Univerzita Karlova v Praze
Fakulta sociálních věd

Institut ekonomických studií

RIGORÓZNÍ PRÁCE

Corporate Credit Risk Under Basel II:

Application of IRB vs. Standardized Approach in the Context of Czech Economy

(Quantitative Study)

Vypracoval:	**Mgr. Martin Kubiček**
Konzultant:	**Prof. Ing. Michal Mejstřík, CSc.**
Akademický rok:	**2005/2006**

Prohlášení

Prohlašuji, že jsem rigorózní práci vypracoval samostatně a použil pouze uvedené prameny a literaturu.

V Praze dne 10. 9. 2006

Mgr. Martin Kubíček

Poděkování

Děkuji Prof. Ing. Michalu Mejstříkovi, CSc. za vstřícnost a cenné poznatky, které mi poskytl při psaní této práce. Dále mé poděkování patří společnosti Creditreform za poskytnutí potřebných dat, bez kterých by tato práce nemohla vzniknout.

ABSTRACT

This thesis presents research on corporate credit risk modeling under the New Basel Capital Accord framework using a real data set. This study provides theoretical foundations of credit risk modeling under the New Basel Capital Accord as well as empirical application of credit risk modeling to a unique data set of Czech companies provided by Creditreform. Several alternative logit regression models are presented, statistically tested and compared. Furthermore, two distinct approaches to calibration of rating classes of a rating system are developed and validated. Finally, the minimum regulatory capital requirements under the standardized approach and the internal ratings based approach of the New Basel Capital Accord are calculated and compared to the capital requirements under the current regulation.

KEYWORDS: Credit Rating, Rating System, Logit Regression, Discriminatory Power, Calibration, New Basel Capital Accord, Capital Requirements

ABSTRAKT

Předkládaná práce se zaměřuje na modelování korporátního úvěrového rizika s ohledem na Nová basilejská pravidla pro kapitálovou přiměřenost. Poskytuje jak teoretické poznatky tak ukázku aplikace na unikátním reálném datovém souboru poskytnutém k těmto účelům společností Creditreform. Na základě těchto dat jsou představeny logistické modely, které jsou dále statisticky testovány a porovnávány. Následně jsou prezentovány a validovány dva rozdílné přístupy ke kalibraci jednotlivých tříd ratingového modelu. V závěrečné části práce jsou spočítány minimální kapitálové požadavky a to dle standardizované metody a metody vnitřních ratingů, které vycházejí ze zmíněných basilejských pravidel. Tyto výsledky jsou pak porovnávány s kapitálovými požadavky vypočtenými podle současných pravidel.

KLÍČOVÁ SLOVA: úvěrový rating, ratingový model, logistická regrese, síla diskriminace, kalibrace, Nová basilejská pravidla pro kapitálovou přiměřenost, kapitálový požadavek

1. Introduction

In order to ensure increased added value to their stakeholders, banks enroll in various activities. Ranking high in the list, financial intermediation, i.e. borrowing money from depositors and lending money to different institutions, individuals and firms is one of most important activities of a bank. As banks play a central role in the economic system, regulation is necessary in order to prevent severe ruptures in the economic system.

This regulation is achieved thanks to various instances and mechanisms. First of all, there is a nation wide regulation, where local banks are subject to a national regulator. Furthermore, in the past decades the Basel Committee on Banking Supervision (BCBS) introduced worldwide regulatory suggestions with the aim to standardize the behavior of all internationally active banks. The Basel Capital Accord (also known as Basel I) introduced in 1988 is now being replaced by the *New Basel Capital Accord* (NBCA) also known as Basel II. In my thesis, I would like to elaborate on the latter, i.e. on the New Basel Capital Accord BCBS (2005a) that is nowadays being implemented by all major banks throughout the world.

Among many other new things, the NBCA introduced new approaches to calculating regulatory capital for *credit risk*, as Basel I was often criticized for not taking the quality of banks' credit portfolio into account. This caveat is overcome in the NBCA by the introduction of a whole new methodology to credit risk, the *internal ratings based* approach (IRB). Under this new approach, banks with good credit portfolio quality should benefit by holding less regulatory capital compared to the approach introduced in Basel I and only slightly modified in the standardized approach of the NBCA. However, the application of the IRB approach is more demanding for banks since it requires that they come up with their own models (rating systems) to assess the *probability of default* (PD) of their obligors.

In line with the Basel I, the NBCA works with segmentation of credit portfolio on sovereign, retail, corporate and so on. As the space constraints of this thesis do not allow focusing on all of these segments, I have chosen to limit myself to the last category. I will, therefore, devote my efforts to the clarification of banks' rating systems for *corporate* clients.

The central hypothesis that I will try to prove empirically in this thesis is that the IRB methodology brings benefits in the form of lower regulatory capital held for credit risk compared to the Basel I methodology and the standardized approach of the NBCA.

There are several ways to achieve my goal. One of them would be the description and analysis of an existing internal rating system used for calculating the regulatory capital for credit risk by the IRB approach. But this would prove difficult as banks do not openly share their systems and hence the background data would be impossible to retrieve. Therefore, amongst all possibilities, I have chosen to clarify the matter by trying to build (construct) a rating system of my own. This approach not only allows a deeper insight of the issue, but also reveals possible drawbacks and dead-ends that banks themselves may encounter when trying to set up the rating system on their own.

The content of my thesis, as partially outlined above, entails some general overview of both of the Basel Capital Accords in the chapter 2. When describing the NBCA, emphasis is put on the *first pillar* that deals with capital requirements.

Chapter 3 is devoted to the detailed analysis of the internal ratings based approach. A precise description of the underlying model and methodology is presented along with theoretical arguments in favor of the IRB approach over the standardized one.

In the next chapter, the data set used in this thesis is presented. I had the chance to get a large (and reliable) data set from Creditreform, a German company founded in 1879. Its main task is to check the reliability of business partners of its clients in order to protect them from irrecoverable claims. As a data set is critical to setting up a valid rating system, data cleaning was carried out, a process which is thoroughly described in the chapter 4.

Chapter 5 deals with the creation of the rating system itself and all the steps of the procedure are explained in detail. The core of the rating system is a *logit regression* model. Two alternative regression models are introduced, statistically tested and eventually the superior one is chosen. One of the next most important steps, which is setting up the rating classes, is performed in two alternative ways since there is no natural optimal solution. Hence, two alternative rating systems are created.

After the rating systems are outlined, the next necessary step is the validation of the rating systems. Hence, the concept of discriminatory power is introduced in the chapter 6. Within the chapter, *cumulative accuracy profile*, followed by *receiver operating characteristics* and *conditional information entropy ratio* are respectively outlined and used for validation of the two rating systems, elaborated in the previous chapter.

After the validation of the rating systems, the next necessary step, their calibration, is also clarified in the chapter 6. Here, *binomial* and *Hosmer-Lemeshow* tests are utilized in order to check the sound calibration of the two rating systems. The results of the tests are compared and the better rating system is then chosen for calculating the capital requirements under the IRB approach.

The validated rating system is used to calculate the regulatory capital for credit risk. In chapter 7, the regulatory capital calculations according to the Basel I and the standardized and internal ratings based approach of the NBCA are conducted.

Chapter 8 concludes this thesis and summarizes all findings.

The whole exercise of building an internal rating system is undertaken with the aim to understand and, hence, present the system of banks' internal mechanisms, used to control and mitigate credit risk. This analytical approach was chosen with the goal of applying the theoretical knowledge to real-life situations and understanding better internal bank procedures.

Still, when compared to banks, the established model based on the Creditreform data disposes of one major advantage. Banks, when building their internal rating system only possess the information about clients to whom the money was actually lent. Therefore, concerning the clients that have been refused, the information about the capacity to pay back debt is missing in the system. As a result, the banks' data sample may not be completely accurate and faces a selection bias. Compared to this sample, the data I had at my disposition has a much larger scale. Therefore, the model set up in this thesis, should be able to discriminate much more accurately than systems used in banks and appears free of this kind of selection bias.

Nevertheless, the aim of this thesis was not advisory (i.e. to set up a better-than-bank's system) – its' only intent was to understand and present the internal ratings based approach *per se*.

2. Introduction to the New Basel Capital Accord

In this chapter, the New Basel Capital Accord (NBCA) know also as Basel II is introduced. The chapter begins with an explanation about the theoretical background and the reasons for which banking regulation is necessary. Further on, the historical context and shortcomings of Basel I (predecessor of the NBCA) are described. The main focus of this chapter is the description of the *there pillars* of the NBCA, emphasis being put on pillar 1 which deals with the minimal capital requirements.

2.1. Regulation of banking system

Three main reasons for capital regulation of financial institutions and banks in particular are widely mentioned in literature such as Pelizzon and Schaefer (2005), Saidenberg and Schuermann (2003) and Dierick et al. (2005) among others. The first reason stems from the *asymmetric information*, where we find depositors on one side and better-informed banks on the other. Here, the role of regulation is to protect depositors from exploitation by an opaque bank.

The second reason is *systemic risk*. Banks are considered to be a source of systemic risk because of their central role in the payments system and, more importantly, in the allocation of financial resources. Banking books are mostly composed of relatively short term liabilities (deposits) and relatively illiquid assets in the form of long term loans to firms and households.

"Deposit insurance is designed to overcome the asymmetry of information in the banking system." (Saidenberg and Schuermann 2003, p.1) A bank is more aware than its depositors about its overall risk profile and exposure. Depositors are left with incomplete information and can, therefore, not distinguish a good bank from a bad one. They also know that banks are highly leveraged institutions. If there was no deposit insurance, depositors would have a strong incentive to withdraw their funds in case of even the slightest doubt about the financial health of a particular bank. Deposit insurance is, therefore, designed to prevent overreaction of depositors to bad news about banks and forms a safety net.

Deposit insurance brings the benefit of overcoming the asymmetry of information, but comes at a cost. The cost here is *moral hazard* of depositors. They no longer have an incentive to monitor banks since they know that they will get their funds back up to a certain coverage limit. Therefore, the presence of the safety net designed to prevent bank panic can lead to the moral hazard. The goal of banking regulation is to maintain balance between these two factors.

Preventing systemic risk is also considered as fundamental rationale for imposing a regulation in the form of capital requirements on banks. "The assumption is that shareholders will not take account of the social costs of systemic risk in their capital decisions and so will tend to hold less capital than if these spillover costs were considered." (Saidenberg and Schuermann 2003, p.3) The *minimal capital requirements* imposed on banks are intended to ensure that shareholders of a bank have at least some minimal level of resources in order to honor their commitments to their customers. The banking business is one of the most indebted businesses in the world since banks borrow money from depositors and lend it to households or firms. Minimal capital requirements should ensure that owners of a bank do not engage in risky lending in order to make more profit. Since the money a bank lends does not belong to the owners but to the depositors, severe moral hazard of the owners of a bank could emerge.

Simply put, a bank is operating with money that does not belong to it. "Capital requirements are intended to mitigate moral hazard by ensuring that the owners of a financial institution have a stake in ensuring that the firm does not engage in fraud and conforms to conduct of business rules, if only to avoid fines or loss of equity value."

(Saidenberg and Schuermann 2003, p.4) In order for the capital requirements to be effective in this task, the capital requirements must be sensitive to the risk exposure.[1]

2.2. History of Basel I

The early 1980's period was influenced by concerns about financial health of international banks and by complaints of unfair competition in the banking industry. Subsequently, The Basel Committee on Banking Supervision (BCBS) initiated world wide discussion on the revision of capital standards. As a result, an agreement was reached in July 1988.

The 1988 Basel Capital Accord, known as Basel I (Basel Committee on Banking Supervision (1988)) was the first international accord of its kind. The Basel Capital Accord considered only credit risk[2] and was entirely focused on capital requirements. Basel I set minimum capital standards for internationally active banks. It was decided that the capital ratio would be 8% of risk-weighted assets.[3]

Even though Basel I was major step towards international financial stability there were many criticisms of the accord. According to Stephanou and Mendoza (2005), one of the main criticisms of Basel I was its "one-size-fits-all" approach.

[1] Sensitivity of the minimal capital requirements to overall risk exposure of a bank is one of the major improvements of the NBCA compared to Basel I.
[2] "Risk of losses in on and off-balance sheet positions resulting from the failure of a counterparty to perform according to a contractual arrangement." (Dierick et al. 2005, p.45)
[3] Basel I as well defined the eligible regulatory capital, assets subject to risk weighting and the risk weighting mechanism.

2.3. *Shortcomings of Basel I*

In order to understand the international effort to come up with the NBCA it is important to comprehend the shortcomings of Basel I. Under Basel I, minimal capital requirements are only moderately related to a bank's risk exposure. This fact is pointed out by Pelizzon and Schaefer (2005) among others. For example, capital requirement on a credit exposure does not differ according to the rating of the obligor. From the regulatory point of view there is no difference whether a bank has a credit exposure with an AAA rated company or with a C rated company.

Saidenberg and Schuermann (2003) argue that this lack of risk sensitivity may distort economic decision making since economic capital[4] and regulatory capital[5] are not in line. It can be more profitable for banks to optimize their portfolios in order to reduce regulatory capital requirements, rather than to optimize their overall risk exposure. Under Basel I, banks also have incentives for capital arbitrage through securitization, for example.

Banks may also be reluctant to invest in better risk management since it can bring no regulatory capital benefits. Sound risk management is costly and, under Basel I, does not bring any benefits, so banks are not motivated to invest in this area.

2.4. *The New Basle Capital Accord*

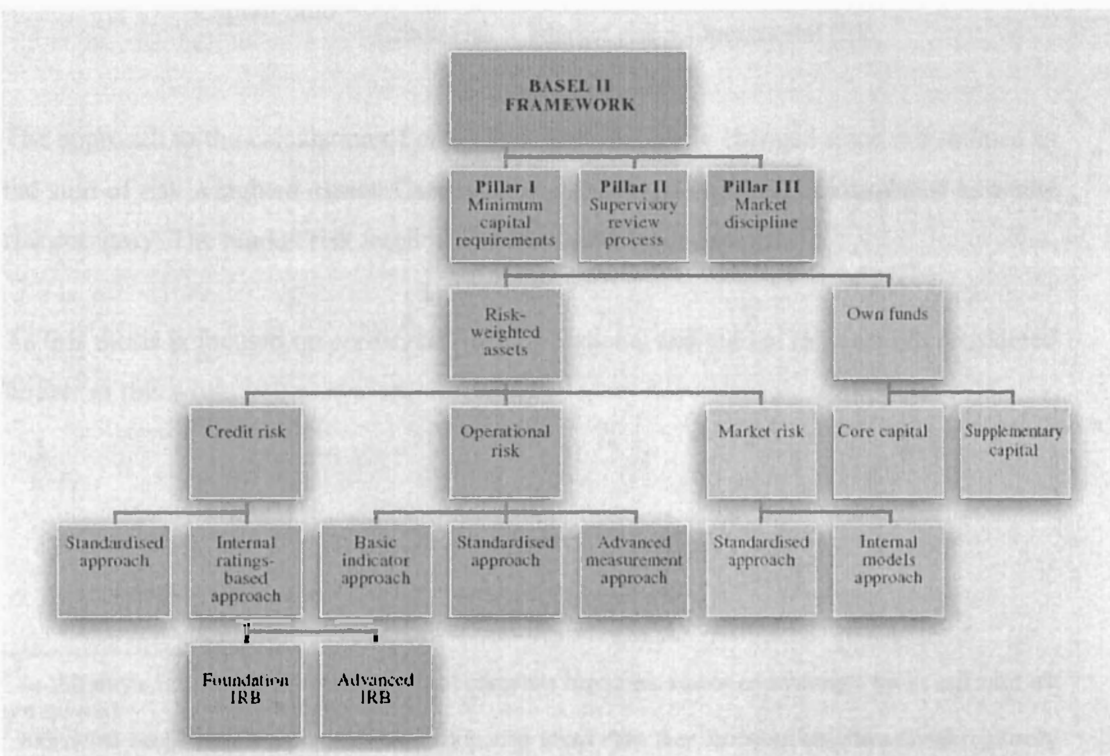
The NBCA aims to alleviate some of the drawbacks of the current Basle Capital Accord. The design of the NBCA is focused on better alignment between regulatory capital and

[4] "Capital held and allocated by the bank internally as a result of its own assessment of risk. It can differ from regulatory capital, which is determined according to supervisory rules." (Dierick et al., p. 45)
5 "Own funds that are eligible to meet the regulatory capital requirements; consist of core capital and supplementary capital, after a number of deductions. (Dierick et al., p. 46)

underlying risks by encouraging better risk management practices in banks. By aligning regulatory capital more closely to a bank's own risk estimates, the NBCA narrows the gap between regulatory and economic capital requirements. The NBCA is based on the three pillars approach. Pillar 1 focuses on minimal capital requirements, pillar 2 defines supervisory review of internal bank assessments of capital relative to risks and pillar 3 increases public disclosure of risks and capital information sufficient to provide meaningful market discipline.

All these pillars are important and self enforcing. "It is significant that for the first time in international capital regulation, supervision and market discipline are placed at the same point in hierarchy as the regulatory minimum." (Pelizzon and Schaefer 2005, p.7) Overview of the New Framework is graphically represented on Picture 1 below.

Picture 1



Source: Dierick et al. (2005)

2.5. *Pillar 1*[6]

In accordance with the theory of *asymmetry of information*[7], pillar 1 provides a menu based approach to credit risk, rather than a uniform "one-size-fits-all" rule.

The minimum capital ratio of 8% remains unchanged. "This ratio expresses the relationship between the bank's regulatory own funds (capital) and its risk weighted assets, a measure of the risk it incurs." (Dierick et al. 2005, p.9) However, the computation of capital requirement has changed substantially. The denominator represents the fundamental changes induced by the NBCA. The numerator in the formula, i.e. what counts toward eligible capital remains unchanged compared to Basel I.

$$\text{Capital ratio} = \frac{\text{Eligible capital}}{\text{Credit risk} + \text{Market risk} + \text{Operational risk}}$$

The approach to the calculation of credit risk has completely changed since it is defined as the sum of risk weighted assets. Capital charge for operational risk is introduced as a new risk category. The market risk fraction remains basically unchanged.[8]

As this thesis is focused on *credit risk* only, operational and market risks are not considered further in this work.

---

[6] As this thesis is concerned with credit risk of corporate exposures, claims on sovereign, banks and retail are not covered.

[7] Individual banks have more and better information about risks they face than an international regulatory body.

[8] The market risk was introduced to Basel I in 1996 when the Basel Capital Accord underwent a major amendment.

2.6. *Credit Risk under Pillar 1*

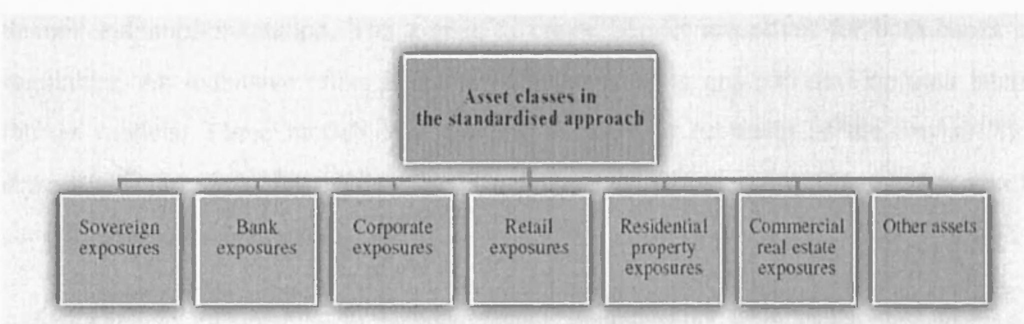
The NBCA reflects the development of credit risk management practices in the financial industry, and pillar I introduces a range of approaches for credit risk assessment. These are the standardized and the internal ratings based (IRB) approaches. The IRB methodology has two subsequent versions – *foundation* and *advanced.*

2.6.1. Standardized Approach

Changes in the standardized approach are modest compared to Basel I. They incorporate risk sensitivities through observable risks measures such as external credit ratings. However, there are several criteria that rating agencies must satisfy in order to be recognized by the banking supervisor as eligible to provide banks with ratings.

Under the standardized approach, exposures are classified into a set of standardized asset classes as depicted on Picture 2. A risk weight is applied to each class reflecting the relative amount of credit risk.

Picture 2



Source: Dierick et al. (2005)

Whereas Basel I assigned 100% risk weight to all corporate exposures, the NBCA offers considerable differentiation in the risk weights according to external ratings. See Picture 3 below.

Picture 3

Credit assessment	AAA to AA-	A+ to A-	BBB+ to BB-	Below BB-	Unrated
Risk weight	20%	50%	100%	150%	100%

Source: BCBS (2005a)

Hence, even the calculation of capital requirements by the standardized approach can bring benefits if a credit portfolio is of a good quality. Nevertheless, obtaining a rating from a recognized rating agency is costly. It can be expected that for many small and medium companies the cost of obtaining a rating would exceed the benefits. Hence, when providing a credit to these companies, in most cases the 100% risk weight will have to be assigned.

2.6.2. Internal Ratings Based Approach[9]

The IRB approach is fundamentally different from the Basel I methodology in concept, design and implementation. The aim is to create correct incentives for both banks and regulators. An extensive effort is expected from banks in order to develop their internal ratings models. These models will provide banks with estimates of the *probability of default* (PD) of their obligors as the PDs are one of the most important inputs into the capital requirements function.

The IRB approach has two versions; these are the foundation approach and the advanced approach. The difference between the two is the extent of internal information of banks

[9] The IRB approach is described in greater detail in chapter 3.

used in the calculation of the capital requirements. In the foundation approach, only PD may be estimated internally, subject to supervisory review (as defined in pillar 2). Loss given default and maturity adjustment is fixed and provided by the regulator. Under the advanced approach, all parameters are determined internally and are subject to supervisory review[10].

On the regulatory side, an extensive effort is also necessary. The national regulator needs to approve and validate rating models of individual banks to ensure that the models are NBCA compliant and appropriate for regulatory capital computations.

2.7. Pillar 2 and 3

Pillar 2 is considered to be an essential element of the NBCA since it promotes the supervisory review process and expects an active role of supervisors. Banks are encouraged to develop internal economic capital assessments in line with their risk profiles for identifying, measuring and controlling risks. The fact that any rule based approach will inevitably lag behind changes in risk profiles of banks is recognized by emphasis on the internal assessments of capital adequacy.

Hence, it is the task of the supervisors to analyze whether a specific bank's capital adequacy assessment is in line with its overall risk profile and business strategy. Supervisors also review if a bank should not hold additional capital against risks that are not (or not fully) covered in pillar 1. "Relative to the present situation, pillar 2 requires supervisors to apply considerably more discretion in their assessment of capital adequacy in individual banks." (Dierick et al. 2005, p.18) Pillar 2, therefore, provides a basis for supervisory intervention to prevent unwarranted decline in a bank's capital.

[10] To be able to use the advanced IRB approach a model for the loss given default (LGD) is necessary. Modeling LGD is beyond the scope of this thesis. Consequently, only the foundation IRB is considered further on in this thesis.

Pillar 3 reflects the effort of the Basel Committee to promote market discipline through greater transparency and improved disclosure of banks across markets. By promoting transparency, pillar 3 attempts to capture the benefits of market discipline. The idea is for banks to tell market participants key parameters of their risk measures, risk management and business profile. For the disclosure of credit risk in particular, portfolio structure, major types of credit exposures and geographical and sectoral distribution should be published. The market's judgment of capital adequacy will influence the banks' share price and access to funding. Therefore, pillar 3 should improve market discipline on banks.

3. Internal Ratings Based Approach

The internal ratings based approach to credit risk represents one of the major innovations in the NBCA. This chapter is devoted to the detailed description of the IRB approach. The formula for computation of capital requirements using the IRB approach is presented as well as the underlying model and methodology the formula is based on.

I would also like to argue why it is beneficial for banks to adopt the IRB approach instead of the standardized one. The argumentation presented in the second part of this chapter is purely of theoretical nature. In later chapters, empirical study is concluded in order to prove the central hypothesis of this thesis, i.e. that adopting the IRB approach yields lower regulatory capital than the standardized approach.

3.1. Description of the IRB Approach

Possible losses from a credit portfolio can be separated in two: the expected loss and the unexpected loss. Banks are expected to cover for the former themselves, whereas regulatory capital stands as a cushion against the latter.

"Although credit losses naturally fluctuate over time and with economic conditions, there is (ceteris paribus) a statistically measured, long-run average loss level." (Stephanou and Mendoza 2005, p.6) For a single exposure, the expected loss (EL) can be estimated as follows[11]:

$$EL=PD*LGD*EAD.$$

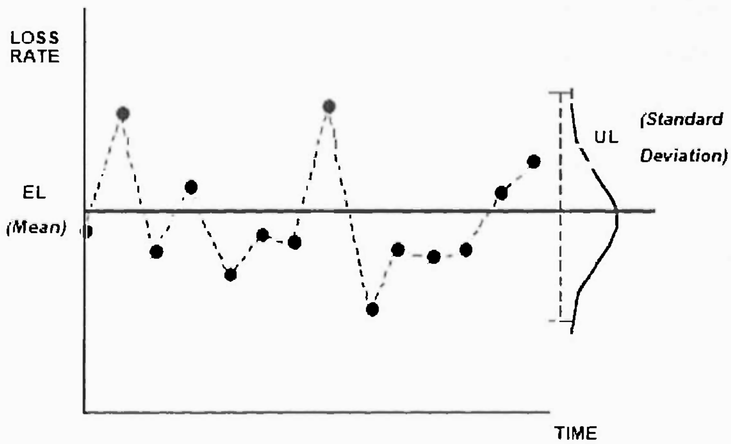
[11] Description of these abbreviations can be found on page 19.

The expected loss from a whole portfolio is then simply the sum of expected losses of the single exposures.

Note that expected losses do not represent risks from a credit portfolio. They should be viewed as the cost of doing business, and capital requirements are not designed to cover expected losses. Banks have to cover their expected losses through pricing, provisions and write-offs.

The regulatory capital requirement is designed to cover for unexpected loss. The unexpected loss is simply the standard deviation of the expected loss. For illustration see Picture 4 below.

Picture 4



Source: Stephanou and Mendoza (2005)

The IRB approach is, in fact, designed to model the variance of losses from a credit portfolio. To achieve this goal, the IRB approach is based on the Value at Risk (VaR) methodology[12] and the key parameters used to estimate credit risk are the following:

[12] For details on the VaR methodology see appendix 1.

PD the probability of default of an obligor over a one year horizon,

LGD the loss given default as a percentage of exposure at default[13],

EAD exposure at default, the nominal amount of credit[14],

M maturity.

For the foundation IRB, only the PD can be assessed internally by banks. The LGD and M are set by the supervisor. "The IRB at heart provides continuous mapping from basic set of four inputs parameters (PD, LGD, EAD and M), plus some other observables such as borrower type, to a minimum capital requirement."(Saidenberg and Schuermann 2003, p.9)

The theoretical basis of the IRB model is the asymptotic single risk factor (ASRF) model. As in many other credit risk models, the probability of a borrower being unable to repay his loan is derived from the distance between the value of its assets and the nominal amount of his debt. "The value of the firm's assets is modeled as a variable which changes over time, in part as a result of the impact of random shocks." (Dierick et al. 2005, p.12) Default occurs when the modeled value of an obligor's assets falls below the amount of outstanding debt. Credit risk is then measured by PD over one year time horizon.

The ASRF model does not take into account obligors' specific risks. These idiosyncratic risks can be diversified away in a large loan portfolio. Hence, the model assumes that banks have large and well diversified portfolios because it measures only the marginal risk contribution of an exposure to such a portfolio. In this well diversified portfolio, every obligor accounts for only a very small share of the total portfolio exposure. Such a portfolio is also well diversified across geographical areas and industry sectors in a large economy.[15]

[13] The paragraph 287 of the NBCA (BCBS 2005a) suggests the LGD to be equal to 45%. This value of LGD is considered through out this thesis.

[14] By definition, the EAD is always known by banks.

[15] It is probable that these assumptions will not be met, for example, by banks specializing in lending to a particular industrial sector. "Hence, under pillar 2, supervisors will analyze potential risk concentrations and may potentially develop appropriate pillar 2 capital buffers against such risk concentrations." (Dierick et al. 2005, p.19)

"The IRB approach therefore contains a deliberate simplification compared with the most advanced techniques currently applied." (Dierick et al. 2005, p.12)

For corporate exposures, the capital charge is equal to:

$$\text{Capital requirement} = 8\% * \text{RWA}$$

The risk weighted assets (RWA) are computed according to the following formula:

$$\text{RWA} = \text{K(PD, LGD, M)} * 12.5 * \text{EAD}$$

where 12.5 is the reciprocal of 8% (the overall level of minimum capital as percentage of RWA). As with the expected loss, all calculations are done at exposure level. Total portfolio capital is then the sum of all exposure level capital charges.

The mapping function K has the following form:

$$K = \left[LGD * \Phi\left[(1-\rho)^{-0.5} * \Phi^{-1}(PD) + \left(\frac{\rho}{1-\rho}\right)^{0.5} * \Phi^{-1}(0.999) \right] - PD * LGD \right] * \left(\frac{1}{1-1.5*b}\right) * \left(1+(M-2.5)*b\right)$$

Where,

$$\rho = 0.12 * \lambda + 0.24 * (1-\lambda)$$

$$\lambda = \frac{1-e^{(-50*PD)}}{1-e^{(-50)}}$$

$$b = (0.11852 - 0.05478 * \ln(PD))^2.$$

$\Phi(.)$ stands for the standard normal cumulative distribution function; ρ is a weighted correlation parameter with weight λ determined by PD and b is a maturity adjustment. The implied correlation ρ among particular exposures is a weighted average between 12% and 24%.

When taking a closer look on the equations stated above, several interesting characteristics can be revealed. First of all, the correlation of losses[16] from single exposures is included, but modeled only as a function of PD. Hence, the IRB approach ignores potentially important portfolio characteristics such as geographical and industrial concentration. It is the task of the national regulator to decide whether diversification in the portfolio is sufficient to comply with the assumptions of the IRB model. If it is not the case, the national regulator may, under pillar 2, request increase of regulatory capital.

Secondly, maturity adjustment M is introduced to cope with potential credit quality deterioration of exposures with longer maturities. The NBCA suggest the average maturity to be 2.5 years. Exposures with shorter maturities are favored whereas exposures with longer maturities are penalized.

Thirdly, the function K is set up in such a way that a bank's regulatory capital should cover unexpected losses with probability 99.9% over a one year horizon. Thus, it can be said that there is one in 1 000 chance that a bank's losses from a credit portfolio over one year will exceed the minimum regulatory capital.

Finally, the IRB approach distinguishes between expected and unexpected losses from a credit portfolio. "Under the IRB approach, banks will need to compare the amount of total eligible provisions [...] with total estimated EL." (Stephanou and Mendoza 2005, p. 20) If the expected loss exceeds the total eligible provisions then banks have to deduct the difference from the regulatory capital. If the opposite is true, i.e. total eligible provisions

[16] Note that there is a parallel with the portfolio theory underlying the CAPM model know from finance. The standard deviation of total credit portfolio depends on pair correlations of all exposures in the portfolio.

exceed the expected loss, then banks may recognize the difference in the regulatory capital up to a certain limit.

3.2. *Advantages of the IRB Approach*

Under the NBCA, banks are free to choose whether they will adopt the IRB approach or whether they will use the standardized approach. Adopting the IRB approach allows banks to use their own models to assess *creditworthiness* (PDs) of their customers. In the following paragraphs, I would like to present reasons why it is beneficial for banks to adopt the IRB approach.

First of all, Schwaiger (2003) argues that internationally active banks with well diversified portfolios of good quality should benefit from the IRB approach by holding less regulatory capital. Secondly, as pointed out by Dierick et al. (2005), the IRB approach allows banks to align their economic and regulatory capital. Thirdly, the decision not to adopt the IRB approach could lead to loss of prestige in the banking community and finally, if a majority of banks on a particular market adopts the IRB approach then rest of the banks would have to follow. Therefore, it can be expected that banks will develop their credit rating systems and adopt the IRB approach to calculate the amount of regulatory capital for credit risk.

The main task for banks before adopting the IRB, is to develop credit rating systems that are statistically powerful. With powerful rating systems, banks will be able to differentiate "good" potential borrowers from "bad" ones, thus overcoming the *asymmetry of information*. A potential borrower always has more information regarding its ability to repay a loan. The analysis of a borrower's quantitative and qualitative information using a rating system allows a bank's risk analyst to make a better decision whether to grant the loan or not.

The *Adverse selection* is also connected with asymmetry of information. „In competitive framework a poor statistical power of a bank's internal rating system will deteriorate the economic performance due to adverse selection, i.e. customers with a better credit quality than assesses by the bank will potentially walk away and leave the bank with portfolio of customers with a credit quality lower than estimated" (Schwaiger et al. 2004, p.2) Hence, customers with good rating quality will take credit from a bank with more powerful rating system that can assess their creditworthiness better. Such a bank can, subsequently, offer these customers better terms on loans. A bank with a less powerful rating system will be left with "bad" customers who know about their low credit quality.

Now that the theoretical foundations of the IRB approach were described, I can turn to the empirical part of this thesis. In the following chapters, I will try to prove the central hypothesis of lower regulatory capital requirements under the IRB approach compared to the Basel I methodology and the standardized approach of the NBCA.

4. Data

In order to prove the central hypothesis of this thesis a valid data set is necessary. This chapter is therefore dedicated to the description of the data set that is used throughout this thesis. First, the data source is mentioned. Secondly, documentation regarding the necessary cleaning of the data is presented. Thirdly, various descriptive characteristics of the data set are mentioned.

4.1. Data source - Creditreform

The data for this thesis was obtained from Creditreform, a German company founded in 1879. Its main task is to check the reliability of business partners of its clients in order to protect them from irrecoverable claims. The Czech branch of Creditreform was established in 1890, closed in 1948 and then reestablished in 1991.

Amongst many other business activities, Creditreform computes *solvency index* for Czech companies. This index can be viewed as a measure of trustworthiness. The main advantage of this solvency index is that it provides a long history of stable data for a large number of companies.

4.1.1. Computation of the Solvency Index

The solvency index embodies 15 aspects of relevant business information and is based on a precise mathematical formula. Table 1 below shows all the information used in order to set up the solvency index.

Table 1

1	mode of payments	9	payment moral of customers
2	credit judgment	10	payment moral
3	order book situation	11	capital turnover
4	business expansion / development	12	legal form
5	number of employees	13	age of company
6	turnover / overall performance	14	shareholder structure
7	productivity / turnover per employee	15	business sector conditions
8	equity		

Source: www.creditreform.cz

This information is then aggregated and the relative weights of all information categories are summarized in the following Table 2. An example of a solvency index calculation can be found in appendix 2.

Table 2

turnover information / financial data	25%
credit judgment	25%
mode of payments	20%
structural data	15%
industry and size	15%

Source: www.creditreform.cz

The range of the solvency index spans from 100 to 600 risk points, 100 risk points representing the most trustworthy entrepreneurships. The index is a continuous scale from 100 to 500 points followed by a stand alone category of 600 risk points. The 600 risk points represent the legal default.

4.2. Data cleaning

The data pool obtained from Creditreform contained observations for 85 169 companies with a time span from 1998 to 2005. The data was ordered into a matrix D, $D = (index_{ij})$, where $i = 1..m$, $j = 1..n$ and $m = 85\ 169$, $n = 8$ (m being the number of companies, n representing years from 1998 to 2005).

As mentioned, the data set I have obtained from Creditreform is comprehensible, fairly large and reliable. Nevertheless, due to various reasons, some data points for particular years and / or firms were missing. Hence, I have faced a bias whether to work with all (and incomplete) data or limit myself to use of data which are thoroughly complete only. As both ways are feasible and yield sound results in a particular context, they could be both, under certain circumstances, employed in this thesis. Hence, my choice could be basically deliberate. Moreover, as the data pool was large enough, I have chosen the latter method, i.e. to narrow down the data pool and use only the complete data. This method is also in line with Kuhn (1977) who advised to work rather with smaller and more reliable pool than vice versa.

Hence, excessive cleaning of the data set was carried out due to missing observations for particular companies for certain years; companies that entered the data pool with 600 risk points and so on. Not cleaning the data pool would inflict severe statistical noise. The cleaning criteria were set up as follows:

- No reestablishment of a defaulted company – once a company received 600 or 500[17] risk points it is considered "dead"[18] and later observations are not taken into account.

[17] For details for the setting of default threshold see chapter 5.1.1.
[18] In line with other literature in this field, such as Lando and Skodeberg (2002) and Schuermann and Hanson (2004), the default class is considered all consuming. Hence, once a company defaults no later observations are considered.

- More than one consecutive observation per company – since transitions from one year to another are needed for creating a rating system.

- Deletion of all observations for year 2005, since the data was not complete at the time of the elaboration of this thesis.

One of the goals of this thesis is to set up a sound rating system that will provide probabilities of default. In order to do so, the solvency indices in year t and year $t+1$ are necessary to estimate the PDs. Thus, it is necessary to have two or more consecutive observations per company.

All 2005 solvency indices were not complete at the time of the writing of this thesis. Using only the available indices could cause a bias in the calibration of the rating system. One can expect that big and well know companies were researched and entered into the database first. Also companies on the brink of bankruptcy would be one of the first in the database. If the 2005 observations were left in the data pool, severe statistical noise would be inflicted.

Application of these three criteria reduced the number of rows in the matrix *D* to 16 363. More than three quarters of the observations did not pass the cleaning and had to be deleted since they brought no useful information for a rating system construction. Following and combining of the cleaning criteria led to the deletion of companies that were new to the database in 2004 and 2005 because there was no transition for these companies. Also, companies that had either 600 or 500 risk points or no value at all were deleted. There were also companies that entered the database in year 1998 as defaulted and were, therefore, deleted from the data set[19]. However, the highest number of companies was deleted because there was only one observation for them.

---

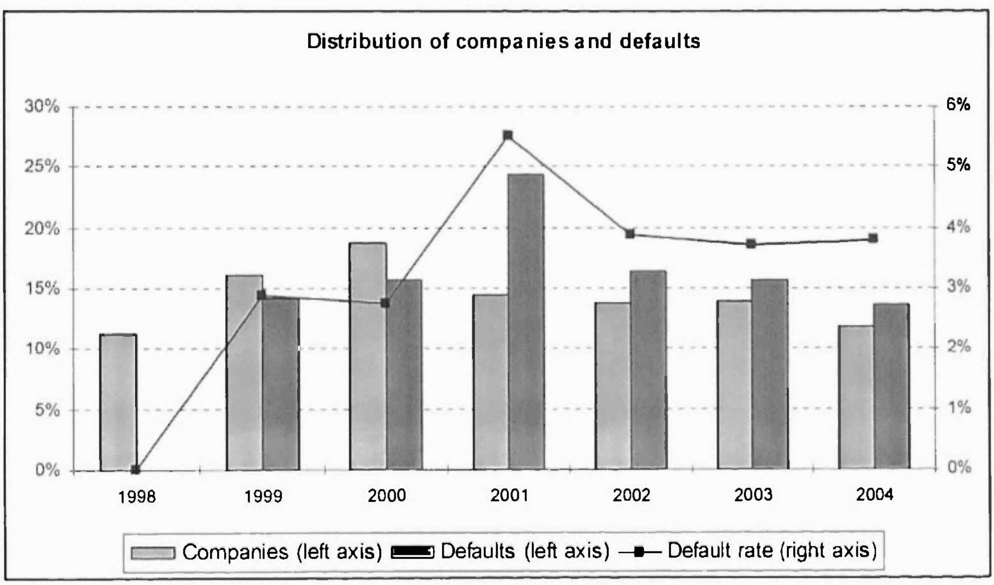
[19] This cleaning resulted in default rate equal to zero for the year 1998.

4.3. Data description

All obligors in the data set are Czech companies. The cleaned sample contains 16 363 companies with the total number of observations reaching 53 489 and covering a time span of 7 years (1998-2004).

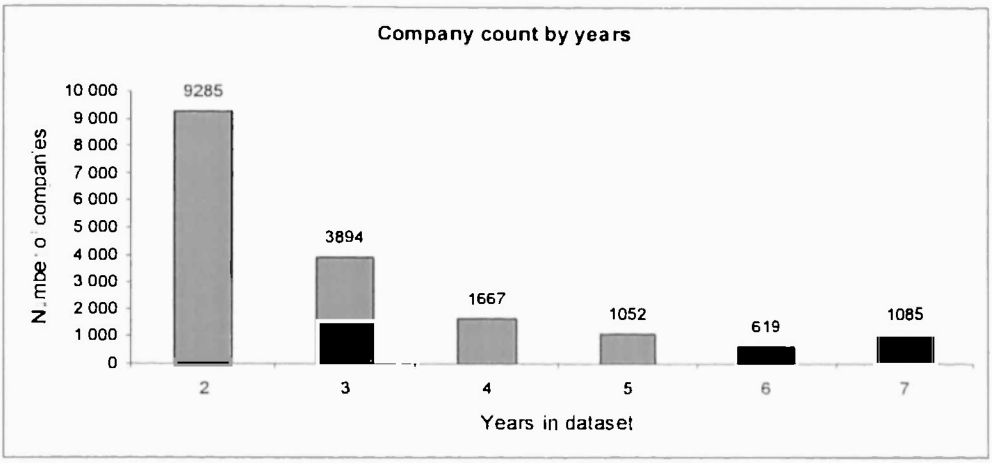
Graph 1 shows that the observations are rather uniformly distributed over the years with a peak in 2000. Defaults are rather uniformly distributed as well with the exception in 2001. The default rate stays below 4% for all years except 2001.

Graph 1



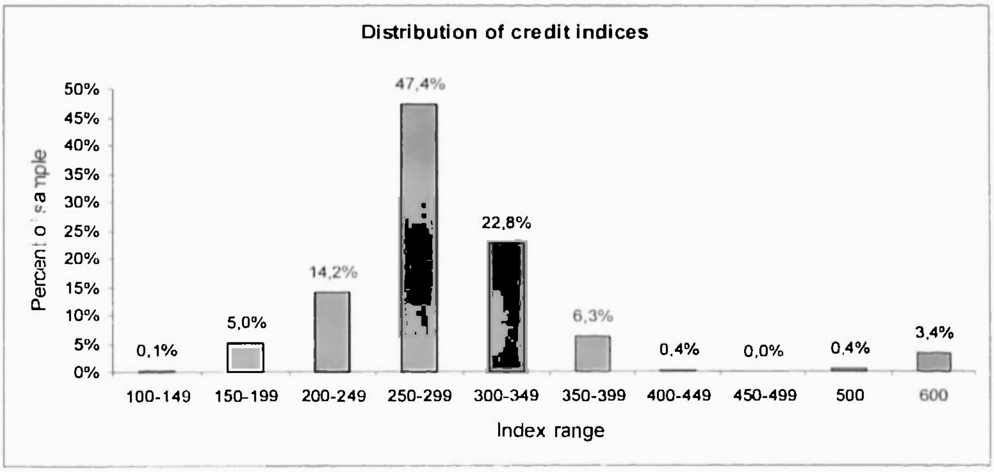
One can easily calculate that the 53 489 observations in our sample represent about 47% of all possible observations if every company was assigned an index every year. For 9 285 companies the data pool contained only two consecutive observations, while for 8 317 companies multiple consecutive observations at different points in time were available. The distribution of the number of observations per company is presented in Graph 2 below.

Graph 2



Graph 3 shows the dispersion of indices within ranges. Almost half of the sample is located within the 250 – 299 range. The data sample has a clear bell shaped distribution where indices are almost equally distributed around the central range.[20]

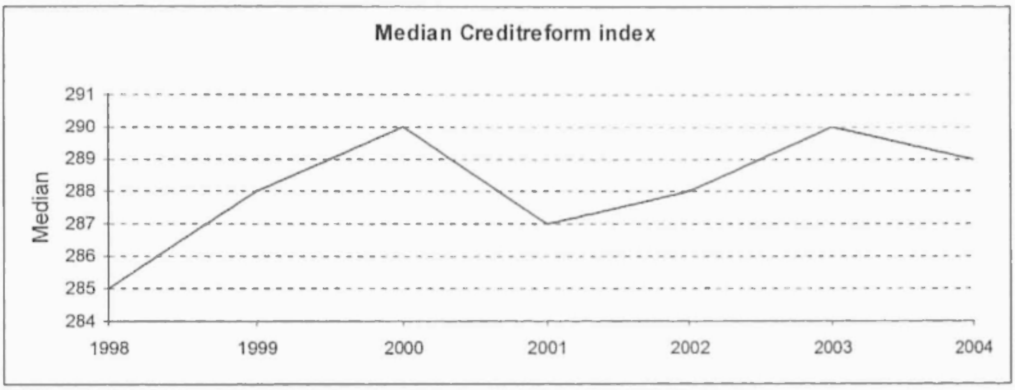
Graph 3



[20] The fact that there are no companies with index values 450-499 was on of the reasons for including the stand alone 500 class in the default category. The discussion of setting the default threshold is presented in chapter 5.1.1.

It is also noteworthy that there appears to be no clear trend in the median Creditreform solvency index. Thus, we can conclude that there is no observable improvement in the credit quality of the data sample.

Graph 4



We can see that the data set is rather large and has reasonable characteristics. Therefore, it is possible to build a valid rating system with this data sample, one that can be used to provide PDs for computation of the minimal capital requirements using the IRB approach.

5. Rating system

There exist many approaches to creating a rating system that are widely discussed in academic literature. If one works with a data set of rated companies, then it is possible to follow the so called cohort (historical) approach as in Lando and Skodeberg (2002), Schuermann and Jafry (2003) and Kadlcakova and Keplinger (2004). Ratings can also be modeled according to Markow chains as described in Lando and Skodeberg (2002). Probabilities of default are then computed according to rating migrations of rated companies.

In the event that companies in the data set are not rated, it is necessary to use a regression model where financial ratios are used as independent variables and the dependent variable marks either default or non default. This approach is used in Fernandes (2005).

The data set used in this thesis allows for both approaches. The Creditreform solvency index can be viewed as 400 + 1[21] ratings. Then it would be possible to use the cohort approach to set up rating classes and assign a PD to every class. The other possibility would be to use the solvency index in a single factor regression model and assign a PD to every index value. Subsequently, rating classes would be set up and PD of every class would equal the expected PD of the rating class.

Since this thesis is focused on the application of the IRB approach in the banking industry, the regression model approach was selected. Building a rating system is the main task that banks willing to adopt internal ratings based approach must undertake. Since most of a bank's obligors are privately held firms with no market data available, banks will probably adopt accounting based credit scoring models.

---

[21] The solvency index has a range from 100 to 500 risk points plus stand alone 600 risk points. For details see chapter 4.

To set up such a model, banks need a history of accounting statements of their clients as well as their delinquency history. Variety of financial ratios can be calculated from the accounting statements and these can be used as inputs to a quantitative model. The model can be a regression model where the delinquency information is used as a dependent variable and the financial ratios as independent variables.

However, all banks, by definition, face a selection bias. Banks can consider the behavior of accepted loans only. There is no way a bank can know what would happen to an obligor it rejected, if it would default or not[22]. Hence, rating models of all banks are affected by such a bias.

The advantage of my rating system is that it is not affected by such a bias, because the data set comes from an external company. Therefore, ceteris paribus, my rating system should provide more accurate results and overcome the selection bias.

The rest of this chapter is organized as follows. First, a quantitative regression model is presented along with the discussion about setting the default threshold and description of the regression data set. An alternative regression model using additional information about the regional division of companies is also introduced. Several statistical tests are applied in order to distinguish which model better fits the underlying data set.

Secondly, the results of the quantitative model are used for calibration of the rating system, i.e. setting rating classes. Since there is no natural solution to the calibration problem two different calibrations are described and subsequently two rating systems are constructed.

[22] When deciding about granting a loan banks can make two types of error. The first error is rejecting a loan to an obligor that would not subsequently default. The cost for the bank coming from this error is the cost of foregone business, i.e. the money the bank could have made if it had provided the loan. The second error happens when a bank grants a loan to an obligor that subsequently defaults. The cost here is lost interest and principal.

5.1. *Quantitative model*

The binary logit regression model is used as the quantitative model underlying the rating system built in this thesis. The model will assign each firm its individual probability of default. These individual PDs will later be used to classify firms into rating classes and to assign a probability of default to each one of these rating classes.

Such a purely quantitative rating system is not compliant with the NBCA because of the absence of human judgment. “Credit scoring models and other mechanical procedures are permissible as the primary or partial basis of rating assignments, and may play a role in the estimation of loss characteristics. Sufficient human judgment and human oversight is necessary to ensure that all relevant and material information, including that which is outside the scope of the model, is also taken into consideration, and that the model is used appropriately.” (BCBS 2005a, par. 417)

However, the methodology applied in this thesis can be used as a building block for a compliant system.

5.1.1. Setting the default threshold

The first task when building the quantitative model is to decide on the value of the Creditreform index that would indicate default. According to paragraph 452 in the NBCA (2005) document: “A default is considered to have occurred with regard to a particular obligor when the obligor is past due more than 90 days on any material credit obligation to the banking group.” The problem here is that the NBCA definition of default is of financial nature whereas the Creditreform solvency index value 600 stands for legal default, i.e. bankruptcy.

The analysis in chapter 4 and the fact that the index value of 600 means legal default and the value of 500 means serious financial difficulties motivated my decision to set the default solvency index value to 499. This definition of default corresponds to one used in Kadlcakova and Keplinger (2004).

5.1.2. Creating the data set for regression

A regression matrix was created according to previous discussions about the setting of the default threshold. In the first column of the matrix were the non defaulting Creditreform solvency indices (CRFO) for year t; the third column being filled by corresponding indices for year t + 1. The default variable D was constructed in the middle column (D=0 if $CRFO_{t+1}<500$ and D=1 if $CR_{t+1}\geq 500$). The matrix is therefore a matrix of all transition pairs obtained from the cleaned data pool. The total number of transitions is 34 384.

With the default definition and the regression data set at hand, it is possible to turn to econometric modeling.

5.1.3. The model

The quantitative model underlying the rating system has the following general form:

$$Y_t = f(\beta, CRFO_t) + e_t .$$

The dependent variable Y is the binary variable that takes value 1 if the company has defaulted at time t + 1 and 0 otherwise. Since we are concerned with one year PDs, variable CRFO is the value of the Creditreform index at time t, i.e. one year before the evaluation of the dependent variable. As mentioned earlier, the logit model is the selected functional form[23].

The regression equation has the following form:

$$Y_i = \beta_0 + \beta_1 * CRFO_i.$$

The results of the logit regression are presented in Table 3 below:

Table 3

	Estimate	Standard error	Wald statistics	p-value
β_0	6.583277	0.176426	1392.392	0.00
β_1	-0.012262[24]	0.000580	447.337	0.00

In the following paragraphs, the results as well as assumptions of the model will be subject to statistical tests (Wald test, Davidson and McKinnon test, Box-Tidwell test and Hosmer-Lemeshow test) in order to ensure the validity and robustness of the regression model.

The significance of each estimated coefficient was tested using the Wald test. Both coefficients are significant at the 95% significance level.

In order to ensure a meaningful application of the Hosmer-Lemeshow goodness-of-fit measures, it is necessary to verify the robustness of the regression model. Since most of the problems for robustness of a logit model stem from heteroskedasticity, the statistical test of Davidson and MacKinnon (1993) is applied. The hypothesis H_0 of homoskedasticity against H_1 of heteroskedasticity is tested.

[23] For details see appendix 3.
[24] The negative sign of the CRFO coefficient stems from the fact that the statistical package used to estimate the model estimates the probability of non defaults. Hence, the higher the CRFO is the lower the odds of not defaulting (the higher the probability of defaulting).

Since the χ^2 test statistic is equal to 7.05 with the χ^2 distribution with one degree of freedom at the 95% confidence level being 3.84, H_0 is rejected in favor of H_1. Therefore, this model setup is not robust, thus resulting in inconsistencies in the estimated coefficients.

Moreover, the assumption of linearity between the CRFO variable and the logit of the dependent variable was checked using the Box-Tidwell test. It turned out that there is a serious non-linear relationship. The true relationship was adequately captured using the fractional polynomial methodology. The revisited regression equation then takes the following form:

$$Y_i = \beta_0 + \beta_1 * CRFO_i^2,$$

that leads to the following results summarized in Table 4:

Table 4

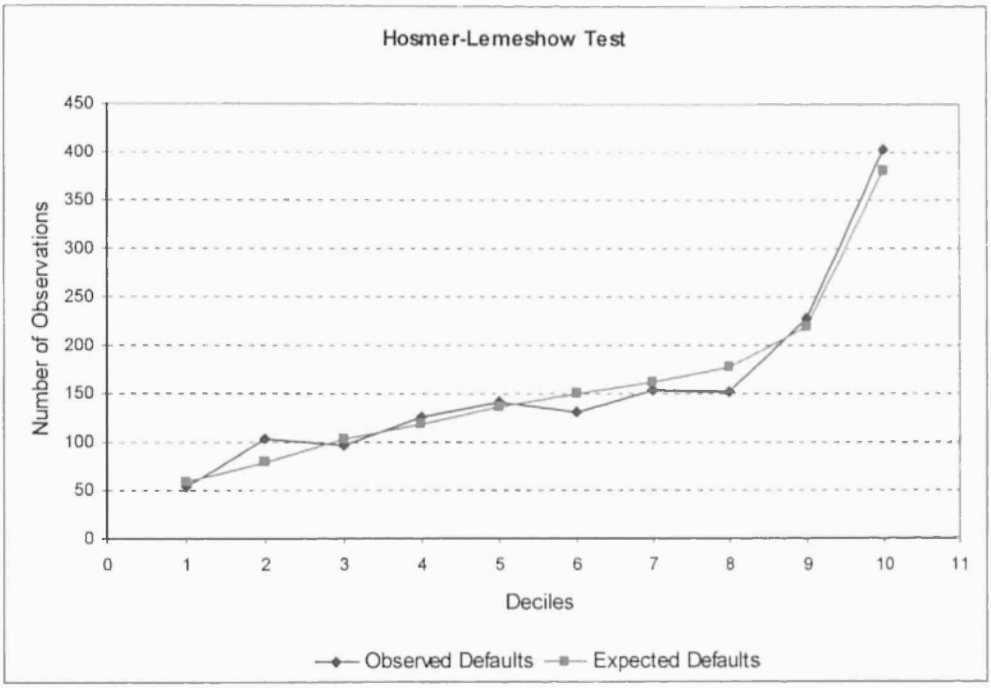
	Estimate	Standard error	Wald statistics	p-value
β_0	4.792372	0.089508	2866.679	0.00
β_1	-0.000020	0.000001	491.382	0.00

Both coefficients are still significant at the 95% significance level and there is no evidence of non-linear relationship.

In this case the H_0 of homoskedasticity cannot be dismissed. The χ^2 test statistic is equal to 3.25 with the χ^2 distribution with one degree of freedom at the 95% confidence level being 3.84. Therefore, this model is robust.

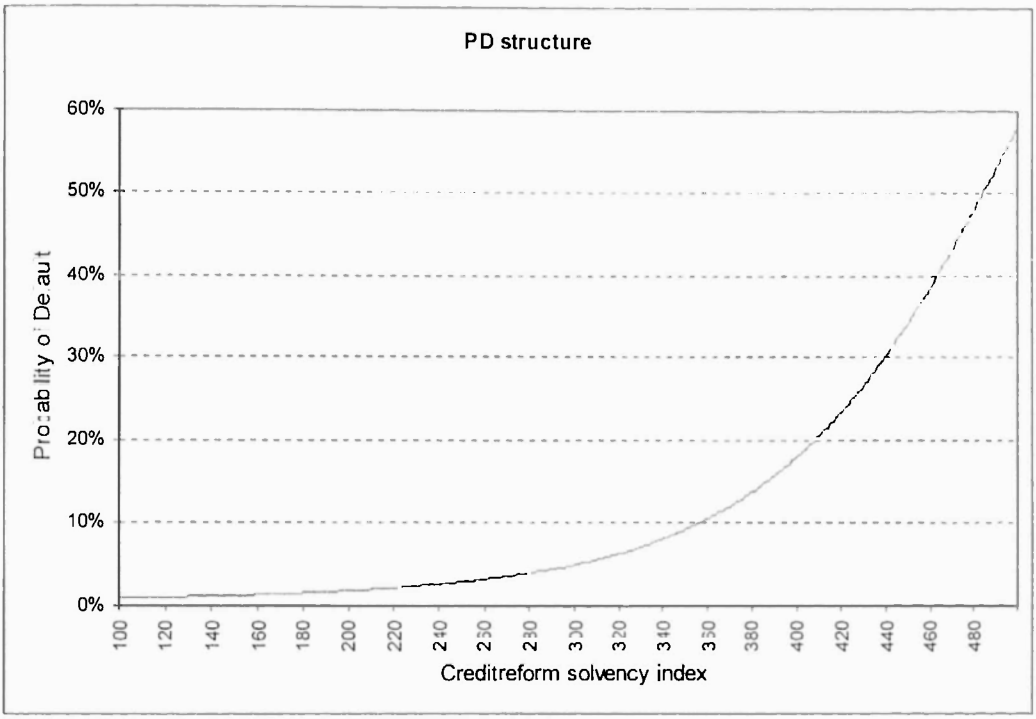
The goodness-of-fit of the corrected model was checked by the Hosmer-Lemeshow test. The Chi-square value of the test is 14.7 and corresponding p-value 0.06. Thus, we cannot reject the H_0 of good model fit at the 95% significance level. Perhaps, the results of the Hosmer-Lemeshow test will appear more clearly on the following Graph 5.

Graph 5



The observed and expected defaults are plotted for each decile. We conclude that this estimated logit regression significantly fits the observed data. The following Graph 6 shows the probabilities of default that the quantitative model assigns to each value of the Creditreform solvency index.

Graph 6



The presented regression model passed all the necessary tests and can therefore be used as the underlying model for a rating system. However, the Creditreform data pool contained additional information that could be used in the regression. To check whether the additional information brings added value is the subject of the following chapter.

5.1.4. The model – an alternative

The Creditreform data set also contains the ZIP code of every company. Thus, it is possible to differentiate companies into 7 regions of the Czech Republic[25]. Moreover, Creditreform does not take into account the regional effects when calculating the solvency index. Since it

[25] These regions are: Prague, Central Bohemia, Western / Southern Bohemia, Northern Bohemia, Eastern Bohemia, Southern Moravia and Northern Moravia.

is possible to use the regional information through dummy variables in a logit model, an alternative logit model was set up to try to make use of the regional information.

The regression equation has the following form:

$$Y_t = \beta_1 * CRFO_t + \beta_2 * R1 + \beta_3 * R2 + \beta_4 * R3 + \beta_5 * R4 + \beta_6 * R5 + \beta_7 * R6 + \beta_8 * R7$$

Where *R1,...,R7* are regional dummy variables assuming value 1 if a company is from the same region and 0 if otherwise. Note that there is no intercept in this regression equation. The reason is that since the dummy variables are mutually exclusive the intercept can be expressed as a linear combination of the dummy variables. If the intercept was left in the equation perfect collinearity between the intercept and the dummy variables would be obtained, resulting in an ill specified model.

The results of the logit regression are presented in Table 5 below:

Table 5

	Estimate	Standard error	Wald statistics	p-value
β_1	-0.011388	0.000579	386.718	0.00
β_2	6.231293	0.176396	1247.893	0.00
β_3	6.512526	0.192545	1144.018	0.00
β_4	6.361327	0.190155	1119.129	0.00
β_5	6.410346	0.197890	1049.334	0.00
β_6	6.340169	0.187973	1137.657	0.00
β_7	6.346346	0.190121	1114.266	0.00
β_8	6.240089	0.184975	1138.041	0.00

As in the previous model setup, the significance of each estimated coefficient was tested using the Wald test. All coefficients are significant at the 95% significance level. Moreover, the joint significance of the region dummies was tested by means of the joint Wald test. The χ^2 statistic yields 1328.7, with the critical value of the χ^2 distribution with seven degrees of freedom at the 95% significance level being 14.07; this implies that the hypothesis that the coefficients of all regional dummies are 0 can be rejected.

The Davidson and MacKinnon test was also applied in this alternative model setup. The results show that the H_0 of homoskedasticity can be rejected since the χ^2 test statistic is equal to 129.08 with the χ^2 distribution with eight degrees of freedom at the 95% confidence level being 15.5. Therefore, the alternative model setup is not robust, resulting in inconsistencies in the estimated coefficients.

Applying the Box-Tidwell test did not show any sign of a non-linear relationship in this model setup.

The alternative model is not robust (as shown by the Davidson and McKinnon test) and no immediate remedy is available[26]. Therefore, the Hosmer-Lemeshow test was not applied since it is not possible to ensure meaningful results of the test.

Hence, the main model specification represents the best model in terms of robustness and significance of the independent variable. Therefore, the main model is used to set up the rating system and the region information is not taken into account.

[26] Further research of this model setup would need to be concluded. However, that is beyond the scope of this thesis.

5.2. *Calibration*

With the regression model in place, the next necessary step is to set up the rating classes, i.e. to find thresholds that would determine the rating classes.

First, it is necessary to determine how many rating classes the rating system should have. According to BCBS (2005a), par. 404: "[...] a bank must have a minimum of seven borrower grades for non-defaulted borrowers and one for those that have defaulted."

Since the number of firms in my data set is sufficiently large, I decided to use nine rating classes for non-defaulters and one for defaulters. As was shown in Schwaiger et al. (2004) using more rating classes presents the advantage of reducing the differences between estimated and true PDs. Individual PDs differ across firms in a particular rating class. The deviations of the individual PDs from the overall PD of the rating class are higher than if the rating class was broken up into several subgroups.

Moreover, if one uses nine plus one rating classes, the results are comparable with rating agencies such as Moody's and S&P since they use rating systems with the same number of rating classes.

As a next step, it is necessary to choose PD boundaries or Creditreform solvency index boundaries to distribute the firms into the rating classes. The only requirement of the NBCA can be found in paragraph 403: "A bank must have a meaningful distribution of exposures across grades with no excessive concentrations, on both its borrower-rating and its facility-rating scales."

There are many different methods in order to set up the boundaries. However, according to Schwaiger et al. (2004) no natural optimal solution exists.

Nevertheless, there are criteria that a sound rating system should comply with. First, as suggested by the NBCA, the concentration of companies among rating classes should be reasonable. Secondly, the PDs of the rating classes should monotonicaly increase as we move towards worse rating classes. Thirdly, the expected PD of a rating class coming from the regression model should be in accordance with the empirical (historical) PD computed from the data set.

I decided to use two more promising approaches as suggested by Schwaiger et al. (2004). First, rating classes are set so that every class has the same number of observations. Thus, in case of nine rating classes, approximately 11% of observations are in every rating class. Secondly, the thresholds are set so that the number of defaults increases linearly from the best to the worst rating class.[27]

When setting up the rating classes all observations in the data sample from 1998 to 2004 are used. Hence, all possible information contained in the data sample is employed.

The PD of a particular rating class is calculated as the average expected PD of all firms within the rating class. To check how the rating system complies with the requirement of accordance, the so called cohort methodology is applied as well. Under the cohort methodology, the PDs of particular classes are calculated according to the following formula:

$$PD_i = \frac{D_i}{N_i},$$

where $i = 1,...,9$; PD_i represents the probability of default of rating class i, D_i is the observed number of defaults in class i and N_i stands for the total number of companies in class i.

[27] See appendix 4 for details.

The results of the calibration of the two rating systems are summarized in the two following chapters.

5.2.1. Uniform distribution of firms in rating classes

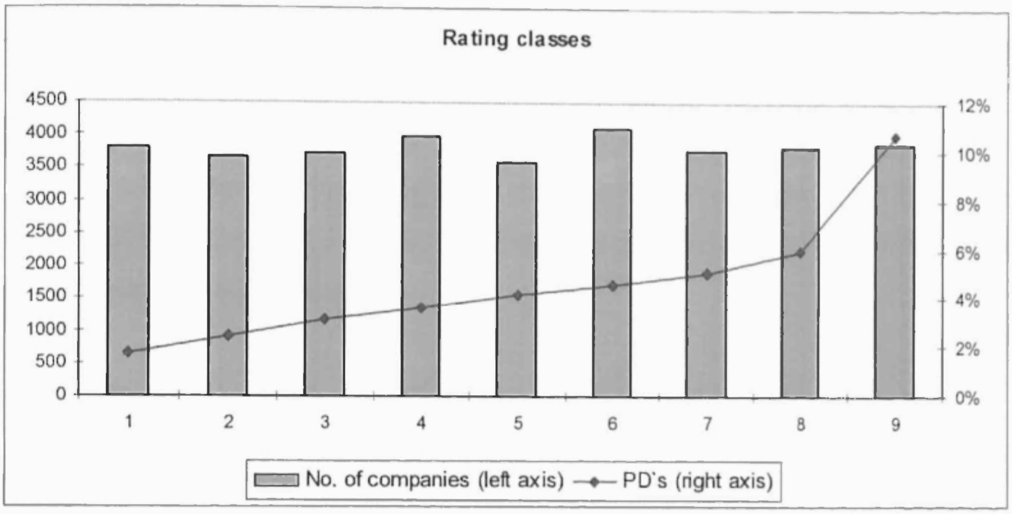
To achieve the uniform distribution of firms in rating classes the thresholds were set as in the following Table 8:

Table 8

	Thresholds
Rating class 1	100-213
Rating class 2	214-248
Rating class 3	249-264
Rating class 4	265-279
Rating class 5	280-289
Rating class 6	290-297
Rating class 7	298-306
Rating class 8	307-334
Rating class 9	335-499

Outcomes of the settings are summarized by the following Graph 7:

Graph 7



The number of observations per rating class is approximately 11% of the total number of observations in the data sample. This represents roughly 3 800 observations per rating class. The Graph 7 shows that PDs are behaving well, i.e. are increasing as we go from better rating classes to worse with a sharp jump from class 8 to class 9. Exact values of the probabilities of default are summarized in the following Table 7:

Table 7

	Probability of default - Logit model	Probability of default - Cohort approach
Rating class 1	1.75%	1.76%
Rating class 2	2.43%	2.69%
Rating class 3	3.09%	3.12%
Rating class 4	3.61%	4.18%
Rating class 5	4.15%	3.62%
Rating class 6	4.57%	4.26%
Rating class 7	5.04%	4.30%
Rating class 8	5.97%	6.11%
Rating class 9	10.71%	11.25%

The table also shows PDs calculated according to the cohort approach. We can see that the logit model fits the data reasonably well as the results are in line with the observed historical PDs[28]. Sophisticated tests of the accordance of the expected PDs and historical PDs will be carried out in the next chapter devoted to rating system validation.

The PD of 1.76% for the first rating class may seem high compared to the best rating classes in banks[29]. However, I would like to point out that the data set used in this thesis does not represent a banking portfolio.

5.2.2. Linear increase in number of defaults

The thresholds are set so that the number of defaults increases linearly from the best to the worst rating class.[30] Under this set up the expected number of defaults in a rating class as predicted by the logit model is following:

Table 8

	Expected number of defaults
Rating class 1	35
Rating class 2	70
Rating class 3	106
Rating class 4	141
Rating class 5	176
Rating class 6	211
Rating class 7	247
Rating class 8	282
Rating class 9	317

To achieve such a distribution of expected defaults, threshold for particular classes were set as following:

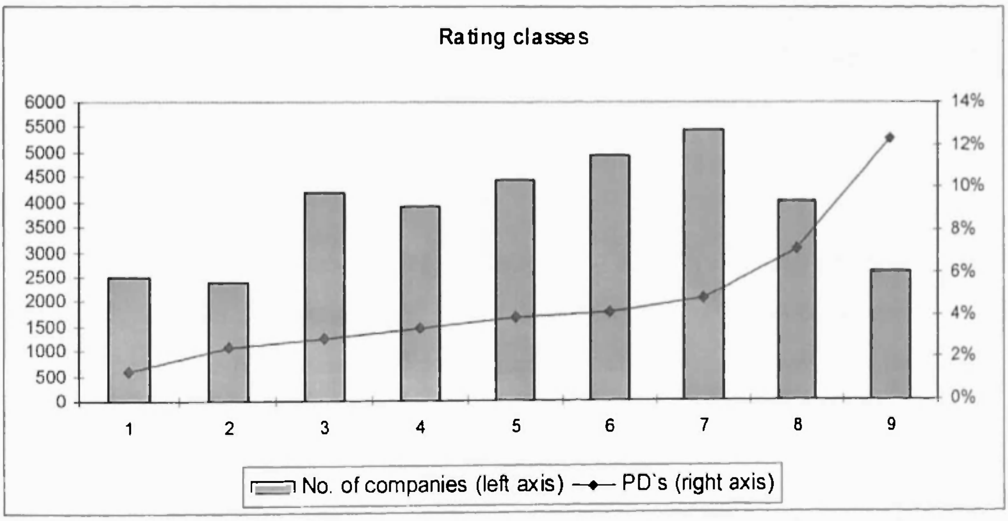
[28] However, the PDs resulting from the cohort approach are not monotonic, since the PD of class 5 is smaller than the PD of class 4.
[29] It is possible to calibrate the rating system in such a way that the first rating class would have PD less than 0.5%. However, then there would be very few companies in the first rating class.
[30] See appendix 4 for details.

Table 9

	Thresholds
Rating class 1	100-198
Rating class 2	199-226
Rating class 3	227-258
Rating class 4	259-273
Rating class 5	274-287
Rating class 6	288-296
Rating class 7	297-308
Rating class 8	309-346
Rating class 9	348-499

Outcomes of the setting are summarized by the following Graph 8:

Graph 8



The distribution of companies in rating classes does not have a clear shape and there is no extensive concentration of observations in any rating class. Moreover, the PDs are still monotonically increasing. Exact values of probabilities of default are summarized in the following Table 10:

Table 10

	Probability of default - Logit model	Probability of default - Cohort approach
Rating class 1	1.39%	1.05%
Rating class 2	2.50%	3.75%
Rating class 3	2.94%	2.59%
Rating class 4	3.42%	3.68%
Rating class 5	3.88%	3.86%
Rating class 6	4.11%	3.96%
Rating class 7	4.76%	3.98%
Rating class 8	7.10%	7.51%
Rating class 9	12.26%	12.94%

When compared with PDs calculated by the cohort approach more differences appear, than in the setup with uniform distribution of companies. Although the rating class PDs according to the logit model are monotonic, PDs calculated according to the cohort method are not. Also note that the historical PDs for rating class 4, 5, 6 and 7 are very similar whereas the PDs from the logit model are not. Therefore, it can be expected that this particular calibration will yield worse results in the calibration tests.

In this chapter, a logit regression model was introduced as the underlying qualitative model for the rating system. The regression model was successfully statistically checked and tested. An alternative regression model using additional information was also introduced, but this regression model failed in the statistical tests.

Based on the regression model, two rating systems were calibrated and checked to see if they complied with the basic requirements of low concentration, monotony and accordance. However, more sophisticated validation of the rating systems is necessary to ensure that they are well calibrated and have explanatory power. The validation of the rating systems will be the subject of the next chapter.

6. Validation

Many rating methods and rating systems were developed in the past years. Therefore, we need to answer the question of which rating methods and systems are preferable to others. After the publication of the second consultative document of the Basel Committee on Banking Supervision (2001), the need to evaluate the quality of rating systems has became increasingly important. This document announced the possibility that an internal ratings based approach could form the basis for setting capital charges to credit risk. "The importance of sound validation techniques for rating systems stems from the fact that rating models of poor quality could lead to sub-optimal capital allocation." (Engelmann et al. 2003, p.1)

The validation techniques employed throughout this chapter are standardized and widely described in literature, for example, BCBS (2005b), Moody's (2001) and Fernandes (2005) among others.

The validation of a rating system consists of two areas. The first one is called *discrimination* and focuses on how good a rating system can, ex ante, distinguish between defaulting and not defaulting obligors, i.e. how powerful the rating system is. The second area of validation called *calibration* focuses on how good the probabilities of defaults of particular rating classes are estimated. This chapter covers both issues.

All that is necessary for assessing the discriminatory power is to have rating scores produced by a rating system. The rating system can be, for example, a logit model that takes into consideration financial statements of an obligor (balance sheet and income statement) as well as certain industry, market and country characteristics, as shown in the previous chapter.

Calibration tests check the accordance of the expected PDs coming from a quantitative model with the historical PDs in the data set. Hence, if one calibrated a rating system using the so called cohort (historical) approach, no calibration tests would be necessary. However, since the rating systems presented in this thesis are based on a regression model it is necessary to test whether the predicted PDs from the model are in line with the historical PDs of the data set.

Both rating systems presented in this thesis are based on the same information, i.e. the Creditreform solvency index. They just represent different regroupings of the underlying indices. Therefore, I expect the two rating systems to yield similar results in the assessment of discriminatory power. Nevertheless, the discriminatory power analysis is carried out to show how powerful the rating systems are.

Calibration tests are also carried out to check the accordance of the expected PDs of the rating classes based on the regression model with the historical PDs. Since the two rating systems are calibrated according to different methods, the tests of calibration should show which of the two rating systems is superior.

This chapter is organized as follows. First, tests for discriminatory power are described and explained, namely, the cumulative accuracy profile (CAP), receiver operating characteristics (ROC) and the conditional information entropy ratio (CIER). Each validation technique is applied to the two rating systems. Secondly, calibration tests are introduced (binomial test and Hosmer-Lemeshow test) explained and applied.

6.1. Discriminatory power

Whether an obligor defaults or not is not known in advance. Therefore, we face a binary classification problem. Rating systems based on present available data work as classification tools in order to determine future status of an obligor. The procedure of applying a classification tool in order to assess future status of an obligor is called discrimination. The discriminatory power of a rating system denotes its ability to, ex ante, distinguish between defaulting and non defaulting obligors.

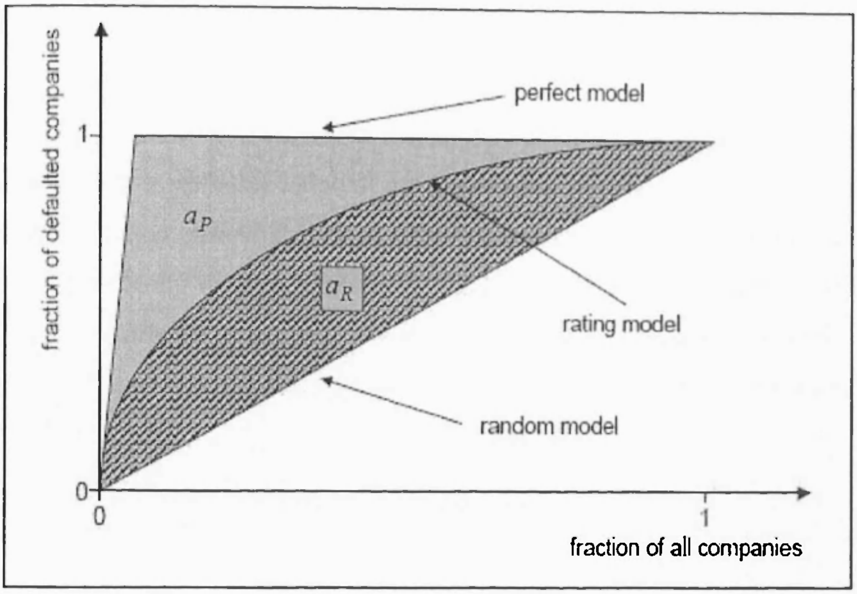
The ideal rating system would consist of only two rating classes where one class would be assigned to the expected defaulters and the other to the expected non defaulters. However, in practice this is not possible. Hence, multiple class rating systems are used[31]. So, a rating system discriminates the better, the more the distribution of defaulters and the distribution of non defaulters differ in particular rating classes.

6.1.1. Cumulative accuracy profile

The cumulative accuracy profile (CAP) is a visual tool used to asses the discriminatory power of a rating system. Accuracy ratio (AR) is the most common summary index of CAP. Picture 5 below illustrates the cumulative accuracy profile.

[31] A good rating system will assign higher share of all defaulters in worse rating classes. Therefore, we can expect that as we move to worse rating classes the relative share of defaulters will increase whereas the relative share of non defaulters will decrease.

Picture 5



Source: BCBS (2005b)

When constructing the CAP, obligors are first ordered from worst to best, i.e. from obligors with the highest probability of default to the obligors with the lowest probability of default. The cumulative percentage of all ordered obligors is displayed on the x axis. The y axis shows the cumulative percentage of defaulters. The CAP maps the fraction of all companies with the worst score onto the fraction of companies within that group.

A perfect rating model will assign defaulters with the worst scores. In this case the CAP is increasing linearly and then staying at one, since in the perfect model no defaulter should have good rating scores. "If the sample contained 10% defaulters, then a perfect model would exclude all those defaulters at 10% of the sample excluded; the 10% of companies with the lowest ranks would consist of the defaults." (Moody's 2001, p.15)

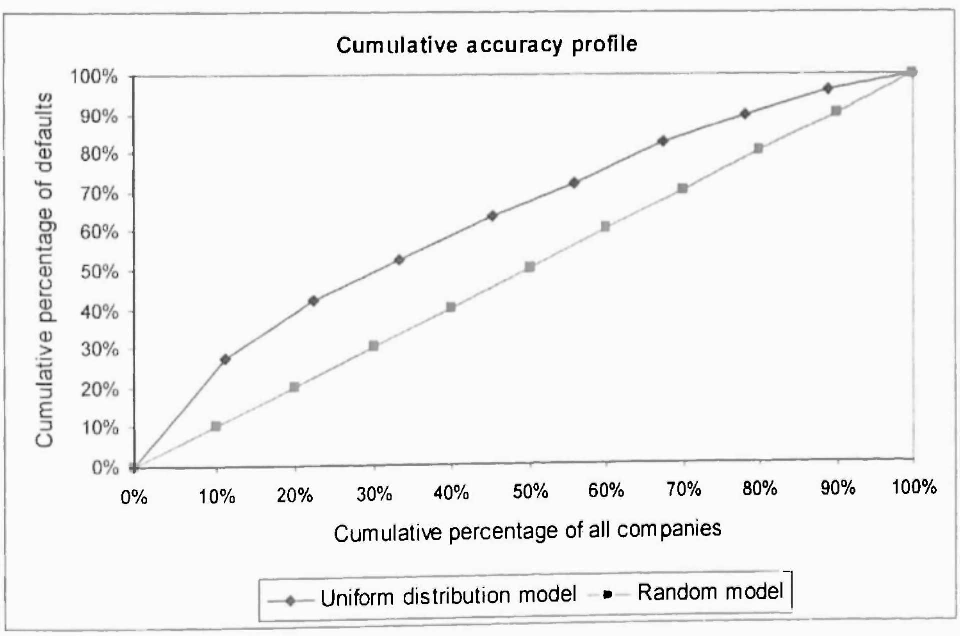
For a random model with no discriminative power a fraction x of all debtors with the lowest scores will contain x percent of all defaulters, thus producing a straight line from coordinates [0,0] to [1,1]. The two extreme cases represent boundaries for real rating systems.

CAP is a visual tool to assess the discriminatory power of a rating system and the accuracy ratio serves as a summary index of the information represented by the CAP. In the rating system validation context, the AR is defined as the ratio of the area between the CAP of the rating model being validated and the CAP of the random model, and the area between the CAP of the perfect rating model and the CAP of the random model. Therefore, the rating model is all the better the closer the AR is to one. According to the notation on Picture 4 the AR is defined as

$$AR = \frac{a_r}{a_p}.$$

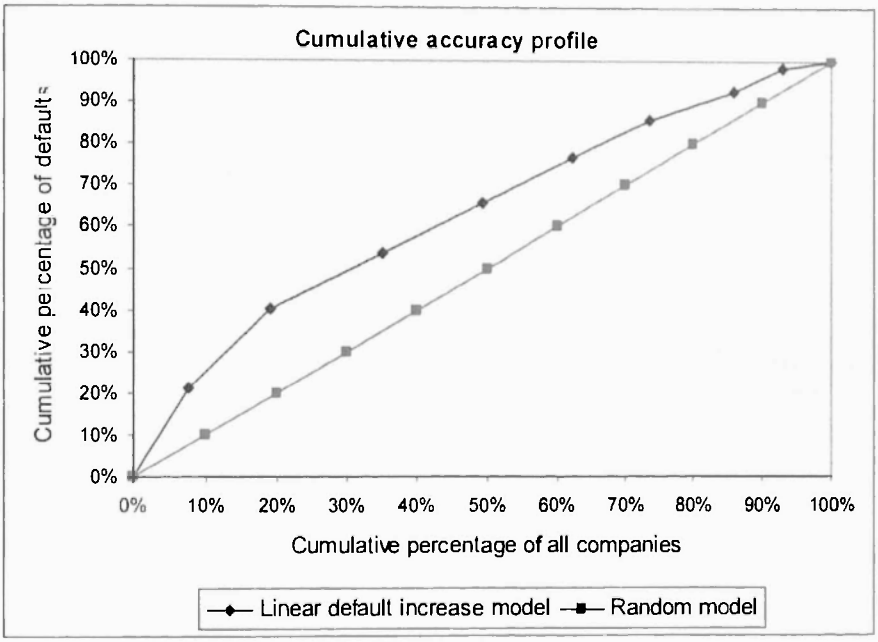
The CAPs for the two rating systems set up in this thesis are presented below.

Graph 9



The Graph 9 above shows the cumulative accuracy profile for the uniform distribution rating system. The accuracy ratio AR is equal to 37.4%. The CAP reveals that 10% of companies with the worst rating scores contain almost 30% of all defaults. Moreover, 50% of total defaults are captured in about 30% of companies with the worst scores.

Graph 10



The CAP for the linear default increase rating system (as shown on Graph 10) is very similar to the previous CAP and the AR is slightly lower: 37.1%.

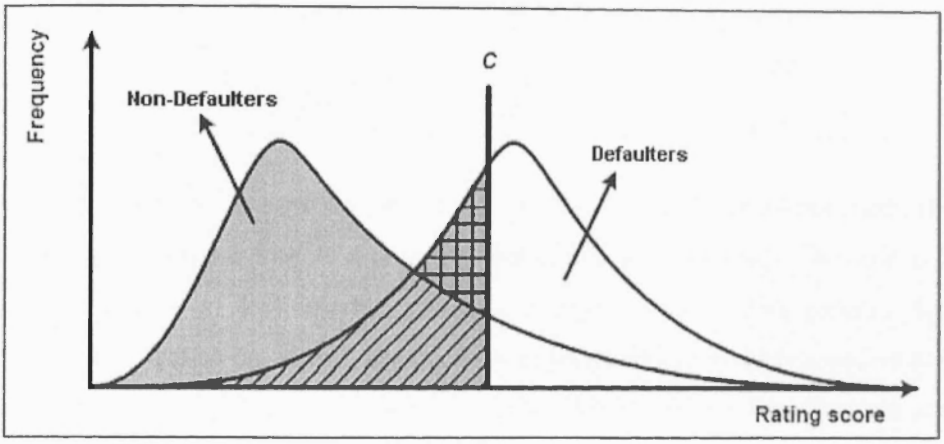
The results of the CAP analysis show that both models are quite powerful. The fact that the results for the two rating systems are very similar is in agreement with my expectations.[32] Validation of rating systems can be done in many ways and the CAP is just one of them. The next chapter introduces another validation method.

[32] For details on this matter see appendix 5.

6.1.2. Receiver Operating Characteristics (ROC)

This validation technique is based on the relative positions of the distribution of defaulters and the distribution of non defaulters. The idea is illustrated on Picture 6 below.

Picture 6



Source: BCBS (2005b), modified by the author

A perfectly discriminating rating system would give defaulters and non defaulters such scores that their distributions would not overlap. However, real rating systems in general cannot achieve perfect discrimination and, therefore, the two distributions will overlap. Picture 5 illustrates the idea.

With the Creditreform solvency index in mind, we face the decision as where to set the threshold value C in order to distinguish between potential defaulters and non defaulters. After setting this threshold all obligors with a rating score lower than C will be marked as potential non defaulters and obligors with rating score higher than C as potential defaulters.

Since we are in a two state world, there are four possible outcomes; two outcomes are correct and the other two are classified as either type I (wrongly classifying non defaulter as defaulter) or type II (wrongly classifying defaulter as non defaulter) error. All possibilities are summarized in Table 11 below.

Table 11

	Obligor subsequently	
	defaults	not defaults
Rating score above threshold C	CORRECT PREDICTION (hit)	TYPE I ERROR (false alarm)
Rating score below threshold C	TYPE II ERROR (miss)	CORRECT PREDICTION (correct rejection)

Making either type of error can be rather costly for a bank. When a bank commits a type II error then it gives a loan to a company that subsequently defaults. The cost in this case takes the form of lost interest, principal, expenses on collection process, bankruptcy proceedings and so on. Hence, banks can be expected to be extremely sensitive to this type of error. Making type I error comes at a cost too, but in this case the cost is not as obvious. The cost in this case is the foregone business (opportunity cost), i.e. the amount of profit the bank could have made if it had correctly given the credit.

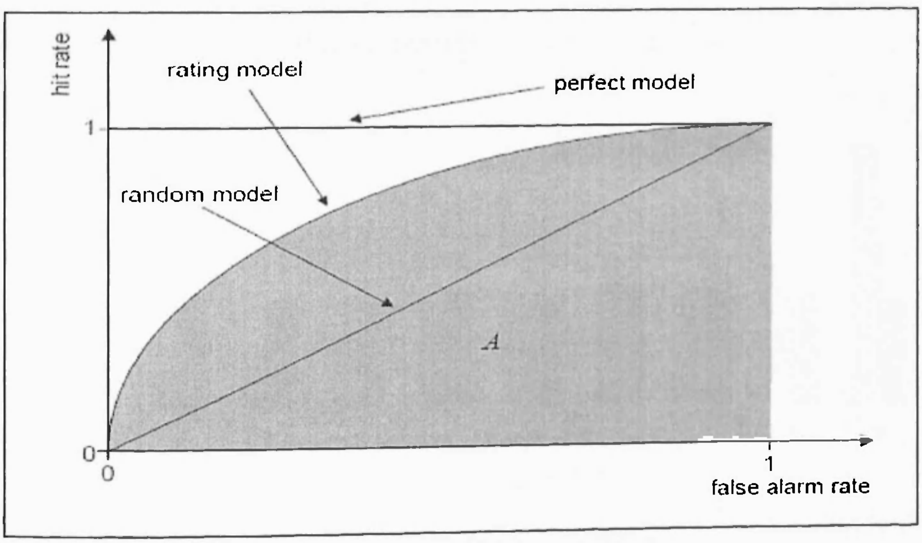
The problem is that a bank can only measure the magnitude of type II error, i.e. the default rate in the banking portfolio. Banks are not able to measure the magnitude of type I error they make. Hence, when building a rating system on a banking portfolio it is necessary to keep this selection bias in mind.

The advantage of my data set is that it comes from an external agency and is, thus, free of the selection bias. On the other hand, the overall default rate in the data set can be expected to be higher. Banks try to differentiate potential clients to "good" ones and "bad" ones and provide credit only to the "good" ones. However, no such differentiation is done by Creditreform.

Let us define the fraction of defaulters that was correctly classified for a given threshold C as the hit rate. The hit rate can be written as $HR(C) = \frac{H(C)}{N_D}$, where $H(C)$ is the number of defaulter classified correctly for a given threshold C, and N_D is the total number of defaulters in the sample. Let us define the fraction of non defaulters that were incorrectly classified as defaulters for a given threshold C to the total number of non defaulters as the false alarm rate. Then the false alarm rate $FAR(C)=\frac{F(C)}{N_{ND}}$, where $F(C)$ is the number of false alarms for a give threshold C and N_{ND} is the total number of non defaulters in the sample.

If we compute *HR(C)* and *FAR(C)* for all possible *C*s and plot them against each other we get the receiver operating characteristic curve. A rating model is all the better the larger the area under the ROC curve is. Let us denote this area by AUC (area under curve). For illustration see the Picture 7 below.

Picture 7



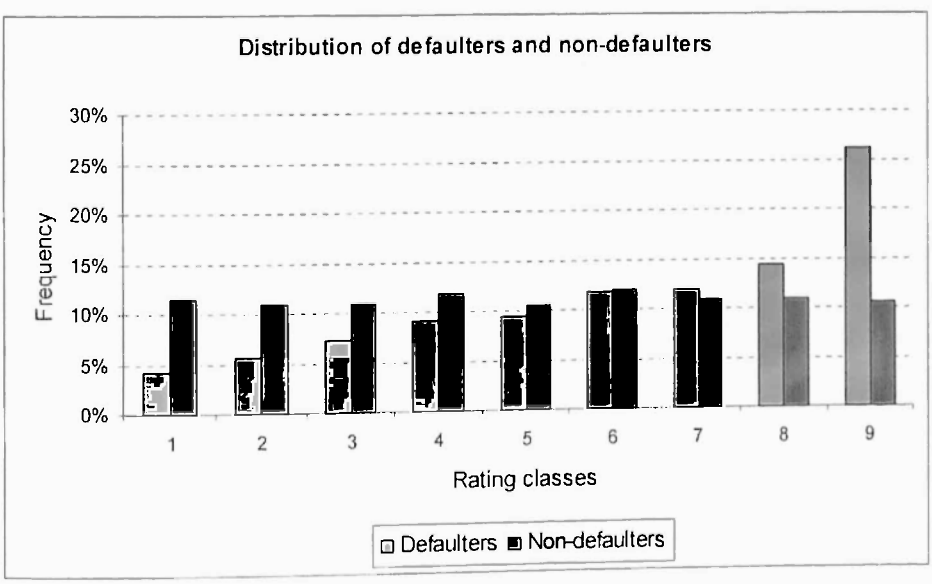
Source: BCBS (2005b)

The area under the ROC curve is equal to the following formula: $AUC = \int_0^1 HR(FAR)d(FAR)$. AUC is equal to 0.5 for a random model with no discriminative power and to 1 for a perfect model. However, as in the AR case, these values represent the extremes. Real rating systems will be in between.

The following paragraphs are devoted to application of the ROC method to the two rating systems of this thesis. The application will be done first for the uniform distribution rating system followed by the linear default increase rating system.

Before calculating the ROC for the uniform distribution rating system, it is useful to have a look on the distribution of defaulters and non-defaulters across the rating classes. A perfect rating system would separate completely the distributions of defaulters and non defaulters. So, Graph 11 below can be used as a visual tool to show the performance of the uniform distribution rating system in separation of defaulters and non defaulters.

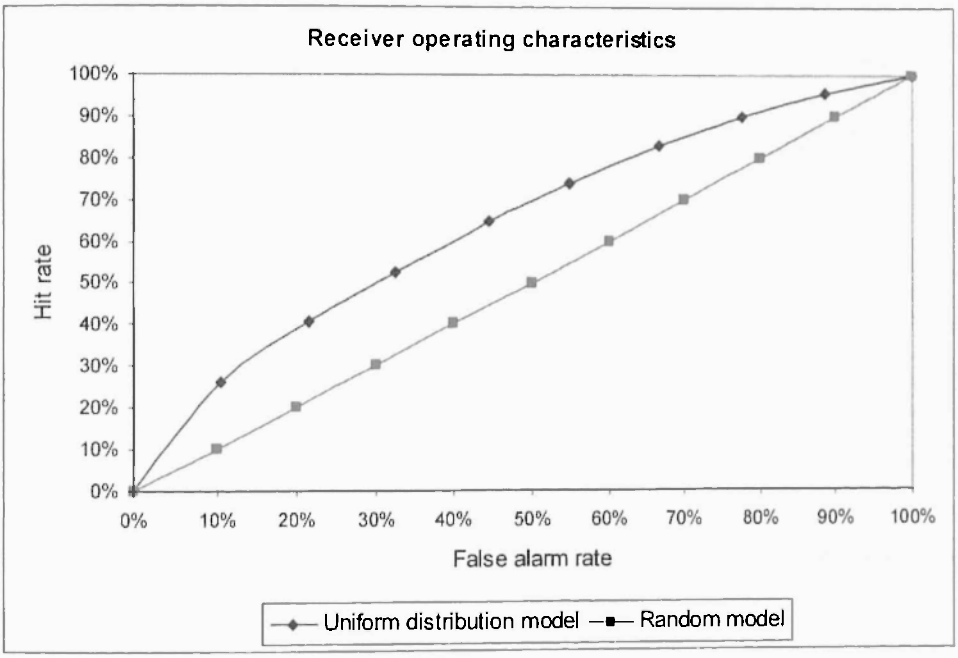
Graph 11



We observe that the number of non-defaulters is roughly constant across all rating classes whereas the number of defaults increases as we move to worse rating classes. The rating system has tendency to put more defaulters in worse rating classes.

Graph 12 below shows the ROC curve for the uniform distribution rating system.

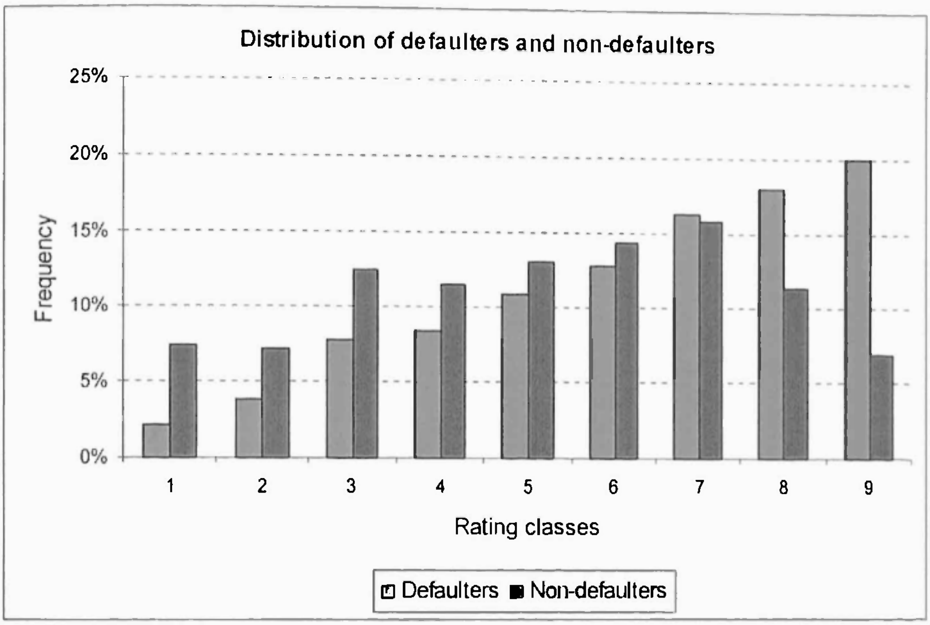
Graph 12



The area under curve is in this case equal to 64.46%.

The same exercise is performed for the linear default increase rating system. First, as is the case for the previous rating system, the distribution of defaulters and non-defaulters is presented. See Graph 13 below.

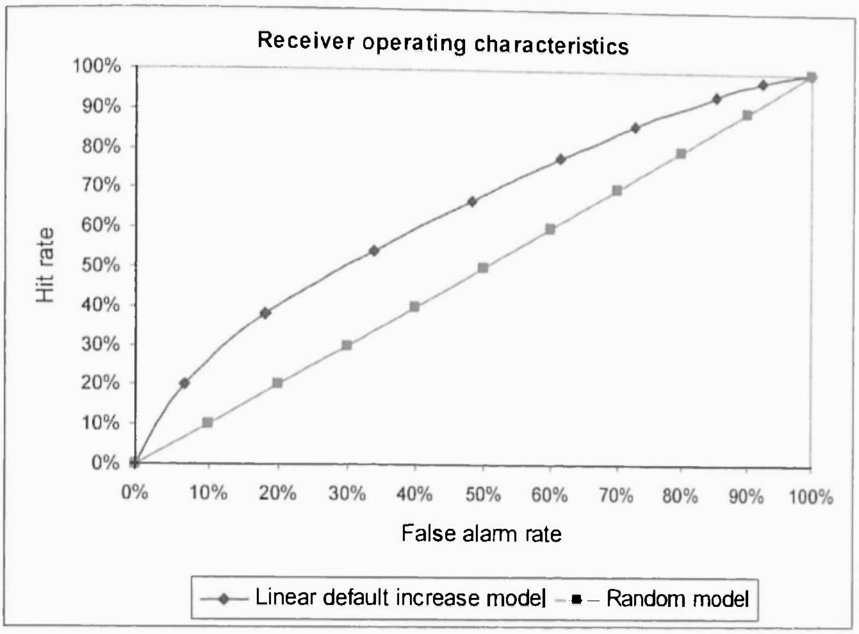
Graph 13



We can see clearly the linear increase in the relative share of defaulters as we move towards the worse rating classes since it was the condition the rating system was built on. The relative share of non-defaulters has no clear shape. Nevertheless, we observe that the rating system has tendency to separate defaulters and non defaulters even thought the separation is not perfect.

The ROC curve for the linear default increase rating system is shown on the following Graph 14.

Graph 14



The receiver operating characteristic for this rating system looks very similar to the other rating system. The area under curve is again slightly lower and is equal to 64.28%.

Again, as when assessing the discriminatory power by the CAP, very similar results are obtained. The reason for the very similar results in CAP and ROC analysis of the both rating system is that these systems are based on the same underlying information. This information is the Creditreform index. The two rating systems are just different regroupings of the Creditreform indices. These rating systems cannot discriminate any better than the Creditreform indices discriminate[33].

Although the CAP and ROC methods are based on different theoretical foundations, the results of these two validation techniques are very similar. We can see the similarities when we compare the Graphs 9 and 12 and Graphs 10 and 14. Engelmann et al. (2003) showed

---

[33] See appendix 5 for details on this matter.

that there is a relationship between the accuracy ratio and the area under curve. The relationship is following: $AR = 2AUC - 1$[34].

6.1.3. Conditional Information Entropy Ratio (CIER)

The conditional information entropy ratio is the last validation technique of the discriminatory power presented in this thesis. The CIER measures the overall amount of uncertainty represented by a probability distribution. In the rating system validation context, it is used to assess how well a rating model reduces uncertainty. We compare the amount of uncertainty in a state of total ignorance[35] (no rating model) to the amount of uncertainty left over after introduction of a rating model[36]. Since the main goal in credit risk management is to reduce uncertainty, the information entropy ratio represents a measure of how well a model is performing.

In a two state world, an obligor has only two future possibilities. Either the obligor will default with a probability p over a certain time period or the obligor will not default with a probability $1-p$ over the same time period. The probabilities of the outcomes provide partial information for the future status. Let us define the $\text{Information} = -\ln(p)$ as the amount of *additional information* needed to completely determine whether or not the obligor will default.

If it is certain that the obligor will default ($p=1$) then the amount of additional information required is equal to $-\ln(1) = 0$. Since, there is no uncertainty about the outcome; there can be no previously unknown relevant information.

---

34 For proof see Engelmann et al. (2003) appendix A.

35 In state of total ignorance only the overall PD of a portfolio can be measured. Since this is the only available information, every obligor in the portfolio is expected to default with this probability.

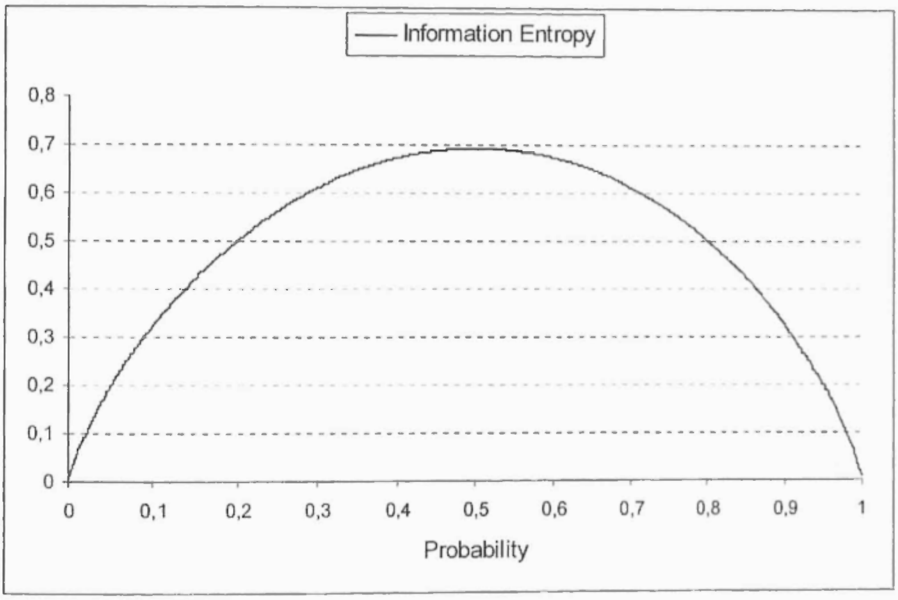
36 Now, it is possible to assign every obligor a PD of a rating class it belongs to.

Having defined this Information we can now define the information entropy $H(p)$ of an event with probability p as:

$$\mathrm{H(p)} = -(p\ln(p) + (1-p)\ln(1-p)).$$

The information entropy visualized as a function of p is presented in the following Graph 15.

Graph 15



It can be seen that the information entropy reaches maximum at $p = 0.5$. This represents the greatest uncertainty, i.e. the highest amount of additional information is needed to decide whether the obligor will default or not. If p equals to zero or one, we know with certainty whether the obligor will default or not.

The above mentioned logic can be applied to a rating system that puts each obligor into a rating class. The purpose of a rating system is to reduce uncertainty present in the portfolio via the separation of obligors into rating classes. Once we have rating classes, we can define the conditional entropy as $H_i(P(D|R_i))$, i.e. the information entropy of the conditional probability of default given rating class R_i. In line with $H(p)$ the conditional entropy can be written as:

$$H_i(P(PD|R_i)) = -(P(PD|R_i)\log(P(PD|R_i))+P((1\text{-}PD)|R_i)\log(P((1\text{-}PD)|R_i)))$$

This expression can be also rewritten as:

$$H_i = -E\left[P(PD|R_i)\log(P(PD|R_i))+P((1\text{-}PD)|R_i)\log(P((1\text{-}PD)|R_i))\right]$$

Now, we can define the Conditional Information Entropy Ratio (CIER) as:

$$CIER = 1-\frac{\sum_{i=1}^{n} H_i}{H_T},$$

where n is the number of rating classes and H_T is the entropy given only by the portfolio itself, i.e. $H_T = -(PD\log(PD)+(1\text{-}PD)\log(1\text{-}PD))$[37].

CIER is simply one minus the ratio of the entropy based on the rating classes to the entropy given only by the data itself. The value of CIER will be all the closer to one the more information about defaults is contained in the rating model. If the rating model were perfectly predictive, the CIER would equal to 1 (no uncertainty). On the contrary, if the model held no predictive power the CIER would be 0.

[37] The PD here is the overall PD in the portfolio.

In other words, the CIER shows how much information we gained by introducing the rating model that enables us to say which obligor will certainly default and which obligor will certainly not default.

Table 12 below summarizes the CIER for the two rating models introduced in this thesis.

Table 12

	Uniform distribution rating system	Linear increase default rating system
CIER	51.9%	46.9%

The CIER is a measure of the reduction of uncertainty associated with a particular model for a given portfolio. Therefore, it provides an unambiguous measure of a model performance under the condition that all the models are applied to the same portfolio. However, the CIERs for the two rating models are, again, similar not allowing to tell which rating system is superior over the other.

From the three validation methods presented so far, the cumulative accuracy profile along with the accuracy ratio are mostly used by other researchers in this field. For example, Fernandes (2005) constructed rating systems with the accuracy ratios ranging from 25% to 43.8%. These results are comparable with the results in this thesis. However, according to Moody's (2001) the accuracy ratio of its professional rating system RiskCalc exceeds 60%.

Having discriminatory power is one of the requirements of a sound rating system. However, a rating system also needs to fit the underlying data well. This is the criteria of accordance. In case of the rating systems set up in this thesis, the theoretical or expected PDs of rating classes estimated by the regression model need to be in line with the real PDs calculated directly from the data set. The analysis of accordance (or calibration) will be the subject of the following chapter.

6.2. Calibration

The focus of the previous chapter was on testing how well the two rating systems perform in discrimination between defaulting and non defaulting obligors. However, in reality, rating systems do not serve just as a basis for deciding whether to grant a loan or not. „Rather, they form the basis for pricing credits and calculating risk premiums and capital charges." (Engelmann et al. 2003, p. 28) Therefore, obligors are grouped into rating classes and the probability of default is quantitatively assessed for every rating class

This chapter is focused on the analysis of how well the estimated probabilities of default for each rating class reflect the true historical probability of default. The correct estimation of PD of rating classes is vital for banks adopting the IRB approach since it, according to Engelmann et al. 2003, influences their amount of regulatory capital. Hence, banks will have to devote a lot of time to calibration of their rating systems. Correct calibration of a rating system means that that the PD estimates are accurate.

Testing of correctness of a calibration can be done separately for each rating class (binomial test) or in a joint test (Hosmer-Lemeshow test). The caveat is that defaults are not independent, they tend to be correlated. As a consequence, standard independence based tests are likely to be biased when applied to credit portfolios.

In this chapter, the two calibration tests – binomial and Hosmer-Lemeshow are explained and subsequently applied.

6.2.1. Binomial test

Undergoing this test, we form the two following hypothesis:

H0: the PD of a rating class is correct

H1: the PD of a rating class is underestimated

Given a confidence level, the H_0 is rejected if the number of historical defaults k in a particular rating class is greater than or equal to a critical value c.

The approximation formula of the binomial test that takes the following form (see Engelmann et al. (2003)) is used for calculations[38].

$$c_i = \Phi^{-1}(q)\sqrt{n_i * PD_i(1 - PD_i)} + n_i * PD_i$$

where $i = 1, \ldots, 9$ represents the rating classes, n_i stands for the number of obligors in class i, PD_i is the probability of default of class i and Φ^{-1} is the inverse of standard normal distribution function.

BCBS (2005b) argues that if correlation effects were accommodated, higher values for c_i would be obtained. However, the problem is the estimation of the correlations. Modeling correlations of defaults is a very complex issue and is beyond the scope of this thesis. Nevertheless, the binomial test can serve as an early warning system. With the assumption of no correlation, c_i values are lower and thus H_0 will be rejected more often. Hence, if H_0 is rejected for a particular rating class, further research is necessary.

Results for the two rating systems are presented in Tables 13 and 14 below:

[38] Note that this set up of the binomial test is simplified since the correlations among default events are not taken into account.

Table 13

Uniform distribution rating system				
Rating class	Estimated PD	Number of defaults - k	Critical value - c	Null hypothesis
1	1.75%	67	85	NOT REJECTED
2	2.43%	99	111	NOT REJECTED
3	3.09%	116	139	NOT REJECTED
4	3.61%	167	172	NOT REJECTED
5	4.15%	130	177	NOT REJECTED
6	4.57%	175	219	NOT REJECTED
7	5.04%	162	221	NOT REJECTED
8	5.97%	234	263	NOT REJECTED
9	10.71%	437	461	NOT REJECTED

Table 14

Linear default increase rating system				
Rating class	Estimated PD	Number of defaults - k	Critical value - c	Null hypothesis
1	1.39%	26	48	NOT REJECTED
2	2.50%	90	78	*REJECTED*
3	2.94%	109	149	NOT REJECTED
4	3.42%	144	160	NOT REJECTED
5	3.88%	171	202	NOT REJECTED
6	4.11%	195	235	NOT REJECTED
7	4.76%	216	295	NOT REJECTED
8	7.10%	301	322	NOT REJECTED
9	12.26%	335	356	NOT REJECTED

The rating system with the uniform distribution of companies within rating classes passes the binominal test for all rating classes. The H_0 of correctness of the estimated PD cannot be rejected for any rating class.

The rating system with linear increases of defaults passes the test in all but one rating class. The PD of rating class no. 2 is underestimated since we reject H_0 in favor of H_1 for this rating class. However, as suggested above, since the correlations of defaults are not taken into account the critical value c_i is underestimated. Thus, it is possible that we could not have been able to reject H_0 if we took the correlations into account.

Based on the binomial test, no clear decision about which of the two rating systems presented in this thesis is favorable over the other can be made. To make a valid judgment the correlations of defaults would need to be estimated to ensure correct critical values c_i. Still, the binomial test serves as an early warning mechanism indicating that the criteria of accordance might not be fulfilled by the linear default increase rating system.

6.2.2. Hosmer-Lemeshow test[39]

Let us define test statistic:

$$T_k = \sum_{i=1}^{k} \frac{(n_i p_i - D_i)^2}{n_i p_i (1 - p_i)}$$

where $k = 9$ (number of rating classes), n_i = number of companies in rating class i, D_i is the number of defaulted obligors in class i, p_i is the forecasted probability of default for rating class i.

The p-value of χ^2_{k+1} test serves as a measure of the accuracy of the estimated default probabilities. However, there is no critical value of p that could be used to determine whether the estimated PDs are correct or not. Simply put, the closer the p-value is to zero the worse the estimation is. Nevertheless, this test can be used for comparison of rating systems.

The results for the two rating systems are summarized in Table 15 below:

[39] For details on this test see BCBS (2005b).

Table 15

	Linear default increase model	Uniform distribution model
p-value	0.001	0.181

We can conclude that the uniform distribution model is superior. This finding was indicated by the binomial test presented in previous chapter.

Before using a rating system for calculating the minimal capital requirements for credit risk it is necessary to validate the rating system. A sound rating system has to comply with criteria of good discriminatory power and good calibration.

The widely accepted and used validation techniques were subject of this chapter. Cumulative accuracy profile, receiver operating characteristics and conditional information entropy ratio were introduced to asses the discriminatory power of the rating systems whereas the binomial and Hosmer-Lemeshow tests were presented in order to check calibration (accordance) of the rating systems.

Both rating systems performed well in the discriminatory tests. In the case of calibration the Hosmer-Lemeshow test revealed that the uniform distribution rating system is favorable over the linear default increase rating system. Therefore, the uniform distribution rating system will be used for calculating the regulatory capital requirement by the internal ratings based approach.

7. Regulatory capital

Under the New Basel Capital Accord (NBCA), banks will be able to use the internal credit risk assessment of their obligors in order to determine the minimal regulatory capital requirements. The first pillar of the NBCA defines two broad methods for calculating the credit risk regulatory capital. These two methods, the standardized approach and the internal ratings based (IRB) approach, were described in detail in chapters 2 and 3 of this thesis.

Up to now, several alternative methods for determining one of the key risk components – the probability of default (PD) were developed. A logit regression model was set up to estimate PDs of all companies in the data set. Several alternative regression models were proposed, tested and the best one chosen. Two alternative approaches to calibration of a rating system were also suggested. Once more, the two calibrations underwent excessive validation and the one with better results was chosen. At this point, we are ready to make regulatory capital estimations according to the standardized and the foundation IRB approaches.

The Creditreform database contains turnover information for all companies that were assigned the solvency index in the year 2004. Under the assumption that every company has a bank credit of 10% of its turnover, the exposure at default (EAD) for every company in the year 2004 was calculated (8 348 companies in total). Hence, the regulatory capital requirements calculated in this thesis are for the year 2005.

In order to prove the central hypothesis of this thesis the regulatory capital was calculated according to both NBCA methods (standardized approach and the IRB approach).

7.1. Standardized approach

Under the standardized approach, every obligor is assigned a risk weight (RW) according to its rating from an external company[40]. See Picture 8 for details. Regulatory capital for every exposure is calculated according to the following formula:

$$\text{Regulatory capital} = \text{EAD}*\text{RW}*0.08$$

The regulatory capital on the whole portfolio is simply the sum of regulatory capital on every single exposure.

Picture 8

Credit assessment	AAA to AA-	A+ to A-	BBB+ to BB-	Below BB-	Unrated
Risk weight	20%	50%	100%	150%	100%

Source: BCBS (2005a)

For the computation of the regulatory capital on the portfolio, we can either consider all obligors as unrated (with 100% risk weight) or we can map the rating classes of the rating system developed in this thesis to ratings of an external rating company. Note that under the assumption of unrated companies, the resulting capital requirements will be the same as if calculated according to the Basel I approach. Both of these methods, i.e. the unrated standardized NBCA / Basel I approach and the mapping method, are applied.

For the mapping procedure, estimates of the probability of default of external ratings are necessary. Hanson and Schuermann (2005) estimated annual probabilities of default for S&P rated U.S. obligors from year 1981 to 2002. For details, see the Table 16 below. These PD estimates are used to map the rating classes to the S&P ratings.

[40] The risk weights are assigned according to the BCBS (2005a), paragraph 66.

Table 16

S&P ratings	PD estimates	S&P ratings	PD estimates	S&P ratings	PD estimates
AAA	0.00%	A-	0.06%	BB-	2.07%
AA+	0.00%	BBB+	0.31%	B+	3.50%
AA	0.00%	BBB	0.36%	B	9.82%
AA-	0.04%	BBB-	0.40%	B-	14.30%
A+	0.05%	BB+	0.55%	CCC	28.53%
A	0.07%	BB	1.16%		

The results of the mapping procedure are summarized in the Table 18 below:

Table 17

	Creditrefom index range	PD	Corresponding S&P rating
Rating class 1	100-213	1.75%	BB-
Rating class 2	214-248	2.43%	B+
Rating class 3	249-264	3.09%	B+
Rating class 4	265-279	3.61%	B
Rating class 5	280-289	4.15%	B
Rating class 6	290-297	4.57%	B
Rating class 7	298-306	5.04%	B
Rating class 8	307-334	5.97%	B
Rating class 9	335-499	10.71%	B

We can directly observe that only obligors in the first rating class will be assigned risk weight 100% (501 companies) while all the other obligors will have 150% risk weight (7 847 companies). The resulting regulatory capital requirements are summarized in the following Table 18.

Table 18

Capital requirements (mio. CZK) - standardized approach	
All obligors unrated / Basel I approach	31 983,30
Mapping to S&P ratings	42 526,71

It is not surprising that under the mapping method, the capital requirements are higher since most of the obligors in the portfolio are assigned risk weight 150% instead of 100% as is the case in the unrated method or the Basel I method. Considering the whole portfolio, the EAD weighted average risk weight is equal to 133%, resulting in higher regulatory capital. The EAD weighted average risk weight for the unrated method is, by definition, equal to 100%.

Note that calculating capital requirements for every exposure and then summing up to obtain the capital requirement for the whole portfolio is equivalent to multiplying the 8% of total EAD for all obligors by the EAD weighted average risk weight.[41]

7.2. *Internal ratings based approach*

As the IRB approach was described in detail in chapter 3, only the calculations are performed in this chapter. The regulatory capital calculated by the IRB approach is equal to *37 206.34* million CZK. All capital requirements are compared in the following Table 19.

Table 19

Capital requirements (in mio. CZK)	
Unrated / Basel I	31 983.30
Mapping to S&P	42 526.71
IRB	37 601.10

The results show that the capital requirements on the portfolio are lowest for the unrated / Basel I approach followed by the IRB approach. When the rating classes are mapped to the

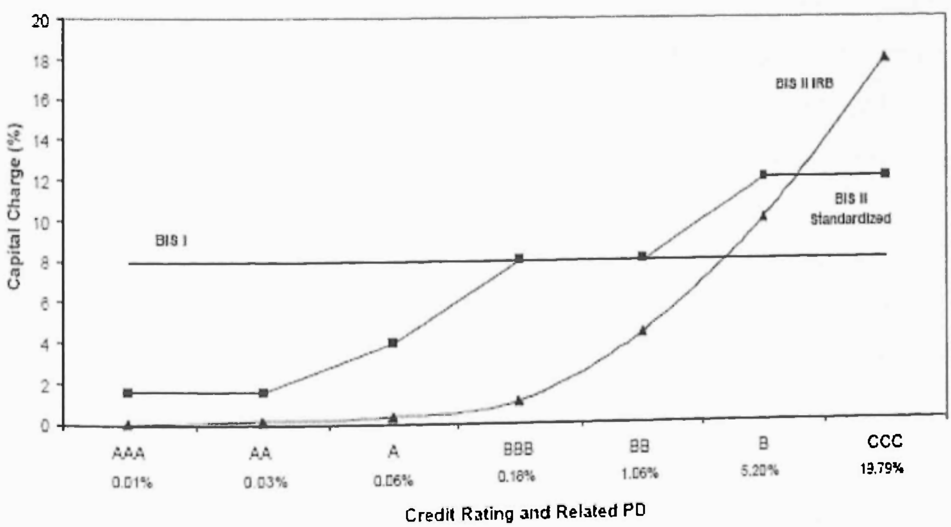
[41] The proof to this statement is straight forward. It is just necessary to realize that the EAD weighted average risk weight is defined as $\sum_i (RW_i * \frac{EAD_i}{\sum_i EAD_i})$.

S&P ratings the highest amount of regulatory capital is obtained. The application of the IRB approach yields higher regulatory capital requirements when compared to the unrated standardized NBCA / Basel I approach. The difference makes 5 617.8 million CZK. However, if the rating classes were mapped to S&P ratings then the IRB approach would result in lower regulatory capital requirements. The difference in this case amounts to 4 925.61 million CZK.

Hence, the data set used in this thesis does not allow for proving the hypothesis that the IRB approach yields lowers regulatory capital requirements than the Basel I approach. In the case of the standardized NBCA approach, the results are ambiguous. Whether the hypothesis is proven or not depends on what is taken for benchmark.

Stephanou and Mendoza (2005) show the capital requirements for a single exposure under the Basel I approach and the standardized and the IRB approach of the NBCA and come to the conclusion illustrated by the following Picture 9.

Picture 9



Source: Stephanou and Mendoza (2005)

Even though Picture 9 shows capital charge for a single exposure under the three approaches, it is possible to generalize the results for a whole credit portfolio if it is considered a structured single exposure. For the data set used in this thesis, this analysis is carried out in the following paragraph.

Under the Basel I approach, the capital charge for the whole credit portfolio will always stay at *8%* since this approach is not sensitive to risks of the portfolio. Considering the mapping approach described in the previous paragraphs, the capital charge equals to *10.64%*[42]. For the IRB approach, it is necessary to find an average PD for the whole portfolio so that the minimal regulatory capital calculated for every single exposure and then summed up would equal the minimal regulatory capital calculated for the total exposure[43]. The average PD for the whole portfolio was found using the iterative method and equals to 3.04%. With the average PD at hand, it is possible to calculate the capital charge under the IRB approach[44]. In this case, the capital charge is equal to *9.41%*. When looking at Picture 8, we see that the data set used in this thesis represents a credit portfolio with a rating somewhere between BB and B (using notation from Picture 9).

Also Schwaiger (2003) shows that a lower quality portfolio will face higher capital charge whereas a higher quality portfolio will face lower capital charges. Hence, the data set used in this thesis contains companies with higher credit risk, resulting in higher regulatory capital charges compared to the unrated standardized NBCA / Basel I approach.

According to the author's knowledge, the research on credit risk modeling using real data is limited, probably because of the lack of public data. From recent studies, Fernandes (2005) calculated minimal capital requirements for a data set of private firms bank loans of a

---

[42] The capital charge is calculated as 8% times the EAD weighted average risk weight (1.33).

[43] Simply put, we need to solve the following equation for PD:

$$\sum_i (8\% * K(PD_i, LGD, M) * 12.5 * EAD_i) = 8\% * K(PD, LGD, M) * 12.5 * \sum_i EAD_i .$$

[44] The capital charge is calculated as $8\% * K(PD, LGD, M) * 12.5 = K(PD, LGD, M)$.

Portuguese bank with the conclusion that the IRB approach yields lower regulatory capital requirements compared to the Basel I approach. Kadlcakova and Keplinger (2004) compare capital requirements for a Creditreform data set under the IRB approach and other credit risk models. However, they do not present results for the standardized NBCA approach.

8. Conclusion

As was first stated in the introduction, the banks' central role in the economic system makes it necessary for them to be subject to an extensive regulation in order to prevent severe distortions in the economy. In order to do so, it was essential to standardize the behavior of all internationally active banks. This was achieved by the Basel Capital Accord (also known as Basel I) introduced in 1988 that is now being replaced by the *New Basel Capital Accord* (also known as Basel II). Both these accords were presented extensively in the past chapters.

The central hypothesis that I tried to empirically prove in this thesis was that the IRB methodology brings benefits in form of lower regulatory capital held for credit risk compared to the Basel I methodology and the standardized approach of the NBCA.

In this thesis, the methodology chosen to achieve this goal was to elaborate a rating system of my own, and in that regard, my proceedings were as follows.

First of all, general overview of both Basel Capital Accords was provided. When describing the NBCA emphasis was put on the first pillar that deals with capital requirements. Detailed analysis of the internal ratings based approach was also conducted along with detailed description of the underlying model and methodology accompanied by theoretical argumentation in favor of the IRB approach over the standardized one.

Finding a valid data set on which to base my rating system was critical, and the data set was used to create the system itself. All steps of the procedure were explained in detail. The core of the rating system is a logit regression model. Two alternative regression models were introduced, statistically tested and the superior one was chosen. One of the most important steps, setting up the rating classes, was performed in two alternative ways since there is no natural optimal solution. Hence, two alternative rating systems were created.

After the rating systems were outlined, the next necessary step was the validation. Here, the concept of discriminatory power was introduced. Cumulative accuracy profile, followed by receiver operating characteristics and conditional information entropy ratio respectively were outlined and used for validation of the two rating systems constructed in this thesis.

After the validation of rating systems, calibration, the next necessary step, was performed. Here, binomial and Hosmer-Lemeshow tests were utilized in order to check sound calibration of the two rating systems. The results of the test for the two rating systems were compared and the better one was chosen for calculation of the capital requirements under the IRB approach.

The validated rating system was used to calculate the regulatory capital for credit risk. Regulatory capital calculations according to the Basel I and the standardized and internal ratings based approach of the NBCA were conducted.

As a conclusion, the central hypothesis of this thesis was not proven, since the capital requirements under the IRB approach yielded higher regulatory capital than the Basel I or standardized NBCA approach (with all companies assumed unrated). However, when the rating classes of the rating system developed in this thesis were mapped to external ratings, higher regulatory capital charges were obtained by the NBCA standardized approach compared to the IRB approach.

The findings in this thesis are in line with other research in this field stating that under the IRB approach lower quality portfolios should face higher capital charges. The data set used in my work contained companies with higher credit risk resulting in higher regulatory capital charges.

The whole exercise of building an internal rating system was undertaken with the aim to understand and, hence, present the system of a bank's internal mechanisms, which are used to control and mitigate credit risk. This analytical approach was chosen with the goal of applying the theoretical knowledge to real-life situation and understanding better internal bank procedures.

9. References

Austrian Financial Market Authority and Oesterreichische Nationalbank (2004). New Quantitative Models of Banking Supervision. Oesterreichische Nationalbank.

Basel Committee on Banking Supervision (2005a). International Convergence of Capital Measurement and Capital Standards (A Revisited Framework). Bank for International Settlements.

Basel Committee on Banking Supervision (2005b). Studies on the Validation of Internal Rating Systems. Working Paper No. 14, Bank for International Settlements.

Basel Committee on Banking Supervision (2001). Basel II: The New Basel Capital Accord - Second Consultative Paper. Bank for International Settlements.

Blochwitz, S., Hohl, S., Tasche, D. and Wehn, C. S. (2004). Validating Default Probabilities os Short Time Series. Working Paper, Deutsche Bundesbank.

Bluhm, C., Overbeck, L. and Wager C. (2003). An Introduction to Credit Risk Modeling. Financial Mathematics Series, Chapman & Hall/CRC, London, UK. ISBN 1-58488-326-X (alk. paper)

Carlig, K., Jacobson, T., Linde, T. and Roznbach, K. (2002). Capital Charges under Basel II: Corporate Credit Risk Modelling and the Macro Economy. Working Paper No. 142, Sveriges Riksbank.

Crouhy, M. (2000). A Comparative Analysis of Current Credit Risk Models. Journal of Banking and Finance No. 24.

Dervitz, A. and Kadlcakova, N. (2001). Methodological Problems of Quantitative Credit Risk Modeling in the Czech Economy. Working Paper No. 39, Czech National Bank.

Dierick, F., Pires, F., Scheicher, M. and Spitzer, K.G. (2005). The New Basel Capital Framework and its Implementation in the European Union. Occasional Paper Series No. 42, European Central Bank.

Dowd, K. (2002). Beyond Value At Risk – The New Science of Risk Management. John Wiley & Sons, West Sussex, UK. ISBN 0-471-97621-0 (hp).

Elizalde, A. and Repullo, R. (2005). Economic and Regulatory Capital in Banking: What is the Difference? CEMFI.
http://www.abelelizalde.com/pdf/economic%20regulatory%20actual.pdf

Engelmann, B., Hayden, E. and Tasche, D. (2003) Measuring the Discriminative Power of Rating Systems. Discussion Paper No. 01/2003, Deutsche Bundesbank.

Fernandes, J. E. (2005). Corporate Credit Risk Modeling: Quantitative Rating System and Probability of Default Estimation.
http://pwp.netcabo.pt/jed_fernandes/JEF_CorporateCreditRisk.pdf

Giesecke, K. (2004). Credit Risk Modeling and Valuation: An Introduction. Cornell University.

Hamerle, A., Liebig, T. and Rosch, D. (2003). Credit Risk Factor Modeling and the Basel II IRB Approach. Discussion Paper No. 02/2003, Deutsche Bundesbank.

Hanson, S. and Schuermann, T. (2005). Confidence Intervals for Probabilities of Default. Federal Reserve Bank of New York.

Henneke, J. and Truck, S. (2005). Capital Requirements for SMEs Under the Revised Basel II Framework. Universitat Karlsruhe.

Jacobson, T., Linde, J. and Roszbach, K. (2004). Credit Risk versus Capital Requirements Under Basel II: Are SME Loans and Retail Credit Really Different? Working Paper No. 162, Sveriges Riksbank.

Kadlcakova, N. and Keplinger, J. (2004). Credit Risk and Bank Lending in the Czech Republic. Working Paper No. 6/2004, Czech National Bank.

Kern, M. and Bernd, R. (2001). Comparative Analysis of Alternative Credit Risk Models – An Application on German Middle Market Loan Portfolios. Working Paper No. 2001/03, Center for Financial Studies an der Johann Wolfgang Goethe-Universitat.

Lando, D. and Skodeberg, T. M. (2002). Analyzing Rating Transitions and Rating Drift with Continuous Observations. University of Copenhagen.

Moody's (2000). RiskCalc™ For Private Companies II: More Results and The Australian Model. Moody's Investor Service.

Moody's (2001). RiskCalc™ For Private Companies: The German Model. Moody's Investor Service.

Moody's (1999). Measuring Private Firm Default Risk. Moody's Investor Service.

Nickell, P., Perraudin, W. and Varotto, S. (2005). Ratings-based Credit Risk Modelling: An Empirical Analysis. Moody's-KMV. Imperial College, UK.

Odera, G., Dacoroga, M. M. and Jung, T. (2002). Credit Risk Models Do They Deliver Their Promises? A Quantitative Assessment. Economic Notes.

Pelizzon, L. and Schaefer, S. (2005). Pillar 1 vs. Pillar 2 Under Risk Management. Working Paper 11666, National Bureau of Economic Research.

Roy, P. V. (2005). Credit Ratings and The Standardized Approach to Credit Risk in Basel II. Working Paper Series No. 517, European Central Bank

Saidenberg, M. and Schuermann, T. (2003). The New Basel Capital Accord and Questions for Research. Working Paper, Federal Reserve Bank of New York.

Saunders, A. (1999). Credit Risk Measurement – New Approaches to Value at Risk and Other Paradigms. John Wiley & Sons, New York, USA. ISBN 0-471-35084-2 (cloth).

Schuermann, T. and Hanson, S. (2004). Estimating Probabilities of Default. Staff Reports No. 190, Federal Reserve Bank of New York.

Schuermann, T. and Jafry, Y. (2003). Measurement and Estimation of Credit Migration Matrices. Federal Reserve Bank of New York.

Schwaiger, Walter S. A., Jankowitsch, R. and Pichler, S. (2004). Modelling The Economic Value of Credit Rating Systems. Vienna University of Technology.

Schwaiger, Walter S. A., Jankowitsch, R. and Pichler, S. (2003). Rating Granularity and Basel II Capital Requirements. Vienna University of Technology.

Stephanou, C. and Mendoza, J. C. (2005). Credit Risk Measurement Under Basel II: An Overview and Implementation Issues for Developing Countries. Working Paper 3556, World Bank.

Tasche, D. (2003). A Traffic Lights Approach to PD Validation. Working Paper, Deutsche Bundesbank.

Appendix 1

The Value at Risk (VaR)[45] methodology was developed in the late 1970s and 1980s by major financial institutions seeking a model to measure aggregate risks across their entire portfolios. VaR is a single summary statistical measure of possible portfolio losses. However, it is necessary to bear in mind that the VaR is based on some estimation procedure and is, therefore, an estimation itself and that it may differ from the true VaR.

One of the advantages of VaR is that it can be applied to a variety of risks such as, credit risk, market risk and operational risk. The resulting VaRs also provide a common and consistent measure. Hence, different types of risk can be directly compared.

Simply put, VaR refers to the maximum amount of money that is likely to be lost over certain period[46], at some confidence level. A 99% confidence level can be chosen, in which case the VaR estimate covers all but the largest 1% of losses. Choosing the 99% confidence level is equal to saying that in one year out of 100, a loss greater than the VaR estimate is expected. Alternatively, it can be said than in 99 years out of 100 the maximum expected loss is less or equal to the VaR estimate.

The model underlying the internal rating based approach works with 99.9% confidence level. Hence, the regulatory capital of a bank calculated by the IRB approach should be sufficient to cover all unexpected losses of the bank in 999 years out of 1000.

In order to employ the VaR measure on a certain portfolio, it is necessary to have a distribution function of expected values of the portfolio. The distribution function can be

---

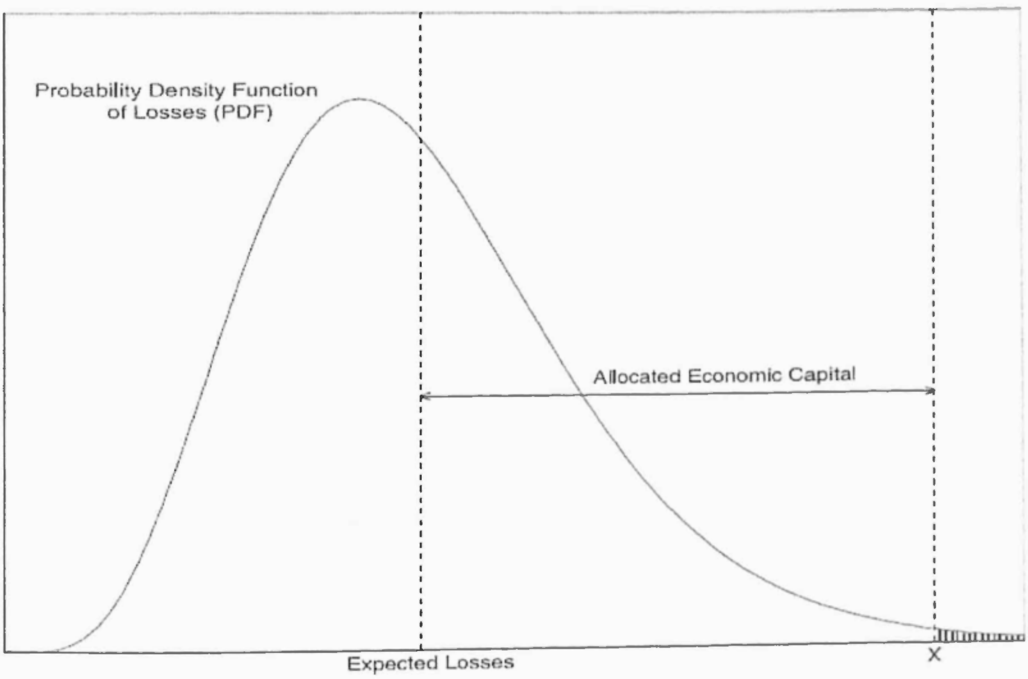
[45] "The Value at risk (VaR) is the maximum expected loss over a given horizon period at a given level of confidence." (Dowd 2002, p. 39)
[46] In the credit risk context, the period is considered to be one year.

based on an analytical approach[47] or simulated by the Monte Carlo simulation. Once the distribution function is known, the value for the selected percentile, for example, the 99th percentile can be calculated to get the VaR measure.

The regulatory capital a bank needs to hold for credit risk then correspond to the VaR minus the mean of the distribution function. The mean of the distribution represents the expected losses and, under the IRB methodology, banks are supposed to cover for expected losses themselves by provisions, etc. Hence, the regulatory capital serves as a cushion against unexpected losses in credit portfolio.

Perhaps it is better to illustrate the VaR by the following Picture 10.

Picture 10



Source: Stephanou and Mendoza (2005)

[47] This is the case of the model underlying the IRB approach.

The probability density function represents the distribution of losses of a bank's credit portfolio and the selected confidence interval corresponds to the loss of value X. As shown on the picture, the regulatory capital is equal to X minus the expected losses.

Appendix 2[48]

An example calculation of the Creditreform solvency index is presented in this appendix. When calculating the solvency index, certain pieces of information concerning a company are taken into account. For details, see Table 20 below.

Table 20

Legal form		Limited liability					
Business sector		Construction					
Age of company		12 years					
Company's business development		Constant					
Order book situation		Good					
Mode of payments		In line with agreed terms					
Credit opinion		Relationship admissible					
Risk factors	Weights in %	Classification					
		1	2	3	4	5	6
Mode of payments	20		40				
Credit opinion	25		50				
Business development	8			24			
Order book situation	7		14				
Legal form	4				16		
Business sector	4			12			
Age of company	4		8				
Turnover	2			6			
Turnover per employee	4			12			
Number of employees	2				8		
Equity	4		8				
Capital turnover	4			12			
Payment moral	4		8				
Payment moral of customers	4		8				
Shareholder structure	4			12			
TOTAL	100		136	78	24		
Solvency Index	**238**						

[48] The example was obtained from the Creditreform webpage - http://www.creditreform.cz/bonity002.htm [10-02-2005] translated by the author.

Appendix 3[49]

Logit regression is useful to model relationships where the dependent variable assumes only two states (in the case of modeling defaults 1 stands for default whereas 0 stands for non default) and the independent variables are of any type. Logit regression estimates the probability of an event occurring. The dependent variable is transformed into a logit variable and maximum likelihood is applied for computation. Compared to ordinary least squares regression, logit regression estimates changes in the log odds of the dependent variable, not changes in the dependent itself.

Let y_i be a binary variable assuming 0, if company i has not defaulted over a given period of time, and 1 otherwise. Let the vector $x_i^1, x_i^2, \ldots, x_i^k$ represent the k explanatory variables. The goal of logistic regression is to model the conditional probability that a company i defaults, i.e. $\Pr(y_i = 1 | x_i^k) = \pi(x_i^k)$ and the conditional probability that the company does not default, i.e. $\Pr(y_i = 0 | x_i^k) = 1 - \pi(x_i^k)$. The odds that company i defaults is then:

$$\text{odds} = P/(1-P) = \frac{\pi(x)}{1-\pi(x)}.$$

Hence, the log odds or logit are equal to:

$$g(x,\beta) = \ln(P/(1-P)) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k.$$

or,

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}.$$

[49] This appendix is based on Fernandes (2005).

Furthermore, after the estimation of the vector of the parameters $\beta_0 + \beta_1 x_1 + ... + \beta_k x_k$ [50], it is necessary to map the vector for $i = 1,...,n$ (n being the number of companies) to a $[0,1]$ space. In the rating system context, this space can be interpreted as the probability of default. The mapping function has the following form:

$$F(x,\beta) = \frac{\exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)}.$$

Tests

Statistical tests relevant for the particular logit regression model used in this thesis are described. Since there is only one independent variable in the logit regression model, tests for multicolinearity are not described.

Significance of coefficients

The Wald Chi-square test can be used for testing the statistical significance of the individual coefficients. The hypothesis H_0 stands for $\beta_i = 0$ and the test statistic takes the following form:

$$W_i = \frac{\hat{\beta}_i^2}{\hat{SE}(\hat{\beta}_i)^2}$$

and follows a Chi-square distribution with one degree of freedom.

[50] In the rating system context, the value of this vector is called the score of a rating system.

The Hosmer-Lemeshow test

The Hosmer-Lemeshow goodness-of-fit test is applied in order to check how effectively the estimated model describes the dependent variable. The predicted probabilities are divided into deciles (g = 10). Let o_i^0 be the observed count of non-defaults for group i and o_i^1 the observed count of defaults. Similarly, let p_i^0 be the predicted count of non-defaults for group i and p^0 the predicted count of defaults. The HL statistic then follows a Chi-square distribution with g - 2 degrees of freedom.

$$HL = \sum_{i=1}^{g} \left[\frac{\left(o_i^0 - p_i^0\right)^2}{p_i^0} + \frac{\left(o_i^1 - p_i^1\right)^2}{p_i^1} \right].$$

The Box-Tidwell Test

A logit regression will lack power if the assumption of linearity is violated. The relationship between the dependent and independent variables will be underestimated, so the Type II error will increase (assuming no relationship when there actually is). The Box-Tidwell test consists of adding term $x_k \ln(x_k)$ to the regression model. If any of these cross-product terms are significant then there is evidence of nonlinearity in the logit regression model.

Appendix 4[51]

For the linear default increase rating system, the thresholds are set so that the number of expected defaults increases linearly from the best rating class to the worst. To achieve this goal, it is necessary to solve the two following equations:

$$\alpha * i = A_i$$

$$\sum_{i=1}^{k} A_i = 1.$$

where i represents the index of a rating class, A_i is the percentage of all defaults in rating class i and k stands for the number of rating classes.

When solving for α, we get:

$$\alpha = \frac{2}{k*(k+1)}.$$

Hence, for nine rating classes α is equal to:

$$\alpha = \frac{2}{9*(9+1)} = \frac{1}{45} = 0.0222.$$

Therefore, the distribution of defaults over rating classes looks as represented on the following Table 21:

[51] This appendix is based on Schwaiger (2004).

Table 21

	Percents of total defaults
Rating class 1	2.22%
Rating class 2	4.44%
Rating class 3	6.66%
Rating class 4	8.88%
Rating class 5	11.10%
Rating class 6	13.32%
Rating class 7	15.54%
Rating class 8	17.76%
Rating class 9	19.98%

Appendix 5

It is possible to apply the discrimination methodologies (CAP and ROC) directly to the Creditreform solvency index. By doing so, we can directly observe the discriminatory power of the solvency index and compare it with that of the rating systems set up in this thesis. The hypothesis is that by aggregating the indices into rating classes, lower discriminatory power is obtained.

Since the two rating systems presented in this paper yield very similar results, the comparison will be carried out only between the Creditreform solvency index and the uniform distribution rating system.

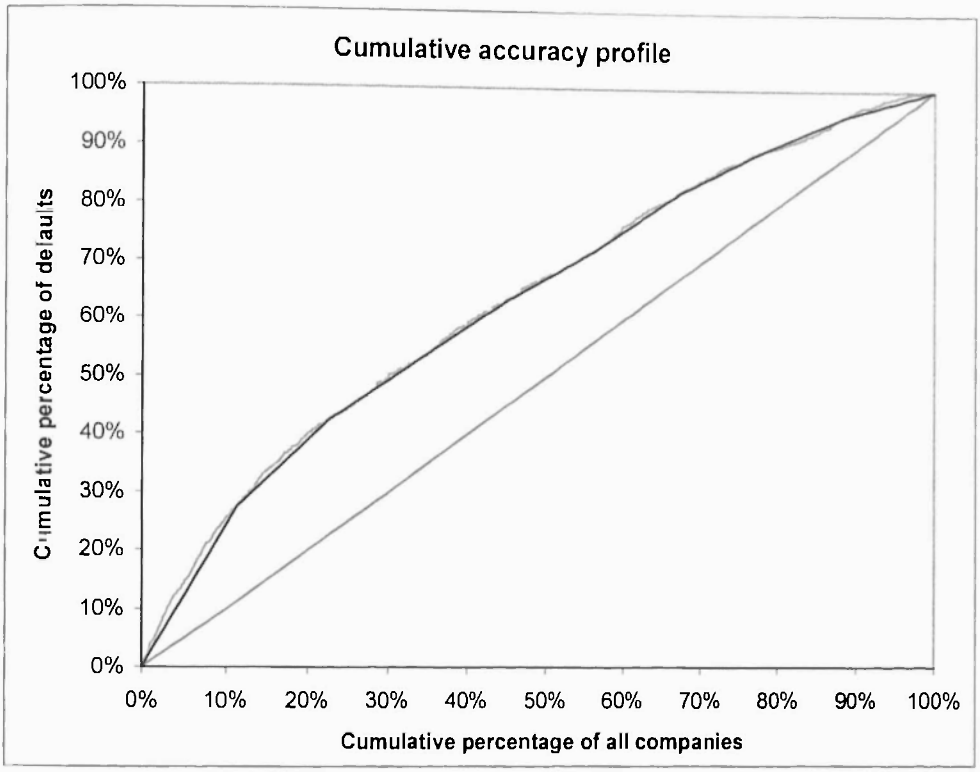
Cumulative accuracy profile (CAP)

The CAP for the Creditreform solvency index is calculated the same way as for the uniform distribution rating system. The results are presented on Graph 16.

However, there is no clear theoretical proof of the hypothesis of lower discriminatory power from aggregating obligors into rating classes. If the CAP of the Creditreform solvency index was concave everywhere along the curve, then it would be straight forward that the CAP curve of the uniform distribution rating system would lie everywhere under the CAP curve of the Creditreform solvency index. Thus, the CAP of the rating system would mark a smaller area, leading to a lower accuracy ratio (AR) and lower discriminatory power.

The problem is that the CAP curve of the Creditreform solvency index is not strictly concave. Hence, whether the AR of the Creditreform solvency index is bigger or smaller than the AR of the rating system depends on the relative concavity / convexity of the Creditreform solvency index CAP curve.

Graph 16



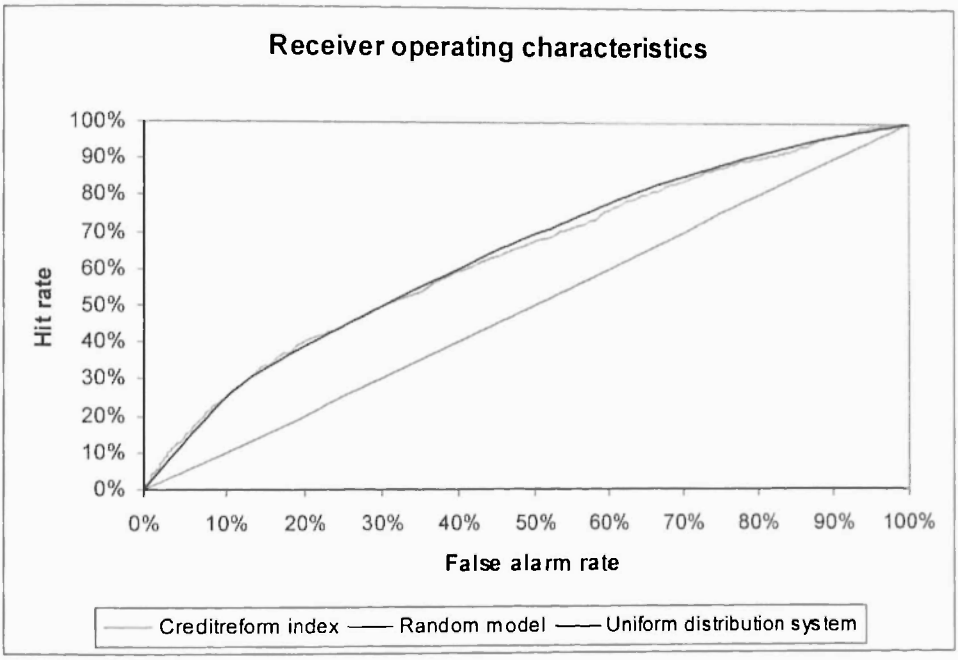
The accuracy ratio of the Creditreform solvency index is 39.11% whereas the AR of the uniform distribution rating system is 37.42%, indicating slight loss of discriminatory power. Thus, the maximum discriminatory power one can achieve when calibrating rating classes is around 39.11%[52].

[52] In accordance with the discussion in the beginning of the appendix, it is possible to set up a rating system so that it would have slightly higher discriminatory power than 39.11%. This is so due to the convexity of the Creditreform solvency index CAP curve in some parts along the curve.

Receiver Operating Characteristics (ROC)

We can make the same analysis for the ROC, and the same discussion applies. The comparison is shown on Graph 17.

Graph 17



The area under curve (AUC) for the whole Creditreform solvency index is 63.96% whereas the AUC for the uniform distribution rating system reaches 64.46%. As opposed to the accuracy ratio comparison, higher AUC for the rating system is obtained.

Testing for discriminative power[53]

A simple test can be performed to access whether the Creditreform solvency index has any discriminative power at all. We test the hypothesis H_0, stating that the Creditreform solvency index has no discriminative power based on the ROC measure, against the alternative hypothesis H_1, stating that the Creditreform solvency index has discriminatory power.

First, we need to define kernel $u_{D\,ND}$ as:

$$u_{D,ND}=\begin{cases} 1, \text{ if } S_D>S_{ND} \\ \frac{1}{2}, \text{ if } S_D=S_{ND} \\ 0, \text{ if } S_D<S_{ND} \end{cases},$$

where S_D represents a defaulter with Creditreform solvency index value equal to S, and S_{ND} stands for a non-defaulter with Creditreform solvency index value equal to S.

Then the test statistics $\hat{U}$ is defined as:

$$\hat{U}=\frac{1}{N_D N_{ND}} \sum_{(D,ND)} u_{D,ND},$$

where the sum is over all pairs of defaulters and non-defaulters (D, ND) in the dataset and the numbers of defaulters and non-defaulters in the sample are respectively denoted by N_D and N_{ND} respectively.

53 For details on this test see Engelmann et al. (2003).

For $N_D, N_{ND} \rightarrow \infty$, the term $T = \frac{AUC - \hat{U}}{\hat{\sigma}_{\hat{U}}}$ is asymptotically normally distributed[54] with mean zero and variance one and $\sigma^2_{\hat{U}}$ is given by:

$$\sigma^2_{\hat{U}} = \frac{N_D + N_{ND} + 1}{12 N_D N_{ND}}.$$

Now, we are ready to test the hypothesis H_0, stating that the Creditreform solvency index has no discriminatory power, i.e. AUC = 0.5 against the alternative hypothesis H_1, stating that the Creditreform solvency solvency index has discriminatory power, i.e. $AUC \neq 0.5$.

The results show that H_0 can be dismissed, since the value of T is -19.75, with the critical values for the standard normal distribution at 5% confidence level being -1.64 and 1.64. Hence, the Creditreform solvency index indeed has discriminatory power.

[54] According to Engelmann et al. (2003), the number of defaults should be at least 50 to guarantee good approximation. The sample used in this thesis contains 1 587 defaults.