CREDIT RISK ASSESSMENT OF CORPORATE SECTOR IN CROATIA

Lana IVIČIĆ*
Croatian National Bank, Zagreb

Article**
UDC 336.71
JEL G12

Saša CEROVAC*
Croatian National Bank, Zagreb

Abstract

The main goal of this paper is modeling credit risk of non-financial businesses entities by assessing the rating migration probabilities and predicting the probability of default over one year horizon on the basis of corporate financial accounts. Our research provides a number of new important insights. Ratings migration matrices are symmetrical in every observed period, which implies that default state is not final terminal state. We find a high degree of rating stability, with the exception of some volatility generated by firms in the middle of the ratings scale. In the period of lower economic growth probabilities of transition between different risks categories are lower than in the period of higher economic growth. Probabilities of default are relatively stable across enterprises operating in different economic activities. After considering a wide range of potential predictors of default, multivariate logistic regression results reveal that the most important are the ratio of shareholders' equity to total assets and the ratio of EBIT to total liabilities, both negatively related to the probability of default. In addition, higher liquidity, profitability and sales as well as construction and real estate sector affiliation all decrease the companies' probability of default in the following year. The model correctly classifies relatively reasonable percentage of companies in the sample (74% of all the companies, 71% of defaulted and 75% of non-defaulted companies) when the threshold is set in such a way to maximize the sum of correctly predicted proportions for both defaulted and non-defaulted companies.

Keywords: credit risk, corporate default, migration matrices, logit

---

* Authors thanks two anonymus referees for valuable comments and suggestions.

** *Received*: June 1, 2009
Accepted: November 3, 2009

1 Introduction

Modeling credit risk is an active area of research of both market participants and regulators, with the later being responsible for the stability of the overall financial system, as well as the task of supervision of individual credit institutions. Therefore, central bankers and other financial regulators have special interest to model risks for the banking sector (Richter, 2007). The purpose of this paper is to model the credit risk of Croatian non-financial businesses entities thus enabling the assessment of the individual banks' risk profile regarding their exposures to the corporate sector and fluctuations in the aggregate risk.

In short, the paper explores the possibilities to forecast exposures to credit risk, at the same time identifying distribution of risk in the banking sector and predicting the probability of the change in capital adequacy. It should be noted that this is first attempt to use CNB's prudential database on credit exposures and corporate balance-sheets (FINA's database) to get an estimate of probability of default at banking sector level for each significant corporate client of the Croatian banks. It provides an overall assessment of the evolution of corporate credit risk determination methods in Croatia and it enables any interested party to use financial micro indicators of any business entity in order to estimate the probability of default or produce a sort of "rating" for any corporation in Croatia. A special emphasis is placed on observance and prediction of changes in banks' corporate portfolios as aggregate measure of credit risk, which is particularly important during the crises period.

This aim is achieved in two ways. The first approach is based on the assessment of the rating migration probabilities. The second approach uses corporate balance sheet data in order to predict the probability of being in default for each company. These forecasting tools could be used to predict the probability of a default and could ultimately yield more accurate assessment of potential losses in the banking sector if these risks materialize.

First we use migration matrix models to gain insight in the stability of ratings and to forecast their transitions, revealing the structure of forecasted probability of default for the non-financial corporate sector. Further on, using additional information from the balance sheets of each debtor enables us to model the risk of default by using multivariate logit regression. Statistical model used here in general permits simulation of various shocks and extraction of important information on the probability of default and loss given default, tracing them back to the individual bank or group of banks. Such an approach is widely used in order to enhance the stress-testing systems (Andersen et al., 2008), used by central banks and other market participants (Figure 1 in Appendix).

The study is organized as follows. Section 2 gives a brief overview of the related previous studies. Section 3 describes the datasets and the necessary pre-treatment of the data. Section 4 explains the main definitions and general concepts of credit ratings and default, while Section 5 presents rating migrations matrices. Modeling credit default, methodology applied, univariate analysis and results for multivariate model are given in Section 6. Finally, Section 7 summarizes the main findings and concludes the study with proposed directions for further research.

2 Literature review

Since 1960's a substantial volume of corporate failure literature has been published. Pioneering papers were written by Beaver (1966), who found that some indicators could discriminate between failed and non-failed firms with univariate analysis, and Altman (1968) who proposed the use of linear multiple discriminant analysis (MDA). Later studies include many extensions to this early methodology, but they were often criticized because restrictive assumptions[1] in multiple discriminant analysis models' were frequently violated.

To avoid some of the problems of the MDA approach, Ohlson (1980) was the first to employ logistic regression to predict company failure using published accounts. He found a negative correlation between the probability of failure and the size, profitability and liquidity of the company, and positive correlation between the probability of failure and company's indebtedness.

Ever since, logistic regression has been extensively used for the development of non-publicly traded companies failure models and a wide range of explanatory variables was tested. Extensions to Ohlson's study include, among others, Platt and Platt (1990) that developed industry specific models and found that probability of failure depends on the sector the company is operating in; similar conclusions were found in Bernhardsen (2001) and Lykke, Pedersen and Vinther (2004).

Bernhardsen (2001) introduced specification of the logit model which allows for flexible rates of compensation as opposed to the common specification applied for the bankruptcy prediction model where the rates at which two variables can substitute another (holding predicted risk unchanged) are constant. The list of explanatory variables contained liquidity, profitability, solidity and indebtedness ratios, age and size of the firm and some industry-specific indicators. In addition to financial ratios, Lykke, Pedersen and Vinther (2004) found qualitative non-financial variables to be significant in explaining the probability of failure in Danish corporate sector. Among others, critical comments from auditors and a capital base reduction increase the probability of failure.

Charitou, Neophytou and Charalambous (2004) examined the incremental information content of operating cash flows in predicting financial distress. They employed both logit methodology and neural networks to develop a prediction model for UK industrial firms and found that the two models could be used for bankruptcy prediction. Their empirical results indicate that operating cash flows (along with two other financial ratios) possess the discriminatory power in predicting company failure.

Work by Jacobson et al. (2008) provides empirical evidence that adding macroeconomic information in simple logistic model with firm-specific factors contributes to explaining the likelihood of defaults. This result suggests that macroeconomic factors shift the default risk distribution over time and thereby are the most important source of the level of default risk.

[1] These assumptions are (a) independent variables are multivariate normal and (b) covariance matrices of two subsamples (failed and non-failed) are equivalent.

Recent years much attention is given to the choice of methodology. Methods like recursive partitioning, neural networks and genetic programming are commonly applied on the bankruptcy prediction problem. However, logit models are still frequently used and central banks from euro area commonly use such models in order to determine eligible collaterals for refinancing operations. The in-house credit assessment system of the central bank of Austria (OeNB)[2] comprises of 4 logit models (1 base model and 3 industry-specific models) as well as of qualitative assessment of credit risk in Austrian companies. Explanatory variables include accounting ratios and some general firm-specific information. Falcon (2007) presents methodology based on logit models used by Banco de España for an in-house credit assessment of non-financial companies. After testing numerous financial ratios, solvency ratios were found as the most powerful factors for default prediction. Some general conclusions are that non-linear logits get significantly better results in terms of predictive power than linear ones, and macroeconomic environment plays a significant role in default prediction, with GDP growth as the best performing variable.

3 Data

For the purpose of our research we use two primary datasets. The relevant information on bank exposures and credit ratings (which is used to construct the default statistics on a quarterly frequency) is extracted from the Croatian National Bank's (CNB's) prudential database which identifies bank's exposures towards significant debtors (for more details see CNB, 2003). This database serves for the analysis of ratings and default risk on the basis of migration matrices. Annual database of corporate financial accounts, provided by Financial Agency (FINA), is used to extract additional information employed in the regression analysis of default risk. The submission of annual report is a legal obligation and FINA states that the reporting enterprises account for the vast majority of operating enterprises with insignificant fraction of enterprises that overlook this obligation. The number of enterprises in the database supports such claims as more than 60 thousand enterprises submitted their reports in 2006 and 2007 (years relevant for our purposes), although analysis revealed some flaws in the database, including both missing enterprises and omissions in the data.

Several steps were done to check the data, clean any errors and omissions that could be identified and adjust the sample before the application of analytical framework.

First, full coverage of the banks in the database of banking exposures as well as detailed information on risk classifications starts from June 2006, thus limiting the analytical timeframe. Second, bank's exposures towards non-residents, non-corporates, non-market oriented firms (public administration and defense) and unidentified debtors as well as aggregated exposures for groups of debtors (other debtors and portfolio of small loans) were removed from the population. Third, exposures towards small identified debtors – those whose amount did not exceed 100,000 kunas (approximately 15,000 euros) – are also omitted in order to reduce the volatility that could arise from small exposures towards debtors with marginal share in total liabilities of the corporate sector. Further on, sample

[2] Winkler (2008)

was stabilized by removal of enterprises entering and/or exiting the database during any observed period (type of left and right censoring). This restriction was relaxed in analysis of quarterly migration matrix as only enterprises not present in the database during two consecutive quarters were removed.[3] Finally, all banking exposures towards each debtor are consolidated according to debtor's ID number and multiple entries are avoided by prioritizing them according to supervisory actions (identified by report abbreviations).

After these adjustments, not including the "censoring", the annual average of 10.670 debtors remains (for the period of eleven quarters). However, the assessment of the rating migration process presented here, after taking into account the fourth step as well, imposes an additional restriction in the sense that one should also control for the possible change of the economic activity[4]. So, firms that migrate cross-sectoraly are omitted. All these actions reduce the number of firms in our quarterly sample adjusted for migration matrices to about 3/4.

This additional procedure was not required for the purpose of constructing a regression default model based on annual frequency. However, the need to combine two different databases (CNB's and FINA's) for this purpose also reduced the number of business entities in the sample, for the most part reflecting the shortcomings of FINA's database. In addition, the sample was stabilized by removing enterprises that are not present throughout the whole year in the CNB's database. Therefore, the final data set is reduced to 7,719 firms present during 2007 and 2008, providing a non-balanced panel consisting of 12,462 observations of binary dependent variable, enterprises for which we have information whether or not they have been in a default during the specific year. This sample accounts for more than 75% of bank's exposures towards market-based corporates.

The dataset of explanatory variables covers a wide set of variables (84 potentially relevant financial indicators) that proved to be successful in predicting default in previous studies and can be grouped in a following manner: liquidity ratios (16), solvency indicators (23), activity ratios (12), efficiency ratios (7), profitability ratios (27) and investment indicators (1) (see Appendix).

Corporate annual financial reports, providing information for the end of the previous year, are used to predict credit risk in the forthcoming period. As accounting data are usually published with a several months time lag, accounting data typically become available during the same year for which the credit risk is assessed. However, even such delayed data are useful since they provide basis for forecasts up until the year's end. Also, as longer time series become available, forecasting horizon should expand.

Outliers were corrected in order to prevent the possible bias. Once identified, extreme values of financial ratios were not removed to avoid the reduction in the sample size (having in mind that these extreme observations most often belong to the troubled companies, removing them would further decrease the number of defaults in the sample). Instead, out-

[3] Unfortunately, it was not possible to check the reason of each specific omission from the database, so it was not possible to discriminate between elimination of the exposure (due to repayment or write-down) or non-reporting.

[4] We grouped the firms into three sectors according to the National Classification of Economic Activities (NACE): industry and agriculture (NACE categories A, B, C, D and E), construction and real estate (NACE categories F and K) and non-financial services (NACE categories G, H and I).

liers were winsorized (typically by the 1st and the 99th percentile of the variable in question, but some variables were winsorized asymmetrically - meaning replacing only the values below the 0.5th or 1st percentile, or only the values above the 99th or 99.5th percentile).[5] This procedure allows keeping the very low and very high values of the affected variables which is useful input to the model, without losing the rest of the information for that observation.

4 Credit rating and default

The CNB's database provides only information on the risk classification of individual exposures (placements and off-balance sheet liabilities), i.e. there is no risk classification of debtors themselves (for more details see CNB, 2009). Since the classification of placements for any debtor by an individual credit institution can be dispersed in several risk categories, which is the case even more often when the debt is summed across the banks for each debtor, the task of classifying a debtor is not a straightforward one.

All the placements are classified into three broad risk categories, depending on the possibility of collection, i.e. on the expected future cash flows:

A – placements for which no evidence of impairment is identified on individual basis (standard);

B – placements for which evidence of partial impairment is identified, i.e. partly recoverable placements (substandard);

C – placements for which evidence of impairment is identified, equal to their carrying amount, i.e. fully irrecoverable placements (delinquent).

Debtors' timeliness in meeting their obligations towards a credit institution is important criteria integrated in the above classification scheme and implies a downgrade from A to B if debtor has overdue liabilities for more than 90 to 180 days, and from B to C if debtor has overdue liabilities for more than 365 days. Also, a sub-category of A that contains overdue liabilities for more than 90 days but secured by eligible instruments of collateral, is reported as A90d. Therefore, we were able to identify the amount of liabilities that fall into the lowest risk category, denoted AX (equal to the differences between A and A90d).

The procedure for classifying debtors into distinct risk categories (R) applied in this paper is based on solving a simple optimization problem derived from the risk distribution of total exposures: we search for the threshold value (T) of a share of successive cumulative amount of exposure (S) that falls into specific range of risk categories (r that go from C to AX) for each debtor so that the amount of exposures classified AX within the group of debtors that would be classified AX and the amount of exposures classified non-AX within all other debtors is simultaneously maximized. More formally,

[5] Winsorization is substitution of detected outliers by non-extreme values (quite common procedure in literature using financial ratios). As a robustness check we constructed series where all extreme values were deleted; the estimation results did not significantly differ from those obtained using the winsorized series.

$$Sr = \sum_{r=C}^{succ.rating(m)} exposure_r \,/\, total\ exposure \geq T \Rightarrow r_m = R_{debtor} \quad (1)$$

We find the optimal threshold to be $T = 0.5$ (graf. 2). The results show that distribution of aggregate exposures that are "correctly" assigned to firms (matching risk category of exposures and debtor's rating) is not very sensitive to the variations of the threshold in close proximity of T.

Defining the event of a default represents a final step that enables us later to identify one of the key parameters in the credit risk assessment - the probability of default. Following the provisions of the *Basel Committee on Banking Supervision* (Basel II Accord), we adopt the definition of default as (*Official Journal of the European Union*, I.177 p. 113):

"A 'default' shall be considered to have occurred with regard to a particular obligor when either or both of the two following events has taken place:

- the credit institution considers that the obligor is unlikely to pay its credit obligations to the credit institution, the parent undertaking or any of its subsidiaries in full, without recourse by the credit institution to actions such as realizing security (if held);
- the obligor is past due more than 90 days on any material credit obligation to the credit institution, the parent undertaking or any of its subsidiaries."

This means that in our case default[6] occurred for debtors rated non-AX. At the first glance, just as one may expect, it is noticeable that in every period under observation the majority of debtors have not defaulted. The ratings distribution of other debtors reveals quite stable rating structure of defaults that occurred from June 2006 to December 2008: close to 8% fall into risk category B, 5% fall into risk category C, and those classified as A90d account for the smallest proportion, merely 2%. The sum of these fractions yields the average default rate for the non-financial corporate sector (15%).

5 Rating migrations and the probability of default

5.1 Migration matrix

A notable feature of the risk evolving process is the formation of rating migration matrices that generate information on probabilities of transition from rating i to j. Following closely Fuertes and Kalotychou (2008: 5-6), let S denote the transition space and $i = 1, 2, ..., k$ risk categories, so that $P(s,t)$ denote the $k \times k$ transition probability matrix generated by continuous Markov chain[7] z. The rating transition in the period between s and t is:

[6] It is important here to distinguish default (a situation when debtor is not capable to fully meet its obligations to a credit institution on the basis of principal, interest, commissions, and other, in the contractual amounts and within the contractual time limits) from delinquency, insolvency, bankruptcy and liquidation.

[7] The basic assumption of the migration matrix estimator is that ratings are cross-time and cross-sectional independent (conditions of Markov property and homogeneity).

$$p_{ij}(s,t) = P(z_t = j | z_s = i), s < t \tag{2}$$

For every transition horizon Δt we estimate the migration matrix $P\Delta_t$ in general form:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{k1} & & \cdots & p_{kk} \end{bmatrix} \tag{3}$$

where $p_{ij} \geq 0 \quad \forall i, j, \quad \sum_{j=1}^{k} p_{ij} = 1 \quad \forall i$.

If $N_i(t)$ is the number of firms i rated at the beginning of the period t, and $N_{ij}(t + 1)$ is the number of firms that migrated from i to j, by the end of that period, then the migration frequency is the ratio $N_{ij}(t + 1) / N_i(t)$. We utilize discrete multinomial estimator:

$$\hat{p}_{ij} = \frac{1}{T} \sum_t \frac{N_{ij}(t+1)}{N_i(l)} \tag{4}$$

which is widely used and in fact represents a special case of the maximum likelihood estimator (MLE) when the number of firms is constant over time.

5.2 Empirical migration frequencies and forecasts

By deriving 1-Year and 1-Quarter migration matrices (Table 1 and Table 2) we generally find high degree of rating stability in the (non-financial) corporate sector, with the exception of some volatility generated by A90d rated firms. To summarize:

- It is clear that the highest and the lowest rated companies (AX and C, respectively) have the lowest migration frequencies, while the volatility of ratings increases in the mid-section of the rating structure: for A90d rated companies and, to lesser extent, for B rated ones. This pattern can be explained by the fact that business environment and financial conditions for the AX and C rated companies are not very likely to change significantly in the short run (in a sense of change that would affect their rating). These risks, however, are more prominent for A90d and B rated firms and so the probability for them to be upgraded or downgraded is naturally higher - in particular so for A90d rated companies whose rating depends to large extent on their collateral.
- The A90d and B rated firms also exhibit asymmetrical migration pattern: the probabilities of their upgrades are higher than probabilities of their downgrades.
- The monotonicity of rating migrations - gradual change in migration frequencies across terminal ratings for each initial rating - is not observed at all in 1-Year migration matrix, but it is partially observed at quarterly frequency, at least for A rated firms.

- The 1-Quarter matrix, that has a more informative content, alters the migration frequencies, so that parameter values on matrix diagonal increase, i.e. it reveals even greater rating's stability. In contrast to 1-Year matrix, the relative degree of migration frequencies for each *i* rated firm at the beginning of observed quarter is altered only for A90d rated firms: $\hat{p}_{A90d \quad A90d} > \hat{p}_{A90d \quad B}$.
- All matrices are symmetrical, i.e. there is no absorbing state *i* that satisfies $p_{ik} = 1, \ \sum_{j=1}^{k} p_{ij} = 1$. The important implication of this is that default state (rating A90d, B or C) is not final terminal state, i.e. every state is reversible so that dependent variable in our logit regression (Section 6) is bi-directionally formed. For example, 1-Year migration matrix shows that the probability of default (*PD*) is equal to 5% and the probability of reversing from a default state (*PR*)[8] is close to 12%. Therefore, the number of firms in default state increases by ΔN_D, depending on both of these probabilities[9]:

$$\Delta N_D = PD * N_{AX} - PR * N_{non-AX} \tag{5}$$

- The structure of PDs (based on annual frequencies) interestingly shows dominant share of B rated firms, A90d being the mid and C its smallest proportion. On the other hand, while A90d still accounts for the most of defaulted firms that remained in this state over one year horizon, the rest of them are almost entirely the lowest rated companies.
- The order of ratings according to their relative shares and migration frequencies behind the PRs are A90d, B and C respectively.
- The ratings and PDs (PRs) exhibit positive (negative) correlation (Figure 3).
- The conditional quarterly matrices give us further insight in the rating migration process. We conditioned the migration frequencies on economic cycle and economic activity.[10]
- It is obvious that the retardation phase increases stability in credit ratings while the PRs are systematically reduced. Looking at the historical time series of default rates (Figure 4), it is also clear that over the last four quarters they tend to increase (across all the sectors), but nothing indicates that they are significantly affected (in relation to their historical values) by recent financial crises, which is hard to expect in the near future.
- The PDs do not appear to be sensitive to economic activity. Non-financial services show the most similarity in relation to the properties of the unconditional ma-

[8] Formally, $PR_i = 1 - D_i, \ \ \forall i \neq AX$; where *D* is probability of remaining in a default state (A90d, B or C).

[9] The PDs "adjusted" for the number of firms reversing from a default state are on average three times smaller (based on quarterly frequencies).

[10] There were no classic economic cycles (contractions of economic activity) during the observed period so we use the growth cycle concept (acceleration phase: 3q2006-3q2007; retardation phase: 4q2007-4q2008; based on the short-run fluctuations of real gross domestic product). The conditions are generally restricted to small number of variations (here two cyclical phases and three categories of economic activity) to avoid the reduction of sub-sample's size. Possible effects of conditioning matrices are portrayed by Figure 5 in Appendix.

trix, but PRs are moderately lower in Industry and higher in Construction where we also find evidence of relatively highest rating's instability.

The insufficient number of observations does not allow us to build an Ordered Dependent Variable Model (Multinomial Logit) that would enable us to generate forecasts of default probabilities conditional on the phase of cycle and economic activity (see e.g. Nickell, Perraudin and Varotto, 2001), so at this point we only use simple unconditional migration ratings model approach. For the $n\Delta t$ horizon (Δt being one quarter or one year depending on the matrix) the migration probabilities are forecasted as $P^{n}_{\Delta t}$. The one-year migrations forecast based on 1-Year matrix ($n = 1$) and 1-Quarter matrix ($n = 4$) is reported in Table 3. The probability of default for 2009 is increased (the PD comes close to 6%) as well as the probability of reversal, most intensively for the A90d rated firms and least for the C rated ones.

6 Modeling credit default

6.1 Multivariate logit regression

The probability of being in default is estimated in this paper using the maximum likelihood method within the framework of logistic regression.[11]

Let the binary observation of a default rate of a company (continuous variable) be:

$$Yi,t = 1 \text{ if the firm } i \text{ is in a default and state}^{12} \text{ 0 oherwise} \tag{6}$$

Binary default variable $y_{i,t}$ is explained by a set of factors X. Therefore, the probability that a company defaults is:

$$P(y_{i,t} = 1 \mid X_i) = F[X_i, \beta] \tag{7}$$

where X_i is the set of K explanatory variables for company i and β is the set of parameters. Using the logit function, the expected probability that a company will default can be written as:

$$F[X_i, \beta] = 1/\left(1 + e^{-w}\right); \quad w = \beta_0 + \beta_{1i,t}x_{1i,t} + ... + \beta_{ki,t}x_{ki,t} \tag{8}$$

The logit model guarantees that F$[X_i,\beta]$ is constrained to interval [0.1].

[11] In the context of credit risk modeling, logit models have several advantages: they do not assume multivariate normality; they are transparent when evaluating the importance of each variable; they allow obtaining a direct estimation of *PD*; they show good predictive results when compared to other techniques and they work well with qualitative explanatory variables (Falcon, 2007).

[12] We differentiate between the probability of being in the state of default (period characterized by debtor's failure to fully meet its obligations to a credit institution) and the probability of default (the event of transition into the state of default).

6.2 The selection of explanatory variables

To identify any possible differences between defaulted and non-defaulted companies, several main descriptive statistics (mean, median, standard deviation, minimum and maximum) were first calculated for each explanatory variable. Several steps followed, aiming to examine their statistical relationship with corporate default and explanatory power.

As a first step all indicators were statistically tested for the equality of the mean with respect to the dependent variable. Table 5 in Appendix shows that defaulting firms in general are less liquid, more indebted, have lower turnover and lower profitability indicators. Performed tests allowed identification of the most promising explanatory variables for the inclusion in the model. Identified variables entered into the next phase of testing, in which type and sign of relationship between the selected variables and probability of default were subject to graphical analysis. For each variable[13] a scatter graph was constructed showing an average default rate for each percentile range of the explanatory variable (see Figure 7 in Appendix). These figures indicated whether there is a meaningful relationship between default and each variable, as well as the monotonicity, shape and sign of this relationship. Further on, it was checked if the sign of relationship for each variable was as expected, and variables showing unexpected sign of relationship as well as variables that lacked any correlation with default risk or those displaying clearly non-monotonic relation were also removed from the set of candidates. Sign of relationship for most explanatory variables was in accordance with prior expectations - for example, default risk seems to decrease as liquidity indicators and economic activity improve; on the other hand, default risk is higher when financial leverage rises.

To identify variables with the highest explanatory power, univariate logit model was estimated for each of the preferred variables, and the corresponding ROC curves[14] were derived. Choice of statistically significant variables with satisfactory levels of Pseudo R^2 and area under the ROC curve led to the selection of 28 candidate variables: 4 liquidity ratios, 10 financial autonomy/financial leverage ratios, 3 economic activity indicators, 1 efficiency ratio, 8 profitability indicators and 2 indicators of firm's size (Table 6 in the Appendix).

Among them, profitability indicators seem to have highest univariate classification ability, with areas under the ROC curve (AUC) ranging from 0.69 to 0.75 (the best one is the ratio of total sales to total liabilities with AUC of 0.75). Regarding liquidity indicators, the best performing is the ratio of cash to total assets. In addition, financial autonomy appears to be a good individual predictor of default too, as ratios of equity capital to total assets and to total liabilities reach the ROC area values above 0.70.

---

[13] Some variables exhibiting high dissipation were transformed in order to be closer to normal distribution; transformation applied was $(\text{sign}(y)*\text{abs}(y)^\wedge\lambda+1)/\lambda$ with $\lambda=0.1$, 0.15 or 0.2.

[14] ROC (Receiver Operating Characteristics) curve is a graphic representation of the relationship between the sensibility and the (1-specificity) of a model for all possible thresholds, where sensibility represents the probability of correctly classifying an individual whose observed situation is "default" and specificity represents the probability of correctly classifying an individual whose observed situation is "no default". Area under the ROC curve is a measure of the model's prediction power: higher area corresponds to higher accuracy of prediction.

6.3 Estimating the multivariate logistic regression model

Following the univariate analysis and taking into account the correlations among variables[15], numerous models including different groups of variables were tested. Problems arising from multi-collinearity as well as additional losses of observations experienced when increasing the number of explanatory variables restricted too wide set of variables. Therefore, final multivariate model was chosen among best performing combinations of three, four, five and six accounting variables (Table 7) along with the dummy variable indicating if the company is in construction and real estate business which proved to be the only significant economic activity indicator.

The guiding principles in selection of the best model among five candidate models were: a) previously proposed and theoretically justified variables, b) statistical significance of estimated parameters, c) the fit of the model in terms of pseudo R^2 [16], d) area under the ROC curve as a measure of models' accuracy and e) overall correct classification ability of the model. It is important to note that all variables in candidate models exhibit the same sign and similar size of estimated coefficients in different specifications, indicating robustness to inclusion of additional variables or their removal from the estimated equation.

More parsimonious candidate models with three, four and five financial ratios were somewhat outperformed by the two models with six financial ratios, especially in terms of correct classification ability. Later two models differ in only one variable and yield very similar results in terms of pseudo R^2, area under the ROC curve and percentage of correctly classified defaulted and non-defaulted companies. However, number of minor differences outweighs our decision in favor of the model 6.1.[17]

The chosen model estimates probability of default based on the size of the firm (measured by total sales), construction and real estate dummy and 5 financial ratios: liquidity ratio (measured as cash to total assets), financial autonomy (shareholders' equity to total assets), activity indicator (accounts receivable turnover in days) and two profitability ratios (EBIT to total liabilities and sales and depreciation to total assets). Resulting equation is:

$$F[X_i, \beta] = \frac{1}{1 + e^{-(-0.17 - 0.28D_{i,t} - 0.63w\_1\_10_{i,t} - 1.96w\_2\_2_{i,t} + 0.09w\_3\_4_{i,t} - 0.14w\_5\_16_{i,t} - 0.37w\_5\_22_{i,t} - 0.01w\_7\_5_{i,t})}} \tag{9}$$

[15] Data from correlation matrix available upon request.

[16] Pseudo R^2 resembles the conventional R^2 measure of linear regression models. It can be interpreted as the degree to which the distribution of predicted probabilities of default for healthy companies does not overlap with the distribution of predicted probabilities of default for firms that defaulted.

[17] First, model 6.1 has slightly lower type I error which is more costly for banks than the type II error (on the basis of the threshold selected to maximize the average between the two proportions of correctly classified observations). Second, in model 6.4, there is somewhat higher correlation between two financial autonomy/leverage indicators (0.44) than between two profitability indicators in model 6.4 (0.39). Next, model 6.1 has greater number of observations than model 6.4, which is related to better availability of the data used. Finally, explanatory variable EBIT/total liabilities in 6.1 is more straightforward in interpretation and more common in related literature than (after tax profit + depreciation)/(debt/365) in model 6.4.

All variables in the estimated model are significant at 1 per cent level and coefficients associated with these variables have expected signs. The magnitudes of the estimated marginal effects[18] are in line with our expectations (see Table 8 in Appendix): the largest marginal effects are associated with ratio of own funding (shareholders' equity to total assets) and profitability indicators.

Higher share of equity funding reduces the probability of default. The marginal effect associated with shareholders' equity to total assets ratio implies that one standard deviation increase in this ratio, at the average level, lowers the probability of default in the following year by 3 percentage points, *ceteris paribus*. Probability of default decreases with profitability gains measured by EBIT to total liabilities and by ratio of sales + depreciation to total assets. A one standard deviation increase from the average in EBIT to total liabilities ratio reduces the probability of default by 3 percentage points, while the same increase in the sales + depreciation to total assets ratio results in probability of default lower by 1.1 percentage points.

The more liquid (measured by the cash to total assets ratio) a firm is, the less likely it is to default. A one standard deviation increase in the cash to total assets ratio from the average level reduces the probability of default in the following year by 2 percentage points, ceteris paribus. Lower activity, measured by the days needed for accounts payable turnover, leads to higher probability of default and one standard deviation increase above the average in number of days increases probability that a firm will default by 0.3 percentage points, ceteris paribus.

We also find that small firms (measured by total sales) are more likely to default than large ones after controlling for all other factors. However, coefficient associated with this variable is very low and 1 standard deviation increase in total sales from the average reduces firm's probability of default by only 0.04 percentage points, holding all other factors constant.

The coefficient on the construction and real estate dummy is negative and significantly different from zero. This implies that if a company is in the construction and real estate sector rather than in agriculture and manufacturing or non-financial services, it's probability of default is 0.9 percentage points lower, ceteris paribus. This is a plausible result given the intensive growth of the construction and real estate sector in 2006 and 2007.

Estimated distributions of predicted probabilities of default for non-defaulted and defaulted companies are given in Figure 8 in Appendix. We used Kolmogorov–Smirnov test (K-S test)[19] to test the null hypothesis that two samples are drawn from the same distribution. As expected, the null hypothesis was rejected at 1 per cent level. The model results in area under the ROC curve of 0.80 (Figure 9 in Appendix). The percentage of "false

[18] The marginal effects are evaluated at the sample means for each variable. For continuous variables the slope of the probability function is calculated to measure the change in the predicted probability for an infinitesimal change in that variable. The marginal effect on the predicted probability from a one-standard deviation change in that variable is then extrapolated out from this. For dummy variable the marginal effect is calculated as the change in the predicted probability of failure when the variable changes from zero to one, with all other variables evaluated at their sample means.

[19] The two-sample K-S test is one of the nonparametric methods for comparing two samples. The K-S test statistic measures a distance between the empirical distribution function of two samples.

negatives" (type I error[20]) and "false positives" (type II error) depends on the chosen default threshold. A decision on which specific value of probability of default to set as a benchmark was made by solving a simple optimization problem where we searched for the threshold value of a default rate that maximizes the sum of percentages of correctly classified defaulted and non-defaulted companies.

We find that benchmark default rate of 0.14 results in the highest sum of correctly classified percentages of non-defaulted and defaulted companies. This threshold is very close to the observed frequency of default in our sample (0.15 in 2007, 0.13 in 2008 and 0.14 in two-years time). Using this threshold we determine the number of eligible (estimated probability of default lower than the threshold) and non-eligible (estimated probability of default higher than the threshold) companies and the number of classification errors. Overall, the model correctly classifies 74.4% of the companies, or 74.9% of non-defaulted and 71.2% of defaulted companies.

Alternatively, a threshold value of default probability can be set to the level that maximizes the percentage of total correctly classified observations. Setting the threshold at 0.55 achieved 88% of correctly classified companies. However, while 99% of non-defaulted companies are correctly classified using that threshold, only 17% of defaulted companies are correctly classified.[21] The reasons for this asymmetry in prediction accuracy is a domination of non-defaulted firms in the sample combined with an overlap between defaulted and non-defaulted firms in terms of their risk profiles, as assessed with the estimated credit risk model. Having in mind that type I error (wrong classification of defaulted companies) is in general more costly, we find the first threshold value more relevant for analysis. The second issue is related to the way banks award risk rating to each exposure and to the construction of firm ratings. Significant overlap in terms of our risk assessment between firms classified as defaulted and non-defaulted suggests a possibility that some ratings are overestimated. This type of selection bias has been addressed in the literature. For example, Hornik et al. (2007) compares ratings awarded by different issuers and look for possible systematic overrating by some institutions. There are also more general procedures used in the literature for correcting selection bias, such as multinomial logit (Bourguignon et al., 2004). However, this extension is left for future research.

7 Summary, conclusion and directions for further research

The main goal of this paper is modeling credit risk of non-financial businesses entities by assessing the rating migration probabilities and predicting the probability of default in one year time on the basis of corporate financial accounts. The relevant information needed for the estimation of the model is obtained from CNB prudential regulation

[20] Binary classification models have one of four possible outcomes: a) "true positive" (non-defaulted company is classified as non-defaulted); b) "false positive" (non-defaulted company is classified as defaulted); c) "true negative" (defaulted company is classified as defaulted) and d) "false negative" (defaulted company is classified as non-defaulted). Type I error is the misclassification of a defaulted firm as non-defaulted and type II error is the misclassification of a non-defaulted firm as defaulted.

[21] In general, if the threshold is lower more observed defaults will be correctly classified (high sensibility), but a lot of false positives will arise. On the other hand, a higher threshold reduces false positives but increases the number of false negatives (defaulted companies classified as eligible).

database and individual companies' financial accounts collected by FINA, while our dependent variable – the state of default – is constructed on the basis of the ratings assigned to corporate debtors by commercial banks.

Our research provides a number of important insights. We found no absorbing state in any of the observed matrices, which implies that default state is not final terminal state, so that dependent variable in our logit regression is bi-directionally formed. We generally find a high degree of rating stability in the corporate sector, with the exception of some volatility generated by A90d rated firms. Also, firms in the middle of the ratings scale exhibit asymmetrical migration pattern as the probabilities of their upgrades are higher than probabilities of their downgrades, suggesting that a number of firms use default during the periods of stress as business strategy, which is tolerated by the banks. We find migration frequencies quite insensitive to the economic activity of enterprises. Economic cycle seems to matter only slightly: in the period of lower growth we found that reversal probabilities (PRs) are systematically reduced. Non-financial services have similar profile as the unconditional matrix, but PRs are somewhat lower in industry and higher in construction where we also find evidence of relatively highest rating's instability.

After considering a wide range of financial ratios and other factors as potential predictors of default, the final model predicts probability of default in the following year using the multivariate logistic regression based on the size of the firm (measured by total sales), economic activity (construction and real estate vs. other sectors) and 5 financial ratios: liquidity ratio (measured as cash to total assets), financial autonomy (shareholders' equity to total assets), activity indicator (accounts receivable turnover in days) and two profitability ratios (EBIT to total liabilities and sales and depreciation to total assets). We find that the most important predictors of default risk are the ratio of shareholders' equity to total assets and the ratio of EBIT to total liabilities (both of the ratios are negatively related to the probability of default). In addition, higher liquidity, profitability and sales as well as operating in construction and real estate sector all decrease the companies' probability of default in the following year. As for the role of the economic sector, only the construction and real estate sector dummy was found to have significant explanatory power for the default risk of the companies. The model correctly classifies relatively reasonable percentage of companies in the sample (74% of all the companies, 71% of defaulted and 75% of non-defaulted companies), when the threshold is set in such a way to maximize the sum of correctly predicted proportions for both defaulted and non-defaulted companies. Nevertheless, the number of non-defaulted companies with probability of default exceeding the preferred threshold overwhelms the number of defaults, suggesting possibility of selection bias in the data.

Also, there are several ways to refine the described model in order to increase its scope and explanatory power. Possible selection bias might be corrected by verification of ratings issued by different raters (Hornik et. al., 2007) or by the use of the multinomial logit approach as proposed in Bourguignon et al. (2004). Moreover, redefinition of the dependent variable (modeling default event instead of the default state) would allow for an easier interpretation and practical application of the model. Finally, our sample is limited in the time-series dimension. Increased sample period (covering also a present phase of economic downturn) would allow the inclusion of macroeconomic variables in the model, which related studies show to be significant predictors of the companies' probability of default.

Since data characteristics differentiate, it is also important to be aware of the fact that any modifications can alter the results in the sense of generating risk factors for different time horizons ("point in time" vs. "trough the cycle"). So, it is not only forecast's improvement that we should seek, but also diverse information depending on our analytical needs.

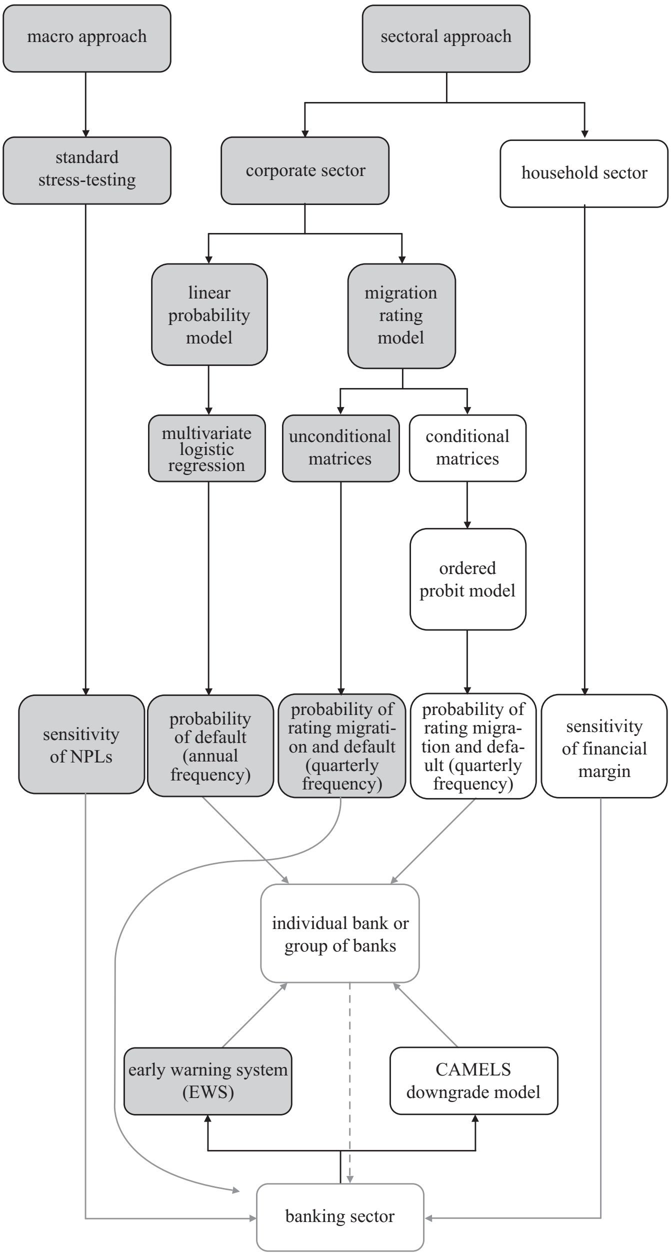
To conclude, this paper contributes to the development of technical infrastructure designed for overall credit risk assessment and offers numerous possible implementations:

- forecasting the probability of default and exposure at default for each bank on a regular basis;
- additional verification of banks' internal credit risk models
- comparison of ratings awarded by multiple banks and identification of possible systematic overrating and comparison of exposure ratings awarded by the banks themselves with the model assessments of credit risk;
- stress-testing of individual banks after identification of relationships between macro-variables and corporate financial indicators (or inclusion of macro variables in the corporate credit risk model when the data time-span permits);
- testing various ad-hoc hypothesis related to credit risk (i.e., the relationship between direct foreign borrowing of domestic companies and credit risk exposure of domestic banks).[22]

[22] One of the major criticisms of the CNB's policy aimed at restraining domestic lending fueled by bank's foreign borrowing was that it deteriorates the quality of banks' balance sheets. The rationale for this view was that it is the more creditworthy borrowers that are more likely to lean on external financing, while domestic banks get stuck with inferior companies in their balance sheets. However, this argument is purely hypothetical and solid empirical attempts to gauge the credit risk effects of stronger foreign borrowing are yet to be submitted.

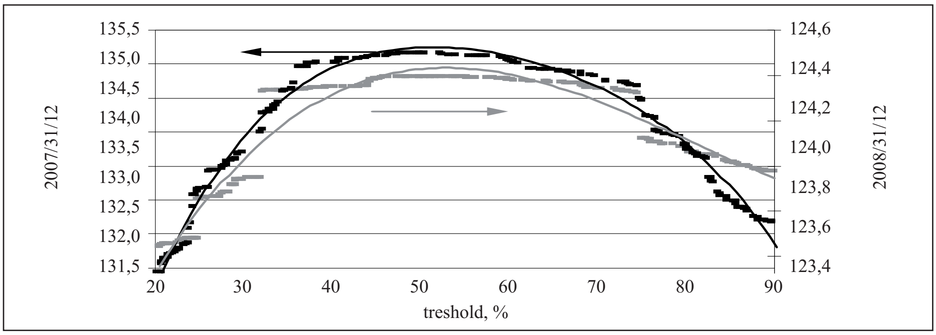
Appendix

Figure 1 Credit risk assessment framework



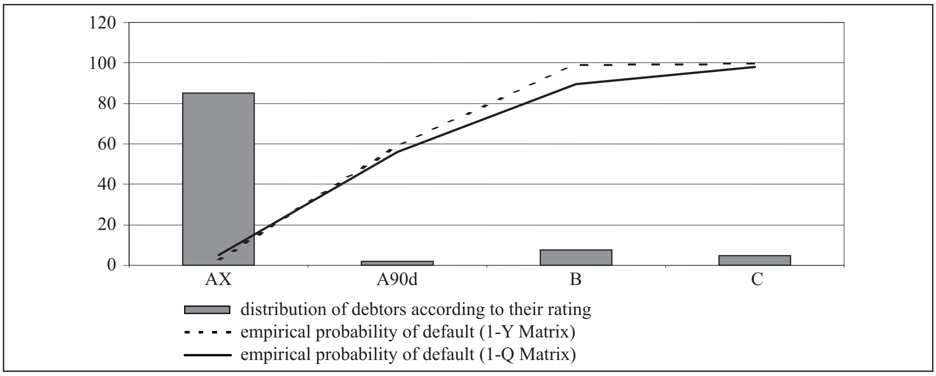
Source: Authors

Figure 2 Total sum of the liabilities classified AX within the group of debtors classified AX and the liabilities classified non-AX within all other debtors depending on the threshold



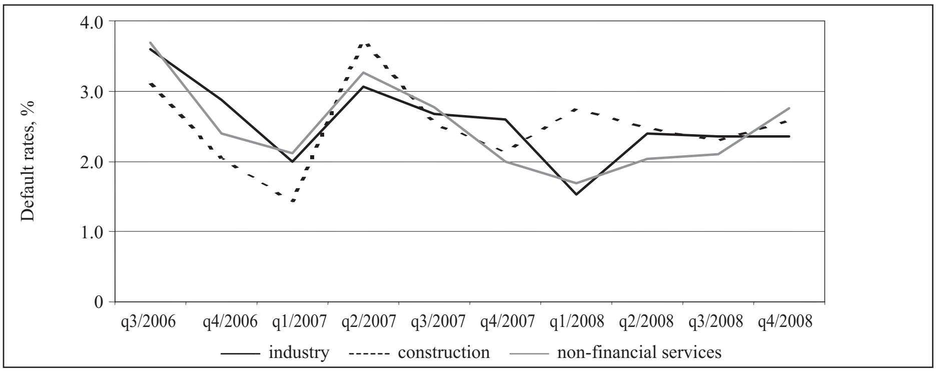
Source: CNB, authors' calculations

Figure 3 Initial ratings and probability of default



Source: CNB, authors' calculations

Figure 4 The evolution of PDs from 2006q3 to 2008q4



Source: CNB, authors' calculation

Table 1 The unconditional migration matrices

1-Year	**AX**	**A90d**	**B**	**C**
AX	95.0	2.0	2.7	0.3
A90d	43.0	22.0	32.3	2.6
B	10.1	1.8	81.9	6.1
C	1.7	0.1	1.3	96.9

1-Quarter	**AX**	**A90d**	**B**	**C**
AX	97.5	1.5	0.9	0.1
A90d	40.6	43.6	14.9	0.8
B	6.0	0.9	90.8	2.3
C	1.5	0.2	0.8	97.5

Note: Initial rating in rows, terminal rating in columns

Source: Croatian national bank, authors' calculations

Table 2 The conditional migration matrices (1-Quarter)

a) Migration matrices conditional on economic activity

Industry	**AX**	**A90d**	**B**	**C**
AX	97.5	1.5	0.9	0.2
A90d	34.6	48.2	16.4	0.8
B	5.3	0.6	91.9	2.3
C	1.2	0.3	0.8	97.7

Construction	**AX**	**A90d**	**B**	**C**
AX	97.5	1.5	0.9	0.1
A90d	46.5	40.8	12.1	0.6
B	8.7	1.5	87.1	2.8
C	1.7	0.0	1.	97.0

Non-financial services	**AX**	**A90d**	**B**	**C**
AX	97.5	1.5	0.8	0.1
A90d	40.9	42.8	15.4	0.9
B	5.6	0.9	91.4	2.1
C	1.6	0.2	0.7	97.5

b) Migration matrices conditional on economic cycle

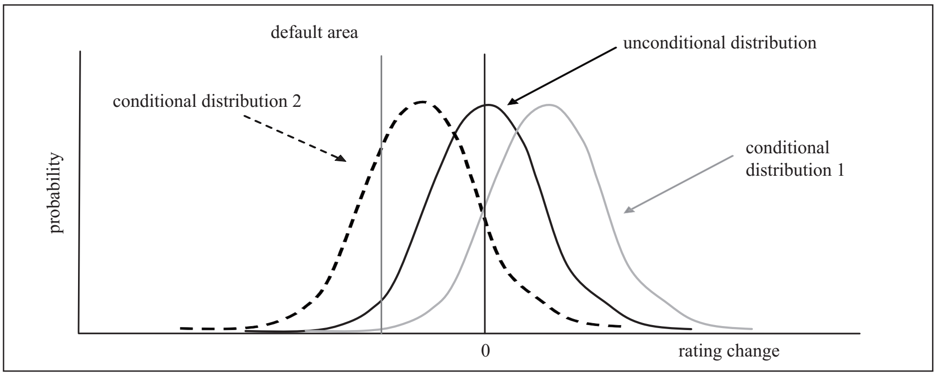
Acceleration phase	**AX**	**A90d**	**B**	**C**
AX	97.2	1.7	0.9	0.2
A90d	45.2	40.2	13.9	0.7
B	6.1	1.0	90.3	2.6
C	2.3	0.2	0.8	96.7

Retardation phase	**AX**	**A90d**	**B**	**C**
AX	97.8	1.3	0.8	0.1
A90d	36.1	47.1	16.0	0.9
B	5.9	0.9	91.2	1.9
C	0.7	0.2	0.9	98.2

Note: a) Initial rating in rows, terminal rating in columns b. Differences in migration frequencies that are statistically significant (5% level) in relation to the parameters of unconditional matrix are in italic. The t-statistics is derived from binominal standard error $\sqrt{\hat{p}_{ij}(1-p_{ij})/N}$, where p_{ij} are population probabilities and $\hat{p}_{ij}$ are sample probabilities (with total number of firms N).

Source: CNB, authors' calculation

Figure 5 Hypothetical distributions of rating upgrades/downgrades



Source: Authors

Table 3 Annual forecast of migration probabilities

Annual Forecast based on 1-Y Migration Matrix

	AX	**A90d**	**B**	**C**
AX	95.0	2.0	2.7	0.3
A90d	43.0	22.0	32.3	2.6
B	10.1	1.8	81.9	6.1
C	1.7	0.1	1.3	96.9

Annual Forecast based on 1-Q Migration Matrix

	AX	**A90d**	**B**	**C**
AX	93.1	2.5	3.9	0.5
A90d	69.6	5.4	22.0	2.8
B	21.8	1.6	68.9	7.8
C	6.2	0.4	2.9	90.5

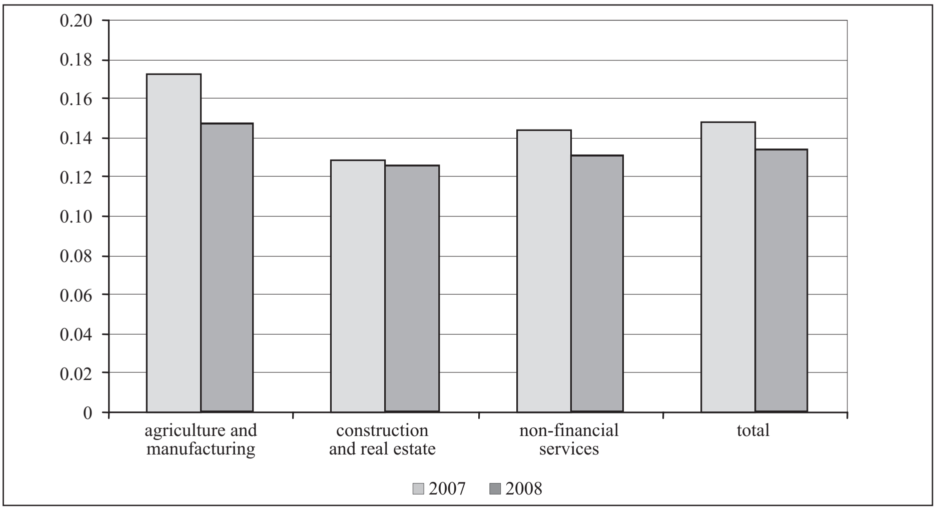
Source: CNB, authors' calculations

Table 4 Number of companies and defaults by year and by economic sector

	2007			2008			Total		
Sector	Companies	Number	Rate	Companies	Number	Rate	Companies	Number	Rate
agriculture and manufacturing	1.887	326	0.173	1.915	281	0.147	3.802	607	0.160
construction and real estate	1.747	225	0.129	2.048	258	0.126	3.795	483	0.127
non-financial services	2.410	346	0.144	2.455	320	0.130	4.865	666	0.137
total	6.044	897	0.148	6.418	832	0.130	12.462	1.729	0.139

Source: CNB, FINA

Figure 6 Number of companies and defaults by year and by economic sector



Source: CNB, FINA

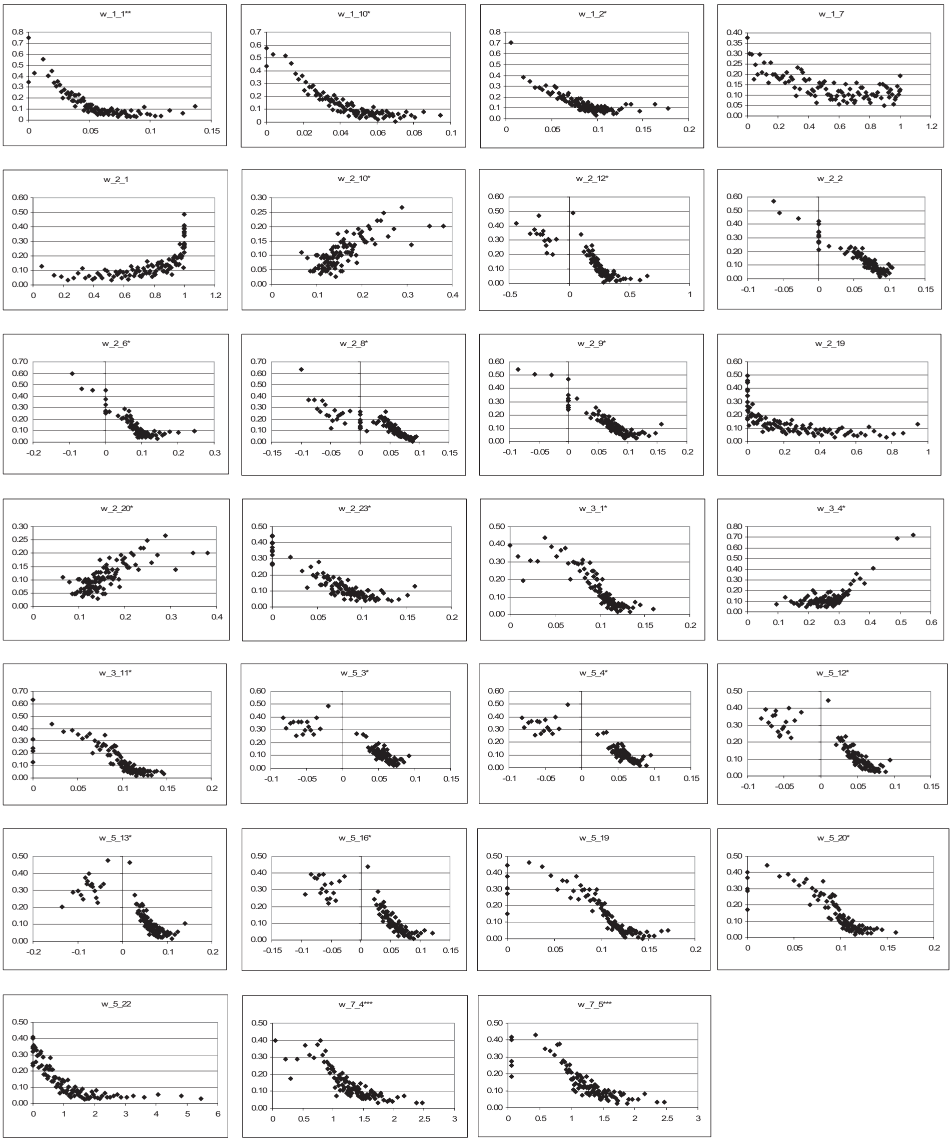
Table 5 Intermediate set of explanatory variables

Varijable	Numerator	Denominator	All companies			Defaulted companies			Non-default companies		
			Mean	Median	Std. Dev.	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
W_1_1**	cash	short-term liabilities	0.19	0.04	0.49	0.11	0.01	0.42	0.21	0.05	0.50
W_1_2*	cash + short-term financial assets	short-term liabilities	0.85	0.42	1.73	0.60	0.19	1.53	0.89	0.46	1.75
W_1_7	short-term assets	total assets	0.55	0.56	0.28	0.46	0.44	0.30	0.56	0.58	0.28
W_1_10*	cash	total assets	0.04	0.02	0.08	0.02	0.00	0.05	0.05	0.02	0.09
W_2_1	total liabilities	total assets	0.75	0.80	0.23	0.85	0.94	0.20	0.73	0.78	0.23
W_2_2	shareholders' equity	total assets	0.22	0.16	0.21	0.11	0.04	0.18	0.23	0.18	0.21
W_2_6*	shareholders' equity	long-term assets	1.81	0.46	6.89	1.11	0.10	6.04	1.92	0.53	7.01
W_2_8*	retained earnings	total assets	0.06	0.03	0.16	-0.03	0.00	0.20	0.08	0.04	0.15
W_2_9*	shareholders' equity	total liabilities	0.53	0.20	1.04	0.26	0.04	0.87	0.57	0.23	1.05
W_2_10*	total liabilities	total assets – total liabilities	15	3	45	25	5	60	14	3	43
W_2_12*	after tax profit + depreciation	debt/365	252	60	898	33	10	585	288	71	935
W_2_19	total assets – total liabilities	total assets	0.25	0.20	0.23	0.15	0.06	0.20	0.27	0.22	0.23
W_2_20*	total liabilities	total assets – total liabilities	15	3	45	25	5	60	14	3	43
W_2_23*	total assets – total liabilities	total liabilities	0.63	0.24	1.16	0.36	0.07	0.96	0.67	0.28	1.19
W_3_1*	total revenue	total assets	1.22	1.00	1.21	0.65	0.42	0.86	1.32	1.10	1.23
W_3_4*	365	accounts receivable turnover	134	71	314	353	100	675	101	69	191
W_3_11*	sales	total assets	1.16	0.96	1.02	0.60	0.37	0.75	1.25	1.06	1.03
W_5_3*	profit after tax + interest expenses	total assets	0.05	0.04	0.10	0.01	0.01	0.10	0.06	0.04	0.10
W_5_4*	profit after tax + interest expenses	total assets	0.06	0.04	0.11	0.01	0.01	0.11	0.07	0.05	0.11
W_5_12*	EBIT	total assets	0.04	0.02	0.11	-0.01	0.00	0.11	0.05	0.03	0.11
W_5_13*	EBIT	short-term liabilities	0.15	0.05	0.68	-0.04	0.00	0.64	0.18	0.06	0.69
W_5_16*	EBIT	total liabilities	0.10	0.03	0.27	-0.01	0.00	0.19	0.11	0.04	0.28
W_5_19	sales	total liabilities	1.83	1.37	1.89	0.80	0.46	1.20	1.99	1.55	1.93
W_5_20*	sales	total assets	1.18	0.96	1.19	0.61	0.37	0.84	1.28	1.06	1.22
W_5_22	sales + depreciation	total assets	1.20	1.01	1.03	0.63	0.40	0.77	1.29	1.11	1.04
W_7_4***	total revenue		43,691,386	10,179,726	118,000,000	18,635,509	3,512,595	66,281,680	47,801,055	11,737,942	124,000,000
W_7_5***	sales		41,764,357	9,577,616	115,000,000	17,220,656	3,031,529	63,267,968	45,790,019	11,165,791	120,000,000

*Note: Variables denoted by *, ** and *** are transformed using the following expression:*
(sign(x)*abs(x)^0.2)/0.2; **(sign(x)*abs(x)^0.1+1)/0.1; *(sign(x)*abs(x)^0.15+1)/0.15.*

Source: CNB; FINA; authors' calculation

Figure 7 Scatter plots of the intermediate set of explanatory variables



Notes:

a) On x-axis: percentile range average of the explanatory variable; on y-axis: average default rate

*b) Variables denoted by *, ** and *** are transformed using the following expression:*

**(sign(x)*abs(x)^0.2)/0.2*

***(sign(x)*abs(x)^0.1+1)/0.1*

****(sign(x)*abs(x)^0.15+1)/0.15*

Source: FINA; authors' calculation

Table 6 Results of the univariate logistic regressions

Variable	Numerator	Denominator	Sign	Pseudo R^2	Area under the ROC curve
W_1_1**	cash	short-term liabilities	negative	0.0875	0.7167
W_1_2*	cash + short-term financial assets	short-term liabilities	negative	0.0498	0.6547
W_1_7	short-term assets	total assets	negative	0.0177	0.5957
W_1_10*	cash	total assets	negative	0.0963	0.7184
W_2_1	total liabilities	total assets	positive	0.0457	0.6791
W_2_2	shareholders' equity	total assets	negative	0.0618	0.7071
W_2_6*	shareholders' equity	long-term assets	negative	0.0767	0.7000
W_2_8*	retained earnings	total assets	negative	0.0606	0.6891
W_2_9*	shareholders' equity	total liabilities	negative	0.0829	0.7060
W_2_10*	total liabilities	total assets – total liabilities	positive	0.0195	0.6153
W_2_12*	after tax profit + depreciation	debt/365	negative	0.0807	0.7270
W_2_19	total assets – total liabilities	total assets	negative	0.0457	0.6792
W_2_20*	total liabilities	total assets – total liabilities	positive	0.0195	0.6153
W_2_23*	total assets – total liabilities	total liabilities	negative	0.0649	0.6788
W_3_1*	total revenue	total assets	negative	0.0706	0.7220
W_3_4*	365	accounts receivable turnover	positive	0.0595	0.6290
W_3_11*	sales	total assets	negative	0.0703	0.7253
W_5_3*	profit after tax + interest expenses	total assets	negative	0.0681	0.6900
W_5_4*	profit after tax + interest expenses	total assets	negative	0.0697	0.6965
W_5_12*	ebit	total assets	negative	0.0771	0.7168
W_5_13*	ebit	short-term liabilities	negative	0.0707	0.7134
W_5_16*	ebit	total liabilities	negative	0.0805	0.7246
W_5_19	sales	total liabilities	negative	0.1021	0.7502
W_5_20*	sales	total assets	negative	0.0703	0.7253
W_5_22	sales + depreciation	total assets	negative	0.0851	0.726
W_7_4***	total revenue		negative	0.0519	0.6731
W_7_5***	sales		negative	0.0557	0.6829

*Note: Variables denoted by *, ** and *** are transformed using the following expression:*
**(sign(x)*abs(x)^0.2)/0.2*
***(sign(x)*abs(x)^0.1+1)/0.1*
****(sign(x)*abs(x)^0.15+1)/0.15*

Source: Authors' calculation

Table 7 Model selection: results of multivariate logistic regressions

	Model 3_1	Model 4_1	Model 5_1	Model 6_1	Model 6_4
C	4.41 (0.22)	-0.41 (0.17)	-0.30 (0.22)	-0.17 (0.22)	-0.06 (0.22)
Construction and real estate dummy	-0.45 (0.06)	-0.26 (0.07)	-0.24 (0.07)	-0.28 (0.07)	-0.30 (0.07)
Cash to short-term liabilities	-0.29 (0.01)				
Cash to total assets		-0.67 (0.04)	-0.67 (0.04)	-0.63 (0.04)	-0.65 (0.04)
Shareholders' equity to total assets			-1.87 (0.19)	-1.96 (0.19)	-2.17 (0.20)
Shareholders' equity to total liabilities	-0.23 (0.01)	-0.27 (0.01)			
After tax profit + depreciation to debt / 365					-0.04 (0.00)
365 / accounts receivable turnover		0.10 (0.01)	0.11 (0.01)	0.09 (0.01)	0.09 (0.01)
EBIT to total liabilities			-0.17 (0.01)	-0.14 (0.01)	
Sales + depreciation to total assets	-0.75 (0.04)	-0.51 (0.05)		-0.37 (0.05)	-0.41 (0.05)
Sales			-0.01 (0.00)	-0.01 (0.00)	-0.01 (0.00)
R^2	0.18	0.19	0.19	0.20	0.20
AUC	0.79	0.79	0.79	0.80	0.80
% of correct 0	71.57	72.37	71.29	74.89	75.90
% of correct 1	73.21	71.20	72.99	71.20	69.50
% of total correct	71.80	72.22	71.51	74.41	75.05

Source: Authors' calculation

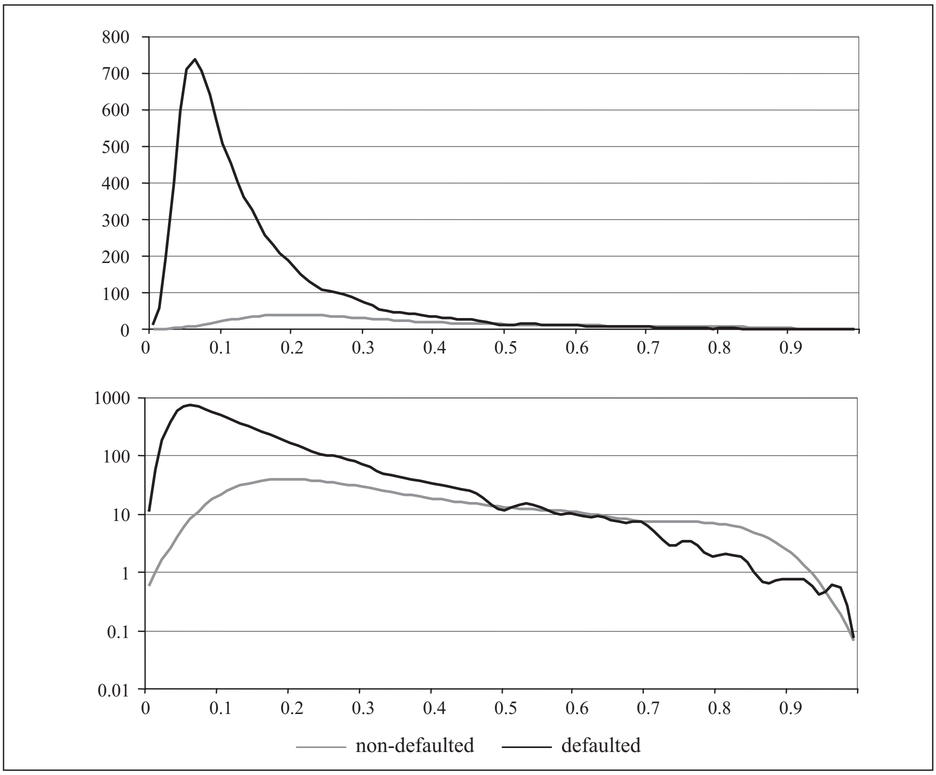
Table 8 Model 6.1 – estimation results

Variable	Coefficient	Standard error	z-statistic	Marginal effect	Marginal effect * 1 std. dev.
C	-0.17	0.22	-0.80		
Construction and real estate dummy	-0.28	0.07	-3.95	-0.020	-0.009
Cash to total assets	-0.63	0.04	-16.21	-0.048	-0.020
Shareholders' equity to total assets	-1.96	0.19	-10.29	-0.149	-0.032
365 / accounts receivable turnover	0.09	0.01	9.32	0.007	0.003
EBIT to total liabilities	-0.14	0.01	-10.47	-0.011	-0.011
Sales + depreciation to total assets	-0.37	0.05	-7.38	-0.028	-0.029
Sales	-0.01	0.00	-4.50	-0.001	-0.0004

Number of obs	11503
Log likelihood	-3589.861
Pseudo R^2	0.1963

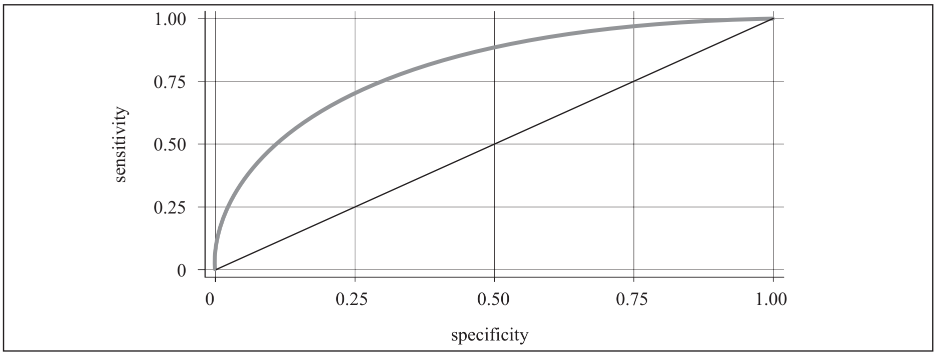
Source: Authors' calculation

Figure 8 Kernel density estimate of default probabilities distribution for defaulted and non-defaulted companies for Model 6.1



Source: Authors' calculation

Figure 9 Model 6.1 – ROC curve



Source: Authors' calculation

LITERATURE

Altman, E. I., 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *Journal of Finance,* 23 (4), 589–609.

Andersen, H. [et al.], 2008. *A suite-of-models approach to stress-testing financial stability*. *Staff Memo,* Norges Bank Financial Stability.

Beaver, W., 1966. "Financial Ratios as Predictors of Failure". *Journal of Accounting Research,* 4, 71-102.

Bernhardsen, E., 2001. "A Model of Bankruptcy Prediction". *Norges Bank Working Paper,* No. 10.

Bourguignon, F., Fournier, M. and Gurgand, M., 2004. "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons". *DELTA (Departement et Laboratoire D'economie Theorique et Appliquee), Working paper,* No. 20.

Charitou, A., Neophytou, E. and Charalambous, C., 2004. "Predicting Corporate Failure: Empirical Evidence for the UK". *European Accounting Review,* 13 (3), 465-497.

Falcon, L. T., 2007. *Logit Models to Assess Credit Risk, Credit Risk Assessment Revisited: Methodological Issues and Practical Implications*. European Committee of Central Balance Sheet Data Offices.

Fuertes, A. and Kalotychou, E., 2005. *On Sovereign Credit Migration: Small-sample Properties and Rating Evolution Faculty of Finance*. London: Cass Business School.

CNB, 2003. *Nadzor nad bankama* [online]. Zagreb: Hrvatska narodna banka. Dostupno na: [http://www.hnb.hr/supervizija/hsupervizija.htm].

CNB, 2009. *Financial Stability Report*, No. 2. Zagreb: Hrvatska narodna banka.

CNB, 2009a. Odluka o klasifikaciji plasmana i izvanbilančnih obveza kreditnih institucija [online]. Zagreb: Hrvatska narodna banka. Dostupno na: [http://www.hnb.hr/supervizija/odgovori-zoki/h-odgovori-klasifikacija-plasmana.pdf].

Hornik, K. [et al.], 2007. "Validation of credit rating systems using multi-rater information". *Journal of Credit Risk,* 3 (4), 3-29.

IMF, 2008. *Republic of Croatia: Financial System Stability Assessment Update.* Washington: IMF; World Bank.

Jacobson, T. [et al.], 2008. "Firm Default and Aggregate Fluctuations". *Sveriges Riksbank, Working Paper,* No. 226.

Lykke, M., Pedersen, K. J. and Vinther, H. M., 2004. "A failure-Rate Model for the Danish Corporate Sector". *Danmark Nationalbank Working Paper,* No. 16.

Nickell, P., Perraudin, W. and Varotto, S., 2001. "Stability of ratings transition". *Bank of England Working Paper Series,* No. 133.

Ohlson J., 1980. "Financial ratios and the probabilistic prediction of bankruptcy". *Journal of Accounting Research,* 18 (1), 109-131.

Richter, F., 2007. *Introduction in Credit Risk Assesment Revisited: Methodological Issues and Practical Implications*. ECCBSDO.

Winkler, G., 2008. *Assessing Credit Risk of the Companies Sector*. Presentation at Annual Regional Seminar On Financial Stability Issues held by IMF and National Bank of Romania.