
MORTGAGE LOAN (MG)				
Internal loan	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Internal loan (Y)	1	0,43%	0,00%	0,00%
Not internal loan (N)	229	99,57%	3,13%	5,24%
Total	**230**	**100,00%**		
Number of instalments	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
0-60	16	6,96%	0,00%	0,00%
61-120	90	39,13%	2,30%	5,56%
121-180	122	53,04%	4,13%	4,92%
181+	2	0,87%	0,00%	50,00%
Total	**230**	**100,00%**		
Kind of property	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Main house (M)	159	69,13%	2,56%	4,40%
Other house (O)	71	30,87%	4,35%	7,04%
Total	**230**	**100,00%**		
Destination	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Purchase (P)	144	62,61%	4,26%	6,25%
Construction (C)	17	7,39%	0,00%	0,00%
Reconstruction (R)	65	28,26%	1,59%	4,62%
Replace other bank (B)	4	1,74%	0,00%	0,00%
Total	**230**	**100,00%**		
Currency	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
BGN	215	93,48%	2,38%	4,65%
EUR	15	6,52%	13,33%	13,33%
USD				
CHF				
GBP				

Other				
Total	**230**	**100,00%**		
Own contribution	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
1-10%	16	6,96%	0,00%	0,00%
11-50%	179	77,83%	3,41%	5,03%
51-65%	25	10,87%	4,17%	8,00%
66%	10	4,35%	0,00%	10,00%
Total	**230**	**100,00%**		
CONSUMER LOAN (CL)				
Internal loan	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
Internal loan (Y)	1321	100,00%	5,93%	8,78%
Not internal loan (N)	0	0,00%	0,00%	0,00%
Total	**1321**	**100,00%**		
Number of instalments	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
12-18	49	3,71%	2,17%	8,16%
19-24	384	29,07%	4,52%	6,51%
25-30	4	0,30%	0,00%	0,00%
31-36	92	6,96%	3,30%	4,35%
37-48	65	4,92%	6,25%	7,69%
49+	727	55,03%	7,29%	10,73%
Total	**1321**	**100,00%**		

Only 230 MG loans (of which 12 bad) were included in the bad/good analysis. Thus, the back testing results are not very reliable. In order to make some changes I will follow the new product characteristics and credit risk management intuition. Regarding MG variables distribution and observed bad rate1:

- **number of instalments variable** is mainly composed of tenors of 61-120 (39.13%) months and of 121-180 months (53.04%), which are the most typical MG loans for the Bulgarian bank. The first group has higher bad rate1 as expected, but bad MG observations are not enough for any significant changes in the scorecards on the basis of the back-testing;
- **the kind of property variable** –as expected indicates that "other house" attribute has higher bad rate1 than "main house";
- **destination variable** is not suitable for any comments as most of the loans are requested for the "Purchase of a house" (62.61%), which is anyway expected to be the least risky category;
- **currency variable** - do not provide much credit risk information as 93.48% of the MG are in local currency. Loans in foreign currency are all given in

Euro. Nonetheless, under the agreement signed with the IMF the exchange rate between the local currency (BGN) and the Euro is fixed. Therefore, this variable should be dropped from the scorecard as it does not give any additional information about the risk;

- **own contribution variable** reveals that most of the loans (about 78%) are in the 11-50% range. Still, their bad rate1 is lower than that of the 51-65 % and 66+% groups. Thus, I suggest splitting the 11-50% group into 2-3 categories on the basis of the sample's distribution.

As far as CL loans are concerned the only variable left (apart from internal loan) is "number of instalments". The currency variable in the CL loan scorecard is not relevant as all loans of this type are granted in local currency and I will drop it for the future analysis. Still, loans with tenor shorter than 18 months (8.16%) have bad rate1 higher than loans with 37-48 months tenor (7.69%), while the opposite is expected. Besides, the bank typically gives loans with tenor of 2 or 5 years.

The second component of the Bulgarian bank scoring model is the debt-ratio. The results below (Table 10) correspond to the current formula for the debt ratio calculation, namely: New Monthly Liabilities / (Couple modified monthly income – Monthly Cost of Living – Existing Monthly Liabilities). Table 10 shows that more than 85% of the sample has positive ability to repay. However, both their bad rate1 and bad rate are lower for people with no ability to repay than for applicants with positive ability to repay. This fact can be explained by the unofficial income that a lot people in the country receive. Again, attention should be paid to the number of clients accepted with no ability to repay (14.64% in the bad/good sample), while according to the scoring model in use these applicants had to be rejected.

Table 10

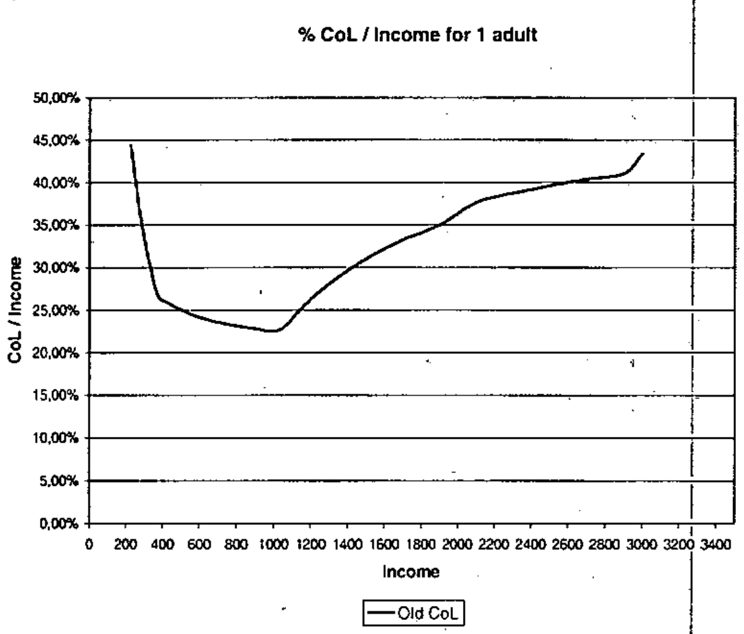
Flag ability to repay	Number of observations	Distribution in %	Bad Rate	Bad Rate1
0	227	14,64%	3,14%	4,85%
1	1324	85,36%	5,92%	8,84%
Total	**1551**			
Debt ratio	**Number of observations**	**Distribution in %**	**Bad Rate**	**Bad Rate1**
0-10	20	1,29%	0,00%	0,00%
11-20	231	14,89%	4,85%	6,49%

21-25	242	15,60%	3,85%	7,02%
26-30	241	15,54%	5,98%	8,71%
31-35	284	18,31%	4,71%	7,39%
36-40	183	11,80%	11,86%	14,75%
41-50	168	10,83%	4,94%	8,33%
51-60	104	6,71%	4,04%	8,65%
61-70	32	2,06%	6,25%	6,25%
81-100	10	0,64%	0,00%	0,00%
101-200	10	0,64%	0,00%	10,00%
201+	9	0,58%	0,00%	0,00%
Total	**1551**	**100,00%**		

Table 10 indicates that about 90% of the observations have debt ratios lower or equal to 50%, where the highest bad rate1 (almost doubled compared to the others) occurred for the persons with debt ratio in the range 36-40%. This evidence supports the hypothesis to determine only two thresholds and to avoid the existing maximum accepted ratio that varies with income. As expected, the lowest bad rate1 corresponds to the low levels of indebtedness. However, as the existing debt ratio formula gives distortion in the results (see paragraph *5.3*), I will apply a new calculation for it.

Concerning the CoL component of the debt ratio, the figure below shows its existing relationship with household income level.

Figure 2



It can be seen that according to the implemented scheme CoL as a percentage of income increases for richer people (income above 1000 BGN), while the opposite should be true. That is, in absolute value CoL should increase with income, but in relative terms it should decrease. Still, the lack of official statistical data (see paragraph 4) explains the inconsistent relationship between household income and CoL. As a result, I will estimate a new CoL.

Finally, Table 11 shows the back-testing results, i.e. the Gini coefficients for each variable in the existing scorecards, calculated on the basis of both "bad rate" and "bad rate1", while Table 12 contains the same information for the second part of the scoring model (couple debt ratio and flag ability to repay). All the variables in Table 11 are treated as discrete; flag ability to repay (Table 12) is a dummy variable, and couple debt ratio (Table 12) is a continuous one.

The final decision on the discriminatory power of each characteristic is based on:

- The level of Gini index for each single indicator,
- The trend of bad rate and bad rate1 by variables' attributes, and
- The number of usable observations - whether the total number of observations per variable is low or high compared to the whole sample.

For example, attention should be paid to "salary drawn in the bank" as it suffers many missing values (low number of usable observation), no matter the high value of its Gini coefficient. Similar is the case of "time at present occupation", where the Gini coefficient is high, but bad rate1 does not follow the expected trend of decrease in risk with increase in time at current job.

Table 11

Variables	Number of Observations	Number of Bad	Gini Coefficient*100 (Flag)	Number of Bad1	Gini Coefficient*100 (Flag1)
Housing rights	1551	83	3,85%	128	2,65%
Age	1551	83	11,25%	128	4,04%
Marital status	1551	83	28,36%	128	28,89%
Time at present occupation	1543	82	32,5%	127	25,56%
Time at present address	1551	83	8,00%	128	4,03%
Salary drawn in the bank	1025	47	26,4%	74	21,40%
Type of contract	1425	74	0,00%	116	0,00%
Kind of employment	1551	83	1,74%	128	0,70%
Level of education	1551	83	5,59%	128	3,60%
Profession	1513	81	6,15%	126	6,25%

Demographic score	**1551**	**83**	**38,96%**	**128**	**32,12%**
Destination MG	230	7	-22,67%	12	-11,12%
Tenor MG	230	7	19,53%	12	11,93%
Own contribution	230	7	-2,63%	12	-13,64%
Kind of property	230	7	12,58%	12	11,39%
Currency MG1	230	7	22,61%	12	10,70%
MG score	**230**	**7**	**18,41%**	**12**	**11,85%**
Demo+MG score	**230**	**7**	**21,04%**	**12**	**15,41%**
Number of instalments	1321	76	14,01%	116	12,78%
CL score	**1321**	**76**	**14,01%**	**116**	**12,78%**
Demo+CL score	**1321**	**76**	**42,94%**	**116**	**36,80%**

The Gini coefficients for bad rate are less significant than the ones for bad rate1 as the problematic credits are less. The above results illustrate that the variables in the current demographic scorecard either do have or do not have any significant discriminatory power (there is no variable in the middle) according to both bad rate and bad rate1:

- Gini coefficient lower than 0.10 (10%) was confirmed for all variables (but three) - "Housing rights", "Age", "Time at present address", "Type of contract", "Kind of employment", "Level of education", "Profession";
- Gini coefficient higher than 0.20 (20%) was verified for "Marital status", "Salary drawn in the bank", and "Time at present occupation".

However, the combination of these individuals' characteristics yields a Gini coefficient of 0.32 (32%) for Demographic score, which suggests that probably most of the variables will remain in the new scorecard, but their weights and attributes (categories) will be reviewed. I will also examine some additional variables, already stored in the system that could be introduced in the new model (see paragraph 5.2).

Product scorecards variables face relatively good discriminatory power. Important fact to mention is the low number of MG used for the back testing and the low number of defaults. Thus, I cannot argue whether the MG scorecard performs well or not. Still, for the sake of coherence among product characteristics, product scorecards, and the bank's credit policy, I will revise and update some of the attributes of the MG variables. Regarding the CL loans, the only variable relevant for the scorecard is "Number of instalments". Besides, it has the same value for the Gini coefficient as CL score. Following the WE bank experience other variables like "Loan purpose" and "Type of purchased good" (new or second hand) could improve the CL loans scorecards discriminatory power.

Table 12

Variables	Number of Observations	Number of Bad	Gini Coefficient*100 (Flag)	Number of Bad1	Gini Coefficient*100 (Flag1)
Couple debt ratio	**1551**	**83**	**-7,64%**	**128**	**-8,37%**
Flag ability to repay	**1551**	**83**	**-5,91%**	**128**	**-6,59%**
Final decision (rating)	**1551**	**83**	**-5,31%**	**128**	**-5,64%**

The relationship between the debt ratio and bad rate1 is negative (as expected), but the Gini coefficient is still below 10%. Clearly, the debt ratio itself is not a predictor of default if it is not compared to the relative thresholds (cut-offs). Therefore, the ability to repay should be the discriminant of good and bad clients. However, its Gini coefficient is insignificant and negative (-6.59 %), while the expected relationship should be the opposite, i.e. if ability to repay is 1, the client is less risky. This evidence firstly supports the hypothesis that the maximum accepted ratio does not have to vary with income (I will substitute it with two thresholds), and secondly that the debt ratio formula should be changed. It should also be kept in mind that 227 applications with no ability to repay have been accepted.

In the end, the discriminatory power of the final rating (black or white) comes out to be quite low according to bad rate1. As the final rating is an intersection between total score (relatively good Gini coefficients between 15.41% and 36.80% depending on the product) and flag ability to repay (controversial Gini coefficient of "– 6.59 %"), it is not surprising that the Gini coefficient of the final rating is even negative "-5.64 %". This can be also explained by two events: a) the low discriminatory power of the "ability to repay" undermines the predictive power of the whole rating (even if it has to strengthen the generic scorecard) and b) 247 applications with black rating have been accepted.

4.2 Scoring model stability back testing - methodology and empirical results

In order to check whether there is any drift in the population characteristics, I make a comparison between the most recent applications sample and the development sample. For the purpose of the stability analysis, both samples should be of similar time span and should contain only accepted applicants. Because of short time history, the development sample has a time span of 13 months (1/08/2003 and 09/09/2004), while the most recent applications sample have a time span of only three months (1/09/2004 and 30/11/2004). Hence, for the sake of homogeneity, I consider only accepted applications during the first four months (1/08/2003 and 30/11/2003) of the development sample and the most recent sample (1/09/2004 and 30/11/2004). Afterwards, I compute the percentage of individuals for each attribute of each variable in the scoring model for both of the samples: development sample of 1728 and most recent sample 3451 (the number is so high compared to the other sample also because of campaigns performed by the bank).

I approach the stability of the system in two ways:

- I compare the two samples distributions by each attribute and examine whether there are any significant differences;
- I calculate the **stability index**.

I firstly compute the stability index for each variable's attribute so as to verify whether the percentage of the population for this attribute in the two samples is significantly different. The stability index formula of the i-th attribute of the j-the variable is as follows:

$$\textbf{Stability index}_{ij}=(a_{ij}-b_{ij})*\ln(a_{ij}/b_{ij})$$

where a_{ij} is the percentage of the most recent population with attribute i-th of variable j-th and b is the percentage of the development sample with attribute i-th of variable j-th.

Secondly, I compute the stability index for every variable (e.g. the j-th variable) that shows the degree of similarity between the two populations:

$$\textbf{Stability index}_j = \sum_{i=1}^{N} (a_{ij} - b_{ij}) * \ln(a_{ij} / b_{ij})$$

where N is the number of attributes of the j-th variable.

The lower the index, the higher the degree of similarity is. Usually, index values below 0.100 mean that the characteristics of the two samples are very similar, whereas values above 0.250 mean that significant differences between the two populations have occurred. Table 13 summarises the stability indexes for all the variables in the demographic scorecard.

Table 13

DEMOGRAPHIC VARIABLES STABILITY INDEX ANALYSIS					
Housing Rights	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
Exclusive ownership	73,15%	71,37%	-0,0178	-0,02460	0,000437282
Rent from state/private	3,88%	4,61%	0,0073	0,17251	0,001259417
Provided by employer	1,62%	0,75%	-0,0087	-0,76581	0,006639286
Other	21,35%	23,27%	0,0191	0,08586	0,001643723
Total	100,00%	100,00%			
			Stability Index =		**0,009979709**
Age	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
low-20	0,58%	0,64%	0,0006	0,09676	5,68866E-05
21-25	6,13%	7,13%	0,0099	0,15019	0,001493083
26-30	13,48%	14,92%	0,0144	0,10143	0,001459985
31-35	17,53%	16,37%	-0,0116	-0,06861	0,000797657
36-40	19,04%	15,50%	-0,0354	-0,20549	0,007267371
41-45	16,44%	16,40%	-0,0003	-0,00208	7,09995E-07
46-50	14,93%	13,94%	-0,0099	-0,06879	0,000682803
51-55	8,85%	10,95%	0,0210	0,21276	0,004466151
56-60	2,49%	3,45%	0,0096	0,32622	0,003131261
61+	0,52%	0,70%	0,0017	0,28913	0,000504871
Total	100,00%	100,00%			
			Stability Index =		**0,019860778**
Marital Status	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
Married/missing (M)	59,95%	59,26%	-0,0070	-0,01167	8,11581E-05
Single (S)	27,84%	28,31%	0,0047	0,01692	8,03677E-05
Divorced (D)	9,20%	9,24%	0,0004	0,00459	1,94092E-06
Separated (X)	0,58%	1,07%	0,0049	0,61663	0,003042773

Widow/er (W)	2,43%	2,12%	-0,0032	-0,13891	0,00043788
Total	100,00%	100,00%			
			Stability	**Index =**	**0,00364412**
Time at address (years)	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
0-5	19,68%	20,31%	0,0064	0,03186	0,000202975
6-10	15,86%	16,46%	0,0060	0,03729	0,000224702
11+	64,47%	63,23%	-0,0124	-0,01941	0,000240653
Total	100,00%	100,00%			
			Stability	**Index =**	**0,00066833**
Time at present occupation (years)	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
0-1	23,26%	22,11%	-0,0115	-0,05089	0,000587491
2	11,46%	10,03%	-0,0143	-0,13353	0,001912455
3-4	13,95%	17,04%	0,0309	0,20023	0,006190691
5-6	10,47%	11,91%	0,0144	0,12840	0,001842564
7-15	26,85%	24,51%	-0,0234	-0,09106	0,002128378
16+	14,00%	14,40%	0,0040	0,02795	0,000110971
Total	100,00%	100,00%			
			Stability	**Index =**	**0,01277255**
Salary drawn in the bank (months)	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
No salary payment	44,33%	18,02%	-0,2630	-0,89994	0,236729081
> 1 months	21,70%	40,92%	0,1921	0,63414	0,121844911
Empty	33,97%	41,06%	0,0709	0,18957	0,01344197
Total	100,00%	100,00%			
			Stability	**Index =**	**0,372015962**
Kind of employment	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
Running own business(B)	5,38%	4,61%	-0,0077	-0,15539	0,001203663
Renter (C)	0,29%	0,17%	-0,0012	-0,50938	0,000588276
Employee (E)	91,20%	92,78%	0,0158	0,01719	0,000271714
Farmer (F)	0,46%	0,29%	-0,0017	-0,46856	0,000811501
Other (O)	0,17%	0,00%	-0,0017	0,00000	0,00000
Retired ®	0,52%	0,84%	0,0032	0,47837	0,001528412
Self Employed (S)	1,97%	1,27%	-0,0069	-0,43387	0,003004985
Students (T)	0,00%	0,03%	0,0003	0,00000	0,00000
Total	100,00%	100,00%			
			Stability	**Index =**	**0,00740855**
Type of contract	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
No term contract (L)	91,55%	92,93%	0,0138	0,01495	0,000206064
empty	8,45%	7,07%	-0,0138	-0,17814	0,002455914
Total	100,00%	100,00%			
			Stability	**Index =**	**0,002661977**
Level of education	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
Elementary (E)	4,46%	3,54%	-0,0092	-0,23148	0,002131529
High school degree (H)	30,96%	27,44%	-0,0352	-0,12067	0,004246667

College (I)	5,32%	4,32%	-0,0101	-0,20954	0,002109005
Master (M)	31,37%	38,95%	0,0758	0,21644	0,016405065
High school no degree (S)	27,89%	25,76%	-0,0213	-0,07955	0,001696618
Total	100,00%	100,00%			
			Stability Index =		0,026588883
Profession	**Development sample (D)**	**Most recent sample (M)**	**(M-D)**	**ln (M/D)**	**(M-D)*ln (M/D)**
agent/solicitor	1,04%	0,90%	-0,0014	-0,14808	0,000212317
architect/designer	0,23%	0,35%	0,0012	0,40691	0,000473011
artist	0,12%	0,09%	-0,0003	-0,28623	8,24624E-05
assistant/secretary	0,81%	0,72%	-0,0009	-0,11188	9,59462E-05
journalist	0,23%	0,46%	0,0023	0,69459	0,001612518
economist	8,10%	10,75%	0,0265	0,28286	0,007491993
pharmacist	0,64%	0,38%	-0,0026	-0,52465	0,001363404
hairdresser/beautician	0,35%	0,20%	-0,0014	-0,53755	0,000776126
public officer	2,08%	3,62%	0,0154	0,55310	0,008511061
computer scientist/programmer	1,22%	1,56%	0,0035	0,25276	0,000883369
engineer	3,82%	4,84%	0,0102	0,23664	0,002413092
cashier/collector	1,39%	1,19%	-0,0020	-0,15618	0,000313655
driver	4,98%	3,33%	-0,0164	-0,40111	0,006596266
clerical	1,56%	1,10%	-0,0046	-0,34995	0,001614564
accountant	4,80%	3,77%	-0,0104	-0,24301	0,002518064
cooker/waiter	1,33%	0,58%	-0,0075	-0,83146	0,006248235
doctor	2,20%	2,61%	0,0041	0,17052	0,000697215
broker/investment adviser	0,17%	0,12%	-0,0006	-0,40402	0,000233129
mariner/fishermen	0,29%	0,38%	0,0009	0,26381	0,000230441
mechanic	3,70%	4,17%	0,0047	0,11923	0,000559192
teacher	4,86%	5,10%	0,0024	0,04797	0,000114576
nurse/midwife	1,85%	1,97%	0,0012	0,06207	7,36126E-05
pilot/stewardess	0,12%	0,12%	0,0000	0,00145	2,42785E-09
policeman	1,74%	1,48%	-0,0026	-0,16107	0,000416012
politician	0,00%	0,03%	0,0003	0,00000	0,00000
doorman/watchman/receptionist	0,75%	0,72%	-0,0003	-0,03777	1,05338E-05
office worker	2,37%	2,09%	-0,0029	-0,12861	0,00036824
scientist	0,35%	0,20%	-0,0014	-0,53755	0,000776126
tourism worker	0,46%	0,58%	0,0012	0,22459	0,000261827
advertising worker	0,58%	0,14%	-0,0043	-1,38485	0,006007716
bodyguard	0,29%	0,23%	-0,0006	-0,22170	0,000127553
lawyer	0,98%	0,61%	-0,0038	-0,48039	0,001802794
psychologist/sociologist	0,00%	0,20%	0,0020	0,00000	0,00000
editor	0,00%	0,06%	0,0006	0,00000	0,00000
non-qualified worker	2,03%	1,91%	-0,0011	-0,05739	6,48387E-05
qualified worker	24,54%	22,49%	-0,0205	-0,08728	0,001789946
farmer/gardener/raiser/grower	0,35%	0,20%	-0,0014	-0,53755	0,000776126
salesman	2,08%	1,51%	-0,0058	-0,32397	0,00186779
fireman	0,17%	0,20%	0,0003	0,15560	4,54793E-05
technician	2,60%	2,64%	0,0003	0,01250	4,093E-06
civil servant	1,33%	1,88%	0,0055	0,34719	0,001918223
professional soldier	1,04%	0,58%	-0,0046	-0,58634	0,002709616

others	11,05%	12,63%	0,0158	0,13367	0,002113018
N/A (empty)	1,39%	1,30%	-0,0008	-0,06309	5,3576E-05
Total	100,00%	100,00%			
			Stability Index		**0,06422776**

The levels of the stability indexes for both attributes and variables (with the exception of "Salary drawn in the bank") are all below 0.01, which demonstrates that no population drift occurred in the last year (2004). Thus, both samples (the first four months of the development sample and the most recent one) have similar distributions. The only exception is "salary drawn in the bank" that has a stability index of 0.37. I explain it with the campaigns run in the recent months where compulsory guarantors were no more asked, but applicants had to receive their salary on account in the bank.

Summarizing paragraphs 4.1 and 4.2, I would say that I extracted the distribution of the attributes for each variable (by checking the DR by each characteristic) and found that in some cases the scores assigned did not correspond to the revealed credit risk (e.g. some of the categories of "age", "time at present occupation", and "CL Number of instalments" variables). Hence, on the basis of the level of the Gini coefficient (CAP validation technique) I demonstrated that creditworthiness assessment of the Bulgarian bank does not have enough discriminatory power and a new improved scoring model should be developed. The demographic scorecard turned out to be stable over the year 2004 (all stability indexes were below 0,100 except one, but the reason is in the introduction of a new product). I performed similar procedure regarding the debt ratio and its parameters, where the results were even more eloquent – the Gini coefficient of the debt ratio is "-8.37%" which outlines its insignificant predictive power. Afterwards, I reviewed the CoL variation with income and found that the usual relationship (it can be seen also in countries where official data exist) between income and CoL is violated. That is, CoL as a percentage of income should decrease whenever income increases as richer people have more money to spend outside their basic living expenses. Finally, I discovered that, even if the final rating was black or the ability to repay was negative, 262 (of which 227 with no ability to repay) applications were accepted and granted a loan.

4.3 Policy rules examination

The empirical analysis that can be done for the revision of the policy rules includes examination of the existing rules and their impact on the final decision. I would like to highlight that, up to now, only Stop and HQ rules were present in the Bulgarian bank individuals' creditworthiness assessment, where Stop rules were preventing the application from further processing and no data were stored. Therefore in my database I have only HQ rules. The examination of the actual rules consists of two three-dimensioned matrices: a) the first one contains rating (see paragraph 2.2), final decision (accepted/rejected), and HQ cases and b) the second one considers rating, final outcome (good/bad), and HQ cases. I will further analyse the number of HQ rules for each applicant and the number of applicants rejected because of poor rating, because of policy rules, and because of both (overlapping of policy rules and rating).

According to Table 38 in Appendix A, and comments in paragraph 3, the overlapping area between policy rules and rating in the rejection region is 21.4%. Table 14 shows that black rating was assigned to 15.86% of the applications and the corresponding bad rate1 is 5.69%, which is quite high relatively to the average bad rate1 of the sample (8.24%). Besides, there are 31 applications rejected by the scorecard and by the policy rules (HQ rules), but accepted by the competent credit analyst. In this manner the results of the creditworthiness assessment are underestimated. Meanwhile, white-rated applications with no HQ rules have bad rate1 higher than the average bad rate1 for the sample, which demonstrates that the current creditworthiness assessment should be improved.

Table 14

Final decision	Flag HQ rule		Bad Rate	Bad Rate1
Black	No	215	3,79%	5,58%
	Yes	31	6,45%	6,45%
White	No	1208	5,90%	8,94%
	Yes	97	4,21%	6,19%
Total		**1551**		

Some of HQ rules have been verified more often than others: e.g. "total working time < 3 months" comprises 2% of the total sample - all of the cases are relative to CL; the "(age + tenor)>retirement age", the "real estate as collateral is not the purchased one" and the "amount of the loan > pre-defined limit" are another set of HQ rules that are often present in MG loan applications. Particular attention should be paid to the fact that these rules are usually in combination with other HQ rules.

Table 15[30] illustrates that one applicant can be subject to more than one HQ rules. As expected, bad rate1 increases with the number of HQ rules. Within the total bad/good sample the higher bad rate1 is verified for clients with two exceptions, however the number of cases is low compared to the case without or with one exception. Bad rate1 is high also for loans without HQ rules, where 236 credits have been granted to applications with black rating, which again demonstrates that the present underwriting system is not applied properly and credit analysts should not ignore the final rating.

Table 15

Total Loans	Decision 1st level	Decision 2nd level	HQ	Bad Rate (HQ)	Bad Rate1 (HQ)
without HQ rules	999	188	236	3,45%	5,08%
with 1 HQ rule	0	0	110	3,70%	5,45%
with 2 HQ rules	0	0	17	11,76%	11,76%
with 3 HQ rules	0	0	0	0,00%	0,00%
with 4 HQ rules	0	0	1	0,00%	0,00%

Finally, in order to check whether the existing cut-offs comprise products riskiness and the Bulgarian bank strategy, I calculate the number of applicants rejected because of poor rating, because of policy rules, and because of both (overlapping of policy rules and rating). Table 16 shows the corresponding rejection rate according to the existing cut-offs for the whole sample.

Table 16

	Frequency	Rejection Rate
Rejected	**10,50%**	**24,40%**
Rejected for Rating (Black)	10,50%	24,40%
HQ	**10,00%**	**14,15%**
Grey for HQ Rules	10,00%	14,15%
Accepted/Proposable	**79,50%**	**1,41%**
Black Rating - HQ Rules	2,94%	26,34%
Final outcome Total sample	**100%**	**5,10%**

[30] See Appendix B table 39 and 40 for HQ rules distribution by product type

As there was no grey rating in the underwriting process in use, the only reason for an application to be sent to HQ was the presence of a HQ rule. It can be seen that for the whole sample the "grey" (HQ) area is only 10% and the "Rejected" area is 10.5%. The only explanation of the low total rejection rate (only 5.10%) is the high number of black rated applications sent to HQ instead of being refused. In fact, it can be argued that the existing "black" area corresponds to the new "grey" one as black-rated applications were rejected or sent to HQ. Also the "Proposable" or directly approvable applications are quite many – 84.65%. The overlapping area between black rating and HQ rules is 2.94%.

As far as the single products are concerned (see Appendix A - Table 41 and Table 42), it can be seen that the "grey" area for MG is much larger than the one for CL. The size of the "Rejection" area is almost the same, while the number of "Proposable" applications is lower for MG, which is expected as the requested loan amounts are higher. The only positive fact is that 28.04% of CL applications rejected by the system are actually rejected, while the corresponding number for MG is 15.48%. Finally, the overlapping area between black rating and HQ rules for MG is almost five times the one for CL.

As a result of the modest discriminatory power of the existing scorecard, the insignificant predictive power of the debt ratio and the final decision, and the low level of rejection rate in the region of black-rated applications with HQ rules, I will develop a new creditworthiness assessment that will substitute the initially created generic model with a statistical one, based on local sample data.

5. New scoring models development – methodology and results

The results of the back testing are the main input for the new scoring model development. The insufficient level of some of the Gini coefficients leads to revision and redefinition of the attributes of the existing scorecard characteristics and the new debt ratio plus insertion of additional scorecard variables (already stored in the system) that I consider relevant for the future analysis.

5.1 New scorecards development – methodology

I carry out a univariate analysis examining all the characteristics (old and new) on a stand-alone basis, I then verify their discriminant power and analyse variables correlation. The decision affecting which variables to be included in the multivariate regression is based on:

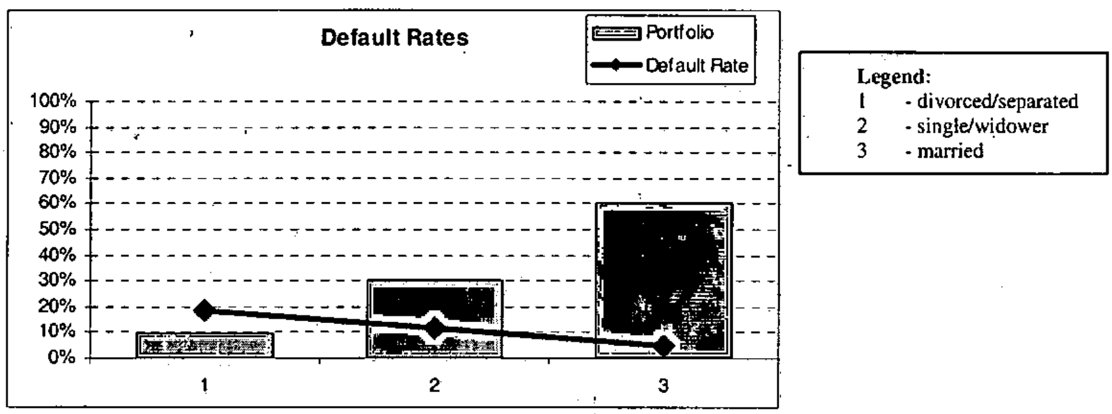
- The level of Gini coefficient for each single indicator;
- The correlation analysis;
- The number of usable observations - whether the total number of observations per variable is low or high compared to the whole sample.

As I mentioned before, I tend to keep out characteristics with Gini coefficients lower than 0.10, but it is not always the case as I also perform checks against the WE bank experience and business sense before I make a final judgement on the usefulness of each variable. I assign a score to each single attributes of every characteristic (variable) in the application form, where the score is related to the risk level, i.e. higher score corresponds to lower riskiness. All the variables that I choose for the next steps of the analysis are transformed and normalized in order to be comparable to each other during the multivariate regression. For that reason I:

- Calculate the DR for each qualitative variable;
- Transform each attribute of the qualitative variable into a score;
- Normalize the transformed scores.

The DR of each attribute of a qualitative variable is equal to the number of defaulted individuals with this attribute divided by the total number of clients that has this attribute. For example, if the sample consists of 1551 customers, of which 930 are married and 46 went into default - the DR for this attribute of the "marital status" variable will be DR=(46/930)*100 = 4.95%. Considering the DR for each attribute of a single variable, I construct the DR line, which is an excellent indicator of the predictive power of a given qualitative variable. A qualitative variable with high predictive power will follow a clear DR trend over its outcomes. If a variable had little power, the DR slope would be much more indeterminate, indicating that there is little relationship between the qualitative variable attributes and default. Figure 3 shows the DR line for the "marital status" variable, where on the x-axis I put the attributes of the variable ordered by their present score: the higher the score of the attribute the lower the riskiness of the client is, i.e. the lower the expected DR. Namely, 1 stays for divorced/separated, 2 – for single/widower, and 3 – for married. The y-axis represents the DR by type of marital status. The grey area in the graph shows the distribution of the whole individual loan portfolio according to the "marital status" attributes. In this case the variable has good discriminatory power, as divorced people experience higher DR than single and married one.

Figure 3



In order to transform and normalize each attribute of a qualitative variable into a score I use the following formula:

$$T(x) = \frac{(-1 \times DR) - \mu}{\sigma} * 50$$

where x is the value of the attribute of the corresponding qualitative variable, DR is the average default frequency for this attribute (in my case bad rate1); μ is the mean of the qualitative variable, and □ is its standard deviation. As a result, each variable has been transformed to a score with mean 0 and standard deviation of 50. Therefore, the regression coefficient of every characteristic can be interpreted as its relative weighting in the final model.

Any time a new scoring model is built, all highly correlated characteristics should be possibly avoided in order to generate a stable statistical model. As a result, before running the multivariate regression analysis, I check the correlations among all the variables candidates for the model through a correlation matrix. Still, for qualitative variables, the level of correlation is a matter of subjective opinion of the academician or the practitioner who is doing the analysis. Thus, I consider variables to be highly correlated if they measure the same risk characteristic and have correlation coefficient higher than 0.500.

One more step before the multivariate analysis is the calculation of the number of usable observations for each variable. I judge whether the total number of observations per variable is low or high compared to the whole sample. Doing so I can have an idea about the possible order of weights of all the variables in the final model no matter what the results of the multivariate analysis are. For example, if a variable has a high Gini coefficient (above 0.20), but suffers a lot of missing values (more than 20%), I will include it in the model, but I will pay attention to its weight. Again, I define subjectively the acceptable level of missing values to be 20%.

I then run logistic regression[31] with forward stepwise selection and backward elimination that screens the available list of independent variables to select only those that are "important" in describing the dependent variable – ln (flag1[32]). I start the analysis with one variable and add/remove variables from the existing scorecard. For this reason, I define a p-entry-value - p_e that is the maximum p-value a variable could have in order to enter the model. Any variable with p-value<0.25 is a candidate for the multivariable model along with all variables of known credit risk management importance. Statisticians (e.g. Hosmer and Lemeshow, 2000) suggest selecting a large significance level so as variables that seem insignificant at a univariate level can come out to be significant while interacting with other variables. Other scholars like Bendel and Afifi (1977) and Lee and Koval (1997) suggest a p_e between 0.15 and 0.20 for forward stepwise logistic regression. At each step I check whether some of the previously added characteristics have become insignificant in

[31] Why logistic regression rather than a liner one? The main pitfall of the linear regression model is that the right-hand side could take values from – □ to + □, but the left-hand side could take only values between 0 and 1. Therefore, both logistic and linear regression models give similar results but until p goes closer to 0 or 1. In other words, the two models produce similar scores except for the cases where the PD is too high or too low, where the logistic approach discriminates better. This is one of the most important factors for the logistic regression being the most successful method among both FI and academicians (Thomas, L., Edelman, D., and Crook, J., 2002). Another fact supporting the use of logistic regression is the highly concentrated distribution in some of the variables.

[32] Flag 1 is a dummy variable that represents the occurrence of bad clients (according to bad rate1), i.e. it takes value 1 if the client is bad and value 0 if the client is good.

the presence of the newly added one. Concretely, I pre-choose a p-removal-value - $p_r > p_e$ that can lead to the exclusion of some previously selected factors. I test the fit of the model after the introduction/elimination of each variable to ensure that the model still adequately fits the data with the help of the Wald test and likelihood ratio chi-square test and it's the corresponding p-values.

The Wald test calculates a z statistic, which squared, yields a Wald statistic with a

$$z = \frac{\hat{E}}{SE}$$

Chi-square distribution.

The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the saturated model (L_1) over the maximized value of the likelihood function for a simpler model (L_0). The log transformation of the likelihood functions leads to a chi-square distributed likelihood-ratio test statistic:

$$\mathbf{-2\log(L_0/L_1) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)}$$

Finally, when all the variables have entered the model or each variable in the model has p-value lower than p_r and higher than p_e, i.e. the introduction of one more variable does not increase significantly the R^2 or increase it in a slower manner, the analysis is completed. Once I come up with a model that includes all the important variables from a statistical and credit risk management point of view, I examine how effectively it describes the outcome variable. Using the Hosmer-Lemeshow statistic methodology (Hosmer and Lemeshow, 2000) I assess the model's goodness-of-fit by dividing the whole sample into 10 ordered groups of obligors and comparing the actual number in the each group (observed) to the number predicted by the logistic regression model (predicted). Thus, I obtain a test statistic that is chi-square distributed, where the non-significant outcome indicates that the model prediction does not significantly differ from the observed one. I create the 10 ordered groups based on their estimated probability - those with estimated probability below 0.1 form one group, and so on, up to those with probability 0.9 to 1.0. I further divide each of these categories into two groups based on the actual observed outcome variable (success, failure). I get the expected frequencies for each of the cells from the model. The null hypothesis in the Hosmer-Lemeshow goodness-of-fit test is that there is no difference between the observed and predicted values of the dependent variable in the test. So if the p-value is large, the null hypothesis cannot be rejected, implying that most of the obligors with success are classified in the higher deciles of risk and those with failure in the lower deciles of risk.

Last but not least, the final model must fulfil the following requirements:

- to have statistically significant predictive power, i.e. must be able to predict future DR;
- to have the structural stability, i.e. to perform well on the N predefined sub-samples of the original data

For this reason, I perform cross validation of the results in order to check the goodness of fit of the model. As there is no testing dataset, I chose the leave-one-out cross validation methodology. In such a way, I set aside a portion of the original dataset as training data and use the remainder of the sample as testing data. The only restriction I put is that the ratio between good and bad borrowers in each sub-sample is homogeneous. I partition the data into 5 predefined equal groups. Four of these groups are training sets, and the last group is the testing set. The idea of the cross validation consists in performing the logistic regression 5 times using the 4/5 of the original data set. At each step I use different portion of the 4/5 of the original data and estimate different coefficients. By making the forecast on the 1/5 of the original dataset using the estimated coefficients I generate the portion of artificial out-of-sample dataset. At the end of the process the out-of-sample dataset had the same size as the original dataset. Finally, I measure the model's predictive power accuracy applying the Gini index on the out-of-sample dataset. I demonstrate the structural stability of the model by comparing the 5 groups of estimated coefficients and finding no significant difference among them.

5.2 New scorecards development – empirical results

5.2.1 Univariate analysis results

The first step of a new scoring model development is the univariate analysis of each single variable candidate for the final model. Therefore, in line with the results of the back testing, I construct new demographic and product scorecards where I:

- include most of the existing model variables;
- regroup into new risk categories[33] the attributes of some of these variables (e.g. age, time at present occupation, etc.) on the basis of bad rate1;
- consider other characteristics (already collected in the system) that could be significant according to credit risk experience and Gini coefficient values.

Before analysing separately each of the possible new scoring model characteristics, I choose the Demographic score (and its components) of the applicant with the highest income[34]. I suggest this change in the model as, from credit risk point of view, the applicant with the highest income is the one that will bear the burden of the monthly-due payment. Table 17 summarizes the values of the Gini coefficients for the existing and new variables according to bad rate1, which is the only one, considered for the new model development (see paragraph 3).

Table 17
The table shows the Gini coefficients for each variable (existing or new) that is candidate for the new model. It further illustrates the risk categories according to the redefined attributes and their corresponding DR (bad rate1).

DEMOGRAPHIC VARIABLES			
Housing rights	Risk Category	Gini Coefficient*100 (Flag1)	DR
Rent from state (RS) Rent from private (RP)	1	2,65%	9,21%
Provided by employer (PE) Other (OO)	2		8,98%
Exclusive ownership (OF) Exclusive ownership (OH)	3		7,99%

[33] The attributes of each variable are grouped into risk categories according to their score, level of risk, and credit risk common sense. Attributes within the same risk category have the same score. Higher risk category stays for higher score, i.e. for lower risk. I will show that variables that do not satisfy this condition have low discriminatory power and do not enter into the new scorecard.

[34] As a result of this change the distribution of the variables, even if not modified, can come out to be different than the initial one. That is, previously, the distribution of each characteristic depended on the maximum Demographic score, while now the score of the applicant with the highest income is will be the one influencing the final decision.

Age	Risk Category	Gini Coefficient*100 (Flag1)	DR
LOW-25	1	7,35%	10,38%
26-30 56-60 61+	2		9,02%
31-35 51-55	3		8,89%
41-45 46-50	4		7,83%
36-40	5		6,62%
Age2	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
LOW-25	1	8,12%	10,38%
26-30 31-35 56-60 61+	2		9,14%
46-50 51-55	3		8,38%
41-45	4		7,25%
36-40	5		6,62%
Marital status	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Divorced (D) Separated (X)	1	28,89%	18,42%
Single (S) Widow/er (W)	2		11,51%
Married/missing (M)	3		4,95%
Time at present occupation (years)	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
0-1 year	1	23,63%	13,45%
2-3 years	2		10,06%
4-10 years	3		5,54%
11+ years	4		5,19%
Time at present occupation2 (years)	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
0-1 years	1	23,27%	13,45%
2-3 years	2		10,06%
4+ years	3		5,37%
Time at present address (years)	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
0-5 years	1	4,03%	9,52%
6-10 years	2		8,47%
11+ years	3		7,84%
Salary drawn in the bank (months)	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
No	1	21,47%	9,44%

1-3 months	2		3,98%
4+ months	3		2,05%
Type of contract	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Seasonal (S) Other (O)	1	-0,46%	0,00% 0.00%
Term contract (T)	2		0,00%
No term contract (L)	3		8,19%
Kind of employment	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Farmer (F) Unemployed (U) Students (T) Renter (C) Other (O)	1	0,76%	0,00%
Running own business (B) Self employed (S)	2		9,38%
Employee (E) Retired (R)	3		8,19%
Level of education	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Elementary (E)	1	4,42%	7,94%
High school no degree (S) High school degree (H)	2		8,88%
College (I) Master (M)	3		7,29%
Profession	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
journalist driver cooker/waiter mariner/fishermen/miners politician doorman/watchman/receptionist bodyguard farmer/gardener/raiser/grower salesman professional soldier sportsman	1	14,99%	11,67%
artist assistant/secretary hairdresser/beautician clerical office worker non-qualified worker fireman	2		10,81%
agent/solicitor cashier/collector broker/investment adviser	3		8,54%

Mechanic nurse/midwife pilot/stewardess policeman tourism worker advertising worker Editor qualified worker technician Others Insurer artisan/merchant/craftsman			
economist public officer Computer scientist/programmer Engineer Scientist psychologist/sociologist civil servant	4		7,87%
architect/designer pharmacist accountant Doctor Teacher Lawyer/Public notary	5		3,09%
Sex	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Male	1	3,04%	8,64%
Female	2		7,68%
Kind of Company	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
ET individual firm COOP cooperative	1	21,44%	12,59%
OOD Limited company SD partnership KD accomandita	2		11,70%
EAD joint stock co. state EOOD Lim.co (1partner) O others	3		7,02%
AD joint stock company	4		4,21%
Kind of Company 2	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
ET individual firm COOP cooperative	1	21,26%	12,59%
OOD Limited company SD partnership KD accomandita	2		11,70%
EAD joint stock co. state EOOD Lim.co (1partner)	3		8,26%

AD joint stock company	4		4,21%
Employer ownership	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Private domestic	1	8,65%	9,11%
Private international Public domestic Others	2		6,50%
Public international State	3		5,32%
Position	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Trainee blue collar worker houseman Clerk other uniform service low post in church hierarchy high post in church hierarchy Other	1	-9,13%	6,54%
specialist specialist in a team uniform officer middle level member of parliament higher post in local administration highest level in government administration	2		8,26%
expert, adviser uniform officer high level	3		6,98%
master Manager Director member of board	4		12,59%
Region	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
085, 086, 087, 0106, 0142, 0143 0162	1	10,64%	17,31%
077, 080, 091, 094, 095, 098, 099, 0107, 0109, 0120, 0125, 0126, 0129, 0131, 0133, 0154, 0165, 0166, 0167, 0169	2		6,43%
081, 083, 084, 090, 0121, 0127, 0128, 0130, 0149, 0157, 0168, 0170	3		7,99%
074, 079, 082, 092, 0105, 0114	4		6,22%
MORTGAGE LOAN VARIABLES			
Destination MG	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Reconstruction (R) Replace by other bank (B)	1	-11,12%	4,35%

Construction (C)	2		0,00%
Purchase (P)	3		6,25%
Tenor MG	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
181+ months	1	11,93%	50,00%
121-180 months	2		4,92%
61-120 months	3		5,56%
0-60 months	4		0,00%
Kind of property MG	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
Other house (O)	1	11,39%	7,04%
Main House (M)	2		4,40%
Cat_currency_MG1	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
EUR/USD/CHF/GBP/other	1	10,70%	13,33%
BGN	2		4,65%
Own contribution MG	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
0 -10%	1	-3,82%	0,00%
11- 20%	2		6,06%
21 - 30%	3		5,97%
31 - 40%	4		5,36%
41%+	5		5,17%
CONSUMER LOAN VARIABLES			
Number of instalment CL	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
60+ months	1	14,86%	11,04%
37-59 months	2		7,38%
0-36 months	3		6,24%
Purpose CL	**Risk Category**	**Gini Coefficient*100 (Flag1)**	**DR**
No purpose travel medical care education other	1	8,48%	10,27%
car, motorbike house maintenance computer	2		9,63%
equipment/furniture	3		8,01%
domestic appliance	4		5,77%

According to the numbers in the above table and the discussion in paragraph 5.1, I can say that:

- **housing rights variable** probably will not enter the final model as its Gini coefficient is 2.65% and most of the answers are highly concentrated in the "own house" category. Because of data collection I could not create any different categories, but I suggest data be collected in a way that allows for future improvement of the discriminatory power of the characteristic, i.e. data should be stored according to the following categories - Exclusive ownership, Rent from state/from private, Living with parents, Right to use/Provided by employer/Cooperative, and Other;
- **age variable** can have two possible regroupings of the existing categories (age and age2), which increases its Gini coefficient to 7.35% and 8.12% correspondingly. Even if the second possibility gives more distinct scores among its categories, it has less sense from credit risk point of view. As I explained in paragraph 4.1, the age variable should follow a U-shaped form. According to the exhibited DR, distribution and scores for each category, I set its minimum at 36-40-age group followed by 41-45 and 46-50. The worst clients come out to be the one with less than 25 years and around the pension age (56 and more);
- **marital status variable** is one of the variables that perform best (its Gini coefficient equals 28.89%). The DR follows the expected path. Some attention should be paid to the lower number of separated and widowers ();
- **time at address variable** experiences a DR that follows the expected trend, but it is not very reliable as the category "11+" contains more than 65% of the sample. I tried to split the last group into: "10-15" and "16+", but the Gini coefficient did not improve and the scores did not follow the expected trend. Therefore, I do not change this variable and I doubt it will enter the final model;
- **time at present occupation variable** is particular as it has high Gini coefficient, but existing scores and groups did not reflect the expected level of risk and common credit risk sense, i.e. clients that had been working for more than 7 years (and less than 10) in the same place had worse payment performance than the ones that did not change their job in the last 3 years. Again, I attempted two possible regroupings (time at present occupation and

time at present occupation2). As the second one has less categories, lower Gini coefficient and higher concentration of clients in the last category (>4 years at current job), I would prefer the first possibility for the multivariate analysis because of its Gini coefficient (23.63%) and better distribution among the categories;

- **salary drawn in the bank variable** underwent some changes of its attributes – the new groups are 0, 1-3, and more than 3 months (4+), where according to the new distribution about 38% of the applicants did not supply this information (see paragraph 5.3.2). Still, the concentration of customers that do not receive their salary in the bank is quite high (67%), but this evidence is going to change with the introduction of the new product requirements (see paragraph 4.2). Apart from that fact, the Gini coefficient is significant 21.47% and the scores follow the expected trend;
- **kind of employment variable** will not enter the new scorecard, because of its insignificant Gini coefficient, high concentration of customers in the employee group, and scores that do not follow the expected trend. However, with the new debt ratio that I am going to calculate, this variable will still affect the scoring model (see paragraph 5.3);
- **type of contract variable** has insignificant Gini coefficient of –0.46% and high concentration in the "long-term contract" employees. Most probably it will not enter the new scorecard, but it affects the new debt ratio;
- **level of education variable** has insignificant Gini coefficient (4.42%), very low number of clients in the "elementary school" attribute and its attributes do not follow the expected trend ("high school degree" and "high school no degree" are riskier than "elementary school"). Probably this evidence will keep the characteristic out of the new model. However, I suggest for the future discriminatory power of this variable that data be collected according to following attributes: Elementary/Primary school, High school no degree, High school no degree, Pre-graduate/College, Bachelor degree, Master/PhD;
- **profession variable** needed some regroupment according to the actual DR (bad rate1) and according to credit risk common sense. Hence, I consider as: riskier than before economists, assistant/secretary, journalist, politician,

hairdresser/ beautician, engineer, cashier/collector, clerical, salesman, professional soldier, cookers/waiters, farmer/gardener/raiser/grower, and doorman/watchman/ receptionist; as less risky than before pilot, scientist, artist, accountant, and agent/solicitor. As the "other group" is too concentrated I would rather to add "public notary" (in the "lawyer" category), dentist (in the "doctor" category), "artisan/merchant/craftsman", "sportsmen", and "insurer" professions after comparisons with the WE bank profession list. The Gini coefficient grows to 15% under the revised risk categories; higher concentration is present in the third and fifth groups, which is normal as the third category contains "qualified worker" that is quite generic and the fifth group includes only professions considered as highly prestigious;

- **sex variable** can become an interesting predictor of default as I found that women tend to be more precise in their credit repayment than men. Even if its Gini coefficient is quite poor (3.04%) in combination with other variables gender can become significant;
- **kind of company variable** has been collected in APS for future purposes, but it not included in the current scorecard. According to the grouping of its attributes suggested by me, "kind of company" achieved significant Gini coefficient of 21.44%, relatively good distribution and scores that follow the expected trend – e.g. Joint stock companies are less risky than limited companies and sole-traders. Finally, only 2% of missing observations are registered. I tried also a second possible grouping of its attributes, where I consider the "other" category as empty but missing values reached 24% which could lead to insignificance in the final model;
- **position variable** has been collected in APS for future purposes, but is not included in the current scorecard. However, I could not to obtain results that were not contradictory for credit risk common sense (scores do not follow the expected trend), i.e. managers and members of the board came out to be the most risky group, which led to negative Gini coefficient of -9.13%. Additionally, the variable suffered 9.3% of missing observations. Hence, it is not possible to use this variable for the new scorecard and it will definitely not enter the final model;

- **employer ownership variable** could be relevant for applicants' stability of income. However, I verify high concentration in the domestic private companies (mainly sole-traders) 67% (which is normal) and 9.3% number of missing values. Even if the Gini coefficient is insignificant (8.65%), it is higher than the one of the other insignificant variables and the scores follow the expected trend;
- **region variable** was identified through the branch code of each application. Usually, the macroeconomic situation of a region (e.g. unemployment level, average salary, etc.) could have an impact on its habitants' payment performance. As branches in Sofia and some of the remaining big cities came out to be the ones with the highest number of problematic credits, it was not possible to obtain groups that reflect the riskiness of each region. Even if the Gini coefficient reached 10.64% and no missing values were present, the scores do not follow the expected trend and the variable cannot be used for the new model.

Regarding product characteristics, I will still define separate scorecards for each kind of product, i.e. one for MG and one for CL loans. In both cases I will not use the "internal loan" variable as bank's employees or managers are not processed in APS. I will drop from the product scorecards also the "currency" variable as it does not give any additional information about the risk, i.e. CL loans are granted only in local currency (BGN) and only 7% of the MG are granted in Euro. Besides, an agreement signed between Bulgaria and the IMF keeps constant the exchange rate between BGN and Euro.

As I verify only 12 problematic MG loans, I am not able to create a statistical scorecard for this product. As a result, I keep the variables in the new MG scorecard almost the same as before. I only regroup the attributes of the "own contribution" variable in order to achieve sensible distribution and expected risk trend.

As far as CL loans are concerned, the only variable that I am left with from the old scorecard is "number of instalments". However, in order to improve its predictive power I regroup its attributes. Another variable that I consider relevant for this product scorecard is "CL loan purpose". It has been collected in APS for future purposes, even if it has not been included in the existing model. Through its elaboration, I achieve scores that reflect the riskiness of the purpose (e.g. buying domestic appliance, furniture, or car is less risky than paying for education or travel). Even if the Gini coefficient is not significant 8.48%, this variable can show important when combined with the "number of instalments".

5.2.2 Correlation analysis

The second step of single variables analysis regards their correlation. For this purpose, I define three correlation matrices (one for each scorecard). The purpose of creating such matrices is to eliminate the highly correlated variables since collinearity reduces the power of the model. I consider only the first possible regrouping for variables "age", "kind of company", and "time at present occupation" (see paragraph 5.2.1). Figure 4 and 5 demonstrate the correlation coefficients for the Demographic and MG scorecards correspondingly. The correlation of the two CL loan variables is 0.014, which is insignificant.

The highest correlation coefficient in the demographic matrix (Figure 4) is between "profession" and "level of education" (0.4468), followed by the one between "position" and "level of education" (0.3614). It is normal that "level of education" "pre-defines" to great extend people's profession. However, the correlation is still below 0.500 and as it will be seen later, "type of education" will not be included in the new scorecard. Besides, all the other coefficients are below 0.30, but such a correlation is not disturbing as it is not significant. Among these, "kind of company" variable shows correlation with "time at present occupation" and "kind of employment".

Figure 4

	Housing rights	Age	Marital status	Time at present occupation	Time at present address	Salary drawn in the bank	Type of contract	Kind of employment	Level of education	Profession	Sex	Kind of Company	Employer ownership	Position	Region
Housing rights	1,00000														
Age	0,22362	1,00000													
Marital status	0,15640	0,07531	1,00000												
Time at present occupation	0,13590	0,26417	0,06768	1,00000											
Time at present address	0,22579	0,07390	0,02963	0,09405	1,00000										
Salary drawn in the bank	-0,04928	-0,01558	0,03991	0,14037	0,01062	1,00000									
Type of contract	-0,02089	-0,06313	-0,04693	0,03732	-0,00191	#DIV/0!	1,00000								
Kind of employment	-0,06614	-0,09339	-0,02182	-0,08339	0,01526	-0,02312	0,28635	1,00000							
Level of education	0,02757	0,05847	-0,00759	0,02400	-0,13969	-0,04313	-0,03563	-0,12784	1,00000						
Profession	0,05697	0,08913	-0,02720	0,10370	-0,02574	-0,00421	-0,02405	-0,09924	0,44068	1,00000					
Sex	0,09593	0,04185	-0,13742	0,01620	0,01583	-0,02131	0,00341	0,02496	0,17358	0,22417	1,00000				
Kind of Company	-0,01111	0,08704	0,03159	0,29562	-0,05020	0,23437	-0,00219	0,25507	-0,03275	0,02305	-0,13387	1,00000			
Employer ownership	-0,02452	0,05603	0,04608	0,25517	-0,04484	0,24065	0,04000	0,03462	0,00040	0,03074	-0,02243	0,28314	1,00000		
Position	-0,02115	0,04961	0,06093	0,01792	-0,08946	-0,07450	-0,03164	-0,08390	0,38144	0,21978	-0,03588	-0,00958	-0,05306	1,00000	
Region	-0,00639	0,00763	0,02561	-0,04726	-0,10716	0,09124	-0,03751	-0,04804	0,08730	0,04806	-0,00211	0,07969	-0,01318	0,03999	1,00000

Figure 5 shows that also correlation among MG variables is low: the highest one is of "- 0.3024" between "tenor" and "destination". Therefore, I can leave all the selected variables in the model.

Figure 5

	Destination MG	Tenor MG	Kind of property MG	Own contribution MG
Destination MG	1,00000			
Tenor MG	-0,30242	1,00000		
Kind of property MG	0,00158	-0,07321	1,00000	
Own contribution MG	0,09590	0,09141	-0,03874	1,00000

5.2.3 Number of usable observations

The last step of single variables analysis consists in verifying the number of usable observations for each characteristic (Table 18). Generally speaking, this is an important measure of variables significance in the final model. Some variables experience higher number of missing values: "salary drawn in the bank" – 37.6%, "kind of company 2" – 24.17%, "employer ownership" and "position" – 9.28%. The most interesting characteristic is "salary drawn in the bank", that suffers many missing values, but has significant Gini coefficient of 21.47%. Therefore, I will pay attention to its weight in the final decision.

Table 18
The table includes number of usable observations (column 2), number of problematic credits (column 3), and Gini coefficients (column 4).

New categories -Demographic variables	Number of Observations	Number of Bad1	Gini Coefficient*100 (Flag1)
Housing rights	1551	128	2,65%
Age	1551	128	7,35%
Age 2	1551	128	8,12%
Marital status	1551	128	28,89%
Time at present occupation	1541	127	23,63%
Time at present occupation 2	1541	127	23,27%
Time at present address	1551	128	4,03%
Salary drawn in the bank	968	71	21,47%
Type of contract	1411	115	-0,46%
Kind of employment	1551	128	0,76%
Level of education	1551	128	4,42%
Profession	1527	127	14,99%
Sex	1549	128	3,32%
Kind of Company	1521	127	21,44%
Kind of Company 2	1176	107	21,26%
Employer ownership	1407	115	8,65%
Position	1407	115	-9,13%
Region	1551	128	10,64%
New categories - CL Loan variables	**Number of Observations**	**Number of Bad1**	**Gini Coefficient*100 (Flag1)**
Number of instalment CL	1321	116	14,86%
Purpose CL	1307	115	8,48%

After applying credit risk common sense and examining the Gini coefficients, the standard deviations, and the number of usable observations, I suggest the following order of the demographic variables according to their weight in the scorecard:

```
1 marital status
2 time at present occupation
3 salary drawn in the bank
4 profession
5 kind of Company
6 age
7 housing rights
8 sex
9 level of education
10 employer ownership
11 time at present address
```

I do not include "Kind of employment" and "type of contract" in the list above as their Gini coefficients are almost 0 and they take part in the debt ratio calculation. "Region" and "position" are not present because the scores do not reflect the level of risk as explained above. I give low importance to "employer ownership" as it faces a lot of missing values compared to the other variables, no matter that its Gini coefficient is even higher than the one for age.

5.2.4 Multivariate analysis – logistic regression - forward stepwise selection and backward elimination

The final step of the credit scoring model (scorecards) development includes the multivariate regression that finds such a combination of characteristics that explains best the default event. For this purpose, I run two separate regressions: one for the demographic and one for the CL loan scorecard. I do not develop any statistical model for MG, as there are only 12 problematic credits of them.

Firstly, I run Demographic logistic regression (relative to the demographic scorecard) considering all variables, where I set p-entry - at 0.25 and p-removal - pr - at 0.35 (see paragraph 5.1). The only variables that enter the model are: marital status, time at present occupation, profession, kind of company, salary drawn in the bank, and sex. The main statistical indicators of the model are: -2LOGL = 795.892 with Chi-square = 87.8649, 6 degrees of freedom and $p < 0.0001$. Such values suggest that the model has good discriminatory power with Gini coefficient of 46.80%. According to the results from the Hosmer-Lemeshow goodness-of-fit test (Chi-square = 4.7532, with degrees of freedom = 8 and $p = 0.7836$), the model fits well the data as the higher the p-value the better the fit is (see paragraph 5.1 for additional comments and Appendix B for the entire output).

I also try different possibilities for the p_e and p_r values. For example, if I increase p_e to 0.45 and p_r - to 0.50, only "level of education", "age" and "type of contract" do not enter the model, but "housing rights" and "employer ownership" have negative coefficients which are contrary to any logic and expectations. Still, this attempt is only for the sake of completeness, as statistically speaking, any variable with p-value above 0.25 is insignificant and not useful for the final model. Hence, through forward stepwise logistic regression only 6 variables come out to be significant. However, according to previous experience "age" is relevant in determining credit risk, so I should include it in the model[35]. Such a choice is supported also by the Gini coefficient of the characteristic, which is the highest (see Table 21) among the insignificant ones (I has already explained that "age 2" and "kind of company 2" are not adequate for further analysis). Hence, I run three regressions, where I set the coefficient for "age" at 0.001, 0.002, and 0.003. Nevertheless, none of the models is convincing as "salary drawn in the bank" receives very high weight (not preferable as it has many missing values) and "time at present occupation" has lower weight than "kind of company".

As a result, I examine another set of models where I change either the coefficient of "salary drawn in the bank" or the one of "time at occupation" or both of them. Appendix B summarizes the comments on their advantages and disadvantages. After examining all the models, I select as final the model where all the weights follow the desired order ("marital status", "time at present occupation", "salary drawn in the bank", "profession", "kind of company", "sex", "age"). The main statistical indicators of the model are: -2LOGL = 800.246 with Chi-square = 61.1844, 5 degrees of freedom and $p < 0.0001$. Hence, the model has good discriminatory power with Gini coefficient of 46.20%. According to the results from the Hosmer-Lemeshow goodness-of-fit test (Chi-square = 4.8221, with degrees of freedom = 8 and $p = 0.7764$), the model fits well the data, i.e. predicted and observed values are very similar. Despite the Gini coefficient and the goodness-of-fit of the first model (without forced variables) are a little bit higher than these ones, a trade-off between statistical results and credit risk management common sense should be made. Therefore, I chose this second model ("age" with 0.002 coefficient and "time at occupation" with 0.085) as the final demographic one.

Secondly, I run *CL loan logistic regression* with p_e=0.25 and p_r=0.35. Both variables come out to be significant. The main statistical indicators of the model are: -2LOGL = 871.687 with Chi-square = 12.0703, 2 degrees of freedom and $p < 0.0024$. Such values suggest that the model has good discriminatory power even if its Gini coefficient is 18.90%, which is acceptable for product variables. As expected,

[35] Remember that academicians argue that one should not follow only the statistical results, but also common sense and experienced knowledge – see Hosmer and Lemeshow, 2000.

"number of instalments CL" has higher weight than "purpose CL". According to the results from the Hosmer-Lemeshow goodness-of-fit test (Chi-square = 5.7807, with degrees of freedom = 2 and $p = 0.5656$), the model fits the data at a moderate level.

I construct a new judgemental scorecard for MG, as the number of problematic credits is very low for any statistical analysis. Thus, on the basis of previous experience of the WE bank and its Bulgarian affiliate, the present product characteristics and policy rules, I include the following variables into the new MG scorecard: "destination", "kind of property", "tenor", and "own contribution". I define the weights and the scores manually considering that most of the defaults in MG loans are identified during the first 2-3 years of their life no matter what the tenor is[36]. That is, if own contribution is high, exposure at default is high.

After the initial transformation and standardization of product scores (MG and CL loan) and Demographic score their means are still 0, however, their standard deviations are different. In order to put them into a new regression and define their weight into the final score, I normalized all the scores to standard deviation 1. Both variables have p-values lower than 0.25. The main statistical indicators of the model are: -2LOGL = 705.965 with Chi-square = 79.8898, 2 degrees of freedom and $p < 0.0001$. Such values suggest that the model has good discriminatory power with Gini coefficient equal to 46.80%. According to the results from the Hosmer-Lemeshow goodness-of-fit test (Chi-square = 5.4593, with degrees of freedom = 2 and $p = 0.7075$), the model fits the data at a significant level.

As no regression is meaningful for MG score and Demographic score together, their weights in the final score are assigned as the ones from the above regression.

Finally, I calculate the scores for each variable and its attributes for the new scorecard.

[36] See also Chiang, R.C., Chow, Y.-F., and Liu, M., 2002

5.2.5 Cross validation

Before saying whether the newly created model is successful, I have to check: first, whether it is able to predict future DR, and second, what is its structural stability (i.e. how does it perform on N predefined sub-samples of the original data). For this reason, I perform three different cross validations: one for Demographic score, one for CL loan score and one for Total score (Demographic plus CL loan). I do not make any cross validation regarding MG score as I do not run any regression concerning it.

For the Demographic model cross validation I divide the whole sample (1551) into 5 sub-samples as explained in paragraph 5.1. Each group has the same DR (bad rate1) of about 9%, where the number of observations is 284 (or 285) good vs. 25 (or 26) bad. The model has better structural stability, as the estimated coefficients of the 5 groups are not significantly different. The predictive power accuracy of the model is 35.44% (its Gini coefficient), which compared to the 46.20% of the model gives positive result.

For the sake of completeness regarding also the discussion in paragraph 3 (for having the number of bad above 100 in each sub-sample), I perform another possible cross validation where I keep the number of bad at 128 and the number of good is 284 (or 285). The model has structural stability, as the estimated coefficients of the 5 groups are not significantly different. The predictive power accuracy of the model is 36.81% (its Gini coefficient), which does not differ a lot from the previous cross validation result and gives again positive result.

For the CL loan model cross validation I divide the whole sample (1321 records) into 5 sub-samples (see again paragraph 5.1). Each group has the same DR (bad rate1) and the number of observations is 241 good vs. 23 (or 24) bad. The model has good structural stability as none of the estimated coefficients of the 5 groups differs significantly from the others. The predictive power accuracy of the out-of-sample is 14.65% (its Gini coefficient), which compared to the 18.90% of the model gives satisfactory results.

Again, I perform second cross validation where I keep always the number of bad at 116 and the number of good at 241. The model has better structural stability than during the previous cross validation, which suggests that the number of bad is quite important (i.e. it should be above 100). The predictive power accuracy of the out-of-sample is significant (18.08%), which is almost the same as the Gini of the model.

I perform cross validation also for the Total score on the already normalized Demographic and CL loan scores. Once more, I use both types of sub-sample division, but the results, as before, are almost the same. The model has significant structural stability, as the estimated coefficients of the 5 groups are very similar. The predictive power accuracy of the out-of-sample is very good – 44.46%, which is very close to the original Gini coefficient of 46.80%.

Not surprisingly, the second cross validation model has higher structural stability, which underlines the importance of the number of bad in each of the sub-samples. The predictive power accuracy of the model is very good - 44.74% (its Gini coefficient), which does not differ a lot from the previous cross validation, result and gives again excellent result.

In conclusion, I can say that the model (demographic and CL loan scorecard) has significant structural stability and predictive accuracy power.

5.3 New Debt ratio

Once I define the new scorecards (demographic and product), I compute and analyse the new debt ratio. In order to correct the CEE banks models, I recalculate the debt ratio as the ratio between total liabilities (existing plus new) and net monthly income net of new CoL. The example in Figure 6 shows the distorting effect of the existing debt ratio in defining the risk of the applicants.

Figure 6

	Applicant 1	Applicant 2
New Monthly liabilities	5	10
Existing Monthly liabilities	6	0
Net monthly income	25	25
Cost of living (monthly)	10	10
Existing debt ratio	0.56	0.67
New debt ratio	0.73	0.67

Notice that according to the old debt ratio, Applicant 1 is less risky than Applicant 2, while according to the new debt ratio it is on the contrary. In reality, Applicant 1 has more liabilities in the banking system and has higher debt to repay so he/she is expected to be more risky than the less indebted Applicant 2.

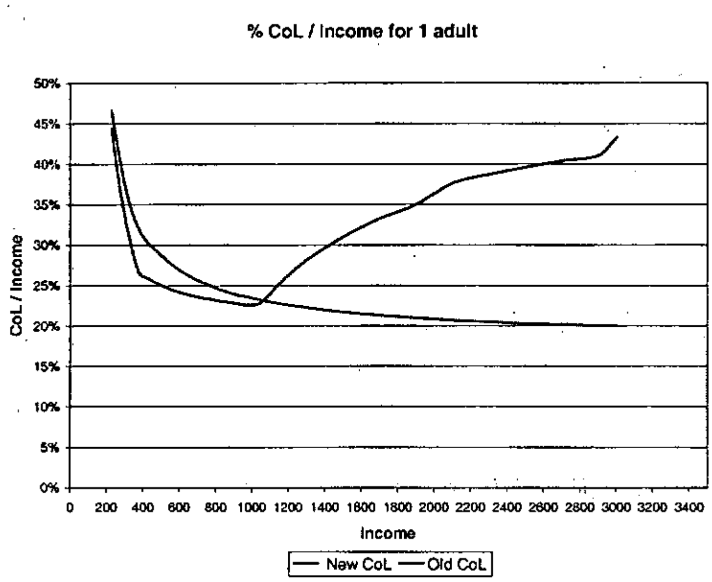
Furthermore, I design a single debt ratio for the whole application where eligible guarantors (parents) are considered (see paragraph 2.1). Eligible guarantors can be only parents as no separation of assets exist and, as I have already mentioned, spouses are obliged to sign the CL/MG loan contract as co-applicants. To improve the coherence of the creditworthiness assessment, I drop the rules regarding the types of income (seasonal, short-term, from dividends, etc.) from the policy rules. I further improve the debt ratio and introduce a table with haircuts for each type of income (e.g. from seasonal contract, from rent, from obligations) when included in the computation (see Vandell and Thibodeau, 1985; Cunningham and Capone, 1991; Jacobson and Roszbach, 2001). I define the haircuts, on the basis of the riskiness of every source of income and after numerous discussions with the employees of the CEE banks. Moreover, up to now, there was slight or no differentiation between various kinds of existing liabilities as necessary official data were available in some of the countries (Poland, Czech Republic and Turkey), while in others (Bulgaria, Croatia, Romania, Slovakia)[37] the only source of such information was applicant's personal declaration. Now, important improvement (even of the WE bank model) is the inclusion of all types of existing liabilities (monthly instalment and non-instalment loans) and their different calculation for the debt ratio. For instance, for existing CL or MG credit, I take the monthly instalment; while for non-instalment loans I use the limit of the loan, multiplied by the average usage rate (AUR) of such a product in the system and by the minimum repayment amount (credit cards). In the

[37] See paragraph 1.4.

case of overdrafts the AUR is multiplied by the product limit and divided by its maturity (usually 12 months). For example, a credit card with 1000 Euro limit, 10% minimum repayment amount, and 40% average usage rate of the limit will contribute to total monthly existing liabilities by 4%x1000=40 Euro. I also take account of applicant's liabilities as guarantor on another loan. In this case his/her liabilities are multiplied by the percentage of non-performing loans (NPL) where a guarantor has intervened (in other words the PD). Thus, if the potential borrower guaranteed a CL loan with monthly instalment of 600 Euro and the above-mentioned percentage is equal to 15, then I will add to his/her total monthly liabilities 600x15%=90 Euro. The loss given default (LGD) cannot be estimated yet, so it is assumed to be 100%.

I also re-estimate CoL after numerous discussions with the employees of the Bulgarian bank. The new CoL follows the expected path, i.e. the CoL as a percentage of household income decreases with income increase. Figure 7 shows the new relationship between CoL and Income.

Figure 7



Once I have defined all the parameters of the new debt ratio I test its predictive power and the following results come out:

Table 25

Variables	Number of Observations	Number of Bad1	Gini Coefficient*100 (Flag1)
Debt ratio with "other income" discounted by 100%	1551	128	-4,57%
Debt ratio with "other income" discounted by 50%	1551	128	-8,30%
Debt ratio with "other income" discounted by 100% and no CoL	1551	128	-16,72%
Debt ratio with "other income" discounted by 50% and no CoL	1551	128	-18,16%

The Gini coefficient suggests that the debt ratio without any consideration of CoL discriminates better than the one with CoL. It is not surprising as CoL numbers are defined subjectively (no official data are available). I will compensate the lack of CoL, though, by tougher cut-offs: while with CoL I could have set the debt ratio cut-offs at 100% and more, now I will set them below 100%. By the way, such a solution (to drop the CoL from the debt ratio) has been recently adopted also by the WE bank. The other point regards the type of income that comes from "other" sources. In the initial and theoretical version of the new debt ratio I has thought not to consider this type of income (i.e. I considered a haircut of 100%). However, the data and the statistical results suggest that "other income" does have some additional value in discriminating between bad and good clients. In fact, it is reasonable, as this income is officially declared, even if its source is not clear. Therefore, I decide to include "other income" in the debt ratio calculation, but with a haircut of 50%. As a result, the final debt ratio (where CoL is neglected and "other" income is considered at 50%) has Gini coefficient of -18,16%, which is more than two times better than the discriminatory power of the existing debt ratio (-8.37%). I expect its discriminatory power to increase further, once data for the relationship between applicants and guarantors is available.[38]

During the final stage of the analysis, after the calculation of the weights of each variable in the new scorecard and the scores assigned to every variable's attribute, I define three colours for the total score – yellow, green, and red - instead of the existing four (see paragraph 2.2) so as to obtain correspondence between total score

[38] In this debt ratio calculation eligible guarantors are not included (only applicants are considered) as no information about the relationship between applicants and guarantors has been collected in the APS.

and final rating (see paragraph 3.1 and Appendix A). I introduce only two thresholds for the level of indebtedness rather than the present multiple ones that vary with income. On the basis of the Gini coefficients of the scorecard and the new debt ratio, and taking into consideration the short time period, I will give a higher weight to the debt ratio in the intersection matrix (Table 26) between score and debt ratio. The final output of the credit-scoring model is white, grey or black rating rather than the existing black and white one (see Table 4).

Table 26

Flag ability to repay / Total Score	Flag ability to repay = 1	Flag ability to repay = 0.5	Flag ability to repay = 0
Green	White	Grey	Black
Yellow	Grey	Grey	Black
Red	Black	Black	Black

5.4 Policy rules review

I revise the existing policy rules and introduce new ones in order to achieve compliance of the Bulgarian bank credit policy with Basel II and the WE bank requirements. As a result, I introduce Black rules that will lead to rejection of the application. However, the application will be still processed till the end and data will be collected for future analysis. The existing Stop rules (regarding wrong data input or non-correspondence) will still prevent the application from further processing and no data will be stored. For the improvement of the policy rules I make numerous comparisons between the WE institution and the Bulgarian bank policy rules. On the basis of environment differences, Black and HQ rates, specific laws and market competition, I decide which rules should be kept, improved, added or discarded. Thus, some of the rules come out to be exactly the same as the WE bank ones (e.g. LTV ratio threshold), while others are adapted to the local reality and therefore differs not only with the WE bank practice, but also among the CEE banks (e.g. the level of minimum accepted income depends on the economic conditions of each country). It is not possible to test some of the new policy rules, as there are no data for doing it. Therefore, the results of the analysis at this stage will be underestimated.

Again I verify that some of the HQ rules are more often encountered than others. It is not surprising as most of the policy rules are the same as before, which means that the comments in paragraph 4.3 are still valid. Table 27[39] illustrates that one applicant can be subject to more than one HQ rule. Bad rate1 does not follow a clear path with the increase of the number of HQ rules. This is a result of the low percentage of applicants with HQ rules (5.86%) and the impossibility to check the effect of some of the new policy rules as no information has been collected (e.g. to check the rules for eligible guarantors).

Table 27

Total loans (MG+CL)	Distribution in %	Bad rate1
without HQ rules	94,13%	8,36%
with 1 HQ rule	4,64%	5,56%
with 2 HQ rules	0,71%	9,09%
with 3 HQ rules	0,45%	14,29%
with 4 HQ rules	0,00%	0,00%
without HQ rules	0,06%	0,00%

As I already mentioned, I can test only some of the defined black rules because of unavailable data in APS. The most frequent black rule is the one regarding the minimum income necessary for whatever type of loan. This rule, however, is more penalizing for CL, as in the case of MG real estate is given as collateral. In fact, I find that only CL suffer black rules (regarding minimum income). I also see that bad rate1 is much higher for clients with one black rule rather than for the ones without any black rule even if only 0.39% of the sample suffers 1 black rule (Table 28).

Table 28

Number of black rules	Distribution in %	Bad Rate1
without black rules	99,42%	8,24%
with 1 black rule	0,39%	16,67%
with 2 black rules	0,19%	0,00%

Finally, Table 29 gives the possible outcomes for the individual creditworthiness assessment after the intersection between applicant's rating and policy rules availability. It shows also which should be the authority level entitled to take the final decision.

[39] See Appendix B - Table 43 and Table 44 for HQ rules distribution by product type

Table 29

It illustrates the possible outcomes of the intersection between applicant's score and policy rules availability and its impact on the decision level. Whenever some policy rules are violated the column "Policy rules" contains "Yes" (correspondingly there can be some HQ, Black or both types of rules violated), while if all policy rules are satisfied it contains "No" (all Black and HQ rules are satisfied). The intersection of Rating and Policy rules can have Black, HQ faculty or Proposable results. According to the latter the application can be accepted / rejected at a Branch/Regional level or sent to HQ for a better examination.

Rating	Policy rules	Black rule	HQ rule	Result	Branch/ Regional	Head quarters
White	Yes	Yes	Yes / No	Black	Reject	//
		No	Yes	HQ faculty	Send to HQ Reject	Accept Reject
	No	No	No	Proposable[40]	Accept Reject	//
Grey	Yes	Yes	Yes / No	Black	Reject	//
		No	Yes	HQ faculty	Send to HQ Reject	Accept Reject
	No	No	No	HQ faculty	Send to HQ Reject	Accept Reject
Black	Any	Any	Any	Black	Reject	//

5.5 Cut-offs definition

Following the discussion in paragraph 1.1, it is not enough to develop a new model and to calculate the scores relative to the new scorecards, but I should decide which clients to accept and which ones to reject. In fact, the final stage of the analysis of the Bulgarian bank individual creditworthiness assessment analysis is the definition of cut-offs that reflect its credit policy and market share objectives. I set these cut-offs on the basis of the targeted rejection rate by the bank under the local market conditions, as no data are still available (even in WE) for the construction of profit maximization and cost minimization function. First, I calculate the new score for all applications in the dataset (accepted and rejected) in order to decrease the "rejection region bias" and to see how many of the previously rejected loan requests would have a white rating now, and how many of the previously accepted applications would have a new black rating. Second, I determine the range of these scores (see table 30) and I apply different cut-off levels (Table 31, Table 32, Table 33, and Table 34) for total score (demographic plus product score) according to product type. Third, I set unique cut-offs for the debt ratio (it can take any value between $-\infty$ and $+\infty$) for all types of credit as the

[40] Proposable means that the application has high rating and no black or HQ rules are abused.

ability to repay of the applicants depends on their total level of indebtedness and not on the currently requested credit.

Table 30

This table shows what applicants' score range is according to the new statistical scorecard (apart from MG where the scorecard is still judgemental) and the requested product':

Score range	Demographic score	Total score = Demographic + CL	Total score = Demographic + MG
Minimum	-2,71	-2,70	-3,07
Maximum	2,32	2,57	2,27

Before setting the cut-offs, I hypothesize the rejection rates for the new white and grey areas to be the same as the ones for the actual white and the black areas (see paragraph 4.3). While the new black area will have a 100% rejection rate. However, I should underline that, as some of the policy rules (Black and HQ) cannot be tested, the grey and black areas will be of slightly different size after some testing period.

According to Table 31 and Table 32, the white area for MG (88.81%) is much broader than the one for CL (60.01%). I adopted this strategy after requests by the Bulgarian bank for stricter cut-offs and higher rejection rate for CL. Correspondingly the grey area (HQ level) for CL is much higher than the one for MG loans. Finally, the effect of the debt ratio on the final black rating is between 1% (for CL) and 2.30% (for MG because of the higher loan amounts).

Legend:
- white area
- grey area
- black area

Table 31

	Score \ DR	DR<50%	50%<=DR<=75%	DR>75%
MG Loans	Green (>-1.7)	88,81%	6,59%	2,28%
	Yellow (>=-2.7 and <=-1.7)	1,88%	0,32%	0,04%
	Red (<-2.7)	0,08%	0,00%	0,00%

Table 32

	Score \ DR	DR<50%	50%<=DR<=75%	DR>75%
CL Loans	Green (>-0.15)	60,01%	0,62%	0,33%
	Yellow (>=-1.2 and <=-0.15)	28,76%	0,33%	0,14%
	Red (<-1.2)	9,62%	0,16%	0,03%

According to the new cut-off levels I achieve higher rejection rate for CL - around 20.15% than for MG - about 7.36% (Table 30a, Table 31a).

Table 33

	Freq	Rejection Rate
Mortgage loans	**2503**	**7,36%**
Reject	**3,20%**	**100,00%**
- Reject only for Rating	2,16%	100,00%
- Reject only for Black Rules	0,80%	100,00%
- Reject for both Rating and Black Rules	0,24%	100,00%
Head Quarter	**20,02%**	**13,57%**
- Grey only for Rating	6,51%	15,48%
- Grey only for HQ Rules	11,31%	11,29%
- Grey for both Rating and HQ Rules	2,20%	19,63%
Proposable	**76,79%**	**1,89%**
Black rating - Rules Overlapping	1,24%	
Grey rating - HQ Rules Overlapping	2,20%	

Table 34

	Freq	Rejection Rate
Consumer Loan	5771	20,15%
Reject	10,69%	100,00%
- Reject only for Rating	10,05%	100,00%
- Reject only for Black Rules	0,42%	100,00%
- Reject for both Rating and Black Rules	0,23%	100,00%
Head Quarter	31,23%	27,98%
- Grey only for Rating	28,47%	28,04%
- Grey only for HQ Rules	1,75%	20,07%
- Grey for both Rating and HQ Rules	1,01%	40,00%
Proposable	58,08%	1,25%
Black rating - Rules Overlapping	0,42%	
Grey rating - HQ Rules Overlapping	1,01%	

A simple comparison to the WE bank rejection rates shows that more CL are rejected in WE (35.64%), while for MG the situation is balanced. I explain the lower overall rejection rate of the Bulgarian bank with: a) high speed of development of retail banking segment in Bulgaria and CEE as a whole and b) tough competition between local and foreign banks. Still, the percentage of directly rejected applicants is 3.20% for MG vs. 10.65% for CL. Even if competition on Bulgarian individual market is severe and the overall rejection rate is lower (compared to WE), I still prefer to keep the Bulgarian bank on alert and to examine dubious applications at a HQ level (20.02% for MG vs. 31.23% for CL go to the Bulgarian bank HQ).

The new cut-offs correspond to local market conditions and the Bulgarian bank credit policy regarding cut-offs differentiation by product type. In such a way, the new creditworthiness assessment considers Bulgarian data sample and is adapted to the Bulgarian market environment.

In the last paragraph, I summarize the strengths and the limits of the new creditworthiness assessment highlighting the importance of its further improvement.

6. Conclusions, implications, and further improvements

In this last paragraph I summarize the results and the main conclusions of the analysis. I then discuss the implications of the study for the development and improvement of retail banking in CEE. Finally, I concentrate on the limitations of the paper and on the propositions for future studies and fine-tuning of individual creditworthiness assessment in CEE.

This paper demonstrates that the currently implemented individual creditworthiness assessment (scoring model and policy rules) in CEE has many pitfalls because of the undeveloped CEE retail banking markets, the lack of any decent local sample data, and the inexistence of external databases with information about clients level of indebtedness in the national banking system (e.g. CB and NCBR). I showed that the implemented generic (not based on local sample data) and judgemental (scores were assigned by subjective personal evaluation and not after statistical calculations) scorecards do not have enough discriminatory power and need further improvements. Even if the performance history of the analysed individual sample was of only thirteen months (which gives a static rather than a dynamic basis for analysis) and I adopted a broader definition of default (more than 60 days of delay), the development of the new model has its strengths and positive policy implications. The newly created model is the first statistical model for individual credit evaluation in CEE. It considers local data sample and represents the local individual loans market conditions in contrast to the firstly implemented model. I calculate and assign the new scores after statistical analysis and computations. I correct and improve the existing debt ratio so as to reflect better the risk of different income and liabilities sources. I create a grey final rating (area separating black (automatically rejected) and white (automatically accepted) rated loan applications) comprising requests that cannot be directly identified as potential bad or good clients, but have to be considered for better examination at a HQ level. Thanks to confidential WE bank information about its own and the CEE individual creditworthiness assessments, I make a complete analysis also of the existing policy rules and their compliance with Basel II requirements. Finally, I set cut-offs that consider both product risk and market competition.

However, even if with significantly superior discriminatory power (46.56 % vs. 32.12% for the new scorecards and –18.16% vs. –8.37% for the new debt ratio), the new credit scoring model can be further improved after the development of CB register and CB score, which are among the best predictors of default in all WE and USA. At present such powerful instrument as CB can be used only at credit policy level, where general rules regarding number of negative records, total past-due amount and cumulative past-due days are introduced. Sophisticated individual underwriting systems in CEE are of prior importance for CEE where individual loans boom, resulting from stable or increasing GDP growth, is expected to lead to increased non-performing loans in the mid-term. Such a dynamic environment characterised by strong competition, elevated demand, and high retail loans growth may translate into weakening of loan eligibility requirements and eventually may cause selection problems and loans quality deterioration. This is also confirmed by fact that in April 2005 the IMF and BNB put severe restrictions on credit growth with a particular attention on individual loans. This paper can be used as a starting point for revision, testing and improvement of individual creditworthiness assessment in CEE. Such validation and fine-tuning will guaranty stability of the banking system and more precisely of the retail market. The obtained discriminatory power of the new model (even if significantly bettered) should give an explicit signal to CEE banking systems about the importance, the utility, and the urgency of a Credit Bureau foundation and CB score computation for the improvement of individual loans banks' portfolios. Longer time span, better data storing and available/reliable CB score in the future will allow the creation of more refined credit scoring models for all type of individual loans. These models will be more compliant with Basel II requirements as data will give the possibility to use various pooling techniques in order to design rating classes also for individuals.

References

Altman, E., 1968. "Financial ratios, Discriminant Analysis and the Prediction of Corporate bankruptcy", *Journal of Finance* 23, 589-609

Altman, E., 2002. "Bankruptcy, Credit Risk, and High Yield Junk Bonds", *Blackwell Publishers*

Altman, E., Marco, G., and Varetto, F., 1994. "Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (Italian Experience)", *Journal of Banking and Finance* 18 (3), 505-52.

Altman, E., Brady, B., Resti, A., and Sironi, A., 2003. "The Link between Default and Recovery Rates: Theory, Empirical Evidence and Implications", *Working paper*

Back, B., Laitinen T., Sere, K., and Van Wezel, M., 1996. "Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit, Analysis, and Genetic Algorithms", TUCS Technical Report No 40, September

Barisitz, S., 2001. "The Development of the Romanian and Bulgarian Banking Sectors since 1990", *Focus on Transition 1/2001*

Bendel and Afifi, 1977. "Comparison of stopping rules in forward "stepwise" regression", *Journal of American Statistical Association,* 72 (357), 46-73

Berlin, M., and Mester, L., 2004. "Credit Card Rates and Consumer Search", *Review of Financial Economics* 13, 179-198

Black, F. and Scholes, M., 1973. "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81 637-659

Boyd, L.A., 1997. "Discrimination in Mortgage Lending: The Impact on Minority Defaults in the Stafford Loan Program", *The Quarterly Review of Economics and Finance* Vol. 37, No. 1, 23-37

Calem, P.S., and Lacour-Little, M., 2004. "Risk-based Capital Requirements for Mortgage Loans", *Journal of Banking and Finance* 28, 647-672

Capozza, D.R., Kazarian, D., and Thomson, T.A., 1997. "Mortgage Default in Local Markets", *Real Estate Economics,* V25, 4, 631-655

Carey, M., 2001. "Some Evidence on the Consistency of Banks' Internal Credit Ratings", Federal Reserve Board

Cavinato I., 2000. "Costruzione di un Modello Interno di Rating: Le Banche della Nuova Europa", Final thesis

CEE Planning Division, 2004. "New Europe Quarterly Economic Update", No 17, October 2004

Chakravarty, S., and Scott, J.S., 1998. "Relationships and Rationing in Consumer Loans", *Working paper*

Chandler, G. G., and Coffman, J. Y., 1979. "A Comparative Analysis of Empirical Versus Judgemental Credit Evaluation", Journal of Retail Banking 1, No.2, 15-26

Chiang, R.C., Chow, Y.-F., and Liu, M., 2002. "Residential Mortgage Lending and Borrower Risk: The Relationship between Mortgage Spreads and Individual Characteristics", *Journal of Real Estate Finance and Economics* 25:1, 5-32

Coats, P. And Fant, L., 1993. "Recognizing Financial Distress Patterns Using Neural network

Crook, J., and Banasik, J., 2004. "Does Reject Inference Really Improve the Performance of Application Scoring Models", *Journal of Banking and Finance,* 28, 857-874

Crosbie, P. And Bohn, J., 2003. "Modelling Default Risk: Modelling methodology", *Moody's KMV Company*

Cunningham, D., and Capone, C., 1990. "The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages", *Journal of Finance* 45(5), 1687-1703

Czech National Bank, 2003. "Banking Supervision 2003"

Damutz, E. C., and Gabbai, D. N., 2001. "The Polish Banking System", *Brown Brothers Harriman & Co.* September 2001

Engelmann, B., Hayden, E., and Tasche, D., 2003. "Testing Rating Accuracy", *www.risk.net (January 2003 risk)*

Falkenstein, E., 2002. "Credit Scoring for Corporate Debt Chapter 8 of "Credit ratings: Methodologies, Rationale and Default Risk" by Ong, M., Risk Waters Group

Fisher, R.A., 1936. "The use of multiple measurement in taxonomic problems", Annals of Eugenics 7, 179-188

Fritz, S.G., Popken, L., and Wagner, C., 2002. "Scoring and Validating Techniques for Credit Risk Rating Systems, Risk Analytics and Instruments", Chapter 9 of "Credit ratings: Methodologies, Rationale and Default Risk" by Ong, M., *Risk Waters Group*

Gardó, S., 2003. "Banking markets in central and eastern Europe (III): Slovakia – from truant to star pupil?", *Die Bank 5/2003*

Goldberg, G.M., and Harding, J.P., 2003. "Investment Characteristics of Low- and Moderate-Income Mortgage Loans", *Journal of Housing Economics* 12, 151-180

Gross, D.B., and Souleles, N.S., 2002. "An Empirical Analysis of Personal Bankruptcy and Delinquency", *The review of Financial Studies* Vol. 15, No. 1, 319-347

Hand, D.J. and Henley, W.E., 1996. "A k-Nearest –Neighbour Classifier for Assessing Consumer Credit Risk", *The Statistician,* Vol.45, No1, 77-95

Hand, D.J. and Henley, W.E., 1997. "Statistical Classification Methods in Consumer Credit scoring: a Review", *Journal of Royal Statistical Society,* Series A (Statistics in Society), vol. 160, No. 3, 523-541

Hillegeist, S., Keating, E., Cram, D., and Lundstedt, K., 2003. "Assessing the Probability of Bankruptcy", *Working paper, September*

Hosmer, D.W., and Lemeshow, S., "Applied Logistic Regression", *John Wiley & Sons, Inc., 2000*

Hsia, D.C., 1978. "Bank Credit Scoring and the Equal Credit Opportunity Act", *Hast. Law J.*, 30, 371-448

International Monetary Fund, 2005 (May). "Bulgaria: First Review under the Stand-By Arrangement and Request for Waiver of Performance Criteria—Staff Report; Staff statement; Press Release on the Executive Board Discussion; and Statement by the Executive Director for Bulgaria", *IMF Country report No.05/ 169.*

Jacobson, T., and Roszbach, K., 2003. "Bank Lending Policy, Credit scoring and Value-at-Risk", *Journal of Banking and Finance* 27, 615-633

Jappelli, T., and Pagano, M., 1989. "Consumption and Capital Market Imperfections: an International Comparison", *The American Economic Review,* Vol.79, No.5, 1088-1105,

Jonkhart, M., 1979. "On the Term structure of Interest Rates and the Risk of Default", *Journal of Banking and Finance* 3 (3), 253-262

Kamakura Corporation. 2004. "A Basel II-Compliant Multiple Models Default Probability Service", *Kamakura Risk Information Services: Credit Risk Overview, www.kamakuraco.com*

Kau, J.B., Keenan, D.C., and Kim, T., 1994. "Default Probabilities for Mortgages", *Journal of Urban Economics*, 35, 278-296

Kealhofer, S., 2003. "Quantifying Credit Risk: default Prediction", *AIMR, January/February*

Knapp, L.G., and Seaks Terry G., 1992. "An analysis of the Probability of Default on Federally Guaranteed Student Loans", *The Review of Economics and Statistics* Vol. 74, No. 3, 404-411

Lacour-Little, M, and Malpezzi, S., 2003. "Appraisal Quality and Residential Mortgage Default: Evidence from Alaska", *Journal of Real Estate Finance and Economics* 27:2, 211-233

Lee, K. I. and Koval, J. J., 1997. "Determination of the Best Significance Level in Forward Stepwise Logistic Regression", *Communications in Statistics: Simulations and Computations*, 26 (2), 559-575

Lown, C., and Persiani, S., 1996. "The Behaviour of Consumer Loan Rates during the 1990 Credit Slowdawn", *Journal of Banking and Finance* 20, 1673-1694

Malhotra, R., and Malhotra, D.K., 2003. "Evaluating Consumer Loans Using Neural Networks", *The International Journal of management science*, Omega 31, 83-96

Martin, D., 1997. "Early Warning of Bank Failure: A Logit Regression Approach", *Journal of Banking and Finance* 13, 249-276

Matherson, J.E., and Winker, R.L., 1976. "Scoring Rules for Continuous Probability Distributions", *Management Science* Vol. 22, No. 10, 1087-1096

Merton R., 1974, "On the Pricing of Corporate Debt", *Journal of Finance* 29 (3), 449-470

Mester, L. 1997. "What's the Point of Credit scoring?", Federal Reserve Bank of Philadelphia, *Business review*, September/October, 3-16

Modelli di scoring (capitolo 9)

Ong, M., 2002. "Credit ratings: Methodologies, Rationale and Default Risk", *Risk Waters Group*

Orgler, Y.E., "A Credit scoring Model for Commercial Loans", *Journal of Money, Credit and Banking* Vol. 2, No. 4, 435-445

Park, S., 1997. "Effects of Price Competition in the Credit Card Industry", Federal Reserve Bank of New York, *Economic Letters* 57, 79-85

Quercia, R.G., and Stegman, M.A., 1992. "Residential Mortgage Default: A Review of the Literature", *Journal of Housing Research,* Vol.3, Issue 2

Quigley, J.M., and Van Order, R., 1992. "More on the Efficiency of the Market for Single Family Homes: Default." *Center for Real State and Urban Economics. University of California, Berkeley. Mimeo.*

Resti, A., and Omacini, C., "Testing for the Consistency of Internal Rating Procedures: An Empirical Exercise Based on Statistical Models", *Working Paper*

Reichert, A. K., Cho, C. C., and Wagner, G.M., 1983. "An Examination of the Conceptual Issues Involved in Developing Credit Scoring Models", *Journal of Business and Economic Statistics* 1, 101-114

Romanian National Bank, 2003. "Annual Report 2003"

Rosenberg, E., and Gleit, A., 1994. "Quantitative methods in Credit Management: A Survey", *Ops. Research,* 42, 589-613

Roszbach, K., 1998. "Bank Lending Policy, Credit scoring and the Survival of Loans", www.defaultrisk.com

Sandor, R., and Sosin,.H., 1975. "The Determinants of Mortgage Risk Premiums: A Case Study of the Portfolio of a Savings And Loan Association", *The Journal of\Business* 48(1), 27-38

Saunders, A., and Allen, L., "Credit Risk Measurement", *John Wiley & Sons, Inc., 2002*

Schardax, R. and Reininger, T., "The Financial Sector in Five Central and Eastern European Countries: An Overview", *Focus on Transition 1/2001*

Scott, J., 1981. "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models", *Journal of Banking and Finance* September, 317-344

Sironi, A. Capitolo 17 "Il credito: evoluzione e innovazione nei criteri di valutazione" da "Corporate and Investment Banking"

Stein, R. M., 2003. "Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation", *Moody's KMV, New York, Technical Report #020305, Revised: 06/13/02*

Süppel, R., 2003. "Comparing Economic Dynamics in the EU and, Accession Countries", European Central Bank, *Working paper series*

Thomas, L., Edelman, D., and Crook, J., 2002. "Credit scoring and Its Applications", *Siam*

Vandell, K.D., 1978. "Default Risk Under alternative Mortgage Instruments", *Journal of Finance*, Vol.33, No.5, 1279-1296

Vandell, K.D., and Thibodeau, T., 1985. "Estimation of Mortgage Defaults Using Disaggregate Loan History Data", *AREUEA Journal,* Vol.13, No.3

Vandell, K.D., 1993. "Handing Over the Keys: A perspective on Mortgage Default Research", *Journal of the American real Estate and Urban Economics association,* Vol.21, No.3, 211-246

Winkler, R.L., 1994. "Evaluating Probabilities: Asymmetric Scoring Rules", *Management Science,* Vol. 40, No. 11, 1395-1405

Yang, T.T., Buist, H., and Megbolugbe, I.F., 1998. "An Analysis of the *Ex Ante* Probabilities of Mortgage Prepayment and Default", *Real Estate Economics,* V26, 4, 651-676

Yang, Z.R., Platt, M.B., and Platt, H.D. 1999. "Probabilistic Neural Networks in Bankruptcy Prediction", *Journal of Business Research,* February, 67-74.

Appendix A

Table 35 contains descriptive statistics for the variables of the whole sample (**8274** observations) used in the back testing and the new model development.

Table 35

Variables	N	Mean	Std. Dev	Minimum	Maximum
Score_house_rights1_2	8274	24,23	9,82	0	30
SC_age_old	8274	-0,92	9,89	-45	10
Score_marit_status1_2	8274	10,82	12,46	-15	20
SC_time_at_occupation_old	8274	0,87	13,23	-20	20
SC_salary_drawn_old	8274	8,86	15,22	0	35
SC_type_of_contract_old	8274	26,71	9,38	0	30
SC_level_of_education_old	8274	4,06	6,10	-15	10
SC_profession_old	8274	7,01	15,49	-25	30
SC_kind_of_employment_old	8274	-1,11	3,43	-30	0
Score_time_at_address1	8274	1,42	11,18	-15	10
Demoscorefinal_old	8274	881,93	43,71	725	995
SC_Destination_MG	2503	16,43	11,20	0	25
SC_Own_contrib_MG	2503	9,51	4,48	0	20
SC_kind_of_property_MG	2503	18,35	11,05	0	25
SC_currency_MG	2503	-2,98	8,98	-30	0
SC_tenor_MG	2503	0,38	18,68	-20	35
SC_internal_loan_MG	7	7,14	12,20	0	25
MGscore	2503	27,16	24,07	-70	90
Demo_MGscore_old	2503	902,16	50,83	735	1085
SC_internal_loan_CL	40	0,63	3,95	0	25
SC_number_of_instalment_CS	5771	-9,69	13,56	-20	15
Clscore	5771	-29,69	13,57	-40	0
Demo_CLscore_old	5771	855,25	46,12	695	990
Couple_debt_ratio1_2	8274	77,97	607,76	0	9999
Max_accept_ratio_1_2	8274	49,06	13,60	30	85
Ability_to_repay1_2flag	8274	0,88	0,33	0	1
Net_monthly_income1	8274	2045,16	124563,42	0	11330785
Net_monthly_income2	5138	314,35	662,01	0	30371
Couple_modif_month_income1_2	8274	2233,64	124562,79	0	11330785
New_monthly_liabilities1	8274	202,93	611,33	21	43733
Total_existing_liabilities1	8274	15,01	168,02	-375	8489
Total_existing_liabilities2	8274	4,12	32,42	-375	1566
Couple_cost_of_living1_2	8274	335,25	357,71	100	2340
Number_exc	8274	0,12	0,40	0	4

The below table gives the distribution of accepted and rejected applicants by decision level (Branch Manager is 1st level, Regional Manager is 2nd level, HQ is Headquarters) and product type.

Table 36

Final decision Total Loans	Decision 1st level			Decision 2nd level			HQ			Total
	A	R	Distribution in %	A	R	Distribution in %	A	R	Distribution in %	
Accepted (A)	5545		72,81%	611		8,02%	1460		19,17%	**7616**
Rejected (R)		292	75,45%		19	4,91%		76	19,64%	**387**
Total	**5545**	**292**		**611**	**19**		**1460**	**76**		**8003**
Final decision CL	**Decision 1st level**			**Decision 2nd level**			**HQ**			**Total**
	A	**R**	**Distribution in %**	**A**	**R**	**Distribution in %**	**A**	**R**	**Distribution in %**	
Accepted (A)	4191		78,70%	433		8,13%	701		13,16%	**5325**
Rejected (R)		219	82,33%		16	6,02%		31	11,65%	**266**
Total	**4191**	**219**		**433**	**16**		**701**	**31**		**5591**
Final decision MG	**Decision 1st level**			**Decision 2nd level**			**HQ**			**Total**
	A	**R**	**Distribution in %**	**A**	**R**	**Distribution in %**	**A**	**R**	**Distribution in %**	
Accepted (A)	1354		59,10%	178		7,77%	759		33,13%	**2291**
Rejected (R)		73	60,33%		3	2,48%		45	37,19%	**121**
Total	**1354**	**73**		**178**	**3**		**759**	**45**		**2412**

The below table gives the distribution of the applicants by total score and ability to repay for the whole sample and for each product.

Table 37

Scoring total loans (MG+CL)	Number of observations	Distribution in %	Flag ability to repay			
			0	Distribution in %	1	Distribution in %
Black	7	0,08%	0	0,00%	7	100,00%
Red	1391	16,18%	210	15,10%	1181	84,90%
Yellow	6241	75,43%	755	12,10%	5486	87,90%
Green	635	7,67%	68	10,71%	567	89,29%
Total	**8274**		**1033**	**12.48%**	**7241**	**87.52%**
Scoring CL	**Number of observations**	**Distribution in %**	**Flag ability to repay**			
			0	**Distribution in %**	**1**	**Distribution in %**
Black	7	0,12%	0	0,00%	7	100,00%
Red	1268	21,97%	185	14,59%	1083	85,41%
Yellow	4349	75,36%	423	9,73%	3926	90,27%
Green	147	2,55%	10	6,80%	137	93,20%
Total	**5771**		**618**	**10,71%**	**5153**	**89,29%**
Scoring MG	**Number of observations**	**Distribution in %**	**Flag ability to repay**			
			0	**Distribution in %**	**1**	**Distribution in %**
Black	0	0,00%	0	-	0	-
Red	123	4,91%	25	20,33%	98	79,67%
Yellow	1892	75,59%	332	17,55%	1560	82,45%
Green	488	19,15%	58	11,89%	430	88,11%
Total	**2503**		**415**	**16,58%**	**2088**	**83,42%**

In the end, Table 38 gives the impact of exceptions on the final decision (rating).

Table 38

Total sample (MG+CL): Final decision	Total sample Flag Exception	Total sample Number of observations	Distribution in %	Total sample Rejection Rate
Black	No	869	10,50%	24,4%
	Yes	243	2,94%	26,3%
White	No	6578	79,50%	1,4%
	Yes	584	7,06%	9,1%
Total		**8274**	**100%**	**5,10%**
Final decision (CL)	**Flag Exception (CL)**	**Number of observations (CL)**	**Distribution in %**	**Rejection Rate**
Black	No	617	10,69%	28,03%
	Yes	80	1,39%	40,00%
White	No	4885	84,65%	1,20%
	Yes	189	3,27%	11,64%
Total		**5771**	**100%**	**4,99%**
Final decision (MG)	**Flag Exception (MG)**	**Number of observations (MG)**	**Distribution in %**	**Rejection Rate**
Black	No	252	10,07%	15,48%
	Yes	163	6,51%	19,63%
White	No	1693	67,64%	1,89%
	Yes	395	15,78%	7,85%
Total		**2503**	**100%**	**5,35%**

Tables 39 and 40 indicate that bad rate1 for CL borrowers without exceptions (5.29%) higher than the corresponding one for MG loans (3.57%), even if the numbers show that MG loans applications are more often accompanied by exceptions (26.5%) than CL ones (5.1%).

Table 39

CL	Decision 1st level	Decision 2nd level	HQ	Bad Rate (HQ)	Bad Rate1 (HQ)
without HQ rule	876	170	208	3,43%	5,29%
with 1 HQ rule	0	0	61	3,39%	6,56%
with 2 HQ rules	0	0	6	16,67%	16,67%
with 3 HQ rules	0	0	0	0,00%	0,00%
with 4 HQ rules	0	0	0	0,00%	0,00%

Table 40

MG	Decision 1st level	Decision 2nd level	HQ	Bad Rate (HQ)	Bad Rate1 (HQ)
without HQ rules	123	18	28	3,57%	3,57%
with 1 HQ rule	0	0	49	4,08%	4,08%
with 2 HQ rules	0	0	11	9,09%	9,09%
with 3 HQ rules	0	0	0	0,00%	0,00%
with 4 HQ rules	0	0	1	0,00%	0,00%

Tables 41 and 42 give the rejection rate by rating and policy rules for the types of product (CL and MG correspondingly).

Table 41

Final outcome CL	Rejection Rate	4,99%
	Frequency	**Rejection Rate**
Rejected	**10,69%**	**28,04%**
Rejected for Rating (Black)	10,69%	28,04%
HQ	**4,66%**	**20,07%**
Grey for HQ Rules	4,66%	20,07%
Accepted/Proposable	**84,65%**	**1,25%**
Black Rating - HQ Rules	1,39%	40,00%

Table 42

Final outcome MG	Rejection Rate	5,35%
	Frequency	**Rejection Rate**
Rejected	**10,07%**	**15,48%**
Rejected for Rating (Black)	10,07%	15,48%
HQ	**22,29%**	**11,29%**
Grey for HQ Rules	22,29%	11,29%
Accepted/Proposable	**67,64%**	**1,89%**
Black Rating - HQ Rules	6,51%	19,63%

Tables 43 and 44 indicate that CL borrowers without exceptions have higher bad rate1 (8.88%) than MG loans without exceptions (4.81%). In general, the numbers show that MG loans applications are more often accompanied by exceptions (18.70%) than CL ones (3.64%).

Table 43

CL	Distribution in %	Bad rate1
without exception	96,37%	8,88%
with 1 exception	3,41%	4,44%
with 2 exceptions	0,23%	33,33%

Table 44

MG	Distribution in %	Bad rate1
without exception	81,30%	4,81%
with 1 exception	11,74%	7,41%
with 2 exceptions	3,48%	0,00%
with 3 exceptions	3,04%	14,29%
with 4 exceptions	0,00%	0,00%
with 5 exceptions	0,43%	0,00%

Appendix B

I summarize here advantages and disadvantages of some of the tentative regressions that I performed. All of the models that I tried had Gini coefficients between 46.10% and 46.80% and undertook either variables categorization or variables coefficients changes.

Thus, the set of models where I used "kind of company 2" has the disadvantage of assigning higher weight to "kind of company 2" rather than to "profession", but the former suffers 24% of missing variables. As I previously explained, because of missing values, this variable should not have significant weight in the final model. All such models had similar parameters and output values, but as I can achieve better categorization for "kind of company 2", I rejected all of them.

In another set of models I tried to assign weights to the variables according to the pre-defined order in paragraph 5.2.3. Therefore, I ran regressions where I forced only "salary drawn in the bank" and "age", but nothing changed as these models kept on giving too much weight to "salary drawn in the bank". Even if the Gini coefficients of all the models varied between 46.60 - 46.80%, I rejected all of these models as "salary drawn in the bank" has the same weight as "kind of company" (and lower one than "profession") even if its discriminatory power is much better.

Therefore, I tried to assign higher weight to "time at present occupation" and lower one to "salary drawn in the bank". In this case all the variables were ordered in the preferred manner, but the Gini coefficient of the models went down to 46.20.

In the end, I ran also regressions where the coefficients for "age" and "time at present occupation" were pre-defined, but the Gini coefficient went down to 46.40% and "salary drawn in the bank" got the same weight as "time at present occupation", which again is not desirable.

Therefore, I select as final the model whose Gini coefficient is significant and the variables' weights follow their discriminatory power, single Gini coefficients and the number of usable observations.