Term structures of default probabilities

Anisa CAJA

Submitted on 17 March 2010

Supervisors:

Frédéric PLANCHET
Professor at ISFA

Olivier BLUMBERGER
Euler Hermes

Jean-François DECROOCQ
Euler Hermes

Contents

Mots clés: Structures par terme des probabilités de défaut, transitions entre classes de notes, Chaîne de Markov continue non-homogène, matrice génératrice, régularisation de la matrice génératrice.

Résumé

Afin de mieux gérer son risque, il serait utile pour un assureur crédit comme Euler Hermes d'étudier l'évolution intra-annuelle des probablités de défaut dans son portefeuille. Avoir une structure par terme des probabilités de défaut lui permettrait d'examiner le lien qui existe entre les probabilités de défaut de deux périodes consécutives, pour ensuite essayer d'expliquer ce lien. Le cycle économique a une influence sur cette dépendance; c'est pour cela qu'il paraît oppurtun de conditionner les termes structures par l'état de l'économie. Dans ce mémoire, nous présentons une manière de calibrer des structures par terme des probabilités de défaut, ceci en utilisant les chaînes de Markov continues et non-homogènes. Ensuite nous présentons la mise-en-oeuvre et les résultats du calibrage avec des données de Standard & Poors. A la fin du mémoire nous proposons une manière de calibrer les structures par terme des probabilités de défaut en tenant compte de la phase du cycle économique. Nous concluons ce mémoire en donnant quelques pistes de recherches sur des problématiques connexes et qui permettraient de faire évoluer le modèle interne d'Euler Hermes pour mieux intégrer les termes structures.

Keywords: Term structures of default probabilities, rating transitions, non-homogeneous continuous time Markov chain, generator matrix, regularization of the generator.

Abstract

In order to better manage its risk, it would be helpful for a credit insurer like Euler Hermes to see how the probabilities of default in its portfolio vary during a one-year period. Having a term structure of default probabilities would make it possible to study the link between default probabilities of two subsequent periods in a year and then try to explain this link. Economic cycles influence this dependence, and this is the reason why conditioning term structures on the business cycle phase would be useful. In this paper, we present a possible way of calibrating term structures of default probabilities by using non-homogeneous Continuous time Markov Chains. Then we calibrate term structures of default probabilities using S & P data. Afterwards we propose a way of conditioning the term structures on the state of the economy. Finally, some further research directions on some closely related topics are given. They would allow to better take into consideration the term structures in the internal model of Euler Hermes.

Introduction

Euler Hermes (EH) insures its clients against losses arising from buyers' insolvencies in their domestic or export markets. Therefore modelling default probabilities is an important feature of its business; a better knowledge of those probabilities helps manage the contracts more appropriately.

There exist many credit risk models. EH has its own internal model. One of the advantages of this model is that it accounts for correlations between defaults. A buyer defaults if its ability-to-pay a debt falls below a certain threshold. The ability-to-pay of a buyer is given by a factor model, i.e. it depends on systematic risk factors and an idiosyncratic risk factor proper to the firm. The correlation between defaults is the result of the fact that a part of the ability-to-pay of each buyer is given by systematic risks, even though the systematic risk factors do not influence the ability-to-pay of each buyer at the same degree. Moreover systematic risks are correlated and every year a matrix with the factor correlations is published. Contracts are subscribed for a lapse of one year so the default probabilities for a horizon of one year are studied and default thresholds are calculated.

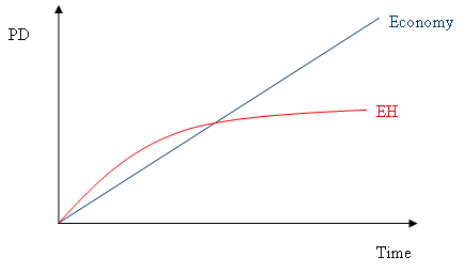
However it would be interesting to know how the probability of default varies within a year. This probability will not be independent during two sub-periods in a year. On the one hand, most probably the economic features will not change radically from one sub-period to the other and on the other hand if we consider the same company, its specific risk will stay almost the same. Thus we should model the fact that probabilities of default for the same company during two sub-periods of a year are not independent.

There are at least two possible ways to deal with this. The first way would be conditioning the probability of default for the next period on the current economic state. This conditioning is done indirectly by actually conditioning the future economic state on the present one. Since the current economic state reflects on the default probability for the current sub-period, conditioning on the current probability of default and conditioning on the current economic state are alike.

The second possible solution would be differentiating between companies of different rating classes and looking at the default probability of a rating class for different time horizons. By definition, a homogeneous Markov chain would not allow to model the dependence between two subsequent periods meanwhile a non-homogeneous Markov chain would, because it depends on time. Indeed,

it has been observed that the best fit of the model to historical data corresponds to the assumption that the credit migration process, which includes defaults, is a non-homogeneous Markov chain.

From EH point of view, it would be interesting to see the effect of risk mitigation on the probabilities of default. Risk mitigation will not be the same in periods of expansion or recession. At the beginning of a recession period the average default rate in the portfolio of EH will be larger than the average default rate in the whole economy. The EH portfolio is a subset of the portfolio composed of all the firms in the economy. One characteristic of this portfolio is that the risk of default of the firms in it is higher than the risk of default of the global economy portfolio. This is why the suppliers of those firms need to buy a cover to protect themselves in case of default of their clients. This phenomenon is called adverse selection (demand for insurance increases with the risk of default or loss). So when the economic situation starts to worsen, the rate of defaulting companies in the EH portfolio is larger than the rate of defaults in the economy. However EH has the power to react and lower granted limits or cancel contracts. So if the recession goes on, the default rates in EH will be lower than those in the economy because the "bad risks" have been evacuated from the portfolio.



Risk mitigation impact on probabilities of default during a recession period

We want to see the impact of risk mitigation and a way of doing this is to compare the PD term structure of the EH portfolio and the PD term structure of the economy portfolio in expansion and recession periods. We will present a way of computing a term structure of PDs by mixing the two approaches above.

The aim of this thesis is to present how a term structure of default probabilities can be introduced in the current EH model. The first chapter will present the credit insurance business and the current framework of the internal model. The second chapter will consist in a reminder of Markov chains, transition and generator matrices. In the literature the credit migration process is modelled with Markov chains and here we will make the same assumption, so we need to remind some results we will use afterwards. In the third part of the thesis we present the possible solutions to the problem in more detail. We calibrate a term structure of probabilities of default with the non-homogeneous Markov chain method using data from S & P. Some further research directions are given in the last chapter.

Chapter 1

Framework

1.1 What is the Credit Insurance business about?

In the credit insurance business the contractual relation involves two parties, namely a credit insurer and a policyholder. However a third party intervenes in this relation and plays an important role, the policyholder's client, named the buyer. The reason why credit insurance contracts exist is because companies (policyholders) buy such contracts to protect themselves in case of default[1] of one of their clients (the buyer).

Thus a credit insurance policy can be considered as a triangle relation. It can be represented as in the figure below:

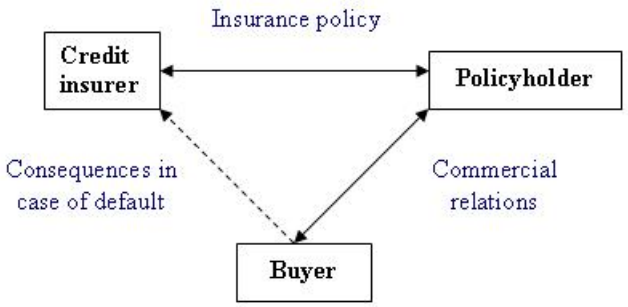


Fig. 1.1: Triangle relation in credit insurance

We should note that a certain buyer may be party in many policies, since it can be client of many policyholders which have bought a credit insurance contract concerning this buyer.

If a buyer defaults during a year and the policyholder declares the corresponding claim, the credit insurer must indemnify the policyholder, after taking

[1] We will define the default event later in this section

into account the different policy parameters. Those parameters make it possible for the credit insurer to *mitigate* the effects of a default.

An effective measurement of credit risk in a portfolio involves three quantities: the amount of financial loss in the event of default of a buyer, the individual probability of default for each buyer, and the correlation between probabilities of default for different buyers. In what follows we present how the credit insurer measures these quantities.

1.2 Effective loss of the insurer in case of default

If a buyer defaults then the credit insurer will probably not reimburse all the amount of the outstandings to the policyholder. He takes measures in order to pay less than the actual amount the buyer owes the policyholder.

Here we present some of the preventive measures that the insurer can take when stipulating the contract and some other measures it takes after the contract is signed.

1. One of the most important parameters of the policy is its *limit*. The limit is defined as the amount of the credit limit granted to the policyholder with regard to the defaulting buyer, so it is the maximum coverage. Credit limits can be reduced or cancelled at any time depending on the creditworthiness of the buyer.

2. The contract might involve *clauses* of uninsured percentage, deductibles, maximum liability or annual aggregate. While the first two terms might be familiar, the notions of maximum liability and annual aggregate need to be defined. The *maximum liability* corresponds to the maximum total amount of indemnification that a credit insurer will pay within a given time horizon. The *annual aggregate* is a deductible that applies to the sum of claims or indemnifications of a policyholder over a certain time period. If the sum is smaller than the deductible nothing is paid; otherwise the amount exceeding the franchise is paid.

3. The credit insurer may choose to share the risk with another insurer, i.e. to be *reinsured*, which is a rather classical move.

4. Recoveries from *recollection* before or after the indemnification.

The last three loss mitigation measures are aggregated into one parameter called *Loss Given Default*. Meanwhile, the limit of a policy intervenes in the definition of the *exposure at default*.

Exposure at default (EAD) is defined as the effective outstandings at the time of default, which represent the used limit at default. In general the effective outstandings during the maturity period of one year are lower than the granted limit itself. Moreover we can observe that in general the granted limit can be decreased several times a year due to the deterioration of the creditworthiness of the buyer whose financial performance worsens. EAD can be estimated by making assumptions on the usage degree of the limit in case of default with

respect to the granted limit. This degree is called *Usage given default (UGD)* and is given by the formula:

$$\text{UGD} = \frac{\text{limit used at time of default}}{\text{granted limit one year before}} \tag{1.1}$$

Thus, formally we have:

$$\text{EAD} \;=\; \text{granted limit} \;\times\; \text{UGD}.$$

We should add that the degree of limit reduction shows the efficiency of the management of the contracts. In the best possible case, the granted limit is cancelled, i.e. the contract is cancelled, before the default of the buyer. In this case the credit insurer is free of any legal obligation and the default of the buyer will be of no consequence on it. In addition to this, when the credit insurer decreases the granted limit, the policyholder will rationally lower the trading activity with that particular buyer. There are two reasons which lead to that. The first reason is that if the insurer decreases the limit, the policyholder will look at this measure as a warning saying that the buyer is less capable of paying off the debt, so the chances it never reimburses the invoice amounts increase. The second reason is that since the granted limit decreases, the amount the policyholder loses in case of default of its buyer is greater.

The figure below sums up how a policy is typically managed and some of the notions defined above.

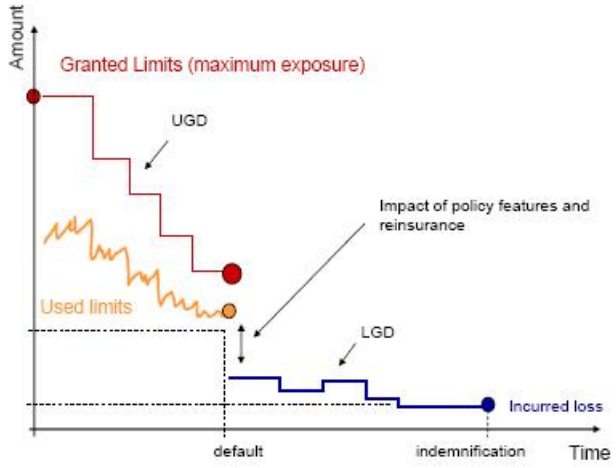


Fig. 1.2:Lowering granted limits within a year.

In what precedes we presented how insurance policies are managed and by doing this we also gave an idea of the amount of loss the insurer faces in case of default of a specific buyer. However an important feature of the credit risk business is the prediction of the mathematical chance that a default event will occur. In the following section we will deal with what is relative to the default probabilities.

1.3 Default probabilities

First let us define what we will call *default event*. There are three different types of default events:

1. **Insolvency** - The default event is the legal insolvency of the buyer;
2. **Opening of a claim file** - The default event is the opening of a claim file for the buyer, either because the policyholder declares a claim, or because an internal process begins, e.g. when a file is transferred from the credit department to the claims department;
3. **Indemnification** - The default event is the declaration of a claim which was already indemnified by the credit insurer.

The probability of default (PD) over one year is then estimated by the empirical frequency of defaults during the last 12 months. So the estimated PD will be:

$$\text{PD} = \frac{\text{number of buyers having defaulted within the 12-month period}}{\text{number of buyers at the beginning of the 12-month period}} \quad (1.2)$$

Since there are 3 types of default events, a probability of default corresponding to each one of the definitions can be computed. Of course those probabilities are not necessarily the same. A type of probability of default will be picked depending on what we want to study. According to the general accounting *principle of prudence*, after having calculated the capital requirements with each type of PD, the greatest one should be chosen.

PDs depend on the creditworthiness of a buyer. Each buyer is given a *rate* (grade) reflecting its current financial situation and ability to meet debt obligations. So *rating classes* (grading classes) gather buyers with the same grade. Default probabilities are defined for each grading class.

1.4 Correlation between probabilities of default

Certainly the individual PDs are important to know. They are used in the computation of capital requirement for an individual buyer. However, in the case of credit portfolios, knowing the risk related to each of the single exposures in the portfolio is not sufficient to evaluate the entire risk of it. Indeed, by doing this we would not consider the fact that the default event for one company is probably correlated to the default of another company in the same country and of the same sector. Studying and integrating the correlation between default events in a risk management model proves to be crucial in order to evaluate the risk faced by the credit insurer correctly. A portfolio composed of low credit risk companies but whose asset values are very correlated with one another, is very risky.

1.4.1 Simple illustration

Let 1 and 2 be two companies of the same sector. Let D_i be the Bernoulli variable indicating if the buyer $i = 1, 2$ defaults. Let ρ be the annual correlation

coefficient between the two companies.
Let us define p_i such that $p_i = \mathbb{P}(D_i = 1)$ for $i = 1, 2$ and p_{12} such that $p_{12} = \mathbb{P}(D_1 = 1, D_2 = 1)$. The correlation coefficient between the default event of the two firms is given by the formula:

$$\rho = \frac{\mathrm{Cov}(D_1, D_2)}{\sqrt{\mathbb{V}(D_1)}\sqrt{\mathbb{V}(D_2)}} \tag{1.3}$$

$$= \frac{\mathbb{E}(D_1 D_2) - \mathbb{E}(D_1)\mathbb{E}(D_2)}{\sqrt{\mathbb{V}(D_1)}\sqrt{\mathbb{V}(D_2)}} \tag{1.4}$$

$$= \frac{p_{12} - p_1 p_2}{\sqrt{p_1(1-p_1)}\sqrt{p_2(1-p_2)}} \tag{1.5}$$

Let us suppose for example that the default probabilities p_1 and p_2 for year N are both equal to 2% and the correlation between defaults is $\rho = 0.1$. Then the probability the two companies default is $p_{12} = 0.24\%$. Let us imagine two possible scenarios for $N + 1$:
(a) Individual probabilities of default p_1 and p_2 increase and become equal to 0.4 and the correlation measured by ρ stays the same.
(b) Individual probabilities of default p_1 and p_2 increase and become equal to 0.4 and the correlation increases and the correlation coefficient becomes equal to $\rho = 0.2$. This scenario is more probable to happen compared to the first one because in general if credit quality decreases the correlation between defaults tends to increase.

In scenario (a) the joint probability of default becomes equal to $p_{12} = 0.544\%$ and in scenario (b) $p_{12} = 0.928\%$. So not taking into account the correlation between defaults would have underestimated the joint probability of default almost by a half, which is quite considerable.

1.4.2 Modelling default correlations

Default correlations are difficult, if not impossible, to be measured directly. A possible way of inferring default correlations is knowing the individual probabilities of default and asset correlations. This seems quite sensible since a company is likely to default if the value of its assets falls below a certain threshold, so two companies will default jointly if the values of their respective assets decrease simultaneously down to a certain degree. The simultaneousness of the two decreases and their degree define the asset correlation of the two companies.

A way to deal with this, is to make the assumption that the joint distribution of default of two companies is a Gaussian copula whose correlation coefficient is the correlation between the asset values of those companies. We denote δ_{ij} the correlation coefficient between the assets of the firms i and j and ρ_{ij} the correlation between the default events of i and j.

Let us denote S_i and S_j the corresponding asset values and d_i and d_j the corresponding default thresholds. We assume that $S_i \sim \mathcal{N}(\mu_i, \sigma_i)$ and $S_j \sim \mathcal{N}(\mu_j, \sigma_j)$.

p_i and p_j are the individual probabilities of default so that: $p_i = \mathbb{P}(S_i < d_i)$ and $p_j = \mathbb{P}(S_j < d_j)$. $p_{ij} = \mathbb{P}(S_i < d_i, S_j < d_j)$ is the joint probability of default. As a result of the Gaussian copula we have:

$$\begin{aligned} p_{ij} &= \mathbb{P}(S_i < d_i, S_j < d_j) && (1.6)\\ &= N_2(N^{-1}(p_i), N^{-1}(p_j), \delta_{ij}) && (1.7)\\ &= \int_{-\infty}^{d_j}\int_{-\infty}^{d_i} \frac{1}{2\pi\sigma_i\sigma_j\sqrt{1-\delta_{ij}^2}} \exp\left(-\frac{z}{2(1-\delta_{ij}^2)}\right) ds_i ds_j && (1.8) \end{aligned}$$

where

$$z = \frac{(s_i-\mu_i)^2}{\sigma_i^2} - 2\frac{\delta_{ij}(s_i-\mu_i)(s_j-\mu_j)}{\sigma_i\sigma_j} + \frac{(s_j-\mu_j)^2}{\sigma_j^2} \quad (1.9)$$

N_2 denotes the Gaussian copula[2] or the bivariate normal distribution. So the asset correlation is an important factor if we want to know the joint probability of default.

After having calculated the joint default probability we can obtain the default correlation between the two firms i and j by the formula given in the previous subsection:

$$\begin{aligned} \rho_{ij} &= \frac{p_{ij}-p_ip_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} && (1.10)\\ &= \frac{N_2(N^{-1}(p_i), N^{-1}(p_j), \delta_{ij}) - p_ip_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} && (1.11) \end{aligned}$$

However, in order to make such a calculation, the asset correlation should be known. The problem is that it is not directly observable.

We present here three possible ways to deal with this problem: approximating asset correlations by equity correlations, deriving them from joint default probabilities and finally the use of factor models.

Using equity correlations as a proxy for asset correlations

It has become common market practice to use equity correlations as a proxy for asset correlations. The underlying assumption is that equity return should reflect the value of the underlying firm, so two firms with highly correlated equity return should have highly correlated assets. Nevertheless, according to Servigny and Renault [2004],equity returns incorporate a lot of noise, like bubbles, and are affected by supply and demand effects (liquidity crunches) that are not related to the firms' fundamentals. The relation between those assets and equities is not linear. This is why the linear correlation coefficient will not be the same for the two of them and this makes equity correlations poor substitutes of asset correlations.

[2]The Gaussian copula does not necessarily represent the true relationship between the default events of two companies. The copula approach becomes quite complicated to be implemented and risks to be over-parametrized if we look at the joint default of more than two companies, so it is not the approach chosen in the internal model.

Default-implied asset correlation

One needs to calculate the individual default probabilities and the joint default probabilities. In order to compute the joint default probability we use the formula:

$$p_{ij} = \sum_t w_t^{ij} \frac{D_t^i D_t^j}{N_t^i N_t^j} \tag{1.12}$$

where w_t^{ij} is the weight representing the relative importance of the sample in year t, i.e. $w_t^{ij} = \frac{N_t^i N_t^j}{\sum_k N_k^i N_k^j}$, D_t^i and D_t^j are the number of defaults for year t in the group where the firm i and j belong respectively. N_t^i and N_t^j are the number of buyers at the beginning of year t in the group where the firm i and j belong respectively. Afterwards, by making the assumption that the asset values follow Gaussian distributions and their joint distribution is a Gaussian copula, we can estimate the *default-implied asset correlation* measured by the coefficient δ_{ij} so that we can have $p_{ij} = N_2(d_i, d_j, \delta_{ij})$.

Multi-factor model

In a factor model it is assumed that there is a "latent variable" and the default event is then defined as being the event of the latent variable falling below a certain threshold. This "latent variable" is commonly named *asset return* because the default event in this case is defined similarly to the case of Merton models which define default as being the event when the value of the firm falls below the value of its liabilities.

A factor model explains the asset return value in terms of values of a set of return drivers or risk factors. We denote Z_i the asset return of company (buyer) i and $\{R_\nu | \nu = 1, ..., k\}$ the risk factors. Then the model would be:

$$Z_i = \omega_{i1} R_1 + \omega_{i2} R_2 + ... + \omega_{ik} R_k + \epsilon_i = \sum_{\nu=1}^{k} \omega_{i\nu} R_\nu + \epsilon_i \tag{1.13}$$

where $\omega_{i\nu}$ is the sensitivity of the asset return of buyer i to the systematic factor ν.

ϵ_i is the specific risk factor, also called idiosyncratic risk factor of buyer i. It is like an error term in a regression model and represents the part of the asset return which is not explained by the systematic risk factors. We assume that $\forall i = 1, ..., n$:

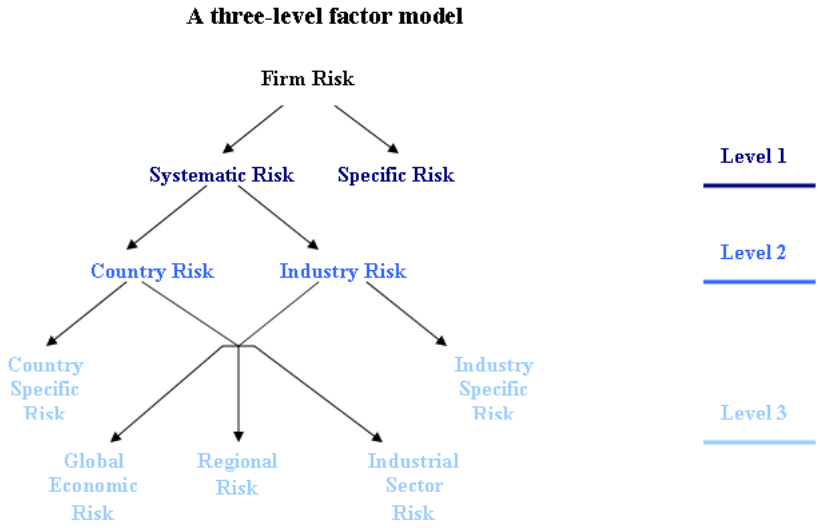
1. $\forall \nu = 1, ..., k \ \rho(R_\nu, \epsilon_i) = 0$, i.e. the idiosyncratic risk factor is not correlated to the systematic risk factors;
2. $\forall i \neq j \ \rho(\epsilon_i, \epsilon_j) = 0$, i.e. the idiosyncratic risk factors of two different buyers are not correlated between them.

A factor model is specified for each buyer. This means that each buyer has its specific risk, and more important, its global risk depends also on systematic factors, which are common to all buyers. For this reason, the correlations between the systematic factors R_ν (together with the respective weights $w_{i\nu}$ and $w_{j\nu}$) define the correlations between the buyers i and j, and the latter is defined *only* by them because idiosyncratic risks for two different buyers are uncorrelated. This model is practical to work with because if we know the covariance matrix between the systematic factors then we can have the correlations between asset returns of different buyers and the number of factors is a lot smaller than the number of buyers.

So factor models define the correlation between buyers. A natural question one can ask would be: What are those factors exactly? Which are the variables that define the correlation between buyers? Which can be the best way of clustering them? An answer to this question is given by the model presented below.

Example of a multi-factor model

This model is a three-level factor model. The picture below represents graphically this type of model.



The first level of the structure makes the difference between firm-specific and systematic risks.

$$Z_i = \alpha_i \phi_i + \epsilon_i \tag{1.14}$$

In the second level, we define the systematic risk ϕ_i as a weighted sum of country and industry factors to which the firm has exposure. So the asset return can be written as:

$$Z_i = \alpha_i \left(\sum_{\nu_C=1}^{k^C} \omega_{i\nu_C} R_{\nu_C} + \sum_{\nu_I=1}^{k^I} \omega_{i\nu_I} R_{\nu_I} \right) + \epsilon_i \tag{1.15}$$

where R_{ν_C} and R_{ν_I} are the systematic country and industry factors respectively and k^C and k^I are the total number of those factors. The weights ω depend on the part of sales or assets of the firm in a particular country or industry.

The systematic risk or undiversifiable risk of firm i is measured by the coefficient α_i and the specific or diversifiable risk by ϵ_i. α_i is estimated by linear regression and is related to the R^2 [3] of the regression of the time series of the asset return Z_i on the systematic factors ϕ_i. We have:

$$R^2 = \frac{\mathbb{V}(\alpha_i \phi_i)}{\mathbb{V}(Z_i)} \tag{1.16}$$

So, we also have:

$$\alpha = \sqrt{\frac{R^2}{\mathbb{V}(\alpha_i \phi_i)}} \mathbb{V}(Z_i) \tag{1.17}$$

For small firms whose market return data is not available, R^2 can be given approximately by comparing it to similar firm whose R^2 is known. A similar firm is a firm of the same size, located in the same country and working in the same industry.

In the third level, the country and industry systematic risk factors are expressed as a sum of systematic and idiosyncratic factors. We have:

$$R_{\nu_C} = \sum_{f=1}^{F} \omega_{f\nu_C} R_f + \epsilon_{\nu_C} \tag{1.18}$$

$$R_{\nu_I} = \sum_{f=1}^{F} \omega_{f\nu_I} R_f + \epsilon_{\nu_I} \tag{1.19}$$

where R_f are common factors for country and industry systematic factors.

Those common factors are divided into three groups and are independent from one another. Their effects are measured by sensibility coefficients, denoted $\omega_{f\nu_I}$ and $\omega_{f\nu_C}$. The common factors are classified as follows:

- Global economic factor which captures the overall effect of the global economy;
- Regional factors which catch the regional economic effects by big geographical area;
- Sector factors which capture the industry effects after the global and economic effects have been removed. The sectors are defined corresponding to the type of service or good produced, like technology, medical, extraction, etc.

[3] the portion of the variance of the asset return of buyer i, $\mathbb{V}(Z_i)$, due to the variance resulting from the impact of systematic factors.

Each country and industry is more or less influenced by those factors. The overall economy of a country is part of the global economy so if it is doing more or less well depending on how the global economy is developing and the degree to which its economy is interacting with the rest of the world. Concerning an industry, the effect of the global economy on a specific industry depends on the object of the industry since people do not have the same behaviour towards all goods and services in times of economic downturn. The region where the country is located plays an important role since it can determine if this country has been subject to any particular natural phenomenon for example. The return of an industry depends also on the region where it is primarily located. Industries which are specific to a certain region (like tobacco to Latin America or oil refining to the Middle East countries) where there has been a natural catastrophe for example will strongly suffer the consequences of that. The returns of the sectors which are the most developed in a country will influence its economy. On the other hand, the returns of an industry depend naturally on the industry sectors it is connected to.

Nevertheless, each country and industry have their specificities, which should be taken into consideration in the model.

Let's remind that our aim is to model default correlations and in order to do that we had to model asset returns. Modelling them by factor models allows us to compute correlations in a rather simple way. The only thing we need in order to model correlations in this framework is the covariance matrix of the systematic factors. Indeed the correlation between two assets i and j is given by:

$$\rho(Z_i, Z_j) = \frac{Cov((Z_i, Z_j)}{\sqrt{\mathbb{V}(Z_i)}\sqrt{\mathbb{V}(Z_j)}} \tag{1.20}$$

$$= \frac{\alpha_i \alpha_j Cov(\phi_i, \phi_j)}{\sqrt{\mathbb{V}(Z_i)}\sqrt{\mathbb{V}(Z_j)}} \tag{1.21}$$

where

$$\mathbb{V}(Z_i) = {\alpha_i}^2 \left(\sum_{\nu_C=1}^{k^C} \omega_{i\nu_C}^2 \mathbb{V}(R_{\nu_C}) + \sum_{\nu_I=1}^{k^I} \omega_{i\nu_I}^2 \mathbb{V}(R_{\nu_I}) + 2 \sum_{\nu_p=1}^{k^C} \sum_{\nu_q=1}^{k^I} \omega_{i\nu_p} \omega_{i\nu_q} Cov(R_{\nu_p}, R_{\nu_q}) \right) + \mathbb{V}(\epsilon_i) \tag{1.22}$$

and

$$Cov(Z_i, Z_j) =$$
$$\alpha_i \alpha_j \left(\sum_{\nu_C=1}^{k^C} \sum_{\nu_C=1}^{k^C} \omega_{i\nu_C} \omega_{j\nu_C} \mathbb{V}(R_{\nu_C}) + \sum_{\nu_I=1}^{k^I} \sum_{\nu_I=1}^{k^I} \omega_{i\nu_I} \omega_{j\nu_I} \mathbb{V}(R_{\nu_I}) \right.$$
$$\left. + \sum_{\nu_C=1}^{k^C} \sum_{\nu_I=1}^{k^I} \omega_{i\nu_C} \omega_{j\nu_I} Cov(R_{\nu_C}, R_{\nu_I}) + \sum_{\nu_C=1}^{k^C} \sum_{\nu_I=1}^{k^I} \omega_{i\nu_I} \omega_{j\nu_C} Cov(R_{\nu_C}, R_{\nu_I}) \right) \tag{1.23}$$

The specification of a correlation structure between systematic risk factors implies that in return we can deal with a correlation structure between asset returns. In order to relate this to default events correlations, we need to make an assumption regarding the dynamics of the asset returns, and subsequently and equivalently an assumption on systematic risk factors and idiosyncratic risk factors. The most common assumption is that asset returns follow a Gaussian distribution, so hereafter we will work under the assumption that asset returns are normally distributed [4].

1.5 Integration of the term structures of PDs in the current framework

1.5.1 Current framework

In the current framework, generally PDs are given for each buyers' group where buyers with similar characteristics are gathered. However there are buyers which, because of their importance in the portfolio, have their own given probability of default. Here, we will start by considering grading classes as groups of buyers and afterwards we will give a possible way of modelling the heterogeneity within a grading class.

In the current model, the probability of default for the rating class j, hereafter denoted p_j, is given by the formula:

$$\mathbb{P}(Z_j < d_j) = p_j \tag{1.24}$$

where Z_j is the ability-to-pay of the rating class j and d_j is a certain threshold such that if the ability-to-pay of a rating class Z_j falls beneath the threshold d_j, the rating class is considered as defaulted on its financial obligations. The ability-to-pay is a "latent variable", just like the asset return variable we defined before. Actually just the name differs here. This "latent variable" is called ability-to-pay in the case of credit insurance because the event of default occurs if a buyer is not able to pay its debt to its supplier (the policyholder). Thus the default event is defined in function of the ability of the buyer to pay the debt.

[4] The internal model deals with non-gaussian asset returns, see Decroocq, Planchet, Magnin (2009).

In the EH Internal Model, the ability-to-pay is given by a factor model, i.e. as a weighted sum of systematic risk factors, denoted R_ν, $\nu = 1, ..., k$ (the state of the economy) and an idiosyncratic risk denoted ε_j (proper to the rating class). The PD of the class j is then given by:

$$\mathbb{P}(Z_j < d_j) = \mathbb{P}(\sum_{\nu=1}^{k} w_{j\nu} R_\nu + \varepsilon_j < d_j) = p_j \tag{1.25}$$

The following assumptions are made about these two types of factors:

- The systematic risk factors vector ${}^t(R_1...R_k)$ follow a multinomial Gaussian distribution with parameters $\mathcal{N}(0, \Sigma)$.
- $\forall i \in J$, $\varepsilon_i \sim \mathcal{N}(0,1)$ and $\forall i \neq j$, $\varepsilon_i \perp \varepsilon_j$.

This definition of the ability-to-pay Z implies that it is a random variable and follows a Gaussian distribution.

For now, for each rating class, default probabilities are given for a horizon of one year. Afterwards they are used to calculate the thresholds d which will then intervene in simulation of default events. $\forall\ j$, d_j is given as a quantile of a Gaussian distribution with mean $\mathbb{E}(Z_j)$ and variance $\mathbb{V}(Z_j)$.

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(\sum_{\nu=1}^{k} w_{j\nu} R_\nu + \varepsilon_j) && (1.26)\\ &= \sum_{\nu=1}^{k} w_{j\nu} \mathbb{E}(R_\nu) + \mathbb{E}(\varepsilon_j) && (1.27)\\ &= 0 && (1.28) \end{aligned}$$

And the variance is equal to:

$$\begin{aligned} \mathbb{V}(Z_j) &= \sum_{\nu=1}^{k} \omega_{j\nu}^2 \mathbb{V}(R_\nu) + 2 \sum_{\nu_p=1}^{k-1} \sum_{\nu_q=\nu_p+1}^{k} \omega_{j\nu_p} \omega_{j\nu_q} Cov(R_{\nu_p}, R_{\nu_q}) && (1.29)\\ &= \sum_{\nu=1}^{k} \omega_{j\nu}^2 \sigma_{\nu_p\nu_p} + 2 \sum_{\nu_p=1}^{k-1} \sum_{\nu_q=\nu_p+1}^{k} \omega_{j\nu_p} \omega_{j\nu_q} \sigma_{\nu_p\nu_q} && (1.30) \end{aligned}$$

The default threshold for the grading class j is then computed by the formula:

$$d_j = \mathbb{E}(Z_j) + \sqrt{\mathbb{V}(Z_j)} \times N^{-1}(p_j) \tag{1.31}$$

where $N^{-1}(p_j)$ is the p_j-quantile of the standard normal distribution.

By simulation, the threshold d_j, together with other parameters such as granted limit, UGD, etc. will give the loss distribution. This makes possible the calculation of capital requirements as empirical quantiles of this distribution.

1.5.2 Integration of term structures of PDs

First method

Here we assume that we have the PD term structures.
Let us denote t_0 the current time. We are looking for the probabilities of default for time horizons $\{t_1, ..., t_N\}$.
For a rating class j, we denote the term structure of probabilities of default for this class $\{p_{jt_1}, ..., p_{jt_N}\}$.

In the current framework, *only* the ability-to-pay *for a year* is specified like in the formula 1.25. Indeed, up to now, only the probabilities of default for one year needed to be computed, so it was not necessary to make an hypothesis on the way the ability to pay changes over time.

The ability-to-pay, by definition, contains information on the correlation structure between the systematic risk factors, which then determines the correlation between abilities-to-pay of different buyers and consequently defaults of those buyers. As a beginning we assume that we do not have a dynamic of the correlations between systematic risk factors. For now, the variance-covariance matrix of the systematic risk factors is only given once a year. A way to circumvent the problem should be found.

We can use the definition of the ability-to-pay of one year and define similarly the ability-to-pay of the time period $[t_i, t_{i+1}]$ in this year. This is equivalent to assuming that the correlations between systematic risk factors are the same for periods of same length $t_{i+1} - t_i$, and thus that the abilities-to-pay during these periods of the same year follows the same distribution. Let's underline the fact that afterwards the definition of the default in one year will change and will not be the same as in the current framework.

So we will make the following assumption:
The correlation structure between systematic risk factors will be the same for periods of same length l during a year.

We consider the ability-to-pay for each time period. We will denote $Z_{j[t_i,t_{i+1}]}$ the ability-to-pay of the buyer j for the time period $[t_i, t_{i+1}]$.
We denote $R_{[t_i,t_{i+1}]\nu}$ the systematic risk factor ν for the time period $[t_i, t_{i+1}]$ and we assume that the systematic risk factors vector ${}^t(R_{[t_i,t_{i+1}]1}...R_{[t_i,t_{i+1}]k})$ follows a multinomial Gaussian distribution with parameters $\mathcal{N}(0, \Sigma_{[t_i,t_{i+1}]})$.
The assumption we made can formally be written as:

$$\forall i = 2, ..., N-1 \ \Sigma_{[t_{i-1},t_i]} = \Sigma_{[t_i,t_{i+1}]} = \Sigma \tag{1.32}$$

We will also assume that the systematic factors weights do not change with time.

$$\forall i = 2, ..., N-1 \ \omega_{[t_{i-1},t_i]} = \omega_{[t_i,t_{i+1}]} = \omega \tag{1.33}$$

This means that:

$$\forall i = 1, ..., N \ Z_{j[t_i,t_{i+1}]} \sim \mathcal{N}(0, \mathbb{V}(Z_j)) \tag{1.34}$$

where the variance $\mathbb{V}(Z_j)$ is given by the formula 4.6.

We denote $\left\{d_{j[t_1,t_2]},...,d_{j[t_{N-1},t_N]}\right\}$ the thresholds for the rating class j such that there is default at time period $[t_i,t_{i+1}]$ for $i=1,...N-1$ if the ability-to-pay of the rating class j falls below the threshold $d_{j[t_i,t_{i+1}]}$.
We denote $p_{j[t_i,t_{i+1}]}$ the probability a buyer **belonging to the rating class j in** t_0 has to default during the period of time $[t_i,t_{i+1}]$ for $i=1,...N-1$. So $p_{j[t_i,t_{i+1}]}$ is a forward default probability. The definition of the default probability at period $[t_i,t_{i+1}]$ is:

$$\forall\ i=1,...,N-1\ \mathbb{P}\left(Z_{j[t_i,t_{i+1}]}<d_{j[t_i,t_{i+1}]}\right)=p_{j[t_i,t_{i+1}]} \tag{1.35}$$

Since we claim that the abilities-to-pay for every period follow the same distribution (see formula 1.34), then the definition of the default probability during a period $[t_i,t_{i+1}]$ for $i=1,...,N-1$ becomes:

$$\mathbb{P}\left(Z_j<d_{j[t_i,t_{i+1}]}\right)=p_{j[t_i,t_{i+1}]} \tag{1.36}$$

Thus the probability of default until t_i for $i=1,...,N-1$ is:

$$\forall\ i=1,...,N-1$$
$$\mathbb{P}\left(\exists k\in\{1,...,i\}\ \text{ such that }\ Z_{j[t_k,t_{k+1}]}<d_{j[t_k,t_{k+1}]}\right)=p_{jt_i} \tag{1.37}$$

The term structure of PDs gives us the probabilities of default until a certain point in time, which we have denoted $\{p_{jt_1},...,p_{jt_N}\}$.
This term structure helps us define the probabilities of default for a buyer in the rating class j in t_0 during a certain time period:

$$p_{j[t_i,t_{i+1}]}=p_{jt_{i+1}}-p_{jt_i} \tag{1.38}$$

Since the abilities-to-pay of each time period follow the same normal distribution $\mathcal{N}(0,\mathbb{V}(Z_j))$, the thresholds $\left\{d_{j[t_1,t_2]},...,d_{j[t_{N-1},t_N]}\right\}$ can be computed by the formula:

$$\forall\ i=1,...,N-1\ d_{j[t_k,t_{k+1}]}=\mathbb{E}(Z_j)+\sqrt{\mathbb{V}(Z_j)}\times N^{-1}(p_{j[t_i,t_{i+1}]}) \tag{1.39}$$

If the economical situation does not change from one sub-period to another then there is no reason why the probability of default within a grade should change. So if the distribution of ability-to-pay for the rating class Z_j stays the same during all periods $[t_i,t_{i+1}]$ then the probability of default should stay the same. Anyway when it comes to practice, when we use default data, we observe that the probability of default is not the same during sub-periods of same length. This is due to the fact that the economical conditions do change.

How can we capture the way they change if we assume that the ability-to-pay has the same distribution? Default thresholds are the ones which get this variation of the economic conditions. However the economic variation effect may be blended with other effects, like for example the changing of the recovery rate of a firm when it defaults. When speaking about defaults in the case of financial products then the default threshold is considered to be a random variable because of the changing recovery rates.

The final aim is to get the most precise as possible loss distribution. Policy features change within a year, so they will be different for different time periods, *e.g.* renewal contracts. After computing default thresholds $\left\{d_{j[t_1,t_2]}, ..., d_{j[t_{N-1},t_N]}\right\}$ with the formula 1.39, loss simulations for each sub-period follow.

We simulate the loss at each time step, basing ourselves *on the data we have at* t_0 concerning each sub-period (how contract parameters change). So we will have the loss for each simulation of systematic risk factors and this for every time period $[t_i, t_{i+1}]$. Let us emphasize that at each time period we apply contract parameters like UGD, LGD[5], proper to the time period. For each simulation, the losses obtained at each period are summed and we have a annual loss simulation. Hence, by considering all the simulations, we have the distribution of the annual loss.

Second method

In the current model, the ability-to-pay in one year follows a normal distribution $\mathcal{N}(0, \mathbb{V}(Z_j))$. Let us divide a one-year period in N sub-periods $[t_i, t_{i+1}]$ with $i = 1, ..., N-1$, where t_N is the time denoting the end of the year. In order to give a dynamic of the ability-to-pay and to be coherent with the first assumption, i.e. $\mathcal{N}(0, \mathbb{V}(Z_j))$, we can assume that the ability-to-pay $(Z_{jt})_{t\in\mathbb{R}^+}$ is a Brownian motion with starting point $Z_{j0} = 0$ and $Z_{jt_N} \sim \mathcal{N}(0, \mathbb{V}(Z_j))$.

1. **Merton approach**
   We can follow a Merton approach and define a default event until t_i only at maturity time t_i. So we only consider the value of the ability-to-pay at the end of the period $[t_0, t_i]$, Z_{t_i}, and see if this value is above or below the default threshold at this time, denoted d_{t_i}. The probability of default until t_i for $i = 1, ..., N$ of a firm in the grading class j is:

   $$\mathbb{P}\left(Z_{jt_i} < d_{jt_i}\right) = p_{jt_i} \tag{1.40}$$

   We state that $t_N = 1$. So $Z_{t_i} \sim \mathcal{N}(0, t_i \times \mathbb{V}(Z_j))$ and then we will have:

   $$\begin{aligned} \mathbb{P}(Z_{jt_i} < d_{jt_i}) &= \mathbb{P}\left(\frac{Z_{jt_i}}{\sqrt{t_i \times \mathbb{V}(Z_j)}} < \frac{d_{jt_i}}{\sqrt{t_i \times \mathbb{V}(Z_j)}}\right) && (1.41)\\ &= N\left(\frac{d_{jt_i}}{\sqrt{t_i \times \mathbb{V}(Z_j)}}\right) && (1.42)\\ &= p_{jt_i} && (1.43) \end{aligned}$$

   So the thresholds can be computed as quantiles of a Gaussian distribution:

   $$\forall\ i = 1, ...N,\ d_{jt_i} = \sqrt{t_i \times \mathbb{V}(Z_j)} \times N^{-1}(p_{jt_i}) \tag{1.44}$$

   The default events will then be simulated by using these thresholds as they are simulated in the current framework.

[5] A new modelling of UGD should follow in order to keep up with the changes of the internal model due to the integration of term structures of PDs.

2. **First passage time approach**
   The Merton approach does not give correct default probabilities, because the default event might occur at any time between t_0 and t_i and not only at maturity. Therefore, the default probability until t_i should be the probability that the ability-to-pay process falls below a certain threshold d_{jt_i}, or equivalently the probability that the minimum of the ability-to-pay process is lower than a certain threshold d_{jt_i}. Formally we have:

$$\begin{aligned} \mathbb{P}\left(\exists t, 0 \leqslant t \leqslant t_i \text{ such that } Z_{jt} < d_{jt_i}\right) &= \mathbb{P}\left(\min_{0\leqslant t\leqslant t_i} Z_t < d_{jt_i}\right) && (1.45)\\ &= p_{jt_i} && (1.46) \end{aligned}$$

   By using a well-known result on the minimum of a Brownian motion, we have:

$$\begin{aligned} \mathbb{P}\left(\min_{0\leqslant t\leqslant t_i} Z_t < d_{jt_i}\right) &= 2N\left(\frac{d_{jt_i}}{\sqrt{t_i \times \mathbb{V}(Z_j)}}\right) && (1.47)\\ &= p_{jt_i} && (1.48) \end{aligned}$$

   Hence we have the following formula for the computation of the thresholds:

$$\forall\ i = 1, ...N,\ d_{jt_i} = \sqrt{t_i \times \mathbb{V}(Z_j)} \times N^{-1}(\frac{1}{2} p_{jt_i}) \tag{1.49}$$

   After computing the thresholds we should simulate the default events. In this case there is default in period $[t_0, t_i]$, $i = 1, ..., N-1$ if the ability-to-pay process falls below the threshold at any time between t_0 and t_i. This means that Brownian paths must be simulated and see for every time period $[t_0, t_i]$, $i = 1, ..., N-1$ if the Brownian path falls below the threshold d_{jt_i}. **This is why the simulation of the default events changes compared to the current framework**, which does not require the simulation of Brownian paths, but simulates values of the systematic risk factors.

In all cases, here we assume that the correlation structure will not change. We have to since we only have the covariance matrix Σ for the current year. A possible improvement is presented in the last chapter.

Chapter 2

Markov chains, transition and generator matrices

Credit risk models assume that a counterparty's rating migrates over a set of possible states. The credit migration process can be modelled as a finite Markov chain where ratings are the states of the chain and the rating of a company changes from one state to another with a certain probability. Those transition probabilities can be globally represented in a matrix, called the transition matrix. This chapter is aimed to give some basic definitions on Markov chains and present some theorems on the generators of such chains. Those concepts will be useful to the following chapter for the calibration of probabilities of default term structures.

2.1 Markov process and Markov chains

Definition 1 (Discrete Markov chain) *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. The values taken by the random variables form a countable or finite set of values denoted S. $(X_n)_{n\in\mathbb{N}}$ is a Markov chain if for all $n \in \mathbb{N}$ they satisfy the Markov property, namely:*

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n,\, X_{n-1} = x_{n-1},\ldots,\, X_0 = x_0)& \\ = \mathbb{P}(X_{n+1} = x_{n+1}|X_n = x_n)& \quad (2.1)\end{aligned}$$

This means that given the present state of the Markov chain, the future state does not depend on the past states of the chain.

Definition 2 (Markov process or continuous Markov chain) *A stochastic process $(X_t)_{t\in\mathbb{R}^+}$ is called a Markov process if for every $t > 0$ and $h > 0$:*

$$\mathbb{P}(X_{t+h} = y|X_s = x_s \text{ for all } s \leqslant t) = \mathbb{P}(X_{t+h} = y|X_t = x_t) \quad (2.2)$$

x_s for $\forall s < t$ is the history of states of the Markov process.

This means that the probability for the Markov chain to be in the state y at time $t+h$, conditioned on the history of states until t, is equal to the probability of having that state but this time conditioned on the state of the process in t.

Conditional on the present, future states are independent from past states.

In credit risk the rating migration process is supposed to be a Markov process. We will denote this process $(R_t)_{t\in\mathbb{R}^+}$. The states of this Markov chain are the rating classes.
In the credit migration process S will be the set of ratings so it will be a finite set of states. In what follows we suppose the companies can be classified in 8 rating classes, namely AAA, AA, A, BBB, BB, B, CCC, CC, C and D = default.

The question might rise: Is it appropriate to model the rating process as a Markov chain? Does the present rating of a company give all the necessary information to predict the future state or the rating path has a explanatory value which is not entirely contained in the present state? The assumption that the rating process is a Markov chain might be restrictive because most likely the rating history influences the future rating of a company. However in practice we observe that the goodness of fit of the Markov chain assumption to the rating process depends on the assumed type of Markov chain (homogeneous vs. non-homogeneous, discrete vs. continuous) and to which process this assumption is applied (differentiating the rating process in expansion and recession periods for example).

Definition 3 (Homogeneous Markov chain) *A Markov chain is homogeneous if the transition probabilities from one state to another do not depend on the time. Formally this means that:*

$$\forall n\ \mathbb{P}(X_{n+1} = i | X_n = j) = \mathbb{P}(X_1 = i | X_0 = j) \tag{2.3}$$

On the contrary, a non-homogeneous Markov chain is a Markov chain which is not homogeneous, which means that the transition probabilities from one state to another depend on time.

2.2 Transition matrix

Definition 4 (Transition matrix) *A transition matrix from t to $t+1$, also called rating migration matrix, is the matrix which gives for each rating class and conditional to the rating in time t, the probability of staying in the same rating class and the probabilities of moving to each of the other rating classes until $t+1$[1]. If we denote $M(t,t+1) = (m_{ij}(t,t+1))_{1\leqslant i\leqslant 8,\ 1\leqslant j\leqslant 8}$ the transition matrix then we have:*

$$\forall i,j\ m_{ij}(t,t+1) = \mathbb{P}(R_{t+1} = j | R_t = i) \tag{2.4}$$

The transition probabilities satisfy:

$$\sum_{j=1}^{8} m_{ij}(t,t+1) = 1 \text{ for } i = \{1,...,8\}$$

[1]In general the transition matrices considered in the literature are yearly, mainly because they use the transition matrices provided by rating agencies, which are given once a year for a period of one year.

$$m_{ij}(t, t+1) \geqslant 0 \text{ for } i, j = \{1, ..., 8\}$$

A matrix whose coefficients satisfy these two conditions is also called a *stochastic matrix*.
We denote $TM(n)$ the set of all transition matrices of type $n \times n$. Let M be the transition matrix for one period: $M = M(0, 1)$.

Remark 5 *The transition matrix of a homogeneous Markov chain does not depend on time and we have:*

$$\forall n \ M(0, n) = M^n$$

2.2.1 Features of a transition matrix

Here is an example of a transition matrix:

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	91,62	7,63	0,53	0,06	0,08	0,03	0,06	0
AA	0,58	90,88	7,79	0,54	0,06	0,09	0,03	0,03
A	0,04	2,04	91,91	5,35	0,4	0,16	0,03	0,08
BBB	0,01	0,15	3,87	90,88	4	0,69	0,16	0,24
BB	0,02	0,05	0,19	5,3	85,42	7,22	0,8	0,99
B	0	0,05	0,15	0,26	5,68	85,02	4,34	4,51
CCC	0	0	0,23	0,34	0,97	11,84	60,96	25,67
Default	0	0	0	0	0	0	0	100

Tab 2.1: Average annual transition matrix from SP, average calculated on 1981-2008

We notice that the last row displays a different characteristic compared to the other rows. According to the definition of the transition matrix the last row is made of the probabilities that a company which has defaulted in t moves to the other rating classes until $t + 1$. We can see that the probability for a defaulted company to move to a non-default class is 0 and the probability of staying in the default class is 1. The last row of the great majority, not to say all, the transition matrices met in the literature is the same, $m_{8j}(t, t + 1) = 0$ for $j = 1, ..., 7$ and $m_{88}(t, t + 1) = 1$. The reason is that default is considered to be an absorbing state; if a company defaults then its rating will not improve during any of the following periods.

As expected default probabilities are higher for lower grades. We can even say that default probabilities increase exponentially with lower ratings, as we can see in the figure below. The likelihood of staying in the same rating class is very high compared to probabilities of rating migrations. If a company does not stay in the same class, it will more probably move to a neighbour rating class rather than jump to a further one. In general the further the rating class is from the actual rating, the lower the probability of migration to this class is, i.e. the further the coefficients of the transition matrix are from the diagonal, the lower they are. This property is known as the "monotonicity" property. However it is not always true. For example for the matrix above the probability of default for a company rated A, BBB, BB or B is higher than the probability of migrating to the immediate lower class. Similarly it is more probable for a company rated AAA to fall to rating BB than to BBB.

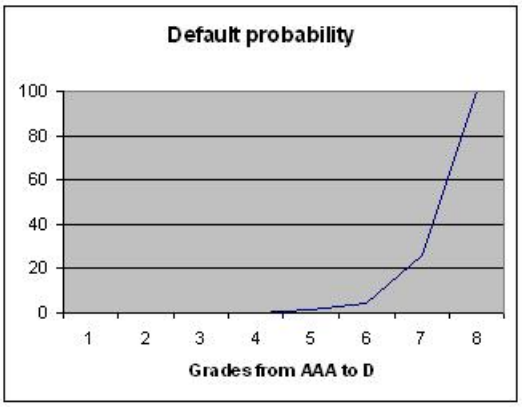


Fig 2.1: Default probabilities of the transition matrix above

2.2.2 How to take account of rating withdrawals?

Transition matrices are obtained by counting the number of transitions among a set of credit states for a given pool of companies over a one-year period. We should note that the published data are not complete because information is lost about companies that were withdrawn from the rating pool due to mergers, repayment of their debt, calling of the debt, etc. This is the reason why in general the row-sums of published transition matrices might be different from 1. Hence, certain assumptions need to be made about the companies that have been removed from the sample and adjust the matrix accordingly. The class of rating withdrawals is also called the "Not Rated" (NR) class. Transitions to this class might be "benign" or "bad". Bad transitions may be due for example to a credit quality deterioration known by the debtor only and this leads the company to bypass a rating agency.

There are at least four methods for removing NRs from the dataset.

- The first method is conservative and proceeds by treating transitions to NR as negative information regarding the change in credit quality of the borrower. Here the probability of transiting to NR is distributed amongst downgraded and defaulted states in proportion to their values by allocating NR values to all cells to the right of diagonal. If we denote $M^{'} = (m^{'}_{ij})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant n}$ then this methods correspond to doing the following adjustment for each row $i = 1, ..., n$:
  - $\forall j \leqslant i \; m^{'}_{ij} = m_{ij}$
  - $\forall j > i \; m^{'}_{ij} = m_{ij} + m_{ij} \dfrac{1 - \sum_{j=1}^{n} m_{ij}}{\sum_{j>i} m_{ij}}$
- The second method is liberal and treats transitions to NR status as benign. The transition probabilities to NR are distributed among all states, except default, in proportion to their values. This is achieved by allocating the probability of transiting to NR to all but the default column. $\forall i = 1, ..., n$
  - $\forall j < n \; m^{'}_{ij} = m_{ij} + m_{ij} \dfrac{1 - \sum_{j=1}^{n} m_{ij}}{\sum_{j=1}^{n-1} m_{ij}}$

– for $j = n$ $m_{ij}^{'} = m_{ij} + m_{ij}\dfrac{1 - \sum_{j=1}^{n} m_{ij}}{\sum_{j>i} m_{ij}}$

- The third method, which has emerged as an industry standard, treats transitions to NR status as non-information. The probability of transitions to NR is distributed among all states in proportion to their values. This is achieved by gradually eliminating companies whose ratings are withdrawn. This method modifies the default probabilities which will then comprise a part of uncertain information. $\forall i = 1, ..., n$

  – $\forall 1 \leqslant j \leqslant n$ $m_{ij}^{'} = m_{ij} + m_{ij}\dfrac{1 - \sum_{j=1}^{n} m_{ij}}{\sum_{j=1}^{n} m_{ij}}$

- A fourth method would consist in considering that the future state of all "not rated" companies would in fact be the same as the current state.

$$\forall i, j = 1, ..., n \; m_{ii}^{'} = m_{ii} + (1 - \sum_{j=1}^{n} m_{ij}) \tag{2.5}$$

2.3 Generator matrices

Definition 6 (Generator matrix) *A Generator matrix* $Q = (q_{ij})_{1 \leqslant i \leqslant 8,\ 1 \leqslant j \leqslant 8}$ *is a matrix whose elements satisfy:*

$$\sum_{j=1}^{8} q_{ij} = 0 \text{ for } i = \{1, ..., 8\}$$

$$q_{ij} \geqslant 0 \text{ for } i, j = \{1, ..., 8\} \text{ and } i \neq j$$

Let $\mathbf{G} \in \mathbf{M}_{8,8}(\mathbb{R})$ be the set of generator matrices.

Proposition 7 *Generators together with the binary operator "+" form a semigroup in the set of matrices.*

This means that the sum of two or more generators is still a generator. The proof is trivial.
The simplest generator associated to a transition matrix M would be $M - I$ where I is the identity matrix of type 8×8.

Proposition 8 *A matrix* Q *is a generator if and only if* $M(t) = \exp(tQ)$ *is a transition matrix.*

Proof
As $t \longrightarrow 0^{+}$ then we have:

$$M(t) = I + tQ + O(t^2) \tag{2.6}$$

$q_{ij} \geqslant 0$ for all $i \neq j$ if and only if $m_{ij} \geqslant 0$ for all i, j and for $t \geqslant 0$ sufficiently small. If M is the transition matrix of a homogeneous Markov chain then $M(t) = M(\dfrac{t}{n})^n$ for all $n \in \mathbb{N}$. So $q_{ij} \geqslant 0$ for all $i \neq j$ if and only if $m_{ij} \geqslant 0$

for all i, j and for all $t \leqslant 0$.
If Q has row sums equal to 0 then so does Q^n for all $n \in \mathbb{N}$.

$$\sum_{j=1}^{n} q_{ij}^{n} = \sum_{j=1}^{n} \sum_{k=1}^{n} q_{ik}^{n-1} q_{kj} = \sum_{k=1}^{n} q_{ik}^{n-1} \sum_{j=1}^{n} q_{kj} = 0 \tag{2.7}$$

Using the Taylor series for the matrix exponential we have:

$$M(t) = \exp(tQ) = \sum_{k=0}^{+\infty} \frac{(tQ)^k}{k} = I + tQ + \frac{(tQ)^2}{2!} + \frac{(tQ)^3}{3!} + ... \tag{2.8}$$

By 2.7 we have that the sum of the row coefficients for the Q polynomials is equal to 0. So the row sum of M(t) is equal to the row sum of I and $\sum_{j=1}^{n} m_{ij} = 1$. On the other hand if M(t) is a transition matrix then the row sums of $tQ + \frac{(tQ)^2}{2!} + \frac{(tQ)^3}{3!} + ...$ must be equal to 0 for all $t \geqslant 0$. For this to be possible the row sum of Q must be equal to 0.

□

Definition 9 (Generator of a Markov process) *Let M(t) be the transition matrix of the Markov process $(R_t)_{t\in\mathbb{R}}$. The matrix Q is called the generator of $(R_t)_{t\in\mathbb{R}}$ if M(t) satisfies:*

$$\frac{d}{dt} M = MQ \tag{2.9}$$

In this case we obtain

$$M(t) = \exp(tQ),\ t \geqslant 0 \tag{2.10}$$

2.4 Embedding and identification problems

Given a finite Markov chain, $(R_n)_{n\in\mathbb{N}}$, the *embedding* problem is posed as follows:

Is it possible to construct a Markov process $(R_t)_{t\in\mathbb{R}}$ in continuous time such that the probability distribution of $(R_t)_{t\in\mathbb{R}}$ at time $t = 1, 2, ...$ is identical to the distribution of $(R_n)_{n\in\mathbb{N}}$?

This is equivalent to determining if a transition matrix is compatible with a true generator Q such that:

$$M = \exp(Q)$$

If the matrix Q exists and is unique, then a continuous-time extention of the transition matrix $\tilde{M}(t)$ can be defined as follows:

$$\tilde{M}(t) = \exp(tQ) \tag{2.11}$$

The *identification* problem consists in looking for the true generator once its existence is established. A transition matrix can have many generators and one must choose between those generators the one that best applies to the problem. Here follow two theorems which give a partial answer to the uniqueness problem of a generator.
Let's define

$$S = \max\{(a-1)^2 + b^2; a + ib \text{ is a complex eigenvalue of M }, a, b \in \mathbb{R}\}.$$

Theorem 10 *Let M be a transition matrix and suppose that $S < 1$. Then the Taylor series for the matrix logarithm*

$$Q = \sum_{k=0}^{+\infty} (-1)^k \frac{(M-I)^k}{k!} = (M-I) - \frac{(M-I)^2}{2!} + \frac{(M-I)^3}{3!} - ... \tag{2.12}$$

converges geometrically quickly (absolute convergence), and gives rise to a matrix Q of same type as M having row-sums equal to 0 such that $\exp(Q) = P$ exactly.

Proof

The numerical radius of a matrix A is given by:

$$\rho(A) = \max\{|\ \lambda\ |: \lambda \in \sigma(A)\}$$

where $\sigma(A)$ is the spectrum (set of eigenvalues) of A. Let's note that if $a + ib$ is the eigenvalue for a matrix M associated to the vector $\mathbf{x}$ then $(a-1) + ib$ is an eigenvalue for $M - I$ associated to the same vector $\mathbf{x}$. Indeed $(M-I)\mathbf{x} = M\mathbf{x} - I\mathbf{x} = (a+ib)\mathbf{x} - \mathbf{x} = [(a-1) + ib]\mathbf{x}$. Thus $S = \rho(M-I)^2$ since $|\ (a-1) + ib\ |^2 = (a-1)^2 + b^2$.

By the *spectral radius formula*

$$\lim_{k \longrightarrow +\infty} \|\ (M-I)^k\ \| = \rho(M-I)^k = S^{\frac{k}{2}} \tag{2.13}$$

If $S < 1$ the series in 2.12 converges geometrically quickly, and converges absolutely. The Taylor *log* expansion for matrices is:

$$\log(M) = \sum_{k=0}^{+\infty} (-1)^k \frac{(M-I)^k}{k!} = (M-I) - \frac{(M-I)^2}{2!} + \frac{(M-I)^3}{3!} - ...$$

where $\|\ .\ \|$ is the operator norm[2]. Since this series in 2.12 converges and this series is equal to Q, we conclude that $Q = \log(M-I)$. We should now prove that the row-sums of Q are equal to 0.

[2]The operator norm $\|\ A\ \|$ of a linear operator $A : V \longrightarrow W$ where V and W are two vector spaces is defined as $\|\ A\ \| = \min\{c : \|\ Av\ \| \leqslant c\ \|\ v\ \| \ \forall v \in V\}$

Lemma 1 *Let A and B be $n \times n$ matrices. Suppose that A has row-sums α and B has row-sums β. Then $C = AB$ has row-sums $\alpha\beta$.*

Proof Lemma

$$\sum_{j=1}^{n} c_{ij} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n} a_{ik}\sum_{j=1}^{n} b_{kj} = \sum_{k=1}^{n} a_{ik}\beta = \beta\sum_{k=1}^{n} a_{ik} = \beta\alpha \quad (2.14)$$

□

Since $M-I$ has row-sums equal to 0, the lemma proves that $\forall k > 0$ $(M-I)^k$ has row-sums equal to 0. Since the series 2.12 is a polynomial of $(M-I)$, Q which is defined as this series has row-sums 0.

□

Definition 11 (Strictly diagonally dominant matrix) *A matrix M is strictly diagonally dominant if its diagonal entries are greater than $\frac{1}{2}$, i.e., $m_{ii} > \frac{1}{2}$ $\forall i$.*

Theorem 12 *If a transition matrix M is strictly diagonally dominant then $S < 1$, which implies that the convergence of the series in theorem 10 is guaranteed. If the generator exists then it is unique*[3]*.*

Proof

Let $m = \min\{m_{ii}\}$. Since according to the assumption M is strictly diagonally dominant then $m > \frac{1}{2}$. Let's write

$$M = mI + (1-m)R,$$

where $R = \frac{1}{1-m}(M - mI)$. Then R is also a transition matrix (row-sums equal to 1 and positive coefficients). We also have:

$$M - I = (1-m)(R-I).$$

Since R is a transition matrix, $\| R \| \leqslant 1$, so that $\| R - I \| \leqslant 2$ by the triangle inequality, i.e. $\| R - I \| \leqslant \| R \| + \| I \| = 1 + 1 = 2$ and $\| (R-I)^k \| \leqslant 2^k$. Hence $\| (M-I)^k \| \leqslant (2-2m)^k$. By the spectral radius formula we have:

$$\rho(M-I) = S^{\frac{1}{2}} = \lim_{k \longrightarrow +\infty} \| (M-I)^k \|^{\frac{1}{2}} \leqslant \lim_{k \longrightarrow +\infty} \{(2-2m)^k\}^{\frac{1}{2}} = 2 - 2m.$$

Since, $m > \frac{1}{2}$, $2 - 2m < 1$, and so $\sqrt{S} < 1$, and $S < 1$.

□

[3]The proof of the second part of the theorem is not presented here

The theorem 10 does not claim that if the series of ln converges than a valid generator exists. It just announces that if this series converges then matrix obtained has row-sums equal to 0. The non-diagonal coefficients of this matrix must be non negative for it to be a valid generator.
Even though the series does not converge the existence of a valid generator is not excluded.

The following theorem gives 3 conditions under which an exact generator does not exist.

Definition 13 (A state accessible from another state) *A state j is accessible from a state i if there is a sequence of states $k_0 = i$, $k_1, k_2, ..., k_m = j$ such that*

$$\prod_{l=0}^{m-1} m_{k_l k_{l+1}} > 0$$

which is equivalent to :

$$m_{k_l k_{l+1}} > 0 \text{ for each } l$$

Theorem 14 *If the transition matrix M satisfies one of the following conditions:*

1. $\det M < 0$,
2. $\det M > \prod_{i=1}^{n} m_{ii}$,
3. *there are states i and j such that j is accessible from i, but $m_{ij} = 0$,*

then there does not exist an exact generator for M.

Proof

1. We know that for a real matrix Q:

$$\det(\exp(Q)) = \exp((Q)) \tag{2.15}$$

Indeed, there exists a basis of eigenvectors of Q where Q is upper-diagonal. Let P be the eigenvectors matrix and T the triangular matrix similar to Q such that:

$$Q = PTP^{-1}$$

Since for two similar matrices have the same determinant $\det Q = \det T$. We also have:

$$\exp(Q) = \sum_{k=0}^{+\infty} \frac{(Q)^k}{k!} = \sum_{k=0}^{+\infty} \frac{(PTP^{-1})^k}{k!} = \sum_{k=0}^{+\infty} P\frac{(T)^k}{k!}P^{-1} = P\exp(T)P^{-1} \tag{2.16}$$

$\exp(Q)$ and $P\exp(T)P^{-1}$ are similar too, so $\det(\exp(Q)) = \det(\exp(T))$.
Since T is a triangular matrix, $\det(\exp(Q)) = \det(\exp(T)) = \exp((Q))$.

If $M = \exp(Q)$ for some matrix Q, wee must have

$$\det(M) = \det(\exp(Q)) = \exp((Q)) > 0.$$

2. We suppose that M has a generator Q. Let $R(t) = \exp(tQ)$. Then $R_{ii} > Q_{ii}R_{ii}$ and $R_{ii}(0) = 1$, so $R_{ii}(t) \geqslant \exp(tQ_{ii})$. Hence $m_{ii} = R_{ii}(i) \geqslant \exp(Q_{ii})$. Using 2.15 we have:

$$\prod_{i=1}^{n} m_{ii} \geqslant \prod_{i=1}^{n} \exp(Q_{ii}) = \exp(\sum_{i=1}^{k} Q_{ii}) = \exp(trace(Q)) = \det(M), \quad (2.17)$$

contradicting condition **2**. Hence assuming condition **2** there is no such generator.

3. It follows from the Lévy Dichotomy.

The Lévy dichotomy states that if a transition matrix M has a proper generator Q then for each pair (i, j) of states we must have either $m_{ij} > 0$ for all $t > 0$ or $m_{ij} = 0$ for all $t > 0$ (where $p_{ij}(t)$ is the ij entry of $M(0, t)$).

Proof of the Levy dichotomy

Suppose that M has a generator. For each state $k = 1, ..., n$ we would have $m_{kk}(s) \longrightarrow 1$ as $s \longrightarrow 0$, so that for sufficiently small s we would have $m_{kk}(s) > 0$ for all states k. If $m_{ij}(t) = 0$ for some $t > 0$, then we must have $m_{ij}(\frac{t}{n}) = 0$ for all sufficiently large integers because otherwise we would have $m_{ij}(t) \geqslant m_{ij}(\frac{t}{n})(m_{jj}(\frac{t}{n}))^{n-1} > 0$. That is the set of zeros of the function $m_{ij}(s)$ has a limit point 0. $m_{ij}(s)$ is an analytic function of s, hence it must be that $m_{ij}(s) = 0$ for all $s > 0$, contradicting our assumption that $m_{ij}(t) = 0$ for some $t > 0$.

Suppose that j is accessible from i. Then we have $m_{ij}(s) > 0$ for some integer s. Hence we must have $m_{ij}(t) > 0$ for all t and in particular $m_{ij}(1) = m_{ij} > 0$ as claimed.

□

The third condition is satisfied for the majority of transition matrices. Actually in the most of cases an exact generator does not exist and thus the need for regularization algorithms emerges.

2.5 Regularization of the generator problem

We would want to find transition matrices for periods shorter than one year so that when raised to a power n we obtain the best approximate the annual transition matrix. Formally this means that we look for $\tilde{X} \in TM(n)$ such that:

$$\| \tilde{X}^n - M \| = \min_{X \in TM(n)} \| X^n - M \| \quad (2.18)$$

where $\| . \|$ is a suitable norm in the space of $n \times n$ matrices.

Since $\tilde{X}$ is raised to a power greater than one, this problem is high dimensional, constrained non-linear optimization problem whose solution is computationally intensive (Kreinin and Sidelnikova [2001]).

2.5.1 Quasi-optimization of the generator

For time continuous Markov chains this problem can be solved by an heuristic approach whose object of regularization is the generator. This method is called the *Quasi-optimization of the generator*.
The problem *Quasi-optimization of the generator* (QOG) is specified as follows: Find Q^* such that

$$\| Q^* - \ln(M) \| = \min_{X \in Q(n)} \| X - \ln(M) \| \tag{2.19}$$

The space of the generator matrices, $Q(n)$, is a Cartesian product of n-dimensional cones. Each row of a generator has the property that its row-sums equal to 0 and non-diagonal elements are non-negative. By permuting the elements of each row, they can be represented as a point in a standard cone, $K(n)$:

$$K(n) = \{(x_1, ..., x_n) \in \mathbb{R}^n, \sum_{i=1}^{n} x_i = 0,\ x_1 \leqslant 0, x_i \geqslant 0 \text{ for } \forall i \geqslant 2\} \tag{2.20}$$

$K(n)$ is contained in the hyperplane $\hat{H}(n)$:

$$\hat{H}(n) = \{(x_1, ..., x_n) \in \mathbb{R}^n, \sum_{i=1}^{n} x_i = 0\} \tag{2.21}$$

This problem can be solved on a row-by-row basis by projecting a point $\mathbf{m} \in \mathbb{R}$, where m is a row of the matrix $\ln(M)$ onto the cone $K(n)$. The problem QOG can be reduced to n independent instances of the following distance minimization problem:
Distance minimization problem for the generator (DMPG):
For a given point $\boldsymbol{m} \in \mathbb{R}, \boldsymbol{m} = (m_1, ..., m_n)$, *find* $\boldsymbol{q^*} \in K(n)$ *such that*

$$\text{dist}(\mathbf{m}, \mathbf{q^*}) = \min_{\mathbf{q} \in K(n)} \text{dist}(\mathbf{m}, \mathbf{q}) \tag{2.22}$$

The optimal solution to problem $DMPG$ can be obtained as follows:

Step 1. Let $\mathbf{b}$ be the projection of $\mathbf{m}$ on the hyperplane $\hat{H}(n)$. For this, set

$$b_i = m_i - \lambda$$

where

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} m_i.$$

Step 2. Let $\hat{m} = \pi(\mathbf{b})$, where π is a permutation that orders the coordinates of $\mathbf{b}$ in a descending sequence.

Step 3. Find l^*, the smallest integer $2 \leqslant l \leqslant n-1$ that satisfies

$$(n - l - 1)\hat{m}_{l+1} \geqslant \hat{m}_l + \sum_{i=0}^{n-(l+1)} \hat{m}_{n-i}. \tag{2.23}$$

Step 4. Let $\Im = \{i : 2 \leqslant i \leqslant l^*\}$. Construct the vector $\hat{\mathbf{q}} \in K(n)$ as follows.

$$q_i^* = \begin{cases} 0 & \forall \mathrm{i} \in \Im \\ \hat{m_i} - \dfrac{1}{n - l^* + 1} \sum_{j \notin \Im} \hat{m_j} & \forall i \notin \Im \end{cases} \tag{2.24}$$

Step 5. Apply the inverse permutation π^{-1} to $\hat{\mathbf{q}}$; $\pi^{-1}(\hat{\mathbf{q}})$ is the solution to the problem $DMPM$.

2.5.2 Other regularization methods

These methods adjust the matrix obtained as the Taylor series expansion of the logarithm $Q = \ln(M)$, in order to construct a valid generator $\tilde{Q}$. For the two methods presented below, namely the *diagonal adjustment* and the *weighted adjustment*, the negative non-diagonal elements are set to 0 and then an adjustment is made so that the row-sum equals 0. The first step is the same for both algorithms. The computation of Q^* is made as follows:

Step 1. For i,j = 1, ..., n, set

$$q_{ij}^* = \begin{cases} 0 & \text{if } i \neq j \text{ and } q_{ij} < 0 \\ q_{ij} & \text{otherwise} \end{cases} \tag{2.25}$$

Step 2a. (diagonal adjustment) Set the diagonal elements to the negative sum of the non-diagonal elements:

$$q_{ii}^* = -\sum_{j=1, j \neq i}^{n} q_{ij} \text{ for } i = 1, 2, ..., n \tag{2.26}$$

Step 2b. (weighted adjustment) Adjust the non-zero elements according to their relative magnitudes:

$$q_{ij}^* = q_{ij} - \mid q_{ij} \mid \frac{\sum_{j=1}^{n} q_{ij}}{\sum_{j=1}^{n} \mid q_{ij} \mid} \text{ for } i, j = 1, 2, ..., n \tag{2.27}$$

Chapter 3

Calibration of PD term structures

This chapter will present two possible ways of introducing a term structure of default probabilities in the current internal model of EH. This project is defined in the continuity of the article Decroocq, Planchet and Magnin [2009].

3.1 The need to introduce term structures of PDs

For a given rating class, why should default probabilities vary within a year? Why should for example the PD of a given company in the first semester be higher or lower than the PD in the second semester of the year? Intuitively we would think that this may either be the consequence of an internal change in the management of the firm or the consequence of a shift in the economic environment - the economy is doing either better or worse than in the first semester and it has an impact on the financial situation of the firm. We can take account of the latter by changing the parameters of the systematic risk factors distribution, i.e. changing the variance and correlation coefficients. This would then imply a change in the correlation structure between PDs of different rating classes, since this correlation is the direct consequence of the correlation between the systematic risk factors which are common to the abilities-to-pay of all rating classes. We will try to capture this economic cycle effect and predict how PDs vary when the economy is doing well or during a downturn. This means that we will introduce a multi-state approach which will be particularly interesting during dramatic economic changes, like the current crisis for instance, which cause the PDs change.
There are several articles (e.g. Bangia et al.[2000], Jones[2005]) which give evidence for the relevance of this approach to the aim of having a term structure for the PDs. In those articles it is assumed that the rating migration process is a homogeneous Markov chain.
Another possible approach would be calibrating a term structure of PDs on historical data over several time periods. In this case the assumption of a homogeneous Markov chain proves restrictive and does not give a good fit to ob-

served data. On the other hand if we suppose that the credit migration process is a non-homogeneous continuous time Markov chain the resulting calibration of the term structure of PDs is highly satisfactory. Thus we do the calibration of term structures of PDs for each grade under this assumption. However this is done by using average data over a certain time period. During an economic downturn the PDs will increase but this will only have a marginal effect on the average PDs used for the calibration. So we should introduce a way of taking into consideration cycle phases more explicitly.
First we will introduce non-homogeneous continuous time Markov chains and how a term structure of PDs can be modelled if we suppose that the credit migration process follows such a process. We also present the calibration of term structures of PDs. Concerning this approach we do not take into account the economic cycle phases yet. Afterwards we will present the approach of Bangia et al. and remark the existence of two economic cycles which impact the credit migration process in different ways.

3.2 Credit migration as a non-homogeneous continuous time Markov chain

In this section we will define PD term structures for each rating class and we will do this by implementing the approach presented in Bluhm and Overbeck [2007] (hereafter [2]). This article supposes the rating migration process is a non-homogeneous Markov chain, so we will start by the definition of this process.

Definition 15 (Non-homogeneous continuous time Markov chain) *The process $(R_t)_{t\in\mathbb{R}^+}$ is a non-homogeneous continuous time Markov chain if its generator is time-dependent and is given by:*

$$Q_t = \Phi(t)Q \tag{3.1}$$

where $\Phi(t) = (\varphi_{ij}(t))_{1\leq i\leq n, 1\leq j\leq n}$ is a diagonal matrix and its diagonal coefficients are functions of time and depend on some parameter values.

In [2], $\varphi_{ij}(t)$ depend on some parameters α and β and are defined as follows:

$$\varphi_{ij}(t) = \begin{cases} \dfrac{(1-\exp(-\alpha_i t))t^{\beta_i - 1}}{1-\exp(-\alpha_i)} & \text{if i=j} \\ 0 & \text{if else} \end{cases} \tag{3.2}$$

φ was chosen of this form because of its properties which are:

1. $\varphi(1) = 1$ this allows to have $M = \exp(Q)$ which should still stay the same after the modification to a non-homogeneous Markov chain.
2. $t\varphi(t)$ is increasing in the time parameter $t \geqslant 0$. This is necessary to have a non decreasing probability of default when the time horizon becomes longer.

3. In the numerator of $t\varphi(t)$, the first factor $(1-\exp(-\alpha t))$ is the distribution function of an exponentially distributed random variable with intensity α; the second factor t^{β} can be considered as a convexity or concavity adjustment term.

This function proves to be sufficiently reasonable to be applied as a modification of the generator Q and the calibration of the term structure on historical data is very satisfactory.

3.2.1 How to calibrate PD term structures?

Calibration using SP credit migration data from 1981-2005

First we need a historical migration data which will be the basis of the calibration. Thanks to this data an average one-year transition matrix is calculated. We also need historical default data for different time horizons. In order to check our results we apply the $C++$ program to the data of Standard and Poors (SP) which are used in [2]. The default probabilities are given for a horizon of 1 year up to 15 years.
The one-year transition matrix is the following:

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	91.68	7.69	0.48	0.09	0.06	0.00	0.00	0.00
AA	0.62	90.49	8.10	0.60	0.05	0.11	0.02	0.01
A	0.05	2.16	91.34	5.77	0.44	0.17	0.03	0.04
BBB	0.02	0.22	4.07	89.72	4.68	0.80	0.20	0.29
BB	0.04	0.08	0.36	5.78	83.38	8.05	1.03	1.28
B	0.00	0.07	0.22	0.32	5.84	82.53	4.78	6.24
CCC	0.09	0.00	0.36	0.45	1.52	11.17	54.06	32.35
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Note: based on S&P data (Standard & Poor's, 2005)

Fig. 3.1 : One-year transition (credit migration) matrix (SP 2005)

Generator matrix and regularization We need to estimate the generator of the rating migration process i.e. Q such that $M = \exp(Q)$, where M is the average one-year transition matrix. In the second chapter we discussed about the embedding and identification problems. Does an exact generator of the rating migration process exist? In this particular case we see that the transition matrix is strictly diagonally dominant (the diagonal coefficients are greater than $\frac{1}{2}$), therefore we can use theorem 12 and calculate the matrix Q as the Taylor expansion for the logarithm of a matrix. In the appendix, we present the matrix logarithm. The theorem 12 does not assure that Q is a valid generator. In fact we can see that there are negative non-diagonal elements. The matrix Q needs to be regularized in order to obtain an approximative valid generator Q^*.

We have implemented the QOG algorithm in C++ inspiring ourselves from the program developed in the framework of the article "Evolutionary models for insertions and deletions in a probabilistic modelling framework" from E. Rivas.

The approximative generator matrix Q^* of the rating migration process is given in the appendix. **We observe that the approximative generator we find is quite similar to the one found in the article.** However the probable roundings might cause the observed difference in some coefficients.

Calibration of PD term structures After the computation of the generator Q^*, the calibration of the parameter vectors of the function φ, i.e. $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8\}$ and
$\beta = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8\}$ follows[1]. The best possible way to calibrate would be finding the parameters vectors α and β that minimize the distance:

$$\sum_{t \in L} \parallel \hat{M}_t - exp(t\Phi(t)Q) \parallel^2 \tag{3.3}$$

where $\hat{M}_t$ is the transition matrix for the time t calculated using the credit migration data and L is the set of period lengths available. Here $L = \{1, ..., 15\}$. However, this would be complicated if there are too many time period lengths, because the number of coefficients in each transition matrix M_t is 64 (8×8), so for 15 periods of time, we would have to calibrate by taking into account 64x15=960 coefficients! For this reason, we will do the calibration of α and β by taking into consideration the probabilities of default for each horizon. Probabilities of default are the most valuable information in a transition matrix and moreover we are looking for a term structure of probabilities of default, so calibrating only on PDs seems sensible. Let remind us that the PDs for a horizon t can be found at the 8^{th} columns of the transition matrix M_t. The minimization problem to be solved in order to find the parameter vectors α and β would then be:

$$\sum_{t \in L} \parallel (\hat{M}_t)_{column(8)} - (\exp(tQ_t))_{column(8)} \parallel^2 \tag{3.4}$$

The Levenberg-Marquardt algorithm was chosen to solve the minimization problem. The description and the algorithm itself are given at the end of this section. The minimization is done under box constraints since the following assumption is made for α and β:

$$\forall i \in \{1, ..., 8\},\ \alpha_i \geqslant 0 \text{ and } \beta_i \geqslant 0$$

We have implemented the calibration of the parameter vectors α and β. The probabilities of default resulting after the calibration for the rating class j can be obtained by the formula:

$$\pi_{t,j} = \exp\left(t\mathbf{Q}_t\right)_{j,\text{column(8)}} \tag{3.5}$$

where $\mathbf{Q}$ is the generator matrix computed with the parameter vectors which minimize the distance in 3.4. The fit of modelled PDs to observed PDs and the

[1] Actually, only the calibration of the first 7 parameters of α and of β is made and the 8^{th} parameter can be fixed at an arbitrary value because the 8^{th} row coefficients of Q which will be multiplied by those parameters are equal to zero.

parameter vectors obtained depend on the starting point of the algorithm, even though not at a significant degree. On the contrary, we observe that they both depend more on the upper bound which must be specified before running the Levenberg-Marquardt algorithm. The best fit is obtained by specifying initial parameter values = 0.4 for all α and β coefficients and an upper bound of 6. In this case we have the following parameters:

	α	β
AAA	0,294	0,678
AA	0,289	0,505
A	0,300	0,377
BBB	0,702	0,632
BB	0,291	0,526
B	0,142	0,281
CCC	5,977	0,466

Tab. 3.1: Parameters resulting from the calibration

In [2] the following parameter vectors are found.

	α	β
AAA	0,340	0,890
AA	0,110	0,260
A	0,810	0,650
BBB	0,230	0,300
BB	0,320	0,560
B	0,230	0,400
CCC	2,150	0,460

Tab. 3.2: Parameter calibration in [2]

The parameters in Tab. 3.1 are different from those in Tab. 3.2. This might be due to the minimization algorithm used by in [2].
It might be interesting noticing that the error made in this case (with starting points equal to 0.4) is lower than the one made when the starting points are the minimization results in [2], given in Tab. 2.3. With the parameters of Tab. 3.2. the error is equal to

$$\sqrt{\sum_{t=1}^{15} \parallel \hat{M}_{t_{column(8)}} - exp(tQ_t)_{column(8)} \parallel^2} \simeq 0,05527$$

and with the starting point in Tab. 3.3 the error is equal to $0,05543$. Moreover in the [2] the error is equal to $0,12811$, higher than the error in our case. Our calibration method seems to be more efficient than the one used in [2]. In order to compare the fitted PDs with the historical ones, we present them in the following graphics. The historical observed PDs are in blue and the modelled PDs are in purple.

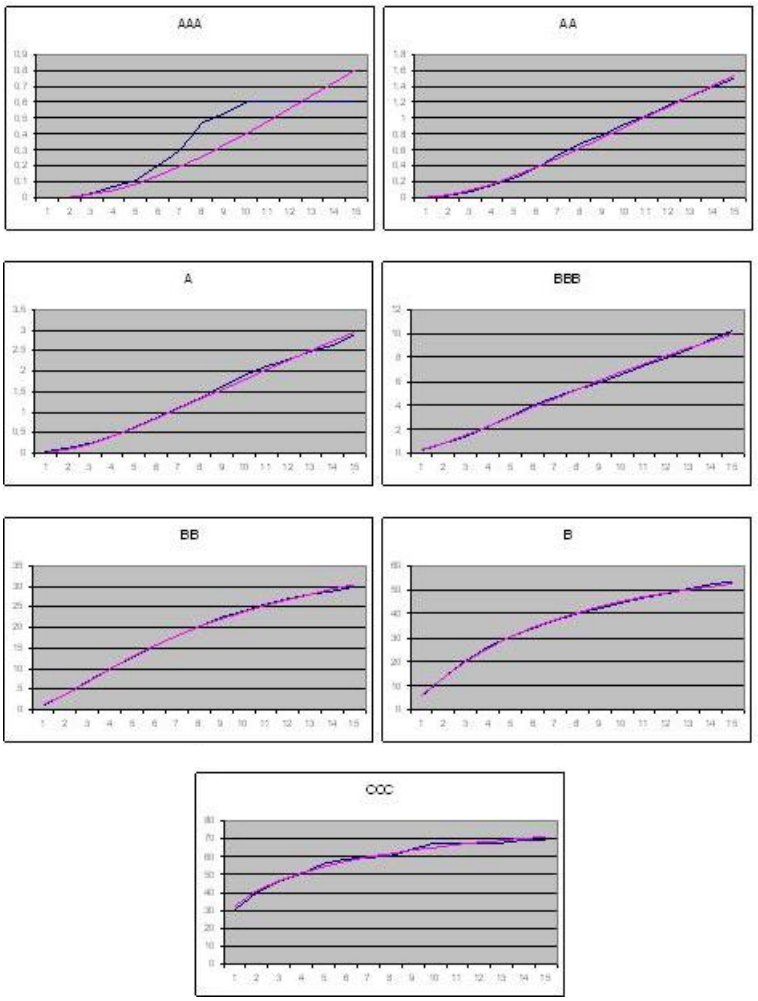


Fig. 3.2 : PD term structures based on a non-homogeneous continuous-time Markov Chain (purple) vs. observed PDs (blue)

Those graphics are to be compared with the ones in [2] which can be found in Fig. A.4 in the Appendix. We can see that except for the AAA and CCC-ratings the fit is quite good. The characteristics of the PDs term structures for the other rating classes are captured by the Markov chain. For the AAA rating class, probably the problem is that the PD does not change during the last years, while the probabilities given by the non-homogeneous continuous time Markov chain are increasing. The Markov chain tends to perform well throughout the 15-year time horizon, which results in underestimating PDs for the first 11 years and then overestimating them. It would be maybe different if the PDs increased year over year, if the shape of the term structure were more regular. However the AAA-rated companies are rare in a portfolio, so the fact that the model does not fit well for this rating class is less problematic as it would be for classes like BB or B with many companies. Concerning the CCC rating class, the fit is quite good, even though not as good as for the other credit classes, due to more irregular PDs, which looks like from concave becomes convex in the 7^{th} year (see Fig 3.2 CCC-graph).

Calibration using SP credit migration data from 1981-2008

In order to see how the method works with another data set, we implemented the same methodology as in the previous paragraph to the SP credit migration data from 1981 to 2008. So we have data on three more years. The corresponding average annual transition matrix is given in Fig. A.5 in the Appendix.
The calibration of the parameter vectors α and β in 2008 will allow us to see how these parameters change with time and if their values are comparable. If we had data to calibrate those parameters at different points in time, we could also see if their values follow a certain function of time in the best case.
The parameter vectors obtained seem to be stable and do not really depend on the initial parameters put in the Levenberg-Marquardt minimization algorithm. We find the following minimization parameters:

	α	β
AAA	0.321648	0.500311
AA	0.304611	0.395415
A	0.146537	0.252930
BBB	0.779988	0.765705
BB	0.296271	0.552484
B	0.399056	0.407457
CCC	1.304356	0.259145

Tab. 3.3: Parameters resulting from the calibration on SP data 2008

The PD term structures by grade calculated with those parameters are presented in Fig. A.6 in the Appendix. The fit of the modelled PDs to the observed ones seems to be very good and in some cases the empirical and modelled PD term structures seem to be exactly the same. The error is equal to :

$$\sqrt{\sum_{t=1}^{15} \parallel \hat{M}_{t\,column(8)} - exp(tQ_t)_{column(8)} \parallel^2} \simeq 0,01983,$$

smaller than the one computed before in the case of the data until 2005. This is probably explained by the fact that the AAA and CCC PD term structure seems to be more regular than the ones with data until 2005.
Concerning the parameter vectors, we cannot distinguish a rule that would determine the calibrated parameters variations.
We must underline the fact that the credit information used for the calibration is average credit migration rates and these averages smooth the economic cycle effects over the years.

3.2.2 Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm (LM) solves non-linear least squares problems. The LM algorithm interpolates between the Gauss-Newton (GN) algorithm, based on the linear approximation of the function to minimize, and the method of steepest descent. After local linearisation of the objective function with respect to the parameters to be estimated, the Levenberg-Marquardt algorithm performs initially small, but robust steps along the steepest descent

direction, and switches to more efficient quadratic Gauss-Newton steps as the minimum is closer. The derivatives are calculated numerically using the perturbation method.
The LM algorithm is more robust than the GN algorithm, which means that even though it starts far away from the final minimum in many cases it finds a solution. However if the function is well-behaved and the starting parameters are reasonable than it tends to be slower than the GN algorithm.
Given a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with $m \geqslant n$, we want to minimize $\parallel \mathbf{f(x)} \parallel$ or equivalently to find $\mathbf{x^*}$=argmin$_{\mathbf{x}}\{F(\mathbf{x})\}$ where

$$F(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m}(f_i(\mathbf{x}))^2 \tag{3.6}$$

Let $\mathbf{J} \in \mathbb{R}^{mXn}$ be the Jacobian of $\mathbf{f}$ i.e.

$$(\mathbf{J}(\mathbf{x}))_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}) \tag{3.7}$$

The step h is defined by the following:

$$(\mathbf{J}^\top\mathbf{J} + \mu\mathbf{I})\mathbf{h} = -\mathbf{g} \text{ with } \mathbf{g} = \mathbf{J}^\top\mathbf{f} \text{ and } \mu \geqslant 0 \tag{3.8}$$

Here $\mathbf{J}$=$\mathbf{J(x)}$ and $\mathbf{f}$=$\mathbf{f(x)}$.

The damping parameter μ has several effects:

1. For all $\mu > 0$ $\mathbf{h}$ is a descent direction towards the vector $\mathbf{x}$ that is the minimum of F($\mathbf{x}$).
2. For large values of μ we get a short step in the steepest descent direction, which is good if the current iterate is far from the solution.
3. If μ is small then the step of LM is close to the step of GN. The algorithm provides a good step in the final stages of the iteration, when $\mathbf{x}$ is close to $\mathbf{x^*}$.

Thus the damping parameter μ influences both the size and direction of the step. μ is modified at each iteration. If F($\mathbf{x}$) decreases during an iteration, μ is lowered and the LM becomes similar to GN. On the contrary if F($\mathbf{x}$) increases this means that $\mathbf{f}$ is not exactly linear in the current region where the algorithm is searching. In this case μ is increased and the LM algorithm becomes similar to the steepest descent algorithm.
An initial damping parameter μ_0 should be chosen relatively to the size of the elements in $A_0 = J^\top(x_0)J(x_0)$, for example by

$$\mu_0 = \tau \times \max_i (a_{ii}^0) \tag{3.9}$$

where τ is chosen by the user. The algorithm is not very sensitive to the choice of τ but in geeral τ is chosen to have a small value.

During the iteration the size of μ can be updated[2]. The update is controlled by the gain ratio ϱ

$$\varrho = \frac{F(x) - F(x_{\text{new}})}{L(\mathbf{0}) - L(\mathbf{h})} \tag{3.10}$$

where the denominator is the gain predicted:

$$L(\mathbf{0}) - L(\mathbf{h}) = \frac{1}{2}\mathbf{h}(\mu\mathbf{h} - \mathbf{g})$$

A large value of ϱ indicates that L($\mathbf{h}$) is a good approximation of F($\mathbf{x}$+$\mathbf{h}$) and we can decrease μ so that the next step is closer to the GN step. If ϱ is small then L($\mathbf{h}$) is a poor approximation so μ should be increased in order to get closer to the steepest descent direction and reducing the step length.
The stopping criteria should reflect that at a global minimum we should have $F'(x*) = g(x*) = 0$, so we can use:

$$\| \mathbf{g} \|_\infty \leqslant \varepsilon_1$$

where ε_1 is a small positive number chosen by the user. Another stop criteria is to stop if the change in $\mathbf{x}$ is too small:

$$\| x_{\text{new}} - x \| \leqslant \varepsilon_2(\| x \| + \varepsilon_2).$$

ε_2 is chosen by the user.

To prevent a infinite loop the user must specify a maximum number of iterations.
The vector of parameters in our case is:

$$\mathbf{x} = (\alpha_1, ...\alpha_8, \beta_1, ..., \beta_8)$$

In practice the LM algorithm converges in much less iterations. However each iteration demands more calculations, in particular the computation of the inverse of the matrix ($\mathbf{J(x)}^\top \mathbf{J(x)}$+$\mu I$). Its use is thus limited to the cases where the number of parameters is not very high.

The function we want to minimize is $\mathbf{f}$ which contains the coefficients:

$$\text{for } i = 1, ..., 8 \; f_{it}(x) = p_{it} - (\exp(\phi(t, x)Q)_{i8}$$

where p_{it} is the probability of default for the rating class i at time t.

[2]For more details look at update of damping parameter in the descent methods

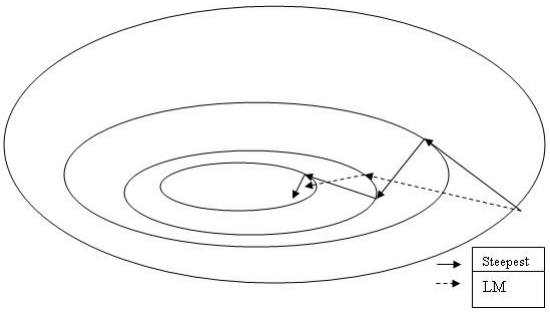


Fig 3.3: The steepest descent and Levenberg-Marquardt algorithms steps on level sets of the function to minimize

Algorithm

begin

k:=0;
ν :=0;
$\mathbf{x}$:=$\mathbf{x}_0$;
$\mathbf{A}$:=$\mathbf{J(x)}^\top$ $\mathbf{J(x)}$;
$\mathbf{g}$:=$\mathbf{J(x)}^\top$ $\mathbf{f(x)}$;
found:=($\| \mathbf{g} \|_\infty \leqslant \varepsilon_1$);
$\mu := \tau \times \max(a_{ii})$

while (not *found***) and (**k<$\mathrm{k}_{\max}$**)**
$k := k + 1$;
Solve($\mathbf{A}+\mu\mathbf{I}$)$\mathbf{h}$=-$\mathbf{g}$;

if $\| \mathbf{h} \| \leqslant \varepsilon_2(\| \mathbf{x} \| + \varepsilon_2)$
found:=**true**

else
$\mathbf{x}_{\mathrm{new}}$:= $\mathbf{x}$+ $\mathbf{h}$
ϱ := ($F(\mathbf{x})$-F($\mathbf{x_{new}}$))/(L($\mathbf{0}$)-L($\mathbf{h}$))

if $\varrho > 0$
$\mathbf{x}$:=$\mathbf{x_{new}}$
$\mathbf{A}$:=$\mathbf{J(x)}^\top$ $\mathbf{J(x)}$;
$\mathbf{g}$:=$\mathbf{J(x)}^\top$ $\mathbf{f(x)}$;
found:=($\| \mathbf{g} \|_\infty \leqslant \varepsilon_1$);
μ:=$\mu \times \max(\frac{1}{3}, 1 - (2\varrho - 1)^3)$;
ν:=2;

else
μ:=$\mu \times \nu$;
ν:=2 $\times\nu$

end

3.3 Economic cycles

In this section we will present how to make the probabilities of default depend on the economic cycle. As mentioned before the articles that use this approach assume that the credit migration process follows a homogeneous Markov chain. The idea of this approach is to calibrate a transition matrix corresponding to expansion periods and another one for contraction (recession) periods.
First one needs to identify expansion and contraction periods. A macroeconomic variable or an index must be used for this purpose. The need of assembling data on a country and perhaps sector criteria rises here, in order to make the choice of an economic index easier and more reliable.
Let us denote t_0 the present instant of time. We consider a first time period denoted $[t_0, t_1]$ and a second time period $[t_1, t_2]$. The considered time periods have the same length, denoted l.
We consider only two economic cycle phases: recession (R) and upturn (U).
In the Economic Cycle approach presented before, we condition the probabilities of default for the period $[t_1, t_2]$ on the economic cycle phase in $[t_0, t_1]$. This is done by estimating average default probabilities and transition probabilities between grades for time periods of length l and when the economic situation was improving or worsening. This implies that we have a transition matrix for a downturn and another for an upturn.
Let D_i be the Bernoulli variable indicating if there is default in the period i.
Let S_i be the economic state for the period i. $S_i = R, U$
The probability of default for the period $[t_1, t_2]$ *conditional* on the economic state in the first period $[t_0, t_1]$ will be:

$$\begin{aligned} \mathbb{P}(D_2 = 1|S_1) \quad &= \mathbb{P}(D_1 = 1|S_2 = R) \times \mathbb{P}(S_2 = R|S_1) + \qquad (3.11) \\ &\quad \mathbb{P}(D_1 = 1|S_2 = U) \times \mathbb{P}(S_2 = U|S_1) \qquad (3.12) \end{aligned}$$

The probabilities $\mathbb{P}(S_{i+1}|S_i)$ are average probabilities of transition between states of economy.

Since the hypothesis of a homogeneous Markov chain is made, the k-year transition matrices are calculated by taking the k^{th} power of the one year transition matrix.

$$M_k = M^k \tag{3.13}$$

The final step is specifying how economic regimes are switched. A 2×2 regime switching matrix can be estimated after having identified the expansion and recession periods. This matrix gives the probability of being in expansion or recession the next period conditional to the regime during the present period. Regime paths can then be simulated.

3.4 Mixing the two approaches

The non-homogeneous Markov Chain approach provides modelled PDs term structures which fit to the term structures observed in the reality. This is a good approach if one wants to have PDs term structures throughout a cycle.

However those term structures do not make any difference between phases of the economic cycle.
So we need to apply the non-homogeneous Markov Chain approach but to default data which capture the economic cycle effect.

How to proceed in practice
We need to condition the term structure of default probabilities on the economic state in the *current* time period. We need the corresponding data to do the calibration of the conditioned term structures.

- We divide time in trimesters (intervals of length l) and define if a trimester is a trimester of recession or upturn (using the GDP variation for example).
- In order to obtain the term structures for an upturn economy, *for each* trimester of upturn economy, we denote the beginning of the following trimester t_0 and we calculate the PDs for every time horizon from 3 months to 3 years with a time step of 3 months, $p_{[t_0,t_i]}$ for i=1,...,12.
- We also calculate the transition probabilities for a horizon of 3 months. We do the same thing for the time periods of economic downturn.
- We calibrate the term structures of PDs for expansion and recession periods using the above data and the approach presented in the section on non-homogeneous Markov chain.

Justification
By doing this we take into consideration, even though ***implicitly***, the **length and amplitude of the economic cycle phase**. The observed term structure at every time step is a possible scenario of what the term structure of the "forward" probabilities of default might be. So if we do the average of the term structures observed at each time step we obtain an average expected evolution of PDs in the future knowing the state of the economy today.
However we need a long history of PDs so that many possible effects of the economic cycle over the probabilities of default are considered. If the history of defaults we have is long enough then it will incorporate many possible scenarios of economical development (cycles of different amplitudes and lengths). Each scenario has its own effect on what happens to the default rates of different time horizons.

This approach can be modified to take into account information about where we are in the cycle phase: beginning, middle or end for example. **If we knew where we are in the cycle phase, we can then compute a weighted average and not a simple average and put more weight on term structures whose beginning is in the same phase as we believe being today.**This can be done either by specifying weights (probabilities) on different scenarios (beginning, middle or end of an expansion or recession phase), depending on our belief in where we are in the economical cycle, or by analysing the economical cycle by time series and predict what will happen in the future. The second approach demands more time and it will only be statistical and will not take

into account information proper to the current period or personal beliefs of the manager based on what he thinks that will happen in the future.
We may also consider the amplitude and length of an expansion and recession period by specifying a probability distribution on the possible amplitudes and lengths. The set of possible amplitudes and lengths would be defined on what has been observed previously. The weighted average PDs will be calculated with weights equal to the probabilities specified. **This requires a long history of default data.**

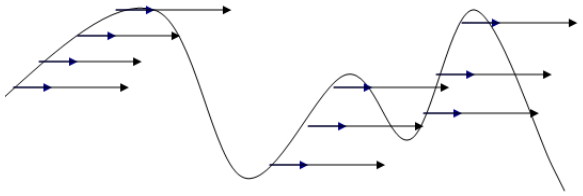


Fig. 3.4: Economical cycle and PDs for two time horizons for expansion periods

Data we need for the calibration
In order to apply the Non-homogeneous Markov Chain approach, first of all we need to determine a time step and the length of the term structure. We will discuss about the time step and length later on. Let's assume for now that the chosen time step is the trimester and the time length is 3 years so that the following explanation will be simpler. The data needed is the following:

- Data which allows to distinguish recession or upturn economy periods. We can use the GDP variations for example. If we want to calculate weighted averages of default, we need to distinguish between beginning, middle and end of an expansion and recession.
- for trimesters of recession or expansion, compute default probabilities with beginning of horizon=end of the trimester and for the time lengths: 3 months, 6 months, 9 months,...,3 years.
- an average transition matrix for the time period of 3 months for recession and expansion.

The choice of the time step is important since it should not be too short because there won't be many, if not any, defaults during the period, especially for the buyers' group with good creditworthiness. If the observed PD term structure looks like a simple function then the Markov Chain will not be able to capture those jumps and will approximate by a continuous function which will give the very general shape of the observed PD term structure. Neither should the time step be too big because then we need a greater length of the term structure and thus more data in order to have enough data for the estimation.

Chapter 4

Conclusion

In this paper we present a way of calibrating term structures of default probabilities. We implemented the calibration algorithm and calibrate the term structures with data from S & P. Our results are quite satisfactory when we compare them to the results of [2].

The data we used are through-the-cycle data, so the term structure we obtain is not conditioned on the state of the economy. In order to have a term structure which depends on the economic cycle, we presented a possible way out. Because of the lack of trimestrial default data specific to different economic cycles we couldn't proceed to the calibration of the term structures conditioned on the economic state. However it would be very interesting to look at the conditioned term structures and to compare them.

We proposed two ways of integrating the term structures in the internal model, and this by assuming that the correlation matrix of the systematic risks is constant throughout the year. This is rather restrictive, so hereby we present some ways of improving the internal model and allowing it to better take into condideration the term structures. In particular we propose having a dynamic correlation matrix.

4.1 Considering heterogeneity by the means of thresholds

Once the default probabilities for each rating class are calculated, the threshold d_j can be calculated as a quantile of the distribution of the ability-to-pay Z_j, which is Gaussian distributed by assumption. By doing so, we can have one threshold for each time t. It would be interesing to get to know the thresholds for each buyer in the group, since there must be some heterogeneity within the rating class that we have not taken into account yet. We can try to see whether the default probability of the rating class k, denoted π_t^k, can be expressed as a function of the default probabilities of each buyer in the rating class, denoted $\pi_{t,j}^k$.

$$\pi_t^k = f(\{\pi_{t,j}^k; j \in J_k\}) \tag{4.1}$$

where J_k is the set of buyers in the rating class k. Conditional to the vector of systematic risk factors the probabilities of default are not correlated so one could reasonably make the assumption that the function f is linear:

$$\pi_t^k = f_{|R}(\{\pi_{t,j}^k; j \in J_k\}) = \sum_{j \in J_k} \lambda_{t,j} \pi_{t,j}^k \tag{4.2}$$

where the coefficients $\lambda_{t,j}$ are to be calculated. We could use the probabilities of default for each time t in order to estimate the coefficients. However the drawback of this method would be that the set of buyers in a given grade changes with time.
Nevertheless, if we find a function f, we can then assume that the threshold of the buyer j in the rating class k, $d_{t,j}^k$ is proportional to the threshold d_k^t of the class k, and :

$$d_{t,j}^k = \alpha_k^t d_j^t \tag{4.3}$$

4.2 Thresholds as random variables

In equation 4.4 we see that default probabilities depend also on a threshold. The threshold can be considered as depending on the recovery rate at default, which is generally considered as a random variable. Therefore random thresholds could be another possible improvement of the model.

4.3 Wishart process and term structure of correlation

The introduction of term structures of PDs in the model makes it necessary to have a time-dependent ability-to-pay process. Indeed, as we saw in the first chapter the probabilities of default of a given firm at different maturities represent the probability that the ability-to-pay of the first falls below a certain threshold (deterministic or random variable) until the maturity. Formally we have:

$$\begin{aligned} \mathbb{P}\{\exists t, 0 \leqslant t \leqslant t_i \text{ such that } Z_{jt} < d_{jt_i}\} &= \mathbb{P}\{\min_{0 \leqslant t \leqslant t_i} Z_t < d_{jt_i}\} & (4.4) \\ &= p_{jt_i} & (4.5) \end{aligned}$$

Up to now, the information on the correlation between systematic risk factors is given by a fixed covariance matrix Σ. The covariance between the systematic risk factors defines the variance of the ability-to-pay random variable since this last is given by the formula:

$$\begin{aligned} \mathbb{V}(Z_j) &= \sum_{\nu=1}^{k} \omega_{j\nu}^2 \mathbb{V}(R_\nu) + 2 \sum_{\nu_p=1}^{k-1} \sum_{\nu_q=\nu_p+1}^{k} \omega_{j\nu_p} \omega_{j\nu_q} Cov(R_{\nu_p}, R_{\nu_q}) & (4.6) \\ &= \sum_{\nu=1}^{k} \omega_{j\nu}^2 \sigma_{\nu_p\nu_p} + 2 \sum_{\nu_p=1}^{k-1} \sum_{\nu_q=\nu_p+1}^{k} \omega_{j\nu_p} \omega_{j\nu_q} \sigma_{\nu_p\nu_q} & (4.7) \end{aligned}$$

There is a considerable literature considering that correlation matrices follow a stochastic process, named *Wishart process*. In other words, this assumption implies that correlation matrices contain not only a part of real information on the correlation structure between systematic risk factors but also a random part which can be separated from the real information. The latter can be treated by the Random Matrix Theory if the correlation matrix is large enough, i.e. if the number of systematic risk factors is big enough [1].

Let us formalize the idea above. First of all we will define a Wishart process.

Definition 16 (Wishart process) *The covariance matrix Σ_t is said to follow a Wishart process if it satisfies the following dynamics:*

$$d\Sigma_t = \left(\Omega\Omega^\top + M\Sigma_t + \Sigma_t M^\top\right) dt + \sqrt{\Sigma_t} dW_t Q + Q^\top \left(dW_t\right)^\top \sqrt{\Sigma_t} \tag{4.8}$$

with Ω, M, Q $k \times k$ matrices, Ω is invertible, and W_t a $n \times n$ matrix Brownian motion. This stochastic differential equation is written under the historical measure. The Wishart process is a affine diffusion process because both the drift and volatility are functions of Σ.

For the volatility to be mean-reverting, the matrix M is assumed to ne negative semi-definite and Ω satisfies:

$$\Omega\Omega^\top = \beta Q Q^\top,\ \beta > n - 1 \tag{4.9}$$

The term $\Omega\Omega^\top$ is related to the expected long-term covariance matrix, Σ_∞, through the solution to the following linear equation:

$$-\Omega\Omega^\top = M\Sigma_\infty + \Sigma_\infty M^\top \tag{4.10}$$

Q is the volatility of the volatility matrix.

In order to take into account leverage effects, it is assumed that the Brownian motions of the asset returns and those of the covariance matrix are linearly correlated.

In order to have a correlation matrix which varies with time, we should estimate the matrices in the equation 4.8 defining the Wishart process. Let us remind that in our case we need the correlation between systematic risk factors. For this reason, first of all, we need to find a sort of representative asset for each country and industry. Then we state a dynamic of this asset and proceed to the calibration of the parameters of this asset dynamic and the Wishart process representing the dynamics of the covariance matrix.

In general, if we denote S_t the n-dimensional risky asset, the asset dynamic is supposed to be the following:

$$dS_t = diag[S_t]\left(\mu dt + \sqrt{\Sigma_t} dZ_t\right) \tag{4.11}$$

[1] In general in finance a correlation matrix between 500 assets is considered as large.

where μ is the vector of returns and Z_t is a vector Brownian motion. In the case of this asset dynamic specification there exist different methods of estimation for the parameters of the Wishart process.
Let us note that the Wishart Affine Stochastic Correlation (WASC) is a continuous process.

The more accurate time-dependent correlation structure we aim to obtain, will allow to have more accurate loss distributions. However, in practice, the accuracy of loss distribution tails depends on the convergence speed of the simulation tools. The internal model of EH is being improved thanks to the implementation of the importance sampling technique for the Monte Carlo simulations. This technique allows to have more stable fat tails and to think about a new correlation modelling.

Bibliography

[1] BANGIA A., DIEBOLD F.-X., SCHUERMANN T., (2000), Ratings Migration and the Business Cycle, With Applications to Credit Portfolio Stress Testing, 00-26, The Wharton Financial Instituions Center.

[2] BOUCHAUD J.-P., LALOUX L., CIZEAU P., POTTERS M. (1999), Random Matrix Theory and Fiancial Correlations, Science Finance (CFM) working paper .

[3] BLUHM C., OVERBECK L., (2007), Calibration of PD term structures: to be Markov or not to be, Credit risk, 98-103.

[4] DA FONSECA J., GRASELLI M., IELPO F. (2008), Estimating the Wishart Affine Stochastic Correlation model using the empirical characteristic function, ESILV, No RR-35.

[5] DECROOCQ J.-F., PLANCHET F., MAGNIN F. (2009), Systematic risk modelisation in credit risk insurance, ASTIN 2009.

[6] GOURIEROUX C., SUFANA R. (2005), Wishart Quadratic Term Structure Models, Les Cahiers du CREF of HEC Montreal Working Paper No. 03-10.

[7] ISRAEL R., ROSENTHAL J., WEI J., (2001), Finding generators for Markov chains via empirical transition matrices with application to credit ratings, Mathematical finance 11(2), 245-265.

[8] JONES M. T.,(2005), Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk, IMF Working Paper No. 05/219.

[9] KREININ A., SIDELNIKOVA M., (2001), Regularization algorithms for transition matrices, Algo Research Quarterly 4(1/2), 2540.

[10] MADSEN K., NIELSEN H.B., TINGLEFF O. (2004), Methods for non-linear least squares problems.

[11] RIVAS E. (2005), Evolutionary models for insertions and deletions in a probabilistic modelling framework, BMC Bioinformatics.

[12] Standard & Poors (2005), Annual global corporate default study: corporate defaults poised to rise in 2005, SP Global Fixed Income Research.

[13] Standard & Poors (2009), Default, Transition, and Recovery: 2008 Annual Global Corporate Default Study And Rating Transitions, SP Global Fixed Income Research.

[14] ZHOU C. (2001), An analysis of default correlations and multiple defaults, The Review of Financial Studies, Vol. 14, No 2, pp. 555-576.

Appendix

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	-8,7152%	8,4440%	0,1483%	0,0684%	0,0649%	-0,0087%	-0,0015%	-0,0003%
AA	0,6788%	-10,1284%	8,9091%	0,3802%	0,0227%	0,1151%	0,0216%	0,0009%
A	0,0459%	2,3701%	-9,3065%	6,3675%	0,3254%	0,1504%	0,0250%	0,0222%
BBB	0,0190%	0,1878%	4,4860%	-11,1692%	5,3886%	0,6538%	0,2200%	0,2141%
BB	0,0443%	0,0761%	0,2447%	6,6776%	-18,7094%	9,6308%	1,1530%	0,8829%
B	-0,0057%	0,0760%	0,2210%	0,1157%	7,0101%	-20,0563%	7,0918%	5,5475%
CCC	0,1264%	-0,0203%	0,4716%	0,5425%	1,6144%	16,5881%	-62,2035%	42,8808%
Default	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%

Fig. A.1: Matrix logarithm of the one-year transition matrix M Fig. 2.1.

We can see that this matrix is not a valid generator because there are negative non-diagonal coefficients. The QOG algorithm is applied to this matrix and the following approximative generator is obtained. We note it Q^*.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	-8,7152%	8,4423%	0,1466%	0,0667%	0,0631%	0,0000%	0,0000%	0,0000%
AA	0,6789%	-10,1284%	8,9092%	0,3803%	0,0228%	0,1152%	0,0217%	0,0000%
A	0,0487%	2,3728%	-9,3065%	6,3702%	0,3282%	0,1532%	0,0278%	0,0000%
BBB	0,0000%	0,1901%	4,4883%	-11,1692%	5,3910%	0,6562%	0,2223%	0,2165%
BB	0,0000%	0,0817%	0,2502%	6,6831%	-18,7094%	9,6364%	1,1586%	0,8884%
B	0,0000%	0,0753%	0,2203%	0,1149%	7,0094%	-20,0563%	7,0911%	5,5467%
CCC	0,1239%	0,0000%	0,4690%	0,5400%	1,6119%	16,5856%	-62,2035%	42,8783%
Default	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%

Fig. A.2: Approximative generator obtained by applying QOG algorithm

The sum of the row coefficients is nearly one for each row and there are no negative non-diagonal elements anymore. This matrix should be compared to the following one from [2], denoted Q^{BO}.

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	−8.73	8.44	0.15	0.07	0.06	0.00	0.00	0.00
AA	0.68	−10.13	8.91	0.38	0.02	0.12	0.02	0.00
A	0.05	2.37	−9.31	6.37	0.33	0.15	0.03	0.02
BBB	0.02	0.19	4.49	−11.17	5.39	0.65	0.22	0.21
BB	0.04	0.08	0.24	6.68	−18.71	9.63	1.15	0.88
B	0.00	0.08	0.22	0.12	7.01	−20.06	7.09	5.55
CCC	0.13	0.00	0.47	0.54	1.61	16.59	−62.22	42.88
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Fig. A.3: Approximative generator found in [2]

The two matrices are very similar. However in the paper matrix the sum of the row coefficients is not always 0 due to the probable roundings. The sum of the absolute value of the differences of the coefficients of Q^*_{ij} and Q^{BO}_{ij} is equal to:

$$\sum_{i=1}^{8}\sum_{j=1}^{8}|Q^*_{ij} - Q^{BO}_{ij}| = 0,00243$$

which can be considered as not significant.

$$\|Q^* - Q^{BO}\|_2 = \sum_{i=1}^{8}\sum_{j=1}^{8}(Q^*_{ij} - Q^{BO}_{ij})^2 = 0,000595$$

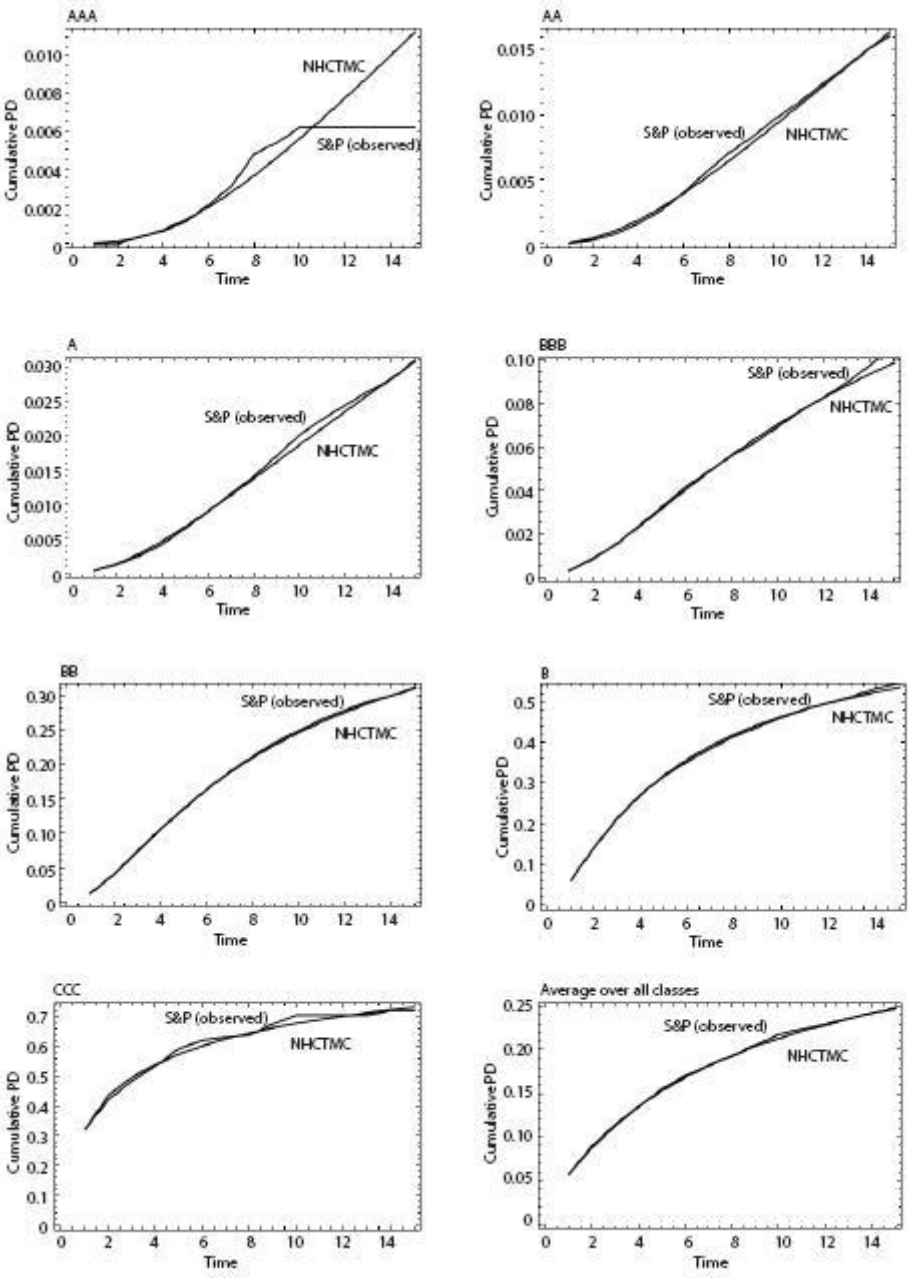


Fig. A.4: PD term structures based on a non-homogeneous continuous-time Markov chain (NHCTMC) approach cf. [2]

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	91,62	7,63	0,53	0,06	0,08	0,03	0,06	0
AA	0,58	90,88	7,79	0,54	0,06	0,09	0,03	0,03
A	0,04	2,04	91,91	5,35	0,4	0,16	0,03	0,08
BBB	0,01	0,15	3,87	90,88	4	0,69	0,16	0,24
BB	0,02	0,05	0,19	5,3	85,42	7,22	0,8	0,99
B	0	0,05	0,15	0,26	5,68	85,02	4,34	4,51
CCC	0	0	0,23	0,34	0,97	11,84	60,96	25,67
Default	0	0	0	0	0	0	0	100

Fig. A.5: SP average annual transition matrix, data 1981-2008

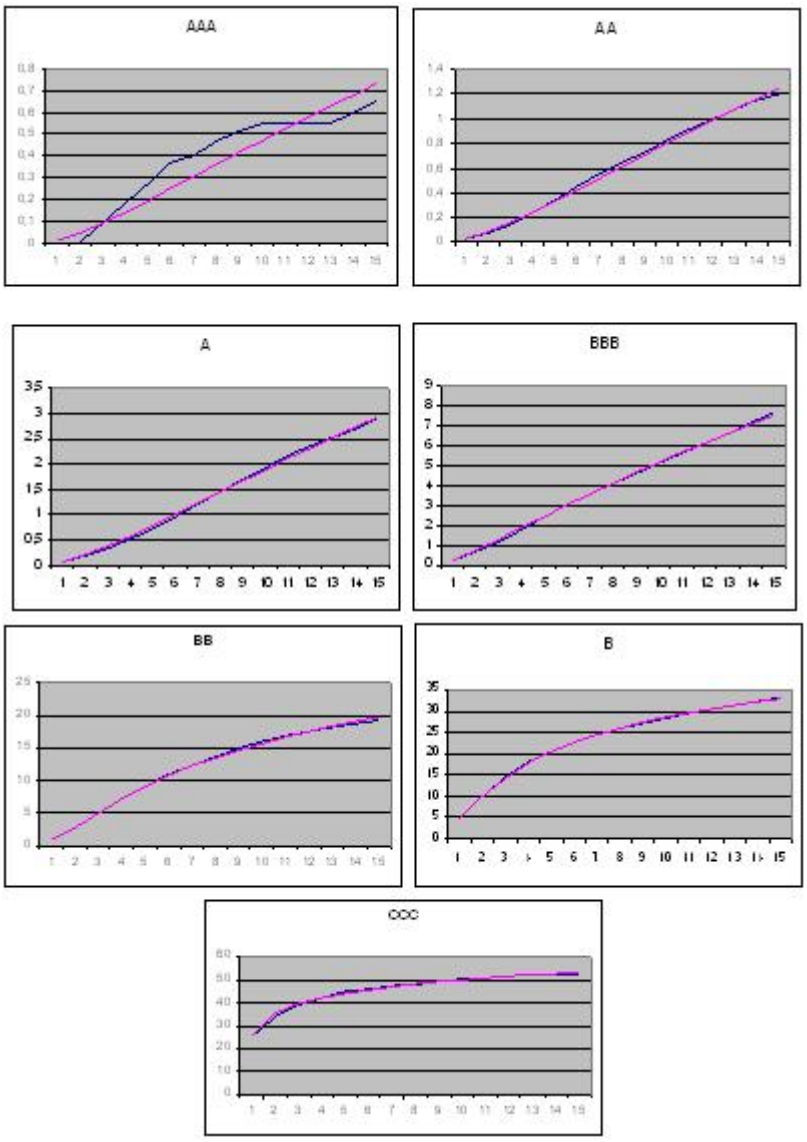


Fig. A.6: PD term structures based on a non-homogeneous continuous-time Markov chain approach for the data until 2008