Estimating Probability of Default Using Rating Migrations in Discrete and Continuous Time

Rickard Gunnvald

September 2, 2014

Abstract

During the financial crisis that began in 2008, even whole countries and very large companies defaulted or were on the verge of defaulting. The turmoil made risk managers and regulators more vigilant in scrutinising their risk assessment. The probability of default (PD) is an essential parameter in measuring counterparty credit risk, which in turn has impact on pricing of loans and derivatives. The last decade, a method using Markov chains to estimate rating migrations, migration matrices and PD has evolved to become an industry standard. In this thesis, a holistic approach to implementing this approach in discrete and continuous time is taken. The results show that an implementation in continuous time has many advantages. Also, it is indicated that a bootstrap method is preferred to calculate confidence intervals for the PDs. Moreover, an investigation show that the frequently used assumption of time-homogeneous migration matrices is most probably wrong. By studying expansions and recessions, specific expansion and recession migration matrices are calculated to mitigate the impact of time-inhomogeneity. The results indicate large differences of estimated PDs over the economic cycle, which is important knowledge to be able to quote correct prices for financial transactions involving counterparty credit risk.

Acknowledgement

First and foremost I would like to thank my supervisor Prof. Boualem Djehiche at KTH for valuable input and encouragement.

As this thesis marks the end of five years at KTH, I would also like to thank the ones that I have spent many long days and nights together with studying to exams and writing reports. Especially, these are Patrik Gunnvald, Magnus Bergroth, Andreas Lagerqvist, Viktor Joelsson, Alexander Keder and Daniel Boros.

Rickard Gunnvald

Contents

Chapter 1

Introduction

During the recent years, the financial markets have been unusually volatile. In the turmoil of the last financial crisis that began in 2008, many companies defaulted on their debt and thereby caused huge credit losses to their counterparties. Furthermore, among the companies that did not default there were many credit rating downgrades. A lower credit rating implies that the probability of default has increased. This causes changes to their Credit Valuation Adjustment (CVA), which is the market value of counterparty credit risk. The higher counterparty credit risk, the more the protection against default of that counterparty should cost, e.g. in form of a credit default swap. During the crisis even very large companies and whole countries, usually seen as safe, were on the verge of defaulting or defaulted. This was something sellers of credit default swaps had not priced in. The effects rippled through the closely entangled financial markets and a real fear of a system collapse spread. The system collapse never materialised, but the crisis served as a wake-up call for many market participants who never would have thought that something like that could happen. As a result, regulators, investors and participants in the financial markets have become more vigilant in dealing with and assessing the credit risks that they face when buying or selling contracts with other market participants.

There are many sources of risk for a financial company, usually divided into three major groups: market risk, credit risk and operational risk. The amount of risk a company faces have an impact on the buffer capital that it is required by regulators to put aside as a cushion in case the risks would materialise. One important input when measuring the credit risk is the probability of default (PD) of a counterparty. This thesis will focus on credit risk and in particular PD. The PD is also an important input when pricing loans and derivatives bought or sold to a specific counterparty. If a value is exposed towards a counterparty (e.g. in form of a loan or derivative contract), it requires putting aside buffer capital to account for the loss given default of that counterparty. Both the PD and the exposure are important when calculating loss given default. Since capital has a cost in form of e.g. interest, it is more expensive to have exposures to counterparties with high PD. This in turn affect the quoted prices. The impact on pricing is just one of the reasons why it is important to companies in general, and financial companies such as banks in particular, to have an accurate estimate of the PD of its counterparties.

There are different ways of calculating or estimating the probability of default. As an example, one can use market implied methods such as backing out the PD from credit spreads. Another example is the Merton's structural model, where assets are modelled as a geometric Brownian motion and debt as a single outstanding bond with a certain face value at a given maturity time

T. If the value of the assets are less than the outstanding debt at time T, then a default is deemed to have occurred. However, this thesis will focus on an approach that is widely used among risk managers. The approach to be investigated uses the credit rating and its migrations to asses the probability of default. Credit ratings are set by rating agencies such as Standard and Poor's or Moody's, but larger banks and financial companies often have their own internal rating system used on its counterparties.

In particular, rating migrations will be estimated using a Markov chain framework, where migration (transition) matrices are used to extrapolate the cumulative transition probabilities forward in time. This approach has been around since the beginning of the 21st century, but has evolved during the years. In short, this approach can be implemented in both discrete and continuous time. One study that is often referred to is the work by Lando and Skødeberg (2002), where they looked at differences between the discrete and continuous method. Also, articles such as the one by Jafry and Schuermann (2004) have been published that suggest different ways of comparing transition matrices to each other. When a PD is estimated, it is also important to know how accurate the estimate is. Work by Christensen et al. (2004) and Hanson and Schuermann (2005) had a focus on trying to estimate confidence intervals (CIs) for PDs. They also compared different methods of estimating CIs. One frequently used assumption is that the transition matrix is time-homogeneous, which is indicated by later research to be a simplification. Therefore, the most recent research has been focused on testing the time-homogeneity assumption and trying to mitigate inhomogeneities or model them. However, there is no consensus regarding how one should account for time-inhomogeneity properly.

This thesis will have a focus on theory, problems and questions that arise when implementing a Markov chain approach to estimating rating migrations and PDs in practice. Therefore, it will take a holistic view on the whole implementation process, meaning it will touch upon many areas that are research fields in their own rights. First of all, a theoretical framework for the Markov chain is presented, as well as its application to the credit migration framework. Theoretical background to the tests performed will also be presented. The areas touched upon range from how to handle data issues to comparing matrices with each other in discrete and continuous time to suggesting a method aimed at mitigating the impact of time-inhomogeneity.

In more detail, some methodology is presented regarding adjustment of the data set. The data set used in this thesis is of course not identical to what other researchers might use, but it still provides useful comments on issues that has to be considered. Also, throughout the thesis, results from using a discrete and continuous calculation method will be compared. Moreover, the full data sample will be divided into subsamples and tested to make sure that the estimated transitions on average are dependent on rating rather than something company specific. One approximation method of calculating in continuous time will also be examined to see how accurate it is. If accurate, it could ease the implementation for those not using computer softwares such as R or MATLAB. Also, confidence intervals will be calculated and compared using two different methods. As a part of a time-inhomogeneity investigation, differences between matrices will be measured. The homogeneity assumption will also be statistically tested using a χ^2 test and a comparison of confidence intervals. Moreover, the inhomogeneity investigation will also be conducted on a subset containing companies from defensive sectors to see if that mitigates impact of inhomogeneities. If so, it would suggest that it would be sufficient to only use a homogeneous migration matrix if exposed to that type of companies. If the homogeneity assumption is accurate, then it would make risk managers' job easier. Finally, a study of cumulative PD curves from annual migration matrices is undertaken to determine what data should be included when calculating expansion and recession migration matrices. The use of expansion and recession matrices are thought to mitigate the impact of inhomogeneities, and will be further elaborated on as a suggestion for further studies.

The analysis performed in this thesis shows that the continuous method is superior to the discrete method in terms of efficiently capturing migrations in the data. It also suggests that the approximation method should only be used on time frames up to 1 year. The splitting into subsets show that migrations on average are not dependent on company specific data. Moreover, the study on confidence intervals suggest that a bootstrap approach is recommended to be used, both in discrete and continuous time. The methods used in the time-inhomogeneity study clearly shows that inhomogeneities are present, and that defensive sectors are exposed to inhomogeneities essentially to the same extent. Finally, the study on cumulative PD curves indicates what years to be included in the expansion and recession migration matrices. The differences between those two are rather striking, and shows the importance of taking time-inhomogeneity into consideration for short-term counterparty exposures.

The outline of this thesis will be as follows: In chapter 2, the theoretical Markov chain framework and its application to credit migrations will be presented. Chapter 2 will also contain some theoretical tools used in testing, as well as an overview on previous research relevant for the thesis. The theory from chapter 2 will later be applied and used on credit rating data. Chapter 3 describes the data set used, tabulating e.g. number of firm years and non-diagonal movements. Moreover, it describes the fields of the data set and how observations are created. In chapter 4, the methodology used when implementing the theory on rating data will be presented. This also includes necessary adjustments to account for certain issues that have occurred along the implementation process. Examples are adjustments to overlapping observations and how to handle defaults that recover. Furthermore, this chapter will contain information on methods to calculate and test some of the results. In chapter 5 the results of this study will be presented with comments describing the outcome. In chapter 6 there is a thorough discussion around the interpretation of the results. Moreover, the main conclusions will be presented in chapter 6, as well as some suggestions for further studies.

Chapter 2

Theory and application

In this chapter, the necessary theoretical framework will be presented. At the end of the chapter, there is an overview of some previous studies in research fields relevant for this thesis.

2.1 Hypothesis testing

Statistical hypothesis testing is a method of statistical inference. Collected data is used to statistically determine which of either a null hypothesis or an alternative hypothesis is accepted in favour of the other given a certain *significance level*. The significance level, often denoted α, is the probability threshold below which the null hypothesis will be rejected. The null hypothesis is commonly denoted as H_0 and the alternative hypothesis is commonly denoted as H_1.

There are different ways to reject or accept a H_0. One way is to use a relevant test statistic T and calculate an observed test statistic t_{obs}. Depending on the significance level and what distribution the data is deemed to follow, T statistics and their values can be found tabulated. The t_{obs} is then calculated from data and depending on its value, the H_0 is either rejected or accepted.

Another way is to calculate the *p-value* from the data. The p-value is the probability of having taken the wrong decision when rejecting H_0. Thus, if a calculated p-value is below the significance level α, then H_0 should be rejected.

Associated with hypothesis testing are the so called *Type I error* and *Type II error*, which can be found explained in Table 2.1.

Table 2.1: Table explaining the Type I Error and Type II Error associated with hypothesis testing.

	H_0 is true	H_1 is true
Reject null hypothesis	Type I Error	Correct decision
Accept null hypothesis	Correct decision	Type II Error

A key part in using hypothesis testing is to meticulously define H_0 and H_1, so that the resulting rejection/acceptance of H_0 is meaningful to the investigated property.

Confidence intervals

One use of confidence intervals is to be able to determine how certain a specific estimation is. As an example, a 95% confidence interval of a parameter is the interval where 95% of the values or outcomes from this parameter will be. Confidence intervals can also be used to determine whether two estimates of the same parameter are statistically different from each other or not. These are the two main purposes that confidence intervals will be used in this report. Moreover, two different methods to calculate confidence intervals will be used and are presented immediately below. One method is the Wald confidence intervals, and the other is a bootstrap method. Finally, a short description of the Kolmogorov-Smirnov two-sample test will be presented. The test is used later in Appendix B to compare Wald confidence intervals to their bootstrapped counterparts.

Wald confidence interval

The Wald confidence interval is an analytic confidence interval where the underlying assumption is that the observed variable follows a binomial distribution.

As a relevant example, let the random variable X be such that it describes if a company defaults or not. This random variable can in discrete time be assumed to follow a binomial distribution. If n is the number of trials, and PD is the probability of default for one time step, then the expected value μ of defaulted companies after one time step is $n \cdot PD$. Moreover, the variance σ^2 is $n \cdot PD(1 - PD)$.

Now consider a situation where we have an observed sample of independent and identically distributed X_is of size n. The X_is have mean μ and variance σ^2 and we are interested in estimating the X_is sample mean, i.e. the estimated probability of default $\widehat{PD}$. Then

$$\widehat{PD} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{2.1}$$

For large enough n, the Central Limit Theorem (CLT) states that the distribution of $\widehat{PD}$ is close to the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. The μ and σ^2 describe the mean and variance of X_i. In the case of estimating the sample mean, i.e. the probability of default, then according to the CLT $\widehat{PD}$ follow the normal distribution below.

$$\widehat{PD} \sim N\left(\widehat{PD}, \frac{\widehat{PD}(1 - \widehat{PD})}{n}\right) \tag{2.2}$$

The construction of a $(1 - \alpha)\%$ confidence interval for $\widehat{PD}$ is now straight forward. The Wald confidence interval, CI_W is then

$$CI_W \pm \kappa\sqrt{\frac{\widehat{PD}(1 - \widehat{PD})}{n}} \tag{2.3}$$

where κ is the $(1 - \frac{\alpha}{2})$ quantile of the standard normal distribution.

Bootstrapped confidence interval

Consider the case where a sample of observed values exists, but it is unknown what distribution they follow. When there is no analytical way to calculate confidence intervals, one option is to use a resampling method called bootstrapping. The empirical distribution of observed values can then be chosen to serve as an approximation of the true distribution, from which values are drawn with replacement. The bootstrapping technique allows for estimation of the accuracy of some distribution parameter, such as the sample mean. This can then can be used to calculate e.g. confidence intervals.

The standard bootstrapping procedure is the one used in this thesis to estimate confidence intervals for the probability of default. Consider having a sample of n observations. Then, out of the original sample, observations are drawn with replacement one at the time to construct a new sample of size n. The new sample gives an estimate of the PD. Then this procedure is repeated N number of times to get N estimates of the PD. These N values now form an estimate of the PD's distribution. Constructing a $(1-\alpha)\%$ two-sided symmetric confidence interval out of this distribution is done by simply ordering the values from the lowest to highest and choosing the $\frac{\alpha}{2}$ percentile and the $(1-\frac{\alpha}{2})$ percentile.

Kolmogorov-Smirnov

The two-sample Kolmogorov-Smirnov (K-S) test can be used to statistically test whether two samples follow the same distribution. The mathematical proof behind the test is not within the scope of this thesis, but an outline of how to use it in hypothesis testing is outlined immediately below.

Let $F_{1,n}(x)$ and $F_{2,n'}(x)$ be the cumulative distribution functions of the two samples with size n and n', respectively. A set of distances between $F_{1,n}(x)$ and $F_{2,n'}(x)$ is obtained by simply calculating $|F_{1,n}(x) - F_{2,n'}(x)|$. Then, the test statistic $D_{n,n'}$ used is the supremum, or loosely speaking the maximum, of the set of differences.

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|$$

The concept behind the test is that if $F_{1,n}(x)$ and $F_{2,n'}(x)$ follow the same distribution, then the $D_{n,n'}$ should converge to 0 as n goes to infinity. The null hypothesis that the two samples follow the same distribution is rejected if

$$D_{n,n'} > c(\alpha)\sqrt{\frac{n+n'}{nn'}} \tag{2.4}$$

where $c(\alpha)$ can be found tabulated for different significance levels α.

2.2 The Markov chain model

In this section, definitions and aspects of the Markov chain theory will be presented. If the reader is familiar with Markov chain theory, then it is possible to jump to section 2.3 where the framework is applied.

2.2.1 Markov chain

Definition 2.1: (Markov chain)

A Markov chain is a stochastic process $\{X_i\}_{i \geq 0}$ that forms a sequence of random variables $X_0, X_1, ...,$ with outcomes $x_0, x_1, ...$ on the finite or countable set S that satisfies the *Markov property*.

Definition 2.2: (State space)

The finite or countable set S forms the *state space* of the Markov chain, i.e. the set of possible outcomes of X_i. Each possible outcome $x_i \in S$ is called a *state*.

Definition 2.3: (Markov property)

For a *stationary* discrete Markov chain, satisfying the *Markov property* means that

$$Pr(X_{n+1} = x \mid X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = Pr(X_{n+1} = x \mid X_n = x_n)$$

for all stages n and all states $x_0, x_1, ..., x_{n+1}$.

Thus, the next stage $n + 1$ only depends on the stage n, creating serial dependence on the adjacent stage as in a "chain". Note the difference between stages and states; stages are the steps with which the Markov chain progresses, whereas the states are the possible outcomes in each stage.

The *Markov property* is sometimes referred to as the *first order Markov condition*, or that a sequence is *memoryless*.

Definition 2.4: (Stationarity or Time homogeneity)

The term *stationary Markov chain*, in a time setting sometimes referred to as a *time-homogeneous Markov chain*, implies that

$$Pr(X_{n+1} = a \mid X_n = b) = Pr(X_n = a \mid X_{n-1} = b) \tag{2.5}$$

Thus, the transition probability is independent of the stage n. Note however that a time-homogeneous Markov chain is *not* independent of the *length* between stages. In a time setting where the stages are time points, this would mean that the Markov chain is independent over time, but not independent of time step length. Naturally, the shorter time step, the less probable it is that the stochastic process has moved during that time.

Definition 2.5: (Transition probability)

The transition probabilities are defined as follows:

$$Pr(X_1 = j \mid X_0 = i) = p_{ij} \quad \text{and} \quad Pr(X_n = j \mid X_0 = i) = p_{ij}^{(n)}$$

corresponding to the single-step transition probability and the transition probability in n steps, respectively. More specifically, p_{ij} is the probability of making a transition (moving) from state

i to state j. In a time setting, each step n could be defined as e.g. one year. Then p_{ij} would be the probability of transitioning from state i to state j in one year's time.

Definition 2.6: (Transition matrix)

For a finite state space S, we now define the transition matrix P over N states as

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$$

where the entries p_{ij} are transition probabilities as in Definition 2.5: Transition probability.

Theorem 2.1: (Properties of the transition matrix)

a) $\sum_{j=1}^{N} p_{ij} = 1$ for $i = 1, 2, ..., N$
b) $p_{ij} \geq 0 \quad \forall\, i, j = 1, 2, ..., N$

The claim in a) follows from the definition of p_{ij}, since the sum of the probabilities of either staying in the current state or moving to any other state in the state space must be equal to one. The claim in b) is obvious since the p_{ij}s are probabilities and therefore non-negative.

Theorem 2.2: (Stage transitions)

Let $P^{(n)}$ be the matrix containing all the state transition probabilities p_{ij}, $i = 1, 2, ..., N$, $j = 1, 2, ..., N$ at stage n. Then, following Enger and Grandell (2006),

a) $P^{(m+n)} = P^{(m)} P^{(n)} \quad m, n \in \mathbb{N}$
b) $P^{(n)} = P^{n} \quad n \in \mathbb{N}$

The formula in a) means that the transition matrix P at stage $m + n$ is the same as multiplying the transition matrix at stage m with the transition matrix at stage n. Note that since m and n are non-negative, it is not possible to run this process backwards through stages or time points.

The formula in b) means that the transition matrix in stage n is obtained by multiplying the one-step (from stage 0 to stage 1) transition matrix P by itself n times. This gives us the tools to calculate transition matrices forward throughout the stages. In a time setting, each stage represents a specific time point, i.e. multiplying P with itself shows the transition probabilities forward in time at different time points. Transitioning through time is an essential result used in thesis. Moreover, the statements in a) and b) are not very intuitive, therefore the proof is found immediately below.

Proof:
a) We prove a) by showing that the elements of $P^{(m+n)}$ are obtained by a matrix multiplication of $P^{(m)}$ and $P^{(n)}$. I.e. for any states i, j in the state space S

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)} \tag{2.6}$$

where the right-hand side of equation (2.6) is in fact the result of multiplying row i in $P^{(m)}$ onto column j in $P^{(n)}$. The whole matrix multiplication and thereby the $P^{(m+n)}$ matrix is obtained by varying i and j. One can also think of the rationale behind the equation (2.6) as follows: Assume we want to calculate the probability of transitioning from state i to state j in $(m+n)$ steps. That probability is the sum of all possibilities of transitioning from state i to an arbitrary intermediary state k in m steps, and then onward from k to state j in n steps. This is precisely equation (2.6). We get

$$\begin{aligned}
p_{ij}^{(m+n)} &= Pr(X_{m+n} = j \mid X_0 = i) = \sum_{k \in S} Pr(X_m = k, X_{m+n} = j \mid X_0 = i) = \\
&= \sum_{k \in S} Pr(X_m = k \mid X_0 = i) Pr(X_{m+n} = j \mid X_m = k, X_0 = i) = \\
&= \{\text{Markov property}\} = \sum_{k \in S} Pr(X_m = k \mid X_0 = i) Pr(X_{m+n} = j \mid X_m = k) = \\
&= \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}
\end{aligned}$$

□

b) We can rewrite

$$\begin{aligned}
P^{(n)} &= P^{(n-1+1)} = \{\text{Theorem } 2.2a)\} = P^{(n-1)} P^{(1)} = P^{(n-1)} P^1 = \\
&= P^{(n-2)} P^2 = \cdots = P^n
\end{aligned}$$

□

As mentioned, when combining a) and b) we see that by simply multiplying the transition matrix by itself m times, the transition probabilities for the next m stages are obtained. Note also that $P^0 = I$ (the identity matrix).

Theorem 2.3: (State distributions) Let $\bar{p}_i^{(n)}$ be the vector containing row i of $P^{(n)}$. I.e. the transition probability distribution for state i at stage n. Then

$$\bar{p}_i^{(n)} = \bar{p}_i^{(0)} P^n \tag{2.7}$$

which describes that the transition probability distribution for state i at stage n is obtained by multiplying the distribution for state i at stage 0 with the transition matrix for stage n. Since

$$\bar{p}_i^{(0)} = [00 \cdots 010 \cdots 0] \in 1 \times N$$

where N is the number of states and the 1 is at column i, this shows that $\bar{p}_i^{(n)}$ is obtained by extracting row i from P^n.

Proof:

$$\begin{aligned}
\bar{p}_i^{(n)} &= Pr(X_n = i) = \{\text{Law of total probability}\} = \\
&= \sum_{k \in S} Pr(X_0 = k) Pr(X_n = i \mid X_0 = k) = \sum_{k \in S} \bar{p}_i^{(0)} p_{ki}^{(n)} = \\
&= \bar{p}_i^{(0)} P^{(n)} = \{\text{Theorem 2.2 b)}\} = \bar{p}_i^{(0)} P^n
\end{aligned}$$

□

Discrete-time Markov chain

For a *discrete-time* Markov chain (DTMC) each stage n corresponds to certain given time points, with constant time step between them. As an example, one can let the time between two time points (stages) be 1 year, so that $p_{ij}(1)$ denotes the probability of moving from state i to state j in one year's time. In general, the probability of transitioning from state i to state j during a time t, will be denoted $p_{ij}(t)$. The transition matrix over a time t will be denoted $P(t)$.

When talking about Markov chains in a time setting we will henceforth talk about stages in the chain as time points and also call the difference between two stages one *time step*.

Continuous-time Markov chain

For a *continuous-time* Markov chain (CTMC), some additional theoretical framework is needed. Instead of considering transition probabilities at fixed time points as in the discrete framework, we now consider a stochastic variable T, the time spent in each state. Moreover, instead of transition probabilities for a fixed time step, we are now considering transition *rates*. The larger the transition rate, the sooner in time the transition is *expected* to take place. In the continuous case, the time spent, T, in each state follows an *exponential distribution*, with the transition rate as rate parameter.

To clarify the difference between discrete and continuous time Markov chains, one can think of how each chain would be simulated. In the discrete case, each state would have certain fixed probabilities to have transitioned to other possible states (including the current state) *at a fixed future time point*. The total probability (including staying in its current state) is of course 1. Thus taking a random number between 0 and 1 could simulate in which state the process will be at the next fixed time point.

For the continuous case, each possible state would be associated with a certain transition rate. To simulate the Markov chain's movements through time, one simply calculate a realisation of the stochastic time spent in its current state before it transitions to each of the other possible states. Thus, a "time spent" $T_1, ..., T_N$ will be obtained for each other possible state $1, ..., N$. The shortest time $min\{T_1, ..., T_N\}$ will decide to which state the Markov chain transitions into, *and how long it takes* before that happens.

Thus, the discrete Markov chain will be said to be in certain states at certain fixed time points, whereas the continuous time Markov chain will move between the states at irregular times. Furthermore, when using the discrete time Markov chain we want estimates of the transition *probabilities*, whereas estimates of the transition *rates* are desired for the continuous time Markov chain. If the transition rates for a CTMC are available, then one can also calculate how the transition probabilities change in continuous time.

Theorem 2.4: (Continuous stage transitions)

A useful result, which is the continuous version of Theorem 2.2: Stage transitions (and proven similarly) is that

$$P(t+s) = e^{(t+s)Q} = P(t)P(s) \tag{2.8}$$

Where Q is the *generator matrix*, as defined immediately below.

Definition 2.7: (Generator matrix)

Let $\{X_t\}_{t\geq 0}$ denote the CTMC, which is a stochastic process in continuous time satisfying the *Markov condition*. Let $P(t)$ be the transition matrix in continuous time, S the state space as in Definition 2.2 and Q the *transition rate matrix*. Q is sometimes also referred to as the *intensity matrix*, the *infinitesimal generator matrix* or simply *generator matrix*.

$$Q = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{pmatrix}$$

where the elements q_{ij} denotes the rate at which the process transitions from state i to state j. A more detailed derivation and description of q_{ij} and q_{ii} is given in Theorem 2.6: Generator and transition matrix relation. The elements p_{ij} of $P(t)$ are defined as $Pr(X_t = j \mid X_0 = i)$, similar to the discrete case.

Theorem 2.5: (Properties of the generator matrix)

The intensity matrix Q should satisfy the following properties:

1. $0 \leq -q_{ii} \leq \infty$
2. $q_{ij} \geq 0$ for all $i \neq j$
3. $\sum_j q_{ij} = 0$ for all $i \iff q_{ii} = -\sum_j q_{ij}$ for all $i \neq j$

Theorem 2.6: (Generator and transition matrix relation)

Consider a time step h. Following the derivations outlined in Enger & Grandell (2006), it follows from the definition of intensities that

$$\begin{aligned} q_{ij} &= \lim_{h \to 0+} \frac{p_{ij}(h) - 0}{h} \quad i \neq j \\ q_{ii} &= \lim_{h \to 0+} \frac{p_{ij}(h) - 1}{h} \end{aligned}$$

or in matrix form:

$$Q = \lim_{h \to 0+} \frac{P(h) - I}{h} \tag{2.9}$$

Noting that $P(0) = I$, the definitions of q_{ij}, q_{ii} and Q are the derivatives of p_{ij}, p_{ii} and P with respect to time.

From Theorem 2.4: Continuous stage transitions, we get that

$$P(t+h) = P(t)P(h) = P(h)P(t)$$

or equivalently

$$P(t+h) - P(t) = P(t)(P(h) - I) = (P(h) - I)P(t)$$

Dividing by h and letting $h \to 0+$ yields

$$P'(t) = P(t)Q = QP(t) \tag{2.10}$$

These are called the *Kolmogorov forward* and *Kolmogorov backward* equations, respectively. The Kolmogorov forward and backward equations are first order differential equations, with unique solution

$$P(t) = e^{tQ} \tag{2.11}$$

Note that e^{tQ} is a matrix exponential, defined as the power series $e^{tQ} \equiv \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}$.

Theorem 2.7: (Infinitesimal definition of CMTC)

The infinitesimal definition of the CTMC is as follows:

Assume that the stochastic process X_t is in state i at time t. Then for $h \to 0$ and $s < t$, X_{t+h} is independent of X_s and

$$Pr(X_{t+h} = j \mid X_t = i) = \delta_{ij} + q_{ij}h + o(h)$$

where $o(h)$ denotes the little-o notation, which implies that the function $o(h)$ goes towards 0 faster than h itself, i.e. $\lim_{h \to 0+} \frac{o(h)}{h} = 0$. The δ_{ij} is the Kronecker delta defined as

$$\delta_{ij} = \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases}$$

Thus, for h small enough

$$Pr(X_{t+h} = j \mid X_t = i) \approx \delta_{ij} + q_{ij}h \tag{2.12}$$

or in migration matrix form

$$P(h) \approx I + Qh \tag{2.13}$$

Note the similarity with the intensity definition in eq. (2.9). One benefit of this approximation is that it allows for computation of the migration matrix P a small time step into the future via the generator matrix Q *without* using the infinite series of a matrix exponential.

2.2.2 Some properties of the Markov chain

In this section, some properties of the Markov chain that later might be referred to will be defined.

Accessibility: A state j is said to be *accessible* from a state i if there is a non-zero probability for a system starting in state i to eventually transition into state j. This is denoted $i \to j$. Note that the process is allowed to pass through several other states along the way.

Communication: A state i is said to *communicate* with a state j if $i \to j$ and $j \to i$. This is denoted $i \leftrightarrow j$. A set of states C is said to define a *communicating class* if all states in C communicates with each other and no state in C communicates with any state outside C.

Irreducibility: A Markov chain is said to be *irreducible* if it is possible to get to any state from any state, i.e. if the Markov chain state space forms one single communicating class.

Transiency: A state i is said to be *transient* if there is a non-zero probability that the Markov chain never will return to state i. If a state is not transient, then it is said to be *recurrent*.

Absorbing: A state i is said to be *absorbing* if it is impossible to leave the state, i.e. if $p_{ii} = 1$ and $p_{ij} = 0$ for $i \neq j$. If every state can reach an absorbing state, then the Markov chain is an *absorbing Markov chain*.

Periodicity: A state i is said to be *periodic* with period k if any return to state i *must* occur in multiples of k time steps, for $k > 1$. If $k = 1$ then it is said to be *aperiodic*, and returns to i can occur at irregular times.

2.3 Credit migration matrices

Credit migration matrices are used to describe and predict the movement that a company (or other rated assets such as bonds) takes through different credit rating classes. This report is, however, focuses on the credit migration of companies. Studying credit migration matrices is at the very heart of credit risk management. The publicly available reports on rating migrations published by Standard & Poor's (S&P) and Moody's are studied frequently by risk managers [16] and rating migration matrices are very important input in many credit risk applications. It is therefore crucial to get an accurate estimation of the migration matrix.

In this chapter the Markov chain theory will be used to show how one can build up a theoretical framework around credit migration. Different methods of estimating credit migration matrices in discrete and continuous time will be presented and compared, as well as further theory and information regarding the rating input and default definitions.

2.3.1 Applying the Markov chain model

The state space S consists of the different credit ratings available. E.g. for S&P's ratings $S = \{$AAA, AA, A, BBB, BB, B, CCC, CC, C, D$\}$, where ratings AA through CCC can be modified with $(+/-)$ to show the relative standing within the rating category. Let N be the number of states for the chosen credit rating framework, i.e. the number of possible ratings. Let M denote the migration matrix. M corresponds to the transition matrix P in the Markov chain theory, where entries p_{ij} in the rating migration framework denotes the probability of making a transition from rating (state) i to rating (state) j during the specified time period.

Also, let G denote the generator matrix, i.e. the matrix corresponding to Q in the CTMC framework. The entries q_{ij} are defined analogously, where the states i and j in the rating migration setting are two different ratings.

The default state D is often assumed to be *absorbing*, so that once a company has entered that state it cannot leave. Praxis is to have the highest rating furthest to the left and then let them descend towards the lowest rating, D, in the rightmost column. Thus

$$M = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1(N-1)} & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2(N-1)} & p_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{(N-1)1} & p_{(N-1)2} & \cdots & p_{(N-1)(N-1)} & p_{(N-1)N} \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and

$$G = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1(N-1)} & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2(N-1)} & q_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ q_{(N-1)1} & q_{(N-1)2} & \cdots & q_{(N-1)(N-1)} & q_{(N-1)N} \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

If the default state is absorbing it will eventually (given enough time) cause all companies to end up in default. As mentioned by Jafry & Shuermann (2003), the time it takes for a credit migration process to end up close to its steady-state is very long in economic terms. It also relies on assumptions such as time-inhomogeneity of the migration matrix, which is questionable over longer time periods. In reality the economic conditions change, and thereby altering the migration matrix, long before a default steady-state implied by an assumed constant migration matrix occurs.

There are different methods of estimating the entries of the M matrix. The two most commonly used and referred to are the so called *cohort* (discrete time) and *duration* (continuous time) methods. This thesis will focus solemnly on these two.

2.3.2 The cohort method

Let $t_0, t_1, ..., t_n$ be discrete time points such that an arbitrary time interval $t_{k+1} - t_k = \Delta t_k$, where Δt_k is constant. As described by Christensen et al. (2004), the estimator of $p_{ij}(t_k)$ over one time period is then

$$\hat{p}_{ij}(t_k) = \frac{n_{ij}(\Delta t_k)}{n_i(t_k)} \tag{2.14}$$

where $n_{ij}(\Delta t_k)$ is the number of companies that have moved from state i to state j between time t_k and t_{k+1}, and $n_i(t_k)$ are the number of companies in state i at time t_k.

If we further assume that the Markov chain considered is time-homogeneous and that data is available from time t_0 to time t_N then e.g. Christensen et al. (2004) have shown that the Maximum Likelihood (ML) estimator is

$$\hat{p}_{ij} = \frac{\sum_{k=0}^{N-1} n_{ij}(\Delta t_k)}{\sum_{k=0}^{N-1} n_i(t_k)} \tag{2.15}$$

The above equation (2.15) describes averaging of the transition probability estimators found in all N time periods of length Δt_k.

From the properties of the migration matrix, Theorem 2.1 a), we know that

$$\hat{p}_{ii} = 1 - \sum_{j \neq i} \hat{p}_{ij} \tag{2.16}$$

The estimations of p_{ij} and p_{ii} then form the migration matrix $M(\Delta t_k)$ for the time window Δt_k used. If $\Delta t_k = 1$ (year), then $M(1)$ is the 1-year migration matrix. The assumption of time-homogeneity is used to aggregate and extrapolate transitions and probabilities over different time periods. To calculate migration probabilities over a 2.5 year interval, extrapolation through matrix multiplication is needed and then one has to interpolate between year 2 and year 3.

If the assumption of time-homogeneity is removed, one can still estimate p_{ij} for a specific time period $[t, T]$ by

$$\hat{p}_{ij}(t, T) = \frac{n_{ij}(t, T)}{n_i(t)} \tag{2.17}$$

where the n_{ij} is the number of companies that have migrated from state i to state j during the time interval $[t, T]$ and n_i is the number of companies in rating category i at time t. However, this type of estimate is not straight forward to aggregate or extrapolate.

Noteworthy is also that there are of course cases where companies go from being non-rated to receiving a rating within a time period, and also a rating withdrawal of a company in the data sample. In the cohort method, there is often the assumption that these types of events are non-informative and the rating data for affected companies is therefore excluded from the sample at those particular times.

2.3.3 The duration method

Following Lando & Skødeberg (2002), one can obtain the ML estimation of M by first obtaining the ML estimation of generator matrix G and then applying the matrix exponential function on this estimate, scaled by time horizon.

Under assumption of time-homogeneity, the ML estimator of elements q_{ij} between time t and T in G is given by

$$\hat{q}_{ij}(t, T) = \frac{n_{ij}(t, T)}{\int_t^T Y_i(s)ds} \quad \text{for} \quad i \neq j \tag{2.18}$$

Where $n_{ij}(t, T)$ is the total number of companies that have migrated from state i to state j during the time period $[t, T]$ and $Y_i(s)$ is the number of companies in rating class i at time s.

From the properties of the generator matrix (see Theorem 2.5) we get that

$$\hat{q}_{ii} = -\sum_{j \neq i} \hat{q}_{ij} \quad \forall\, i \tag{2.19}$$

The estimations of q_{ij} and q_{ii} form the elements of the generator matrix G, from which the migration matrix for an arbitrary time t, $M(t)$, is calculated as

$$M(t) = e^{tG} = \sum_{k=0}^{\infty} \frac{(tG)^k}{k!} = I + (tG) + \frac{(tG)^2}{2!} + \cdots \tag{2.20}$$

Just as with the cohort method, disregarding time-homogeneity in the duration method is not straight forward and requires more theoretical and empirical work. Transitions to and from the unrated category are seen as non-informative.

2.3.4 Comparison of the cohort and duration methods

One drawback with the cohort method is that the estimators give probability zero to an event if there are no records of such an event in the data. As mentioned in Lando (2004), this makes the estimators poor in capturing rare events.

The advantages of choosing the duration method over the cohort method have been mentioned in a number of papers, e.g. Lando & Skødeberg (2002). As stated earlier, the cohort method assigns zero probability to events not present in the data. However, by using the duration method and the generator matrix one gets small but non-zero probability for these events. It is of course relevant from a risk perspective to be able to capture rare events even if they are not present in the data set.

Another benefit from using the duration method is that there is no problem with the so called *embedding problem*. The embedding problem is the problem of finding a generator G that is consistent with M, i.e. that $M(t) = e^{tG}$ exactly. The problem occurs because not every discrete Markov chain can be realized as a continuous-time chain interpolated from a discrete transition matrix. Israel et al. (2001) states some conditions under which a generator G does *not* exist, and the problem is further elaborated on in Lando (2004). One of the conditions in Israel et al. (2001) occurs frequently with real data, and that is:
There exists states i and j such that $i \to j$, but $p_{ij} = 0$.

This is not reasonable since a continuous Markov chain must have either $p_{ij} \geq 0$ for all t, or $p_{ij} = 0$ for all t. As stated in Lando (2004), what can happen over a period of time may also happen over arbitrarily small periods of time.

A further benefit of using the duration method is that e.g. the cumulative probability of default for an arbitrary time t can be calculated directly through the formula $M(t) = e^{tG}$. It also allows to calculate the cumulative probability of default down to a specific day (depending on data used in estimations), whereas interpolation is inevitable with the cohort method. This is of course good for practical purposes.

Furthermore, the duration method allows use of an arbitrary length of the estimation window no matter the time period length of the desired migration matrix. This is not the case with the cohort method, where e.g. 1-year estimation window(s) are used to estimate a 1-year migration matrix. One could technically choose to mimic this duration method benefit with the cohort method by estimating and interpolating a large amount of 1-day migration matrices, however this is not very practical.

Finally, the duration method captures migrations in continuous time. One example can be that non-rated companies gets a rating and enters the data set, then the $Y_i(s)$-term in equation (2.18) "reacts" to this faster. It also better captures how many companies there are in a certain rating class i, since there is a time integral looking at actual time spent in the rating class rather than a fixed observation of the number of firms at the start of an estimation window. Therefore the duration method more efficiently uses all the data in the data set. The ability to choose estimation window arbitrarily with the duration method further enhances the efficiency from a practical point of view.

One potential drawback with the duration method is that calculating the matrix exponential e^{tG} requires calculating an infinite series expansion, which is not possible in practice. It can also be cumbersome if someone is forced to use lesser refined computer programs. Computer software such as MATLAB have very accurate approximations of matrix exponentials that are fast to compute. However, if one have to use e.g. Excel (that has no fast approximation of the matrix exponential), a remedy to the somewhat unwieldy infinite series expansion might be the infinitesimal definition of the CMTC, as defined in Theorem 2.7. This is because it allows calculation of a migration matrix from a generator matrix without using the infinite series expansion. Later on, the impact on the estimated results will be examined when using the exact definition compared to the infinitesimal definition (which is an approximation).

2.3.5 Default definition

The definition of default may have some differences between companies, but the European Union and the Basel Committee publishes legislative acts and regulations on how to calculate certain capital requirements and when to conclude that an obligor is in default. In *Regulation (EU) No 575/2013*,[6], one can under *Article 178* find a definition of when a default should be considered to have occurred.

To put it simple, an institution should consider a default to have occurred if:

a) the obligor is unlikely to pay its credit obligations (principal, interest or fees) to the institution in full

b) the obligor is past due more than 90 days on any material credit obligation to the institution.

Another way to define a default when dealing with swaps and derivative contracts is to look at what is said to be a "credit event" that would trigger a settlement under a Credit Default Swap (CDS) contract. These events are stated in the International Swap Dealers Association (ISDA) agreements. The most common credit events (see [18]) are the following:

i) Bankruptcy - The entity has filed for relief under bankruptcy law (or equivalent law)

ii) Failure to pay - The reference entity fails to make interest or principal payments when due, after grace period expires (if grace period is applicable)

iii) Debt restructuring - The configuration of debt obligations is changed in such a way that the credit holder is unfavourably affected (maturity extended and/or coupon reduced)

The take-away point from these definitions is that, depending on the internal definitions, reasons that a company defaults may vary. A default may occur for reasons ranging from suspicion of not being repaid in full or being late with payments to filing a bankruptcy. With this in mind, there is the possibility that a company that has been given a default rating may recover and receive a performing rating again. This, of course, contradicts the assumption that the default state is absorbing, which is further elaborated on in chapter 4.3.

2.3.6 Understanding the rating input

In this report, probability of default will be estimated using credit rating input from a wide range of different companies. Even if one were to make some adjustments later on due to macro conditions or company profile, it is relevant to know how the rating input is determined to begin with. Understanding the rating input might also help to interpret the results.

There are two important distinct classifications of rating systems; through-the-cycle (TTC) and point-in-time (PIT).

The PIT rating describes the actual creditworthiness of a company for a certain time period. This makes it dependent on e.g. macroeconomic cycles, since obviously the rating should generally be better for a majority of companies (and PD less) if there are good times ahead compared to if there is a recession ahead. The PIT rating should evaluate all available information at the time, and then set a PD that is constant over the considered time period ahead.

If a rating is TTC, the aim is that companies should have the same rating through the whole economic cycle. As mentioned in Andersson & Vanini (2010), the TTC ratings are sometimes referred to as stressed ratings since they should stay the same over time, especially during a period of financial distress. In contrast to the PIT PD, the TTC PD should vary over time for a certain rating grade.

Thus, one expects a migration matrix where the ratings are TTC to be heavier on the diagonal than the corresponding migration matrix estimated with PIT ratings. Therefore, one should in theory see more migrations between performing (non-defaulted) ratings in a PIT migration matrix over time, whilst the PIT PD should change little over time. On the other hand the opposite is reasonable for a TTC migration matrix, i.e. not that many observations of movements between rating grades, but a movement of the PD over time within each rating class.

In reality though, most rating models are a mix between the two and it is of course very hard to get a model to be 100% TTC.

There is also a difference between calculated and approved ratings. Calculated rating is something that a model suggests based on input parameters. However, often an expert judgement can override the quantitatively set rating. The expert might take into account other more soft values about the company's management, or in other ways use his or hers deeper knowledge about the company in question. The impact of this is however not the focus of this report.

Finally, there might be different rating models that feeds rating data to the same database. The level TTC versus PIT may not always be determined for each specific model and it may also vary between the models. The reasons for having different rating models is of course that different companies might need to have different input parameters or parameter weights. Nevertheless, the goal of the different rating models is the same, namely to estimate an as accurate rating as possible. Therefore, the actual ratings should not behave particularly different between the models. However, it is still important to keep in mind in case there are some special limitations to a certain model.

2.4 Time-inhomogeneity

One often made assumption is that the Markov process is time-homogeneous. That implies that the migration matrices will stay the same over time, which makes the estimations easy to extrapolate.

However, there is evidence that rating migrations are not time-homogeneous. The degree of PIT/TTC is one explanation but there may be other reasons as well. For instance, rating models develop over time and becomes more sophisticated and better to discriminate between good and bad borrowers. Both Bangia et al. (2002) and Rachev & Trueck (2009) shows that default rates vary over time, and that different migration matrices are obtained if they are estimated during recession or expansion. Also the Annual Global Corporate Default Study from Standard & Poor's shows that the default rates vary a lot over time (see e.g. Chart 21 in the 2012 report [24]).

Even though evidence of time-inhomogeneity has been present in the academic literature for some time, there is no standard way to try to mitigate or account for it. In this report, the internal data set will be tested for time-inhomogeneity.

2.5 Short on previous studies

This section will briefly go through the evolution of some previous studies within the field of credit migrations related to this thesis.

In their work on credit risk spreads, Jarrow, Lando and Turnbull (1997) were the first to model transition probabilities and defaults using a Markov chain framework on a finite state space that represented different rating classes. Their work increased the attention on using transition matrices and Markov chains to model credit migrations. Already at this time, a generator matrix was proposed to create a homogeneous Markov chain in continuous time.

In 2001, Israel, Rosenthal and Wei published an article on how to find generators for Markov chains via empirical transition matrices. Their article focus much on when a generator exist. As an example, they found and formulated conditions regarding the so called embedding problem.

That means a generator matrix does not exist in certain cases, and therefore cannot be computed via the relation $M(t) = e^{tG}$. This is discussed more in section 2.3.4.

Lando and Skødeberg published their article *Analyzing rating transitions and rating drift with continuous observations* in 2002, which has been frequently referred to in later research. In their article, they look at both discrete and continuous time Markov chains and describe some differences. They also note that the embedding problem often occurs in real data. Moreover, they find evidence of non-Markov behaviour such as rating drift. One reason to rating drift can be that rating agencies are reluctant to downgrade several rating grades at one time, but rather downgrade the rating one step two or three times in a rather short interval. Rating drift is not the focus of this thesis, but is alongside time-inhomogeneity a topic that has been popular to investigate in more recent times. Note that a non-Markov behaviour causes time-inhomogeneity, which is a somewhat broader area.

Bangia, Diebold and Schuermann publish an article in 2002, focused on rating migrations through the business cycle. They try to reject the Markov assumption through eigenvalue analysis, but find that hard. However, they introduce eigenvalue analysis as a way of comparing and measuring migration matrix differences as they evolve over time.

In 2004, Lando publishes his book *Credit Risk Modeling - Theory and Applications*, which is a good overview of different models, results etc. at that time. A chapter on rating migrations via Markov chain models can be found but no ground breaking new steps are taken. Some concepts are elaborated on a little further.

At this time, focus is also put on more practical problems with the Markov chain models. One example is Christensen et al. (2004), who try to estimate confidence intervals for rating transition probabilities with a special focus on rare events. They suggest a bootstrap procedure, where they use a model to simulate fictive rating histories. Furthermore, they look at the non-Markov behaviour that Lando and Skødeberg found evidence of in 2002. Moreover, they note that real data sets often suffers from a lack of data, which makes confidence set estimations difficult for rare events. One example of a rare event can be the default of an investment grade rated company, i.e. a migration from an investment grade rating directly to the default rating. In 2005, Hanson and Schuermann also publish a thesis where they look at different ways of estimating confidence intervals. In one part they look at analytical options, such as the Wald confidence interval, Agresti-Coull confidence interval and Clopper-Pearson confidence interval. However, they also look at bootstrapping procedures and finds that they are in most cases tighter than the analytical options. The only advantage they see in e.g. the Wald interval is that by using it one is able to derive genuine (analytical) confidence intervals. They suggest bootstrapping on actual rating histories rather than simulating them as Christensen et al. (2004). Moreover, Trueck and Rachev (2005) also publish an article where they estimate confidence intervals with the purpose of calculating credit Value-at-Risk. They use the methods proposed by Christensen et al. (2004) and Hanson and Shuermann (2005). As it turns out, they also speak in favour of bootstrapping since e.g. Wald intervals depend so heavily on the number of firm years, which is evident in equation B.1.

Also in 2004 Jafry Schuermann publish their article *Measurement, estimation and comparison of credit migration matrices*, focused on measuring differences between matrices. They present different types of norms, such as the L_1 norm, the L_2 norm and different variations of these. One example of a variation is to subtract the identity matrix from the migration matrix, something they introduce as "mobility matrix". The mobility matrix roughly resembles the generator matrix. They also develop a measure based on singular value decomposition.

In 2009, Trueck and Rachev publish a book called *Rating based modeling of credit risk* that put together much of the current findings and results, much like Lando's book did five years earlier. The book presents a good overview over a number of areas.

Different attempts of detecting, testing and modelling non-Markov behaviour and time-inhomogeneity has been done in more recent years. Kiefer and Larson (2006) test time-homogeneity using the χ^2 test originally introduced by Anderson and Goodman (1957). They find that time-homogeneity is easy to reject over longer time periods.

Bluhm and Overbeck (2007) tries to drop the homogeneity assumption of the generator matrix, and allows it to evolve over time. By calibrating parameters after observed data, they alter the generator matrix and get a PD term structure that fits well with the observed PD term structure. However, as they point out, their method is an interpolation approach rather than extrapolation approach. This is because their method only allows for a fit within an observed time period.

Andersson Vanini (2010) attempts to account for time-inhomogeneity by estimating direction and speed of migrations within the migration matrix, to create a regime-shifting migration matrix. Moreover, regarding the speed and direction, they provide a small discussion on the differences of Point-in-time and Through-the-cycle ratings. Their work is somewhat based on previous work by Andersson in 2007 and 2008. The regime shifting matrix can keep static generators by using two markov chains, one for upgrades and one for downgrades. They aim at an application in form of credit derivatives, and therefore introduce stochastic time changes and dynamics to the Markov chains. Their focus is out of scope of this thesis, but their ideas are nevertheless very interesting.

The academic research today is much focused on trying to find ways to describe and model non-Markov behaviour and time-inhomogeneity stemming from e.g. the different economic conditions over a business cycle. To this date, there is no real consensus regarding exactly how to handle the problem of time-inhomogeneity when using the Markov chain approach to estimating rating migrations.

Chapter 3

The data set

In this chapter, a description of the data set and some data statistics will be presented.

3.1 Data description

The dataset is a time series of rating changes taken from an internal database of business counterparties. The rating input is believed to be a mix of TTC and PIT. The estimated distribution between TTC and PIT is not possible to disclose in this report, and is of no great importance since estimations has to be made on this type of TTC/PIT mix regardless of the distribution between them. Furthermore, it consists solely of approved ratings. The data set also contains numerous rating models. The different models are aimed at rating different types of companies, e.g. financial institutions or real estate companies. Other models might be of an older type that is no longer used. Another fact to have in mind is that the models have evolved over time and have become more and more sophisticated.

The database of course contains a lot of information, but there has been a first round of filtering from the database to receive what will be referred to as "the original data" in this report.

The original data only contains counterparties that have received an internal rating. The counterparties that receive an internal rating are all legal entities with liabilities towards the company that exceeds a certain threshold. The threshold is set rather low, meaning that even small companies are included.

The different data fields that exist in the data set, and that each data point (each observation) has, are:

`customer_id` - *A code used to identify each different customer*
`nace_code` - *(Nomenclature des Activités Économiques dans la Communauté Européene), a code that classifies which sector the economic activity of the customer belongs to*
`rating_model` - *The internal rating model used to when rating was calculated*
`rating_value` - *The actual rating grade the customer received*
`date_from` - *The date when the rating was set*
`date_to` - *The date where the rating set at* `date_from` *ceases to be valid*

The `customer_id` is there to be able to distinguish between different counterparties and to be able to get each customer's rating history throughout the years.

The `nace_code` will be used to classify which companies that are so called "defensive sector companies" and which are not. This knowledge will later be used to test whether business sector has impact on the results, which is relevant information regarding the time-homogeneity assumption. It might be so that some sectors are more or less sensitive to the economic cycles, which makes the time-homogeneity assumption less or more accurate. Whether the time-homogeneity assumption is accurate is of course relevant for investors keen on using it, since homogeneity is more straight forward and allows for an easier implementation.

The `rating_model` is relevant for cleansing and sorting data, which is further elaborated on in section 4.1.4: Different rating models. It is also relevant if one want to exclude certain rating models due to that they are outdated, or maybe that some issues have been found with certain models.

For obvious reasons, the `rating_value`, `date_from` and `date_to` are relevant for the migration estimations.

A new data point is only created when there is a change in `customer_id`, `nace_code`, `rating_model` or `rating_value`. The `date_from` is set as either the initial date (first observation), or as the new date of the observation when something changed. The `date_to` is initially set by the ones that determined the rating, so the rating is valid until that date. However, if a change appears this is treated as newer and more accurate information and therefore overwrites the old observation even if the change occured before the initially set `date_to`.

Below is an example that clarifies how the date transformation works when the previously mentioned fields change. Note how the `date_to` of the first observation changes when it is overlapped by the new more accurate observation containing a rating downgrade.

Example 3.1: A company is being downgraded before the end date of the previous rating.

customer_id	nace_code	rating_model	rating_value	date_from	date_to
U1212	38	the_model	AA	25/11/2007	25/11/2008
U1212	38	the_model	A	19/07/2008	19/07/2009

Figure 3.1: An example of two observations before date transformation.

customer_id	nace_code	rating_model	rating_value	date_from	date_to
U1212	38	the_model	AA	25/11/2007	**19/07/2008**
U1212	38	the_model	A	19/07/2008	19/07/2009

Figure 3.2: An example of the same two observations after date transformation.

The data set in its original form contained about one million data points spread over the years 2002-2013. However, due to the quality of the data, the data set had to be cleansed even after the first round of filtering from the database. More details about the process of cleansing and adjusting the data set can be found in section 4.1 in the methodology chapter. After the cleansing and adjustments the final data set which is the foundation for all estimates contained about 90% of the original amount data points. In the section immediately below, descriptive statistics of the final data set for the cohort and duration estimations is presented. The statistics are focused on non-diagonal movements as actual movements on the diagonal are not measured for the duration method.

3.1.1 Descriptive statistics

Table 3.1: Descriptive statistics over final data during years 2002-2007, split by estimation method. Cohort year 2002 is set as index 100 for firm years and non-diagonal movements.

	Year	2002	2003	2004	2005	2006	2007
Cohort	Firm years	100	304	364	386	397	457
	Non-diagonal movements	100	289	207	320	344	327
	Upgrade movements (%)	54.3%	53.3%	60.0%	59.1%	55.6%	56.8%
	-of which recoveries	0.2%	1.6%	2.5%	1.6%	2.6%	1.3%
	Downgrade movements (%)	45.7%	46.7%	40.0%	40.9%	44.4%	43.2%
	-of which defaults	5.5%	6.9%	4.5%	3.2%	2.6%	3.5%
Duration	Firm years	169	340	399	449	492	652
	Non-diagonal movements	124	313	497	386	460	499
	Upgrade movements (%)	50.4%	51.6%	54.9%	58.5%	56.8%	56.3%
	-of which recoveries	0.9%	1.9%	2.2%	1.8%	2.6%	1.7%
	Downgrade movements (%)	49.6%	48.4%	45.1%	41.5%	43.2%	43.7%
	-of which defaults	7.6%	6.7%	3.9%	3.2%	2.6%	3.0%

Table 3.2: Descriptive statistics over final data during years 2008-2013, split by estimation method. Cohort year 2002 is set as index 100 for firm years and non-diagonal movements.

	Year	2008	2009	2010	2011	2012	2013
Cohort	Firm years	493	483	548	626	650	18
	Non-diagonal movements	278	369	381	355	348	4
	Upgrade movements (%)	47.5%	30.4%	54.0%	57.4%	50.3%	38.7%
	-of which recoveries	0.8%	0.5%	0.9%	1.2%	0.9%	0.4%
	Downgrade movements (%)	52.5%	69.6%	46.0%	42.6%	49.7%	61.3%
	-of which defaults	4.0%	5.8%	4.2%	4.6%	6.6%	1.3%
Duration	Firm years	661	605	662	723	723	712
	Non-diagonal movements	500	638	576	521	498	371
	Upgrade movements (%)	46.9%	30.9%	50.8%	53.6%	47.1%	51.2%
	-of which recoveries	1.1%	0.9%	1.1%	1.3%	1.1%	1.7%
	Downgrade movements (%)	53.1%	69.1%	49.2%	46.4%	52.9%	48.8%
	-of which defaults	4.1%	6.1%	4.6%	4.5%	6.6%	4.8%

Chapter 4

Methodology

This chapter describes the process from obtaining a data set to producing the results. It describes adjustments that has been done to the data set, as well as methods that has been used to test and measure the results. Furthermore, it gives more background as to why certain adjustments, tests and methods have been used.

4.1 Adjusting the data set

As mentioned in chapter 3, the data set had to be cleansed even after the first round of filtering from the database. The below sections describe how and why this cleansing was done.

4.1.1 Removing problematic observations

Some observations during the years 2002 and 2003 contained ratings of a different type, and not the type of ratings used otherwise. The ones clearly of another type only accounted for 0.08% of the observations originally within year 2002 and 2003. However, due to the features of the ratings, there could also be a few ratings hidden within the data that were not possible to sort out. The number of rating observations possibly hidden was calculated to be about 0.005% of the observations during years 2002 and 2003. Because of the small number and the fact that they were likely to be spread out evenly within the different ratings, this will not cause a problem. Also, a few observations were found where different ratings were given to the same company the same day. To mitigate this, some rules were set up to decide which observation is the most likely to be correct, e.g. to check which observation that had the newest set of so called NACE codes (for an explanation of what a NACE code is, see section 4.1.3).

4.1.2 Mapping ratings

Due to the sensitive nature of the internal dataset, the original ratings have been mapped to ratings similar to those used by Standard and Poor's (S& P). This it to somewhat disguise the original features of the internal data set. Exactly how the mapping was done is not possible to disclose in this report, but the highest to lowest ratings were mapped into the new ratings {AAA, AA, A, BBB, BB, B, CCC, D}. The highest rating is AAA, and then they follow in descending order to CCC and finally D which is the defaulted rating. The rating names are inspired by Standard and Poor's rating system, but should in no respect be taken to fulfil requirements set by S& P. However, I will refer to "investment grade ratings" as ratings AAA to BBB, and "speculative grade ratings" as BB to CCC, as in the S& P framework.

4.1.3 Mapping NACE codes

The NACE code (Nomenclature des Activités Économiques dans la Communauté Européene) is a standardised way of classifying economic activities within the European union. Therefore, it is useful when separating companies into different industries.

The NACE code is formatted as the following example: $Q86.23$, where

$$\begin{aligned} \textit{Section}: &\quad Q &&\quad \textit{Human Health \& Social Work Activities} \\ \textit{Division}: &\quad 86 &&\quad \textit{Human Health Activities} \\ \textit{Group}: &\quad 86.2 &&\quad \textit{Medical \& Dental Practice Activities} \\ \textit{Class}: &\quad 86.23 &&\quad \textit{Dental Practice Activites} \end{aligned}$$

Thus, Q is the most general classification, which is then followed by integers, and up to two decimals as the most detailed description of economic activity.

The code number has however been revised from what is known as revision 1.1 (rev. 1.1) to revision 2 (rev. 2) during the period that the data was collected. Since the original data has been collected from several sources, this caused some observations that were recorded more than once, since the different NACE codes was treated separately when sourcing the observations from the database. To be able to later split the data set, it is essential that NACE codes follow the same framework. I.e. it is necessary to translate all rev. 1.1 codes to rev. 2 codes.

Even though Eurostat stated a date from when the new revision 2 should be implemented, this sharp date is not clearly present in the data. Furthermore, some businesses have actually changed their business over the years which rightfully should give them different NACE codes. Unfortunately, some NACE code numbers exist in both rev. 1.1 and rev. 2, but mean completely different things. Moreover, it is desirable to get rid of observations that are copies apart from the NACE code, since it will give the impression of two migrations when it actually is one. Unless mitigated, this will distort the results. Getting as many correct NACE codes as possible is also desirable when looking into time-inhomogeneity for non-cyclical counterparties, since they decide the data set split. Therefore, using correspondence tables from Eurostat, the following steps was taken to mitigate this issue to the greatest extent possible:

- One-to-one list mapping

  The one-to-one list mapping used the fact that sometimes, one unique number in rev. 1.1 had one unique counterpart in rev. 2. This meant that these could safely be mapped regardless of the `date_from` and `date_to`.

- Unique list mapping

  The unique list mapping meant that some numbers were found only in rev. 1.1 but not in rev. 2. These were mapped according to the correspondence table regardless of the `date_from` and `date_to`.

- Pre-effective date mapping

  The pre-effective date mapping used the effective date of the new revision 2, Before the effective date of rev. 2, NACE codes could safely be mapped with the correspondence tables, as these early reported numbers should have used revision 1.1. A closer look at the data revealed that some rev. 1.1 NACE codes were lagging behind even after the effective date. These are the ones that we want to mitigate with the one-to-one and unique list mapping.

- Cleansing of codes $99, 98$ and 0

  Cleansing of NACE codes 99, 98 and 0, meant in practice that observations with these codes were removed if there was an identical copy of it, except that it had a NACE code other than these three. It was discovered while manually cleansing the data for problematic ratings that often the codes 99, 98 and 0 occurred as copies. Furthermore, the codes 99 and 98 are more of a "miscellaneous-type" code, and the code 0 does not even exist in the NACE framework.

- Reduction from class codes to division codes

  The last bullet point, reduction from class codes to division codes, is due to the fact that reporting of NACE codes could sometimes vary slightly for the same business. One example is "Q88.10 - Social work activities without accommodation for the elderly and disabled" and "Q88.99 - Other social work activities without accommodation". These are essentially the same (especially in an economic cycle point of view) even though they have different codes. Since there are a lot of different codes and not enough data to get good estimates of all of them, only the first two numbers were used in the final stage of NACE code cleansing. The fact that some NACE codes probably still are wrong due to data quality makes this rationalisation even more reasonable when aiming for stable results. This also makes sense in the economic cycle point of view and furthermore allows discarding of "copies" such as the example in this bullet point.

As mentioned, all mapping was done via correspondence tables provided by Eurostat (see [7]). However, since the rev. 2 is more detailed than rev. 1.1, there will inevitably be some one-to-many transformations. After investigating this a bit further, it was concluded that a map to one of the many class choices is in the absolute majority of the cases enough to put the mapping into the right division. For the purpose of the NACE codes, namely selecting divisions that are sensitive to the economic cycles, this was deemed to be sufficient.

The cleansing and mapping of NACE codes above solved about 65% of all problematic observations. It was further noted in the correspondence table that the mapping between rev. 1.1 and rev. 2 was from a higher to lower number in 70% of the cases. Thus, by sorting correctly and then correcting overlapping dates, the overlapping should have removed the observation desired to remove in 70% of the remaining problematic cases. Therefore, it is at this stage estimated that at least 90% of the NACE codes were mapped correctly. Furthermore, some of the last problematic cases were probably correct already, which may increase the number of correct mappings a slight bit more.

4.1.4 Different rating models

The data set contains ratings where the calculated rating came from different rating models, 14 in total. Some models are designed to rate different types of companies and some have replaced older models over the years. Because of this, some observations might have been identical apart from the rating model, since rating model also was treated separately in the database sourcing. Therefore, it is important to know which models that should be prioritised when they overlap or are copies apart from the rating model name.

The prioritisation was made by looking at the data set too see the minimum, maximum and average date for each model, to be able to get a view on which models were newer and more accurate. The final prioritisation was done by also consulting internally. The final list was used for sorting observations just before overlapping observations was fixed (see section 4.1.5), thus the higher prioritised models overwrote the lesser prioritised ones.

Furthermore, after a first round of estimations, some strange patterns that had been seen in the data set emerged more clearly in the migration matrices. This was found to stem from a certain rating model which had some special features, e.g. less granular rating possibilities. As an example, this could mean that a company only could receive rating AAA and A, but not AA. The exact features of this model can not be publicly disclosed, but nevertheless they were deemed to not be suitable for estimation of migration matrices. The observations from this model was removed, which shows the importance of investigating each rating model's properties.

4.1.5 Overlapping observations

Overlapping observations can distort the results and is not desirable. After all cleansing of the data set was done, the final step taken was to make sure that the `date_from` and `date_to` did not overlap. Sorting of the data had been done to make sure that higher prioritised rating models and NACE codes should overwrite lesser prioritised observations if they overlap. The overwriting was then done as described in chapter 3, Example 3.1. Thus, the final data set allows for individual companies to have gaps in their rating history, but never parallel rating histories.

4.2 Transitions to and from non-rated state

In practice, companies can transition from having rating "Non-Rated" or "Unrated" to be given a rating defined in the rating migration matrix (in the state space) and vice versa. This can occur if the ratings that are examined, AAA to D, have certain conditions that has to be met in order to receive a rating. The conditions can e.g. be based on size of the company or, if the company is a counterparty with outstanding loans or trades, an exposure risk threshold.

These transitions to and from non-rated states are therefore out of scope of the ratings migrations under examination. More formally, these types of transitions does not occur within the state space S that the model is used for. Therefore, they are seen as non-informative and those transitions are removed from the data set. Worth to mention is that this affects the duration method estimates less compared to the cohort method, since ratings are observed much more frequently and therefore more information is kept.

4.3 Handling defaults that recover

As discussed regarding the definitions of default in the regulations, a default does not imply that a company in reality is in such a bad shape that it will never operate again. As mentioned previously, delayed payments might set a company in default. This in turn also affects it's subsidiaries or parent companies as well as outstanding loans and/or derivative contracts that these companies might have, in accordance with EU regulations.

The applied Markov chain model found in theory normally treats the default state as *absorbing*, i.e. once a company has defaulted it cannot leave the defaulted state. In practice, though, a defaulted company recovers every once in a while. It can e.g. be so that it has found new investors that can cover its cash flows for an extended period of time, and therefore is back in business. The company then receives a new performing (non-defaulted) rating.

This of course collides with the absorbing state assumption. However, the company that recovers from default have been reassessed based on its new (economic) conditions, conditions that did not exist before. One can therefore think that the reassessed company has been given the rating that best fits its current conditions, and is therefore different to the company that deserved the default rating. Due to this, it is still reasonable to view recovered defaulted companies as new observations and therefore there is no contradiction in using the absorbing assumption.

Furthermore, the absorbing assumption ensures that the cumulative probability of default is increasing with time. As described in Lando (2004) [15], removing the absorbing state would change the interpretation of the estimated PD values. Removing the absorbing condition would result in an estimation of *being* in default at a certain time, whereas keeping the absorbing condition would result in an estimation of *having defaulted* at a certain time. The latter describes the cumulative PD, which is the desirable estimation to use from the risk manager's point of view, see e.g. Lando (2004).

4.4 More weight to recent years

As mentioned earlier, rating models evolve over time. With this in mind, more recent ratings produced from recent models might be more relevant and accurate. Thus, one might want to give more weight to recent observations or recently estimated migration matrices since they might be a better estimate of the future in terms of rating model output. Furthermore, estimates where more data points have been used are more certain than estimates where less data points have been used. As an example, the cohort estimate for 2013 is not very certain, since the data points in that sample is far smaller than any other year.

By looking at the firm years found in descriptive data statistics in Chapter 3, it is evident that the observation universe grows larger with time. When estimating an average migration matrix, this will automatically give more weight to more recent years where more recent rating models have been used. If the number of firm years changed drastically combined with a drastic change of economic conditions, this might give too large weights to unlikely events and thereby skew estimates for the average matrices. However, for the data set used in this report there has been a steady build-up of firm years. There has also been a mix of recessions and expansions, spread out over the time scope of the data set both for recent years and for the whole time span.

Therefore, by giving each single observation point equal weight, more weight is gradually given to more sophisticated rating models. It is also makes the weight proportional to the number of observations, which is reasonable from a probabilistic point of view. Equal weight to each single observation is therefore used in this report, but as mentioned, this is something that at least should be considered before calculating migration matrices.

4.5 Estimation and validation set

After the final cleansing, the data set was divided temporarily into two, one estimation set and one validation set. All companies were listed by a company id and then sorted A to Z. From this list, every second company was put into the validation set. The remainder formed the estimation set. This kind of separation was made to make the separate sets have observations reasonably even spread out over the whole estimation window, ensuring that they were not in totally different parts of the economic cycle.

The purpose of the validation test is to check that estimations from one set of companies have a good enough fit on a completely different set of companies. The two sets are taken from roughly the same 12-year time period, and using roughly the same rating models. This means that the economic and rating model conditions should be about the same between the two sets. Thus, the setting of this test is that either the time-homogeneity assumption is on average valid, or that one have modelled the state of the economy etc. roughly correct. If the estimations between the two sets are very similar, then we have shown that the estimates are not company-specific on average, which of course is desirable.

In this report, the average migration matrices for both cohort and duration method will be compared. These estimates could then be seen as quite general since the past 12 years has seen both expansions and recessions and therefore might be a good average representation of times to come. This approach of comparing averages is more plausible than comparing specific years when one does not want the possible effect of economic cycles.

4.6 Estimation in Matlab

When all data had been cleansed and adjusted, the actual estimation of migration matrices were done in MATLAB. Furthermore, estimating their confidence intervals via a bootstrap method requires MATLAB (or similar software), since it will involve calculating several thousands of migration matrices.

4.6.1 Migration matrices

Depending on what migration matrix that was supposed to be estimated, data was split into different Excel files that was then read into MATLAB where all the calculations were done. The theory stating exactly how to calculate is found in Chapter 2.

Briefly, the entries of the cohort migration matrix was estimated by

$$\hat{p}_{ij}(t_k) = \frac{n_{ij}(\Delta t_k)}{n_i(t_k)}$$

where i and j are rating states and t_k is January 1st each observed year k. Furthermore $n_{ij}(\Delta t_k)$ is the number of migrations between ratings i and j between year t_k and t_{k+1}, and finally $n_i(t_k)$ is the number of companies in rating i at time t_k. The calculations result in a one year migration matrix since $\Delta t_k = 1$ year.

For the average matrix, all data from 2002 to 2013 was used, where observations of rating had been done each January 1st.

When calculating using the duration method, migration rates in the generator matrix were estimated using equation (2.18) and equation (2.19). To calculate the matrix exponential, the MATLAB function `expm` was used. As with the cohort method, different Excel files were read into MATLAB depending on the time frame used. For an annual matrix, only data for a single year was used, whereas all available data was used for the average matrix.

The results were then checked against estimations that had previously been done more manually step-by-step in Excel to verify that the code worked properly.

4.6.2 Confidence intervals

Two methods of calculating confidence intervals (CIs) are presented in this thesis, see section 2.1. The first method is the analytical Wald CI. The second, however, is calculation via a bootstrap technique. Hanson and Shuermann (2005) suggest bootstrapping on realised rating histories, as they regard a rating history as "the basic data unit from the perspective of PD estimation". The method used here contrasts to both theirs and to the method proposed by Christensen et al. (2004) by using *one rating migration* as the basic data unit and use realised rating migrations. This also ensures that the same amount of data is used in each bootstrap, as a contrast to Hanson and Shuermann (2005) where the number of firm years could vary slightly.

The bootstrap procedure is done via MATLAB, in accordance with the theory presented in section 2.1. In every confidence interval calculation, a sample containing 1000 bootstrap replications is used. To clarify, one observed empirical sample is bootstrapped 1000 times (creating 1000 bootstrap replications). From each bootstrap replication, one migration matrix is estimated. Thus there are 1000 estimated migration matrices that form the confidence interval for each rating migration. There is no predetermined number of bootstrap replications that has to be used, but is rather dependent on the computer power and time available.

4.7 Investigating time-inhomogeneity

The investigation of time-inhomogeneity will be separated in two main sections, *detecting* time-inhomogeneity and *testing* time-inhomogeneity. Detection is thought of as how one could see evidence of time-inhomogeneity in the data sample, whereas the testing part is how to actually say something statistically about the time-homogeneity assumption. Since this report is more focused on application, testing is more relevant. Two ways of statistically being able to say something about the time-homogeneity assumption will be presented, each with a little different angle.

4.7.1 Detecting time-inhomogeneity

There are several ways of detecting that time-inhomogeneity is present, where most of them are based on comparing annual matrices to the average matrix. Either distances between matrices (norms) or their eigenvalues/eigenvectors are compared, as in Jafry and Shuermann (2004), Yang (2008), Trueck and Rachev (2009) or Lencastre et al (2014). Jafry and Shuermann (2004) also develop a measure based on singular value decomposition. Moreover, there are also several small variations of the above concepts.

Developing and comparing measures of how the matrices differ is a substantial task in its own right. Furthermore, what these measures have in common is that they give relative values of how matrices compare to each other, but no real statistical test to determine if there is a significant difference. Moreover, Yang (2008) also point out that they don't discriminate between upgrades and downgrades, and does not give useful economic information about a rating movement. For these reasons, investigating different ways of detecting time-inhomogeneity is not the focus of this report. However, we will present one way of doing so by calculating the L_2 norm. The method is describe immediately below.

Matrix L_2 norm

One way of detecting time-inhomogeneity is to measure the cell-by-cell distance between the annual matrices and the average matrix. The cell-by-cell distance can be calculated with e.g. different norms such as the L_2 norm, as mentioned in Jafry and Shuermann (2004). The L_2 norm is easy to understand as the sum of all norms described below in equation (4.1) is the standard deviation of the annual matrices. The L_2 norms serve as a baseline estimation. Ideally (if the homogeneity assumption is correct) the distance should be zero for each cell in each year, but at least it is desired to only show small variations. The major drawback is that there is no real way of measuring how small is "small", or what the difference should be between two L_2 norms for it to be large enough to indicate time-inhomogeneity. Neither can it state how certain that indication is, as with most of the measurement techniques (see Yang (2008)).

Nevertheless, this type of distance measuring at least shows a relationship between the different migration matrices. Since variations of distance measuring between matrices is seen quite often in the literature, the L_2 norm which is easy to understand will be calculated and presented in chapter 5. As described in Jafry and Shuermann (2004), Yang (2008) and Trueck and Rachev (2009), the L_2 norm between two matrices M_1 and M_2 is calculated as

$$\Delta M_{L2}(M_1, M_2) \equiv \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(M_{1,i,j} - M_{2,i,j})^2}}{N(N-1)} \tag{4.1}$$

where N is the number of ratings and i, j are rating states. The -1 in equation (4.1) is because the defaulted state is excluded from the calculation, since it always is equal between migration matrices.

4.7.2 Testing for time-homogeneity

The assumption of time-homogeneity is common in practice, as several papers previously have mentioned, e.g. Lando and Skødeberg (2002). However, authors such as Jafry and Schuermann (2004), Kiefer and Larson (2006) note that the assumption of time-homogeneity probably isn't valid, and it is known that it is a somewhat simplified view. Nevertheless, it is important to test the assumption as its accuracy will be different for each data sample.

Two tests are presented in this thesis, one χ^2 test and one based on bootstrapped confidence intervals. Each test has a different approach to the time-homogeneity assumption. If homogeneous, then every element in the migration matrix should be constant. The χ^2 test is designed to test whether we can reject a null hypothesis that the annual migration matrices are equal to the average migration matrix, which they should be given time-homogeneity. The test involves calculating differences of every element in the matrices.

The other test is to compare 95% confidence intervals of PD estimates to see if they give significantly different estimates between two adjacent years. If two estimations differ significantly, then this means that it is very unlikely that they could produce the same estimate. As we just mentioned, time-homogeneity implies that estimates are constant. If two estimates cannot be the same, then it cannot be constant. Because of the rather cumbersome task of comparing every element and its confidence interval in several matrices, this test will focus on the PD estimates.

Thus, one test compares whole annual matrices over the whole time span to see if they can be the same as the average matrix. The other test looks at PD estimates of adjacent years to see if they possibly could produce the same estimates.

Chi-square test

A way of statistically testing time-homogeneity is to use a χ^2 test. The test was originally stated by Anderson and Goodman (1957), but several other authors have used it more recently too. e.g. Kiefer and Larson (2006) and Trueck and Rachev (2009).

Following the approach described by Anderson and Goodman (1957), the test is designed to see whether transition probabilities are constant. The full sample are divided into T independent subsamples, which in our case means $t = 1, 2, .., 12$, for the $T = 12$ years from 2002 to 2013. We now want to test if the migration matrices from the subsamples differ significantly from the average migration matrix calculated using the whole sample.

Let $\hat{p}_{ij}(t)$ denote the transition probability from rating i to rating j during year t. Let also $\hat{p}_{ij}$ denote the corresponding transition probability estimated with the full sample. Furthermore let n_i be the number of firm years for rating i. Our null hypothesis, H_0, is

$$H_0 : \hat{p}_{ij}(t) = \hat{p}_{ij} \quad \forall i = 1, 2, \cdots, m \quad \forall j = 1, 2, \cdots, m, \quad \forall t = 1, 2, .., T \tag{4.2}$$

and the test statistic

$$\chi^2 = \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} n_i(t) \frac{(\hat{p}_{ij}(t) - \hat{p}_{ij})^2}{\hat{p}_{ij}} \tag{4.3}$$

is shown by Anderson and Goodman (1957) to follow the χ^2-distribution with $m(m-1)(T-1)$ degrees of freedom. In our case, $m = 8$ since there are eight rating states. Thus, if our observed test statistic calculated in equation (4.3) above exceeds the tabulated $\chi^2(7{\cdot}8{\cdot}11)$-value at a certain significance level, then we can reject the null hypothesis at that significance level. Rejecting the null hypothesis means that we can reject the assumption of time-homogeneity.

Confidence intervals

By estimating migration matrices for specific years and using the confidence intervals associated with each estimate, it is possible to see if the estimations differ significantly from each other. If they do, it is likely that time-inhomogeneity is present. However, it gives us no exact number on how likely it is that there is time-inhomogeneity, such as a p-value.

Nevertheless, the approach of using confidence intervals gives us another interesting angle on time-homogeneity. The χ^2 test is used to see whether the annual matrices differ from the "true" matrix, which is estimated by the average matrix. They are not supposed to be differ significantly if the time-homogeneity assumption is correct. However, by the definition of a homogeneous Markov chain, the estimated probabilities in each cell of the migration matrix should be constant. Thus, if estimations of e.g. $p_{BBB,D}$ differ significantly from each other between different years, then we can say that time-inhomogeneity is present, since they should be able to produce the same estimate.

Therefore, another test for time homogeneity will be a bootstrap method together with confidence interval comparisons. Naturally, one source of time-inhomogeneity is that rating models evolve over time as mention in Lencastre et al (2014) and discussed in section 4.1.4. To focus on the state of the economy and to avoid a possible effect of rating models to the greatest extent possible, only adjacent years will be considered. The years considered will also be recent in time. This is because we have seen volatility in recent years (increases probability to find inhomogeneity), but also because the rating models in recent years are more relevant. Furthermore, it is sufficient to find one significantly different estimation to say that we have found time-inhomogeneity. However, comparing every single cell's confidence intervals is cumbersome, and we will therefore focus on the most important estimations for this thesis, namely the PD estimations. Confidence intervals will be compared for both the cohort and duration method.

4.7.3 Defensive sectors and time-homogeneity

As described by a recent article in Forbes (see [8]), the traditional advice to prepare for possible market corrections is to invest in defensive stocks or sectors. The idea is that no matter the state of the economy, people will have to eat, drink and take their medicines. Thus, companies that produce these kinds of products should be better off than others in downturns, at the cost of maybe not giving as high returns when the market is up. Defensive companies are sometime also called "non-cyclical", as they are supposedly less correlated to the economic cycle compared to the average company. By applying this logic to rating migrations, companies within these types of sectors might be experiencing less migrations over the business cycle since they have a more stable cash flow. Therefore, it is possible that the time-homogeneity assumption is more applicable when looking at defensive sectors. To test whether this assumption is true is relevant to those that have credit risk exposures towards defensive sectors. If it turns out that the time-homogeneity assumption is true for these sectors, then this would ease the burden of applying a Markov chain migration approach, since they would not have to care about how to account for possible time-inhomogeneity.

In this thesis, it will therefore be investigated if companies within defensive sectors are less exposed to time-inhomogeneity. A subset of defensive companies is formed and will go through the same time-inhomogeneity investigation as the full data sample.

To be able to construct a subset, we look at the NACE code which describes the economic activity of a company. By using the NACE codes, companies within agriculture, manufacture of food products, pharmaceuticals, health care, certain real estate activities etc. was chosen to the subset. The size of the defensive companies subset is about 55% of the full sample. It is desirable to not make the subset too small, as this would cause problems when splitting it further into annual data sets. For a complete list the chosen defensive sectors, see Appendix C.

4.7.4 Recession and expansion matrices

If time-inhomogeneity is shown to be present in the data, and if past data is deemed as a good representation of times to come, then it will likely be present in the future as well. Solving the issue of time-inhomogeneity is not straight forward, some attempts have been done as we saw in section 2.5: Previous studies. However, there is no consensus about a feasible method.

As this report is focused on the application of a Markov chain credit rating migration approach, the suggestion to solution must be somewhat easy to implement. Deep analysis of time-inhomogeneity behaviour and modelling of such is a research field in its own right and not within the scope of this thesis. A suggestion to mitigate the impact of probable future time-inhomogeneity is to use different migration matrices for different states of the economy.

The average matrix might be a good estimation for longer term estimations, or estimation in "average" economic climate. The average matrix is based on 12 years of data between 2002 and 2013. During that time the economy has seen both recessions and expansions, low and high volatility climates and therefore has been exposed to a variety of conditions likely to appear in the future as well. However, for time periods shorter than an average business cycle, say 5 years, the future economic climate is likely to have an impact of credit migrations. Getting a proper estimation of cumulative PD curves is of course essential from a risk perspective, but also from an economic perspective in terms of holding the right amount of capital.

For recessions and expansions, however, the suggestion is to use matrices estimated only using data from recession and expansion periods. Together with rules on when to use them, they could probably mitigate the impact of time-inhomogeneity. This is further elaborated on in section 6.3: Suggestion for further studies. To choose what data to use, focus is put on the main goal of rating migration estimations, namely the cumulative PD curves. A PD curve for a certain rating will be calculated for each annual matrix estimated using data from 2002 to 2013, i.e. 12 PD curves in total. Note that every element of the matrix is used to calculate the PD curves, which is another reason for looking at them rather than e.g. just the PD column of 1-year transition matrices. Therefore, the PD curves will be used to decide what data to use. The decision regarding which rating to look at fell on rating B. One reason for using that rating is because observations to the defaulted state has been seen more frequently there than in higher ratings. Thus, estimations of PDs are more certain. Furthermore, companies that are not very well off rating wise might be more sensitive to fluctuations in the economic climate and therefore make a clearer distinction between the PD curves. Also, the number of observed firm years are larger for rating B than for e.g. rating CCC. A larger number of firm years is of course better to make sure that the accuracy of the estimations are better.

To conclude, recession and expansion matrices will be calculated to possibly mitigate the impact of future time-inhomogeneity. The PD curves for rating B, calculated from each annual matrix will be used to decide what data should be used to estimate these matrices.

Chapter 5

Results

In this chapter, the results from the estimations are presented, with some details deferred to Appendix A: Confidence intervals, and Appendix B: Wald and bootstrapped CI comparison. Migration matrices have been coloured to make it easier for the reader to see the matrix structure. All migration matrices are 1-year migration matrices. When reading migration matrices, ratings in the rows to the left are the previous ratings, and the columns are the ratings that counterparties migrate to. Thus, when looking at figure 5.1, the number 10.69% with row rating A, and column rating BBB means that a company with a rating A has a 10.69% probability of being in rating BBB in one year's time.

The chapter is divided into a first part showing the migration matrices for the estimation and validation sets. Then, average matrices for the whole data set are calculated and compared. Moreover, PDs and their confidence intervals are tabulated. Furthermore, cumulative PD curves are then calculated for the average matrices, where some results associated to them are shown in tables. The third part of the results chapter shows the study on the infinitesimal approximation, which is a way of calculating migration matrices from continuous data without using the infinite series expansion.

The fourth part is the time-inhomogeneity investigation, which has several subsections including measurements and statistical tests. Then, the same investigation on the defensive companies subset follows. Finally, a study of annual PD curves for rating B is presented, and recession and expansion migration matrices are calculated.

5.1 Performance on validation set

This section contains the matrices for the estimation and validation data sets, where data on their bootstrapped confidence intervals is deferred to Appendix A. The purpose is to show whether the methodology of estimating migration matrices to estimate rating movements of unknown companies seems reasonable. The economic conditions and the rating models will be reasonably similar between the two sets, but they contain unique sets of companies.

Consider figures 5.1 and 5.2. In the terminology of Markov chains, both matrices have the default state D as *absorbing*. Apart from the defaulted state, all the other states except AAA to CCC in the validation set matrix are *communicating*. In the validation data set, there are no observations of a AAA rated company being downgraded directly to CCC, but CCC is *accessible* from AAA.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.32%	1.53%	0.13%	0.35%	0.54%	0.11%	0.01%	0.01%
AA	3.07%	82.81%	12.32%	0.91%	0.84%	0.01%	0.02%	0.02%
A	0.15%	3.47%	83.64%	10.69%	1.82%	0.15%	0.02%	0.04%
BBB	0.08%	0.37%	7.55%	80.36%	10.33%	0.90%	0.11%	0.30%
BB	0.16%	0.14%	1.10%	13.42%	79.07%	4.49%	0.69%	0.94%
B	0.14%	0.05%	0.36%	3.02%	15.63%	73.55%	3.25%	3.99%
CCC	0.03%	0.09%	0.27%	1.34%	5.37%	11.26%	73.80%	7.84%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.1: Migration matrix where all data in the *estimation* data set has been used. Calculated with the cohort method.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.35%	1.47%	0.19%	0.30%	0.58%	0.05%	0.03%	0.02%
AA	2.98%	82.56%	12.59%	0.91%	0.91%	0.01%	0.00%	0.04%
A	0.15%	3.37%	83.95%	10.45%	1.85%	0.17%	0.01%	0.05%
BBB	0.08%	0.36%	7.36%	80.66%	10.17%	0.97%	0.09%	0.30%
BB	0.18%	0.11%	1.08%	13.38%	79.13%	4.42%	0.69%	1.01%
B	0.11%	0.05%	0.33%	3.11%	15.78%	73.27%	3.46%	3.88%
CCC	0.04%	0.06%	0.34%	1.22%	5.31%	11.87%	73.81%	7.35%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.2: Migration matrix where all data in the *validation* data set has been used. Calculated with the cohort method.

This is one of the conditions that cause the embedding problem, which we see present here in the cohort estimation. The probability is zero even though the validation set contains about several hundred thousand observations spread over 12 years. In the migration matrix where the estimation set has been used, all states (except D) communicate with each other. The Markov chain is not *irreducible* because of the absorbing default state. All states except D are *transient* since there is a non-zero probability that they never return to that state, because of the non-zero probability that they eventually end up in the absorbing state.

As can be seen from the estimations, the matrices are very similar, which is exactly what we were hoping for. This means that on average, the estimations depend on rating rather than company specific features, which is the aim of this Markov chain estimation approach. This test was done first, to motivate that we can now allow the estimation and validation sets to be joined. When doing all other estimations we of course want to have as many observations as possible at our disposal and use all our data.

To further motivate that the validation and estimation data sets give (statistically) the same estimations, a calculation of their 95% confidence intervals was made. This was done via a bootstrapping procedure, as described in section 4.7.2. The tables showing the intervals are found in Appendix A, and it can be seen that the confidence intervals for every every single distinct p_{ij} overlap. Thus, we cannot statistically make a difference between those estimates.

The estimation and validation data sets were also used to calculate migration matrices with the duration approach.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.02%	1.40%	0.27%	0.49%	0.65%	0.14%	0.02%	0.02%
AA	2.49%	81.24%	13.31%	1.63%	1.18%	0.08%	0.03%	0.04%
A	0.18%	3.41%	83.37%	10.28%	2.33%	0.29%	0.05%	0.09%
BBB	0.10%	0.56%	7.12%	80.88%	9.76%	1.11%	0.16%	0.31%
BB	0.16%	0.31%	1.67%	12.31%	79.22%	4.62%	0.71%	1.01%
B	0.14%	0.13%	0.63%	3.92%	14.54%	71.30%	4.31%	5.03%
CCC	0.06%	0.19%	0.43%	1.91%	5.95%	10.22%	70.54%	10.71%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.3: Migration matrix where all data in the *estimation* data set has been used. Calculated with the duration method.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.02%	1.40%	0.35%	0.41%	0.61%	0.16%	0.03%	0.03%
AA	2.45%	81.13%	13.51%	1.57%	1.20%	0.08%	0.01%	0.04%
A	0.17%	3.35%	83.48%	10.28%	2.26%	0.30%	0.05%	0.11%
BBB	0.10%	0.55%	6.96%	81.17%	9.62%	1.14%	0.15%	0.31%
BB	0.16%	0.28%	1.61%	12.26%	79.30%	4.62%	0.71%	1.06%
B	0.12%	0.12%	0.65%	3.94%	14.57%	71.19%	4.30%	5.11%
CCC	0.06%	0.17%	0.49%	1.90%	5.77%	10.62%	70.37%	10.62%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.4: Migration matrix where all data in the *validation* data set has been used. Calculated with the duration method.

Again, both matrices look similar. A closer look at the bootstrapped 95% confidence intervals found in Appendix A reveals that all the estimated entries are statistically indistinguishable between both matrices.

Furthermore, we can see that all states (apart from the *absorbing* defaulted state) are *communicating* and are *transient*. Moreover, the condition which creates the embedding problem is not present when using the duration method. Henceforth, description of migration matrices using the terminology of Markov chains will be left to the reader.

Since the matrices are similar for both cohort and duration methods, the data set is now joined to one for further use. This data set is the one referred to as the "final data set" in Chapter 3.

5.2 Average matrices

The average matrices presented below are the ones where all data has been used, for both cohort and duration method.

The two methods give somewhat different estimations, as can be seen in figures 5.5 and 5.6, which shows the migration matrices. To get a better feeling of how the estimations differ, a matrix where the cohort estimations are divided by the duration estimations was calculated, see

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.34%	1.50%	0.16%	0.32%	0.56%	0.08%	0.02%	0.02%
AA	3.02%	82.69%	12.45%	0.91%	0.87%	0.01%	0.01%	0.03%
A	0.15%	3.42%	83.80%	10.57%	1.84%	0.16%	0.02%	0.04%
BBB	0.08%	0.36%	7.46%	80.51%	10.25%	0.94%	0.10%	0.30%
BB	0.17%	0.12%	1.09%	13.40%	79.10%	4.45%	0.69%	0.98%
B	0.13%	0.05%	0.35%	3.07%	15.71%	73.41%	3.36%	3.93%
CCC	0.04%	0.08%	0.30%	1.28%	5.34%	11.57%	73.80%	7.60%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.5: Migration matrix where all data in the has been used, calculated with the cohort method.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.02%	1.40%	0.31%	0.45%	0.63%	0.15%	0.02%	0.03%
AA	2.47%	81.19%	13.41%	1.60%	1.19%	0.08%	0.02%	0.04%
A	0.18%	3.38%	83.43%	10.28%	2.29%	0.29%	0.05%	0.10%
BBB	0.10%	0.56%	7.04%	81.03%	9.69%	1.13%	0.15%	0.31%
BB	0.16%	0.30%	1.64%	12.29%	79.26%	4.62%	0.71%	1.03%
B	0.13%	0.13%	0.64%	3.93%	14.56%	71.25%	4.30%	5.07%
CCC	0.06%	0.18%	0.46%	1.90%	5.86%	10.42%	70.46%	10.66%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.6: Migration matrix where all data in the has been used, calculated with the duration method.

figure 5.7. Figure 5.7 have some numbers in bold font, which indicates where the cohort method and duration method have given statistically different estimations. By statistically different, we mean that the 95% bootstrapped confidence intervals of the cohort and duration estimates do not overlap. For reference, the confidence intervals can be found in Appendix A.

The matrix created by dividing the cohort and duration migration matrices show that 34 out of 56, or roughly 61% of the entries, are statistically different estimations. One can also see that the cohort method estimates have a heavier diagonal and much lighter tails.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	100%	107%	**52%**	72%	90%	54%	86%	58%
AA	**122%**	**102%**	**93%**	**57%**	**73%**	**14%**	42%	77%
A	84%	101%	100%	103%	**80%**	**56%**	**38%**	**43%**
BBB	84%	**65%**	**106%**	**99%**	**106%**	**83%**	**68%**	97%
BB	106%	**42%**	**66%**	**109%**	100%	**96%**	97%	95%
B	101%	**41%**	**54%**	**78%**	**108%**	**103%**	**78%**	**78%**
CCC	62%	**43%**	**65%**	**67%**	91%	**111%**	**105%**	**71%**

Figure 5.7: Matrix where cohort average estimation is divided by the duration average estimation. Numbers in bold describe where the cohort and duration methods have given a statistically different estimation. The default row has been left out since they are identical by construction.

5.2.1 Probability of default

This subsection focuses on what is the main goal of the estimations - to be able to calculate probability of default and cumulative probability of default curves. The estimated PDs can be found in the last column of the migration matrices. In tables 5.1 and 5.2, PD estimates and their calculated confidence intervals are presented. The table containing estimates by the cohort method have both Wald and bootstrapped confidence intervals. In table 5.1, it can be seen that the Wald CIs and bootstrapped CIs look very similar. For the sake of comparison and to make it easier for the reader, only bootstrapped CIs will henceforth be tabulated for the cohort and duration methods. A closer study on exactly how similar Wald CIs and their bootstrapped counterparts are can be found in Appendix B.

Table 5.1: Table over PD estimates from the cohort method, including estimated 95% confidence intervals and interval lengths.

	Cohort				
Rating	PD estimate	CI_{Wald}	CI_{Boot}	Length CI_{Wald}	Length CI_{Boot}
AAA	0.02%	[0.00%,0.03%]	[0.00%,0.04%]	0.03%	0.04%
AA	0.03%	[0.01%,0.05%]	[0.01%,0.05%]	0.03%	0.04%
A	0.04%	[0.03%,0.05%]	[0.03%,0.05%]	0.02%	0.02%
BBB	0.30%	[0.28%,0.32%]	[0.28%,0.32%]	0.04%	0.04%
BB	0.98%	[0.94%,1.02%]	[0.94%,1.02%]	0.08%	0.08%
B	3.93%	[3.79%,4.08%]	[3.79%,4.08%]	0.29%	0.29%
CCC	7.60%	[7.22%,7.97%]	[7.22%,7.94%]	0.74%	0.72%

Table 5.2: Table over PD estimates from the duration method, including estimated 95% confidence intervals and interval lengths.

	Duration		
Rating	PD estimate	CI_{Boot}	Length CI_{Boot}
AAA	0.03%	[0.01%,0.04%]	0.03%
AA	0.04%	[0.03%,0.05%]	0.02%
A	0.10%	[0.09%,0.11%]	0.02%
BBB	0.31%	[0.30%,0.32%]	0.03%
BB	1.03%	[1.00%,1.06%]	0.06%
B	5.07%	[4.95%,5.21%]	0.26%
CCC	10.66%	[10.28%,11.08%]	0.80%

By multiplying the cohort migration matrix with itself and each time extracting the PD-column, cumulative PDs for each year are obtained. These are plotted up to year 10 in figure 5.8. For the duration method, cumulative PDs are extracted from the migration matrix calculated via the formula $M(t) = e^{tG}$. The continuous cumulative PDs for up to year 10 are then plotted in figure 5.9.

As indicated by the migration matrices and PD tables, the cumulative PDs are higher for the duration method than for the cohort method, especially for the lowest rating B and CCC. The estimated PDs for all years using the average matrices, and also the ratio duration/cohort PD estimation can be found in table 5.3, table 5.4 and table 5.5.

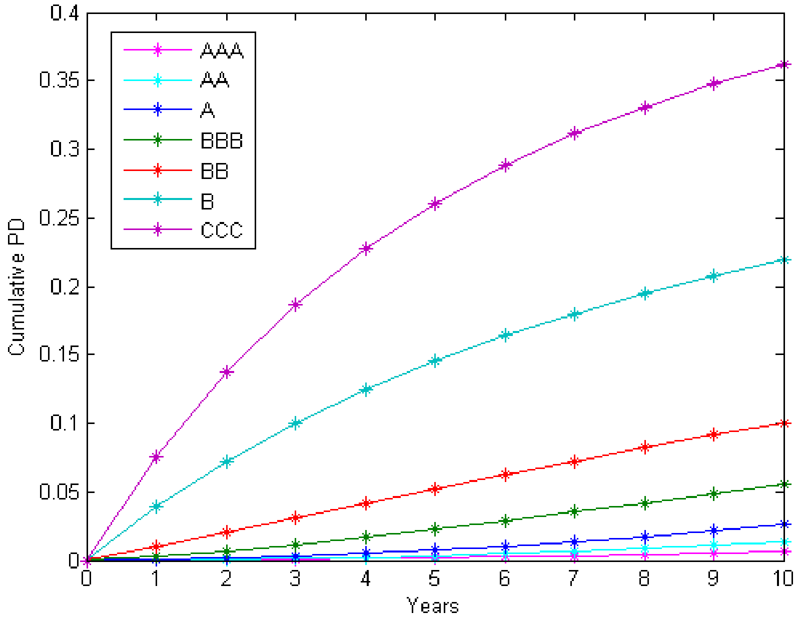

Figure 5.8: Graph showing cumulative PDs during a 10-year period for the cohort method. Based on the average cohort migration matrix.

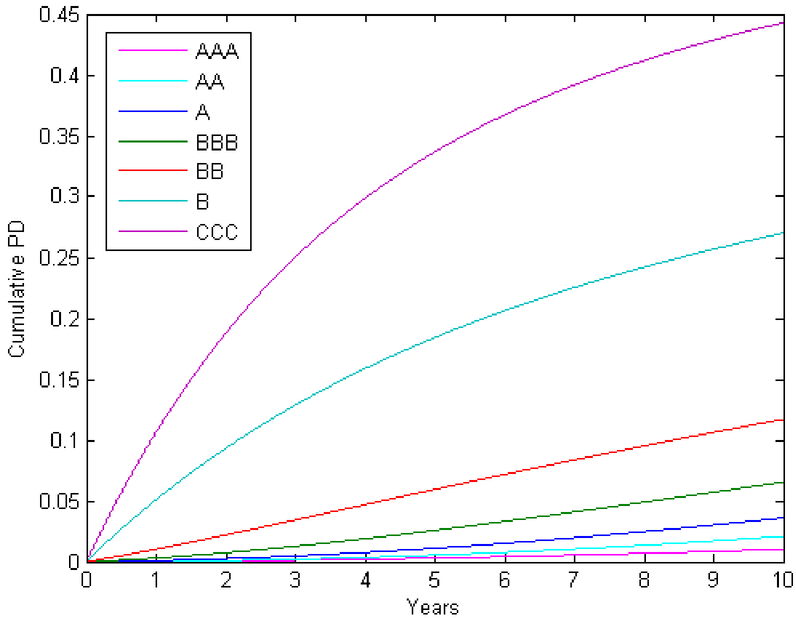

Figure 5.9: Graph showing cumulative PDs during a 10-year period for the duration method. Based on the average duration migration matrix.

Table 5.3: Table showing estimated cumulative PDs for 10 years into the future using the average matrix obtained by the cohort method.

Year	1	2	3	4	5	6	7	8	9	10
AAA	0.02%	0.04%	0.08%	0.13%	0.19%	0.26%	0.34%	0.43%	0.53%	0.64%
AA	0.03%	0.07%	0.14%	0.22%	0.33%	0.47%	0.64%	0.84%	1.07%	1.33%
A	0.04%	0.14%	0.28%	0.48%	0.73%	1.03%	1.37%	1.75%	2.17%	2.62%
BBB	0.30%	0.69%	1.15%	1.68%	2.25%	2.87%	3.51%	4.17%	4.85%	5.54%
BB	0.98%	2.02%	3.08%	4.15%	5.20%	6.23%	7.23%	8.20%	9.13%	10.04%
B	3.93%	7.24%	10.05%	12.46%	14.55%	16.38%	18.00%	19.44%	20.74%	21.92%
CCC	7.60%	13.71%	18.67%	22.72%	26.05%	28.81%	31.13%	33.09%	34.76%	36.20%

Table 5.4: Table showing estimated cumulative PDs for 10 years into the future using the average matrix obtained by the duration method.

Year	1	2	3	4	5	6	7	8	9	10
AAA	0.03%	0.07%	0.13%	0.21%	0.31%	0.42%	0.55%	0.69%	0.85%	1.02%
AA	0.04%	0.11%	0.21%	0.36%	0.54%	0.77%	1.04%	1.36%	1.71%	2.10%
A	0.10%	0.26%	0.48%	0.76%	1.11%	1.52%	1.97%	2.48%	3.02%	3.61%
BBB	0.31%	0.74%	1.28%	1.90%	2.58%	3.32%	4.10%	4.90%	5.72%	6.55%
BB	1.03%	2.20%	3.44%	4.69%	5.94%	7.17%	8.36%	9.52%	10.63%	11.71%
B	5.07%	9.31%	12.86%	15.86%	18.42%	20.62%	22.53%	24.20%	25.68%	27.01%
CCC	10.66%	18.77%	25.00%	29.85%	33.66%	36.70%	39.16%	41.18%	42.87%	44.29%

Table 5.5: Table showing the ratio between estimated cumulative duration PDs and estimated cumulative cohort PDs.

Year	1	2	3	4	5	6	7	8	9	10
AAA	172%	169%	166%	165%	164%	163%	162%	161%	161%	160%
AA	131%	149%	158%	163%	164%	164%	163%	161%	160%	158%
A	234%	189%	170%	159%	152%	147%	144%	141%	139%	138%
BBB	104%	108%	111%	113%	115%	116%	117%	117%	118%	118%
BB	106%	109%	111%	113%	114%	115%	116%	116%	116%	117%
B	129%	129%	128%	127%	127%	126%	125%	124%	124%	123%
CCC	140%	137%	134%	131%	129%	127%	126%	124%	123%	122%

5.3 Infinitesimal approximation comparison

In this section, the results of using an infinitesimal approximation of the continuous-time Markov chain will be compared to the results when the exact definition is used. The difference lies within the calculation of the migration matrix M.

The exact definition uses $M(t) = e^{tG}$ for each t, where e^{tG} is a matrix exponential defined by the power series $e^{tG} \equiv \sum_{k=0}^{\infty} \frac{(tG)^k}{k!}$.

On the other hand, the infinitesimal definition of a CMTC, which is an approximation, defines the migration probability between rating i and j, p_{ij}, for a short time step h as $p_{ij} = \delta_{ij} + g_{ij}h$.

Below in figure 5.10 and figure 5.11 are graphs of cumulative PD-curves calculated with the exact and the approximative infinitesimal definition. Note that the approximations are giving linear cumulative PD curves. The ratings are split to two graphs due to the different magnitude of their PDs. The time step was chosen to be 1 day, which essentially is the smallest measurement unit in the data sample. The G matrix and it's entries is estimated using the full sample, i.e. the G matrix is the one corresponding to the average M duration matrix.

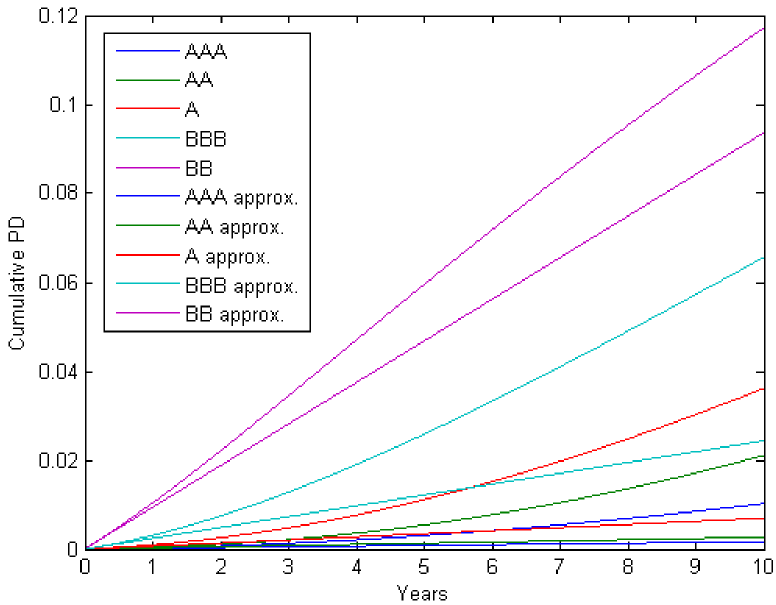

Figure 5.10: Graph showing cumulative PDs for ratings AAA through BB during a 10-year period for the duration method. Calculations of PDs are done via both the exact method and the approximative infinitesimal definition of a CMTC.

As seen in both the formulas in this section and the figures 5.10 and 5.11, the infinitesimal approximation is a linear approximation. Cumulative PD-curves calculated exactly have concave shaped curves for the lowest ratings and convex shaped curves for the highest ratings. In between lies fairly linear shaped PD-curves. This means that over time, the ratings that are fairly linear (such as rating BB here) will be relatively better estimated by the approximation method.

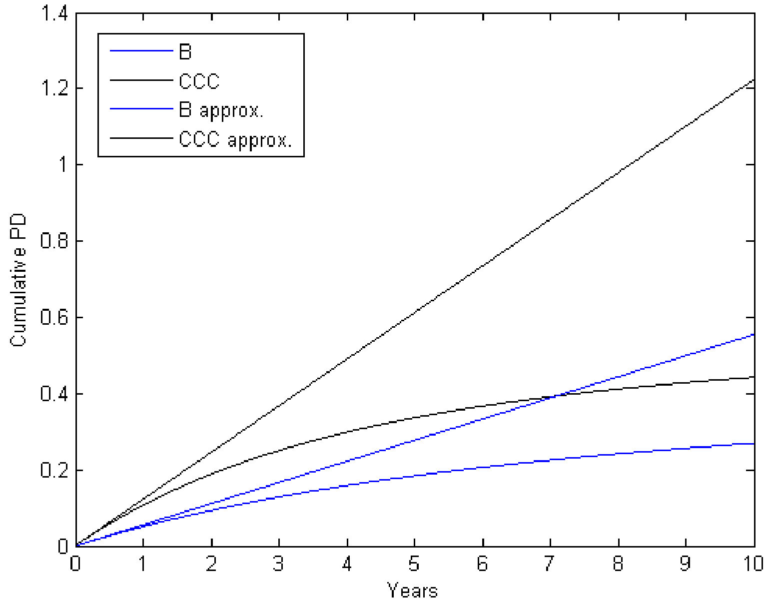

Figure 5.11: Graph showing cumulative PDs for ratings B and CCC during a 10-year period for the duration method. Calculations of PDs are done via both the exact method and the approximative infinitesimal definition of a CMTC.

However, we can still see that already after 5 years the approximation PD curves that are relatively closest to their corresponding exact PD-curves (e.g. rating BB) have a relative difference of about 15 to 20%. After about 8 years, the CCC linear approximation hits the probability roof, $PD = 1$, of defaulting. This is of course not realistic and highlights a flaw of using the linear approximation.

5.4 Investigating time-inhomogeneity

In the following sections, results from the time-inhomogeneity investigation will be presented. This includes measures for detecting inhomogeneities and testing the time-homogeneity assumption. The methods to detect and test will be used on the full data sample, but later also on a subsample of companies from defensive sectors.

5.4.1 Detecting time-inhomogeneity

The measure used to detect time-inhomogeneity in this report is the L_2 norm, for reasons stated in section 4.7.1. The L_2 norms between the average matrix and annual matrices for years 2002-2013 for both the cohort and duration method are plotted in figure 5.12. Note that for the cohort method, the number of firm years during 2002 and 2013 are substantially lower than the other years, see table 3.1 and table 3.2.

Figure 5.12 shows that some years are less similar to the average matrix than the other years. Possibly 2005, but especially 2009 is a clear example of an annual matrix more different that

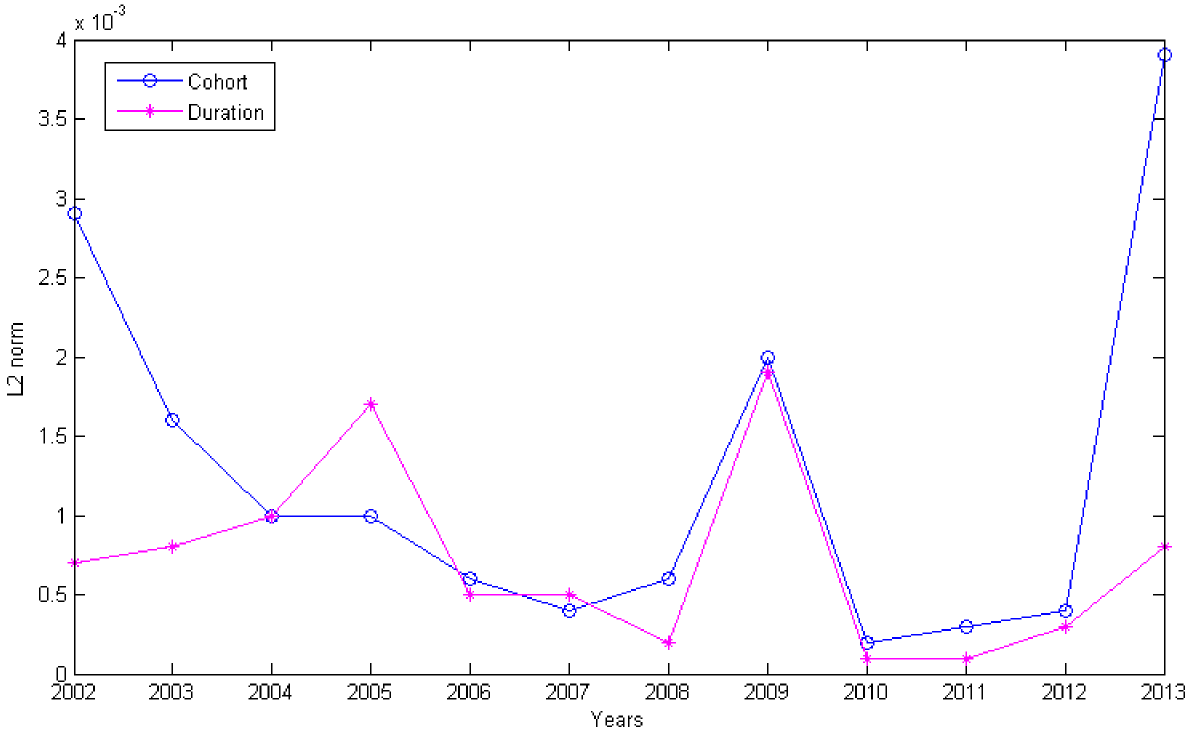

Figure 5.12: Graph showing the L_2 norm between annual matrices and the average matrix, for both the cohort and duration method.

others for both methods. Year 2002 and 2013 is very different when the cohort method is used, probably due to the low amount of observations during these years.

5.4.2 Testing for time-homogeneity

This section contains results related to testing of time-homogeneity. The first test is a χ^2 test as designed by Anderson and Goodman (1957). The second approach to test for time-homogeneity is to examine the confidence intervals of the estimations calculated from two adjacent years' data.

χ^2 test

In the methodology chapter, section 4.7.2, the theory of the χ^2 test for time-homogeneity is presented. If the calculated χ^2 value exceeds the tabulated value with $7 \cdot 8 \cdot 11 = 616$ degrees of freedom, then we can reject the null hypothesis that transition probabilities are constant over time.

The tabulated value of the χ^2 distribution at the 99% level with 616 degrees of freedom is $\chi^2_{0.99}(616) \approx 701$, obtained by the Matlab function `chi2inv(0.99,616)`. Thus, if the values calculated by formula (4.3) exceeds 701, then we can reject the null hypothesis.

Table 5.6: Table of calculated (observed) χ^2 values when comparing annual matrices to the average matrix. Values are calculated based on matrices estimated with both the cohort and the duration method using the full data sample.

Calculation method	Cohort	Duration
Observed χ^2 value	25666	25981

As shown in table 5.6, the observed values are much larger than the tabulated $\chi^2_{0.99}(616)$ value. Thus we can reject the null hypothesis that migration matrices are constant at the 99% level. We also note that the test statistic number calculated with duration estimates are somewhat larger. Further elaboration on the results will be found in the discussion chapter in section 6.1.3.

Confidence intervals

In this section, bootstrapped 95% confidence intervals for PD estimates during the years 2009, 2010, 2011 and 2012 will be presented. Year 2013 is not included due to the small sample size for the cohort method. Adjacent years are compared to minimise impact of rating models.

Below are tables containing the bootstrapped confidence intervals and a "Yes" or "No", indicating if they are statistically different or not.

Table 5.7: Table over bootstrapped 95% PD confidence intervals for years 2009 and 2010. Calculations are based on the full sample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2009	Cohort 2010	Different	Duration 2009	Duration 2010	Different
AAA	[0.00%, 0.00%]	[0.00%, 0.09%]	No	[0.01%, 0.02%]	[0.01%, 0.13%]	No
AA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.07%, 0.11%]	[0.02%, 0.03%]	Yes
A	[0.06%, 0.15%]	[0.01%, 0.05%]	Yes	[0.22%, 0.31%]	[0.05%, 0.10%]	Yes
BBB	[0.61%, 0.80%]	[0.18%, 0.30%]	Yes	[0.95%, 1.12%]	[0.28%, 0.37%]	Yes
BB	[1.60%, 1.92%]	[0.83%, 1.06%]	Yes	[2.10%, 2.40%]	[1.06%, 1.26%]	Yes
B	[4.83%, 6.00%]	[4.55%, 5.62%]	No	[7.51%, 8.68%]	[5.98%, 7.02%]	Yes
CCC	[6.37%, 8.87%]	[4.72%, 6.63%]	No	[12.00%, 14.70%]	[8.88%, 10.92%]	Yes

Table 5.8: Table over bootstrapped 95% PD confidence intervals for years 2010 and 2011. Calculations are based on the full sample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2010	Cohort 2011	Different	Duration 2010	Duration 2011	Different
AAA	[0.00%, 0.09%]	[0.00%, 0.00%]	No	[0.01%, 0.13%]	[0.00%, 0.00%]	Yes
AA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.02%, 0.03%]	[0.01%, 0.01%]	Yes
A	[0.01%, 0.05%]	[0.01%, 0.06%]	No	[0.05%, 0.10%]	[0.04%, 0.08%]	No
BBB	[0.18%, 0.30%]	[0.13%, 0.21%]	No	[0.28%, 0.37%]	[0.19%, 0.27%]	Yes
BB	[0.83%, 1.06%]	[0.75%, 0.96%]	No	[1.06%, 1.26%]	[0.80%, 0.97%]	Yes
B	[4.55%, 5.62%]	[3.75%, 4.73%]	No	[5.98%, 7.02%]	[5.37%, 6.31%]	No
CCC	[4.72%, 6.63%]	[6.80%, 8.90%]	Yes	[8.88%, 10.92%]	[9.08%, 11.24%]	No

Table 5.9: Table over bootstrapped 95% PD confidence intervals for years 2011 and 2012. Calculations are based on the full sample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2011	Cohort 2012	Different	Duration 2011	Duration 2012	Different
AAA	[0.00%, 0.00%]	[0.00%, 0.07%]	No	[0.00%, 0.00%]	[0.01%, 0.12%]	Yes
AA	[0.00%, 0.00%]	[0.00%, 0.05%]	No	[0.01%, 0.01%]	[0.01%, 0.02%]	Yes
A	[0.01%, 0.06%]	[0.01%, 0.05%]	No	[0.04%, 0.08%]	[0.08%, 0.14%]	Yes
BBB	[0.13%, 0.21%]	[0.51%, 0.66%]	Yes	[0.19%, 0.27%]	[0.27%, 0.36%]	Yes
BB	[0.75%, 0.96%]	[1.12%, 1.38%]	Yes	[0.80%, 0.97%]	[1.10%, 1.30%]	Yes
B	[3.75%, 4.73%]	[5.36%, 6.46%]	Yes	[5.37%, 6.31%]	[9.52%, 10.88%]	Yes
CCC	[6.80%, 8.90%]	[6.61%, 8.90%]	No	[9.08%, 11.24%]	[13.12%, 15.85%]	Yes

As apparent by tables 5.7, 5.8 and 5.9, significantly different estimates can be found with both methods between all years. Furthermore, the duration method has easier to pick up significant differences, especially for the highest ratings.

5.5 Defensive companies investigation

In this section, results from the time-inhomogeneity investigation of the subset containing defensive companies will be presented. The results will be compared to those of the whole sample, to see if there is the time-homogeneity assumption is better for the defensive subset.

5.5.1 Average defensive matrices

By using the whole subset spanning 2002 to 2013, the following average matrices was calculated using the cohort and duration methods, respectively. The are similar to the average matrices of the full sample, as is to be expected since they measure migrations over a whole business cycle.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.49%	1.45%	0.12%	0.32%	0.51%	0.09%	0.02%	0.01%
AA	3.13%	82.00%	12.89%	0.99%	0.96%	0.00%	0.00%	0.03%
A	0.15%	3.35%	83.93%	10.51%	1.83%	0.17%	0.02%	0.04%
BBB	0.09%	0.38%	7.51%	80.39%	10.30%	0.93%	0.09%	0.30%
BB	0.17%	0.13%	1.10%	13.53%	78.98%	4.40%	0.70%	0.98%
B	0.12%	0.06%	0.37%	3.11%	15.89%	73.24%	3.19%	4.02%
CCC	0.02%	0.03%	0.36%	1.42%	5.35%	11.49%	73.89%	7.45%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.13: Average migration matrix calculated with the cohort method, based on the defensive companies subset.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.22%	1.37%	0.28%	0.39%	0.57%	0.13%	0.02%	0.03%
AA	2.50%	80.74%	13.75%	1.61%	1.27%	0.08%	0.02%	0.03%
A	0.18%	3.28%	83.61%	10.25%	2.25%	0.30%	0.05%	0.09%
BBB	0.10%	0.55%	7.11%	80.94%	9.71%	1.12%	0.16%	0.31%
BB	0.17%	0.32%	1.68%	12.28%	79.23%	4.58%	0.71%	1.03%
B	0.13%	0.13%	0.67%	3.93%	14.53%	71.02%	4.42%	5.16%
CCC	0.07%	0.17%	0.48%	1.94%	5.97%	10.30%	70.48%	10.60%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.14: Average migration matrix calculated with the duration method, based on the defensive companies subset.

5.5.2 Detecting time-inhomogeneity

Analogously with section 5.4.1 for the whole sample, the L_2 norm between the average matrix and the annual matrices will be calculated based on observations from the defensive companies subsample. These are plotted for the cohort and duration method, together with the corresponding results from section 5.4.1 where the whole sample was used.

From figure 5.15, it can be seen that the L_2 norms calculated with the defensive subsample is practically similar to those calculated with the full sample.

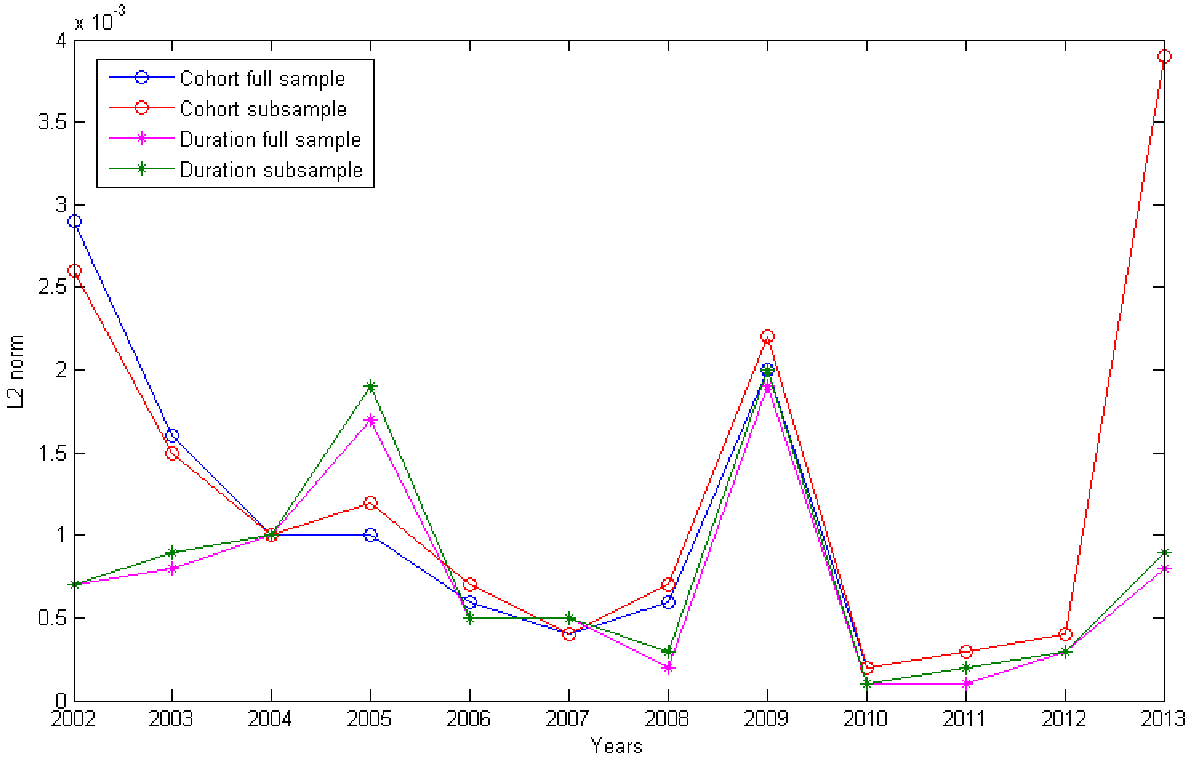

Figure 5.15: Graph showing the L_2 norm between annual matrices and the average matrix, for both the cohort and duration method. The lines labelled subsample is the numbers based on the defensive subsample.

5.5.3 Testing for time-homogeneity

χ^2 test

The χ^2 test for the defensive companies subsample is performed exactly as the one where the full sample was used. The test statistic is calculated using formula (4.3) to see whether we can reject that annual matrices and the average matrices are the same, i.e. to reject time-homogeneity. To reject the null hypothesis that transition probabilities are the same on the 99% level, the calculated χ^2 value has to exceed the tabulated value $\chi^2_{0.99}(616) \approx 701$.

Table 5.10: Table of calculated (observed) χ^2 values when comparing annual matrices to the average matrix. Values are calculated based on matrices estimated with both the cohort and the duration method using the defensive companies subsample.

Calculation method	Cohort	Duration
Observed χ^2 value	15288	17246

From the table 5.10, it is evident that we can reject the null hypothesis using both the cohort and duration method. Yet again the cohort method gives a somewhat lower χ^2 value compared to the duration method.

Confidence intervals

Comparison of confidence intervals for the defensive companies subsample is performed analogously with the time-inhomogeneity investigation for the full sample. That is, two 95% bootstrapped CIs for a certain PD estimated with two adjacent years data are compared to decide whether they are statistically different or not. If they are different, then it is likely that time-inhomogeneity is present. The results are shown in tables 5.11, 5.12 and 5.13).

Table 5.11: Table over bootstrapped 95% PD confidence intervals for years 2009 and 2010. Calculations are based on the defensive sectors subsample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2009	Cohort 2010	Different	Duration 2009	Duration 2010	Different
AAA	[0.00%, 0.00%]	[0.00%, 0.15%]	No	[0.01%, 0.01%]	[0.00%, 0.13%]	No
AA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.07%, 0.10%]	[0.03%, 0.04%]	Yes
A	[0.03%, 0.14%]	[0.01%, 0.08%]	No	[0.22%, 0.35%]	[0.05%, 0.12%]	Yes
BBB	[0.63%, 0.91%]	[0.17%, 0.31%]	Yes	[0.97%, 1.22%]	[0.27%, 0.39%]	Yes
BB	[1.49%, 1.92%]	[0.80%, 1.11%]	Yes	[2.01%, 2.40%]	[1.04%, 1.31%]	Yes
B	[4.71%, 6.28%]	[4.99%, 6.51%]	No	[7.25%, 8.66%]	[6.02%, 7.39%]	No
CCC	[5.75%, 8.92%]	[4.22%, 6.79%]	No	[10.67%, 14.27%]	[8.87%, 11.66%]	No

From tables 5.11, 5.12 and 5.13 it is clear that time-inhomogeneity is present as several estimations of adjacent years are significantly different. The duration method captures differences more often than the cohort method. Compared to the calculations with the full sample, the duration calculation based on the defensive subset has a little harder to find significant differences.

Table 5.12: Table over bootstrapped 95% PD confidence intervals for years 2010 and 2011. Calculations are based on the defensive sectors subsample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2010	Cohort 2011	Different	Duration 2010	Duration 2011	Different
AAA	[0.00%, 0.15%]	[0.00%, 0.00%]	No	[0.00%, 0.13%]	[0.00%, 0.00%]	Yes
AA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.03%, 0.04%]	[0.01%, 0.01%]	Yes
A	[0.01%, 0.08%]	[0.00%, 0.07%]	No	[0.05%, 0.12%]	[0.03%, 0.09%]	No
BBB	[0.17%, 0.31%]	[0.12%, 0.24%]	No	[0.27%, 0.39%]	[0.19%, 0.30%]	No
BB	[0.80%, 1.11%]	[0.69%, 0.96%]	No	[1.04%, 1.31%]	[0.76%, 0.97%]	Yes
B	[4.99%, 6.51%]	[3.69%, 4.98%]	Yes	[6.02%, 7.39%]	[5.56%, 6.86%]	No
CCC	[4.22%, 6.79%]	[6.91%, 9.85%]	Yes	[8.87%, 11.66%]	[9.32%, 12.13%]	No

Table 5.13: Table over bootstrapped 95% PD confidence intervals for years 2011 and 2012. Calculations are based on the defensive sectors subsample, and are tabulated for both cohort and duration methods.

Rating	Cohort 2011	Cohort 2012	Different	Duration 2011	Duration 2012	Different
AAA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.00%, 0.00%]	[0.00%, 0.12%]	Yes
AA	[0.00%, 0.00%]	[0.00%, 0.00%]	No	[0.01%, 0.01%]	[0.01%, 0.02%]	Yes
A	[0.00%, 0.07%]	[0.01%, 0.06%]	No	[0.03%, 0.09%]	[0.05%, 0.11%]	No
BBB	[0.12%, 0.24%]	[0.46%, 0.66%]	Yes	[0.19%, 0.30%]	[0.27%, 0.38%]	Yes
BB	[0.69%, 0.96%]	[1.12%, 1.46%]	Yes	[0.76%, 0.97%]	[1.14%, 1.41%]	Yes
B	[3.69%, 4.98%]	[5.38%, 6.98%]	Yes	[5.56%, 6.86%]	[9.75%, 11.65%]	Yes
CCC	[6.91%, 9.85%]	[6.02%, 8.82%]	No	[9.32%, 12.13%]	[13.34%, 16.87%]	Yes

5.6 Expansion and recession matrices

In this section, estimates of PD curves for different annual matrices will be presented to determine and calculate migration matrices to be used in expansion and recession economic climates.

As discussed in section 4.7.4, cumulative PD curves for rating B will be used to decide what data should be the base for estimation of expansion and recession matrices. The resulting PD curves for the cohort and duration methods are found i figures 5.16 and 5.17. As a clarification, PD estimates at year 10 is tabulated in table 5.14.

Looking at the plotted PD curves, we note that the 2013 curve is way lower than any other year's curve in the cohort method plot. However, that year has very few observations so the estimate is not reliable and can be ignored. Also, 2002 has a substantially lower amount of observations than the other years.

From the PD curves, clusters containing high PD and low PD emerge. The cohort method suggests 2002, 2009 and 2012 as recession years and 2004, 2005, 2006, 2007, 2008 as expansion years. The clustering in the duration method plot suggests 2009 and 2012 as recession years and 2005, 2006 and 2007 as expansion years. Note that the duration method register movements more efficiently. Moreover, the years suggested by the duration method were also practically the ones suggested by the cohort method. Therefore, data for 2009 and 2012 will be used to calculate the recession matrices, and the expansion matrices will be calculated using data from 2005, 2006 and 2007.

Table 5.14: Table showing the year 10 rating B cumulative PDs, calculated with different annual matrices using both the cohort and duration method.

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Cohort	35.8%	25.8%	16.3%	16.0%	14.2%	15.5%	15.1%	34.9%	24.6%	21.9%	32.8%	1.9%
Duration	34.0%	25.6%	21.5%	16.6%	15.0%	14.3%	21.8%	47.4%	30.2%	28.6%	45.8%	34.3%

The recession and expansion matrices for the two methods can be found in figures 5.18, 5.19, 5.20 and 5.21. It can be seen from these figures that the PD column differ substantially between recession and expansion, as well as the ratings around the diagonal.

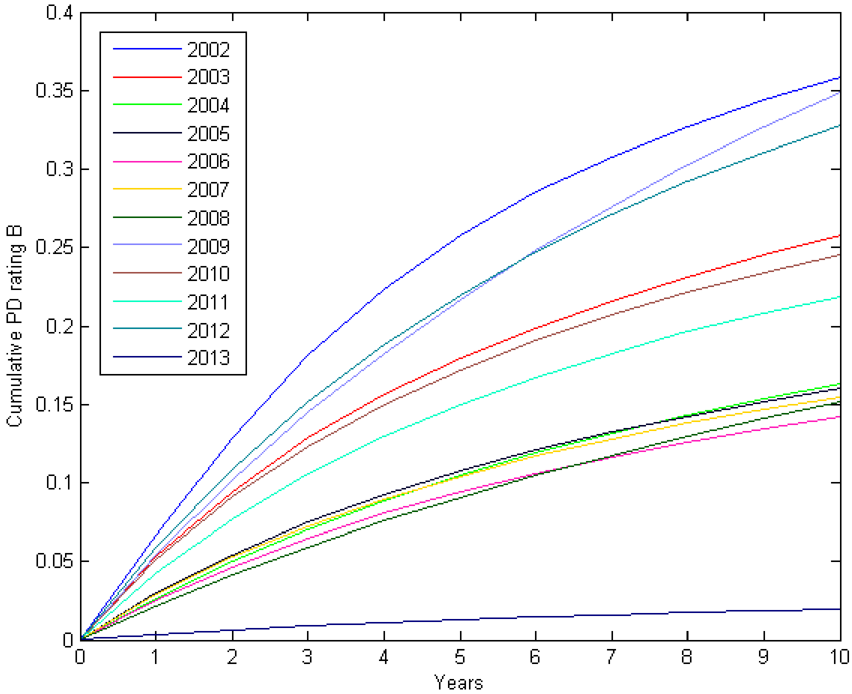

Figure 5.16: Graph showing cumulative PDs during a 10-year period for rating B. The curves are based on the 12 annual migration matrices calculated with the cohort method.

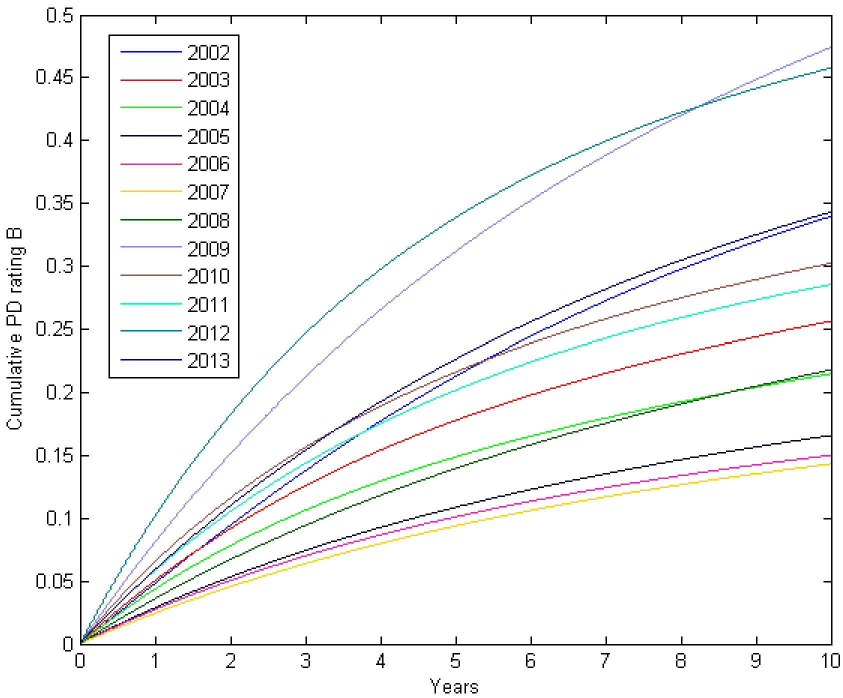

Figure 5.17: Graph showing cumulative PDs during a 10-year period for rating B. The curves are based on the 12 annual migration matrices calculated with the duration method.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	98.28%	1.03%	0.15%	0.14%	0.35%	0.02%	0.03%	0.02%
AA	6.54%	74.55%	17.01%	1.66%	0.22%	0.00%	0.00%	0.01%
A	0.07%	2.73%	83.09%	12.09%	1.83%	0.12%	0.01%	0.06%
BBB	0.14%	0.11%	5.44%	81.53%	11.07%	0.94%	0.13%	0.64%
BB	0.34%	0.06%	0.86%	10.21%	82.18%	4.31%	0.56%	1.48%
B	0.22%	0.07%	0.28%	1.52%	9.95%	78.87%	3.40%	5.68%
CCC	0.10%	0.03%	0.16%	0.37%	3.38%	13.16%	75.12%	7.68%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.18: Recession matrix calculated with the cohort method using data from 2009 and 2012.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	93.39%	3.67%	0.23%	0.87%	1.47%	0.35%	0.00%	0.03%
AA	3.09%	89.56%	6.36%	0.67%	0.29%	0.01%	0.01%	0.00%
A	0.35%	3.77%	81.66%	11.46%	2.41%	0.27%	0.03%	0.05%
BBB	0.13%	0.30%	9.49%	77.47%	11.21%	1.12%	0.12%	0.16%
BB	0.31%	0.15%	1.34%	15.63%	76.21%	5.09%	0.58%	0.69%
B	0.21%	0.07%	0.51%	4.72%	18.72%	69.20%	3.77%	2.80%
CCC	0.04%	0.10%	0.58%	1.60%	6.59%	10.75%	73.67%	6.67%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.19: Expansion matrix calculated with the cohort method using data from 2005, 2006 and 2007.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.78%	0.95%	0.35%	0.33%	0.48%	0.05%	0.02%	0.03%
AA	5.01%	67.93%	22.34%	2.55%	1.97%	0.11%	0.04%	0.05%
A	0.16%	2.45%	80.99%	12.92%	2.89%	0.33%	0.09%	0.17%
BBB	0.16%	0.27%	5.77%	80.18%	11.32%	1.45%	0.21%	0.63%
BB	0.32%	0.45%	1.46%	10.52%	79.98%	4.80%	0.77%	1.71%
B	0.18%	0.11%	0.56%	2.50%	9.89%	72.20%	5.45%	9.10%
CCC	0.12%	0.10%	0.26%	0.91%	4.08%	11.56%	69.06%	13.91%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.20: Recession matrix calculated with the duration method using data from 2009 and 2012.

Rating	AAA	AA	A	BBB	BB	B	CCC	D
AAA	98.08%	0.11%	0.18%	0.81%	0.78%	0.03%	0.00%	0.00%
AA	25.67%	64.03%	8.05%	1.47%	0.73%	0.03%	0.01%	0.01%
A	0.68%	1.91%	86.29%	8.70%	2.09%	0.22%	0.04%	0.08%
BBB	0.24%	0.26%	8.78%	79.93%	9.25%	1.22%	0.16%	0.15%
BB	0.43%	0.06%	1.71%	12.85%	78.63%	5.01%	0.75%	0.56%
B	0.13%	0.02%	0.65%	4.67%	15.26%	72.99%	3.41%	2.87%
CCC	0.06%	0.04%	0.22%	1.90%	5.70%	8.86%	75.48%	7.74%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Figure 5.21: Expansion matrix calculated with the duration method using data from 2005, 2006 and 2007.

Chapter 6

Discussion

In this chapter, the findings and results obtained in chapter 5 is analysed. First, a thorough examination and interpretation of the results is presented. Thereafter, the main conclusions and implications of this thesis is reported. Then, the final part contains some suggestions for further studies that is deemed to be interesting and relevant.

6.1 Interpretation of results

In this section, a discussion of the results is presented to interpret the findings and more clearly point out what was seen in the results chapter.

6.1.1 Cohort and duration methods in general

First of all, we note that for both methods the migration matrices for estimation and validation sets indicated that, on average, the Markov chain method does not depend on company specific properties but rather the internal rating. This is obviously necessary for the method to be used at all. The results motivated a merge of the two data sets into one full data sample. On a side note, also splitting the data into a defensive companies subset gave on average practically the same estimations, which further strengthens the arguments for a merge.

Looking at tables 3.1 and 3.2 in the data description section, we can see that firm years and off-diagonal movements are very different between the methods, a clear indication of the better data efficiency of the duration method. In fact, the number of off-diagonal movements are about 52% higher with the duration method, even after excluding years 2002 and 2013.

An impact of the duration method's higher efficiency is that the estimations between the two methods differ. In many cases, as we saw in figure 5.7, the estimations differ significantly. Furthermore, the duration method gives higher estimations to migrations further off the diagonal. We also note that, as seen from table 5.5, the duration method consistently gives higher cumulative PD values for all ratings. This is not entirely in line with [9] who only found this behaviour at the lowest rating levels. However, since the duration method gives higher estimates far off the diagonal and registers more movement in general than the cohort method, it seems logical that the PDs should be higher.

Moreover, the embedding problem which could be seen in several annual cohort matrices and the cohort validation matrix is not present when using the duration method. The duration method is also capable of estimating confidence intervals for all migration probabilities, since all the entries in the migration matrix are non-zero.

The suggestion is therefore to, if possible, use the duration method. The only drawback is that it is harder to calculate without a software such as Matlab or R. The infinitesimal approximation that could have mitigated this problem was shown to be quite crude. The approximation works best for mid rated companies as their PD curves are more linear. The figures 5.10 and 5.11 show how the approximation diverge from the exact calculation method. Depending on the level of desired accuracy, figures 5.10 and 5.11 can serve as a guideline for how long timespan the approximation method can substitute the exact method. Based on the figures, a general suggestion is to use the approximation method only if the desired time span is no longer than one year.

6.1.2 A note on Wald and bootstrapped CIs

In Appendix B, a study that compares Wald confidence intervals (CIs) to bootstrapped CIs is presented. A note on the results is that for the ratings BBB to CCC, the bootstrapped CIs were close to Gaussian and had a good fit with their analytical Wald counterparts. The results are in line with Hanson and Schuermann (2005) who find the same for the speculative ratings. For the highest ratings, many PD observations will be found the be zero when using the cohort method, i.e. there will be many bootstrapped replications stacking up at zero probability of default. The ordinary Gaussian distribution with a mean very close to zero can of course (depending on the standard deviation) produce negative values. For this reason, a Wald confidence interval must be used with some care for these cases, and one has to take into account that a PD of zero is the floor.

Furthermore, where there are no observations of default in the whole data set, both the Wald CIs and the bootstrap method will give degenerate CIs, i.e. it will just be mean zero and a zero interval length. As seen in Appendix B, table B.3, there are a number of fairly low p-values. This means that there is still a decent chance of having done the right decision if rejecting that the Wald CI distribution and the bootstrap CI distribution are the same. The Wald CI essentially only takes one value is account, namely the average PD of the sample. The remaining results are based on defaults having a binomial distribution and that we have iid observations, which might not be true considering the inhomogeneities found. The bootstrap on the other hand uses the actual empirical sample in every replication. Several previous researchers such as Christensen et al. (2004), Trueck and Rachev (2005) and Hanson and Schuermann (2005) speak in favour of the bootstrap method. The only advantage with the Wald CIs is that it is fast and easy to compute without any computer software such as R or Matlab. Considering the discussion immediately above, using bootstrap is the suggestion we will give in this thesis too. However, this decision should not have to be taken very often, as the duration approach is the preferable way of estimating rating migrations. With the duration approach, there is no analytical alternative, which leaves only the bootstrapping alternative.

To sum up the findings regarding Wald and bootstrapped intervals, Wald CIs and bootstrapped CIs are very similar for the ratings BBB to CCC. This is in line with previous findings from Hanson and Schuermann (2005). However, the Kolmogorov-Smirnov two-sample tests showed that there is a decently large probability that the Wald CI distribution and the bootstrapped CI distribution are not the same. Thus, we prefer to use bootstrapped intervals as this approach uses the whole empirical sample rather than only the average PD and assumptions regarding the distribution, plus it can be used analogously also on investment grade ratings.

6.1.3 Time-inhomogeneity investigation

The presence of time-inhomogeneity has been investigated with three methods, L_2 norm, χ^2 test and comparison of confidence intervals. The methods have been used on both the full data sample and the defensive companies subsample. Immediately below follow a discussion on the results found with these investigation methods.

L_2 norm

The L_2 norm was calculated both for the whole sample and for the defensive companies subsample. It was also calculated using both cohort and duration methods to estimate migration matrices. The norm is a measure of the cell-by-cell difference between the annual matrices and the average matrix. It does not put any weights depending on distance from the diagonal, but rather measuring sheer differences. The formula 4.1 describes how to calculate the L_2 norm. Figure 5.15 show all calculations. It can be noted that the norms were fairly similar between the full sample and the defensive companies subset. Also, years 2005 and 2009 seemed to be less similar to the average matrix than other years. For the cohort method, 2002 and 2013 were very different too, but it is reasonable to believe that this is due to the low amount of data. Moreover, we note that the defensive companies subset is roughly 40% smaller, which probably can cause larger differences. However, the differences in L_2 norms are hard to draw conclusions from other than that there are differences. Therefore, the investigation of time-inhomogeneity is moved on to χ^2 tests and comparison of confidence intervals.

χ^2 test

Below is a discussion of the results from the χ^2 tests found using both the full sample and the defensive companies subsample.

As stated in table 5.6 and 5.10, the observed test statistic was far larger than the tabulated value. Thus, the null hypothesis that annual matrices and the average matrix are the same could be easily rejected (and thereby the homogeneity assumption). One reason for the relatively large observed values is the sheer size of the data set which makes us very "sure" about our estimations. This can be seen from the formula 4.3, where the number of firm years for a certain rating, $n_i(t)$, is multiplied with a ratio $(\hat{p}_{ij}(t)-\hat{p}_{ij})^2/\hat{p}_{ij})$ that quantifies the difference. Thus the annual matrix estimations does not have to differ much from the average matrix for the χ^2 test statistic in formula (4.3) to produce large numbers.

The results are in line with Kiefer and Larsson (2006) who used the same test on S&P corporates and municipal bonds. As an example, they had 20 years of data on S&P corporates. If they made only 4 transitions with a 5-year transition matrix over the 20 years, they could still reject time-homogeneity with a p-value of essentially zero.

In the calculations done in this thesis, a large data set and 12 annual transitions are used. Thus, the high numbers are reasonable. The χ^2 test thereby suggests that time-inhomogeneity is present within the data, regardless of using all data or just the defensive companies subset.

With regards to the difference in observed χ^2 values between the full sample and the defensive subsample, much can be explained by the difference in sample sizes. The observed χ^2 values with the full sample was 53% (cohort) and 73% (duration) larger. However, the full sample is about 73% larger in terms of firm years which scale up the matrix differences linearly. Since the defensive companies sample is smaller, it is likely to have a little bit more sparse matrices. By taking these both factors into account, it seems reasonable that the full sample and the defensive subsample has roughly the same level of time-inhomogeneity present.

Confidence intervals

Tables 5.7 to 5.9 and 5.7 to 5.13 show comparisons of adjacent years' confidence intervals. If the confidence intervals do not overlap, then the estimations can be said to be statistically different. Statistically different estimations indicate that time-inhomogeneity is likely to be present.

The tables show that both cohort and duration methods find statistically different estimations in all adjacent year pairs investigated. This is also true for the defensive companies subset. One finding is also that the duration method is better at picking up differences than the cohort method. This is reasonable since the duration method is more sensitive to rating changes and more efficiently use the data. Another visible pattern is that the cohort method has difficulties picking up differences for the highest ratings AAA to A, whereas the duration method register statistically different estimates in most of those cases.

With regards to comparing the full sample and the defensive sample, we can observe four cases where the duration method can't see a difference in estimations using the defensive subsample that it previously saw using the full sample. These are ratings B and CCC when comparing 2009 to 2010, rating BBB when comparing 2010 to 2011 and finally rating A when comparing 2011 to 2012. The question is now, is it a result of the companies within the defensive subset or is it a consequence of its smaller sample size?

The main point is still that we clearly saw inhomogeneity in both samples, but it is interesting to at least make some approximative calculation of how the answers would change if the subsample had the same number of observations as the full sample. Of course the intervals are dependent on actual observations, but it is also dependent on sample size. Following two suggestions could approximate the answer:

- Assume that, since the actual PD estimates are quite similar, the intervals of the subsample might have been as tight as those in the full sample.
- Use the study of Wald intervals, and assume that the similarities between Wald CI and bootstrap CI that were found in ratings A to CCC still hold for the duration approach. The Wald interval size is heavily dependent on the number of firm years. By calculating the number of firm years for the specific cases, we can change the factor $1/\sqrt{n_R}$ in the Wald CI formula accordingly to account for the smaller sample size.

Both the above approaches gave the same result, namely that we now saw a clear difference on rating B 2009/2010 and a slight difference for rating BBB 2010/2011. The duration method still couldn't see any statistically significant difference regarding rating CCC 2009/2010 and rating A 2011/2012. Thus, it seems as when using the defensive companies subset, the duration method has a little harder to obtain statistically significant differences. In 2 out of 17 cases where saw a difference using the whole sample, we did not see it in the defensive companies subsample. From the CI comparison point of view, this means means a little less time-inhomogeneity. Nevertheless, the defensive companies subsample saw significant differences in PD estimations between all years, using both cohort and duration method. Therefore, we can say that time-inhomogeneity is present in the defensive companies subsample.

To conclude the whole section on time-inhomogeneity investigation, the investigations of L_2 norms, χ^2 values and confidence intervals all point in the same direction. It shows that time-inhomogeneity is present and that even sectors thought of as defensive are exposed to these fluctuations in credit migration probabilities to essentially the same extent as other sectors.

6.1.4 Expansion and recession matrices

To be able to decide which years' data to be used as a base to calculate expansion and recession matrices, a study of cumulative PD curves was used. There are two main reasons why the decision was made based on PD curves. The first is that it is hard to see exactly which annual matrices that describe recessions and expansions. When calculating cumulative PD curves every cell is used. This makes it easier to see the impact on estimated PD from all migration probabilities within the matrix. The second reason why we are looking at cumulative PD curves is because they are the final product that is used to calculate for instance expected losses. In other words, it is the cumulative PD curves that matters for the risk managers. A fairly low rating (B) was chosen to be examined. One reason for choosing rating B was that lower rated companies probably are more sensitive to the economic cycle, which then could distinguish this effect more clearly in the graphs. Furthermore is has more observations in total than e.g. rating CCC, but also a fair amount of observations that defaulted. This make the estimations in this region more certain.

After examining the PD curves, the clustering of cumulative PD after 10 years made the decision which years to include in what matrix. The decision fell on 2009 and 2012 as recession years, and 2005, 2006 and 2007 as expansion years. These years were pointed out by both the cohort and duration method.

The expansion and recession matrices was calculated to be used as tools to mitigate the effect of time-inhomogeneity. It also clearly shows the difference in PD estimations obtained when different years are used as a base for migration matrix calculations. Moreover, it gives a hunch on how big the difference in migrations can be during recessions and expansions. This is important when PD estimates are used to calculate e.g. expected loss on an exposure with shorter time horizon. In times of expansion, less capital could be hold as a buffer for expected losses, while it is the opposite in times of recession. The amount of capital to be held affect the prices of e.g. derivatives transactions, which is why this knowledge is important to be able to quote correct prices.

6.2 Conclusions and implications

This section summarises the main implications and conclusions that could be drawn from the methods, results and discussions in this thesis.

The splitting into an estimation and a validation data set, but also into a defensive companies subsample showed that migration probabilities are on average not dependent on company specific data, but rather on internal rating. This is of course desired for the Markov chain approach, and shows a usefulness in making PD estimations based on internal rating. The result motivated a merge between the estimation and validation subsets into one whole data set.

Data gathered on e.g. firm years and off-diagonal movements clearly shows the data efficiency gains when using the duration method, compared to the cohort method. The suggestion is therefore to use the duration method if possible, which likely requires some software such as R or MATLAB. The investigation of an infinitesimal approximation to avoid using softwares as the above showed that this approach is only feasible for shorter time horizons, depending on desired accuracy. The suggestion is to only use the approximation when considering time frames of at most 1 year into the future.

An effect of the different data efficiency was that migration matrices calculated with the cohort and duration methods gave several statistically different estimations, even though calculations were based on the same time frame. As expected, the duration method migration matrix was heavier further from the diagonal compared to the cohort migration matrix. As a result of larger migration probabilities further off the diagonal, caused by data efficiency gains in registering more migrations, the duration approach gave higher cumulative PD estimations for all ratings.

The comparison of Wald CIs and bootstrapped CIs showed that for ratings BBB to CCC, the bootstrapped CIs fit the analytical Wald CIs very well. This was in line with previous research. However, the fact that bootstrapped intervals uses every empirical observation to a greater extent, plus its usefulness when estimating higher rated companies speaks in favour of using the bootstrap method as a first choice. Moreover, when using the duration approach, no analytical Wald CIs can be computed. Therefore, using bootstrap on both cohort and duration migration matrices is the preferred choice and by using the same method it makes comparisons easier.

All three methods of investigating time-inhomogeneity (L_2 norm, χ^2 test and comparison of CIs) suggest that time-inhomogeneity is present in the data. This applies to the full sample, but also to the defensive companies subsample. The defensive subsample showed on a bit less inhomogeneity when looking at the χ^2 test and the confidence intervals. However, much of that effect can be referred to the fact that the defensive companies subset had a smaller amount of data. Nevertheless, inhomogeneities were present in all compared years when using migration matrices and their CIs based on both cohort and duration calculation methods. The implication is that this is something that has to be taken into account by all risk managers, even those who only have exposures towards defensive sectors.

Using recession and expansion matrices is suggested as a tool to mitigate the impact of time-inhomogeneity. This approach is further elaborated on in section 6.3. A study of cumulative PD curves was used to decide what data should be the base for calculating recession and expansion matrices. When calculated, the recession and expansion matrices clearly showed the effect of economic cycle on e.g. estimated 1-year PDs. The differences are important, as the estimated PDs dictate the amount of capital buffer needed to put aside for e.g. expected losses. This knowledge is therefore important to be able to quote correct prices on transactions such as loans and derivatives that requires buffer capital to be put aside to cover the risk of counterparty credit risk.

6.3 Suggestions for further studies

This report has taken an application approach to estimating cumulative PDs through matrix migrations, and has a broad focus on appeared issues and necessities when implementing the Markov chain framework to rating migrations in practice. Many of the application components such as estimating matrices, means of measuring differences between matrices, calculating confidence intervals and statistically testing outcomes are research fields in their own right. However, the issue of time-inhomogeneity is a pressing one. The current researchers try to find solutions to this problem at the moment, and it is certainly an important field. To be able to take the next big leap in the Markov chain approach to estimating credit migration, this has to be solved. We have seen from previous studies that there has been several attempts to solve this issue, but there is no real consensus on a feasible approach.

From an application point of view, an extension of this analysis is to build upon the time-inhomogeneity issue and find better ways to estimate or simulate migration matrices, since this is one of the main issues with this approach. However, the first step would be to find some solution that is relatively easy to implement in practice. One way that is easy to implement is simply to, as suggested, use different migration matrices for expansions and recessions. First off, one would like to test expansion and recession matrices on future data. As discussed in section 6.1.4, the different PD estimations in time of recession and expansion can make a huge impact on the price of e.g. derivatives. Therefore, it is essential to make the right prediction of PD, which we have seen differ a lot from expansion to recessions. To test the mitigating effect of using these matrices would be highly interesting. This was not possible in this thesis due to the limited number of years in the data set. The feasibility of using expansion and recession matrices is of course dependent on how good the decisions can be *when* to use them. The issue to be solved is therefore how to decide *when* these matrices are to be used. In the longer run, for instance 8-10 years which roughly can be considered one economic cycle, it might be a good approximation to just go with the average matrix (tests on future data would be desirable also here). The time scope where the recession and expansion matrix decision issue has to be solved is therefore rather up to 5 years.

The first and quite blatant suggestion is that the solution could be based on internal views. One might simply have an own view on how the economy (and thereby rating migrations) might change the coming years and choose migration matrices accordingly. Otherwise, if this is to be implemented in a large institution, it might be possible to incorporate an internal macro group's or an internal credit research group's views to decide upon what type of economic climate that lies ahead. Migration matrices are then picked accordingly.

The second suggestion is to use some kind of market implied approach or macro factors to determine the migration matrices. One approach could be to find some publicly published variables that can serve as proxies to estimate the future economic climate. Suggestions could be e.g. industrial production or industrial capacity utilisation. Also, if one believes that the stock market is correlated with rating migrations, it might be feasible to try to use stock indices or the VIX index which moves quicker than stock indices and is a little forward looking. Possibly, a downturn in stock prices might happen before companies are in such bad shape that they get downgrades or default. However, it might be so that these variables are not capable to predict rating migrations 4 or 5 years from now. Therefore, I suggest that credit spreads or prices of credit default swaps (CDS) could be examined. These have predetermined maturities, for CDSs most lie between 1 to 10 years (5 is typical, see [28]). Thus, it should be possible to find data for the relevant time span. They also target credit quality more specifically, which is more closely related to credit rating migrations. Furthermore, examining credit spreads or CDSs could make it easier to conduct industry-specific research. One simply splits the data into industry categories, and thereby get better specific estimations for counterparties belonging to a given industry. In my personal point of view, I think that the credit spread or CDS alternative is interesting and would have that approach as my main suggestion for further studies.

To conclude, a suggestion for taking the whole field of calculating rating migrations using Markov chains further and thereby also the application approach further, is to investigate the time-inhomogeneity issue. From an application point of view, a first step would be to test if the approach of using recession and expansion matrices is feasible. If so, then the problem is to find decision rules on *when* to use them. A first but blatant suggestion to solution is to use internal views. It can be either the investor's own view or some internal expert group's, such as a credit research group. The second and more interesting suggestion is to use either publicly available factors as proxies for the future macro environment, or to use a more market implied approach with credit spreads and credit default swaps. However, the macro proxy variables might have more difficult to forsee the farther end of the 5-year interval than e.g. CDSs which should have data over a whole economic cycle. Moreover credit spreads and CDSs probably are more closely related to credit migrations. Therefore, I suggest investigating their relation to credit migrations and their usefulness in trying to predict credit migrations for time horizons shorter than 10 years.

Appendix A

Results - confidence intervals

This appendix contains result tables over bootstrapped 95% confidence intervals for whole matrices that were deemed to be a bit too unwieldy to have in the results chapter. The purpose is to make the reader able to verify some of the claims that were made when presenting the results, e.g. that cohort and duration matrices gave significantly different migration probability estimations in a large number of cases.

Table B.1: 95% two-sided cohort method confidence interval where all data in the *estimation* data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	96.99%	97.63%	1.30%	1.79%	0.06%	0.20%	0.23%	0.47%	0.40%	0.70%	0.05%	0.17%	0.00%	0.03%	0.00%	0.03%
AA	2.85%	3.30%	82.31%	83.29%	11.88%	12.78%	0.77%	1.04%	0.72%	0.95%	0.00%	0.02%	0.00%	0.04%	0.00%	0.05%
A	0.12%	0.18%	3.33%	3.61%	83.35%	83.94%	10.46%	10.93%	1.73%	1.93%	0.13%	0.18%	0.01%	0.04%	0.02%	0.05%
BBB	0.07%	0.10%	0.34%	0.40%	7.40%	7.68%	80.15%	80.57%	10.17%	10.48%	0.85%	0.95%	0.10%	0.13%	0.28%	0.33%
BB	0.14%	0.19%	0.12%	0.16%	1.04%	1.15%	13.24%	13.61%	78.84%	79.28%	4.37%	4.59%	0.64%	0.73%	0.89%	0.99%
B	0.10%	0.18%	0.03%	0.08%	0.30%	0.42%	2.85%	3.21%	15.24%	16.04%	73.10%	74.03%	3.06%	3.43%	3.79%	4.21%
CCC	0.00%	0.07%	0.04%	0.16%	0.17%	0.38%	1.10%	1.56%	4.91%	5.84%	10.61%	11.95%	72.90%	74.69%	7.34%	8.36%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Table B.2: 95% two-sided cohort method confidence interval where all data in the *validation* data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	97.03%	97.72%	1.23%	1.72%	0.11%	0.28%	0.19%	0.42%	0.44%	0.75%	0.01%	0.10%	0.00%	0.07%	0.00%	0.05%
AA	2.74%	3.20%	82.04%	83.08%	12.17%	13.03%	0.78%	1.03%	0.78%	1.04%	0.00%	0.03%	0.00%	0.00%	0.01%	0.07%
A	0.12%	0.18%	3.23%	3.51%	83.66%	84.20%	10.23%	10.68%	1.76%	1.96%	0.14%	0.20%	0.00%	0.02%	0.03%	0.06%
BBB	0.06%	0.09%	0.33%	0.39%	7.24%	7.51%	80.47%	80.86%	10.01%	10.32%	0.92%	1.03%	0.08%	0.11%	0.27%	0.32%
BB	0.16%	0.20%	0.09%	0.13%	1.02%	1.13%	13.18%	13.55%	78.91%	79.36%	4.32%	4.53%	0.64%	0.73%	0.96%	1.07%
B	0.08%	0.15%	0.03%	0.08%	0.27%	0.40%	2.93%	3.29%	15.38%	16.18%	72.80%	73.72%	3.26%	3.65%	3.68%	4.08%
CCC	0.01%	0.09%	0.02%	0.12%	0.22%	0.46%	1.02%	1.43%	4.86%	5.76%	11.23%	12.50%	72.92%	74.67%	6.81%	7.88%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Table B.3: 95% two-sided duration method confidence interval where all data in the *estimation* data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	96.73%	97.29%	1.23%	1.61%	0.22%	0.33%	0.39%	0.58%	0.54%	0.77%	0.09%	0.19%	0.01%	0.03%	0.01%	0.04%
AA	2.32%	2.66%	80.76%	81.67%	12.95%	13.70%	1.55%	1.72%	1.08%	1.27%	0.06%	0.10%	0.02%	0.04%	0.03%	0.06%
A	0.16%	0.20%	3.31%	3.53%	83.11%	83.63%	10.10%	10.47%	2.25%	2.39%	0.26%	0.31%	0.04%	0.06%	0.08%	0.10%
BBB	0.09%	0.11%	0.53%	0.58%	6.99%	7.24%	80.68%	81.10%	9.62%	9.89%	1.08%	1.15%	0.15%	0.17%	0.29%	0.33%
BB	0.14%	0.18%	0.28%	0.33%	1.62%	1.72%	12.14%	12.48%	78.98%	79.44%	4.53%	4.71%	0.67%	0.74%	0.97%	1.05%
B	0.11%	0.17%	0.11%	0.16%	0.58%	0.68%	3.78%	4.06%	14.15%	14.91%	70.70%	71.87%	4.13%	4.48%	4.83%	5.23%
CCC	0.03%	0.09%	0.13%	0.26%	0.36%	0.52%	1.74%	2.10%	5.60%	6.28%	9.72%	10.76%	69.47%	71.60%	10.13%	11.27%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Table B.4: 95% two-sided duration method confidence interval where all data in the *validation* data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	96.72%	97.29%	1.22%	1.58%	0.28%	0.43%	0.32%	0.50%	0.49%	0.73%	0.11%	0.22%	0.01%	0.05%	0.01%	0.06%
AA	2.28%	2.61%	80.59%	81.63%	13.10%	13.93%	1.49%	1.66%	1.10%	1.30%	0.07%	0.10%	0.01%	0.02%	0.02%	0.06%
A	0.15%	0.19%	3.25%	3.47%	83.22%	83.74%	10.10%	10.47%	2.19%	2.33%	0.28%	0.32%	0.04%	0.05%	0.09%	0.12%
BBB	0.08%	0.11%	0.53%	0.58%	6.84%	7.07%	80.97%	81.40%	9.48%	9.75%	1.10%	1.18%	0.14%	0.16%	0.29%	0.33%
BB	0.14%	0.18%	0.27%	0.31%	1.57%	1.66%	12.10%	12.44%	79.05%	79.53%	4.52%	4.72%	0.67%	0.74%	1.01%	1.10%
B	0.09%	0.15%	0.10%	0.14%	0.60%	0.70%	3.80%	4.07%	14.24%	14.93%	70.63%	71.71%	4.13%	4.47%	4.92%	5.31%
CCC	0.03%	0.10%	0.12%	0.23%	0.41%	0.59%	1.74%	2.07%	5.41%	6.12%	10.03%	11.15%	69.24%	71.46%	10.09%	11.22%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Table B.5: 95% two-sided cohort method confidence interval where all data in the final data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	97.11%	97.55%	1.34%	1.67%	0.11%	0.22%	0.25%	0.40%	0.47%	0.67%	0.05%	0.12%	0.01%	0.04%	0.00%	0.04%
AA	2.86%	3.18%	82.38%	83.07%	12.14%	12.73%	0.82%	1.01%	0.78%	0.96%	0.00%	0.02%	0.00%	0.02%	0.01%	0.05%
A	0.13%	0.17%	3.32%	3.51%	83.60%	83.99%	10.40%	10.74%	1.77%	1.91%	0.14%	0.18%	0.01%	0.03%	0.03%	0.05%
BBB	0.07%	0.09%	0.34%	0.39%	7.36%	7.55%	80.37%	80.65%	10.14%	10.36%	0.90%	0.97%	0.09%	0.12%	0.28%	0.32%
BB	0.16%	0.19%	0.11%	0.14%	1.05%	1.13%	13.27%	13.54%	78.94%	79.25%	4.38%	4.54%	0.65%	0.72%	0.94%	1.02%
B	0.10%	0.16%	0.03%	0.07%	0.30%	0.39%	2.93%	3.19%	15.41%	15.96%	73.09%	73.77%	3.22%	3.50%	3.79%	4.08%
CCC	0.01%	0.07%	0.04%	0.12%	0.23%	0.39%	1.13%	1.44%	5.01%	5.67%	11.10%	12.00%	73.20%	74.42%	7.22%	7.94%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Table B.6: 95% two-sided duration method confidence interval where all data in the final data set has been used. Based on 1000 bootstrap replications.

Rating	AAA (lower,upper)		AA (lower,upper)		A (lower,upper)		BBB (lower,upper)		BB (lower,upper)		B (lower,upper)		CCC (lower,upper)		D (lower,upper)	
AAA	96.81%	97.21%	1.27%	1.53%	0.27%	0.36%	0.38%	0.52%	0.55%	0.72%	0.11%	0.19%	0.01%	0.04%	0.01%	0.04%
AA	2.36%	2.59%	80.83%	81.52%	13.13%	13.70%	1.54%	1.67%	1.12%	1.25%	0.07%	0.09%	0.01%	0.03%	0.03%	0.05%
A	0.17%	0.19%	3.30%	3.45%	83.23%	83.60%	10.15%	10.41%	2.24%	2.34%	0.28%	0.31%	0.04%	0.05%	0.09%	0.11%
BBB	0.09%	0.10%	0.54%	0.57%	6.96%	7.12%	80.88%	81.18%	9.58%	9.78%	1.10%	1.15%	0.15%	0.16%	0.30%	0.32%
BB	0.15%	0.17%	0.28%	0.31%	1.61%	1.67%	12.16%	12.39%	79.09%	79.42%	4.55%	4.68%	0.68%	0.73%	1.00%	1.06%
B	0.11%	0.15%	0.11%	0.14%	0.60%	0.68%	3.83%	4.04%	14.33%	14.79%	70.88%	71.58%	4.18%	4.42%	4.95%	5.21%
CCC	0.04%	0.08%	0.14%	0.22%	0.40%	0.52%	1.78%	2.02%	5.62%	6.12%	10.04%	10.81%	69.64%	71.23%	10.28%	11.08%
D	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%

Appendix B

Wald and bootstrapped CI comparison

This appendix shows a brief study of the Wald confidence intervals compared to the bootstrapped confidence intervals for the cohort method. The purpose is to suggest whether a bootstrap method is necessary or if the analytical Wald confidence interval is enough when using the cohort method. The study is deferred to the appendix pages as practitioners probably will use the duration method more, plus the fact that different confidence interval calculation methods are not the main focus of this thesis.

We have already seen in table 5.1 that the Wald CIs and the bootstrapped CIs for the average cohort matrix were very similar. Confidence intervals were also calculated for the annual matrices 2009, 2010, 2011 and 2012 that was part of the time-inhomogeneity study. These years will be studied again as they are the most recent and has the highest number of observations. As can be seen from the tables B.1 and B.2, the 95% confidence intervals yet again look very similar.

Table B.1: Table over Wald CIs and bootstrapped 95% CIs for years 2009 and 2010. Calculations are based on the full sample.

Rating	Bootstrap 2009	Wald 2009	Bootstrap 2010	Wald 2010
AAA	$[0.00\%, 0.00\%]$	$[0.00\%, 0.00\%]$	$[0.00\%, 0.09\%]$	$[0.00\%, 0.09\%]$
AA	$[0.00\%, 0.00\%]$	$[0.00\%,0.00\%]$	$[0.00\%, 0.00\%]$	$[0.00\%, 0.00\%]$
A	$[0.06\%, 0.15\%]$	$[0.05\%, 0.15\%]$	$[0.01\%, 0.05\%]$	$[0.00\%, 0.05\%]$
BBB	$[0.61\%, 0.80\%]$	$[0.60\%, 0.80\%]$	$[0.18\%, 0.30\%]$	$[0.19\%, 0.30\%]$
BB	$[1.60\%, 1.92\%]$	$[1.60\%, 1.93\%]$	$[0.83\%, 1.06\%]$	$[0.82\%, 1.05\%]$
B	$[4.83\%, 6.00\%]$	$[4.84\%, 6.02\%]$	$[4.55\%, 5.62\%]$	$[4.57\%, 5.61\%]$
CCC	$[6.37\%, 8.87\%]$	$[6.31\%, 8.83\%]$	$[4.72\%, 6.63\%]$	$[4.67\%, 6.63\%]$

According to the theory of Wald confidence intervals, the estimated PD for a rating R should follow a normal distribution as in equation (B.1)

$$\widehat{PD}_R \sim N\left(\widehat{PD}_R, \frac{\widehat{PD}_R(1-\widehat{PD}_R)}{n_R}\right) \tag{B.1}$$

Table B.2: Table over Wald CIs and bootstrapped 95% CIs for years 2011 and 2012. Calculations are based on the full sample.

Rating	Bootstrap 2011	Wald 2011	Bootstrap 2012	Wald 2012
AAA	$[0.00\%, 0.00\%]$	$[0.00\%, 0.00\%]$	$[0.00\%, 0.07\%]$	$[0.00\%, 0.07\%]$
AA	$[0.00\%, 0.00\%]$	$[0.00\%, 0.00\%]$	$[0.00\%, 0.05\%]$	$[0.00\%, 0.04\%]$
A	$[0.01\%, 0.06\%]$	$[0.01\%, 0.05\%]$	$[0.01\%, 0.05\%]$	$[0.01\%, 0.05\%]$
BBB	$[0.13\%, 0.21\%]$	$[0.12\%, 0.21\%]$	$[0.51\%, 0.66\%]$	$[0.51\%, 0.66\%]$
BB	$[0.75\%, 0.96\%]$	$[0.75\%, 0.96\%]$	$[1.12\%, 1.38\%]$	$[1.12\%, 1.37\%]$
B	$[3.75\%, 4.73\%]$	$[3.77\%, 4.71\%]$	$[5.36\%, 6.46\%]$	$[5.32\%, 6.47\%]$
CCC	$[6.80\%, 8.90\%]$	$[6.78\%, 8.95\%]$	$[6.61\%, 8.90\%]$	$[6.63\%, 8.91\%]$

where the n_R is the number of firms that began that year in rating R.

To determine whether this is a good assumption for percentiles other than the ones we just studied, we can test if the bootstrapped sample appears to be from a distribution as described by equation (B.1). We will use a Kolmogorov-Smirnov (K-S) two-sample test, as described in section 2.1. One sample will be the bootstrapped empirical sample, the other will be a generated normal distribution sample with the mean and variance dictated by the Wald theory.

Ratings AAA and AA will not be tested, as the bootstrapped samples will have piles of stacked observations at zero probability which will blur the test results. The null hypothesis is that the bootstrapped empirical sample and the generated normal distribution sample have the same distribution. Both samples have the same size, 1000 observations.

To K-S two-sample test is calculated using the MATLAB function `kstest2`, with significance level 5%. That means that if the calculated p-value is less then 0.05, then the null hypothesis is rejected. Thus, we only reject the null hypothesis (that they come from the same distribution) if the risk of taking the wrong decision when rejecting null hypothesis is less than 5%. In table B.3, H and p-values are tabulated. If $H = 1$, then the null hypothesis is rejected. If $H = 0$, then the null hypothesis is accepted.

Table B.3: Table over Kolmogorov-Smirnov test outputs for years 2009, 2010, 2011 and 2012. Calculations are based on the full sample.

Rating	2009 [H,p]	2010 [H,p]	2011 [H,p]	2012 [H,p]
A	$[0, 0.23]$	$[1, 0.00]$	$[1, 0.00]$	$[1, 0.00]$
BBB	$[0, 0.18]$	$[0, 0.49]$	$[0, 0.79]$	$[0, 0.98]$
BB	$[0, 0.72]$	$[0, 0.13]$	$[0, 0.39]$	$[0, 0.95]$
B	$[0, 0.26]$	$[1, 0.00]$	$[0, 0.33]$	$[0, 0.64]$
CCC	$[0, 0.88]$	$[1, 0.01]$	$[0, 0.95]$	$[0, 0.09]$

From table B.3, we can see that the null hypothesis is rejected in 25% of the cases, with most rejections on the rating with lowest PD. However, there are a couple of more ratings with relatively low p-value. To clarify the similarities and differences between Wald and bootstrapped CIs, cumulative distribution functions from two rejected and two accepted tests are shown in figures B.1, B.2, B.3 and B.4. They represent the best and worst p-values for both rejected and accepted null hypothesis. Whether the analytical confidence intervals are good substitutes for the bootstrapped ones is depending on each investor's or researcher's preferences. At least the Wald CIs seem to have a good fit to the bootstrap for the ratings BBB to CCC. These results are in line with Hanson and Schuermann (2005) who also find that the speculative grade rating distributions are "surprisingly close to normal (Gaussian)". The Wald CIs are also easy and fast to compute.

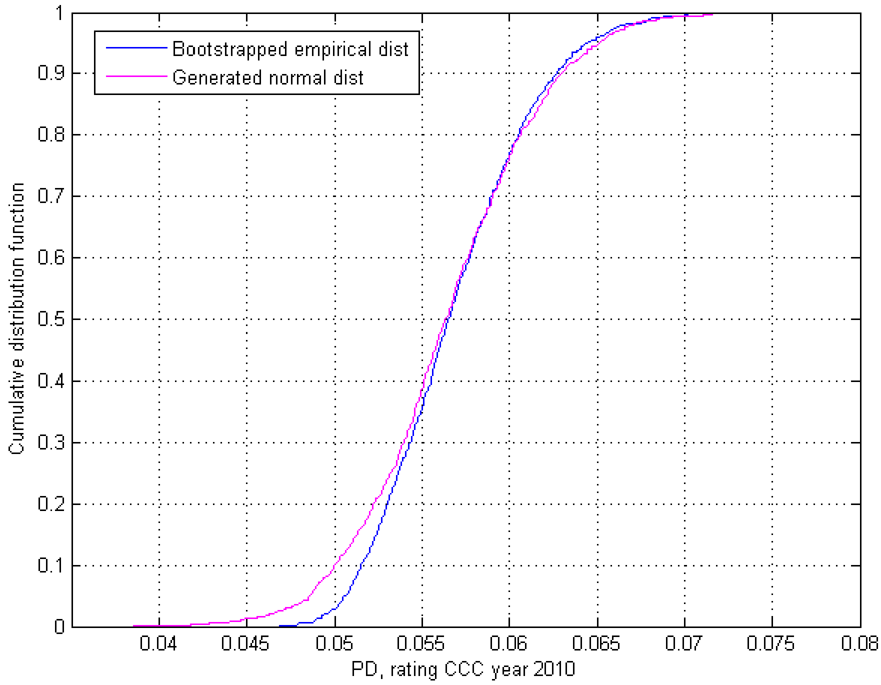

Figure B.1: Cumulative distribution functions for a bootstrapped PD estimation sample and its generated Gaussian counterpart.

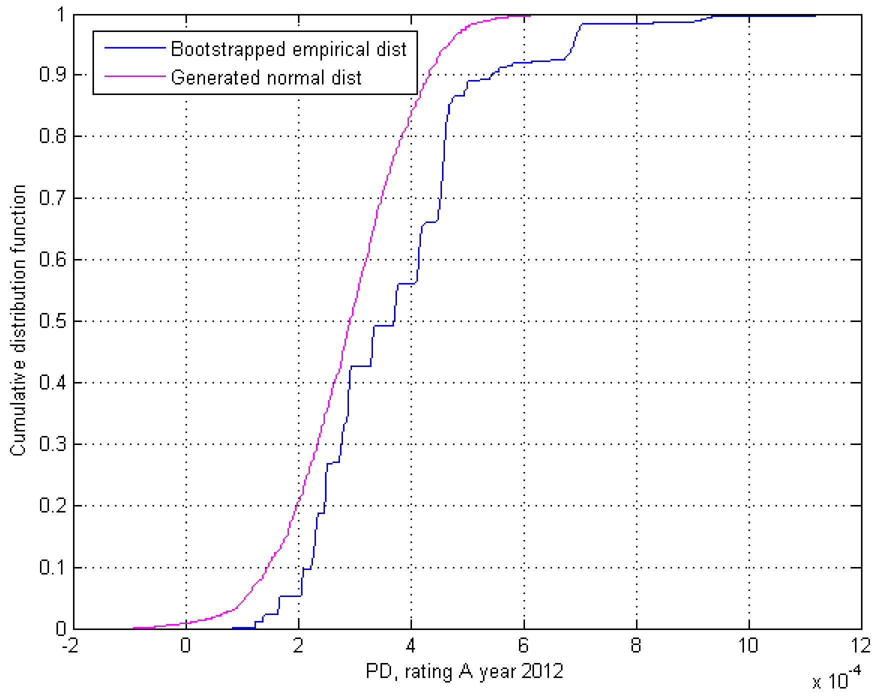

Figure B.2: Cumulative distribution functions for a bootstrapped PD estimation sample and its generated Gaussian counterpart.

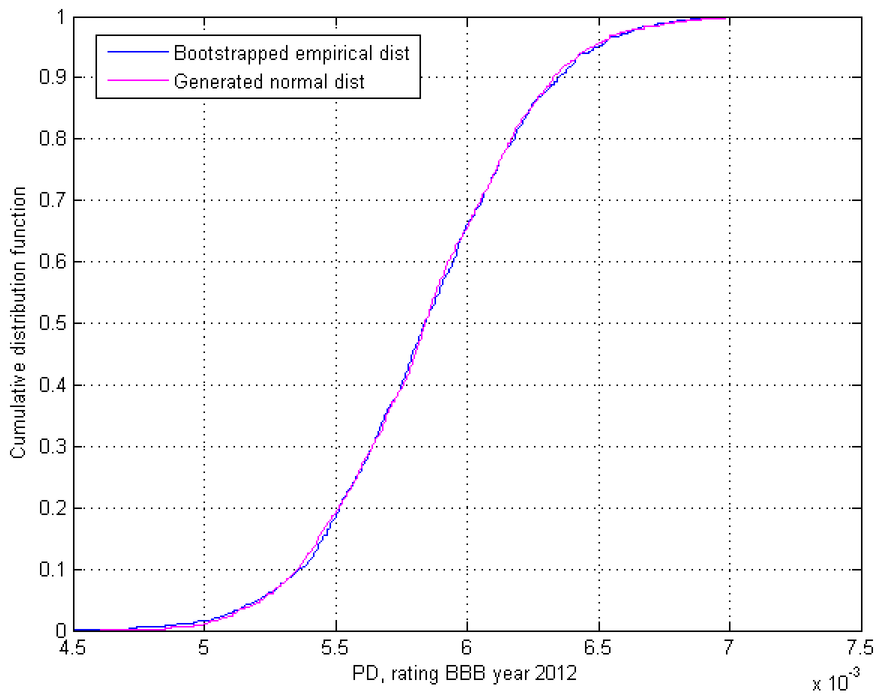

Figure B.3: Cumulative distribution functions for a bootstrapped PD estimation sample and its generated Gaussian counterpart.

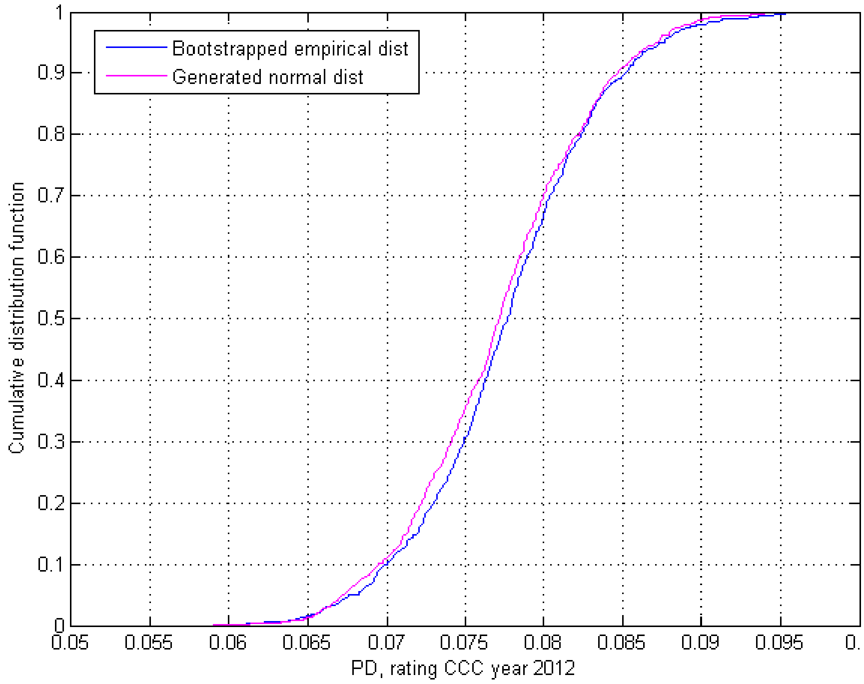

Figure B.4: Cumulative distribution functions for a bootstrapped PD estimation sample and its generated Gaussian counterpart.

Appendix C

Defensive sectors

Below follow the list of NACE sections and divisions that were chosen to form the defensive companies subset.

A AGRICULTURE, FORESTRY AND FISHING	1 Crop and animal production,hunting and related service activities
A AGRICULTURE, FORESTRY AND FISHING	2 Forestry and logging
A AGRICULTURE, FORESTRY AND FISHING	3 Fishing and aquaculture
C MANUFACTURING	10 Manufacture of food products
C MANUFACTURING	11 Manufacture of beverages
C MANUFACTURING	12 Manufacture of tobacco products
C MANUFACTURING	13 Manufacture of textiles
C MANUFACTURING	14 Manufacture of wearing apparel
C MANUFACTURING	15 Manufacture of leather and related products
C MANUFACTURING	16 Manufacture of wood and products of wood and cork;except furniture;manufacture of articles of straw and plaiting materials
C MANUFACTURING	17 Manufacture of paper and paper products
C MANUFACTURING	18 Printing of reproduction of recorded media
C MANUFACTURING	19 Manufacture of coke and refined petroleum products
C MANUFACTURING	20 Manufacture of chemicals and chemical products
C MANUFACTURING	21 Manufacture of basic pharmaceutical products and pharmaceutical preparations
C MANUFACTURING	22 Manufacture of rubber and plastic products
C MANUFACTURING	23 Manufacture of other non-metallic mineral products
C MANUFACTURING	31 Manufacture of furniture
C MANUFACTURING	32 Other manufacturing
C MANUFACTURING	33 Repair and installation of machinery and equipment
D ELECTRICITY,GAS,STEAM AND AIR CONDITIONING SUPPLY	35 Electricity,gas,steam and air conditioning supply
E WATER SUPPLY;SEWERAGE,WASTE MANAGEMENT AND REMEDIATION ACTIVITIES	36 Water collection, treatment and supply
E WATER SUPPLY;SEWERAGE,WASTE MANAGEMENT AND REMEDIATION ACTIVITIES	37 Sewerage
E WATER SUPPLY;SEWERAGE,WASTE MANAGEMENT AND REMEDIATION ACTIVITIES	38 Waste collection,treatment and disposal activities;materials recovery
E WATER SUPPLY;SEWERAGE,WASTE MANAGEMENT AND REMEDIATION ACTIVITIES	39 Remediation activities and other waste management services
F CONSTRUCTION	42 Civil engineering

G WHOLESALE AND RETAIL TRADE;REPAIR OF MOTOR VEHICLES AND MOTORCYCLES	46 Wholesale trade,except of motor vehicles and motorcycles
G WHOLESALE AND RETAIL TRADE;REPAIR OF MOTOR VEHICLES AND MOTORCYCLES	47 Retail trade,except of motor vehicles and motorcycles
H TRANSPORTATION AND STORAGE	53 Postal and courier activities
J INFORMATION AND COMMUNICATION	58 Publishing activities
J INFORMATION AND COMMUNICATION	59 Motion picture, video and television programme production, sound recording and music publishing activities
J INFORMATION AND COMMUNICATION	60 Programming and broadcasting activities
J INFORMATION AND COMMUNICATION	61 Telecommunications
J INFORMATION AND COMMUNICATION	63 Information service activities
K FINANCIAL AND INSURANCE ACTIVITIES	65 Insurance, reinsurance and pension funding, except compulsory social security
L REAL ESTATE ACTIVITIES	68 Real estate activities
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	69 Legal and accounting activities
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	70 Activities of head offices; management consultancy activities
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	72 Scientific research and development
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	73 Advertising and market research
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	74 Other professional, scientific and technical activities
M PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	75 Veterinary activities
N ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	77 Rental and leasing activities
N ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	80 Security and investigation activities
N ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	81 Services to buildings and landscape activities
N ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	82 Office administrative, office support and other business support activities
O PUBLIC ADMINISTRATION AND DEFENCE;COMPULSORY SOCIAL SECURITY	84 Public administration and defence;compulsory social security
P EDUCATION	85 Education
Q HUMAN HEALTH AND SOCIAL WORK ACTIVITIES	86 Human health activities
Q HUMAN HEALTH AND SOCIAL WORK ACTIVITIES	87 Residential care activities
Q HUMAN HEALTH AND SOCIAL WORK ACTIVITIES	88 Social work activities without accommodation
R ARTS, ENTERTAINMENT AND RECREATION	90 Creative, arts and entertainment activities
R ARTS, ENTERTAINMENT AND RECREATION	91 Libraries, archives, museums and other cultural activities
R ARTS, ENTERTAINMENT AND RECREATION	92 Gambling and betting activities
S OTHER SERVICE ACTIVITIES	95 Repair of computers and personal and household goods
T ACTIVITIES OF HOUSEHOLDS AS EMPLOYERS;UNDIFFERENTIATED GOODS- AND SERVICES-PRODUCING ACTIVITIES OF HOUSEHOLDS FOR OWN USE	98 Undifferentiated goods-and services-producing activities of private households for own use

References

[1] T. Anderson, L. Goodman, 1957, *Statistical Inference about Markov Chains*, Annals of Mathematical Statistics, Vol 28 No. 1, 89-110

[2] A. Andersson, P. Vanini, 2010, *Credit-migration Risk Modeling*, The Journal of Credit Risk, Vol 6 No. 1, 3-30

[3] A. Bangia, F. Diebold, T. Schuermann, 2002, *Rating Migrations and the Business Cycle, with Application to Credit Portfolio Stress Testing*, Journal of Banking & Finance, Vol 26, 445-474

[4] J. Christensen, E. Hansen, D. Lando, 2004, *Confidence sets for continuous-time rating transition probabilities*, Journal of Banking & Finance, Vol 28 No. 11, 2575-2602

[5] J. Enger, J. Grandell, 2006, *Markovprocesser och köteori*, available at `http://www.math.kth.se/matstat/gru/markovkomp.pdf`, accessed at February 12, 2014

[6] European Parliament, *Regulation (EU) No 575/2013 on prudential requirements for credit institutions and investment firms*, available at `http://eur-lex.europa.eu/legal-content/EN/ALL/;jsessionid=8VMkT1GNhBfhNnThS2lPp6yjPjD27mmmNdYbNT14X0VlbwjLhXmv!-1268091117?uri=CELEX:32013R0575`, accessed at May 23, 2014

[7] Eurostat, *Correspondence tables Nace Rev. 1.1 - Nace Rev. 2*, available at `http://epp.eurostat.ec.europa.eu/portal/page/portal/nace_rev2/correspondence_tables`, accessed at June 3, 2014

[8] Forbes.com, *Defensive sectors are no help in market downturns*, available at `http://www.forbes.com/sites/sharding/2014/05/09/defensive-sectors-are-no-help-in-market-downturns/`, accessed at August 12, 2014

[9] S. Hanson, T. Shuermann, 2005 *Confidence Intervals for Probabilities of Default*, available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=766345`, accessed January 19, 2014

[10] R. Israel, J. Rosenthal, J. Wei, 2001, *Finding generators for Markov chains via empirical transition matrices*, Mathematical Finance, Vol 11 No. 2, 245-265

[11] Y. Jafry, T. Schuermann, 2003, *Metrics for Comparing Credit Migration Matrices*, Financial Institutions Center, Wharton

[12] Y. Jafry, T. Schuermann, 2004, *Measurement, estimation and comparison of credit rating matrices*, Journal of Banking & Finance, Vol. 28, 2603-2639

[13] N. Kiefer, C. Larson, 2004, *Testing Simple Markov Structures for Credit Rating Transitions*, available at http://www.occ.gov/publications/publications-by-type/occ-working-papers/2008-2000/wp2004-3.pdf, accessed at August 6, 2014

[14] N. Kiefer, C. Larson, 2006 *A Simulation Estimator for Testing the Time Homogeneity of Credit Rating Transitions*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=956623, accessed at August 4, 2014

[15] D. Lando, 2004, *Credit Risk Modeling - Theory and Applications*, Princeton University Press, Princeton

[16] D. Lando, T. Skødeberg, 2002, *Analyzing rating transitions and rating drift with continuous observations*, Journal of Banking & Finance, Vol 26, 423-444

[17] P. Lencastre, P. Lind, F. Raischel, T. Rogers, 2014 *Are credit ratings time-homogeneous and Markov?*, available at http://arxiv.org/pdf/1403.8018v1.pdf, accessed at July 30, 2014

[18] Markit, *Credit Derivatives Glossary*, available at http://www.markit.com/assets/en/docs/products/data/indices/credit-index-annexes/CDS_glossary.pdf, accessed 23 May, 2014.

[19] J. Norris, 1997, *Markov Chains*, Cambridge University Press, Cambridge.

[20] S. Rachev, S. Trueck, 2005, *Credit Portfolio Risk and PD Confidence Sets through the Business Cycle*, available at http://www.ams.sunysb.edu/~rachev/publication/tr_credit_portfolio_risk.pdf, accessed July 29, 2014

[21] S. Rachev, S. Trueck, 2009, *Rating Based Modeling of Credit Risk - Theory and Application of Migration Matrices*, Academic Press, San Diego

[22] SICCODE.com - World Business Directory, *What is a NACE Code?*, available at https://siccode.com/en/pages/what-is-a-nace-code, accessed June 3, 2014

[23] Standard & Poor's, *About Credit Ratings*, available at http://www.standardandpoors.com/aboutcreditratings/RatingsManual_PrintGuide.html, accessed February 10, 2014

[24] Standard & Poor's, *2012 Annual Global Corporate Default Study And Rating Transitions*, available at http://www.nact.org/resources/NACT_2012_Global_Corporate_Default.pdf, accessed 26 May, 2014

[25] L. Thomas, D. Edelman, J. Crook, 2002, *Credit Scoring and Its Applications*, SIAM, Philadelphia

[26] Wikipedia.org *Kolmogorov-Smirnov test*, available at http://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test, accessed at August 16, 2014

[27] Wikipedia.org, *Statistical hypothesis testing*, available at http://en.wikipedia.org/wiki/Statistical_hypothesis_testing, accessed August 4, 2014

[28] Wikipedia.org *Credit Default Swap*, available at http://en.wikipedia.org/wiki/Credit_default_swap, accessed at August 6, 2014

[29] L. Yang, 2008, *Credit Migration Index*, available at http://summit.sfu.ca/item/724, accessed at August 5, 2014