Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Credit Risk Monitoring in the Czech Banking Sector: Early Warning Model

Author: **Bc. Pavel Mužíček**

Supervisor: **Doc. Ing. Oldřich Dědek, CSc.**

Academic Year: **2009/2010**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 12, 2010

Signature

Acknowledgments

The author is grateful especially to Credit Monitoring department for providing necessary knowledge, ideas and data.

Special thanks belongs to Ondřej Strecker for his useful suggestions and comments regarding the econometric part, and to banking managers Petr Hladký and Petr Weiss regarding theoretical part.

Finally, this thesis would never look in the way it does without wonderful help of my girlfriend, family, friends and colleagues from work.

Abstract

The aim of my thesis is in the first place to show how to deal with a credit risk, and which tools the Czech banking sector uses to minimize it (based on the adequate literature and own experience). In the second place, the aim is to find out the reliable logit model estimating the probability of default during the short period based on available data (in the time of economic crisis in the Czech environment). In the first part of this thesis I am describing the development of Non-performing loans before and during the current financial crisis together with the results of the CNB's stress tests. Next chapter describes the credit risk with the emphasis on the credit monitoring, including the most frequently used monitoring tools. Practical part turns us to the most important EWM model. The strictly confidential banking data (credit account turnovers, credit contract), together with data from the financial statements and CRU registry are the inputs to the Model. The Model should work as an early warning signal detection thanks to the estimate of probability of default (more specifically the watch loan classification or worse) during the next three months.

Keywords	Monitoring, Probability of Deafult, Credit Risk, Logistic Regression

Author's e-mail	pavelmuzicek@seznam.cz
Supervisor's e-mail	dedek@fsv.cuni.cz

Abstrakt

Cílem mé práce je v prvé řadě poukázat na to, jak se vypořádat s úvěrovým rizikem a jaké nástroje český bankovní sektor používá k minimalizaci tohoto rizika (na základě vybrané literatury a mých vlastních zkušeností). V druhé řadě pak vyvinout spolehlivý ekonometrický model (v době finanční krize v českém prostředí) určující pravděpodobnost insolvence v krátkém časovém období na základě dostupných dat. V první části této práci popisuji vývoj úvěrů v selhání před a během současné finanční krize spolu s výsledky zátěžových testů ČNB. Další kapitola přibližuje kreditní rizika, zejména pak jeho monitoring, včetně nejčastějších nástrojů, které jsou ke sledování používány. V praktické části se dostávám k nejdůležitějšímu EWM modelu. Do modelu vstupují citlivá bankovní data (obraty na účtech, úvěrová smlouva), stejně tak jako záznamy z CRU a finančních výkazů. Model by měl fungovat jako včasný varovný signál vzhledem k odhadu pravděpodobnosti defaultu (respektive úvěrová klasifikace sledovaná a horší) během následujících tří měsíců.

Klíčová slova	Monitoring, Pravděpodobnost Insolvence, Úvěrové riziko, Logistická regrese
textbfE-mail autora	pavelmuzicek@seznam.cz
E-mail vedoucího práce	dedek@fsv.cuni.cz

Contents

List of Tables

List of Figures

Acronyms

NPL Non Performing Loans

LLP Loan Loss Provision

EWM Early Warning Model

CNB Czech National Bank

CZK Czech Crown

LGD Loss Given Default

YoY Year-to-Year

EaD Exposure at Default

OLS Ordinary Least Squares

PD Probability of Default

Master Thesis Proposal

Author	Bc. Pavel Mužíček
Supervisor	Doc. Ing. Oldřich Dědek, CSc.
Proposed topic	Credit Risk Monitoring in the Czech Banking Sector: Early Warning Model

Topic characteristics In a broad sense my thesis will deal with the credit risk as the whole. The aim of my thesis will be to create a model estimating a probability of default that will help to take certain action to minimize the risk of bank arising from already granted loans. Based on the available literature, the theoretical part will cover the process describing the origin of the credit risk and closely connected processes. The description of real lending process from the bank point of view will follow.

In the practical part, I will use the bank data to provide a logistic model that will estimate the probability of default. I will have two sources of data, the corporate financials (as the part of the internal review) in the first place, and the data collected by the bank in order to monitor the client's behaviour (credit account turnovers, CRU and other warning signal). In the opposite of the common behavioural models estimating a default during 1 year period, this model will try to predict the default of a debtor during the following 3 - 6 months. The model should help mainly in the case of internal review of the already existing loans.

Hypotheses Is there any difference in the credit risk interpretation from the academic and inside-banking point of view? In the time of economic crisis in the Czech environment, does the reliable logit model exist estimating the probability of default during the short period based on the past behavior of the client and its financials?

Methodology Comparison of the academic and inside-banking view on the credit risk.

The academic point of view will be based on the available literature (IES, ČNB and NKP library and web server scholar.google.com) and will be compared to the practical experience that I have gained during my employment in one of the domestic bank.

The analysis of the bank data with the standard econometric methods. The model of probability of default will be estimated by the logistic regression that, in the opposite to standard linear regression, deals with restricted values of explained variable (values 0 or 1). The preparation and the selection of the explanatory variables for further analysis will be also very important part of the model building. The bank data include all clients from corporate segment that had the loan and were not classified to certain date (i.e. that the loan is considered as standard according to the Czech National Bank rules). All statistical operations will be computed via appropriate and available statistical software. The quality of the final model will be confirmed by the standard econometric methods (e.g. Wald test, measurement of the discriminatory power, etc.) and after that verified on the data from different date.

Outline

1. Theoretical Background
    - Credit risk
    - Loan Process
    - Current Scoring Models
    - Intended Explanatory Variables
2. Practical Part
    - Logistic Regression
    - Description of Bank Data
    - Modeling of Probability of Default
3. Summary of the achieved results

Core bibliography

1. Crouby, M. & D. Galai & R. Mark(2000): "A comparative analysis of current credit risk models." *Journal of Banking & Finance* **24**: pp. 59–117.

2. Dluhošová, D.(2008): "Finanční řízení a rozhodování podniku: Analýza, Investování, Oceňování, Riziko, Flexibilita." *Ekopress*, Praha.

3. Green, W.(1998): "Sample selection in credit-scoring models." *Department of Economics, Stern School of Business, New York University* **Working Paper No. 194 presented at The Japan-U.S. Technical Symposium in Japan**: November 13, 1995.

4. Kadlčáková, N. & J. Keplinger(20004): "Credit risk and bank lending in the Czech Republic." *Czech National Bank*, Prague.

5. Orgler, Yair A. (1970): "A Credit Scoring Model for Commercial Loans." *Journal of Money, Credit and Banking* **2**: pp. 435–445.

Author Supervisor

Chapter 1

Introduction

When we are talking about banking sector and its dependence on economic crisis, it is not a big surprise which segment is the first to react. When there is a shock to a world-wide economy, the corporate sector is affected first. The impact to retail sector is a little bit delayed. Furthermore, the impact is smoothed in time in the retail sector. It implies that any financial institution specialized to retail (mortgages, consumer loan, etc.) has better initial starting point when the crisis breaks out. This is really in the contrast with the fact, that it was the mortgage (mortgage based security) that has caused current financial crisis. But as we are talking about Czech banking sector, the evidence is obvious. Financial institutions that are more dependent on interest revenues from business sector have more difficulties with coping with current loan situation in the beginning of the crisis. In fact, these corporate banks have to take appropriate actions to minimize credit and other risks and to decrease costs as much as possible. On the other hand, banks having their customer's portfolio spread equally between retail and corporate sector (or prevailing retail customers) are more likely to go through the beginning of the current crisis without bigger difficulties (because of relatively stable revenues from retail segment).

The point of previous paragraph and of my thesis is the way how to deal with the credit risk, and to show which tools the Czech banking sector (based on the adequate literature and own experience) uses to minimize it. The most important part of this thesis then will be the early warning model that will try to detect the potential problem clients.

It is possible to work out the chosen topic in different ways. In the first place, I have been concerning the question what area of credit risk to cover. I have decided (after the agreement with the supervisor) to conclude the whole bank's credit risk and chosen only the specific area of the credit risk monitoring. The most difficult part has been the choice of the variables for the Early warning model – the choice of a definition for corporate client's default (i.e. dependent variable) and variables that should explain this default (i.e. explanatory variables). As I am trying to detect the early warning signal, the client's default will be meant to be a classification watch (according to CNB classification) and worse during the following three months. The selection of explanatory variables is then one of the most important parts of this thesis resulting in the final Early warning model.

It will be very difficult to model the probability of default without exact and appropriate data. In this place, I am very grateful to use the internal confidential data and share the public results. The most difficult part of the own research has been the data collection and its further preparation to the final appearance. I have tried to get as many explanatory variables as possible and then let the logistic model choose the best ones.

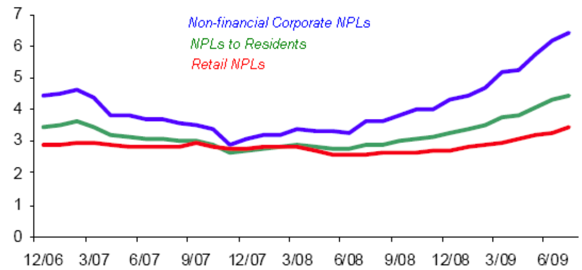
In the first part of this thesis (**Chapter 2**) I am describing Czech banking environment during the crisis (mainly from the non-performing loans and loan classification point of view). The **Chapter 3** deals with the Monitoring of Credit risk, starting with credit risk in general, continuing with the credit risk monitoring process, credit risk management, and the meaning of credit collateral in the credit process. In the **Chapter 4** I am trying to prepare the theoretical background for the Early warning model – the introduction to the scoring models, the most popular Altman's Z-score as a possible alternative to current scoring models and the choice of appropriate estimator for the regression (logistic regression) with associated discrimination methods and fit statistics. The **Chapter 4** ends with the description of the intended variables for the model.

The most important part of this thesis is the **Chapter 5 – Early Warning Model**. It begins with the recapitulation of the data and the process of explanatory variables reduction, so that we can get final model (based on random 50 % observations). The output of the final model is adequately described and fit statistic are explained. The model with all observations included is also presented to compare the results with more robust final model. Finally, I am trying to sumurize the most important results in the **Chapter 6** to state the final conclusions for given hypothesis.

Chapter 2

Czech banking environment during the crisis

Figure 2.1: Non Performing Loans



Source: ČNB.

The Vice-Governor of the Czech National Bank (ČNB), M. Singer, has described development of the crisis in the Czech banking environment regularly within his lectures. One way how to recognize the degree of recession in the financial sector is to analyze the development of bad loans (see the definition of bad loans/clients in the Definition 4.1) ratio compared in time. One can then see that a country in which are non-standard or loss e.g. 50 % of all loans granted, then the recession is really deep, because every second debtor is not capable to meet its obligation. In fact, the loss in financial sector, and thus recession in the banking sector, begins as soon as the aggregate sum of Loss Given Default[1] (LGD) overweights the aggregate margin from the financial trades. This implies that costs (LGD, see the Definition 2.2) overweight returns (margin/interest), and hence the business is loss-making.

[1]Loss given default is the credit loss incurred by an obligor default.

Definition 2.1 (Exposure at Default). The Exposure at default is the expected amount of exposure that a bank would be exposed to, when a debtor defaults on a loan from that bank.

Definition 2.2 (Loss Give Default). The Loss given default represents the percentage of Exposure at default (EaD) which is expected to be lost if a counter party goes into default. In other words LGD is a credit loss incurred if an debtor defaults. The LGD can then be interpreted within the bank portfolio as follows:

$$LGD_P = \sum_i^P \frac{Total\_losses_i}{EAD_i} \tag{2.1}$$

where the i represents the i^th loan in the bank's portfolio (with the sum equal to P). As we are trying to express the total losses *ex ante*, the LGD can be redefined as follows:

$$LGD_P = \sum_i^P \frac{Total\_commitment_i \times PD_i \times recovery\_factor_i}{EAD_i}^2 \tag{2.2}$$

It is quite difficult for CNB to monitor the aggregate LGD, instead the development of non-performing loans [3] (NPL) is better to analyze. (Singer 2009, pg. 1-30) describes the development of non-performing loans ratio as follows. The peak itself, positively meant, was in November 2007 (compared to all loans in the banking sector). Its value was on 3 % (previous negative peak was 15 % in 10/2002) level from all loans in each segment (non-financial corporate, retail and loan for residents). Then the negative boom of NPL ratio started (see Figure 2.1). Not surprisingly, corporate NPL curve looks like an exponential function. In June of 2009 it is almost on the 7 % level, while retail curve has risen to only 3.5 % and residents curve to 4.5 % out of all loans in the segment. This fact supports the idea that corporate based financial institutions have found more difficulties when coping with current crisis. Compared to other European countries, Czech Republic's NPL to all loans ratio is a sort of average

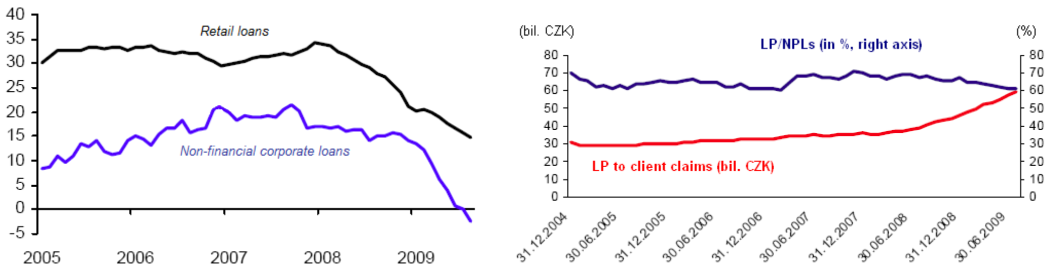
[2] Total commitment is meant to be the current bank's total exposure to the client, PD is the probability of default and recovery factors is not directly determined and depends on the collateral value and the bank's general ability to workout the credit.

[3] According to the national bank's classification, the loans are called non-performing whenever the commercial bank classifies the loan as substandard, doubtful or loss. In other words, whenever the fear from the client's default appears in bank's liabilities as the risk costs with more then 20 % of total debtor's loan volume.

[4].

As we are talking about the total volume of loans, the year-to-year (YoY) difference in the loan volume for non-financial corporate clients has in 8/2009 fallen to negative numbers for the first time since 2005 (-2.6% on YoY bases). Loans in the remaining two segments have also fallen but the values are still in the black numbers (see the left Figure 2.2). The overall volume of loans is thus decreasing. In this context it is worth to show the quality of the Czech loan portfolio. The ratio of the loan provisions to total volume of loans granted is the generally known measure. The right Figure 2.2 shows the given trend. The coverage ratio[5] has a little bit decreased since 2004 (from 70 % to 60 % in 2009), while the loan provisions at all have doubled its size to nearly 60 bil. CZK in 2009. This means that the loan provisions are increasing but compared to the non-performing loans (NPLs) the level is slightly decreasing.

Figure 2.2: The YoY change in the loan volume (left) and the loan provisions (right)



Source: ČNB.

For the retail sector in the Czech Republic ČNB predicts (Singer 2009, pg. 1-30)that more dramatic worsening will appear in 2010, mainly because of the consumer loan defaults. This is obvious, because everyone can imagine the situation where the people working in the companies that go bankruptcy in 2009 are becoming potential non-payers in 2010. This is why the financial institutions have to take now the actions that will ensure them sort of insurance in the predicted bad times in 2010. We are talking mainly about raising of equity. Furthermore, the increase of equity is often part of the intra-group policy as the most of the Czech banks are the subsidiary companies of the multinational not only European mothers. These contra-crisis actions are sometimes overreacted but have its logic. The potential customers in the world need to know that

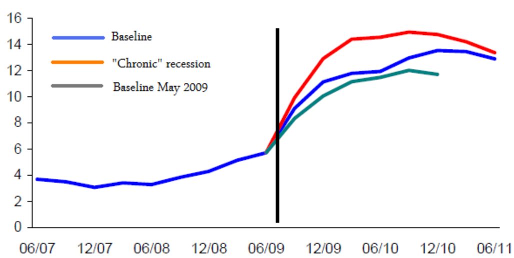
[4]The extremes are Ukraine (24 %), Romania (14 %), Luxemburg (almost 0 %) and Denmark (0.5 %).

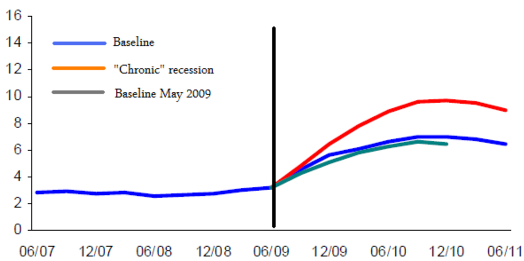
[5]The loan provisions to non-performing loans.

their financial institutions are healthy and will get over any shocks that will appear in the future. They simply want to know that their savings are safe. Adding, a bank with a "strong" capital signals its health. But it is not only the capital requirement financial institution have to undertake when preparing for future problems. Reducing of costs and searching for less risky profits are other decision areas that will need to be taken into account.

All of these precautions create a big pressure on increasing of bank's capital. This pressure means of course decrease of new business, because banks and other institutions would miss capital (that has to be raised) when offering new financial products (e.g. loans). And financial experts are going to even bigger extreme. It is estimated that 1% increase of capital implies 1.2% decrease of sales financing within next four years[6]. Although, capital increase is meant to support people's confidence it is, unfortunately, rather contra-cyclical when it slows down whole economy by cuts in sales financing. Anyone can think of which effect (retail vs. corporate) is prevailing.

Figure 2.3: The ČNB's stress tests: The development of the default rate in the corporate (left fig.) and the retail sector (in %)





Source: (Singer 2009).

ČNB has been regularly designing a set of stress tests on the banking sector since 2003. The results are then published in the Reports about the financial stability[7]. The stress tests try to show how will local banks get over different macroeconomic scenarios. The main prediction, based on the official macroeconomic prognosis of the ČNB, is called Baseline (Table 2.1 shows also the scenario for the "Chronic" recession[8]).

The logic part of the stress tests has been the prediction of the development of the bank market in the following years (see Figure 2.3). The corporate

[6]J.Kunert (3.11.2009, CRO lectures, Prague).
[7]For further information go to ČNB's web page.
[8]For this scenario is assumed longer and deeper decrease of the GDP in comparison with the Baseline.

baseline (i.e. the most probable scenario for the corporate sector) assumes default rate 12 % in 2009 and 11 % in 2010 (i.e. slowly decreasing effect), while the retail (based on the same assumptions) is supposed to be 5.5 % in 2009 and 6 % in 2010 (i.e. slightly increasing which supports the idea of smoothing the effect of the crisis to the retail sector).

Table 2.1: The development of the macroeconomic variables

Type of the scenario	**Baseline**		**“Chronic” recession**	
	2009	2010	2009	2010
Real GDP growth (%, YoY)	-3.8	0.8	-5.2	-1.1
Inflation rate (%, YoY)	1.2	1.1	1.2	1.2
Interest rate 3M PRIBOR (%)	2.1	1.8	2.3	2.0
Exchange rate CZK/EUR	26.6	25.7	27.0	27.0

Source: (Singer 2009) and LaTeXtable.

2.1 Summary of the stress tests

In the half of the 2009, the Czech national bank has published the results of the stress tests. The Czech banking sector looks quite healthy, has sufficient liquidity, own resources and capital. The Czech Republic is thus relatively resistant against external shocks like the current financial crisis. Despite this good position, in the nearest future is expected:

- the increase of NPL,
- decrease of growth of total loan volume and
- the increase of the credit risk (Singer 2009, pg. 29).

As the Czech economy is closely connected to the foreign countries (mainly to the West part of the Europe), the Czech GDP growth will be then highly correlated to the development of the crisis abroad. Nevertheless, the growth economic potential of the country will be suppressed. The reason is obvious. Anyone can see the current behavior of the bank market. The central bank is pushing down the interest rate as much as possible to get the economy started (and to avoid negative inflation), but the interest rate for the company is held on the same level, no matter what the repo rate is. The position of the companies when negotiating a suitable loan will be then probably very difficult.

2.2 Estimate of the NPLs shares in particular industrial sectors of the Czech economy

This section is based on currently published (January 2010) CNB's statistics. Based on the development of volume of non-standard loans in particular industrial sectors of the Czech economy, it is possible to derive the percentage shares of these loans relatively to the total credit exposure of the banks in the given sectors (NPLs ratio). The values of the shares of NPLs for the given sectors and also for the retail and small business sector are shown in the Table 2.2.

Table 2.2: The monthly shares of non-performing loans in particular sectors (to 31.12.2009)

Sector total	**5.2 %**
Manufacturing	12.7%
Production and distribution of electricity, gas,...	3.6%
Construction	13.4%
Wholesale and retail	9.9%
Automotive and stock-keeping	6.8%
Finance and insurance	0.6%
Real estate	3.6%
Professional, scientific and other activities	6.0%
Non-financial corporates - residents	7.8%
Retail sector - residents	4.1%

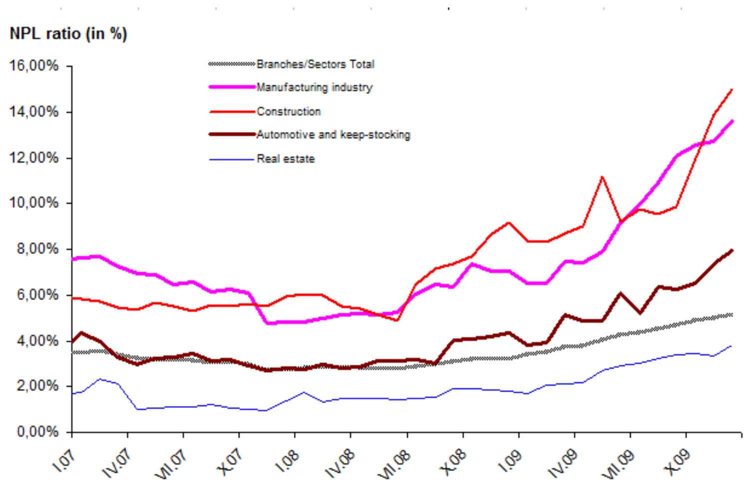
Source: ARAD statistics (CNB) and LaTeXtable.

Following the beginning of the Chapter 2, the Table 2.2 is a sort of a connection bridge between this chapter and the final model estimating the probability of default, as the sector, respectively the CZ-NACE code[9], should play one of the key role in denoting the probability of default. As can be seen in the Table 2.2, the companies running its business in the sector of construction or manufacturing has had in the end of 2009 the biggest problems with the payment their bank loans.

One can object that a view on the one month of some period might be misleading and it is generally truth. But not in this case. When looking at the development of the non-performing loans in each industrial segment for the period of 2002 to 2009 (see the Figure 2.4 and Figure 2.5), anyone can notice

[9]Classification of economic activities. The CZ-NACE code has substituted the OKEC code (Sector classification of economic activities) since 1.1.2008.

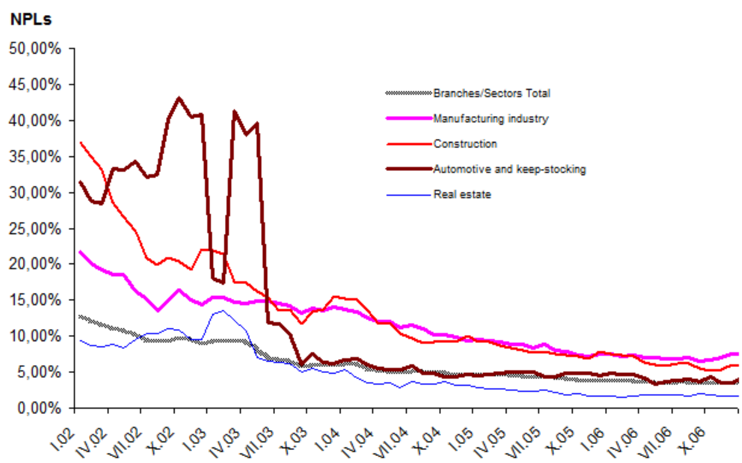
Figure 2.4: Development of the shares of non-performing loans in particular sectors from 1.1.2007 to 31.12.2009



Source: ARAD statistics (CNB) and own computation.

three important facts. The first one is that the actual NPLs correspond to long-term situation since the beginning of the current crisis (late 2008). The second one refers to the comparison of both figures concerning the problematic sectors - while manufacturing and construction have been the worst segments during the both economic crisis (2002 and current), the automotive sector has behaved extraordinarily throughout both crisis. This behavior is evidently caused by the government interventions in the European economic space.

Figure 2.5: Development of the shares of non-performing loans in particular sectors from 1.1.2002 to 31.12.2006 (in %)



Source: ARAD statistics (CNB) and own computation.

The scrap subsidy [10] that was rejected by the Czech president Václav Klaus but implemented in several European countries in 2009 is a sufficient example. The last notice concerning development of the non-performing loans is quite depressing. At the end of the first decade of the 21^{th} century, the crisis in the Czech republic does not seem to go to its end. None of the shares of NPLs has decreasing tendency.

Nevertheless, when looking at both figures, one can predict the development of the NPLs in automotive industry. As the subsidy policies of European countries will come to its end, the negative effect will probably appear, even more drastic than without them. This artificial help (very often approved by government because of visible positive influence on the votes to parties that have given the appraisal) is rather contra-cyclical, because the long-term negative effect overweight the short-term positive effect. Anyone can think of breaking the Smith's *invisible hand of the market* by the government shots to the economy.

2.3 Loan Classification

Definition 2.3 (Classified loan). The loan is considered to be classified, whenever the the provision rate is greater than 1 %. There exist many loans that are classified but the loan provisions are not created as the credit exposure is fully covered by the loan collateral. In term of classification classment, classified loans are watch, non-standard, doubtfull and loss.

Definition 2.4 (Non-performing loan). The loan is considered to be non-performing, whenever the the provision rate is greater than or equal to 20 %. There exist many loans that are classified but the loan provisions are not created as the credit exposure is fully covered by the loan collateral. In term of classification classment, classified loans are non-standard, doubtfull and loss.

When the bank is not sure whether the client will meet its (re)payment obligations, the loan provisions has to be created. The duty is given by the central bank and the exact shares are given by the Basel II Accord. The

[10] In the Czech it is used the word "šrotovné" for the financial contribution from the state for the purchase of new personal car with the obligation to transfer the old one to car breaker's yard in order to let it ecologically liquidate. This state intervention tries to moderate the results of the economic crisis indicated by the decrease in new car purchases.

numbers valid for the Czech Republic is shown in the Table 2.3. One can also notice how much can bank use as a tax deductible item[11][12].

Table 2.3: Credit Classification and Provisioning due to CNB's Guidelines

Category	Past Due Days	CNB's LLP	Tax's LLP
Standard	< 31 days	0 %	0 %
Watch	30–90 days	1–19.9 %	0 %
SubStandard	90–180 days	20–49.9 %	1 %
Doubtful	180–360 days	50–99.9 %	10 %
Loss	> 360 days	100 %	20 %

Source: ČNB (2010).

The bank needs to consider the question of the tax optimization. On the one side, the loan provision creates costs, on the other side the tax deductible decreases the tax base. There is also some evidence that the bank's management often use the provision as a tool to control the net profit. More information about the loan provisions, earnings management and signaling effects in e.g. Ahmed *et al.* (1999).

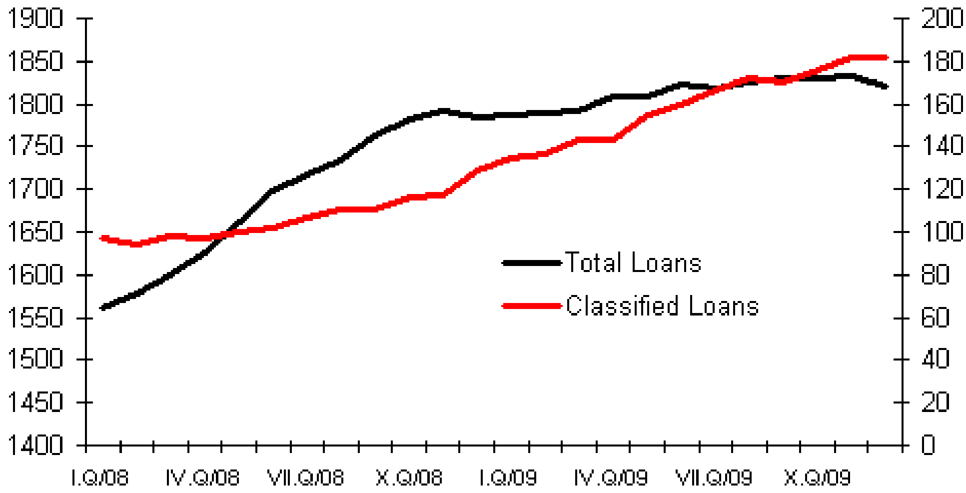
2.3.1 Classified loans in the Czech Republic at the turn of decade

Zavadilová (2010) published in March 2^{nd}, 2010 (Magazin E15) that the classified loans in the Czech Republic have exceeded the 10% threshold. One can see in Figure 2.6 that classified loans have reached cca 180 bil.CZK in comparison to the 1800 bil.CZK of total loans granted. The situation of classified loans has really worsened during last 2 years, when the ratio was cca 6.5 % (100 bil.CZK classified to total 1550 bil.CZK). The article also supports the theory mentioned before that the financial institutions with prevailing retail/household business have had better initial starting position when the crisis brought out. With the data at the end of 2009, there was only 4.8 % of classified loans in Česká spořitelna, while Komerční banka reported 10.4 % of classified loans. The Unicredit Bank reported 8.6 % in avarage, but when purged by the corporate sector, the retail achieved only 5.8 %. All of these data support the idea, that the corporate sector is the first to react to the crisis. "The biggest

[11] Mejstřík (2005)

[12] The Loan Loss Provision permitted by the Tax Authority.

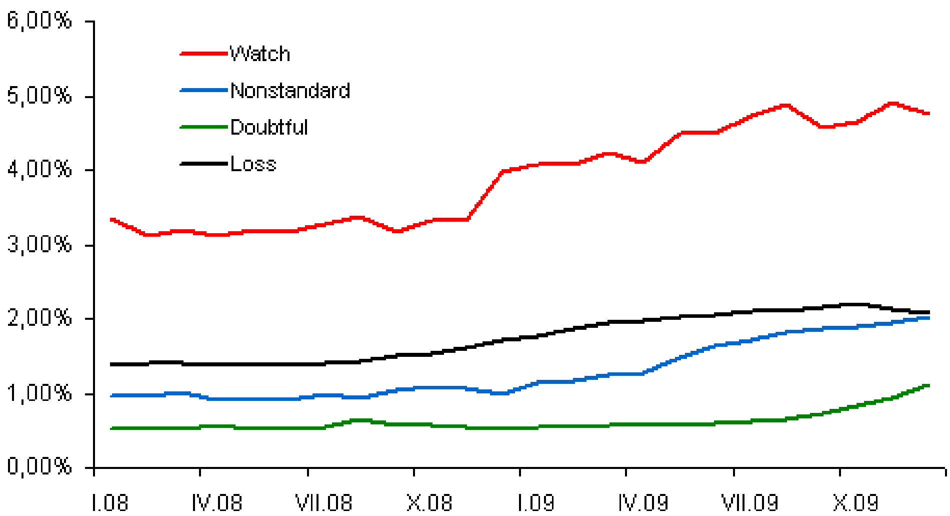
Figure 2.6: Development of the total and classified loans from 1.1.2008 to 31.12.2009 (in bil. CZK)



Source: ARAD statistics (CNB) and Zavadilová (2010).

volume of risk loans fall on the firms (128 bil.CZK), on households then 70 bil.CZK. The worst payment discipline is in manufacturing industry, trade and real estate, on contrary, the health service and educatinal field have appeared as the best one"[13].

Figure 2.7: Development of the classified to total loans ratios from 1.1.2008 to 31.12.2009 (in %)



Source: ARAD statistics (CNB) and own computations.

It is worth to mention the proportion of each classification categories com-

[13] (Zavadilová 2010, pg. 8).

pared to the total loans granted (see Figure 2.7). Watch loans (i.e. past due days between 30 and 90 days) have increased from 3 % in 2008 to nearly 5 % in late 2009. In fact, it has been the biggest increase from all risk categories, and it really responds to the financial crisis reality. Most of the firms have been affected by the crisis, but have survived it. This is why the watch loans have the biggest portion of classified loans. Adding, anyone can imagine the situation, where the debtor is not really in trouble in case of delayed payments for 30 days. This is also the reason, why the CNB classifies the loan as the non-performing after the past due days achieve 90 days.

2.4 Chapter Summary

This chapter has showed the negative development of NPLs in the Czech banking sector before and during the current financial crisis. The CNB's stress tests has been mentioned in order to show the prediction of NPL's further development, negative in short-term horizon, but positive in the long-run. Then the 2010 January's sector-decomposition of NPLs has been mentioned to maintain the possible risk in the automotive and the machinery industry. Finally, the March's article has confirmed the theory mentioned in the Chapter 1 about the better initial position of banks oriented to the retail sector.

Important notes

On the contrary to the general loan classification, in the Early Warning model will be bad client considered as the one with **watch classification** or worse. The reason is obvious. The EWM should detect problems with comfortable advance and not assign firms that already have the problem. Another thing to point out is that this thesis is aimed to explain corporate default, so the ideas and descriptions that follows are meant to describe **corporate sector** (unless it is mentioned in other way). According to equation concerning LGD ($LGD = Total\_commitment \times PD \times recovery\_factor \div EaD$), the EWM model will estimate current probability of default. With appropriate recovery factor and EaD, one can easily compute the exact LGD. The estimated probability of default, on contrary to internal rating, is supposed to be more actual. While the probability of default based on the internal rating is generally computed once or twice a year, the EWM model is supposed to compute the PD on the monthly basis with regard to actual client's behavior (credit account turnovers, debits, etc.).

Chapter 3

Monitoring of Credit Risk

... BANKING IS A BUSINESS WITH THE RISK

This chapter provides information about credit risk, the origin of this type of risk and finally, the very important part of this thesis, the **Monitoring of Credit risk**.

The chapter itself is closely connected to the practical part of this thesis, as the Early Warning Model represented in the following chapters should point out those clients, that are likely to default and thus need to be checked. The contents of this check is then described in Section 3.6 and Section 3.7.

3.1 Risk in the financial sector

The main function of every commercial bank is collecting deposits and granting loans. It also represents the most typical products, but whereas the trade with standard non-financial products ends at the same time as the product has been sold (in the most cases), the trade with financial product ends with its settlement (maturity), which can be from overnight to decades. In the meantime, financial institutions face several types of risk. We can divide them into following groups (for more information about each risk see e.g. Mejstřík *et al.* (2008)):

1. a risk that is caused by the creditor itself (e.g. operational, liquidity or insolvency risk),

2. a risk that is caused by the debtor (e.g. credit risk),

3. a risk that is caused by the third part of a contract(e.g. market, monetary or interest rate risk).

For further description, the first and the third group will be omitted so that the chapter will deal only with the the risk that can arise on the debtor side (thus the debtor can influence it). It means that we will *ceteris paribus* assume that the risk has its origin only on the client's side.

3.2 Credit risk

Every creditor (financial institution) faces the credit risk (expressed by LGD) associated with a credit contract. Credit risk arises because there is a chance that the debtor will not repay the granted credit (i.e. the debt), interest from the loan and other loan fees[1]. For this reason, the banks are carefully verifying every potential or current customer and use different collateral instruments (guarantees, assignment of claims or pledge of movables or immovables) to secure themselves against the potential loss.

There are two reasons of the credit risk. Internal and external. The internal one goes hand in hand with the wrong decision of the management of the credited company throughout its running (e.g. wrong business deal) leading to inability to pay, while the external one is independent of company's decision and is given by the development of the whole economy, political situation or other external shocks (in fact the external reason of credit risk is very similar to a group of risk that is caused by the third part of a contract and thus will be left).

The (in)ability of a client to meet its obligations will then be treated carefully.

3.3 The success of a business

The success of a bank (at least most of them) is based on an ability to take care of a current loan portfolio. A bank needs to identify the risk as soon as possible, monitor its portfolio, review every loan regularly and raise the data

[1] Mostly the administration fees associated with a loan settlement and its later running. Furthermore, when the debtor is somehow overdue, there is a penalty interest that is computed as a given portion of the overdue amount.

quality. They should also structure the trade in the right way[2], prepare the loan contract carefully and decide whether to go for the business or not. All of these precautions are the only way how to minimize the potential risk.

When talking about the business itself, the revenue side of the business is very important for the bank, especially the new trade. As we are then thinking about connection between the risk and the trade, the risk is just *bias of the expected revenue from the business.*

3.4 Purpose of the Risk Monitoring

As soon as the loan (or other form of the loan relationship) is approved, it is necessary to monitor whether the loan is being used for its primary purpose and paid regularly. In the most cases, the quality of the loan is not constant during its existence. In other words, the risk associated with the loan is not constant over the time and thus has to be under control of the creditor. The financial situation of the debtor, the situation of the whole sector or the collateral value is influenced by many factors that change in time and after that, change the quality of the loan itself. This is why the banks often check the level of the amount drawn. They monitor the overdrafts or simply regulate the level of drawing.

3.5 Credit Risk Monitoring

Some literature (e.g. (Navrátil 1994, pg. 185)) mentions that the credit administration, the credit review and the credit monitoring are frequently used as synonyms, but the reality is different. Each of them has its own place in the banks hierarchy. Following definitions are based on author's own experience (so it cannot be generalized to every bank).

Definition 3.1 (Credit Administration). The Credit Administration department is a part of a financial institution that is responsible for the administration of the loan. It involves processing of all documents associated with the loan (e.g. the loan application, appraisal or review). In other words, the credit administration is the process of loan documents computerizing.

[2] The business is meant to be well-structured when the risk is well-balanced over the whole portfolio.

Definition 3.2 (Credit Review). The Credit Review involves a process of checking the debtor and after that re-fitting the loan settlement. After the loan contract has been settled, there are regular reviews in order to check the debtor's solvency. The more risky the client is, the more often is review done. Credit review might also mean new pricing of the loan, adjustment of payments or credit renewal (the renew credit contract after the previous loan have been payed). The business team involving relationship manager and credit specialist is responsible for processing of credit review (again, cannot be generalized to every bank).

To sum up, all of these terms are connected in some way, but have different meaning. After the loan has been granted, the department responsible for the Credit administration (see Definition 3.1) enters the given documents into appropriate PC application, transforms them into bank's ledger documents, and mainly the realization of credit drawing and repayments. On the regular basis, the credit reviews (see Definition 3.2) are done to check whether the debtor is capable to meet its obligations and whether the loan contract is still actual. The department responsible for the Credit monitoring (see Definition 3.3) processes the data in order to monitor/detect potential non-payers.

This thesis will try to focus mainly on the last activity, the *Credit monitoring*. This part of the credit risk management side of the bank is not so public known, but interested readers should know that it is very important part of every financial institution. Credit monitoring department (or department responsible for credit monitoring) minimizes the risk costs by detecting potential default clients, or even the fraud cases. One can object that this "work-flow" should be covered by the relationship manager (direct contact with a client), credit officer (direct communication with relationship manager or the client) or given underwriter (accepting/rejecting credit application), but this evidently clear and warranted control of the business has its little fault, the information asymmetry.

All of these bank officers are working properly and decide on their best own, but the final decisions might be influenced by the so-called *"Inattentional/perceptual blindness"*[3]. Nevertheless, the department responsible for the

[3] For every living person, there is a possibility of suffering from a certain type of blindness as a consequence of long-term stay in unchangeable environment. This "blindness" is not dependent only on location, but mainly on the character of the working area. For the banking environment, it is very difficult for the officer to see beyond her common job description, because of e.g. huge range and quick development of the target business.

credit monitoring should supply this working restriction[4]and fill in the possible gap in the business circle consisting of the bank officers mentioned above.

Nevertheless, the primary aim of the department responsible for the Credit monitoring is to monitor current loan portfolio. The processing[5] the database of all debtors and preparing given reports (e.g. overdue report or current account turnovers report) that summarize the most risky one is then one of the tool how to achieve positive results. Another job description is to detect the potential bad clients and following communication with given business team or, in worse cases, with Credit Workout and Restructuring department.

Definition 3.3 (Credit Monitoring). A phrase Credit Monitoring will be used in the sense of a continuous monitoring of client's activity and creditworthiness during the time of a loan maturity between the timely discrete credit activities (credit renewals and reviews).

Monitoring of particular credit is done in order to ensure the persistent control on the capability of the client to meet its credit obligations. Credit officer thus considers the degree of a risk associated with the loan, and whether the loan is sufficiently secured. The result of this procedure is then the assessment of the fact, whether the client is in the final effect capable to fully repay the principal and interest. Financial institutions are monitoring not only particular credit, but also the quality of the whole credit portfolio.

The purpose of the portfolio monitoring is to reduce the risk caused by the combination of particular credits, and of course the aim is to keep the balanced portfolio with the strategical goals of the bank. The following paragraphs will try to present the process of the credit monitoring process. The main sources are two: Navrátil (1994) and Polidar (1999). Finally, the author's own experience is added.

3.6 Credit monitoring process

This section is based mainly on Navrátil (1994).

[4]Sholarpedia denotes the inattentional blindness as the "failure to notice a fully-visible, but unexpected object because attention was engaged on another task, event, or object". For more information see e.g. Simons (2009).

[5]Independent IT department is responsible for its running.

Phases of the credit monitoring process

Monitoring of particular client's creditworthiness will be described in this section. Practically, monitoring involves two types of activities:

- current monitoring of the relevant information and factors in the way that is going to be described in the Subsection 3.6.1.
- detailed checks consisting of formal and complete updating of degree of the credit risk in the way described in the Subsection 3.6.2.

3.6.1 Current monitoring

It is obvious that at least two independent employees (within two stages of the financial institution) should monitor the loan in order to keep objectivity of the decisions. In the first place, it is the relationship manager, who creates the business. It is he/she, who has something like the connection bridge between the client and the risk side of the bank, and who is from the very beginning supposed to continuously monitor the client. In the second place, there should be another employee that revises the same loan. The bank then should ensure his/her full independence on the primary decision (of granting a loan). This function might be undertaken by the credit officer.

The relationship manager should ideally participate directly on the primary decision (of granting a loan, understand well the character of the client and its business sector, and communicate with the client regularly. Current check of the loan is focused to the following tasks:

- the check of the incoming payments and monitoring of the principal and interest amortization on the basis of the conditions stated in the loan contract;
- behavioral analysis of the client in the bank;
- visits at the client in order to contact client directly and evaluate the management and business on the continuous basis;
- evaluation of the impact of the new credit application to the current loans;
- gathering the financial information directly from the client.

The last task is very important for further monitoring of the loan by the Credit Monitoring department. It involves mainly the transformation of the financial statements into the form of given tables (that make up a part of the database needed for further monitoring) and analysis of the financial ratios in the way that will be described in the Subsection 4.5.4.

Other important task is to update the information about the subject of collateral, incl. the valuation review according to the subscribed loan contract. The relationship manager should also continuously monitor the economic, industrial and political factors that could potentially threaten solvency of the customer.

Finally, according to the administration area, the task is to file the loan documents properly and keep its quality so that it enables continuous check of the credit. These documents according to Navrátil (1994) involve: credit application, financial statements and appropriate tables, credit questionnaire, notes from the meeting with client/about the project, correspondence, memorandum, legal documents, newspaper clippings or reports from visits.

The task that goes beyond the bank's "frontier", but should be also part of the client's check-out, is to find out all client's commitments out of the bank if exist. One possibility is that the client informs RM on his own or on demand and RM fully trusts it. Another possibility is to check regularly the Central registry of the loans (CRU) provided by the Czech national bank for corporate debtors.

3.6.2 Formalized detailed check: The review

Formalized detailed check, or simply the loan review, should be planned regularly for every loan (in standard cases the period of the regular checks is usually once a year). The exceptions are the loans with high credit risk. It is obvious that these risky loans are not easy to detect and occur in different segments. Due to this fact, it is not easy to choose the loans, to which the bank's officers should pay more attention. In fact, most of the financial institutions set the loan criterion which divides the granted loans into several subgroups. The criterion can differ within the whole banking sector. To be clear, not all clients are controlled on the regular basis, the reviews are worked out regularly by the relationship managers only in case that the loan has not been yet classified as watched or worse. In these cases, the loans are within the bank transfered from normal business to the department that is responsible for credit workout or debt restructuring.

According toNavrátil (1994) the limit concerning the overall amount granted to the company (in case of group of economically connected companies the limit is just the some over the single loans provided to its group members). This limit can be e.g. €[6] 5 mil. (or its Czech Crown equivalent) and means that every company that has borrowed money over this sum has to be reviewed regularly more often (e.g. twice a year) than those, whose exposure have not reached the given limit. Of course, this methodology can be misleading as most of the clients (concerning the overall count) are the individuals whose loans are far away from this limit. But, once again, in this thesis we are interested in corporate sector (i.e. companies, conglomerates and big multinational groups) and, what is even more important, the exposition to individuals (or generally the retail segment) is usually far away from the sum that is granted to a company. In this place, it is worth to mention that it would be very expensive and uneconomical to check in detail every bank's client more often that once a year. From the author's point of view, this type of limit is quite out-of-date as the literature source has its origin in times of beginning market environment[7].

With a quick link to the previous paragraphs, the Early warning model (EWM) that is going to be discussed in Chapter 5 is thought to substitute these limits and effectively determine which loan should be reviewed more often (based on the estimated probability of default). In other words, the EWM is supposed to indirectly decrease the cost of the risk by:

- minimizing the cost of review and
- decreasing the Exposure at Default by early detection of potential bad clients.

These two types of limit determines whether it is necessary to review the client more often than it is common in the standard cases, but does not say anything about the concrete frequency of the checks.

[6]The Euro currency is mentioned because of the fact that all of the Czech commercial banks are the members of the multinational companies.

[7]For more information about starting conditions in the Czech banking sector see e.g. Mužíček & Mejstřík (2007).

3.7 Credit risk management

When the socialism fell down in the Czechoslovakia in late 1980's, the situation in the banking sector was quite unbalanced. Nobody can be surprised that the bankers were old-fashioned with insufficient know-how of modern banking as the Czechoslovakia was isolated from the advanced Western market. The Czechoslovak Trade Bank (ČSOB) was the only local financial institution that provided services in the area of foreign trade financing and convertible currency operations[8]. This sort of privilege ensured bankers working in ČSOB that they were very valuable in the labor force. There were many financial institutions founded in the new decade and they all needed experienced bankers. The demand for experienced bankers exceeded supply. One of the possible solution was to teach and educate current non-experienced bankers with a suitable literature. The author has got to know[9] that almost every banker that wanted to learn about modern banking practices was supposed to study the Management of Bank and Banking Trades written by Polidar(Polidar (1991) and Polidar (1992)). In the following paragraphs will be mentioned the principle of the credit risk management as it is described in Polidar (1999).

Credit risk

The risk is associated with almost every banking activity. There is a certain chance that the client will not repay the loan provided by the bank and the agreed interest and charges. This chance is represented by the credit risk and is rising proportionally with longer maturity of the loan. As have been described earlier, the bank faces several types of credit risk. Credit risk can be then minimized in each part of the trade. In the time of the loan negotiation, the bank is supposed to check the current and future situation of the potential customer. After the loan has been granted, the bank has to regularly check the client's creditworthiness in order to minimize all mentioned risk. Another way to restrict or even to fully exclude the credit risk from the trade is the security of the loan. When the credit collateral is negotiated, the agreement has to also include the instruments related to the case of the forced recovery of the debts.

[8]ČSOB, a.s., (online).

[9]The idea based on the consultations with the member of board of directors of one of the local banks that is responsible for the risk.

The purpose of credit collateral

The primary use of credit collateral is for the following reason:

- The ability of the bank to recover its claims even in the case of default of the debtor during the maturity of the loan.

3.7.1 Instruments for credit risk management

Banks eliminate the origin of the risk situations by:

- The check of the creditworthiness of the debtor,
- the credit limitation, and
- the credit control of the debtor.

The check of the creditworthiness is the necessary part of each credit application and is a logic instrument used before the loan settlement. The purpose of the check is to globally evaluate all possible risk associated with the money lent. Based on the results of the check the bank decides whether, and - in case of positive risk manager's opinion - under what conditions the loan will be approved and granted.

The check of the creditworthiness involves:

- The check of the legal status of the credit applicant,
- the check of the personal credibility of the credit applicant, and
- the check of the credit applicant's economic and financial situation[10].

Legal relation of the credit applicant

The following legal conceptions have to be clearly stated in the credit application:

- The legal entity: It has to be stated, whether the client is a company of individual (physical entity running a business on her own), trading company or corporate enterprise, syndicate or other legal entities and physical individuals. The bank needs to verify the client's legal entity in order to make sure that the client fulfills the legal conditions to lie under an obligation resulting from economic deal.

[10] (Polidar 1999, pg. 213)

- Identification of the legal and factual existence: This is very important to check the legal and factual existence, because there exist many cases, where the company of the credit applicant does not really exist and thus the bank lends its money to "nobody". This is one of the typical fraud case[11]. As almost every bank does not provide any loan to an unknown client, this identification is checked when initiating first business connections[12].

- Property relations: It is necessary for the creditor to know the property relations of the debtor. This holds especially for the small business clients and individuals (with unlimited liability) in case of default of the client[13]. Polidar (1999)'s experience from the foreign banking point a big interest on the property relations of the small business and individuals as this segment of clients creates in average a numerous and meaningful part of the bank's portfolio. In addition, this segment belongs rather to the non-stable part of the business sector.

- Other legal contingencies: Other events that might play a significant role from the legal point of view are the entitled person (a legal capacity of the representative person) and the representation (a bank should negotiate only with a person that is entitled to representation of the company) of the company. The bank should not make a loan settlement or any other negotiations with a not authorized person.

The personal credibility of the credit applicant

The credibility of the applicant represents the character of the client as a bank's reliable business partner. The client can be successful and powerful, but if the bank has not a previous good relationship and experience with a customer, the loan will never be granted. The characters of credibility due to Polidar (1999) are the stability of the financial management (represented by the rentability,

[11]Fraud case is the most scary thing for every bank. There exists many sample fraud cases that are presented during the obligatory learning lectures to every employee of a bank. Fraud case is for a bank like a bank heist.

[12]It is generally when the client wants to open a current account.

[13]The following example can be given: The client entrepreneur was granted a loan for which he used a flat as a collateral. The bank did not figure out properly the use of property in the shared ownership and did not realize that he was living with a common-law wife, so that he had a choice to change the ownership of the flat to his companion. And he really had done so, as far as he felt that his business was going wrong. The only collateral for the bank was then not enforceable by law. This implied a net loss for the bank as the debtor was not able to meet his obligation and did not leave any collateral.

trade sales, etc.); exact and on time fulfillment of all the obligations (incl. repayment of loans, interests, cheques, etc.); absence of or minimal reclamations on the supplied goods and made services; timeliness, integrity and the credibility of the financials; or the sufficient current account balance in times of no loan application.

It is important to maintain that all characters mentioned above are just relative and thus they might be incomplete and distorted. Anyone can imagine the purposeful willingness of the company to build up a good name in order to gain the credit. The behavior after the loan settlement can never be predicted with certainty. In spite of that, the foreign banking considers the credit applicant's credibility as a meaningful character when the loan is negotiated.

The economic and financial situation of the credit applicant

This part of the check of the creditworthiness of the debtor is obviously the most important. For the purpose of quick and reliable evaluation, every bank has a certain statistic model computing the internal rating of the potential client based on the available data (credit application, own experience, etc.). Based on the internal rating, the bank then decide whether to provide a credit, and if so, under what conditions. The general computation and its interpretation has already been mentioned in Subsection 3.6.2. (Polidar 1999, pg. 217-225) summarizes areas that are evaluated. In fact, these are also some of the data needed to compute the internal rating:

- The economic standing of the client divided into the following areas
  - evaluation of the firm's business standing (e.g. firm's growth potential, the market position, the market share, the price level compared to the competition and also the legal form of the company),
  - evaluation of the firm's line of business (whether the client is running its business that is more prone to be influenced by the economic crisis, e.g. the automotive industry), and
  - country evaluation (e.g. politically instable and indebted country with a high inflation rate),
- the financial standing of the client divided into the following areas
  - rentability analysis (e.g. ROA, ROE, ROS, etc.),

- cash-flow analysis
- credit indebtness analysis and capital structure,
- liquidity analysis, and
- evaluation of the business plan.

The data source are then usually the last available financials (balance sheet, profit-loss statement, cash-flow statement and changes in equity, if available), the last company's economic reports, foundation charter and articles, the list of the property, the information about the situation of the orders, the investment and financial plan of the company, the planned property and capital balance for the next years, the planned creation of the financial sources, special business market research and other source materials as necessary[14].

The credit limitation

After the bank assigns the client's rating, it is then divided into several categories. Generally, the groups are three, the credit eligible ones (clients with the best rating and thus above-average firms with the lowest or none credit risk), the problematic ones (with an average rating that will probably get the loan, but with further limits), and the credit ineligible ones (clients with the worst rating and thus below-average firms with the highest credit risk). There will be probably no credit granted for the credit ineligible group, while the credit eligible group will gain the loan with no other limitations.

There are several possibilities how to limit the loan for the middle group, i.e. the problematic clients. The common tool is to limit the conditions of the loan. They are mainly the collateral limit (i.e. the minimal amount of the collateral - a share of the market price equivalent - needed to grant a loan) and the maturity limit (i.e. the maximal maturity of the loan acceptable by the bank for the given risk client). Another way is to limit the maximal amount of the loan that is the bank willing to offer to the client with respect to the given risks.

The construction of the bank's internal rating is strictly confident and any bank will never unclassify the way of computation, neither the incoming variables or their exact share in the equation. Polidar (1999) mentions the general way of the client's global rating construction as the sum of the numeric valuation of the particular risk factors, or expressed mathematically:

[14] (Polidar 1999, pg. 224)

$$X = x_1 + x_2 + x_3 + ... + x_n \tag{3.1}$$

where X = coefficient of the total credit risk that is decisive for the risk group classification and $x_{1,2,3...n}$ = risk factors.

This conception of the risk classification is in fact just another way how to express the probability of default. It means that these risk factors and the final coefficient mentioned above are very similar to the explanatory variables of the model that will be described in Section 4.5. The connection of the model to the risk classification is then obvious. If the model estimates that the client will default (with 95% probability), then it should be worth it to reconsider the risk classification and further business cooperation with the client. As the statistic model is not able to give 100% certainty that the default will really happen, the reconsideration should happen in the way of the extraordinary check of the client. The bank should do that in order to minimize potential loss and to ensure its invested money. This idea is then the basic mission of this thesis.

The banking control

Here comes the real sense of the credit monitoring process as being described in Section 3.5. It is insufficient to check only the credit applicant's economic eligibility and to limit the loan, because there is always the chance that the bank will have to write off the credit claim or at least get the repayments after the maturity. These two steps banks have to do in order to decide for or against the business. In case of positive opinion, it is necessary to continuously monitor/check the client in order to figure out the ability to meet its obligation. The process of these checks is more or less the same as the entrance one and the aim of the checks is to figure out:

- whether the debtor's economic and financial situation is not deteriorating,
- whether the collateral value is not decreasing (by the client's will)
- whether the loan is being used for the agreed purpose[15].

Again, the Early Warning Model should help to assign which client should be immediately checked to answer the questions above.

[15] (Polidar 1999, pg. 227)

3.7.2 Author's point of view

Navrátil (1994) is the literature originated in Audit company Pricewaterhouse-Coopers and seems a little bit old-fashioned. In reality, something still holds, but there often exist more effective tools in the monitoring process. On contrary, Polidar (1999) describes and maintains the most important parts of credit ensuring and when compared to the financial world of 2010, it will be difficult to find out some differences. When considering the loan limitation, according to author's own experience, the most common is the limit concerning the total cost of risk associated with the company or the economic group. This type of limit is more useful when considering the cost of review with respect to the expected loss (representing the cost of credit risk). The expected loss then determines which client should be reviewed more regularly and which not. The expected loss is in these days represented by the internal rating. In fact, due to Basel II convention, the expected Total losses (TL) is computed as follows:

$$TL = Exposure \times \overbrace{Probability\_of\_default}^{Rating} \times Recovery\_factor \tag{3.2}$$

The probability of default is commonly represented by the rating. The rating itself is commonly known and very often used by financial institutions to determine the credit risk of the client. In our case (concerning the criterion of frequency of working out the reviews) it is one of the best tools how to minimize the cost of reviews and maximize the count of detected companies that bring to the bank the biggest potential cost of risk.

With regards to the Basel II convention and the LGD, every bank is closely monitoring the changes in the client's PD in order to improve bank's position in case of client's inability to pay. There are two possibilities of the improvement in case of the worsening of client's PD:

1. decrease (or partial complication) of credit drawing (reduction, regulation)= decrease of LGD by the decrease of Exposure, and
2. improvement of collateral or at least the collateral check = decrease of recovery factor.

3.8 Credit collateral

The credit collateral issues are very important for the bank, but also for the behavior of the client. The following paragraph will concern only the bank's point of view[16]. There are many studies that concern the role of collateral in the credit contract. E.g. Jiménez & Suarina (2003) found that "collateralised loans have a higher probability of default, loans granted by savings banks are riskier and, finally, that a close bank–borrower relationship increases the willingness to take more risk". Anyway, we cannot forget that the primary purpose of collateral is to prevent the bank from the loss in the way that the collateral creates a possibility to stake the claim for the credit and interest claims. There exist many possibilities how to "collateralise" the loan. Polidar (1999) mentions two subgroups of credit collateral, the personal (e.g. guarantees, bills or admission of obligation) and the real collateral (deposit, transfer of ownership, assignment of claims or mortgage).

3.9 Monitoring and commitment in bank lending behavior

Monitoring and commitment in bank lending behavior is also a subject of study of several papers. Blavy (2005) shows that "within a framework of asymmetric information between lenders and borrowers and under costly termination of lending arrangements, commitment may explain the accumulation of non-performing loans". In the Chapter 5 we will try to find out the statistical correlation between the commitment (expressed in the total amount) and default of the client.

3.10 Chapter summary

In this place, we have enough information about economic side of the model. The reader is capable to imagine what a bank officer has to do, when his/her client has been detected as the risk one.

Monitoring of credit risk is irreplaceable activity to prevent bank from a loss. The credit risk management includes several instruments for the elimination of

[16] We will left the client's point of view behind with addressing the adequate literature, e.g. Arnoud *et al.* (1994) for the problematic of moral hazard or Murray & Maksimovic (2005) for the problematic of adverse selection.

the risk situation. The credit limitation and the credit control of the debtor are some of them. The client needs to be also checked for the creditworthiness, which is a part of a regular credit review. The overall debtor's situation or the use and the value of the collateral are other areas to be checked. All of these help the given credit officer to decide for a new credit strategy (increase or decrease of commitment, hold, or exit) and set new risk classification to reflect the real risk.

We will conclude this chapter with the statement that everything matters (within the relationship between the lender and the debtor). This forwards us to the next chapter that will deal with the statistical background for Chapter 5: Early Warning Model.

Chapter 4

Theoretical background

4.1 Scoring model

All financial institutions face up a dilemma when they are lending money. On the one side, there is a wish to maximize its profit by granting a loan to almost every firm that exhibits an interest. Let's call it business wish. On the other side, with each granted loan rises the bank's risk that the firm will not repay the debt. Let's call it credit risk. Every financial institution then has to set its own balance between the business wish and credit risk, so that she is capable of making the right decision (accept vs. reject) when there is a new potential client asking for a loan. For this purpose each financial institution develops its function that estimates the probability of default of each new client. Whereas the explanatory variables like the size of the firm or net profit are continuous, the explanatory variable takes only two values, default or not. Given function then assigns to every loan applicant given final score, where, based on the the score value, the financial institution decides whether to go for the business or not. The choice of the function is up to every institution and the function differs in the explanatory variables. Unfortunately, the most popular (because of its simplicity) linear regression based on the ordinary least squares (OLS) method is inappropriate.

The OLS function cannot be used because the variance of the residuals is inconsistent so that the estimate is heteroscedastic (Baltagi 2002, pg. 331)[1]. The logistic regression is then the logic solution of the problem.

[1]More information about this problem will appear in Section 4.3.

4.2 Credit scoring models

Credit scoring model is a bank's common tool to calculate the probability of default, or to sort the clients into different risk category. The calculation is based on the data on observed debtor's characteristics. Based on the final score, the bank's officer (the risk manager), with respect to known economic and financial characteristics, should be able to:

- "Numerically establish which factors are important in explaining default risk.
- Evaluate the relative degree or importance of these factors.
- Improve the pricing of default risk.
- Screen high-risk loan applicants.
- Calculate any reserves needed to meet expected future loan losses."[2]

4.2.1 Altman's Z-Score

Altman (1968) developed the well known Z-score model. The final Z is then a measure of debtor's default risk classification. The given Z is described by the debtor's characteristics and chosen financial ratios. The weighted variables then gives the Z value.

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5{}^3. \tag{4.1}$$

The coefficients were derived based on data from manufacturing firms in the United States and thus cannot be applicable world-wide, or even branch-wide. The Equation 4.1 is then derived from the discriminant analysis model (for more information about this type of analysis, see Altman (1968), Lachenbruch (1979), or simply Wikipedia (2010)). The effect of Z-score is: "The higher the value of Z is, the lower the borrower's default risk classification."[4]. citeCornett1999 names four general problems that appear with using Altman's model:

[2] (Cornett & Saunders 1999, pg. 225).

[3]The coefficients X_{1-5} represent working capital/total assets ratio, retained earnings/total assets ratio, EBIT/total assets ratio, market value of equity/book value of long-term debt ratio and sales/total assets ratio respectively.For more, see (Altman 1985, pp. 473-555)

[4] Altman (1985)

1. the model discriminates only between default and no default, while in the real world, there are many stages of a firm's default (e.g. from late payment to insolvency or account execution, etc.),

2. estimated coefficients, and this holds for any scoring model, are not constant over time and the model needs to be regularly re-calibrated in order to achieve significant/confidential estimates,

3. this model, and most of other scoring models, does not include important factors that are hardly to quantify: e.g. the reputation of the debtor, the relationship between the lender and the debtor or the phase of the business cycle,

4. the model data do not have to be complete (not every bank is willing to offer the data associated with the client's default).

Similarly to Altman's Z-score, the Early Warning model should sort the firm into different risk classification based on the value that reflects in some way the probability of default. On contrary, Altman's Z-score uses only the publicly known data, while the EWM uses, in addition, the behavioral data and the data from non-public sources (e.g. the Central Registr of Loans). This turns us to the biggest difference between scoring models and the EWM. The EWM estimates probability of default after the loan has been granted, while the scoring models generally are supposed to assign whether to provide the loan or not. So the usage of the models is different. Scoring models are used by risk officers (e.g. underwriters), while the EWM should be used by the monitoring department to take care of already granted loans.

4.2.2 Sample selection problem in credit-scoring models

Due to Section 3.7, the EWM model should assign, which clients needs extra check. In other words, the EWM model gives to the firm the score on the *ex post* basis (based on its previous behavior), while the scoring models gives to the firm the score on the *ex ante* basis (based on the data from already grated loans. Hence, a firm asking for the loan is successful or not by the model that was developed on the firms that has succeeded. It means that the model are being developed on the subsample of all firms asking for the loan. In other words, the scoring models are developed on different sample (only the successful firms) than it is then applied (all firms). The final score then might *de facto* produce bias predictions. There are several articles that deal with this problem and name it as a sample selection problem in credit-scoring models (e.g. Green (1998)). The most important for this thesis is the fact that the EWM model will not be influenced by the sample selection as it is developed on the same sample as it is then applied.

4.2.3 Recent advances in credit risk modeling

When considering the risk models as general, the loan risk is not the only risk that is nowadays being modeled in the financial markets. Capuano *et al.* (2009) divides the recent credit risk models into structural and reduced-form models. "As is well known, two main approaches are in use for modeling the default risk of a single issuer: the intensity-based or reduced-form and structural approaches."[5]. It says that reduced-form approach assumes that the timing of a default depends on the exogenous stochastic process and that the default is not connected to the observable characteristics of the firm. On the other hand, the structural approach[6] connects the default with the inability to service its debt. The Early Warning Model can be classified in this way as the structural approach as it estimates the default based on the previous behavior (the bad behavior very often signals inability to service the debt). More about the latest approaches in the credit risk modeling in Capuano *et al.* (2009).

4.3 Appropriate estimator for the regression

Proposition 4.1 (OLS inconsistency). *When estimating binary dependent variable, the use of OLS method leads to inconsistent estimate.*

Proof. Let's have a function of binary dependent variable denoted by y taking the value 0 and 1. The regression is then applied on the Equation 4.2.

$$y_i = x_i^T\boldsymbol{\beta} + u_i \tag{4.2}$$

When the y_i is 1 (default of the client i) the residual will be then $u_i = 1 - x_i^T\boldsymbol{\beta}$. Let's denote probability of this situation as Π_i. If y_i is 0 (the client i does not have any problem with payment) the residual then will be $u_i = -x_i^T\boldsymbol{\beta}$.

[5] (Capuano *et al.* 2009, pg.4)
[6] Black & Sccoles (1973) and Merton (1974) were first to mention this approach.

Let's denote probability of this situation as $1 - \Pi_i$. The mean of residual must be equal to zero (see Equation 4.3).

$$E(u_i) = 0 \tag{4.3}$$

$$\Pi_i(1 - x_i^T\boldsymbol{\beta}) + (1 - \Pi_i)(-x_i^T\boldsymbol{\beta}) = 0 \tag{4.4}$$

$$\Pi_i = x_i^T\boldsymbol{\beta} \tag{4.5}$$

When substituting the Equation 4.5 into the Equation 4.6, one get that $Var(u_i)$ is not constant and thus heteroscedastic.

$$Var(u_i) = \Pi_i(1 - \Pi_i) \tag{4.6}$$

$$Var(u_i) = x_i^T\beta(1 - x_i^T\boldsymbol{\beta}) \tag{4.7}$$

Equation 4.7 implies that OLS estimate is not consistent (Baltagi 2002, pg. 332).

□

4.4 Logistic regression

The logistic regression is a special form of a regression that predicts a binary variable. So we have a dichotomic dependent variable and we try to analyze the effect of the finite number of explanatory variables that can be both, numeric and categorical type. The following paragraph shows how it works.

We do not estimate the probability of 0 or 1 but the probability of the chance of "success" (or event[7]). If P is the probability of the event, the (1-P) is the probability that the event will not happen. The chance of the event is then $P/(1 - P)$. The chance of event is called the Odds ratio.

Simultaneous effect of explanatory variables to the chance is then

$$Odds = \frac{P}{1 - P} = e^{\alpha + \sum_{i=1}^{p} \beta_i X_i} \tag{4.8}$$

If we take the logarithm of both sides of the Equation 4.8, we will get

$$\log \frac{P}{1 - P} = \log e^{\alpha + \sum_{i=1}^{p} \beta_i X_i} \tag{4.9}$$

that get us to the final Equation 4.10.

[7] In our case it is the default of the client

$$LogitP = \alpha + \sum_{i=1}^{p} \beta_i X_i \tag{4.10}$$

Coefficients $\beta_1, \beta_2, ..., \beta_p$ are then minimizing the function of maximal likelihood (more about likelihood function in e.g. Anděl (2007)). Cipra (2008) denotes Logit in terms of probabilities:

$$P(y_t = 1|x_{t\cdot}, \beta) = 1 - F(-x_{t\cdot}\beta) = 1 - \frac{e^{-x_{t\cdot}\beta}}{1 + e^{-x_{t\cdot}\beta}} = \frac{e^{x_{t\cdot}\beta}}{1 + e^{x_{t\cdot}\beta}} \tag{4.11}$$

which is based on the distribution function of the logistic distribution (on contrary to Probit that is based on normal distribution). It is worth to mention that the probability density of logistic distribution is $f(x) = e^x/(1 + e^x)^2$.

4.4.1 Discrimination ability of the model

One of the most important characteristics of the model is its ability to discriminate between good and bad clients. Ideally, there would be one exact scoring level, over which the clients will be classified as the bad ones, and the good ones *vice versa*. In reality, it is very difficult (if not impossible) to denote this level. The client classified as good can have problems with payment and evidently good client can be classified as the bad one.

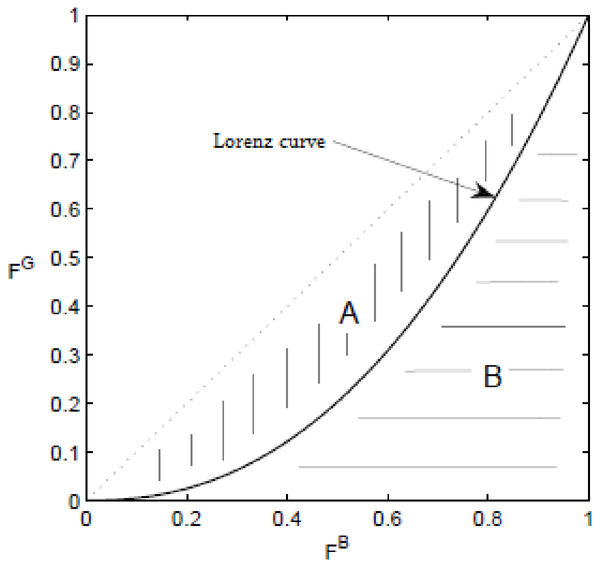
Discrimination methods

One of the most popular method to graphically illustrate the discrimination ability is the **Lorenz curve** (see e.g. Rychnovský & Charamza (2008)). This curve is well-known in economics, because it is used to illustrate the inequality of the income distribution. Closely associated with the Lorenz curve is the **Gini coefficient**. It is usually defined as the area between the Lorenz curve and a diagonal of the unit square compared to the whole area under the diagonal (see Figure 4.1 – area A/(A+B))[8]. As the area (A+B) is the half of the unit square, then the Gini coefficient equals just to the area of 2A. In ideal state of world, the Lorenz curve will copy the legs of the right triangle and the Gini coefficient will be equal to 1.

In our model we will use (instead of Lorenz curve) the **Receiver Operating**

[8] F^G stands for the probability function that the randomly chosen good client will have score less then the scoring level, while F^B stands for the probability function that the randomly chosen bad client will have score less then the scoring level.

Figure 4.1: Lorenz curve and Gini coefficient



Source: Rychnovský & Charamza (2008).

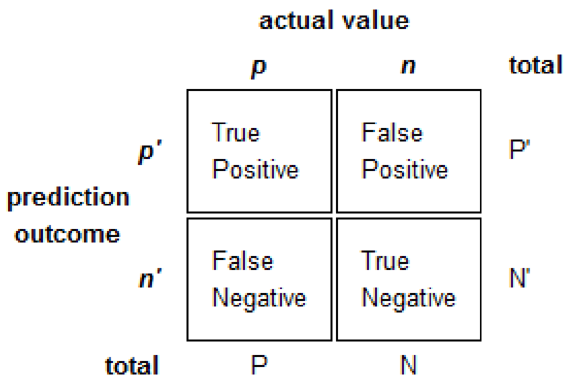
Characteristics (ROC) curve[9], which is a graphical plot of the sensitivity (or true positives) vs. (1-specificity)(or false positives - see Figure 4.2)[10]. The final shape of the ROC curve will be shown within the results of EWM (see e.g. Figure 5.5). We will not use Gini coefficient itself, instead it will be used the **Somers' D statistics**. It has been already proved that the Gini coefficient can be estimated through the Somers' D statistic[11]. Nevertheless, our goal will be to find the model with the Somers' D statistic close to 1 and ROC curve close to the legs of the triangle.

4.4.2 Other Fit statistics

Akaike's Information Criterion (AIC) and the **Schwarz Criterion** (SC) are methods that compare alternative specifications by adjusting Estimated sum of squares (ESS) for the sample size (n) and the number of coefficients in the model (K). These criteria can be used to decide if the improved fit caused by an additional variable is worth the decreased degrees of freedom and increased

[9] The area behind the ROC curve is denoted by the c statistics and in fact it is very similar to the Somers' D statistics.
[10] P for positive instances and N for negative instances.
[11] Rychnovský & Charamza (2008)

Figure 4.2: ROC curve: Confusion matrix



Source: Hosmer & Lemeshow (1989).

complexity caused by the addition[12]. The statistical expression is as follows:

$$AIC = \log(ESS/n) + 2(K)/n \tag{4.12}$$

$$SC = \log(ESS/n) + \log(n)(K)/n \tag{4.13}$$

4.5 Description of the variables used in the model

This section provides basic information about probable variables that are thought to be part of the probability model.

4.5.1 Dependent variable

The purpose of this model is to detect potential risk client. But how to specify a risk client? There are many possibilities how to define this type of the debtor, some of them are less severe, some of them are more. For the purpose of this thesis I have chosen the following definition of the "bad" client.

Definition 4.1 (Bad client). The debtor is assumed to be a bad one if the financial institution is forced to create a loan provision within the following 3 months (i.e. client is risky). This definition includes also the situation when the loan is classified as watch or worse and within 3 months changed back to standard.

Our sample consists of 1908 clients with the data available to the end of September 2009. The dependent variable is denoted as Y200909 and take the

[12] Hofler (2001)

value 1 whenever it fulfills the Definition 4.1 and take the value 0 when it fulfills the Definition 4.2.

Definition 4.2 (Good client). The debtor is assumed to be a good one if the financial institution do not have to create a loan provision within the following 3 months (i.e. client is classified as standard all the 3 months).

Definition 4.3 (Bad rate). The bad rate corresponds to the rate of bad clients to the all clients in the given sample.

From this point of view, our sample consists of 82 bad cases and 1826 good ones. This implies the bad rate (see the Definition 4.3) of the whole sample equals to 4.3 %. There is a empiric condition that if anyone wants to create a robust logistic model, she needs to have the bad rate of at least 5 %. As our bad rate is closed to this level, we will assume that this condition has been fulfiled. In this phase, one can admit that the bad rate is very low in comparison with the NPL mentioned in the first chapter. The reason is obvious. As the NPLs count the amount of problematic loan to all loans granted, our bad rate counts amount of problematic clients to all clients in the portfolio. Hence, our bad rate does not take into account the amount of the loan granted to the single client[13]. However, this should not be the limiting factor when trying to estimate the default of the single borrower.

In the following paragraphs, the variables that are going to be used in the model to describe the dependent variable, will be introduced . We can divide them into several subgroups (e.g. qualitative or quantitative). For better outline we will label them corresponding to the data source as the base facts, the behavioral ones and those given from financial statements.

4.5.2 Explanatory variables: Basic loan facts

When a company asks for a loan, the detailed loan application has to be filled in. There are not only basic information about given company (e.g. company name or line of business), but also more "confidential" data like equity structure, other creditors position or list of suppliers and main customers. Of course, there is also the information about loan itself. The overall amount the client

[13]Imagine the situation of default of the strategic client with almost e.g. one tenth of bank's exposition, where bank has 100 clients. Then the NPL then will be 10 %, while the bad rate will be only 1 %

wants to borrow, collateral value or the purpose of use of money. The final data about the loan then gives the loan appraisal.

- *log(Total commitment)*: The logarithm of Total amount of money that the bank is reserving for a client (i.e. all types of loan, account overdrafts and also guarantees).
- *log(Onbalance)*: The logarithm of the amount of money currently used (drawn) by a client (i.e. current account overdraft drawing, loans and guarantees, if used)[14].
- *Pomer*: The rate of onbalance to Total commitment shows the amount of money currently used by a client relatively to the total amount of what a client could use. The rate then has to be between zero (no use of money) and one (full draw), rarely bigger than one in case that the client has drawn the loan down and is overdue with payments, so that the sanction fee arise (and is not yet classified as watch or worse).
- *Segment*: The most of the financial institutions divide its corporate clients into four segments: Small Business (SMB), Middle Corporates (MID), Large Corporates (LRG) and finally Real Estate Companies (REE). These segments will create dummy variable.

4.5.3 Behavioral Explanatory variables

All financial institutions take care about their client portfolio in order to detect possible non-payers. They simply watch the client's behavior and try to push the client to behave like the banks wish. They have several tools to do that. Watching credit account turnovers or current account debits are some of them. Here is the list of variables that are supposed to be a part of the model. They can be divided into several categories:

1. Overdue

    - *count1*: A count of overdue days within past one month, i.e. how many days have been the client debit on his accounts in the past 30 days. It is supposed that the variable will be insignificant but from that implying dummy variable (overdue or not) should be significant. Both possibilities will be discussed later.

[14]If the client is not overdue then the onbalance has to be smaller or equal to total commitment

- *log(avgOverdue1)*: The logarithm of an average amount of overdue within pas one month, i.e. a sum of overdue/debit amount over the past month divided by a variable *count1*.
- *count3*: A count of overdue days within past three months, i.e. how many days have been the client debit on his accounts in the past 90 days. It is supposed that the variable will be insignificant but from that implying dummy variable (overdue or not) should be significant. Both possibilities will be discussed later.
- *log(avgOverdue3)*: The logarithm of an average amount of overdue within pas three months, i.e. a sum of overdue/debit amount over the past three months divided by a variable *count3*.

2. Central registr of Loans (CRU)

    - *CRUcreditors*: A number of all client's bank creditors.
    - *DaysS*: The maximum of a count of overdue days in the whole banking sector. The number itself is based on the central register of loans (CRU, from the Czech Centralni registr uveru), where every financial institution is obliged to deliver actual informations about its debtors on a month basis. It is supposed that the variable will be insignificant but from that implying dummy variable (overdue or not) should be significant. Both possibilities will be discussed later.
    - *log(CRUoverdue)*: The logarithm of a sum of overdue amount in the whole banking sector out of the domestic bank.
    - *log(CRUbalance)*: The logarithm of the total amount that has the client at disposal from all financial creditors.

3. Credit Account Turnovers

    - *log(YoYturn)*: The rate of the credit account turnovers on the year basis calculated as a month average from the past three months (e.g. a month average in the period of December 08 to February 09 compared to the same period of a year before).
    - *log(TurnoverTH)*: A turnover threshold representing the amount that a client has to monthly pay in order to meet its financial obligation (long[15] as well as short[16] "money").

[15] All type of bank loans that have its maturity over one year, mostly the investment loan.
[16] All financial products that have its maturity within one year, mostly account overdrafts.

- *3mAvg_Turnovers*: The 3 months average of credit account turnovers, i.e. the sum of all income credit payments during the last 3 months, divided by three.
- *Overdraft_level*: The threshold of overdraft. This typical banking product is usual for the operational financing. Of course, not every corporate client has overdraft, the missing values for other type of financing are substituted with the median.
- *Turnover_months*: This number is based on *3mAvg_Turnovers* and *Overdraft_level*, and denotes the time of how long will it take to fully repay (with the incoming credit payments) the overdraft credit (e.g. number 5 denotes that the client will probably have fully repay the overdraft within next 5 months). As the overdraft is an operational type of financing (and thus short-term), this number should not be larger than 6 months in case of standard trouble free client.

4. Others

- *ExpiredRD*: The number represents the count of day for how long is the loan review expired. E.g. if the review is set to be done before 31.12.2009 and is not yet done when the model is estimated (e.g. to 28.2.2010). Then the *ExpiredRD* will be 59. If positive, with increasing *ExpiredRD*, the company's probability of default should increase.
- *CZ-NACE*: The combination of 6 numbers represent the Classification of economic activities, where the first two figures go for the particular branch. As mentioned in Section 2.2, this categorical variable is supposed to significantly influent the final probability of default. Of course, the variable has to be transformed as the single variable has no logic meaning and, what is more important, no linear relationship with the response variable. This turns us to further analysis of this variable in order to create several dummy variables.
- *Rating class*: Value of Rating class lies between 1 for the best to 10 for the worst. This variable is quite controversial. As the employee of the bank, it is not allowed to reveal the backgroung of computation or anything else. The *Rating* itself represents the probability of default (see Figure A.1) based on the data gained during the credit appraisal or review. The EWM should instead include behavioral

variables and that is the difference between the rating and EWM. As the rating itself do not take the numeric value (e.g. 5+), the variable *poradie* is employed (it assigns the exact numeric value to the given rating).

4.5.4 Explanatory variables given from the financial statements

Following list of variables is very often meant to be the strongest and irreplaceable variables that cannot miss any scoring model. But the contrary might be the case. Many of the beginning statisticians are often excited to involve financial ratios or other variables from financial statements into their model. If the quality of the data is very solid, there would be probably no problem and estimated coefficients will be significant. But in the reality, there is a problem of very poor quality of the data. Companies do not deliver their financials at the same time. This causes that the model could have been used until the time of delivering of all financials which can be even e.g. half a year after the end of accounting period. And accounting period itself brings about other quality decrease. Not every company has the end of this period set to December 31^{th}. Hence, if we suppose to use the intended model regularly, e.g. on the monthly basis, the variables given from financial statements should have appropriate quality and be well-prepared for further analysis (i.e. filling in of the missing data by the mean or the median, etc.).

As the simple absolute number in the financials does often say anything about the solvency of the client (the only exception might the net profit), in the estimated model will appear financial ratios, that are calculated either on the basis of one financial statement (e.g. 2009) or on the basis of year-to-year analysis of the absolute number of financial ratios.

All of the variables are chosen from the bank's point of view, which means that are important for assess the current customer's ability to meet its obligations. They can be divided into several categories due to their computation are due to Blaha & Jindřichovská (2006) as follows:

1. Absolute figure

    - *Net profit* as the only absolute number. If positive, with increasing net profit, the company's probability of default should decrease.

2. Ratios given from the single financial statements

- *Indebtedness*
$$\frac{Net\ Financial\ Debt}{EBITDA} \tag{4.14}$$
as the ratio showing the overall indebtedness of the client. If positive, with increasing indebtedness, the company's probability of default should also increase.

- *Interest coverage*
$$\frac{EBIT}{Interest\ Paid} \tag{4.15}$$
as the second ratio concerning debt. It shows, whether the generated profit is big enough to pay at least the interest. If positive, with increasing Interest coverage, the company's probability of default should decrease.

3. YoY figures and ratios

- *Change in equity ratio*
$$\frac{Equity - Intangible\ Assets}{Total\ Assets - Intangible\ Assets} \tag{4.16}$$
as the ratio concerning equity. If negative, with increasing change in equity ratio, the company's probability of default should increase.

- *Quick liquidity*
$$\frac{Current\ Assets - Inventory}{Total\ Short\ Term\ Liabilities - Advances} \tag{4.17}$$
as the only ratio concerning liquidity situation of the company. If positive, with increasing quick liquidity, the company's probability of default should increase.

- *Change in Account payables payment period*
$$\left(\frac{Short\ and\ long\ term\ trade\ payables}{Cost\ in\ raw\ mater., cons.\ \&\ purch.\ merch. + Serv.\ exp.}\right)\times 365^{17} \tag{4.18}$$
as the ratio showing the company's ability to pay its commitments in days. If positive, with increasing change in account payables payment period, the company's probability of default should increase.

[17] This formula is being used in case of Total Cost format. In case of Cost of Sales format will differ the denominator.

- *Change in Net Sales*

$$\frac{Sales_t - Sales_{t-1}}{Sales_{t-1}} \tag{4.19}$$

as the relative ratio showing the year-to-year development in the revenue side of the company. If negative, with increasing change in net sales, the company's probability of default should increase.

- *Change in EBITDA*

$$\frac{EBITDA_t - EBITDA_{t-1}}{EBITDA_{t-1}} \tag{4.20}$$

as the relative ratio showing the year-to-year development of the operational profit (Earnings before interest and taxes, depreciation and amortization). If negative, with increasing change in EBITDA, the company's probability of default should increase.

4.5.5 Transformation of the variables

Every statistician, when getting a set of draft data, has to analyze all the variables case by case. First of all, the exact definition of the variable that is going to be explained is needed. The number of overall observations then should correspond to the explanatory variable that has the lowest number of observation. If there is missing any observation in the explanatory variable, it is up to the statistician whether to delete surplus response observations or whether to fill the missing value by the mean or the median (or another logic value). As we have the balanced set of data with the same number of observations in all used variables we can proceed to other analysis.

Secondly, the outlying observations in each explanatory variable (the so-called outliers) need to be excluded or substituted by the mean or the median. Otherwise, the estimate of given variable is threatening to be insignificant. The outlier can be caused e.g. by the error in the data set (while inputing the value into the data file). The process of finding the outliers within the tens of explanatory variables can be relatively time demanding, so the possible solution can be to exclude the upper and lower 1 % of all observations in each variable (i.e. to cut-off the first and the hundredth percentil), or again, to compute the median and substitute the outlying values.

Thirdly, every explanatory variable has to be carefully analyzed not to loose any important information. As we are trying to estimate probability of default using the logistic regression, which is the linear type, the relationship between the response and explanatory variable, if it is thought to be significant, has to be also linear. This fact brings to us very important restriction. While the given variable is supposed to explain default, at least partially, the logit output can show something else (the insignificance). However, when we are persuaded that the variable should stay in the model, we should try to transform it somehow to gain the linear relationship. There are several tools how to do that and if it still will not help, it is obvious that the variable is insignificant and thus has to be left from the model. One can apply logarithmic function (mainly for amount variables) or transformation to dummy variable (e.g. overdue or not). However, the most important part is to study the development of the variable concerning the default rate (i.e. the share of bad clients in the given subset), looking for extremes and trying understand the logic of the distribution of bad clients within the explanatory variable. If there is any logic consequence, then try to make the relationship linear.

The relationship between response and explanatory variable is linear, if the count of bad clients is increasing or decrearasing in the explanatory variable.

Table 4.1: The example of linear relationship between a binary response and a continuous explanatory variable ($\mathbf{x}$)

Decile of x	**# of good clients**	**# of bad clients**	**bad rate**
1.	99	1	1 %
2.	96	4	4 %
3.	95	5	5 %
4.	93	7	7 %
5.	92	8	8 %
6.	89	11	11 %
7.	90	10	10 %
8.	87	13	13 %
9.	82	18	18 %
10.	77	23	23 %
Total	900	100	10 %

Source: Own computation.

If it is not possible to make it linear, then treat the distribution carefully to create a several subgroups with significantly different bad rate resulting in new dummy variable. The dummy variable then should be equal to one if the given event happens and zero if not. The example of this transformation might be as follows. The explanatory variable *ExpiredRD* is not significant when applied as a continuous variable. When trying to transform it to gain the linear relationship, we cannot get any reasonable output. This forwards us to treat the distribution concerning the bad rates. In the final chapter it will be shown that choice of one dummy variable should be the best way how to use this variable in the model. Concretely, Dummy *ExpiredRD* will be equal 1 if the review date has been expired more then 6 days and 0 any else. This is because the bad rate for the first subgroup is significantly higher then for the second one[18].

4.6 Multicollinearity problem

It is generally known that highly correlated predictors might cause some problems in regression models or regression type models (e.g. logistic regression). These problems affect the confidence intervals and interpretation of estimated coefficients in the model. The usual solution is to remove one or several collinear variables or to reduce the amount of redundant variables in the data by the factor analysis, or by the principal component analysis. In the world of banking, the collinear variables are more than usual. Financial statements or data from CRU registry are closely connected with the internal banking data like credit account turnovers, the level of drawing, etc. This is why it should have been appropriate to take care about collinearity.

However, the multicollinearity itself does not break any of the OLS assumptions, so that the estimate of OLS parameters remains the best linear unbiased estimator. Achen (1982) says that the multicollinearity might cause the significant decrease of "statistical power" of the tests, because there can remain too little variability for the confident estimate of separate effects of each variable (from the total amount of explained variability of two predictors and independent variable). The implications is that the multicollinearity needs bigger data sample to reach the same level of statistical significance.

Thus, the only problem of multicollinearity is the fact that the coefficient estimates with small standard errors are harder to estimate. This effect, how-

[18]The selected subgroup of 6 days and more has significantly higher bad rate then any other subgroup tested.

ever, can also appear in the model with few observations[19]. Adding, we might loose some independent or unique variation by removing the highly correlated predictor. The important conclusion for the EWM model is then the fact that the multicollinearity appears, but its relevance is not so important for the final model as the whole.

4.7 Chapter summary

The financial institutions develop their function that estimates the probability of default of each new client in order to accept/reject the application for a certain type of credit. This function is called a scoring model. The possible alternative to this scoring model is the well-known Altman's Z-score. The Early warning model does not have to take care about the sample selection problem, as the probability of default is estimated on the same sample as it is developed, in opposite to common credit scoring models.

The OLS is not an appropriate estimator for binary response variable, as the estimate of variance is inconsistent and thus heteroscedastic. The logistic regression with maximum likelihood function is the logic solution. For this case, the analogy of Lorenz curve, the Receiver Operating Curve, can be used for a graphical discrimination ability of the model. Gini coefficient, or the Somer's D statistics respectively, is then the numerical alternative of the discrimination ability. AIC and SIC are another criteria to choose the best fit model.

Our data sample consists of data from different sources. They are the data from credit application, or credit appraisal respectively, including financial statements, internal rating, etc. The CRU registry and credit account flows are other data source. These sources make together many basic variables, from which are many others created (logarithm or dummy variable) to reach the demanded 5% significance level. Most of the variables are chosen according to the author's own intention, only the choice of variables concerning the financial statements are based on adequate literature – they are supposed to cover all main risk areas in the company (debtness, equity, etc.).

All of these variables together create the possible danger of multicollinearity. Adequate literature makes us sure that this fear is not in place, as we are concerned in the quality of the whole model. Multicollinearity influence the

[19] More information about this problem in Achen (1982).

estimate of confidence interval, but do not punish the coefficient estimate itself. To sum up, we have enough information to build the final Early warning model.

Chapter 5

Early Warning Model

5.1 Data recapitulation

The data set consists of 1908 observations, included 82 bad clients. The variables can be sorted due to their source – CRU, current account and loan flows (balance, exposition, overdue), the financial statements and basic information about the client from the credit application. Many of the variables are just derived from the basic ones in order to reach the significance (if needed) – the log or dummy variables are created in that case. Due to the missing observations, the filling in the median is employed (in order not to loose the variable or the observation). For the model selection have been used the random choice of 50 % of the observations (the rest then used for the validation of the model). 71 variables are used to explain the default of the client within the following 3 months – **the end of September 2009** is used as the baseline. As have been mentioned earlier, the default of the client is meant to be whenever the loan is classified as **Watch** or worse.

5.2 Logistic regression

Despite the Linear regression with Ordinary least squares estimator, the logistic regression uses the Maximum likelihood estimator. The reason is obvious, we are estimating limited variable (0 or 1), thus the estimated result has to be also limited (values between 0 and 1). We have already defined the response variable and the estimator has been chosen, the selection of the explanatory variables is the task to be solved. We use all selection methods to make sure that we have not remove any variable that could have any significant information.

Definition 5.1 (Chi-square statistics). Simply defined, the Chi-square statistic measures the difference between the expectation and what will really happen:

$$(\chi^2) = \frac{(Expected - Observed)^2}{Expected} \tag{5.1}$$

where expected value is the estimated one, while the observed value is the real one. If the χ^2 value is big, then the p-value associated with the χ^2 statistics is small. The p-value then represents the probability that the event is happening randomly (that the statistical significance is random). The χ^2 statistics has the χ^2 distribution with *(k-1)* degrees of freedom, where k is a number of explanatory variables in the model.

For the selection of the variables are combined different approaches (selection methods):

- **Forward selection**– testing of the univariate[1] value of the χ^2, while the best predictor is chosen followed again by the computing of the univariate value of the χ^2 of the rest variables with the condition that the first one has been chosen (i.e. a sort of the conditional probability).
- **Stepwise selection** – very similar to the forward selection with the exception that after the variable is added, the multivariate value is computed in order to check whether all the chosen variables are still significant. If not, the insignificant variable is removed from the model.
- **Backward elimination** – starting with all variables and computing the multivariate value in order to remove in each step the least significant variable. The process ends when all variables left are significant on the chosen probability level (5 % by default).
- **Score selection** – creates the models combining the all possible subsets of the variables and choosing that one/ones with the highest score derived from the function of maximal likelihood (χ^2).

5.2.1 Explanatory variables reduction

Hosmer & Lemeshow (1989) recommend to treat all variables case by case, but if there are more then 50 variables, then it is good to reduce them. The

[1]Univariate χ^2 value is considered in the model with the only explanatory variable that is being tested, while the multivariate χ^2 value is computed for the tested explanatory variable with the condition that all other variables are considered.

empirical rule recommends to calculate univariate chi-square value (see Definition 5.1) χ^2 value and let the variables with the p-value lower than 0.5 (because others have very low predicative power). Due to Rud (2001), the stepwise selection has been employed in SAS system with the exception that only one step is permitted, so that the SAS will choose the most significant variable based on the univariate χ^2 value.

5.2.2 Model building

The model selection

The SAS statistic package has been employed.

The stepwise selection procedure creates the list of the rest variables with their univariate chi-square value (and with given p-value). This selection process leaves only variables that might have strong predicative power. The probability level is chosen to be 15 %, either for entry to the model or for stay in the model. In fact, 27 variables have been removed due to the low chi-square value (and high p-value). For the rest 44 (in italics in Figure A.2), the backward elimination and the forward and stepwise selection process is employed (see Figure A.4, Figure A.5 and Figure A.3 in Appendix). The score selection process is applied afterwards to find out the model with the most effective χ^2 value to gain the best combination of the variables. Finally, within the simple logistic procedure, the insignificant variables are removed from the model with regards to p-value, AIC and SC criteria and also ROC curve with Somers' D and c statistics.

The backward elimination has left only 32 variables for further modeling. It is important to notice that 15 % probability level is chosen to let the variable to stay in the model. Finally, the score selection is applied and the model with the best score value is chosen. It is supposed to choose the model, where the marginal score starts to decrease rapidly. This happens between 27 and 28 variables. However, when applied simple logistic procedure, there are to many insignificant variables. The model is not worsening, if we remove the insignificant ones, so that the final model of only 9 variables is chosen. Surely, it is far, far away between 27 and 9 variables, but it is important not the forget that the simplier, the better the model is. So if we are able to leave only significant variables (on the 5 % significance level), then we should get the best model.

5.3 EWM: Results

5.3.1 SAS output and coefficients interpretation

Figure 5.1: Early Warning Model output

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-9.0263	1.4959	36.4101	<.0001
pomer	1	1.8717	0.9077	4.2519	0.0392
D_seg_ree	1	2.4481	0.7558	10.4924	0.0012
D_nace_ree	1	-1.5989	0.7887	4.1099	0.0426
overduecelkem0	1	3.5893	1.3523	7.0446	0.0080
D_overdue3_1800more	1	1.7943	0.4343	17.0651	<.0001
log_turnover_months	1	0.8232	0.1513	29.6183	<.0001
fs_702b	1	-0.0174	0.00584	8.8509	0.0029
rating_class	1	3.1739	0.9648	10.8213	0.0010
poradie	1	-0.8926	0.3101	8.2822	0.0040

Source: SAS statistical package and own computation.

Intercept : If all other dummy variables used in the model are equal to zero, than the estimated coefficient means that the client is not in real estate or machinery branch/segment, has been overdue in monthly average less than 1800 CZK and the change in equity ratio is not worse than -100 %. In this sense, the estimated coefficient is negative as expected, and what is even more important, is the most significant from the all estimated coefficients in the final model.

Pomer = The rate of *onbalance* to *Total commitment* shows the amount of money currently used by a client relatively to the total amount of what a client could use. The estimated coefficient is positive as expected, thus significantly increase the estimated probability of default.

D_seg_ree and D_nace_ree = both dummy variables to define the real estate segment but differ in their source (D_seg_ree from the bank's own definition of segment – REE/LRG/MID/SMB/RET, while the D_nace_ree is from the exact branch code from the CZ_NACE definition). The sum of estimated coefficients together positive as expected, but the D_nace_ree negative in spite of positive expectations. Hence, coefficients together significantly increase the estimated probability of default. When trying to let only one variable in the model, the other became insignificant.

OverduecelkemO = continuous variable to define the overdue amount in mil. (within all creditors). The estimated coefficient positive as expected, thus significantly increase the estimated probability of default.

D_overdue3_1800more = dummy variable to define whether the average amount in overdue for the last 3 months is grater then 1800 CZK or not. The estimated coefficient positive as expected, thus significantly increase the estimated probability of default.

Log_turnover_months = logarithm of continuous variable to define the average amount used divided by the average credit payments within the last 3 months. The estimated coefficient positive as expected, thus significantly increase the estimated probability of default. *The most significant variable in the model.*

fs_702b = continuous variable to define the change in equity ratio. The estimated coefficient negative as expected, thus significantly decrease the estimated probability of default.

Rating_class and poradie = both continuous variables to define the risk category given by the bank's internal rating, where the Rating_class refers to classical rating category (from 1 to 10) and *poradie* refers to the exact rating (from 1 to 26). E.g. when the exact rating is 5+, then the rating_class is 5 and the poradie is 13. The sum of both estimated coefficients together is positive as expected, but the poradie is negative in spite of positive expectations. Hence, coefficients together significantly increase the estimated probability of default. When trying to let only one variable in the model, the other became insignificant.

As the logistic procedure is not estimating the coefficients itself, but the odds ratio, Figure 5.2 shows the confidential limits for the estimated odds ratios.

Figure 5.2: Odds Ratio Estimates

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
pomer	6.500	1.097	38.506
D_seg_ree	11.566	2.630	50.874
D_nace_ree	0.202	0.043	0.948
overduecelkem0	36.209	2.557	512.775
D_overdue3_1800more	6.015	2.568	14.091
log_turnover_months	2.278	1.693	3.064
fs_702b	0.983	0.972	0.994
rating_class	23.901	3.607	158.376
poradie	0.410	0.223	0.752

Source: SAS statistical package and own computation.

Any other variable has reached the chosen significance of 5 %. We might consider the model with 11 variables (see Section A.1 for the SAS output), but these two extra variables (D_EBITDA and D_NACE_Machin) have not passed the 5 % significance level (p-level 0.11 and 0.06 respectively). Furthermore, after removing these variables, the model has not got worse. The Akaike Informaton Criteria has a little bit worsened (slight increase from 206.3 to 208.7), but the Schwarz Criteria has improved (decrese from 264.2 to 256.2, see Figure 5.3 and Figure A.14). Other variables in the model have remained significant, this is why we choose the model with only 9 variables as the final one.

Testing Global Null Hypothesis (BETA=0): all tests (Likelihood ratio, Score and Wald) refuse the null hypothesis of all variables being equal to zero.

When considering the ROC curve and the Somers' D statistics, we might be satisfied with the rate of discrimination power. The ROC curve is very

Figure 5.3: EWM: Statistics

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	330.450	208.070
SC	335.267	256.238
-2 Log L	328.450	188.070

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	140.3797	9	<.0001
Score	158.4398	9	<.0001
Wald	73.6551	9	<.0001

Source: SAS statistical package and own computation.

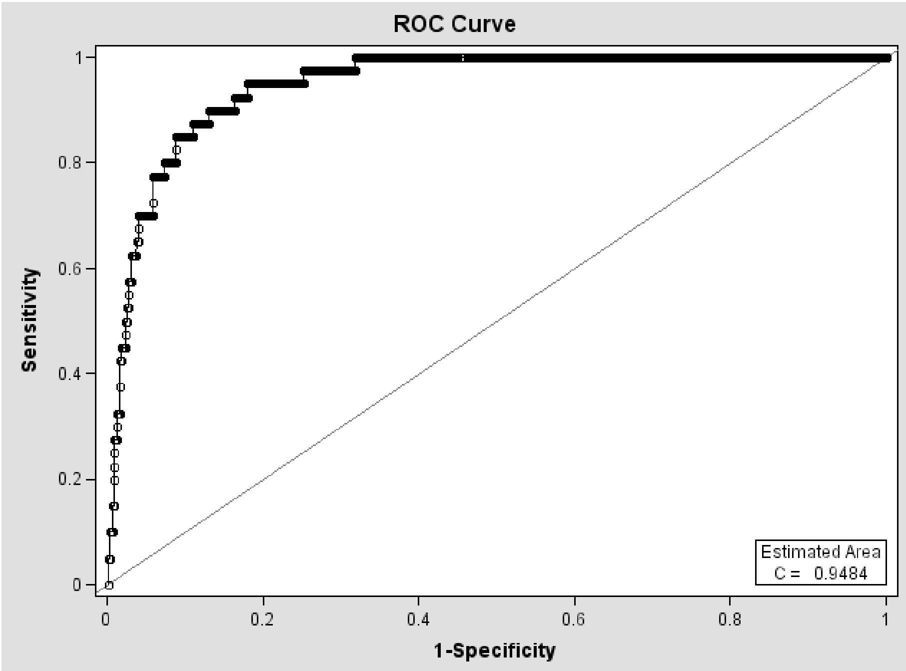
Figure 5.4: Association of Predicted Probabilities and Observation Responses

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	94.7	Somers' D	0.897
Percent Discordant	5.0	Gamma	0.899
Percent Tied	0.2	Tau-a	0.075
Pairs	34920	c	0.948

Source: SAS statistical package and own computation.

close to the legs of the triangle and Sommers' D reaches 0.897 (furthermore, the c statistic equals 0.948). If compared to perfect model (value of 1.00), the discrimination power of the selected model reaches very good result (see Figure 5.5.

Figure 5.5: Receiver Operating Characteristic Curve



Source: SAS statistical package and own computation.

Concretely, if we look behind the curve and study the data of sensitivity and 1-specificity, we would obtain the following result: Curve shows e.g., that if we choose the worst 4.3 % of all observations (70), the subset will contain 80 % of all bad clients. The example chosen is on 17.44% probability level and says that if we select the clients with predicted probability over this level, the subset will contain 70 clients, of which the 32 were predicted correctly as events and 38 nonevents predicted as events. Good result achieved as to the statistic point of view. When considering the economical part, the chosen probability level will mean that 70 client will be marked as bad, of which only 32 will really default. In other words, immediate check of 38 clients will be waste of time and money (administration costs). For this purpose can be chosen grater probability level, e.g. 68 %, where only 10 clients are marked as bad ones with the portion of clients 7 to 3. This choice will certainly be more transparent for other users and more suitable for real business use.

For the purpose of model validation, the decile analysis has been employed.

Figure 5.6: Decile Analysis - origin sample

	Prospects	Predicted probability	Default rate
Decile			
0	91	0.33144	0.34065934
1	91	0.05951	0.06593407
2	91	0.02436	0.02197802
3	92	0.01097	0.01086957
4	91	0.00620	0.00000000
5	91	0.00351	0.00000000
6	92	0.00206	0.00000000
7	91	0.00097	0.00000000
8	91	0.00035	0.00000000
9	92	0.00005	0.00000000
Total	913	0.04381	0.04381161

Source: SAS statistical package and own computation.

The data (used for the model development) is sorted descending by the predicted probability and then divided into ten subgroups, i.e. into deciles. From the statistical point of view, the model seems to be very good, as the worst 50 % of all observations contains all (i.e. 100 %) bad clients. Furthermore, there exists the clear cut-off point, under which we can say that the client will (with probability close to 1) not default within next 3 month. This might be with no doubt used as a efficient credit risk monitoring tool.

As have been mentioned earlier, the rest of random 50 % of the origin observations are used to validate the final model. As can be seen in Figure 5.7, the similar implications can be used for the validation sample and thus we can say that the model seems to be robust! Another task is to ensure the robustness of the model in time, which might be a theme for the rigorous thesis...

The final model contains 9 variables (the 10^th one is the constant). In fact, we can consider only the seven of them as there are two couples that explain almost the same (but have to be included due to the (in)significance of the model). There is a couple of the dummy variables (D_seg_ree and D_nace_ree) representing the sphere of business (credit application's data) and a couple of continuous variables (rating_class and poradie) representing the *ex ante* probability of default based on the financial statements and other data delivered by the credit applicant (internal confidential data). The other sources that are represented in

Figure 5.7: Decile Analysis - validating sample

	Prospects	Predicted probability	Default rate
Decile			
0	99	0.30414	0.26262626
1	99	0.05779	0.08080808
2	100	0.02248	0.03000000
3	99	0.01183	0.01010101
4	100	0.00636	0.02000000
5	99	0.00337	0.02020202
6	100	0.00181	0.00000000
7	99	0.00084	0.00000000
8	100	0.00032	0.00000000
9	100	0.00005	0.00000000
Total	995	0.04072	0.04221106

Source: SAS statistical package and own computation.

the model are the CRU registry (overduecelkemO), the internal data - concerning overdue and turnovers (D_overdue3_1800more and log_turnover_months), and finally the financial statements - that can be meant to be as the part of the credit application's data (fs_702b).

5.4 EWM: A model with all observations included

The model with all observations included has been also considered. However, it is recommended to treat the model on exact subsample and then test the robustness of the model on the rest subsample (let's denote it as the Statistical model), it is worth it to compare the "robust" results with the "all-observations" model (let's denote it as the All model). The Figure 5.8 shows the given results.

In this place, we will comment only the extra variables, because the variables concordant for both models have very similar estimated coefficients (and thus the interpretation).

D_nace_machin = dummy variable to define the machinery segment based on the CZ-NACE code. The estimated coefficient positive as expected.
Estimate: 0.81 with p-value 0.0487

D_daysS = dummy variable (based on the CRU registry) to define whether

Figure 5.8: Early Warning Model output - all

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.3407	0.9744	91.8936	<.0001
pomer	1	1.0304	0.4952	4.3306	0.0374
D_seg_ree	1	2.8103	0.5465	26.4465	<.0001
D_nace_machin	1	0.8060	0.4089	3.8856	0.0487
D_nace_ree	1	-1.3709	0.5591	6.0128	0.0142
D_daysS	1	2.0050	0.4916	16.6383	<.0001
D_overdue3_1800more	1	1.3264	0.2811	22.2688	<.0001
log_turnover_months	1	0.6480	0.1117	33.6729	<.0001
fs_702b	1	-0.00788	0.00386	4.1749	0.0410
D_EBITDA	1	0.8606	0.3112	7.6483	0.0057
rating_class	1	2.6648	0.5926	20.2222	<.0001
poradie	1	-0.6585	0.1947	11.4413	0.0007

Source: SAS statistical package and own computation.

the debtor is overdue or not (exactly whether the period for which has been the debtor overdue, is positive). The estimated coefficient positive as expected.
Estimate: 2.01 with p-value < 0.0001.

D_EBITDA = dummy variable to define whether the year-to-year change[2] in Operational profit (expressed by EBITDA) is less then minus 100 %. The estimated coefficient positive as expected.
Estimate: 0.86 with p-value 0.006.

Compared to the Statistical model, the All model contains 11 variables on the 5% significance level. The extra dummy variables are D_nace_machin, D_daysS and D_EBITDA. On the other hand, the variable overduecelkemO has disappeared. Any other variable has reached the chosen significance level of 5 %.

Figure 5.9: EWM: Statistics - all

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	678.547	438.651
SC	684.101	505.297
-2 Log L	676.547	414.651

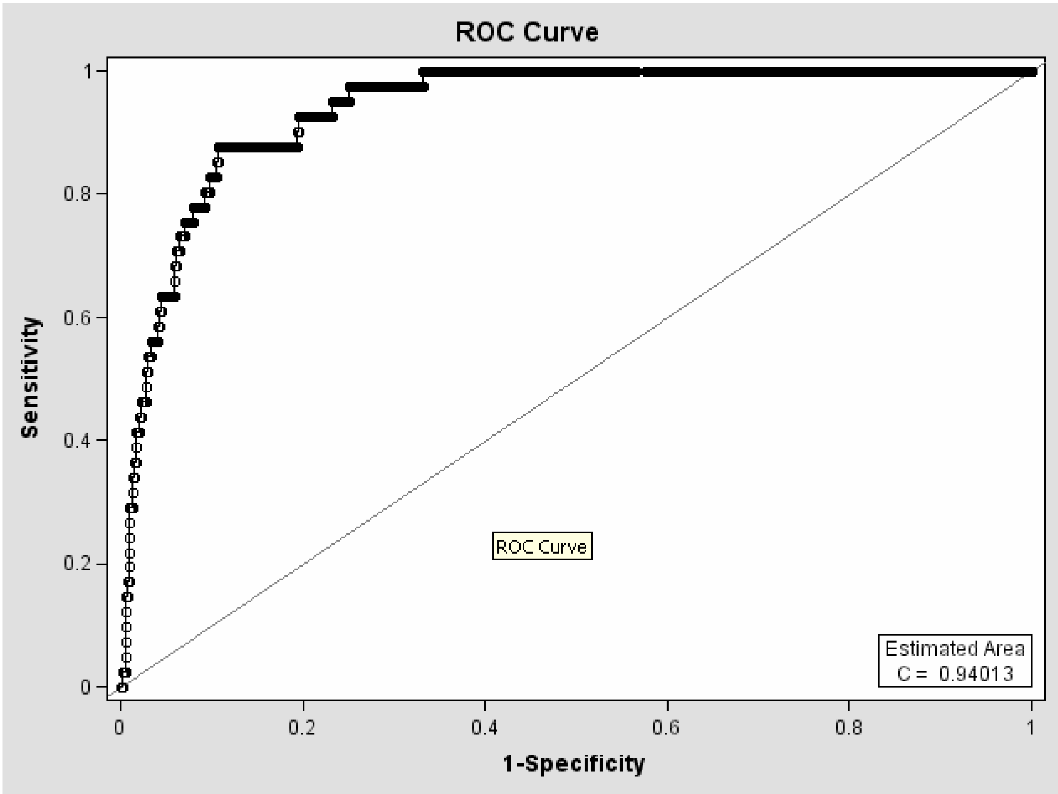
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	261.8963	11	<.0001
Score	316.3356	11	<.0001
Wald	148.9132	11	<.0001

Source: SAS statistical package and own computation.

[2]The case, when the previous EBITDA is negative, while the current one is positive, is also considered.

Comparison of AIC or SC criteria is out of the sense as these criteria compares the same models regarding the number of explanatory variables. When testing the global null hypothesis (all coefficients equal to zero), all three tests (Likelihood ratio, Score and Wald tests) reject the null even on the 0.01% significance level. This confirms us that the explanatory variables have its reasonable place in the model.

Figure 5.10: Receiver Operating Characteristic Curve - All



Source: SAS statistical package and own computation.

The Receiver Operating Characteristic curve is another criterion to be compared. While the estimated area beyond the curve is 0.9401 for the All model, it is slightly higher for the Statistic model (c=0.948).

The Sommer's D that can be approached as a Gini coefficient is again higher for the Statistic model (0.897 compared to 0.880).

Compared to the Statistical model, the All model contains 3 extra variables, but one is missing. The data sources have remained the same, but the composition has slightly changed. The total amount of overdue (overduecelkem0) representing the CRU registry data has been replaced by the dummy variable assigning the duration of overdue (D_daysS) and two variables have joined the

Figure 5.11: Association of Predicted Probabilities and Observation Responses - all

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	93.9	Somers' D	0.880
Percent Discordant	5.9	Gamma	0.882
Percent Tied	0.2	Tau-a	0.072
Pairs	149732	c	0.940

Source: SAS statistical package and own computation.

Figure 5.12: Decile Analysis - all observations

	Prospects	Predicted probability	Default rate
Decile			
0	190	0.30704	0.32105263
1	191	0.06275	0.05759162
2	191	0.02691	0.04188482
3	191	0.01434	0.01047120
4	190	0.00896	0.00000000
5	191	0.00525	0.00000000
6	191	0.00315	0.00000000
7	191	0.00170	0.00000000
8	191	0.00072	0.00000000
9	191	0.00015	0.00000000
Total	1,908	0.04298	0.04297694

Source: SAS statistical package and own computation.

group of data based on credit application (D_nace_machin) and financial statements (D_EBITDA).

5.5 Chapter summary

The data set consists of 71 explanatory variables and 1 response variable in 1908 observations. The bad rate for the whole sample is 4.3 % (baseline in September 2009 with 3 months outline), which is lower than the avarage of NPL ratio[3] in the Czech banking sector to the end of 2009 (see Table 2.2).

Four different selection method (stepwise, backward, forward and score selection processes) are used to reduce the explanatory variables. The final model contains only 9 explanatory variables combining three dummy variables (two for real estate sector and one for 3 months average overdue higher then 1800 CZK), five continuous variables (the rate of amount drawn to total commitment, one for overdue amount within the whole banking sector, one for YoY change in equity ratio and the last two defining the internal rating). There are two doubled variables that explain almost the same information (real estate sector and internal rating), but none of them can be removed from the model, unless we want the model to get worse.

[3] As the NPLs do not contain the Watch loans, the level of our bad rate looks even smaller.

Chapter 6

Conclusion

The aim of this thesis has been to find out, whether there is any difference between the theory and reality in the Monitoring process of a credit risk, and whether exists the reliable logit model estimating the probability of default. I can summarize the results into following answer: "Both tasks connect the same problem - that the financial sector belongs to the one of the most evolving sectors in the economy, even more in the time of the financial crisis. The literature regarding the credit risk is trying to keep the pace with the real banking world, but can never reach the same level. The logit model can be also relatively easy to find, but the robustness in time will be hard to reach. The possible solution should have been a frequent recalibration of the model (let's say on quarterly basis)."

It is not a big surprise that the Czech banking sector has noticed the negative development of NPLs before and during the current financial crisis. Adding, the CNB's stress tests have showed what the prediction of NPL's further development is, negative in short-term horizon, but positive in the long-run. The January 2010's sector-decomposition of NPLs ratio has been mentioned to maintain the possible risk in the automotive and the machinery industry. **Chapter 2** ends with an introduction to loan classification problematics. The client's default generally starts after 90 days in overdue, which corresponds to loan classification non-standard or worse according to CNB and Basel II. Oppositely to non-performing loans (loan classification non-standard, doubtful or loss), the Early warning model tries to reflect all loans with loan loss provision, which enlarges NPLs by loans with Watch classification.

Chapter 3 introduces the risk in the banking sector, especially the credit risk. Monitoring of credit risk is irreplaceable activity to prevent bank from a loss. The credit risk management includes several instruments for the elimination of the risk situation. The credit limitation and the credit control of the debtor are some of them. The client needs to be also checked for the creditworthiness, which is a part of a regular credit review. The overall debtor's situation or the use and the value of the collateral are other areas to be checked. All of these help the given credit officer to decide for a new credit strategy (increase or decrease of commitment, hold, or exit) and set new risk classification to reflect the real risk.

Chapter 4 sets the needed theoretical background for the practical part of this thesis. The financial institutions develop their function that estimates the probability of default of each new client in order to accept/reject the application for a certain type of credit. This function is called a scoring model. The possible alternative to this scoring model is the well-known Altman's Z-score. The Early warning model does not have to take care about the sample selection problem, as the probability of default is estimated on the same sample as it is developed, in opposite to common credit scoring models.

The OLS is not an appropriate estimator for binary response variable, as the estimate of variance is inconsistent and thus heteroscedastic. The logistic regression with maximum likelihood function is the logic solution. For this case, the analogy of Lorenz curve, the Receiver Operating Curve, can be used for a graphical discrimination ability of the model. Gini coefficient, or the Somer's D statistics respectively, is then the numerical alternative of the discrimination ability. AIC and SIC are another criteria to choose the best fit model.

Our data sample consists of data from different sources. They are the data from credit application, or credit appraisal respectively, including financial statements, internal rating, etc. The CRU registry and credit account flows are other data source. These sources make together many basic variables, from which are many others created (logarithm or dummy variable) to reach the demanded 5% significance level. Most of the variables are chosen according to the author's own intention, only the choice of variables concerning the financial statements are based on adequate literature – they are supposed to cover all main risk areas in the company (debtness, equity, etc.).

All of these variables together create the possible danger of multicollinearity. Adequate literature makes us sure that this fear is not in place, as we are concerned in the quality of the whole model. Multicollinearity influence the estimate of confidence interval, but do not punish the coefficient estimate itself.

The final **Chapter 5** describes the model building. The data set consists of

71 explanatory variables and 1 response variable in 1908 observations. The bad rate for the whole sample is 4.3 % (baseline in September 2009 with 3 months outline), which is lower than the avarage of NPL ratio[1] in the Czech banking sector to the end of 2009 (see Table 2.2).

Four different selection method (stepwise, backward, forward and score selection processes) are used to reduce the explanatory variables. The final model contains only 9 explanatory variables combining three dummy variables (two for real estate sector and one for 3 months average overdue higher then 1800 CZK), five continuous variables (the rate of amount drawn to total commitment, one for overdue amount within the whole banking sector, one for YoY change in equity ratio and the last two defining the internal rating) and one continuous variable in logarithmic scale (variable denoting the time of how long will it take to fully repay the overdraft credit with the incoming credit payments). There are two doubled variables that explain almost the same information (real estate sector and internal rating), but none of them can be removed from the model, unless we want the model to get worse.

The main contribution of this work is closely associated with the equation regarding the Loss Given Default, which equals to the bank's Commitment to the client *times* probability of default *times* recovery factor, all *divided* by Exposure at Default. The Early Warning Model is in the first place estimating the probability of default, in the second place should assign which corporate client needs to be checked for the creditworthiness. The credit strategy from the implying credit review then modifies the bank's commitment to the client, so that the final LGD might decrease (or increase). The check of collateral influences the recovery factor, which can again decrease the LGD even more in case of client's default.

The Early Warning Model consists of variables representing all available data sources. Only the variables defining the real estate segment are opposite to expectations. The Table 2.2 shows the prevailing negative development in the automotive and the machinery industries, so that one can expect a given dummy variable in the model. But the opposite is true. These variables have been insignificant on the given probability level. In my opinion, this problem has two reasons, the chosen data set and bank's specialization. As the baseline for the data has been September 2009, there is a chance that only a few corporate clients have defaulted in the automotive or the machinery industry,

[1] As the NPLs do not contain the Watch loans, the level of our bad rate looks even smaller.

while many of them defaulted in the real estate one. Furthermore, there is also a chance, that the data are not 100 % true in this way. For this reason, it will be worth to try to model the data based on another month, or rather year.

The Model has of course its imperfections. The decile analysis has proved that the Model is robust on the given sample, but does not ensure the robustness in time. This is exactly the area of possible thesis extension. One can imagine the rigorous thesis dealing with the Early Warning Model with the time parameter to make the Model robust in time. Another area might be the extension of the explanatory variables. I have not surely covered all the variables that might significantly describe the client's default, so the explanatory variable research might be another way for the extension of this thesis.

The final words are addressed to the usage of this model in reality. I would be really happy if the Early Warning Model would become a part of the monitoring process, with a decrease of Loss Given Default (or let's say costs of risk) by detection of risk clients on the one side, and with decrease of administrative costs by the check of only risk clients on the other side.

Bibliography

Achen, C. (1982): *Interpreting and Using Regression.* Beverly Hills: CA:SAGE.

Ahmed, A. S., C. Takeda, & S. Thomas (1999): "Bank loan loss provisions: a reexamination of capital management, earnings management and signaling effects." *Journal of Accounting and Economics* pp. 1–25.

Altman, E. I. (1968): "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *Journal of Finance* .

Altman, E. I. (1985): "Managing the Commercial Lending Process." *Handbook of Banking Strategy* .

Anděl, J. (2007): *Základy matematické statistiky.* Praha: Matfyzpress.

Arnoud, W., A. Boot, & W. T. Anjan (1994): "Moral Hazard and Secured Lending in an Infinitely Repeated Credit Market Game." *International Economic Review* **35**.

Baltagi, B. H. (2002): *Econometrics, Third edition.* Berlin: Springer.

Benešová P., C. P. (2008): "Rozlišovací schopnosti ruzných skóringových funkcí."

Black, F. & M. Sccoles (1973): "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* **81**.

Blaha, Z. S. & I. Jindřichovská (2006): *Jak posoudit finanční zdraví firmy, 3.rozšířené vydání.* Prague: Management Press.

Blavy, R. (2005): "Monitoring and Commitment in Bank Lending Behavior." *IMF Working Paper, Western Hemisphere Department* .

Capuano, C., J. Chan-Lau, G. Gasha, C. Medeiros, A. Santos, & M. Souto (2009): "Recent Advances in Credit Risk Modeling." *International Monetary Fund Working Paper* **31**.

Cipra, T. (2008): *Finanční Ekonometrie.* Praha: Ekopress.

Cornett, M. & A. Saunders (1999): *Fundamentals of Financial Institutions Management.* United States of America: The McGraw-Hill Companies. Wall Street Journal edition.

Crouby, M., D. Galai, & R. Mark (2000): "A comparative analysis of current credit risk models." *Journal of Banking and Finance* **24**: pp. 59–117.

Dluhošová, D. (2008): *Finanční řízení a rozhodování podniku: Analýza, Investování, Oceňování, Riziko, Flexibilita.* Prague: Ekopress.

Green, W. (1998): "Sample selection in credit-scoring models." *Japan and the World Economy* **10**: pp. 299–316.

Hofler, R. A. (2001): "Akaike's Information Criterion and the Schwarz Criterion." *University of Central Florida* .

Hosmer, D. W. & S. Lemeshow (1989): *Applied Logistic Regression.* New York: John Wiley & Sons.

Jiménez, W. & J. Suarina (2003): "Collateral, type of lender and relationship banking as determinants of credit risk." *Journal of Banking and Finance* **28**.

Kadlčáková, N., K. J. (2004): "Credit risk and bank lending in the Czech Republic." *Czech National Bank* .

Lachenbruch, P. A., G. M. (1979): "Discriminant Analysis." *International Biometric Analysis* .

Mejstřík, M. (2004): *Cultivation of financial markets in the Czech republic.* Prague: The Karolinum Press.

Mejstřík, M. (2005): "Bankovnictví." *Europeum* Lectures IES FSV UK.

Mejstřík, M., M. Pečená, & P. Teplý (2008): *Basic Principles of Banking.* Prague: The Karolinum Press.

Merton, R. C. (1974): "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *The Journal of Finance* **22**: pp. 449–470.

Mužíček, P. & M. Mejstřík (2007): "System hospodareni ceskeho bankovnictvi v porovnani se strední Evropou."

Murray, Z. & V. Maksimovic (2005): "Trade Credit, Collateral, and Adverse Selection." *Journal of Banking and Finance* .

Navrátil, V. (1994): *An Introduction to Credit Risk Management.* Prague: Management Press. Origin translated from: Price Waterhouse.

ČNB (2010): "Česka Národní Banka." *http://www.cnb.cz/cs/index.html* .

Orgler, Y. E. (1970): "A Credit Scoring Model for Commercial Loans." *Journal of Money, Credit and Banking* **2**: pp. 435–445.

Polidar, V. (1991): *Management Úvěrových Obchodů Bank.* Prague: Economia.

Polidar, V. (1992): *Management Úvěrových Obchodů Bank.* Prague: Economia. Druhé vydání.

Polidar, V. (1999): *Management Bank a Bankovních Obchodů.* Prague: Ekopress. Druhé aktualizované a rpzšířené vydání.

Rowan, M., P. Stump, R. Cantor, R. Solomon, & K. Elmery (2006): "Probability of Default Ratings and Loss Given Default Assessments for Non-Financial Speculative-Grade Corporate Obligors in the United States and Canada." *Technical report*, Moody's Investors Service.

Rud, O. P. (2001): *Data Mining.* Prague: Computer Press.

Rychnovský, M. & P. Charamza (2008): "Postupná výstavba modelu ohodnocení kreditního rizika."

Simons, D. J. (2009): "Visual Cognition Lab." *University of Illinois* .

Singer, M. (2009): "Bankovní sektor v době krize z hlediska finanční stability (6.10.2009)." *Technical report*, Czech national bank.

Sinkey, J. F. (1998): *Commercial Bank Management in the Financial Services Industry, 5th ed..* New Yersey: Prentice-Hall, c1988.

WIKIPEDIA (2010): "Linear Discriminant Analysis." *http://en.wikipedia.org/wiki/Linear_discriminant_analysis* .

ZAVADILOVÁ, T. (2010): "Rizikové úvěry v bankách překročily deset procent." *E15: 15 Minutes for the Economics & Business* .

Appendix A

Other models considered

Figure A.1: The relationship between the Moody's rating and the probability of default given by the Expected Loss rate

Idealized Loss Rates and Default Rates Used in Rating Structure Finance Securities

Corporate Family Rating	Four-Year Idealized Expected Loss Rate	Probability of Default Rating	Four-Year Idealized Default Probability (Assumes 50% Average Expected LGD)
Aaa	0.0010%	Aaa	0.0020%
Aa1	0.0116%	Aa1	0.0232%
Aa2	0.0259%	Aa2	0.0518%
Aa3	0.0556%	Aa3	0.1112%
A1	0.1040%	A1	0.2080%
A2	0.1898%	A2	0.3796%
A3	0.2970%	A3	0.5940%
Baa1	0.4565%	Baa1	0.9130%
Baa2	0.6600%	Baa2	1.3200%
Baa3	1.3090%	Baa3	2.6180%
Ba1	2.3100%	Ba1	4.6200%
Ba2	3.7400%	Ba2	7.4800%
Ba3	5.3845%	Ba3	10.7690%
B1	7.6175%	B1	15.2350%
B2	9.9715%	B2	19.9430%
B3	13.2220%	B3	26.4440%
Caa1	17.8634%	Caa1	35.7268%
Caa2	24.1340%	Caa2	48.2680%
Caa3	36.4331%	Caa3	72.8662%
Ca	50.0000%	Ca	100.0000%
C	80.0000%	C	100.0000%

Source: Rowan *et al.* (2006).

EWM: The model with 13 variables and all observations included

A.1 EWM: The model with 50 % random observations and 11 variables

Figure A.2: EWM: 42 variables left after Stepwise selection procedure - univariate Chi-square

Effect	DF	Score Chi-Square	Pr > ChiSq	Effect	DF	Score Chi-Square	Pr > ChiSq
pomer	1	25.3402	<.0001	log_avg_balance	1	3.5942	0.0580
onbalance	1	0.4365	0.5088	D_balance	1	2.6452	0.1039
log_onbalance	1	11.3297	0.0008	avg_kredity	1	0.0936	0.7597
total_exposure	1	0.0155	0.9010	D_kredity	1	8.3155	0.0039
log_total_exposure	1	0.3792	0.5381	log_avg_kredity	1	15.9533	<.0001
D_seg_ret	1	30.6301	<.0001	Limit_C	1	0.2454	0.6203
D_nace_machin	1	5.7135	0.0168	log_Limit_c	1	0.0646	0.7994
D_nace_ret	1	6.6223	0.0101	avg_usage	1	21.1144	<.0001
D_nace_auto	1	1.6151	0.2034	D_usage	1	25.3442	<.0001
creditors	1	0.1743	0.6763	turnover_months	1	67.5715	<.0001
dayss	1	5.9760	0.0145	log_turnover_months	1	29.5539	<.0001
D_days5	1	25.0649	<.0001	D_turnover_months	1	34.1463	<.0001
overduecelkem0	1	30.9997	<.0001	REFERENCE_YEAR	1	0.2051	0.6506
log_overduecelkem0	1	0.0044	0.9472	D_FS	1	0.0819	0.7747
total_balance0	1	0.2571	0.6121	D_2008	1	0.1417	0.7066
log_total_balance0	1	0.0367	0.8480	D_rev	1	1.5657	0.2104
onbalance0	1	0.0352	0.8513	fs_170	1	0.1059	0.7449
log_onbalance0	1	0.2131	0.6443	fs_702b	1	25.6102	<.0001
pomer0	1	1.1552	0.2767	D_EQ	1	20.1552	<.0001
D_pomer0	1	9.7934	0.0015	fs_711	1	0.0379	0.8457
count1	1	21.1214	<.0001	D_QL	1	4.2575	0.0391
D_count1	1	37.7303	<.0001	fs_713	1	5.6765	0.0172
avg_overdue1	1	2.2071	0.1374	D_Debt	1	2.8675	0.0904
log_avg_overdue1	1	38.7909	<.0001	fs_716	1	22.6402	<.0001
count3	1	26.5055	<.0001	D_receivables	1	5.0420	0.0247
D_count3_4more	1	42.6561	<.0001	fs_721	1	0.1463	0.7021
D_count3_1to3	1	10.0901	0.0015	D_Cover	1	1.5467	0.2136
log_avg_overdue3	1	64.3597	<.0001	fs_733	1	0.7192	0.3964
D_overdue3_1500less	1	0.2897	0.5904	fs_733_new	1	0.0823	0.7742
D_overdue3_1500more	1	75.6522	<.0001	D_EBITDA	1	13.7071	0.0002
propadle_rev_days	1	0.2145	0.6430	fs_731	1	0.0226	0.8805
D_RD	1	0.0045	0.9467	rating_class	1	25.7013	<.0001
D_RD10	1	0.0002	0.9888	rating_miss	1	0.5259	0.4683
D_balance_miss	1	0.2665	0.6057	poradie	1	22.3704	<.0001
avg_balance	1	0.2051	0.6506	rozdil_notches	1	0.0155	0.8999
				rating_manual	1	0.4524	0.4573

Source: SAS statistical package and own computation.

Figure A.3: EWM: model building - stepwise selection univariate

Summary of Stepwise Selection							
	Effect						
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	D_overdue3_1800more		1	1	75.6822		<.0001
2	log_turnover_months		1	2	22.9816		<.0001
3	fs_702b		1	3	16.9365		<.0001
4	D_seg_ree		1	4	10.7515		0.0010
5	overduecelkem0		1	5	9.5844		0.0020
6	rating_class		1	6	9.3331		0.0023
7	poradie		1	7	7.7176		0.0055
8	pomer		1	8	3.4503		0.0632
9	D_nace_machin		1	9	4.1121		0.0426
10	D_nace_ree		1	10	3.7320		0.0534
11	D_EBITDA		1	11	2.6216		0.1054
12	rating_manual		1	12	1.5588		0.2118

Source: SAS statistical package and own computation.

Figure A.4: EWM: model building - backward elimination

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	D_nace_auto	1	46	0,0000	0,9960
2	turnover_months	1	45	0,0008	0,9768
3	rating_miss	1	44	0,0026	0,9595
4	D_rev	1	43	0,0028	0,9578
5	D_count3_1to3	1	42	0,0016	0,9679
6	log_avg_overdue3	1	41	0,0194	0,8894
7	log_avg_balance	1	40	0,0270	0,8695
8	log_onbalance	1	39	0,0358	0,8498
9	D_turnover_months	1	38	0,0402	0,8411
10	power0	1	37	0,0470	0,8284
11	D_balance	1	36	0,0462	0,8299
12	D_Debt	1	35	0,0940	0,7591
13	fs_733	1	34	0,2141	0,6436
14	D_EQ	1	33	0,2202	0,6389
15	D_usage	1	32	0,2594	0,6105
16	log_total_exposure	1	31	0,2570	0,6122
17	fs_716	1	30	0,5505	0,4581
18	count1	1	29	0,5771	0,4474
19	count3	1	28	0,6464	0,4214
20	D_count3_4more	1	27	0,1405	0,7078
21	D_receivables	1	26	0,7040	0,4014
22	log_avg_kredity	1	25	0,8150	0,3667
23	fs_713	1	24	0,9209	0,3372
24	D_kredity	1	23	0,9294	0,3350
25	D_daysS	1	22	1,0719	0,3005
26	D_Cover	1	21	0,9079	0,3407
27	avg_overdue1	1	20	0,8223	0,3645
28	D_QL	1	19	0,9256	0,3360
29	D_power0	1	18	0,9958	0,3183
30	dayss	1	17	0,8470	0,3574
31	avg_usage	1	16	0,6926	0,4053

Source: SAS statistical package and own computation.

Figure A.5: EWM: model building - forward selection

Summary of Forward Selection					
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	D_overdue3_1800more	1	1	75.6822	<.0001
2	turnover_months	1	2	107.2267	<.0001
3	fs_702b	1	3	22.5346	<.0001
4	pomer	1	4	13.9729	0.0002
5	D_nace_machin	1	5	8.0808	0.0045
6	overduecelkem0	1	6	6.6505	0.0099
7	rating_class	1	7	4.4811	0.0343
8	poradie	1	8	7.5119	0.0061
9	D_rev	1	9	4.7851	0.0287
10	log_turnover_months	1	10	4.2782	0.0386
11	D_seg_ree	1	11	7.4064	0.0065

Source: SAS statistical package and own computation.

Figure A.6: Early Warning Model output - all (13 variables)

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-9.0710	0.9750	86.5580	<.0001
pomer	1	1.2612	0.5258	5.7530	0.0165
D_seg_ree	1	2.7934	0.5582	25.0450	<.0001
D_nace_machin	1	0.8796	0.4113	4.5734	0.0325
D_nace_ree	1	-1.3086	0.5814	5.0658	0.0244
D_daysS	1	1.9882	0.5096	15.2229	<.0001
D_overdue3_1800more	1	1.3909	0.2881	23.3163	<.0001
propadle_rev_days	1	0.00387	0.00241	2.5872	0.1077
log_turnover_months	1	0.6462	0.1110	33.9120	<.0001
fs_702b	1	-0.0106	0.00398	7.1044	0.0077
D_receivables	1	-0.5206	0.3222	2.6106	0.1062
D_EBITDA	1	1.0377	0.3229	10.3264	0.0013
rating_class	1	2.7304	0.6025	20.5353	<.0001
poradie	1	-0.7050	0.1987	12.5885	0.0004

Source: SAS statistical package and own computation.

Figure A.7: Odds Ratio Estimates - all (13 variables)

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
pomer	3.530	1.259	9.892
D_seg_ree	16.336	5.471	48.783
D_nace_machin	2.410	1.076	5.397
D_nace_ree	0.270	0.086	0.844
D_daysS	7.302	2.690	19.825
D_overdue3_1800more	4.019	2.285	7.067
propadle_rev_days	1.004	0.999	1.009
log_turnover_months	1.908	1.535	2.372
fs_702b	0.989	0.982	0.997
D_receivables	0.594	0.316	1.117
D_EBITDA	2.823	1.499	5.315
rating_class	15.339	4.709	49.966
poradie	0.494	0.335	0.729

Source: SAS statistical package and own computation.

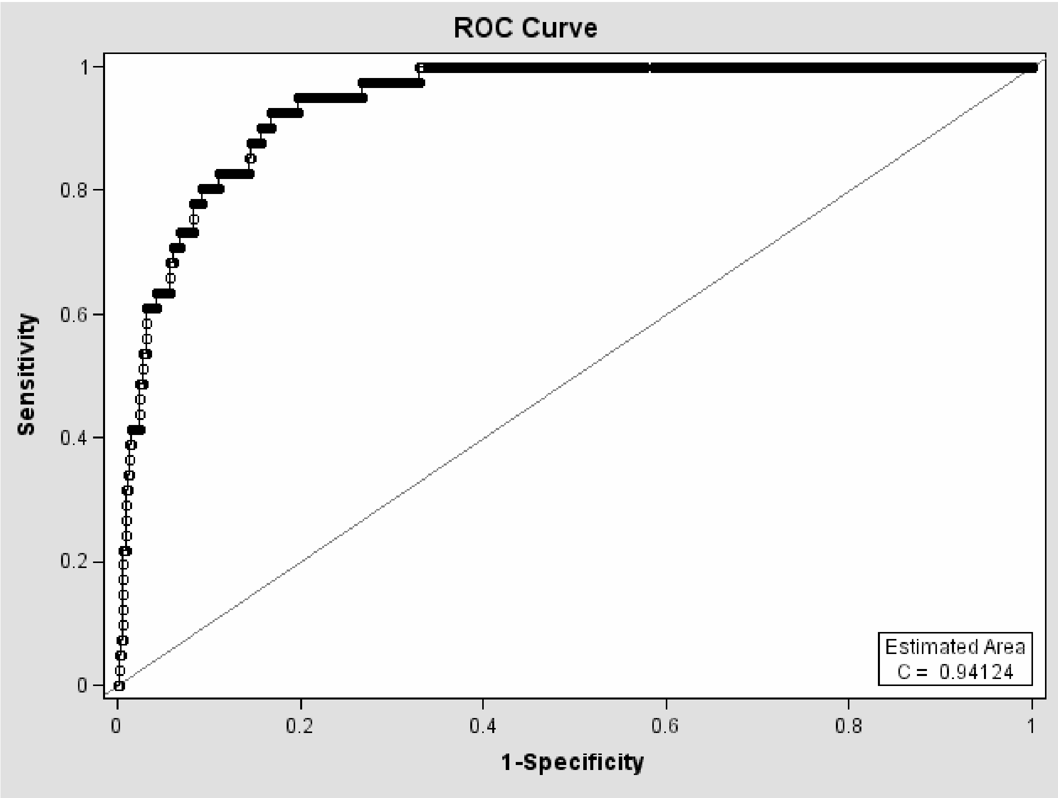
Figure A.8: EWM: Statistics (13 variables)

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	678.547	436.762
SC	684.101	514.515
-2 Log L	676.547	408.762

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	267.7853	13	<.0001
Score	324.0389	13	<.0001
Wald	148.3345	13	<.0001

Source: SAS statistical package and own computation.

Figure A.9: Receiver Operating Characteristic Curve (13 variables)



Source: SAS statistical package and own computation.

Figure A.10: Association of Predicted Probabilities and Observation Responses - all (13 variables)

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	94.0	Somers' D	0.882
Percent Discordant	5.7	Gamma	0.885
Percent Tied	0.3	Tau-a	0.073
Pairs	149732	c	0.941

Source: SAS statistical package and own computation.

Figure A.11: Decile Analysis - all observations (13 variables)

	Prospects	Predicted probability	Default rate
Decile			
0	190	0.31119	0.31578947
1	191	0.06123	0.08376963
2	191	0.02656	0.02094241
3	191	0.01401	0.01047120
4	190	0.00812	0.00000000
5	191	0.00469	0.00000000
6	191	0.00290	0.00000000
7	191	0.00151	0.00000000
8	191	0.00065	0.00000000
9	191	0.00013	0.00000000
Total	1,908	0.04298	0.04297694

Source: SAS statistical package and own computation.

Figure A.12: Early Warning Model output - 11 variables

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-8.8794	1.4726	36.3555	<.0001
pomer	1	2.0888	0.9425	4.9120	0.0267
D_seg_ree	1	2.6893	0.7935	11.4869	0.0007
D_nace_machin	1	1.0917	0.5848	3.4853	0.0619
D_nace_ree	1	-1.7242	0.8252	4.3664	0.0367
overduecelkem0	1	3.9220	1.3687	8.2110	0.0042
D_overdue3_1800more	1	1.6764	0.4397	14.5382	0.0001
log_turnover_months	1	0.8216	0.1540	28.4522	<.0001
fs_702b	1	-0.0189	0.00592	10.2214	0.0014
D_EBITDA	1	0.7693	0.4794	2.5749	0.1086
rating_class	1	2.9560	0.9705	9.2780	0.0023
poradie	1	-0.8530	0.3133	7.4108	0.0065

Source: SAS statistical package and own computation.

Figure A.13: Odds Ratio Estimates - 11 variables

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
pomer	8.075	1.273	51.217
D_seg_ree	14.721	3.108	69.716
D_nace_machin	2.979	0.947	9.374
D_nace_ree	0.178	0.035	0.899
overduecelkem0	50.503	3.454	738.545
D_overdue3_1800more	5.346	2.258	12.656
log_turnover_months	2.274	1.681	3.075
fs_702b	0.981	0.970	0.993
D_EBITDA	2.158	0.843	5.523
rating_class	19.221	2.869	128.778
poradie	0.426	0.231	0.788

Source: SAS statistical package and own computation.

Figure A.14: EWM: Statistics - 11 variables

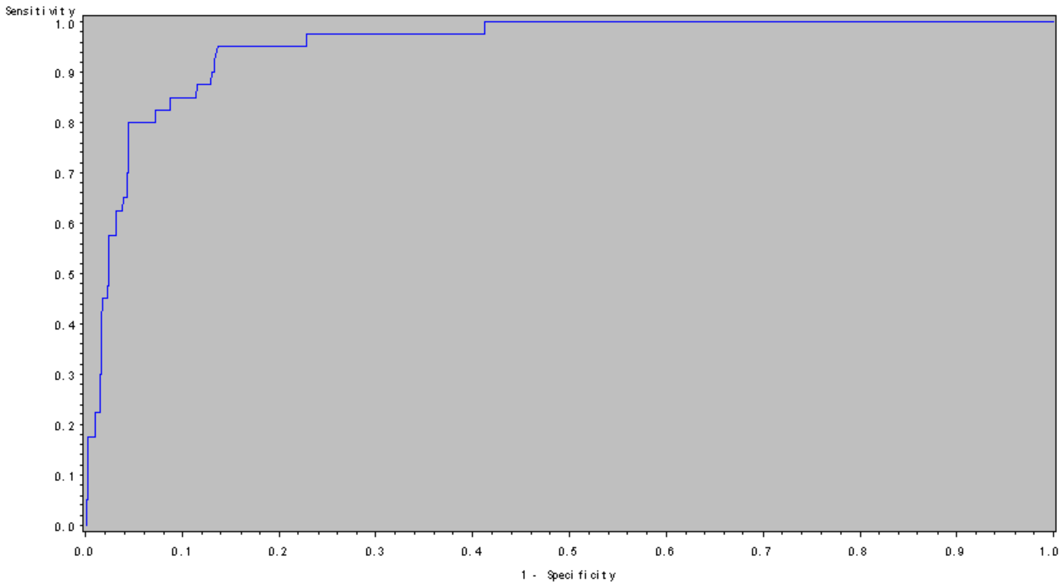
Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	330.450	206.371
SC	335.267	264.172
-2 Log L	328.450	182.371

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	146.0792	11	<.0001
Score	165.4492	11	<.0001
Wald	73.3750	11	<.0001

Source: SAS statistical package and own computation.

Figure A.15: Receiver Operating Characteristic Curve - 11 variables



Source: SAS statistical package and own computation.

Figure A.16: Association of Predicted Probabilities and Observation Responses - 11 variables

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	94.9	**Somers' D**	0.900
Percent Discordant	4.9	**Gamma**	0.903
Percent Tied	0.3	**Tau-a**	0.075
Pairs	34920	**c**	0.950

Source: SAS statistical package and own computation.

Figure A.17: Decile Analysis - origin sample - 11 variables

	Prospects	Predicted probability	Default rate
Decile			
0	91	0.34037	0.35164835
1	91	0.05615	0.06593407
2	91	0.02161	0.01098901
3	92	0.01004	0.00000000
4	91	0.00545	0.01098901
5	91	0.00299	0.00000000
6	92	0.00171	0.00000000
7	91	0.00079	0.00000000
8	91	0.00028	0.00000000
9	92	0.00004	0.00000000
Total	913	0.04381	0.04381161

Source: SAS statistical package and own computation.

Figure A.18: Decile Analysis - validating sample - 11 variables

	Prospects	Predicted probability	Default rate
Decile			
0	99	0.29796	0.27272727
1	99	0.05175	0.07070707
2	100	0.02249	0.03000000
3	99	0.01046	0.00000000
4	100	0.00552	0.05000000
5	99	0.00296	0.00000000
6	100	0.00140	0.00000000
7	99	0.00067	0.00000000
8	100	0.00027	0.00000000
9	100	0.00004	0.00000000
Total	995	0.03919	0.04221106

Source: SAS statistical package and own computation.

Appendix B

Content of Enclosed DVD

There is a DVD enclosed to this thesis which contains the SAS and LaTeXsource codes. The empirical data are not enclosed as they are strictly confidential.

- Folder 1: Source codes
- Folder 2: LaTeXsource codes
- Folder 3: Thesis in PDF format