ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ ΑΝΑΛΟΓΙΣΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΔΙΟΙΚΗΤΙΚΗ ΚΙΝΔΥΝΟΥ

ΠΙΝΑΚΕΣ ΜΕΤΑΒΑΣΗΣ ΣΤΟΝ ΠΙΣΤΩΤΙΚΟ ΚΙΝΔΥΝΟ

Σαράντης Καπόλος

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων του για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου

Πειραιάς

Ιούλιος 2020

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμόν ……………………….. συνεδρίαση του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Αναλογιστική Επιστήμη και Διοικητική Κινδύνου.

Τα μέλη της επιτροπής ήταν:

- Πολίτης Κωνσταντίνος
- Κούτρας Μάρκος (Επιβλέπων)
- Ανθρωπέλος Μιχαήλ

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα

UNIVERSITY OF PIRAEUS

**DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN ACTUARIAL SCIENCE
AND RISK MANAGEMENT**

TRANSITION MATRICES IN CREDIT RISK

Sarantis Kapolos

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Actuarial Science and Risk Management.

Piraeus, Greece

July 2020

This thesis was approved unanimously by the three-member Commission of Inquiry appointed by the Department of Statistics and Insurance Science of the University of Piraeus No. ………………………… meeting with the by laws of the Postgraduate Program in Actuarial Science and Risk Management.

Committee members were:

- Politis Konstantinos
- Koutras Markos (Supervisor)
- Anthropelos Michail

The approval of the thesis by the Department of Statistics and Insurance Science of the University of Piraeus does not imply acceptance of opinions of the author

Contents

ABSTRACT

During the last decades credit risk has been proved one of the greatest risks an institution is facing. This fact has force banks and the regulators all over the world to reassess the importance of credit risk and establish it as a core concept on their everyday activities. This thesis presents credit risk measurement approaches, using transition probability matrices, which have a substantial effect on loan pricing. Through the recent years a method using Markov chains, in order to calculate transition probabilities and probability of default has evolved and has been established as an industry standard for both continuous and discrete time periods. Moreover this thesis investigates the different assumptions, that are crucial for our model implementation and statistical measures that help us to investigating whether these assumptions hold true. Finally we will present other measures related to transition probability matrices which are, crucial for decision making in the banking sector.

Chapter 1

Introduction

Risk management is a core concept in the financial sector because of it's substantial effect not only on the behavior of the financial institution, but also to the economy of a country as well as to the entire world. For that reason, risk management nowadays attracts attention on all levels of an organization over the world. In addition, as the financial industry becomes more competitive and complex, bankers and financial mangers have moved away from the traditional way of making profit. In other words it is not only sufficient to earn high returns from investments but also it is important to know if the earned return corresponds to the risk the financial institute accepted. This is why quantifying risks and trying to find an optimal mix between taking risks and maximizing returns is a very important procedure.

However risk management is a broad and complex process and it is defined as: "the identification assessment and prioritization of risks followed by coordinated and economical application of resources to minimize, monitor and control the probability or impact of unfortunate events or to maximize the realization of opportunities".

The main risks faced by typical financial institutions fall into the broad categories: credit, market, operational, liquidity risks. Among the risks that a financial institution faces, credit risk is one of the most important ones. Proof for that statement is the fact that credit risk is mentioned in both Basel III (banking) and solvency II (insurance) where a financial institution has the obligation to hold capital for a variety of risks. Among those risks is credit risk as well.

Having mentioned all the above it is time to define credit risk. Credit risk is the probable risk of loss resulting from a borrower's failure to repay a loan or meet contractual obligations. Traditionally, it refers to the risk that a lender may not receive the owned principal and interest, which results in an interruption of cash flows and increased costs for collection.

Although it is impossible to know exactly who will default on obligations, properly assessing and managing credit risk can decrease the severity of loss. Even if potential losses come from defaults , in credit risk it is not clear yet what default exactly means. For some organizations such as country rating agencies default means that the obligor is unlikely to pay its credit obligations (principal, interest or fees) to the institution in full, while for others like banks it means that the obligor is past due more than 90 days on any credit obligation to the institution.

No matter whether the default is complete or partial it is a highly unpleasant phenomenon in today's world for both the borrower and the lender. Due to the fact that losses in credit risk come from defaults, we will present different types of defaults in order to understand what credit risk really is.

Sovereign defaults: Sovereign default is the failure of a government or a country to pay back it's debt in full or on time. One example is Greece, which defaulted on an IMF loan in 2015 and paid it's debt at the end of the month. In such cases, the defaulting country and the creditor are more likely to renegotiate the interest rate, length of the loan, or the principal payments. Sovereign default can occur for many reasons such as:

1. Poor macroeconomic performance
2. Political instability
3. Unwise lending
4. Rollover risk
5. Poor credit history

6. Weak revenues

7. Rising interest rates

Orderly default: In times of acute insolvency crises, it can be advisable for regulators and lenders to preemptively engineer the methodic restructuring of a nation's public debt—also called "orderly default" or "controlled default". Experts who favor this approach to solve a national debt crisis typically agree that a delay in organizing an orderly default would wind up hurting lenders and neighboring countries even more.

Strategic default: A strategic default is the decision by a borrower to stop making payments on a debt , despite having the financial ability to make the payments. This is particularly assosiated with residential and commercial mortages, in which case it usually occurs after a substantial drop in the house's price so that the debt owed is greater than the value of the property. The property has a negative equity and is expected to remain so for the forseeable future, such as following the bursting of a real estate bubble.

Consumer default: Consumer default frequently occurs in rent or mortgage payments, consumer credit, or utility payments. It is the default of borrower who, for reasons of inability he can not pay his obligations to the borrower. Contrary to strategic default where the borrower intentionally refuses to pay his debt, here default occurs due to consumer's finance inability. A European Union wide analysis identified certain risk groups, such as single households, being unemployed – even after correcting for the (significant) impact of having a low income, being young (especially being younger than around 50 years old), being unable to rely on social networks, etc. Even internet illiteracy has been associated with increased default, potentially caused by those households being less likely to find their way to the social benefits they are often entitled to. While effective non-legal debt counseling is usually the preferred option, more economic and less disruptive, consumer default can end up in legal debt settlement or consumer bankruptcy procedures, ranging from 1-year procedures in the UK to 6-year procedures in Germany.

Business default: Business default is the situation where a business is unable to pay obligations to clients. There are many examples such as:

1. A company takes a loan or issues a bond and bankrupts before it repays it
2. A bank can not return funds to depositors
3. An insurance company is not able to pay a policy obligation
4. An option derivative agreement gets canceled because of bankruptcy

For the reasons mentioned above we can create a credit control system which will hedge credit risk in order to minimize our losses. There is a big variety of ways we can do that such as

- Risk-based pricing. This classical technique uses higher interest rates to borrowers with higher probability of default.With this way we can collect money faster and minimize our losses in case of default
- Collateral. Banks might ask for collateral, if necessary
- Avoid risk. Lenders can decide not to give a loan if expected losses are relatively big
- Credit insurance. Lenders and bond holders may transfer the risk to another counterpart(insurer) by paying a premium.
- Deposit insurance. Governments can establish deposit insurance to guarantee bank deposits in order to prevent a bank default

All companies face risk, without risk there is no reward. The flip side of this is that too much risk can lead to business failure. Risk management allows a balance to be struck between taking risks and reducing them. Effective risk management can add value to any organization. In particular, companies operating in the investment industry rely heavily on risk management as the process that allows them to withstand market crashes. An effective risk management framework seeks to protect an organization's capital base and earnings without hindering growth. Furthermore, investors are more willing to invest in companies with good risk management practices. This results in lower borrowing costs, easier access to capital for the firm and improved long-term performance. There are four crucial components that should be considered when creating a risk management framework.

- risk identification
- risk measurement
- risk mitigation
- risk reporting

We provide a short description of each one of these components:

Risk identification The first step in identifying the risks a company faces is to define the risk universe. The risk universe is simply a list of all possible risks. Examples include IT risk, operational risk, regulatory risk, legal risk, political risk, strategic risk and credit risk. After listing all possible risks, the company can then select the risks to which it is exposed and categorize them into core and non-core risks. Core risks are those that the company must take in order to drive performance and long-term growth. Non-core risks are often not essential and can be minimized or eliminated completely. In credit risk it is obvious that it is quite easy to identify the risk which is loss due to failure of a borrower to meet their obligations.

Risk measurement Risk measurement provides information on the quantum of either a specific risk exposure or an aggregate risk exposure, and the probability of a loss occurring due to those exposures. When measuring specific risk exposure it is important to consider the effect of that risk on the overall risk profile of the organization. Some risks may provide diversification benefits while others may not. Another important consideration is the ability to measure an exposure. Some risks may be easier to measure than others. For example, market risk can be measured using observed market prices, but measuring operational risk is considered both an art and a science.

Risk Mitigation Having categorized and measured its risks, a company can then decide on which risks to eliminate or minimize, and how much of its core risks to retain. Risk mitigation can be achieved through an outright sale of assets or liabilities, buying insurance, hedging with derivatives or diversification.

Risk reporting It is important to report regularly on specific and aggregate risk measures in order to ensure that risk levels remain at an optimal level. Financial institutions that trade daily will produce daily risk reports. Other institutions may require less frequent reporting. Risk reports must be sent to risk personnel who have that authority to adjust (or instruct others to adjust) risk exposures.

In this thesis we are going to focus on calculating quantities that will provide assistance in measurement of credit risk. Although hedging the credit risk is very useful in our analysis we need to quantify credit risk in order to apply the hedging techniques we mentioned above. We can do that by using a credit migration matrices approach. Through these matrices we may quantify variables such as credit worthiness, probability of default, expected time until default, expected losses ect.These quantities can be used to protect shareholders from credit events and increase firms value and credibility by reducing it's credit losses. The present thesis is going to describe how to calculate these quantities . Specifically:

Chapter 2 defines the Markov property as well as credit migration matrices and their applications. We shall also define multiple step transition matrices and give details on how we can use them in order to decide about our client's future credit worthiness.

Chapter 3 deals with the construction of a transition matrix and how to test whether a very useful property of a transition matrix called "time homogeneity" holds true. Moreover we our going to introduce different methods, which can help us compare different transition matrices with each other.

Finally in Chapter 4 we present several associated with quantities from transition matrices such as time until default, number of visits in a specific rating scale, absorbing probability, ect .

Chapter 2

CREDIT MIGRATION MATRIX IN DISCRETE AND CONTINUOUS TIME

2.1 Markov chain

In this chapter we are going to present a brief introduction to Markov chains. Markov chains were introduced by Andrey Markov in 1906. The Markov chain is a stochastic tool describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. Roughly speaking a stochastic process satisfies the Markov property if we can make predictions for the future based on the present state, ignoring any other information from the past. So, conditional on the present state of the system, its future and past states are independent. The books and papers that are used in this chapter are from O.Chrysafinou, R.Gunnvald, K.Chung, M.Koutras.

Nowadays the Markov property is used in a variety of applications such as credit risk, exchange rates of currencies, queues or lines of consumers arriving in a bank, population growth in biology, wifi connection, gambling, sampling (Monte carlo Markov chain) ect. Markov chain can be defined either in discrete or in continuous time. Now we are ready to give a formal definition for the Markov chain.

Discrete time Markov chain: A discrete Markov chain is a sequence of random variables $X_0, X_1, ...$ with the property, that the probability of moving to the next state depends only on the present state and not on the past states. Namely

P(X_n =j|X_{n-1} =i_{n-1},X_{n-2} =i_{n-2},....,X_0 =i_0)= P(X_n =j|X_{n-1} =i_{n-1})

The possible values of X_i from a countable set S called the state space of the chain which can be either discrete or continuous.
Moreover it is very important to point out that Markov chain can be combined with many useful properties such as time homogeneity. By that we imply that the probability of transition from state i to state j is independent of n.

P(X_n =j|X_{n-1} =i)=P(X_{n-1} =j|X_{n-2} =i)= ... $= P(X_1 = j|X_0 = i)$.

From now on we define $p_{ij}^{(n)}$=$P(X_n = j|X_0 = i)$ as the n-step transition probability. For n=1 we write p_{ij}=$p(X_1 = j|X_0 = i)$.
If time homogeneity is valid then we can write that:

p_{ij}=$P(X_1 = j|X_0 = i) = ... = P(X_n = j|X_{n-1} = i)$

2.2 Rating migration matrix definition

The rating migration matrices contain the probability that an entity(individual/company/country) will migrate from one rating scale to another in a specific time period. Banks and markets are using these matrices in order to adjust interest rates in loans and bonds. Different credit rating agencies use different rating scales. For the purposes of this thesis we are going to use the following rating scale: AAA, AA, A, BBB, BB, B, CCC, CC, D where D is the default state. The default state can be either absorbing, meaning that we can not recover again if we enter there or it can be with recovery, meaning that if we enter to this state the chain can exit the particular rating. Most of our examples are constructed under the absorbing state assumption. No matter whether the default state is absorbing or not the methodology we are going to follow is identical.

Now we are ready to define the first step transition matrix. A transition matrix P has the following form:

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

where p_{ij} are the first step transition probabilities for all i=*1,2,..N* and all j=*1,2,..N*
The transition matrix has the following properties

1. $\sum_{j=1}^{N} p_{ij}$=1 for all i=*1,2,..N*
2. $p_{ij} \geq 0$ for all i=*1,2,..N* and all j=*1,2,..N*

If $P^{(n)}$ is the n-step transition probability matrix of a time homogeneous discrete Markov chain. Then:

1. $P^{(n+m)} = P^{(n)} P^{(m)}$ (2.1)
2. $P^{(n)} = P^n$ (2.2)

Example 2.1 Suppose we have the next first step transition matrix with three states A,B,C where C is the default absorving state

$$P = \begin{pmatrix} 97.5\% & 1.5\% & 1\% \\ 12.5\% & 81.5\% & 6\% \\ 0\% & 0\% & 100\% \end{pmatrix}$$

Then the second step transition matrix is

$$P^2 = \begin{pmatrix} 95.25\% & 2.69\% & 2.1\% \\ 22.38\% & 66.61\% & 11.923\% \\ 0\% & 0\% & 100\% \end{pmatrix}$$

The fifth step transition matrix is

$$P^5 = \begin{pmatrix} 89.59\% & 4.91\% & 5.56\% \\ 40.93\% & 37.19\% & 23.61\% \\ 0\% & 0\% & 100\% \end{pmatrix}$$

The 10-step transition matrix is

$$P^{10} = \begin{pmatrix} 82.26\% & 6.23\% & 11.70\% \\ 51.89\% & 15.84\% & 34.67\% \\ 0\% & 0\% & 100\% \end{pmatrix}$$

Finally the 100-step transition matrix is

$$P^{100} = \begin{pmatrix} 22.9\% & 2.00\% & 75.10\% \\ 16.70\% & 1.47\% & 81.80\% \\ 0\% & 0\% & 100\% \end{pmatrix}$$

All the above matrix calculations have been performed in python. The code is given below:

```
import numpy as np
from numpy.linag import matrix_power
P=np.array([97.5%,1.5%,1%],[12.5%,81.5%,6%],[0%,0%,100%])
P2=matrix_power(P, 2)
P5=matrix_power(P, 5)
P10=matrix_power(P, 10)
P100=matrix_power(P, 100)
```

It is also very important not only to construct transition matrices but also extract useful information from them. For example we can see from the 5 year transition matrix to the 10 years transition matrix that the firms in category B are moving with high probability either in category A either in category C, due to the fact that transition probability from state B to state B has decreased from year five to year 10. This information gives us the feeling that the economy will face a big future change in which companies will either default or will grow very strong. That can happen for many reasons(for example some type of industries might get stronger and others not). Moreover we can observe from the matrix P^{100} that as we move in time the probability of default increases. The particular phenomenon occurs to all transition matrices and it can be explained simply as the fact that in bigger time lengths there is a high number of default trigger events, resulting in higher probability of default. The big question that raises also is whether the probability of default can reach to one, as time passes. This matter is going to be discussed in detail in chapter 4.

2.3 Transition matrices in continuous time

Due to the fact that the economy is facing a lot of challenges nowadays there is a need to construct transition matrices in shorter time horizons than one year. For this reason we would wish to construct matrices in continuous time. This leads to the need of constructing Markov chains in continuous time.
Let X_t be the random variable describing the state of the process at time t and assume the process is in state i at time t. Then if $X_t = i$, X_{t+h} is independent of the previous values (X_s:s<t) and as h→0 ,

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

where δ_{ij} is the Kronecker delta, o(h) is a function of h so as $\lim_{h\to 0} \frac{o(h)}{h} = 0$ and q_{ij} indicates how quick the transition from i to j happens.

Moreover if

$$P(X_t = j|X_s = i) = p(X_{t+h} = j|X_{s+h} = i)$$

for every t,s then the Markov chain is called time homogeneous.
For a continuous time Markov chain with the time homogeneity property we will be using the notation $P(X_t = j|X_0 = i) = p_{ij}(t)$

Just like the discrete transition matrices we can construct matrices in continuous time. This is very helpful especially when we want to monitor and update our predictions for credit risk very often in non integer time periods. So as follows using $p_{ij}(t)$ we can construct the transition matrix $P^{(t)}$,t>0

$$P^{(t)} = \begin{pmatrix} p_{11}(t) & p_{12}(t) & \dots & P_{1N}(t) \\ p_{21}(t) & p_{22}(t) & \dots & p_{2N}(t) \\ \dots & \dots & \dots & \dots \\ p_{N1}(t) & p_{N2}(t) & \dots & P_{NN}(t) \end{pmatrix}$$

Now the main problem is how are we going to calculate $p_{ij}(t)$. If the time period we want to calculate is an integer number then we have no problem calculating $p_{ij}(t)$ due to the fact that this probability can be calculated just by finding P^t. But that can not be applied for non-integer t>0. The solution to this problem is to use a matrix called generator matrix.
The generator matrix is a matrix which, instead of having probabilities as entries, it contains elements that represent how fast we move in time from one state to another. Below we can see a matrix.

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

Q is a NxN matrix. Moreover Q is a generator matrix if and only if $q_{ij}=\lim_{h\to 0}\frac{p_{ij}(h)-0}{h}$ for all i≠j and $q_{ii}=\lim_{h\to 0}\frac{p_{ii}(h)-1}{h}$ where $p_{ij}(t)$ is the transition probability from state i to j during time t.
The quantityq_{ij} represents the rate a continuous time Markov chain moves from state i to state j.

properties of the generator matrix

1. $q_{ij}\geq 0$ for all $i\neq j$
2. $q_{ii} \leq 0$
3. $\sum_{j=1}^{N} q_{ij} = 0$ for all i

Although we defined matrix Q it is not clear yet how are we going to calculate q_{ij}. In order to calculate the generator matrix we are going to need some additional useful quantities such as the exponential matrix, the logarithm matrix and the matrix norm.

A matrix norm (denoted as $||\ \ ||$) is a function $||\ \ ||$:$C^{mxn} \to$ R (where C^{mxn} is the vector space of all matrices of size mxn) that satisfies the following properties: For all matrices A and B in C^{mxn}

- $|| A ||\geq 0$

- $\| A \| = 0$ if and only if $A = 0_{m,n}$
- $\| aA \| = |a| \| A \|$ (absolutely homogeneous)
- $\| A + B \| \leq \| A \| + \| B \|$ (triangle inequality)

So a matrix norm is a vector norm whose elements represent the distance between a given matrix and the zero vector.

The calculation of the matrix norm is the follwing: Let A be an NxN matrix with entries a_{ij} *i,j=1,2,...,N.*
Then $\| A \| = max(\sum_{j=1}^{N} |a_{ij}|)$

Matrix norm is going to help us define the exponetial matrix and the logarithm matrix.When we are dealing with real numbers we know that

- exp(x)=$\sum_{k=0}^{\infty} \frac{x^k}{k!}$
- log(x)=$\sum_{k=0}^{\infty} \frac{(-1)^k (x-1)^k}{k!}$.

Making the above expansions we can define the same quantities for matrices. Namely for a square matrix Q , we define:

- exp(Q)=$\sum_{k=0}^{\infty} \frac{Q^k}{k!}$
- log(Q)=$\sum_{k=0}^{\infty} \frac{(-1)^k (Q-I)^k}{k!}$, $||Q - I||$<1

From all the above we can calculate $P^{(t)}$ by using the next formula

$$P^{(t)}=\exp(tQ) \quad (2.8)$$

where $Q = log(P)$ is a generator matrix

Also we have to mention that for t=1, we have $P^{(1)}$=P=e^Q and therefore $log(P) = Q$

Example 2.2

In the following example we are going to calculate a 6 month transition matrix using formula (2.8). After that we are going to observe that the particular relationship does not give very accurate results and for this reason we will propose an algorithm in order to make our results applicable. Let as start with matrix 1 as a first order transition matrix.

Matrix 1 Average first order transition matrix from Standar and Poors(in %) from 1930 to 2006

	1	2	3	4	5	6	7	8	D
1	91.1200	7.8020	0.8779	0.1743	0.00251	0.0000	0.0000	0.0000	0.0000
2	1.3430	90.7400	6.8850	0.7316	0.1864	0.394	0.0021	0.0043	0.0671
3	0.0859	3.1110	90.230	5.6180	0.7349	0.1145	0.0202	0.0085	0.0806
4	0.454	0.3170	4.9960	87.7800	5.5250	0.8395	0.1623	0.0173	0.3170
5	0.0079	0.0921	0.5346	6.6460	82.7100	7.8360	0.6256	0.0573	1.4870
6	0.0080	0.0613	0.1965	0.7155	7.1460	81.1600	5.6910	0.5702	4.4490
7	0.0000	0.0317	0.0418	0.244	1.0240	10.0800	70.9900	4.0120	13.5700
8	0.0000	0.0000	0.1338	0.0000	0.5466	3.7900	8.8770	63.7900	22.91
D	0.0000	0.0000	0.0000	0.000	0.0000	0.0000	0.0000	0.0000	1

$\| P - I \| \leq 1$

Next we calculate the generator matrix Q=log(P) by using the first 100 terms of the series . This way we obtain the next approximation of matrix Q

	1	2	3	4	5	6	7	8	D
1	–9.3630	8.5740	0.6387	0.1426	0.0132	–0.0022	–0.0002	–0.0003	–0.0033
2	1.4750	–9.9110	7.5980	0.5729	0.1653	0.0312	–0.0008	0.0050	0.0642
3	0.0680	3.4320	–10.5900	6.2890	0.6411	0.0739	0.0135	0.0095	0.0657
4	0.0463	0.2540	5.6010	–13.4600	6.4520	0.6779	0.1563	0.0137	0.2605
5	0.0059	0.0838	0.3879	7.7820	–19.6500	9.5460	0.4454	0.0295	1.3730
6	0.0083	0.0618	0.1888	0.4980	8.7120	–21.7800	7.4710	0.5705	4.2700
7	–0.0009	0.0340	0.0233	0.2322	0.7298	13.0600	–35.1500	5.9400	15.02
8	–0.0002	–0.0071	0.1691	–0.0595	0.4852	4.3130	13.0600	–45.3700	27.42
D	0	0	0	0	0	0	0	0	0

The calculation of the generator matrix has been perfomed using Python. Below we present the related code:

```
import numpy as np
from scipy.linalg import logm
P=np.array([91.1200,7.8020, 0.8779,0.1743,0.00251,0,0,0,0],
  [1.3430,90.7400,6.8850,0.7316,0.1864,0.394,0.0021,0.0043,
0.0671],
  [0.0859,3.1110,90.230, 5.6180,0.7349,0.1145,0.0202,0.0085,0.0806],
  [0.454,0.3170,4.9960,87.7800,5.5250,0.8395,0.1623,0.0173,0.3170],
  [0.0079,0.0921,0.5346,6.6460,82.7100,7.8360,0.6256,0.0573,1.4870],
  [0.0080,0.0613,0.1965,0.7155,7.1460,81.1600,5.6910, 0.5702,4.4490],
  [0.0000,0.0317,0.0418,0.244,1.0240,10.0800,70.9900,4.0120,13.5700],
  [0.0000,0.0000,0.1338,0.0000,0.5466,3.7900,8.8770,63.7900,22.91],
```

```
   [0.0000,0.0000,0.0000,0.000, 0.0000,0.0000,0.0000,0.0000,1])
Q=scipy.linalg.logm(P)
```

Now by using the equation (2.8) we are able to calculate transition matrices in continuous time.Just for illustrating the method we are going to calculate a 6 month transition matrix. Applying (2.8) we may write $P^{(0.5)}{=}e^{0.5Q}$ and the right hand side provides the following result

Matrix 2 6 month transition matrix

	1	2	3	4	5	6	7	8	D
1	95.4400	4.0890	0.3824	0.0789	0.0096	–0.0003	–0.0001	–0.0001	–0.0008
2	0.7036	95.2100	3.6150	0.3286	0.0878	0.0176	0.0004	0.0023	0.0328
3	0.0387	1.6330	94.9100	2.9710	0.3462	0.0472	0.0084	0.0045	0.0365
4	0.0229	0.1438	2.6440	93.5900	2.9830	0.3840	0.0793	0.0078	0.1444
5	0.0035	0.0439	0.2336	3.5930	90.7900	4.3200	0.2750	0.0222	0.7154
6	0.0040	0.0307	0.0959	0.3091	3.9410	89.8900	3.2560	0.2887	2.1860
7	–0.0002	0.0163	0.0166	0.1182	0.4503	5.7570	84.0700	2.4390	7.1340
8	–0.0001	–0.0016	0.0749	–0.0139	0.2598	2.0130	5.3770	79.7800	12.5100
D	0	0	0	0	0	0	0	0	1

The python code for matrix 2 is given below:

```
import numpy as np
from scipy import linalg
P0.5=linalg.expm(0.5*Q)
```

Looking at the last matrix we can see that the 6 month transition matrix does not fulfill all the properties of a transition matrix due to the fact that it contains negative probabilities(although they are very close to zero). Moreover, apart from that, there is another reason why the 6 month transition matrix needs a correction. In the annual transition matrix we may see that several transition probabilities are equal to zero like for example from rating grade 1 to rating grade 8 and vice versa. This means that we can not move from grade 1 to grade 8 or from grade 8 to 1 in a year. Apparently in the 6 month transition matrix, the same should apply, because if we can not move from grade 1 to 8 in one year then we should not be able to move from 1 to 8 in 6 months as well. On the other hand, if there is a positive probability moving from one rating grade to another in a year it does not necessarily means that the same should apply for the 6 month matrix because of the fact that the transition can not happen in the first semester but can happen in the second. As a conclusion if something happens in 6 months it could also happen in a year but the opposite is not always true. An idea for resolving this problem was proposed by Kreinin and Sidelnikova (2001) using a regularization algorithm. This algorithm has the following 2 steps:

1. replace all the negative elements by zero
2. all the zero elements remain zero
3. to all non-negative elements of the same row add the quantity $[b_i p_{ij}^{(t)}]/a_i$ where $p_{ij}^{(t)}$ is the (i,j) element of the matrix $P^{(t)}$, b_i is the sum of the negative elemenst of row i in matrix $P^{(t)}$ and a_i is the sum of the positive elements in the same row

With the help of the algorithm we now have the next 6 month transition matrix

Matrix 3 regulized 6 month transition matrix

	1	2	3	4	5	6	7	8	D
1	95.42759	4.088468	0.38395	0.078889743	0.0095987	0	0	0	0
2	0.7036	95.2100	3.6150	0.3286	0.0878	0.0176	0.0004	0.0023	0.0328
3	0.0387	1.6330	94.9100	2.9710	0.3462	0.0472	0.0084	0.0045	0.0365
4	0.0229	0.1438	2.6440	93.5900	2.9830	0.3840	0.0793	0.0078	0.1444
5	0.0035	0.0439	0.2336	3.5930	90.7900	4.3200	0.2750	0.0222	0.7154
6	0.0040	0.0307	0.0959	0.3091	3.9410	89.8900	3.2560	0.2887	2.1860
7	0	0.016299	0.016599	0.118199	0.450299	5.75698	84.06983	2.43899	7.13398
8	0	0	0.07489	0	0.25976	2.0127	5.3761	79.787	12.508
D	0	0	0	0	0	0	0	0	1

This matrix of course may not fulfill relationship (2.8) but on the other hand it is a matrix with no significant difference from the original one. Also this matrix is not a transition matrix anymore because $\sum_{j=1}^{N} p_{ij} \neq 1$

Summarizing we give again the two 6 month matrices for comparison so as we can observe there is not a significant difference between the 6 month transition matrix and the regulized 6 month transition matrix

1	–9.3630	8.5740	0.6387	0.1426	0.0132	–0.0022	–0.0002	–0.0003	–0.0033
2	1.4750	–9.9110	7.5980	0.5729	0.1653	0.0312	–0.0008	0.0050	0.0642
3	0.0680	3.4320	–10.5900	6.2890	0.6411	0.0739	0.0135	0.0095	0.0657
4	0.0463	0.2540	5.6010	–13.4600	6.4520	0.6779	0.1563	0.0137	0.2605
5	0.0059	0.0838	0.3879	7.7820	–19.6500	9.5460	0.4454	0.0295	1.3730
6	0.0083	0.0618	0.1888	0.4980	8.7120	–21.7800	7.4710	0.5705	4.2700
7	–0.0009	0.0340	0.0233	0.2322	0.7298	13.0600	–35.1500	5.9400	15.02
8	–0.0002	–0.0071	0.1691	–0.0595	0.4852	4.3130	13.0600	–45.3700	27.42
D	0	0	0	0	0	0	0	0	0

1	95.42759	4.088468	0.38395	0.078889743	0.0095987	0	0	0	0
2	0.7036	95.2100	3.6150	0.3286	0.0878	0.0176	0.0004	0.0023	0.0328
3	0.0387	1.6330	94.9100	2.9710	0.3462	0.0472	0.0084	0.0045	0.0365
4	0.0229	0.1438	2.6440	93.5900	2.9830	0.3840	0.0793	0.0078	0.1444
5	0.0035	0.0439	0.2336	3.5930	90.7900	4.3200	0.2750	0.0222	0.7154
6	0.0040	0.0307	0.0959	0.3091	3.9410	89.8900	3.2560	0.2887	2.1860
7	0	0.016299	0.016599	0.118199	0.450299	5.75698	84.06983	2.43899	7.13398
8	0	0	0.07489	0	0.25976	2.0127	5.3761	79.787	12.508
D	0	0	0	0	0	0	0	0	1

In addition we can also compare the annual matrix with the regulized 6 month matrix

$$\begin{pmatrix}
91.1200 & 7.8020 & 0.8779 & 0.1743 & 0.00251 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
1.3430 & 90.7400 & 6.8850 & 0.7316 & 0.1864 & 0.394 & 0.0021 & 0.0043 & 0.0671 \\
0.0859 & 3.1110 & 90.230 & 5.6180 & 0.7349 & 0.1145 & 0.0202 & 0.0085 & 0.0806 \\
0.454 & 0.3170 & 4.9960 & 87.7800 & 5.5250 & 0.8395 & 0.1623 & 0.0173 & 0.3170 \\
0.0079 & 0.0921 & 0.5346 & 6.6460 & 82.7100 & 7.8360 & 0.6256 & 0.0573 & 1.4870 \\
0.0080 & 0.0613 & 0.1965 & 0.7155 & 7.1460 & 81.1600 & 5.6910 & 0.5702 & 4.4490 \\
0.0000 & 0.0317 & 0.0418 & 0.244 & 1.0240 & 10.0800 & 70.9900 & 4.0120 & 13.5700 \\
0.0000 & 0.0000 & 0.1338 & 0.0000 & 0.5466 & 3.7900 & 8.8770 & 63.7900 & 22.91 \\
0.0000 & 0.0000 & 0.0000 & 0.000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1
\end{pmatrix}$$

$$\begin{pmatrix}
95.42759 & 4.088468 & 0.38395 & 0.078889743 & 0.0095987 & 0 & 0 & 0 & 0 \\
0.7036 & 95.2100 & 3.6150 & 0.3286 & 0.0878 & 0.0176 & 0.0004 & 0.0023 & 0.0328 \\
0.0387 & 1.6330 & 94.9100 & 2.9710 & 0.3462 & 0.0472 & 0.0084 & 0.0045 & 0.0365 \\
0.0229 & 0.1438 & 2.6440 & 93.5900 & 2.9830 & 0.3840 & 0.0793 & 0.0078 & 0.1444 \\
0.0035 & 0.0439 & 0.2336 & 3.5930 & 90.7900 & 4.3200 & 0.2750 & 0.0222 & 0.7154 \\
0.0040 & 0.0307 & 0.0959 & 0.3091 & 3.9410 & 89.8900 & 3.2560 & 0.2887 & 2.1860 \\
0 & 0.016299 & 0.016599 & 0.118199 & 0.450299 & 5.75698 & 84.06983 & 2.43899 & 7.13398 \\
0 & 0 & 0.07489 & 0 & 0.25976 & 2.0127 & 5.3761 & 79.787 & 12.508 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

In matrix 3 we can see that the probability of default for each grade is lower in the 6 month matrix than in the annual one.

As a conclusion from chapter 2 we can state that the construction of the first order transition matrix is crucial to our analysis. Having this matrix we can calculate the generator matrix from which we can calculate continuous transition probabiliies. In chapter 3 we are going to focus on multiple ways we can construct first order transition matrix as well as testing if time homogeneity assumption holds for a variety of transition matrices.

Chapter 3

CONSTRUCTION OF TRANSITION MATRICES AND TIME HOMOGENEITY TESTING

There are different methods of estimating the entries of a transition matrix. The two most commonly used are the so called cohort (discrete time) and duration (continuous time) methods.We shall next present these two methods. The books and papers that are use for this chapter include the following authors: Y.Jafry- T.Schuermann, D.Lando, P.Lencastre-P.Lind- F.Raischel- T.Rogers, K. Papaioannou, T. Ferendinos.

3.1 cohort method

Cohort is a very popular method of estimating transition probabilities due to it's simplicity. The cohort method has been widely used as it applies simple calculations, although sometimes the results are less efficient. It is not an advanced stochastic method, instead it simply uses frequencies for estimating transition probabilities. Moreover the particular method is distribution free and calculations can be made for any discrete time horizon.

Let T be a specific time period. Then the estimator of $p_{ij}(T)$ is given by the expression :

$$p^*_{ij}(T) = \frac{n_{ij}(T)}{n_i(T)} \quad \textbf{(3.1)}$$

where

- $n_{ij}(T)$ is the number of entities that moved from rating i to rating j at the end of period T
- $n_i(T)$ is the number of entities that where in rating i at the beginning of the period T

When applying the cohort method , one has to know that there are certain advantages and disadvantages.

Advantages of the cohort method

1. Simple method for calculations. The Cohort method is just a simple ratio method and for that reason there is no need for complicated calculations, like using stochastic processes and probability theory.
2. No time homogeneity assumption needed. So we can always proceed to the estimation results no matter whether there is time homogeneity or not.
3. It calculates directly the first order transition matrix without having to calculate the generator matrix first.

Disadvantages of the cohort method

1. One drawback of the cohort method is that the estimators assign zero probability to an event if there are no records of such an event in the data. This makes the estimators poor in capturing rare events such as moving from very high rating grades to very low rating grades
2. The cohort method uses the grades at the beginning and at the end(discrete time method) of the time period. That means that the model does not take into consideration the total migrations happening to each company in all the duration of time period, a fact that can affect the model's accuracy.
3. We can not calculate transition matrices in continuous time, as this method applies for discrete time periods.

4. Sometimes data are not in an accessible form. By that we mean that there is not a rating from the beginning of the period a company or client. In that case the observation can not be used by this model

3.2 Duration method

In the duration method, instead of estimating the transition matrix, we try to estimate the generator matrix Q. The basic assumption here is that there is time homogeneity. The estimator of $q_{ij}(T)$ is:

$$q^*_{ij}(T)=\frac{m_{ij}(T)}{\int_0^T Y_i(s)ds} \quad \textbf{(3.2)}$$

where

- $m_{ij}(T)$ is the number of transitions from grade i to j during period T
- $Y_i(s)$ number of firms with rate i at time s

Just like the cohort method, the duration method has some benefits and drawbacks.

Advantages of the duration method

1. Solves the zero probability problem for rare events that occurs in the cohort method. As a consequence the estimations for the first step transition matrix become more realistic.
2. Estimates the generator matrix Q directly without calculating the series expansion of log making this way easy calculation for any transition matrix.
3. It takes into account all the transitions made during the period, extracting more realistic results.
4. The duration method is a continuous time method. So if a company has a grade during time period T but not at the beginning of the period it still can be used in our model. In other words we do not need to have a rating for all entities at time t=0 as far as we have some rating until period T.

Disadvantages of the duration method

1. Time homogeneity assumption required. This means that if time homogeneity does not hold the method becomes useless and can not be used.
2. If we are talking about the annual transition matrix, the cohort method is faster because it estimates directly the matrix and the duration method needs to calculate e^Q

3.3 Time homogeneity testing

As already mentioned the duration method works only if time homogeneity exists. Moreover all the results from chapter 2 work also under the time homogeneity assumption. For this reason it is crucial to test whether time homogeneity is justified by our data. In the next section we are going to present 3 methods for testing time homogeneity. The chi-square test, the coefficient intervals method and the metrics method.

3.3.1 coefficient intervals

The basic idea in this method is that by spiting time period T into smaller equal time periods T_1<T_2<........<T, and by calculating $p_{ij}(T_k)$ for every k=1,2,... and constructing a confidence interval, then we can check if these probabilities belong to the confidence interval. If they do, then we have time homogeneity, because the transition probabilities in equal time periods have no statistically important difference.
For example if T=12 years then we can construct 12 different time periods with length one year each. Then we calculate p_{ij} for each period using the cohort method and compare these probabilities with the coefficient interval's we have constructed. The coefficient intervals are made by using our data under the assumption that the time homogeneity assumption is valid. Two method exist for constructing the confidence interval, the historical simulation method and the binomial method.

a. Historical simulation or bootstrap method

In this method we construct coefficient intervals without any distributional assumption. The empirical distribution constructed by the use of observed values serves as an approximation of the true distribution, from which values are drawn with replacement. The bootstrapping technique allows for the estimation of the accuracy of some distribution parameter, such as the mean. This can then be used to calculate confidence intervals.
The standard bootstrapping procedure is the one used in this thesis to estimate confidence intervals for every p_{ij}. Consider having a sample of k observations, which represent the various rating transitions within a specific time period. Then, out of the original sample, observations are drawn with replacement one at a time to construct a new sub-sample of size n. The new sample gives an estimate of the p_{ij} using either cohort or duration method. Then this procedure is repeated N times to get N estimates of the PD. These N values now form an estimate of the p_{ij} distribution. Constructing a (1-a)%=95% two-sided symmetric confidence interval out of this distribution is done by simply ordering the values from the lowest to the highest and choosing the $\frac{a}{2}$ percentile and the 1-$\frac{a}{2}$ percentile.

b. Binomial method

In this method we construct coefficient intervals of every transition probability of the transition matrix based on binomial distribution. After that we shift from the binomial to the normal distribution using the central limit theorem.
Due to the fact that the transition probabilities can be written as sum of other random variables, if p_{ij} is the first step transition probability from i to j then:

$$p_{ij}=\frac{X_1+X_2+...X_n}{n},$$

where

$$X_k=\begin{Bmatrix} 1, \ if\ the\ k-th\ company\ with\ rating\ grade\ i\ moves\ to\ j\ in\ one\ year \\ 0, \ if\ the\ k-th\ company\ with\ grade\ i\ doesn't\ move\ to\ j\ in\ one\ year \end{Bmatrix}$$

Then X_k~Bernouli(p^*_{ij}) and assuming that $X_1, X_2, ..., Xn$ are independent random variables and for "large" n, we obtain by the central limit theorem

p_{ij}~N(p^*_{ij},$\frac{p^*_{ij}(1-p^*_{ij})}{n}$). Therefore we can construct confidence intervals for all p_{ij} $i,j=1,2,3,\ldots N$

Taking many different samples or by using Monte Carlo simulation we may create a vector with multiple p_{ij}. Then the confident interval is

$$(p_{ijavg}\text{-}z_{a/2}\text{S},p_{ijavg}\text{-}z_{a/2}\text{S}) \quad \textbf{(3.3)}$$

where

- p_{ijavg} is the average value of all different p_{ij}
- S is the standard deviation of the observations
- $z_{a/2}$ is the percentile of the standard normal distribution.

The matrices below show annual transition matrices using the cohort and the duration method respectively.

Example 3.1

97.35	1.47	0.19	0.3	0.58	0.05	0.03	0.02
2.98	82.56	12.59	0.91	0.91	0.01	0	0.04
0.15	3.37	83.95	10.45	1.85	0.17	0.01	0.05
0.08	0.36	7.36	80.66	10.17	0.97	0.09	0.3
0.18	0.11	1.08	13.38	79.13	4.42	0.69	1.01
0.11	0.05	0.33	3.11	15.78	73.27	3.46	3.88
0.04	0.06	0.34	1.22	5.31	11.8	73.81	7.35
0	0	0	0	0	0	0	1

Matrix 1 Annual transition matrix calculated by cohort method

97.02	1.4	0.27	0.49	0.05	0.14	0.02	0.02
2.49	81.24	13.31	1.63	1.18	0.08	0.03	0.04
0.18	3.41	83.37	10.28	2.33	0.29	0.05	0.09
0.1	0.56	7.12	80.88	9.76	1.11	0.16	0.31
0.16	0.31	1.67	12.31	79.22	4.62	0.71	1.01
0.14	0.13	0.63	3.92	14.54	71.03	4.31	5.03
0.06	0.19	0.43	1.91	5.95	10.022	70.54	10.71
0	0	0	0	0	0	0	1

Matrix 2 Annual transition matrix calculated by duration method

Now having two transition matrices one calculated by the cohort method and one using the duration method we are going to test time homogeneity with the binomial approach. So what we will do

is using the probabilities from the duration method(it can be done also with probabilities from the cohort method), we will construct 95% interval coefficients for each p_{ij}. Then we are going to test if the probabilities from the cohort transition matrix belong to these intervals. Due to the fact that we do not have any real data available we are going to generate random numbers from the normal distribution because from the binomial model we know that p_{ij} follow the normal distribution. Also this is the advantage of the binomial method whereas in the historical simulation we need real data. Because the whole process can take long time, due to the fact that we have to calculate coefficient intervals for every rating scale, we are going to perform the test only for the default state.
Using the Anaconda Python program and by simulating 1 million random numbers from the normal distribution with parameters different from each p_{iD} and using a sample of 50.000 firms we have the following results(in %):

$-$	$default$	p_{avg}	S	$minimal$	$maximum$
1	0.02	0.02000003	$4.01 * 10^{-5}$	0.0806211	0.0993774
2	0.04	0.0399998	$7.98 * 10^{-5}$	0.03994421	0.04005776
3	0.09	0.0899994	0.1342003	0.08062126	0.09993774
4	0.31	0.3100186	0.02487375	0.2926443	0.323939
5	1.01	1.010039	0.04967205	0.9788353	1.041242
6	5.03	5.029985	0.009558851	5.023308	5.03662
7	10.71	10.71001	0.01914027	10.69664	10.72338

The above calculations, have been performed using Anaconda Python. The code is given below:

```
import numpy as np
from numpy.random import randn
import scipy.stats
probability_default=np.array([0.0002,0.0004,0,0009,0,0031,0.0101,0.0503,0,1071])
p_avg=[0]*7
S=[0]*7
lowerl_bound[i]]=[0]*7
upper_bound[i]=[0]*7
n=50000
z_0.025=scipy.stats.norm.sf(0.025,0,1)
```

```
for i in range(0,6,1):
p_avg[i]=(np.sqrt((probability_default[i]*(1-probability_default[i]))/n)*randn(10000)
+probability_default[i]).mean()
S[i]=(np.sqrt((probability_default[i]*(1-probability_default[i]))/n)*randn(10000)
+probability_default[i]).std()
lowerl_bound[i]=p_avg[i]-S[i]*z_0.025
upper_bound[i]=p_avg[i]+S[i]* z_0.025
```

Comparing now this with the entries of the cohort matrix we can see that time homogeneity exists only in rating $1D, 2D, 4D, 5D$. So the Markov chain does not have time homogeneity. Repeating the same process for each p_{ij} we give again the cohort matrix pointing out in which cells time homogeneity does not exist.

$$\begin{pmatrix}
97.35 & 1.47 & \mathbf{0.19} & 0.3 & 0.58 & 0.05 & 0.03 & 0.02 \\
\mathbf{2.98} & \mathbf{82.56} & \mathbf{12.59} & \mathbf{0.91} & \mathbf{0.91} & \mathbf{0.01} & \mathbf{0} & 0.04 \\
0.15 & 3.37 & 83.95 & 10.45 & \mathbf{1.85} & \mathbf{0.17} & \mathbf{0.01} & \mathbf{0.05} \\
0.08 & \mathbf{0.36} & \mathbf{7.36} & \mathbf{80.66} & \mathbf{10.17} & \mathbf{0.97} & \mathbf{0.09} & 0.3 \\
0.18 & \mathbf{0.11} & \mathbf{1.08} & \mathbf{13.38} & 79.13 & \mathbf{4.42} & 0.69 & 1.01 \\
0.11 & \mathbf{0.05} & \mathbf{0.33} & \mathbf{3.11} & \mathbf{15.78} & \mathbf{73.27} & \mathbf{3.46} & \mathbf{3.88} \\
0.04 & \mathbf{0.06} & \mathbf{0.34} & \mathbf{1.22} & 5.31 & \mathbf{11.8} & \mathbf{73.81} & \mathbf{7.35} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Matrix 3 Entries in bold show as the ratings where there is no time homogeneity in those transitions

The only model weak point of this method is that we consider X_i as independent variables something that is not always true, because default events are triggered not only for a company's unique characteristics but from market conditions and macroeconomic factors as well.

3.3.2 The Chi-square test

The basic idea of this test is to see if the transition probabilities are constant. The full sample is divided into equal length smaller sub samples. We now wish to check if the migration matrix calculated by the sub samples has any statistical difference from the transition matrix calculated using the whole sample.
Let $p^*_{ij}(t)$ be the transition probability from rating i to j during year

t and p^*_{ij} the transition probability calculated using the whole sample. Moreover let $n_i(t)$ be the number of firms with rating i during time t. Our null hypothesis is

$$H_0{:}p^*_{ij}(t) = p^*_{ij} \qquad H_1{:}p^*_{ij}(t) \neq p^*_{ij}$$

and the test statistic is:

$$X^2 = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} n_i(t)\frac{(p^*_{ij}(t)-p^*_{ij})^2}{p^*_{ij}} \quad \textbf{(3.4)}$$

where X^2 follows the x^2 distribution with $N(N-1)(T-1)$ degrees of freedom. In our case we have 9 states, so N=9.
If $X^2 > X^2_{72}$ then we reject the null hypothesis of time homogeneity

3.3.3 The metrics method

Another method to test time homogeneity is by using a metric. The metric is a function that has all the properties of distance. Time homogeneity is tested by measuring the distance between the cohort matrix and the duration matrix. In the case where this distance is zero or relatively close to zero there is evidence that time homogeneity exists
A metric on a set X is a function M:$X \times X$->[0,∞) so that, the following conditions are satisfied:

1. $M(x,y) \geq 0$
2. $M(x,y) = 0 \iff x = y$
3. $M(x,y) = M(y,x)$
4. $M(x,z) \leq M(x,y) + m(y,z)$

There are many metrics we can use for our analysis. Each one of them has its own benefits and drawbacks and it is up to the user to decide which one he/she will use.

3.3.4 The euclidean distance

One of the metric that we can use is the euclidean distance.With this metric the distance between two matrices P=(p_{ij}) and G=(g_{ij})of the same dimension can be defined as

$$M_{euc}(P,G) = \frac{\sqrt{\sum_{j=1}^{N}\sum_{i=1}^{N}(p_{ij}-g_{ij})^2}}{N^2} \qquad \textbf{(3.5)}$$

So what we need to do is to compare different transition matrices with each other and try to investigate which matrices have similar behaviour(which matrix is closer to another matrix).

Example 3.2

In this example we consider three different transition matrices P, G, K which they represent the migrations in consumer loans, credit cards, house loans,with 4 ratings A,B,C,D where D is the absorbing state

$$P = \begin{pmatrix} 0.9 & 0.06 & 0.035 & 0.0005 \\ 0.1 & 0.7 & 0.15 & 0.05 \\ 0.0003 & 0.047 & 0.65 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix} G = \begin{pmatrix} 0.91 & 0.05 & 0.025 & 0.015 \\ 0.1 & 0.8 & 0.05 & 0.05 \\ 0.0003 & 0.047 & 0.6 & 0.35 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$K = \begin{pmatrix} 0.9 & 0.04 & 0.03 & 0.03 \\ 0.1 & 0.7 & 0.06 & 0.04 \\ 0.0001 & 0.019 & 0.65 & 0.33 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Using (3.5) we have that $M_{euc}(P,G) = 0.00996$ and $M_{euc}(P,K) = 0.006537$

So matrix P is closer to matrix K than matrix G which means that consumer loans and credit cards have more similar behaviour than house loans within a year. That does not necessarily mean that the long term behaviour is similar. In order to say something like that we need to do the same analyses using the n-step transition matrices But someone in this stage might think the euclidean metric is the most accurate for our analysis or there is a more effective one?In order to answer to this question let's first make some observations. Firstly we can see that the euclidean metric "treats" all entries as having the same significance. In reality this is not very practical due to the fact that the off-diagonial entries play a more significant role in our analysis.

3.3.5 Introducing the M_I metric

Since the migration matrix, by definition, determines how a given state vector (or probability distribution) will migrate from one epoch to the next, a central characteristic of the matrix is the amount of migration (or "mobility") imposed on the state vector from one epoch to the next. We can highlight this characteristic by simply subtracting the identity matrix before proceeding with further manipulations. This apparently trivial observation turns out to be crucial . The identity matrix (of the same order as the state vector) corresponds to a static migration matrix, i.e. the state vector is unchanged by the action of the matrix from one epoch to the next. Subtracting the identity matrix from the migration matrix leaves only the dynamic part of the original matrix, which reflects the "magnitude" of the matrix in terms of the implied mobility.
Another metric we can use is the M_I metric which is a special case of the euclidean metric Let $I = (\delta_{ij})$ be the identity matrix and $P = (p_{ij})$ a NxN matrix. We define the metric $M_I(P)$ as follows :

$$M_I(P) = \frac{\sqrt{\sum_{j=1}^{N}\sum_{i=1}^{N}(p_{ij}-\delta_{ij})^2}}{N^2} \quad \textbf{(3.6)}$$

We can use this metric in order to count a matrix mobility. Matrix I is a matrix with all the properties of the transition matrices. Moreover the identity matrix has all the diagonal entries equal to one. That means that if we start at a specific rating we stay in the same rating for ever. So matrix I is a matrix which represents the perfect stability and therefore the "closer" the matrix is to the identity matrix the less are the transitions inside the matrix. On the other hand if the distance between a transition matrix and matrix I is large, then there is a lot of mobility in the matrix which means that things tend to change in our Markov chain.

Example 3.3

Suppose we have the next 3 transition matrices with two states $P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix}$ $G = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$ $K = \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}$

By using the euclidean metric we have that $M_{euc}(P,G) = M_{euc}(G,K)$

Now lets see how our new metric will work
$M_I(P) = 0.1838$ $M_I(G) = 0.3606$ $M_I(K) = 0.2062$
That means that matrices P and K are far more similarwith the perfect imobility matrix than matrix G. Moreover matrix G has a lot more mobility than the other two matrices.

Of course we have to mention that when we calculate the M_I metric we pay more attention in the off- diagonal elements but in reality each element and generally each row of our matrix does not give us the same amount of information. So it would be even better if we could give different weights in our rows. The amount of information that each row gives us can be assessed by matrix eigenvectors. For this reason we can use eigenvectors as weights.

3.3.6 the M_{SDV} metric

Suppose we have a transition matrix P and I is the identity matrix, then we denote the matrix $P_I = P - I$. We call matrix P_I as the mobility matrix of matrix P, due to the fact that this matrix is a measure of distance between matrix P and the identity matrix, which it represents the absence of mobility. So we want to create a metric which will count the mobility of a matrix. For this reason we define the $M_{SDV}(P)$ as follows:

$$M_{SDV}(P) = \frac{\sum_{i=1}^{N} \sqrt{\lambda_i} P_I P_I'}{N^2} \quad \textbf{(3.7)}$$

where λ_i $i = 1, 2, .., N$ are the eigenvalues of matrix P_I. Note that (3.7) applies only when $\lambda_i \geq 0$

So as we mentioned before this metric is a measure of mobility. But that is not very clear from the beginning when someone observes the particular metric and for that we are going to give more details about this metric. We can see that formula (3.7)involves a product between the mobility matrix and the transpose mobility matrix using also as weights the eigenvalues of matrix P. The multiplication of those matrices represents the distance between transition matrix and identity matrix, so they seem pretty reasonable. The only thing it is not yet clear is why we use the eigenvalues as weights. Generally eigenvalue of a vector is the number which when multiplied with the

vector gives us a new vector which will be just stretched and not rotated. In a transition matrix eigenvalues have also the physical meaning that the new rotated vectors will be uncorrelated. Because transition probabilities might be correlated with each other (for example transition probabilities of low rating firms are usually correlated) by multiplying with eigenvalues makes data uncorrelated. As a result we do not add information that already exists in our metric. M_{SDV} also has a set of properties that make our metric very "attractive"
Let P and G be tranistion matrices of the same dimension then:

1. $M_{SDV}(P) \geq 0$
2. $M_{SDV}(P) \geq M_{SDV}(G)$ if $p_{ij} \geq g_{ij}$ for all$i \neq j$ (Monotonicity)
3. If $p_{ii} = g_{ii}$ and $p_{ij} \neq g_{ij}$ for every $i \neq j$ then $M_{SDV}(P) \neq M_{SDV}(G)$ (distribution discriminatory)

3.3.7 Additional metrics

Some other metrics that can be used are the following:

1. $M_T(P) = \frac{1}{N-1}(N - tr(P))$
2. $M_D(P) = 1- \mid det(P) \mid$
3. $M_E(P) = \frac{1}{N-1}(N - \sum_{i=1}^{N} \mid \lambda_i(P) \mid)$, where λ_i are the eigenvalues of matrix P

Note that in the special case when all the eigenvalues of P are non-negative then $M_E(P) = M_T(P)$.

The reason why we use those metrics is due to their simplicity. More specifically note that $M_T(P)$ measures the distance between $N = tr(I)$ and tr(P). For this reason the particular metric indicates the tendency of the matrix to move to another rating. Keep in mind that the identity matrix is a transition matrix which indicates the absence of mobility in a Markov chain. So what we try to quantify is the mobility of a matrix and compare the results with other matrices. The disadvantage of this metric is that it gives a lot of significance in the diagonal entries. The problem with that is that in case two transition matrices have the same diagonal then the metric fails to measure the difference between them. The same applies for $M_E(P)$ with the only difference being that instead of the trace we use the sum of the eigenvalues which is pretty similar with the trace

of the matrix.
Last but not least the $M_D(P)$ metric gives another perspective into our analysis because it counts the distance between the determinant of the identity matrix and the determinant of the given transition matrix. In linear algebra the determinant can be viewed as the volume scaling factor of the linear transformation described by the matrix. That is a good measure of a matrix mobility due to the fact that we are comparing the volume scaling factor of the identity matrix and the given transition matrix P. On one one hand we solve the problem with the diagonal entries but from the other the metric starts to loose it's meaning when det(P)=0.

Moreover these metrics have a very useful property. If $M(P) \geq M(G)$ then $M(P^n) \geq M(G^n)$ (period consistency)
This means that if matrix P has greater mobility than matrix G then the same applies for their higher order matrices. This is a very convenient property because once we check the mobility of two matrices and find out that mobility of matrix P is greater than mobility of matrix G, then the same applies for all their higher order transition matrices.

Example 3.4

Now we are going to compare all the different metrics using some matrices as examples in order to see pros and cons of each metric and see in which cases it is wiser to use one instead of another.

Let us consider the following two first order transition matrices

$$P = \begin{pmatrix} 0.5 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.2 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.5 \end{pmatrix} G = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{pmatrix}$$

Notice that the two matrices have the same diagonal entries. Now we are going to calculate all the metrics for each matrix and compare the results

$M_{SDV}(P) = 0.50208 \qquad M_{SDV}(G) = 0.5785$
$M_I(P) = 0.5060 \qquad M_I(G) = 0.6325$

$M_T(P) = 0.625$ $\quad$ $M_T(G) = 0.625$
$M_E(P) = 0.625$ $\quad$ $M_E(G) = 0.625$
$M_D(P) = 0.9808$ $\quad$ $M_D(G) = 1$

M_{SDV} and M_I discriminate between P and G (with a larger value for the more extreme matrix,G) whereas M_T and M_E are "blind" to the variations in the distribution of the off-diagonal elements.M_D is different for each matrix but still very close to each other.

From the above examples, it is clear thatM_{SVD} and M_I are preferable to the others from the distribution discriminatory point of view. However, it is not immediately apparent which is preferable between M_{SVD} and M_I. To answer this, consider the following two matrices which differ only in the permutation of the non-diagonal entries within each row

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.7 & 0 \\ 0 & 0.4 & 0.7 \end{pmatrix} \; G = \begin{pmatrix} 0.8 & 0 & 0.2 \\ 0 & 0.7 & 0.3 \\ 0.4 & 0 & 0.6 \end{pmatrix}$$

Then:
$M_{SDV}(P) = 0.3463$ $\quad$ $M_{SDV}(G) = 0.34072$
$M_I(P) = 0.3590$ $\quad$ $M_I(G) = 0.3590$

Since M_{SVD} distinguishes between these matrices (satisfies distribution discriminant property) whereas M_I does not, we therefore prefer M_{SVD} over M_{euc} on the grounds that it satisfies distribution discrimination property more generally than does M_I

3.3.8 testing time homogeneity through the metric

We can use the M_{SDV} metric in order to test time homogeneity. What we actually are going to do is to separate our sample in sub samples (for example annual samples) and calculate for each year the transition matrices using both the homogeneity assumption and the inhomogeneity assumption with the duration and the cohort method. Then what we actually are going to do is calculate the quantity $DM^i_{SDV} = M^i_{SDV}(P_{homogeneous}) - M^i_{SDV}(P_{inhomogeneous})$ for each year i. Then using the bootstrap method we can create 95% confidence interval. Finally by using by the whole sample we can

see:
$DM_{SDV} = M_{SDV}(P_{homogeneous}) - M_{SDV}(P_{inhomogeneous}) = 0$. So we investigate if 0 belongs to the interval. If it does it means that the matrices have no statistical difference so time homogeneity holds if not then the opposite applies.

Chapter 4

EXTRACTING ADDITIONAL INFORMATION FROM A MARKOV CHAIN

In the previous chapter we saw how to use metrics in order to test time homogeneity of a transition matrix. In this chapter we are going to introduce useful quantities that we can extract from a transition matrix.This way we can transform the probabilities into more meaningful quantities, which can lead us to take more informative decisions. In the sections below we are going to calculate quantities such as occupancy times, expected time to default, absorbing probability, first entrance probabilities and other various concepts as well. These quantities will assist us understand how we can connect transition probabilities with other measures that can be extracted from these probabilities. The books, papers and notes used for our analysis are from: I.Dimitriou, D.Lando.

4.1 Occupancy times

Occupancy is a quantity which indicates how many times we are going to visit a particular state for a particular time horizon. This way we know in a specific time horizon the percentage of time the chain is going to be in a particular state.
Suppose we have a discrete time Markov chain (DTMC) $(x_n)_{n \in N}$

and $v_j^{(n)}$ is a random variable which represents the number of visits in rating scale j within n years with

$$v_j^{(0)}=\begin{Bmatrix} 1 & if & x_0 = i \\ 0 & if & x_0 \neq i \end{Bmatrix}.$$

Then we define $M_{ij}^{(n)} = E(v_j^{(n)}/x_0 = i)$ as the expected number of visits in rating j within n years given that we start from rating i . Also we define as $M^{(n)} = (M_{ij}^{(n)})_{i,j \in 0,1,..N}$

Then we can also write that

$$M_{ij}^{(n)} = \sum_{r=0}^{n} p_{ij}^{(r)} \ \mathbf{(4.1)}$$

or alternatively we can write equation (4.1) as:

$$M^{(n)} = \sum_{r=0}^{n} P^{(r)}$$

Occupancy is a great tool which can help us see more things than just probabilities. For instance we can see more clear how the Markov chain is going to move until it defaults. So if we have a loan or a bond we can see until maturity time how the loan is going to move.

Example 4.1

Suppose we have the following annual transition matrix P under the time homogeneity assumption. The particular matrix is going to be used in order to give loans to companies with 10 years maturity.

	1	2	3	4	5	6	7	8	D
1	91.1200	7.8020	0.8779	0.1743	0.00251	0.00029	0.00033	0.0002	0.00033
2	1.3430	90.7400	6.8850	0.7316	0.1864	0.394	0.0021	0.0043	0.0671
3	0.0859	3.1110	90.230	5.6180	0.7349	0.1145	0.0202	0.0085	0.0806
4	0.454	0.3170	4.9960	87.7800	5.5250	0.8395	0.1623	0.0173	0.3170
5	0.0079	0.0921	0.5346	6.6460	82.7100	7.8360	0.6256	0.0573	1.4870
6	0.0080	0.0613	0.1965	0.7155	7.1460	81.1600	5.6910	0.5702	4.4490
7	0.0008	0.0317	0.0418	0.244	1.0240	10.0800	70.9900	4.0120	13.5700
8	0.0003	0.00055	0.13352	0.0446	0.5476	3.733	8.8690	63.7900	22.89
D	0.0000	0.0000	0.0000	0.000	0.0000	0.0000	0.0000	0.0000	1

Calculating the matrices $P^2, P^3, ...P^{10}$ and using (4.1) we have the following matrix $M^{(10)} =$

7.31572815	2.571469	0.812303	0.212175	0.05454	0.016059	0.0028767	0.00068850	0.015347
0.44447470	7.4021756	2.271970	0.592870	0.16564	0.053311	0.0099289	0.00232827	0.057430
0.07084623	1.0297912	7.399492	1.752830	0.45785	0.147850	0.0295595	0.00542818	0.107659
0.0263789	0.2480643	1.551356	6.766750	1.44008	0.508631	0.1051244	0.01609544	0.336774
0.00834951	0.0789891	0.426348	1.710459	5.66486	1.693230	0.3229349	0.04968715	1.040451
0.0046170	0.0372813	0.148823	0.508441	1.53889	5.382146	0.9002047	0.14826999	2.329558
0.00160912	0.0166287	0.053973	0.182726	0.52877	1.614444	3.6796049	0.38992530	4.530755
0.00080850	0.0087834	0.043202	0.085679	0.26843	0.785141	0.8912101	2.82951276	6.085752
0.0000000	0.00000	0.000000	0.000000	0.0000	0.00000	0.000000	0.0000000	11.00000

(4.1) can be applied in situations where there is no time homogeneity.
So in this case $M^{(n)} = P^{(1)} + P^{(2)} + ... + P^{n)}$ In addition if $M = \lim_{n\to\infty} M^{(n)}$, is the long-term matrix of the expected number of visits from one rating to another one, then:

$$M = \sum_{r=0}^{\infty} P^{(r)} \quad \textbf{(4.2)}$$

Finally if I is the identity matrix and I-P is invertible then

$$M = (I - P)^{-1} \quad \textbf{(4.3)}$$

Note that formula (4.2) equals with formula (4.3) when there is time homogeneity.

4.2 Expected time of default

Another quantity we can use in order to make our analysis more efficient is the expected time of default. The expected time of default is the number of steps (time periods) needed in order to default.
So suppose we have a DTMC and $T = min(n : x_n = D)$ is the time

of default. Then $m_i = E(T|x_0 = i)$ is the expected time of default given that we begin from rating i$\neq D$. For i=D $m_D = 0$

$$\text{Furthermore if } m = \begin{pmatrix} m_1 \\ m_2 \\ \ldots\ldots \\ m_{N-1} \end{pmatrix} \text{ Then}$$

$$m = l + Bm \quad \mathbf{(4.4)}$$

where B is the transition probability matrix without the default vector and row and l is a vector with all it's entries equal to one.

Example 4.2

Using the same transition matrix with Example 4.1 we may form matrix B as follows

	1	2	3	4	5	6	7	8	D
1	91.1200	7.8020	0.8779	0.1743	0.00251	0.00029	0.00033	0.0002	0.00033
2	1.3430	90.7400	6.8850	0.7316	0.1864	0.394	0.0021	0.0043	0.0671
3	0.0859	3.1110	90.230	5.6180	0.7349	0.1145	0.0202	0.0085	0.0806
4	0.454	0.3170	4.9960	87.7800	5.5250	0.8395	0.1623	0.0173	0.3170
5	0.0079	0.0921	0.5346	6.6460	82.7100	7.8360	0.6256	0.0573	1.4870
6	0.0080	0.0613	0.1965	0.7155	7.1460	81.1600	5.6910	0.5702	4.4490
7	0.0008	0.0317	0.0418	0.244	1.0240	10.0800	70.9900	4.0120	13.5700
8	0.0003	0.00055	0.13352	0.0446	0.5476	3.733	8.8690	63.7900	22.89
D	0.0000	0.0000	0.0000	0.000	0.0000	0.0000	0.0000	0.0000	1

and solving the system of equations we have

$$m = \begin{pmatrix} 117.0996 \\ 107.591941 \\ 96.91762 \\ 80.7484 \\ 39.2482 \\ 21.98833 \\ 13.55768 \end{pmatrix}$$

We can use this information in order to decide about maturity time. For example it would not be wise to buy a firm bond belonging to rating 8 with more than 13 years maturity or give a loan with more than 13 years duration due to the fact that these type of company has an expected default time approximately equal to 13,5 years, as seen from the vector above.

4.3 First entrance probabilities

Another useful quantity for our analysis is the first entrance probability. The first entrance probability will help us determine in which time an entity is going to pass through a specific rating scale. Let $f_{ij}^{(n)} = p(x_n = j, x_r \neq j \ r = 1, ..., n-1 | x_0 = i)$ be the first entrance/re-entrance probability in rating j in n-steps given that we start from rating i. Then $f_{ij}^{(n)}$ is computed by solving the following system of equations
$p_{ij}^{(n)} = f_{ij}^{(n)} + f_{ij}^{(n-1)} p_{jj}^{(1)} + f_{ij}^{(n-2)} p_{jj}^{(2)} + ... + f_{ij}^{(1)} p_{jj}^{(n-1)}$ for every n=1,2,....

Giving a more analytical form we can write the system as
$f_{ij}^{(0)} = 0 \ p_{jj}^{(0)} = 1$
$f_{ij}^{(1)} = p_{ij}$
$p_{ij}^{(2)} = f_{ij}^{(2)} + p_{jj} f_{ij}^{(1)}$
$p_{ij}^{(3)} = f_{ij}^{(3)} + p_{jj} f_{ij}^{(2)} + p_{jj}^{(2)} f_{ij}^{(1)}$
....
$p_{ij}^{(n)} = f_{ij}^{(n)} + f_{ij}^{(n-1)} p_{jj}^{(1)} + f_{ij}^{(n-2)} p_{jj}^{(2)} + ... + f_{ij}^{(1)} p_{jj}^{(n-1)}$

Example 4.5

In our example we are going to use the first order transition matrix used in Example 2.2

We are going to calculate $f_{12}^{(3)}$. So we need to solve the following system
$f_{12}^{(1)} = p_{12}$
$p_{12}^{(2)} = f_{12}^{(2)} + p_{22} f_{12}^{(1)}$
$p_{12}^{(3)} = f_{12}^{(3)} + p_{22} f_{ij}^{(2)} + p_{22}^{(2)} f_{12}^{(1)}$
In our case by calculating P^2 and P^3 we have that $p_{22} = 0.9074$
$p_{22}^{(2)} = 0.826589 \ \ p_{12} = 0.07802 \ \ p_{12}^{(2)} = 0.142111 \ \ p_{12}^{(3)} = 0.19447$
So as a result we have
$f_{12}^{(1)} = 0.07802$
$0.14221 = f_{12}^{(2)} + 0.9074 f_{12}^{(1)}$
$0.19477 = f_{12}^{(3)} + 0.9074 f_{ij}^{(2)} + 0.826589 f_{12}^{(1)}$

The final results are
$f_{12}^{(1)} = 0.07802$

$f_{12}^{(2)} = 0.07141$
$f_{12}^{(3)} = 0.068123$

Although first entrance probabilities in n steps do not have a very significant practical meaning for our analysis they will help us analyze the general first entrance probability $f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ This probability is very important especially in the scenario where i=j. It i very useful to examine whether f_{jj}^* is equal to one or not. If it is equal to one that means that we are sure that we are going to return to rating j. Especially in models with re-defaults f_{DD}^* is an important quantity. Of course the calculation of f_{ij}^* is not easy using the series.
Moreover the expected number of visits is connected with the first entrance probability. Through the formula $M_{ij} = f_{ij}^* M_{jj}$
We can also write an equation which relates the first entrance probabilities to the first order probabilities:

$$f_{ij}^* = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj}^* \quad \textbf{(4.5)}$$

So in case we have matrix M we can calculate f_{ij}^* for every $i \neq j$ and for i=j we can also calculate f_{jj}^*

Finally keep in mind that $f_{iD}^* = u_i$, where u_i is the long term probability of default given that we start from rating i.

Example 4.6
Here we have the same first order transition matrix used in Example 2.2

In our example we are going to find f_{i2}^* for every i=1,2,...D
Using formula (4.5) we have the following system of equations

$$
\begin{aligned}
f_{12}^* &= p_{12} + p_{11}f_{12}^* + p_{13}f_{32}^* + p_{14}f_{42}^* + p_{15}f_{52}^* + p_{16}f_{62}^* + p_{17}f_{72}^* + p_{18}f_{82}^* + p_{1D}f_{D2}^* \\
f_{32}^* &= p_{32} + p_{31}f_{32}^* + p_{33}f_{32}^* + p_{34}f_{42}^* + p_{35}f_{52}^* + p_{36}f_{62}^* + p_{37}f_{72}^* + p_{38}f_{82}^* + p_{1D}f_{D2}^* \\
&\;\;\vdots \\
f_{82}^* &= p_{82} + p_{81}f_{12}^* + p_{83}f_{32}^* + p_{84}f_{42}^* + p_{85}f_{52}^* + p_{86}f_{62}^* + p_{87}f_{72}^* + p_{88}f_{82}^* + p_{8D}f_{D2}^*
\end{aligned}
$$

Solving the 9x9 system of equations we have all f_{i2}^* for every i=1,2,...D

4.4 Absorbing probability

In this subsection we are going to discuss absorbing probabilities which will help us define long term probabilities of default. Suppose we have a DTMC and $T = (min\ n : x_n = D)$ and also $v_i(n) = p(T > n | x_o = i)$ is the absorbing probability in more than n steps(years) given that we start from rating i$\neq D$
Also let

$$
V(n) = \begin{pmatrix} v_1(n) \\ v_2(n) \\ \dots\dots\dots \\ v_{N-1}(n) \end{pmatrix}.
$$

Then

$$
V(n) = B^n l \quad \textbf{(4.6)}
$$

Example 4.7 Again we give the same transition matrix as in example 4.2 and example 4.1

Then by using (4.6)for $n = 10$ we have the matrix $V(10)$

$$
V(10) = \begin{pmatrix} 0.9959166 \\ 0.9870675 \\ 0.974318 \\ 0.9261154 \\ 0.7999061 \\ 0.6061614 \\ 0.3525641 \\ 0.2068218 \end{pmatrix}
$$

Using $v_i(n)$ we can calculate the long term probabilities of default. But before we move to this procedure we need to give the definition of some terms.
Suppose we have a transition matrix with N different states. Let i be a state in the transition matrix.

1. We say that i is recurrent if and only if $f_{ii}^* = 1$
2. We say that i is transient if and only if $f_{ii}^* < 1$
3. We say that i is positive recurrent if and only if it is recurrent and $\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)} < \infty$
4. We say that i is zero recurrent if and only if is recurrent and $\mu_i = \infty$

Another additional term in order to understand how a Markov chain works is the period of a state. We call period(d_i) of a state i the greatest common divisor(GCD) of all integers $n \geq 1$ so that $p_{ii}^{(n)} > 0$. If

1. $d_i = 1$ then state i is called aperiodic
2. $d_i = m > 1$ then the chain returns to i in times multiple of m

Moreover it is also very important in order to continue our analysis to define the accessibility between two states. So suppose we have two different states in a Markov chain i and j. State j is accessible from state i if and only if there exists an $n \geq 0$ so as $p_{ij}^{(n)} > 0$.We symbolize this as $i- > j$. Having in mind the above if we have two different states in a Markov chain i and j, then state i and j communicate with each other if and only if $i- > j$ and$j- > i$. More analytically if and only if there are $n, m \geq 0$ so as $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. We denote that $i < - > j$. When two states communicate with each other then all the properties owned by one of the state(aperiodic, positive recurrent, transient ect...) are automatically transferred to the other one. So if are give a family of communicative states then all the properties of one of the states, are shared automatically with all the other states of the family.

Example 4.8

Let P be the transition matrix with 3 states 1,2,3

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Then $p_{11} = 0.5 > 0$, $p_{11}^{(2)} = p_{11}p_{11} = 0.025 > 0$, $p_{11}^{(3)} = p_{11}p_{11}p_{11} + p_{12}p_{22}p_{23} > 0$
$GCD(1, 2, 3, ...) = 1 = d_1$. So state 1 is not periodic and because 1$< - >$2$< - >$3 then 2 and 3 are also not periodic.

From the above we can understand that if $p_{ii}^{(1)} > 0$ then state i is not periodic.

example 4.9

Let P be a transition matrix with 3 states 1,2,3

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$p_{11} = 0, p_{11}^{(2)} = 0, p_{11}^{(3)} = p_{11}p_{12}p_{13} = 1 > 0, p_{11}^{(4)} = 0, p_{11}^{(5)} = 0, p_{11}^{(6)} = 1 > 0$
So $GCD(3, 6, 9, 12, ...) = 3 = d_1 = d_2 = d_3$

Suppose we have a Markov chain.We say that state j is ergodic if and only if it is positive recurrent and not periodic.

Suppose we have a Markov chain (x_n) and state i has the ergodicity property. Let $v_i = \lim_{n\to\infty} v_i(n) = p(T = \infty | x_0 = i)$ be the non absorbing long term probability and

$$V = \begin{pmatrix} v_1 \\ v_2 \\ .... \\ v_{N-1} \end{pmatrix}.$$

In that case vector V is given by solving the following equation.

$$V = BV$$

Example 4.10

Once more we are using the same matrix as in Example 2.2.

From the transition matrix we can see that $p_{11} = 91.1200 > 0$ so state 1 is aperiodic. Due to the fact that all states except D communicate with each other then the not periodic property gets transferred to all the other states except D.$p_{DD} = 1 > 0$ so state D is also not periodic.
Also $f^*_{DD} = 1$, indicating that state D is a positive recurrent state. For the other states $f^*_{ii} < 1$, so they are transient. After solving the system from theorem 4.6

$$V = \begin{pmatrix} 51.65772 \\ 51.62803 \\ 51.63025 \\ 51.55889 \\ 51.06355 \\ 49.73983 \\ 43.61159 \\ 35.54558 \end{pmatrix}$$

This vector can give us useful information about the probability of default for a "very big n"($n \to \infty$).With this matrix we are in a position to know how many companies are going to default in a long term period from each rating just by calculating the quantities $u_i = 1 - vi = p(T < \infty | x_0 = i)$ which is the probability of default in a long term time horizon. So

$$U = \begin{pmatrix} 48.3423 \\ 48.37197 \\ 48.37975 \\ 48.44111 \\ 48.93645 \\ 50.26062 \\ 56.38841 \\ 64.45442 \end{pmatrix}$$

Furthermore it is very useful to give some relationships about transition probabilities in an infinite number of steps, in order to understand the chain's long term behaviour.
Let (x_n) be a Markov chain which represents the state of an entity in time n. Also let $\lim_{n\to\infty} p^{(n)}_{ij}$=$\lim_{n\to\infty} p(x_n = j | x_o = i)$. Then:

1. $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ if i and j are transient states.
2. $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ if i is an absorbing state and j a transient state
3. $\lim_{n\to\infty} p_{ij}^{(n)} = v_i$ if i is a transient state and j the absorbing state

4.5 A model with re-defaults

As we mentioned in Chapter 1 the word default does not always mean bankruptcy. There are many times where a new investor comes and pays the debt of a firm and the firm starts again it's function. In practice that seems a more realistic model. In our model where there is an absorbing state when a firm defaults if this firm re-opens again then we record that firm as a whole new firm we give a starting rating and we proceed as mentioned in the previous chapters. Moreover many times by the term default we mean a delay of a particular payment usually for over 90 days

What's really interesting in the transition matrices whether they have an absorbing default state or not is to analyze their asymptotic distribution and observe the differences from the transition matrices with absorbing states. Another thing important to mention is that all the results that we found in the transition matrices with absorbing state(M_{ij},$f_{ij}^{(n)}$,f_{ij}^{*}) can be also applied in transition matrices with re-defaults. Of course now quantities such as m_i,v_i have no meaning.
Moreover what is different when we have chains with re-defaults is that if (x_j) is a time homogeneous Markov chain, P the annual transition matrix and $\lim_{n\to\infty} p_j^{(n)} = \lim_{n\to\infty} P(x_{n=j}) = \pi_j$,then
$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$

The statements above prove that in case of re-defaults the initial state i is not significant at all for the value of the long-term probability. So the chain tends to forget from where it started. But how can we find π_j? The way we can find it is by using the vector π, which contains all π_i . Then by solving the equation

$$\pi = \pi \text{P} \quad \textbf{(4.7)}$$

we are able to compute all π_i, where P is the annual transition matrix. Note that in order to use this equation we need a time homogeneous Markov chain with the ergodicity property .

Example 4.11

Let P be an annual transition matrix with time homogeneity,4 rating states,where state D stands for a delay of a payment or a series of payments for over 3 months

	1	2	3	4	5	6	7	8	D
1	91.1200	7.8020	0.8779	0.1743	0.00251	0.00029	0.00033	0.0002	0.00033
2	1.3430	90.7400	6.8850	0.7316	0.1864	0.394	0.0021	0.0043	0.0671
3	0.0859	3.1110	90.230	5.6180	0.7349	0.1145	0.0202	0.0085	0.0806
4	0.454	0.3170	4.9960	87.7800	5.5250	0.8395	0.1623	0.0173	0.3170
5	0.0079	0.0921	0.5346	6.6460	82.7100	7.8360	0.6256	0.0573	1.4870
6	0.0080	0.0613	0.1965	0.7155	7.1460	81.1600	5.6910	0.5702	4.4490
7	0.0008	0.0317	0.0418	0.244	1.0240	10.0800	70.9900	4.0120	13.5700
8	0.0003	0.00055	0.13352	0.0446	0.5476	3.733	8.8690	63.7900	22.89
D	0.0002	0.00044	0.011	0.0446	0.5	4.2	12.7437	22.4	60.1

Firstly the chain has the ergodicity property because:

1. all states are positive recurrent because all states communicate with each other.
2. all states are not periodic because $p_{11} > 0$ so state 1 is not periodic and because all states communicate the whole chain is not periodic

So because of the above we can compute π by using that $\pi = \pi\mathrm{P}$ and also that $\sum_{j=1}^{D} \pi_j = 1$, we have that:

$$\pi' = \begin{pmatrix} 2.180018\% \\ 7.774842\% \\ 15.20319\% \\ 16.1964\% \\ 13.76484\% \\ 16.18434\% \\ 10.51168\% \\ 7.846051\% \\ 10.3335\% \end{pmatrix}$$

The quantity π_j has another meaning apart from the long term probability. It also represents the amount of time that our chain is going to be in state j.

Chapter 5

Summary and future research

5.1 Synopsis

Although quantifying credit risk is a continuous process, this thesis aims to present some major problems concerning credit risk. First we construct transition probability matrices for different time frames in both discrete and continuous time, and try to explain their physical meaning. Secondly we explain how a first order transition matrix is calculated using two different methods and analyze their advantages and disadvantages for each one of them. Furthermore we present different methods for testing the time homogeneity property, which is crucial for transition matrices construction. Finally we explain how transition matrices are used in order to create alternative default measures.

Bibliography

[1] R.Gunnvald (2014), *Estmating Probability of Default Using Rating Migration in Discrete and Continuous Time*

[2] Y.Jafry, T.Schuermann (2004), *Measurement, Estimation and Comparison of Credit Migration Matrices*

[3] D.Lando (2004), *Credit risk Modeling -Theory and applications*

[4] P.Lencastre, P.Lind, F.Raischel, T.Rogers (2014), *Are Credit Ratings Time Homogeneous and Markov?, Cornell university edition*

[5] Standard and Poor's (2012), *Annual Global Corporate Default Study And Rating Transitions*

[6] K. Papaioannou, T. Ferendinos (2000), *Statistics Theory, Stamoulis editions*

[7] O. Chrysafinou (2016), *An Introduction to Stochastic Processes, Sofia editions*

[8] I.Dimitriou (2017), *Stochastic Processes, class notes from Mathematics department of University of Patras*

[9] M.Koutras(2018), *Introduction to Credit Risk, class notes from actuarial science and risk management M.S.c of university of Piraeus*

[10] K.Chung (1960), *Transition matrix: basic properties*