LISBOA SCHOOL OF ECONOMICS & MANAGEMENT

MASTER

MATHEMATICAL FINANCE

MASTER'S FINAL WORK

DISSERTATION

HIDDEN MARKOV MODELS
FOR CREDIT RISK

LEONOR MARQUES POMPEU DOS SANTOS

OCTOBER-2015

MASTER

MATHEMATICAL FINANCE

MASTER'S FINAL WORK

DISSERTATION

HIDDEN MARKOV MODELS
FOR CREDIT RISK

LEONOR MARQUES POMPEU DOS SANTOS

SUPERVISION:

JOSÉ PEDRO GAIVÃO

OCTOBER-2015

Abstract

Credit Risk measurement, the evaluation of the risk of default or reduction in market value caused by changes in credit quality, has been a broadly studied subject over the last thirty years and is now more relevant than ever, when the world is still suffering the consequences of the break of a financial crisis in its genesis induced by a false observation of this kind of risk.

Just like some of the previous studies, the model presented in this dissertation assumes that default events are directly connected to risk state variables, starting from a very simple model that assumes defaults to follow a two-state Binomial Hidden Markov Model, considering only two different risk categories to fully explain default occurrence, and approximating it to a Poisson Hidden Markov Model, with all the computational simplifications brought by this approximation, trying, at the same time, to translate the model into a less extreme framework, with the addition of an intermediate risk level, a "normal" risk state.

Keywords: Hidden Markov Models, Poisson Hidden Markov Models, Credit Risk, Credit Risk Models, Maximum Likelihood, Risk State, Default Modeling.

Resumo

A análise do Risco de Crédito, a avaliação do risco de *default* ou de redução do valor de mercado causado por alterações na qualidade de crédito, tem sido um tema vastamente estudado ao longo dos últimos trinta anos e é hoje mais relevante que nunca, com o mundo ainda a recuperar das consequências de uma crise financeira, na sua génese induzida por uma observação imperfeita deste tipo de risco.

Tal como alguns dos modelos apresentados anteriormente, o modelo apresentado nesta dissertação assume que os eventos de *default* estão directamente ligados a uma variável associada ao risco, partindo de um modelo simples que assume que o *default* segue um Modelo Oculto de Markov Binomial de dois estados, ou seja, um modelo que considera apenas dois "estados de risco" possíveis para explicar na totalidade a ocorrência de *default*, e aproximando-o a um Modelo Oculto de Markov Poisson, com todas as simplificações computacionais associadas a esta aproximação, tentando, ao mesmo tempo, traduzir o modelo para um cenário menos extremo, com a inclusão de um nível de risco intermédio.

Palavras-chave: Modelos Ocultos de Markov, Modelos Ocultos de Markov Poisson, Risco de Crédito, Modelos para o Risco de Crédito, Máxima Verosimilhança, Risco, Modelação do Incumprimento.

Acknowledgments

· Agradecimentos ·

Ao professor José Pedro Gaivão, meu orientador, por toda a paciência, por toda a ajuda e por não ter desistido de mim mesmo nos meus momentos de maior preguiça.

Aos meus pais, por todo o apoio desde sempre, por sempre me terem incentivado a querer ser o melhor possível.

Às minhas colegas de mestrado, pelos bons e maus momentos, pelos momentos de estudo e de menor motivação, foi um prazer partilhar esta viagem convosco!

Aos Filipes, à Rita, ao Pedro, à Susana, à Marta, à Sara e a todos os meus amigos e família pelo apoio e pela paciência nesta fase de stress e por todas as outras que já partilhámos ou ainda viremos a partilhar.

Por fim, à minha irmã, "a outra peça", por estares lá mesmo quando não estás, por me apoiares incondicionalmente e por continuares a ser tudo aquilo que sempre foste: a "minha pessoa".

Contents

Introduction

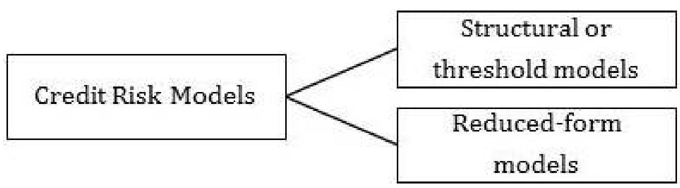
Credit Risk can be defined as the risk of variation in portfolio value caused by unexpected changes in the credit quality of issuers or trading partners evaluated through credit rating. This type of risk includes not only the effects of downgrades in internal and/or external rating systems but also the effects of defaults, having a relevant presence in the portfolio of any financial institution.

The recent development of the market for credit derivatives and the publication of the three Basel Accords[1] have made the subject of quantitative credit risk modeling into a very pertinent and very active sub-field of quantitative finance and risk management.

Credit risk models can be applied in two main areas of knowledge: **credit risk management**,where models are used to determine the loss distribution of a loan or bond portfolio over a certain time period with the goal of finding the best risk-capital allocations (*static models* - focus on loss distribution over time) and **credit risk securities analysis**, where models are used to determine the payoff at the exact time of default, usually resorting to a risk-neutral probability measure in order to build an adequate pricing model (*dynamic models* - focus on the evolution of risk over time using a stochastic process to describe it).

The purpose of this dissertation is to study the distribution of defaults from a risk management perspective, modeling its occurrence over time with the help of static models, a subject that proves to be an interesting and pertinent subject within the field of Financial Mathematics.

Credit Risk Models can be split into two groups:



Structural, firm-value models or latent variable models, first introduced in 1974 by Merton, consider that default occurs whenever a stochastic variable, generally

[1]Basel I, II and III Accords - Global regulatory frameworks on bank capital necessities, stress testing and market liquidity risk issued by the Basel Committee on Banking Supervision (BCBS) in 1988, 2004 and 2013 respectively.

representing an asset value, falls below a threshold that represents liabilities, which is why, when applied at portfolio level, this type of models is commonly denominated threshold models. A popular implementation of this type of models is the commercial KMV model.

Reduced-form models, on the other hand, recognize default time to be modeled as a non-negative random variable whose distribution depends on economic co-variables, while the precise mechanism leading to default is left unspecified. In this dissertation the Mixture Model approach will be briefly introduced as a common example of static portfolio versions of reduced form models.

The purpose of this dissertation is, thus, to use a simple form of Mixture Models to study discrete time Hidden Markov Models as the simplest possible explanation for observed default events.

The usual rationale behind the application of a Hidden Markov Model in of financial time series modeling rests on the assumption that the market may switch from time to time between a “quiet” state (a state of low volatility) and a “turbulent” state (an unstable, high volatility state), allowing for the study of the influence of the state of the market on the behavior of the financial series in a certain time spam.

A similar rationale is followed in our model, assuming, instead of two possible states of the market, two possible positions for a financial instrument: a normal risk state and a high risk state corresponding, in a simplified manner, to the existence or absence of an outside risk trigger associated to third party default. By resorting to a Markov-type model, it is ensured that the risk state variable will have no memory of its prior behavior, in other words, prior changes between risk states will have no predictive power for the direction of future fluctuations, a crucial point in our study.

All in all, the hypothesis of default event estimation using a Poisson Hidden Markov Model will be tested starting from the simple case with 2 possible risk states and moving on to a slightly more complex scenario where a third, intermediate, “normal state” variable, is added. As a fundamental resource in our estimation, the Standard & Poor’s default data library for the years 1981 to 2000 available in R was put to use.

1 Introduction to Credit Risk Models

1.1 Structural Models of Default

Structural or firm value models intend to derive the likelihood of a firm's default over any time horizon from its capital structure and assumptions concerning its value process, according to the designated conditions determining default.

This category of Credit Risk Models presents itself as a method of explanation of the mechanism behind the occurrence of default using a generic stochastic process (X_t) to value the process in continuous time, assuming that default will occur whenever the value of the stochastic variable (or, in the case of dynamic models, the stochastic process) falls below a certain boundary, explicitly defined for each model.

Two of the main limitations associated with this group of models are the presumption of observability of firm value, something that is very far from the truth in the real world, since the value of a firm's assets is never completely traded, even if the firm possesses traded equity, and the assumption that default can only happen at time of maturity (T). Meanwhile this group has the advantage of allowing for the formulation of relatively simple realistic models for default correlation between issuers.

1.1.1 The Merton Model

Based on a stochastic process, $\{V_t\}$, to mimic a firm's asset value, the Merton Model assumes that at a given time t a firm possesses a certain level of equity, S_t and a certain level of debt consisting on a single debt obligation or zero coupon bond with face value B_t and maturity T. This means that, for a frictionless market, the value of a firm's assets at t can be obtained from the sum $V_t = S_t + B_t$, assuming that the firm cannot pay dividends or issue new debt.

When interpreting the model it is considered that default occurs if the firm misses a payment to its debt holders, which can only occur at maturity T, leaving us with two possible scenarios at the maturity date:

- $V_T > B_T \Rightarrow$ Firm assets exceed liabilities. Debt holders receive the total value of debt B_T and shareholders receive the remaining value $S_T = V_T - B_T$
- $V_T \leqslant B_T \Rightarrow$ Firm cannot meet its financial obligations. Either shareholders

agree on providing new equity capital or they exercise their "limited-liability option", handing over control of the firm to debt holders, that is, $B_T = V_T$ and $S_T = 0$

From these two scenarios it is easy to understand how the value of a firm's equity can be seen as the payoff of a European call option:

$$S_T = max(V_T - B, 0) \tag{1.1.1}$$

And the value of debt can be interpreted as the subtraction of the payoff of a European put option to the nominal value of the liabilities:

$$B_T = min(V_T, B) = B - max(B - V_T, 0) \tag{1.1.2}$$

For which there exists a number of available valuation formulas such as the one proposed in the Black-Sc holes model, making it possible to determine the default probability of a firm by computing the probability $P(V_T \leq B)$ under the real-world measure P [2].

This is of course a very simplified representation of a company's debt and the market in general but still it is an interesting starting point for credit risk modeling.

Remark 1. *In order to use this model at portfolio level, a multivariate version of the Merton Model may be applied, contemplating a multivariate asset value process* $\{V_t\}$ *with* $V_t = (V_{t1}, V_{t2}, \ldots, V_{tm})$, $\mu_V = (\mu_{V1}, \mu_{V2}, \ldots, \mu_{Vm})'$ *and* $\sigma_V = (\sigma_{V1}, \sigma_{V2}, \ldots, \sigma_{Vm})'$, *where* m *represents the number of assets in the portfolio.*

The same logic may be employed when discussing default risk for portfolios using other Structural Models for Default.

1.1.2 The KMV Model

Widely used in industry, the KMV Model is a very straightforward extension of the Merton Model, with the addition of a very important variable: the *Expected Default Frequency* (EDF), which expresses the defined probability of a given firm defaulting within a year.

[2] See McNeil, Frey and Embrasures (2005)[9] pp331 to 336.

The EDF is, therefore, a function of the current asset value V_t, the annualized mean μ_V and volatility σ_V and the default threshold $\tilde{B}$, estimated using an empirically estimated decreasing function.

In the Merton Model, default and, consequently, bankruptcy occurs if the value of the firm's assets falls below the value of its liabilities, leading to a very simplistic and less than realistic relationship between asset value and default probability. In order to deal with this issue, KMV introduces a new state variable, the *distance to default* (DD), which represents the "number of standard deviations a company is away from its default threshold", calculated based on the company's asset value at a certain time period, its annualized volatility and the default threshold $\tilde{B}$. It is considered that firms with similar DD have the same default probability.

Furthermore, the KMV Model does not assume that the asset value V_t is observable, since it would be naive to contemplate asset value as the firm's market value, considering the latter represents investor expectations about the business prospects of the firm and not all equity and debt are actively traded. To address this question, KMV uses an iterative technique to estimate asset value from the firm's equity value, which, conjugated with the non-assumption of adequacy of the EDF to a Gaussian distribution, makes this a more accurate method to model default.

The problem that arises from the use of the KMV Model is that, despite the fact that it may be better at reacting to changes in firm prospects than some other less responsive models (see, for example, the next section, regarding Credit Migration Models), it tends to be rather sensitive to global under- and overreaction of equity markets, with the breaking of market bubbles, for example, provoking a drastic increase in EDFs, even if the outlook of a given company has not changed very much.

1.1.3 Models based on Credit Migration

The Credit Migration approach is based on the hypothesis that each firm can be assigned to a credit-rating category at any given time according to its credit quality, including a level of default, assuming that current rating completely determines its probability of default. A matrix is then created from these rating categories with the aim of presenting a visual representation of the probability of moving from one credit rating to another over a given time horizon.

For major companies and sovereigns, credit rating and transition-matrices are provided by rating agencies such as Moody's or Standard & Poor's (S&P), while for other less representative companies proprietary rating systems internal to a financial institution can be used. In both situations transition rates are estimated as an average calculated from historical data over long time horizons, which makes them insensitive to the effects of current economic environment, causing for them to be less accurate in the estimation of "point-in-time" default probabilities as needed, for example, in the pricing of short-term loans.

A possibly more suitable procedure to model default probabilities would be to combine the premises of credit migration models with the ones regarding general firm-value models, considering the log-normally distributed asset value process $\{V_t\}$ of a firm that is matched to a certain rating category.

Instead of just considering the situation when asset value is not enough to pay for its liabilities one could define a number of different thresholds to represent "boundaries" between rating positions, allowing us to access to which rating category the firm belongs and to estimate migration probabilities. Two of the preeminent market applications of credit migration matrices are the ones provided by Moody's and Standard & Poor's, as can be seen in Figures 1 and 2.

From Rating \ *To Rating*	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca-C	*WR*	*DEF*
Aaa	**92.4**	2.2	1.3	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.8	0.0
Aa1	4.3	**84.4**	4.6	2.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.4	0.0
Aa2	1.4	5.7	**78.4**	5.8	1.2	0.4	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	0.0
Aa3	0.1	1.0	4.1	**83.6**	3.5	1.3	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	0.0
A1	0.0	0.1	0.4	5.7	**82.9**	4.1	1.7	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	0.0
A2	0.0	0.0	0.2	1.0	5.0	**82.9**	4.1	1.4	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.5	0.0
A3	0.1	0.1	0.1	0.2	0.9	7.6	**78.7**	4.5	2.1	0.7	0.2	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.5	0.0
Baa1	0.1	0.1	0.1	0.1	0.2	1.7	5.9	**79.8**	5.0	1.7	0.4	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	4.4	0.1
Baa2	0.1	0.1	0.1	0.1	0.1	0.7	1.6	5.5	**78.6**	5.1	1.2	0.7	0.4	0.2	0.1	0.1	0.0	0.0	0.0	0.0	5.2	0.0
Baa3	0.1	0.0	0.0	0.0	0.2	0.5	0.6	2.2	8.3	**76.2**	3.3	1.9	1.0	0.4	0.2	0.1	0.1	0.0	0.0	0.0	4.6	0.2
Ba1	0.0	0.0	0.1	0.1	0.5	0.4	0.4	1.2	4.1	13.8	**59.8**	3.9	3.4	1.3	0.9	0.5	0.2	0.1	0.0	0.1	9.2	0.2
Ba2	0.0	0.0	0.1	0.1	0.2	0.3	0.2	0.5	1.2	4.5	11.2	**58.6**	4.8	2.9	2.1	1.2	0.7	0.2	0.1	0.2	10.6	0.5
Ba3	0.0	0.0	0.1	0.1	0.0	0.3	0.2	0.2	0.6	1.2	3.8	9.5	**60.6**	4.1	4.3	1.9	0.7	0.4	0.2	0.1	11.1	0.5
B1	0.0	0.0	0.0	0.0	0.1	0.2	0.1	0.3	0.2	0.4	1.0	2.9	8.5	**63.1**	5.4	4.1	1.6	0.9	0.3	0.2	10.0	0.8
B2	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.4	0.8	2.5	7.8	**65.8**	5.3	3.0	1.6	0.6	0.4	9.4	1.6
B3	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.3	0.5	3.0	6.2	**64.9**	4.1	3.6	1.0	1.1	11.5	3.1
Caa1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.2	0.2	0.7	2.3	5.8	**55.0**	4.5	3.6	2.5	17.1	7.7
Caa2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.2	0.0	0.5	0.6	0.8	2.8	3.4	**51.4**	2.3	4.2	19.2	14.0
Caa3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.0	0.1	1.4	0.5	1.4	3.3	4.8	**40.1**	5.7	23.9	18.3
Ca-C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.3	0.7	0.4	1.2	2.2	**35.3**	35.2	24.4

Figure 1: Moody's 2007 forecast of rating transitions for the global rated universe over the following year

1.1.4 Threshold Models

Inspired in the firm-value models presented before, the defining attribute ascribed to Threshold Models is the concept that company i will enter in a situation of default

	--One-year transition rates--								
From/to	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC/C**	**D**	**NR**
AAA	87.17	8.69	0.54	0.05	0.08	0.03	0.05	0.00	3.38
	(9.11)	(9.13)	(0.86)	(0.31)	(0.25)	(0.20)	(0.40)	(0.00)	(2.66)
AA	0.54	86.29	8.36	0.57	0.06	0.08	0.02	0.02	4.05
	(0.55)	(4.90)	(3.99)	(0.75)	(0.25)	(0.24)	(0.07)	(0.07)	(1.91)
A	0.03	1.86	87.26	5.53	0.36	0.15	0.02	0.07	4.71
	(0.13)	(1.15)	(3.47)	(2.10)	(0.49)	(0.35)	(0.07)	(0.11)	(1.91)
BBB	0.01	0.12	3.54	85.09	3.88	0.61	0.14	0.22	6.39
	(0.06)	(0.23)	(2.31)	(4.62)	(1.82)	(1.02)	(0.24)	(0.26)	(1.79)
BB	0.02	0.04	0.15	5.18	76.12	7.20	0.72	0.86	9.71
	(0.06)	(0.16)	(0.39)	(2.35)	(5.02)	(4.63)	(0.92)	(1.04)	(2.84)
B	0.00	0.03	0.11	0.23	5.42	73.84	4.40	4.28	11.68
	(0.00)	(0.13)	(0.37)	(0.33)	(2.50)	(5.30)	(2.52)	(3.32)	(2.98)
CCC/C	0.00	0.00	0.16	0.24	0.73	13.69	43.89	26.85	14.43
	(0.00)	(0.00)	(0.70)	(1.01)	(1.29)	(8.42)	(12.62)	(12.48)	(7.19)

Figure 2: Standard and Poor's multiyear global corporate one-year transition matrix (1981-2012) – number in parentheses are standard deviations.

when some critical random variable $X_i := X_{T,i}$ lies below a certain threshold D_i at its maturity date T.

In Merton's Model, one of the models in the basis of the creation of Threshold Models, X_i represents the log-normally distributed asset value, while d_i represents liabilities; in CreditMetrics, on the other hand, X_i is a normally distributed random variable interpreted as a change in logarithmic asset value.

Notation

Considering a portfolio of m obligors and a fixed time horizon T and defining $S_i := S_{T,i} \in 0, 1, 2, \ldots, n$ as the state indicator for obligor i at time T, where 0 represents a state of default and the remaining values represent increasing credit quality. Assuming that at time $t = 0$ no obligor is in a situation of default, one can have a random vector $\mathbf{X} = (X_1, X_2, .., X_m)$ defined with respect to the deterministic matrix $\mathbf{D}$ such that:

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1i} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2i} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1} & d_{i2} & \cdots & d_{ii} & \cdots & d_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mi} & \cdots & d_{mn} \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_m \end{pmatrix} \tag{1.1.3}$$

Setting the elements of the each row included in matrix $\mathbf{X}$ to be increasing thresholds, i.e. $d_{i1} < \ldots < d_{in}$, then, going back to the state variable:

$$S_i = j \Leftrightarrow d_{ij} < X_i \leqslant d_{i(j+1)} \quad j \in \{0, \ldots, n\}, \quad i \in \{1, \ldots, m\} \tag{1.1.4}$$

And the pair $(\mathbf{X}, \mathbf{D})$ can be defined as a threshold model for the state vector $\mathbf{S} = (S_1, \ldots, S_m)'$.

In the majority of the models inserted in this subclass a binary variable is considered to evaluate the existence of default in particular, the default indicator variable:

$$Y_i = \begin{cases} 1, & S_i = 0 \\ 0, & S_i > 0 \end{cases} \tag{1.1.5}$$

Originating the vector of default indicators for the portfolio $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ and the joint probability function $p(\mathbf{y}) = P(Y_1 = y_1, \ldots, Y_m = y_m), \mathbf{y} \in \{0, 1\}$.

These models are the first to actively try to model default or event correlations, making use, for that purpose, of the correlation between default indicators $\rho(Y_i, Y_j)$.

Another interesting prospect associated with this type of models is the definition of criteria of equivalence between models in terms of the marginal distributions of the state vector $\mathbf{S}$ and the copula function of $\mathbf{X}$[3] , a factor that allows the analysis of structural similarities between distinct industry models for portfolio credit risk management.

Some relevant examples of Threshold Models are the portfolio versions of the KMV and CreditMetrics Models and also Li's Model.

Model Risk

Model Risk represents the risk associated with handling misspecified models, in this case models that serve as poor representations of the mechanism behind default and/or migration in credit quality. Assuming that individual default probabilities have been satisfactorily determined, there is still a possibility that the models in use are not adequate to reality.

One of the main factors that can cause for a Threshold Model to be inadequate is the adoption of a specific faulty copula function. By assuming a Gaussian dependence structure, for example, the probability of large joint movements of risk

[3] In a Threshold model the copula function determines the link between marginal probabilities of migration. For more on Threshold Models and Copulas see McNeil, Frey and Embrechts (2005)[9].

factors may be underestimated, possibly leading to drastic implications on the performance of the risk-management model. When opting for a t distribution, on the other hand, the higher level of dependence in the joint tail of the t copula may create the expectation of occurrence of a superior number of joint defaults.

Another fundamental driver for Threshold Model Risk is the impact of the factor structure of asset returns on joint default events and hence on the tail of the distribution function of the number of defaults. In other words, the recognition of a certain level of asset correlation is directly linked to the level of systematic risk that is associated to the obligors in the portfolio and may have pronounced consequences on the appearance of the distribution of the number of defaults over time and on the allocation of firms to different credit ratings.

1.2 Mixture Models

As static portfolio versions of reduced-form models, Mixture Models demonstrate as an example of models where the mechanism leading to default is left unspecified, being launched from the expectation of dependence between an obligor's default risk and a set of stochastically modeled economic factors, assuming that default of individual firms is independent considering a certain range of realized factors.

1.2.1 Bernoulli Mixture Models

Considering the random vectors:

$$\boldsymbol{\Psi} = \begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_p \end{pmatrix} \quad \text{and} \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} \forall p < m \tag{1.2.1}$$

If $\mathbf{Y}$ follows a Bernoulli Mixture Model with factor vector $\boldsymbol{\Psi}$, then for $\mathbf{y} = (y_1, \ldots, y_m)' \in \{0,1\}^m$:

$$P(\mathbf{Y} = \mathbf{y} | \boldsymbol{\Psi} = \psi) = \prod_{i=1}^{m} p_i(\psi)^{y_i}(1 - p_i(\psi))^{1-y_i} \tag{1.2.2}$$

And the unconditional distribution of the default indicator $\mathbf{Y}$ can be obtained from the integration of the conditional distribution over the distribution of the factor vector $\mathbf{\Psi}$ in order to recover the default probability of company i from its default indicator: $p_i = P(Y_i = 1) = \mathbb{E}(p_i(\mathbf{\Psi}))$.

A possible extension to this model would be to consider a strictly increasing link function $h : \mathbb{R} \to (0, 1)$, such as a Gaussian or Logit distribution function, such that:

$$p_i(\mathbf{\Psi}) = h(\mu + \beta \mathbf{x}_i + \sigma \mathbf{\Psi}) \tag{1.2.3}$$

Which would allow for covariates for individual firms to influence default probability, favoring the contemplation of dependency between defaults[4].

It is important to realize that although Mixture Models seem to present a very different structure from the Threshold Models presented before, the majority of the Threshold Models can be rewritten as Bernoulli Mixture Models, with a number of advantages associated with the second's formulation.

1.2.2 Poisson Mixture Models

Another ordinary appearance for Mixture Models is the application of a Poisson distribution. Considering a vector $\mathbf{\Psi}$ defined as above and a random vector $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_m)'$ that follows a Poisson Mixture Model with factor vector $\mathbf{\Psi}$, one can define a new random variable $\widetilde{M} = \sum_{i=1}^{n} \widetilde{Y}_i$ to be, under certain conditions, approximately equal to the number of defaulting companies at time i and a set of functions $\lambda_i = \mathbb{R}^p \to (0, \infty), 1 \leqslant i \leqslant m$ such that, conditional on $\mathbf{\Psi} = \psi$ the random vector $\widetilde{\mathbf{Y}}$ is a vector of independent Poisson distributed random variables with rate parameter $\lambda_i(\psi)$. Given the factors, $\widetilde{M}$ will be the sum of conditionally independent Poisson variables and thus its distribution satisfies:

$$P(\widetilde{M} = k | \mathbf{\Psi} = \psi) = exp\left(-\sum_{i=1}^{m} \lambda_i(\psi)\right) \frac{\left(\sum\limits_{i=1}^{m} \lambda_i(\psi)\right)^k}{k!} \tag{1.2.4}$$

A widely used industry model for credit risk following Poisson Mixture Models is the CreditRisk+ model, created by Credit Suisse Financial Products in 1997 based on Vasicek's Large Homogeneous Portfolio Model[5]. The particularity of this model is that it assumes $\mathbf{\Psi}$ to consist on a set of p independent gamma-distributed random

[4] For an asymptotic view concerning Bernoulli Mixture Models applied to large portfolios consult section 8.4.3 of McNeil, Frey and Embrechts (2005) [9]

[5] See McNeil, Frey and Embrechts (2005)[9].

variables, allowing for the computation of the distribution of M using distributions that are thoroughly studied and are well-known in actuarial mathematics.

1.2.3 Linear Models

Observing a more general view over the class of Mixture Models, a group of generic models where the probability function is not predetermined can now be analyzed, providing a universal view over this type of models.

Linear Models come as the simplest models lying in the basis of the development of Linear Mixed Models, being used in a vast group of branches of statistics with diverse theoretical and practical uses.

The basic equation of the general model is:

$$\mathbb{E}[Y] = \mu \tag{1.2.5}$$

Where Y is a *random vector* consisting on N Gaussian variables $(Y_1, Y_2, ..., Y_N)$ and follows the distribution:

$$Y \backsim N(\mu, \Sigma) \tag{1.2.6}$$

with *variance-covariance matrix* Σ (i.e. $Var(Y) = \Sigma$) and where μ is usually of the form $\mu = X\beta$ for Y of the form:

$$Y = X\beta + \epsilon \qquad \epsilon : \Omega \to \mathbb{R}^n, \quad \epsilon \backsim N(0, \Sigma) \tag{1.2.7}$$

Where X is a matrix of *known values*, β is a vector of *unknown variables* (to be estimated) and ϵ is the *error vector* composed by independent and identically distributed values, causing this equation to be a matrix equation.

Estimation

Linear Models may be estimated through more than one method. This estimation is frequently performed using Ordinary Least Squares (OLS) or Generalized Least Squares (GLS), both methods not demanding of an underlying distribution for Y but neither of them providing an estimate for the variance of the *random variable* Y.

As it seems a more interesting approach to obtain a more complete result, the method of Maximum Likelihood under Normality may be the best option, providing not only estimates for the variables constituting vector β but also presenting a credible estimate for the value of Σ.

In this sense, starting from a multivariate normal distribution of Y:

$$Y \sim N(X\beta, \Sigma) \tag{1.2.8}$$

The Maximum Likelihood Estimation method could be applied and after a number of computations the Maximum Likelihood estimators for β and σ would be retrieved:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \text{if} \quad (X'X)^{-1} \quad \text{exists and} \quad \hat{\sigma} = (Y - X\hat{\beta})'(Y - X\hat{\beta})/N \tag{1.2.9}$$

An important issue with this type of estimator is that $\hat{\beta}$ exists only if $(X'X)^{-1}$ exists, requiring $X_{N\times p}$ to have full column rank p, which can be quite restrictive.

In a credit risk estimation scenario in particular, a Linear Model could present as:

$$\mathbb{E}[Y_i] = \mu + \alpha_i \quad i = (1, 2, ...) \tag{1.2.10}$$

Where the elements of vector β referred before are the elements of the vector $[\mu \;\; \alpha_1 \;\; \alpha_2 \;\; ...]$, μ represents the general mean, α_i represents effects on the response variable and all of the elements are fixed effects associated with relevant economic factors that can influence credit risk.

1.2.4 Linear Mixed Models

Linear Mixed Models come as a way of adding complexity to Linear Models, following the same principles as the linear regression model:

$$Y = X\beta + \epsilon \tag{1.2.11}$$

Where y is a vector of observations, X is a matrix of known covariates, β the vector of unknown regression coefficients (which we want to find) and ϵ is a vector of unobservable random errors. This model typically considers fixed regression coefficients, a property that tends to change only when there is correlation between observations, causing the coefficients to be random.

Being a generalization of Linear Models, the basic appearance of Linear Mixed Models presents itself as the previous plus an additional parameter associated with the random effects exclusively included in this type of model. The expression for the general model can be displayed as:

$$Y = X\beta + Zu + \epsilon \tag{1.2.12}$$

Where Z is the known (model) matrix, u is a vector of random effects that occur in the data vector y and the general mean is of the form $\mu = X\beta$ with X and β as defined in the prior section. In this case the model obviously depends on the realized (unobservable) random values therefore it would not make sense to consider the simple expected value as opposed to the conditional expected value for the model:

$$\mathbb{E}[Y|U = u] = X\beta + Zu \tag{1.2.13}$$

$$u \sim (0, D), \quad \text{with} \quad \mathbb{E}[u] = 0 \quad \text{and} \quad Var(u) = D$$

Remark 2. *Setting the expected value of u to zero may seem demanding but in fact it makes no difference in the formulation. Supposing that we had a different value to this parameter, i.e.* $\mathbb{E}[u] = \tau$*, we would get:*

$$\mathbb{E}[Y|U = u] = X\beta + Z\tau + Z(u - \tau) \tag{1.2.14}$$

Then, creating three new parameters $u^* = u - \tau$*,* $X^* = [X\ Z]$ *and* $\beta^* = [\beta'\ \tau']$ *we would obtain:*

$$\mathbb{E}[Y|U = u] = X^*\beta^* + Zu^*\ , \quad \mathbb{E}[u^*] = 0 \tag{1.2.15}$$

Which has the exact same structure as the previous conditional mean.

Knowing the variance of u we have the conditional variance of Y:

$$Var(Y|U = u) = R \tag{1.2.16}$$

Causing for the distribution of Y to present as:

$$Y \sim (X\beta, ZDZ' + R) \tag{1.2.17}$$

Showing, as it would be expected, that the fixed effects enter only on the mean, whereas the random effects model matrix and variance affect only the variance of Y.

This type of model can appear as:

$$\mathbb{E}[Y_{ij}] = \mu + a_i + \beta_j + c_{ij} \tag{1.2.18}$$

Where a_i represents the random effect of element i, β_j represents the fixed effect associated with a certain element j and c_{ij} mimics the random effect of the interaction.

An even more general view over Linear Models is the Generalized Linear Mixed Model approach, considering the existence of random effects as a convenient way to specify a correlated data model and a method for adding complexity in order to include systematic factors for each time period, capturing patterns of variability in the responses that cannot be explained by the observed covariates alone, loosely described as the state of the economy. A relevant example of this model is the broadly used Probit Model. [6]

[6] For more information regarding Generalized Linear Mixed Models feel free to read chapter 8 of McCulloch and Searle (2001) [8].

2 Hidden Markov Models for Default

As defined in MacDonald and Zuchinni (2009) [13]"Hidden Markov models (HMMs) are models in which the distribution that generates an observation depends on the state of an underlying and unobserved Markov process."

Hidden Markov Models present themselves as an interesting hypothesis for the modelling of time series in diverse areas of knowledge thanks to its simplicity, accessible mathematical foundations and the reasonably effortless computation of its likelihood, being used in the most varied set of fields, from environmental studies to bioinformatics, having a relevant place in the analysis of financial series such as daily returns of a certain financial instrument.

Before approaching the theme of Hidden Markov Models itself it would be noteworthy to address two important topics indispensable in the description of the models: Independent Mixture Models and Markov Chains.

2.1 Independent Mixture Models

An independent mixture distribution consists on the combination of a finite number of component, discrete or continuous, distributions and a "mixing distribution" whose job is to select from this group of possible distributions.

As an example to facilitate understanding of this type of models we may consider a counting process, frequently modeled using the Poisson distribution, a distribution which is specially convenient thanks to its relatively simple probability function:

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!} \tag{2.1.1}$$

Another interesting feature of this distribution is the property of parity between variance and mean. The result obtained from using such a distribution as a counting process proves to suit well for series of uniform numbers, allowing for simple predictions depending only on parameter λ.

The problem arises when, trying to use the Poisson distribution to model a certain event, there is a large discrepancy between sample mean and variance, thanks to the verification of a high level of dispersion between results.

One possible method to address the issue of over-dispersed observations would be the division of the observations between coherent groups, creating a multimodal distribution modeled by a mixture model.

The resulting independent mixture distribution would then involve a finite number of Poisson component distributions and a "mixing distribution" which would have to select which distribution to apply in each case. The mixture distribution can thus be defined as a function of the probabilities assigned to the different components $\delta_1, \dots, \delta_m$ and their probability or density functions $p_1, \dots, p_m$ and is given by:

$$p(x) = \sum_{i=1}^{m} \delta_i p_i(x) \tag{2.1.2}$$

Which, for the discrete case, translates as:

$$P(X = x) = \sum_{i=1}^{m} P(X = x|C = i)P(C = i)$$

$$\Downarrow$$

$$\mathbb{E}[X] = \sum_{i=1}^{m} P(C = i)\mathbb{E}[X|C = i] = \sum_{i=1}^{m} \delta_i \mathbb{E}[Y_i] \tag{2.1.3}$$

Considering Y_i to correspond to the random variable with probability function p_i. A generalization may be applied in order to obtain the order k moment of X in both the discrete and the continuous cases:

$$\mathbb{E}[X^k] = \sum_{i=1}^{m} \delta_i \mathbb{E}[Y_i^k], \quad k = 1, 2, \dots \tag{2.1.4}$$

The same result does not hold for central moments as, for example, the variance of X is not a linear combination of the variances of its components Y_i.

2.2 Markov Chains

A sequence of random variables $\{C_t : t \in \mathbb{N}\}$ is a Markov Chain if it satisfies the Markov "lack of memory" property:

$$P(C_{t+1}|C_t, \dots, C_1) = P(C_{t+1}|C_t) \tag{2.2.1}$$

Considering $\{C_t\}$ to be dependent in a way that is mathematically relevant. From this formulation we can now define the transition probabilities associated with a certain moment in time t:

$$\gamma_{ij}(s,t) = P(C_{t+s} = j | C_t = i) \tag{2.2.2}$$

If these probabilities are independent of the moment in time t the Markov Chain can be denominated **homogeneous**. Usually a Markov Chain is assumed to be homogeneous unless stated otherwise.

The transition probabilities can be written in the form of a square matrix $\mathbf{\Gamma}$:

$$\mathbf{\Gamma}(s) = \begin{pmatrix} \gamma_{11}(s) & \cdots & \gamma_{1m}(s) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(s) & \cdots & \gamma_{mm}(s) \end{pmatrix} \tag{2.2.3}$$

Where m denotes the number of states the Markov Chain can take, the sum of row values is equal to 1 and $\mathbf{\Gamma}$ can be designated by one-step **transition probability matrix** or simply **transition matrix**. Finite state space homogeneous Markov Chains satisfy the Chapman-Kolmogorov equations:

$$\mathbf{\Gamma}(t+u) = \mathbf{\Gamma}(t)\mathbf{\Gamma}(u) \tag{2.2.4}$$

Implying that

$$\mathbf{\Gamma}(t) = \mathbf{\Gamma}(1)^t \tag{2.2.5}$$

Which, translated to words, means that the matrix of the t-step transition probabilities is equivalent to the tth power of the one-step transition probability matrix.

The unconditional probability $P(C_t = j)$ can be defined as the probability of the Markov Chain presenting in a given state j at time t and can be written as the vector:

$$\mathbf{u}(t) = (P(C_t = 1), \ldots, P(C_t = m)), \forall t \in \mathbb{N} \tag{2.2.6}$$

This probability can, thus, be obtained using an iterative technique from the initial distribution of the Markov Chain $\mathbf{u}(1)$ or simply from the last calculated probability by multiplying by the transition matrix $\mathbf{\Gamma}$:

$$\mathbf{u}(t+1) = \mathbf{u}(t)\mathbf{\Gamma} \tag{2.2.7}$$

A Markov Chain is said to have **stationary distribution** δ if $\delta\Gamma = \delta$ and $\delta\mathbf{1}' = 1$ where δ is a row vector with non-negative elements and $\mathbf{1}'$ is a column vector where all the elements are equal to 1. A Markov Chain started from a stationary distribution will continue to have that distribution, being referred to as a **stationary Markov Chain**.

An **irreductible** Markov Chain is an homogeneous, discrete-time, finite state-space Markov Chain, being characterized for having a unique, strictly positive, stationary distribution.

Autocorrelation Function

Another characteristic associated with a Markov Chain that can be interesting to analyze is its **autocorrelation function**. In order to do it we first need to describe its covariance function, defining vector $\mathbf{v} = (1, 2, ..., m)$ and matrix $\mathbf{V} = diag(1, 2, ..., m)$, then for all non-negative integers k we have:

$$Cov(C_t, C_{t+k}) = \delta\mathbf{V}\mathbf{\Gamma}^k\mathbf{v}' - (\delta\mathbf{v}')^2 \tag{2.2.8}$$

Where δ is, as defined earlier, the stationary distribution of the Markov Chain. Then, for $\mathbf{\Gamma}$ diagonalizable with eigenvalues ω_i, $\mathbf{\Omega} = diag(1, \omega_2, \omega_3, ..., \omega_m)$ and $\mathbf{U}$ a matrix whose columns are corresponding right eigenvectors of $\mathbf{\Gamma}$ the following holds:

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{-1} \tag{2.2.9}$$

For all non-negative integers k the covariance function of C_t can be given by:

$$\begin{aligned} Cov(C_t, C_{t+k}) &= \delta\mathbf{V}\mathbf{U}\mathbf{\Omega}^k\mathbf{U}^{-1}\mathbf{v}' - (\delta\mathbf{v}')^2 \\ &= \mathbf{a}\mathbf{\Omega}^k\mathbf{b}' - a_1 b_1 \\ &= \sum_{i=2}^{m} a_i b_i \omega_i^k \end{aligned} \tag{2.2.10}$$

where $\mathbf{a} = \delta\mathbf{V}\mathbf{U}$ and $\mathbf{b}' = \mathbf{U}^{-1}\mathbf{v}'$, from where we can conclude that the variance of the Markov Chain is given by:

$$Var(C_t) = \sum_{i=2}^{m} a_i b_i \tag{2.2.11}$$

And that for all non-negative integers k the autocorrelation function of C_t is given by:

$$\rho(k) = Corr(C_t, C_{t+k}) = \frac{\sum_{i=2}^{m} a_i b_i w_i^k}{\sum_{i=2}^{m} a_i b_i} \tag{2.2.12}$$

The final step in the scrutiny of a Markov Chain would be the estimation of the corresponding transition probabilities. The likelihood of the $m^2 - m$ parameters $\gamma_{ij}, i \neq j$ conditioned on the first observation is:

$$L = \prod_{i=1}^{m} \prod_{j=1}^{m} \gamma_{ij}^{f_{ij}} \tag{2.2.13}$$

Where f_{ij} denotes the number of transitions observed from state i to state j. Making the log-likelihood:

$$l = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} f_{ij} log \gamma_{ij} \right) \tag{2.2.14}$$

Defining l_i to be $l_i = \sum_{j=1}^{m} f_{ij} log \gamma_{ij}$ we can rewrite l as:

$$l = \sum_{i=1}^{m} l_i \tag{2.2.15}$$

Now we can maximize l by maximizing l_i individually. Deriving l and making it equal to zero we obtain the maximum likelihoods:

$$\gamma_{ii} = \frac{f_{ii}}{\sum_{j=1}^{m} f_{ij}} \quad \text{and} \quad \gamma_{ij} = \frac{f_{ij}\gamma_{ii}}{f_{ii}} = \frac{f_{ij}}{\sum_{j=1}^{m} f_{ij}} \tag{2.2.16}$$

Obtaining the estimator:

$$\hat{\gamma_{ij}} = \frac{f_{ij}}{\sum_{k=1}^{m} f_{ik}} \quad \forall \quad i, j \in \{1, \ldots, m\} \tag{2.2.17}$$

Alternatively, this valuation can be done making use of a set of observations, by computing the transition counts and from those values calculating the transition probabilities.

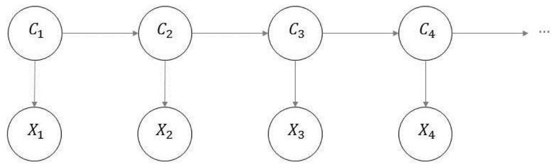


Figure 3: Outline of a basic Hidden Markov Model

2.3 Hidden Markov Models

Also known as "Hidden Markov Processes", "Markov-Dependent Mixtures" or "Markov Mixture Models", **Hidden Markov Models** represent a particular kind of dependent mixture, innovative by allowing for the existence of serial dependence among observations, a characteristic common to most time series in the real world.

The Hidden Markov Model $\{X_t : t \in \mathbb{N}\}$ can be divided in two parts: an unobserved parameter process $\{C_t : t = 1, 2, \ldots\}$ that satisfies the Markov property as described in 2.2.1 and a state-dependent process $\{X_t : t = 1, 2, \ldots\}$ whose distribution depends only on the current state C_t, as schematized in Figure 3, satisfying the Hidden Markov Model property:

$$P(X_t|C_t, C_{t-1}, ..., X_{t-1}, X_{t-2}, ...) = P(X_t|C_t) \tag{2.3.1}$$

Considering a Markov Chain $\{C_t\}$ with m states we can refer to $\{X_t\}$ as an m-state Hidden Markov Model and we may define for both discrete and continuous-valued observations:

$$p_i(x) = P(X_t = x|C_t = i) \quad \forall \quad i = 1, 2, \ldots, m \tag{2.3.2}$$

Although the interpretation differs between the two cases. For discrete-value observations p_i can be interpreted as the probability mass function of X_t if the Markov Chain is in state i at time t. Meanwhile, in the case of continuous-value observations, p_i represents the probability density function if the Markov Chain is in state i at time t. In both cases the set $p_1, \ldots, p_i, \ldots, p_m$ contains all the **state-dependent distributions** of the model.

In order to derive necessary conclusions from the model, it may be useful to obtain the **distribution** of X_t that, for discrete-valued observations, the less complex case on which this dissertation is focused, can be defined as:

$$P(X_t = x) = \sum_{i=1}^{m} P(C_t = i)P(X_t = x|C_t = i) = \sum_{i=1}^{m} u_i(t)p_i(x) \tag{2.3.3}$$

$$\text{where } u_i(t) = P(C_t = i) \quad \forall \quad t \in \{1, \dots T\}$$

Which, translated to matrix notation can be written as:

$$P(X_t = x) = (u_1(t), \dots, u_m(t)) \begin{pmatrix} p_1(x) & 0 & \cdots & 0 \\ 0 & p_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_m(x) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{2.3.4}$$

$$= \mathbf{u}(t)\mathbf{P}(x)\mathbf{1}'$$

But from equation 2.2.7 we know that $\mathbf{u}(t) = \mathbf{u}(1)\mathbf{\Gamma}^{t-1}$, so the previous equation translates as:

$$P(X_t = x) = \mathbf{u}(1)\mathbf{\Gamma}^{t-1}\mathbf{P}(x)\mathbf{1}' \tag{2.3.5}$$

If we assume the Markov Chain to be stationary with stationary distribution δ the result will be:

$$P(X_t = x) = \delta\mathbf{P}(x)\mathbf{1}' \tag{2.3.6}$$

As we have defined before that for stationary Markov Chains the following holds:

$$\delta\mathbf{\Gamma}^{t-1} = \delta \tag{2.3.7}$$

It may also be interesting to compute the moments of the distribution of X_t:

$$\mathbb{E}[X_t] = \sum_{i=1}^{m} \mathbb{E}[X_t|C_t = i]P(C_t = i) = \sum_{i=1}^{m} u_i(t)\mathbb{E}[X_t|C_t = i] \tag{2.3.8}$$

That, in the stationary case can be written as:

$$\mathbb{E}[X_t] = \sum_{i=1}^{m} \delta\mathbb{E}[X_t|C_t = i] \tag{2.3.9}$$

Which, generalizing for $\mathbb{E}[g(X_t)]$ and $\mathbb{E}[g(X_t, X_{t+k})]$ translates as:

$$\mathbb{E}[g(X_t)] = \sum_{i=1}^{m} \delta_i \mathbb{E}[g(X_t)|C_t = i] \tag{2.3.10}$$

and

$$\mathbb{E}[g(X_t, X_{t+k})] = \sum_{i,j=1}^{m} \mathbb{E}[g(X_t, X_{t+k})|C_t = i, C_{t+k} = j]\delta_i \gamma_{ij}(k) \tag{2.3.11}$$

where $\gamma_{ij}(k) = (\mathbf{\Gamma}^k)_{ij} \forall k \in \mathbb{N}$. We may also be interested in a function g that factorizes $g(X_t, X_{t+k}) = g_1(X_t)g_2(X_{t+k})$, making the previous equation:

$$\mathbb{E}[g(X_t, X_{t+k})] = \sum_{i,j=1}^{m} \mathbb{E}[g_1(X_t)|C_t = i]\mathbb{E}[g_2(X_{t+k})|C_{t+k} = j]\delta_i \gamma_{ij}(k) \tag{2.3.12}$$

These expressions allow the calculation of covariates and correlations, very useful in the use of Hidden Markov Models in practice.

2.3.1 Estimation of HMM using the method of Maximum Likelihood

As a final result, it may be relevant to develop an explicit formula for the likelihood L_T of T consecutive observations $x_1, x_2, \ldots, x_T$ generated by an m-state Hidden Markov Model with initial distribution δ, transition matrix $\mathbf{\Gamma}$ and state-dependent probability (or, in the continuous case, density) functions p_i, defining the probability of observing such a sequence of observations.

$$L_T = P(X_1 = x_1, X_2 = x_2, \ldots, X_T = x_T) = \delta \mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2)...\mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}' \tag{2.3.13}$$

Which, in the case the initial distribution δ represents the stationary distribution of the Markov Chain, translates as:

$$L_T = \delta \mathbf{\Gamma}\mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2)...\mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}' \tag{2.3.14}$$

Proof. As we know the likelihood of the Hidden Markov Model can be obtained from the computation of:

$$\begin{aligned} L_T &= P(X_1 = x_1, X_2 = x_2, ..., X_T = x_T) \\ &= \sum_{c_1,...,c_T=1}^{m} P(X_1 = x_1, X_2 = x_2, ..., X_T = x_T, C_1 = c_1, C_2 = c_2, ..., C_T = c_T) \end{aligned} \tag{2.3.15}$$

But from Bayes Theorem we know that for a bivariate distribution the following holds:

$$P(Y_t, Y_{t+k}, V_t, V_{t+k}) = P(Y_t)P(Y_t|V_t)P(V_{t+k}V_t)P(Y_{t+k}|V_{t+k}) \tag{2.3.16}$$

Therefore we obtain:

$$\begin{aligned} &P(X_1 = x_1, ..., X_T = x_T, C_1 = c_1, C_2 = c_2, ..., C_T = c_T) = \\ &\qquad = P(C_1) \prod_{k=2}^{T} P(C_k|C_{k-1}) \prod_{k=1}^{T} P(X_k|C_k) \end{aligned} \tag{2.3.17}$$

Which finally translates to:

$$\begin{aligned} L_T &= \sum_{c_1,...,c_T=1}^{m} \left(\delta_{c_1}\gamma_{c_1,c_2}\gamma_{c_2,c_3}...\gamma_{c_{T-1},c_T}\right) \left(p_{c_1}(x_1)p_{c_2}(x_2)...p_{c_T}(x_T)\right) \\ &= \sum_{c_1,...,c_T=1}^{m} \delta_{c_1}p_{c_1}(x_1)\gamma_{c_1,c_2}p_{c_2}(x_2)\gamma_{c_2,c_3}...\gamma_{c_{T-1},c_T}p_{c_T}(x_T) \\ &= \delta\mathbf{P}(x_1)\mathbf{\Gamma}\mathbf{P}(x_2)...\mathbf{\Gamma}\mathbf{P}(x_T)\mathbf{1}' \end{aligned} \tag{2.3.18}$$

□

In order to simplify the notation we may now consider a matrix $\mathbf{B}_t$ defined as $\mathbf{B}_t = \mathbf{\Gamma}\mathbf{P}(x_t)$, making the likelihood present as:

$$L_T = \delta\mathbf{P}(x_1)\mathbf{B}_2\mathbf{B}_3...\mathbf{B}_T\mathbf{1}' \tag{2.3.19}$$

And in the case when δ is the stationary distribution of the Markov Chain:

$$L_T = \delta\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3...\mathbf{B}_T\mathbf{1}' \tag{2.3.20}$$

Inserting another variable in order to facilitate computation we can define the vector α_t as:

$$\alpha_t = \delta\mathbf{P}(x_1)\mathbf{\Gamma P}(x_2)...\mathbf{\Gamma P}(x_T) = \delta\mathbf{P}(x_1)\prod_{s=2}^{t}\mathbf{\Gamma P}(x_s) \quad (2.3.21)$$

i.e.

$$\alpha_t = \delta\mathbf{P}(x_1)\prod_{s=2}^{t}\mathbf{B}_s \quad (2.3.22)$$

Making the likelihood:

$$L_T = \alpha_T\mathbf{1}' \quad \text{and} \quad \alpha_t = \alpha_{t-1}\mathbf{\Gamma P}(x_t) \quad \text{for } t \geqslant 2 \quad (2.3.23)$$

We can now present the elements necessary for the computation of formula 2.3.13:

$$\alpha_1 = \delta\mathbf{P}(x_1) \quad (2.3.24)$$

$$\alpha_t = \alpha_{t-1}\mathbf{\Gamma P}(x_t) \quad \text{for } t = 2, ..., T \quad (2.3.25)$$

$$L_T = \alpha_{\mathbf{T}}\mathbf{1}' \quad (2.3.26)$$

Meaning that for every value of t there will be m elements of α_t to be computed and each of those will be a sum of m products of an element of α_{t-1}, a transition probability γ_{ij} and a state-dependent probability (or, for the continuous case, density) $p_j(x_t)$.

Alternatively, one could estimate Hidden Markov Models using methods such as the EM algorithm, which resorts to forward and backward probabilities and is also used for decoding and state prediction.

2.3.2 Estimation using the Baum-Welch algorithm

A possible alternative to the direct application of Maximum Likelihood Estimation is the use of the Baum-Welch algorithm, a method that merely consists on an instantiation of the more general Expectation-Maximization (EM) algorithm and that works by maximizing a proxy to the log-likelihood, iteratively updating the model in order to become closer to the "optimal model".

Also known as the Forward-Backward algorithm, the **Baum Welch algorithm** is an iterative procedure for estimating the parameters θ of a Hidden Markov Model which, in our case, would result in the estimation of a vector containing expected values, λ, and a transition probabilities matrix, Γ [7].

By maximizing a proxy to the log-likelihood and updating the current model to be closer to the optimal model, this model guarantees the escalation of the log-likelihood of the data in each iteration, making the convergence to the optimal solution highly likely (although, just like in all estimation methods, an optimal solution is never guaranteed).

Considering R to be the hidden state variable, X to be the observation sequence and θ to be the parameters array, the Baum-Welch method will work by repeatedly computing the following steps:

1. Calculation of $Q(\theta, \theta^s) = \sum_{r \in \mathbb{Z}} log\,[P(X, r, \theta)]\; P(r|X; \theta^s)$;

2. Setting $\theta^{s+1} = \underset{\theta}{argmax}\; Q(\theta, \theta^s)$.

Further in this dissertation this algorithm will be computed along with the Maximum Likelihood Estimation method applied to our model in order to draw conclusions related to the adequacy of the use of the MLE technique in model estimation.

2.4 Description of the Model

After this brief presentation of the support contents, it is time to present the model in the center of the development of this dissertation: the model of Giacomo Giampieri, Mark Davis and Martin Crowder, from now on referred to as the GDC model.

The basic premise of the GDC discrete time Hidden Markov Model (HMM) for credit default lays on the so called *paradox in infection modeling*, a theory that goes as follows:

- Suppose that there are two portfolios, A and B, containing assets from n different obligors.

- The only difference between the two portfolios is that portfolio B has an additional asset issued by a distinct obligor $n + 1$.

[7] Further detail on this choice of parameters will be provided in the next sections.

- Previous studies on infection modeling show that in the case where obligor $n+1$ would enter a situation of default it would be expected that the remaining issuers in portfolio B would enter a "high risk state", while for portfolio A, considered in isolation for not having witnessed any default events, no reaction would be expected.

- The interesting factor that can easily be recognized is that, after the default of $n + 1$, portfolios A and B would be exactly the same, except that the first is considered to be in a considerably less risky situation than the second, proving the necessity for consideration of outside world influences in infection modeling.

In order to address this type of situations, the GDC model assumes occurrence of defaults to be modeled as a Hidden Markov process where the hidden variable represents a risk state assumed to be common to all bonds within a particular sector, geography or, as will later be analyzed, rating position.

The GDC discrete time Hidden Markov Model assumes, as the name suggests, a discrete time spam, accepting the fact that the current state is not observable but can be evaluated in each time step through a certain output to which a predetermined probability distribution is associated.

Considering N_0 to represent the initial number of bonds in the portfolio, N_t to be the number of bonds in the portfolio at time t and X_t to be the number of defaults happening at time t, it is easy to understand that $X_t \leqslant N_t$ and that, considering a closed portfolio with no possibility of addition of assets, the number of bonds in the portfolio at time $t + 1$ will be given by: $N_{t+1} = N_t - X_t$.

Remark 3. *Considering a portfolio with possibility of entry of new bonds, such as the market, it is necessary to add a new variable M_t that will influence the value of N_t. The number of bonds in the portfolio at time $t + 1$ will, thus, be given by: $N_{t+1} = N_t - X_t + M_{t+1}$. Comprehensibly, the behavior of this new variable will not be studied, as it is influenced by exterior forces.*

The model assumes N_t and consequently X_t to be influenced by a hidden state variable associated with the previously referred effects of external influences which are directly connected to the level of risk held by a certain asset and whose value

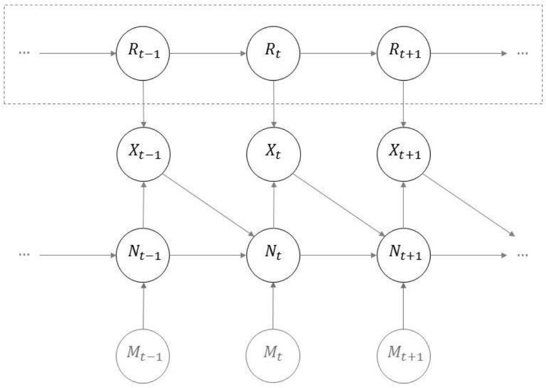


Figure 4: Hidden Markov Models Scheme

is given by the observations registered in a Markov Chain. The model can be schematized as shown in figure 4.

Where the variables inserted in the dashed rectangle, the denominated **Hidden Markov Models** R_i, are not observable in the market, being defined as values drawn from a Markov Chain that verify the "lack of memory" Markov Property:

$$P(R_{t+1} = j | R_0 = i_0, \dots, R_t = i_t) = P(R_{t+1} = j | R_t = i_t) \tag{2.4.1}$$

The general model, thus, defines the number of defaults X_t to be given by a certain function that depends on the number of bonds in the portfolio N_t and on the risk state i associated to each of these bonds, i.e.

$$P(X_t = x | R_t = i) = \varphi(x, i, N_t) \tag{2.4.2}$$

Where $\{R_t\}$ is an m-state Markov Chain.

2.4.1 The GDC Binomial Model

In the specific case of the GDC Model presented in [5] $\{R_t\}$ is assumed to be a 2-state Markov Chain such that:

$$R_t = \begin{cases} 0 & \text{low risk level category} \\ 1 & \text{high risk level category} \end{cases}$$

And the probability of occurrence of a certain level of defaults x is given according to the Binomial Distribution formulation:

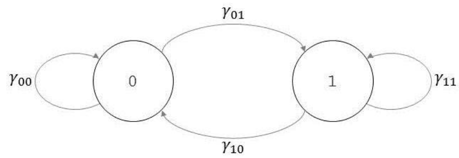


Figure 5: Transition Probabilities scheme.

$$P(X_t = x|R_t = i) = \binom{N_t}{x} p_i^x (1 - p_i)^{N_t - x} \tag{2.4.3}$$

$$X_t|R_t \sim Binomial(p_i, N_t)$$

Conditional on the risk state i.

The particularity that distinguishes this model from other models that resort to a Binomial distribution to model default is the recognition of the fact that default probabilities are not constant in time and between risk levels, considering p_i to be the parameter for each time step, where:

$$p_i = \begin{cases} p_0 & \text{low risk level category} \\ p_1 & \text{high risk level category} \end{cases}$$

As a formulation simplification p_0 can be relabeled as p, i.e. $p_0 = p$, and as $p_1 = p * k$ where k represents a multiplication factor such that $k \geqslant 1$. It is relevant to stress that the situation where $k = 1$ represents the special case when risk level does not cause for changes in default probability, which means that $P(X_t = x|R_t = 0) = P(X_t = x|R_t = 1)$.

The transition matrix can now be defined as Γ, the matrix containing the probability of transition from one state to another conditional on the initial state. Since only two risk level categories are at the time being considered, for a unit time step the transition matrix will be 2×2, being presented as:

$$\Gamma = \begin{pmatrix} \gamma_{00} & \gamma_{01} \\ \gamma_{10} & \gamma_{11} \end{pmatrix} \tag{2.4.4}$$

Where γ_{ij} for $i \neq j$ represents the probability of transitioning from risk level category i to risk level category j and γ_{ii} represents the probability of remaining in the same risk level category, $i, j = \{0, 1\}$ (see figure 5).

The model is, thus, fully described by the set of variables (p, k, Γ, N_0).

2.4.2 A new approach

As it is known, the Binomial and Poisson distributions are similar, both of them being used to measure the occurrence of certain events (or, as commonly stated in statistical publications, to "count the success cases") within a certain frame.

The main divergence between the two is associated with the fact that the Binomial distribution is based on discrete events, requiring as an input the number of attempts N and the probability of success p, while the Poisson distribution is based on "endless" episodes, that is, the number of attempts is assumed to tend to infinite and the chance of success to be infinitesimal. The particularity that connects the two is, then, intuitive: if we assume $N \to \infty$ and $p \to 0$ in such a way that $Np \to \lambda$, then the Binomial distribution approaches a Poisson distribution with parameter λ (the "rate of success").

The new approach suggested in this dissertation is, thus, connected to this matter: the approximation of the probability of occurrence of a certain level of defaults x, previously modeled resorting to a Binomial distribution, to a Poisson distribution, assuming that the number of "attempts", i.e. the number of bonds considered in the portfolio, is large enough and the probability of default small enough to make this approximation valid.

The probability of occurrence of a certain level of defaults x will, thus, according to our new approach, be given by:

$$P(X_t = x | R_t = i) = \frac{e^{-\lambda_i} \lambda_i^x}{x!} \tag{2.4.5}$$

Where

$$X_t | R_t \sim Poisson(\lambda_i) \quad \text{and} \quad \lambda_i \approx p_i N_t$$

One other innovation that could be interesting to test is the possibility of consideration of an intermediate risk level, a "normal risk level category", allowing for the evaluation of

$$R_t = \begin{cases} 0 & \text{low risk level category} \\ 1 & \text{normal risk level category} \\ 2 & \text{high risk level rating category} \end{cases}$$

2.5 Estimation of the Model

2.5.1 Baum-Welch algorithm versus MLE

One other aspect in which the method followed in this dissertation differs from that used in the development of the GDC model is in the estimation approach: while Giampieri, Davis and Crowder chose to use the Baum-Welch algorithm to test the fitting of the model to real data, our choice was to apply the method of Maximum Likelihood Estimation, making it necessary to demonstrate that results using both techniques would be identical and both good estimators for our model, and that, therefore, our results will not be clouded by our choice in estimation technique.

In order to prove that the Maximum Likelihood Estimation Method could be applied without loss of accuracy, a simple example was used, testing the estimation procedure for a simulated 2-state Hidden Markov Model with parameters:

$$\lambda = (1 \quad 23) \quad \text{and} \quad \mathbf{\Gamma} = \begin{pmatrix} 0.80 & 0.20 \\ 0.25 & 0.75 \end{pmatrix}.$$

This computation resulted on the following Baum-Welch and Maximum Likelihood estimators, respectively:

$$\hat{\lambda}_{BW} = (0.946 \quad 22.701) \quad \text{and} \quad \hat{\mathbf{\Gamma}}_{BW} = \begin{pmatrix} 0.843 & 0.157 \\ 0.267 & 0.733 \end{pmatrix}.$$

$$\hat{\lambda}_{MLE} = (0.946 \quad 22.700) \quad \text{and} \quad \hat{\mathbf{\Gamma}}_{MLE} = \begin{pmatrix} 0.841 & 0.159 \\ 0.265 & 0.735 \end{pmatrix}.$$

Demonstrating that both methods can be used successfully on the estimation of the parameters of a Hidden Markov Model. [8]

2.5.2 Maximum Likelihood Estimation of the 2-state HMM

In order to carry out with the estimation of the model, the R library of Standard & Poor's default data for A, BBB, BB, B and C-rated companies for the years 1981 to 2000 was consulted and used.

[8]See Appendix B.1 for further details on the computation method and Appendix C for the output of the experience in R.

Because this is a series of data withdrawn from the real market over a period of 20 years, the total number of companies is not stable, making the data very inconstant and quite difficult to use in practice. In order to address this question, before starting to implement the model, a series of manipulations were done in order to provide relative default values for a hypothetical set of 1000 companies per rating, per year.

Starting with the computation of the method, the first step was to apply the Maximum Likelihood Estimation method to estimate the parameters of the Hidden Markov Model applied to real data. Starting by the definition of a set of 2 states, a "low" risk state and a "high" risk state, and by the definition of the initial distribution of the Markov Chain. Since the data starts in year 1981 with zero defaults for the hole portfolio, the initial distribution is assumed to be deterministic, being defined as being given by the vector $(1 \quad 0)$, i.e. the Markov Chain presents in the low risk state with 100% probability.

The next step is to give an initial guess for the parameters λ and Γ, which were defined as:

$$\lambda(0) = (\mu_{SP} \times \epsilon \quad \mu_{SP} \times (1+\epsilon)) \quad \text{and} \quad \mathbf{\Gamma}(0) = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$$

Where μ_{SP} is the mean of the default series considered at the moment (the hole portfolio or just one of the ratings at a time) and ϵ is a disruptive factor that decreases expected default in the low risk state and increases expected default in the high risk state. In a first stage the value 0.5 was tested for ϵ. In the case where the entire Standard & Poor's data series is considered, the initial guess for λ will, thus, be:

$$\lambda(0) = (125.25 \quad 375.75)$$

It is important to highlight that, being an iterative method, Maximum Likelihood will start from these guesses to find the Maximum Likelihood estimators, making these first guesses, although important, not decisive in the outcome of our estimation.

The Maximum Likelihood Estimators obtained for the total series of data from the Standard & Poor's library where:

$$\hat{\lambda}_{Total} = (124.22 \quad 353.82) \quad \text{and} \quad \hat{\mathbf{\Gamma}}_{Total} = \begin{pmatrix} 0.4734 & 0.5266 \\ 0.4313 & 0.5687 \end{pmatrix}$$

Proving that our considered disruptive factor must be quite accurate ($\hat{\lambda}$ is quite close to $\lambda(0)$) and that, for the total set of data, there is not a pattern when it comes to transition probabilities, as the probability of transition between risk states resembles the probability of remaining in the same state.

After obtaining the Maximum Likelihood Estimators, these values were used as inputs for the estimation of a new Hidden Markov Model series with the aim of comparing the real data series with a Hidden Markov Model obtained from this series' expected value and transition probability parameters. Plotting a graph including both series, see figure 6, and comparing both series can be observed that the real data series (black) and the estimated data series (red) demonstrate, in most of the cases, a similar behavior, having several common growth and decrease periods.

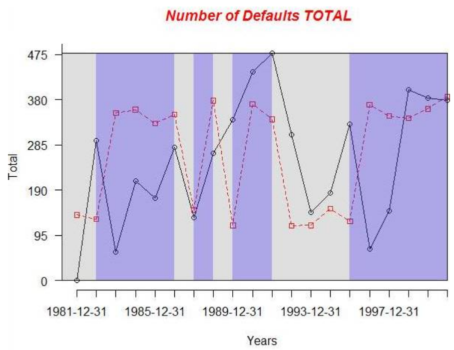


Figure 6: Comparison between the real data series (black) and the estimated series (red) for the total S&P data series

It can also be observed that, for this experience, the periods estimated as high risk periods, indicated with the blue rectangles, coincide with the majority of the high default occurrence periods, pointing to the idea that our model can be used to successfully estimate default events.

The same method can be administered in order to test the existence of infection effects originated by the risk state variable between companies within the same rating category if instead of applying the Maximum Likelihood Estimation approach for the entire data set, the Maximum Likelihood Estimates were obtained for the parameters for a set of data from a particular rating [9].

By observing, for example, the output of the Maximum Likelihood estimation for companies in rating A:

$$\hat{\lambda}_A = (0.141 \quad 2.151) \quad \text{and} \quad \hat{\mathbf{\Gamma}}_A = \begin{pmatrix} 0.8174 & 0.1826 \\ 1.0000 & 4.32 \times 10^{-9} \end{pmatrix}$$

[9] See Appendix A for a clear view of the available S&P data used in our estimation.

The results appear trustworthy: being, in theory, a low risk level rating category, it seems legitimate that the low risk level rating category should be bulged, resulting on a very high probability of, starting from either risk state, transiting to (or remaining in) the low risk state, a phenomenon that can actually be observed in the transition matrix $\mathbf{\Gamma}$.

Observing the graphic combining the real data series of default and the estimated series, once again it seems to point to the conclusion that not only the model seems to be fairly good at estimating defaults, a deduction made from the proximity between peaks in both lines, but there also seems to be a link between the high risk level periods and the observation of a higher number of default events.

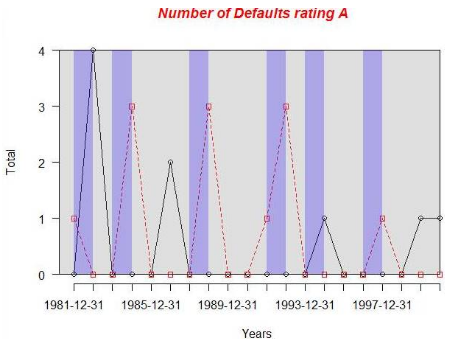


Figure 7: Comparison between the real data series (black) and the estimated series (red) for rating A. The blue areas indicate the high risk state periods.

Scrutinizing the results obtained for the Maximum Likelihood estimators for the remaining rating categories[10] the conclusion that the higher the risk associated with the issuer's rating the more likely it is that once the obligor enters a high risk state it will stay there for another period continues to be pointed, a result that would be expected and seems like a good indicator for the quality of the model.

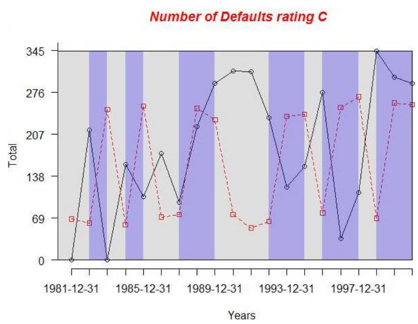


Figure 8: Comparison between the real data series (black) and the estimated series (red) for rating C.

One other phenomenon that can be verified for all rating categories is the overlap of high level of default periods with the estimated high risk level periods which, combined with the relative approximation between the estimated default and the observed default events series, leads to the conclusion that although this is a very simple model, it is capable of estimating default events with relative success. It is, however, important to emphasize that further testing would be necessary in order to truly prove the reliability of these conclusions.

[10]See Appendix C.

2.5.3 Maximum Likelihood Estimation of the 3-state HMM

A possible flaw associated with the 2-state model presented in the previous section is associated with the exclusive consideration of extreme situations, assuming that an obligor can only be subjected to either a high risk state or a low risk state, possibly causing for a bigger "resistance" against transition between risk states. An evolution that could, therefore, be suggested is the insertion of a third, intermediate, risk state, connected to situations in which risk level is neither high nor low.

Proceeding in the same way as for the 2-state Hidden Markov Model, the first step of the estimation process is the determination of the initial distribution of the Markov Chain which, for the same reasons as before, is assumed to present in the low risk level category, making the initial distribution $(1 \quad 0 \quad 0)$.

The next step will, like before, be the definition of the initial guesses for the parameters:

$$\lambda(0) = (\mu_{SP} \times \epsilon \quad \mu_{SP} \quad \mu_{SP} \times (1+\epsilon)) \quad \text{and} \quad \mathbf{\Gamma}(0) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Where $\lambda(0)$ is defined using the same logic as before, only adding the intermediate level expected value as the mean of the series, and the transition probabilities inserted in $\mathbf{\Gamma}(0)$ are assumed to be unknown variables, being, therefore, defined as the equitable division of the probability between the different risk levels.

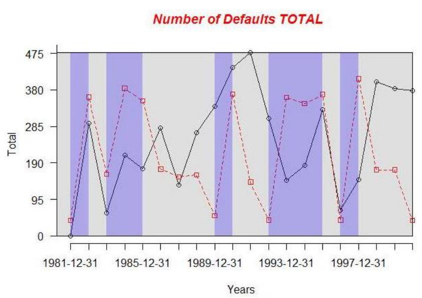


Figure 9: Comparison between the real data series (black) and the estimated series (red) for the total S&P data series.

After obtaining the Maximum Likelihood Estimators and using them as inputs in the estimation of a Hidden Markov Model to simulate a new series of defaults, the combined graph for the entire S&P data series was plotted. Once again, the blue areas represent the high risk level periods and it is easy to verify, not only in this graph but also for the different rating groups [a], that there is a relation between these areas in the graph and the pikes in the number of defaults, united with an evident resemblance between the estimated default and the verified default lines.

[a] See Annex C.

3 Conclusions

An extension of the model proposed by Giampieri, Davis and Crowder was analyzed, testing the hypothesis of modeling of the occurrence of defaults within a certain portfolio as a simple Hidden Markov process in which the hidden state variable is assumed to be a risk state common to all bonds within a portfolio, a market or a rating category.

In an initial stage, a situation where only two risk states were available was analyzed, a high risk and a low risk levels, and, after computation of the model assuming a real set of data, it was concluded that the results seem to point to the hypothesis that not only the model is accurate in the estimation of default events but also there is a direct link between the "observation" of this high risk level and the observation of notably higher levels of default.

As an evolution to this model, the insertion of an intermediate risk state was suggested, aiming towards the diminution of the theoretical gap between the available risk states. Altogether, this evolution demonstrated that, as expected, firms that are in a normal risk state seem to tend to stay there, while in general the probability of an issuer migrating to lower risk states tends to be bigger for the best rating categories (issuers that, theoretically, are less risky) than for the lower ratings. Additionally, the obtained results seem to point towards a direct relation between the estimated data and real data, once more directing towards a perception of success in the estimation. It is important to stress that, although these results seem satisfactory and encouraging, further testing would be necessary in order to undeniably verify the accuracy of these results.

A limitation that could be pointed regarding the use of this model is related to the obligation of providing initial guesses for the parameters to consider. This problem is minimized when it comes to transition probabilities, since the possibilities are limited to numbers between 0 and 1, but is more critical for the expected value vector λ, since, as verified, the universe of possibilities is very wide. This detail is specially problematic since expected defaults are not easy values to obtain.

In conclusion, the model introduced in this thesis is certainly very simple, and with no regard for certain aspects such as economic cycles, but it manifests as intended: it is a procedure with reduced computation complexity that, in its simplicity seems to be able to rather accurately model default events.

References

[1] Altman, E.I. and Saunders, A., "Credit risk measurement: Developments over the last 20 years", *Journal of Banking & Finance*, pp.1721-1742, 1998.

[2] Bahr, R. and Hamori, S., *Hidden Markov Models. Applications to Financial Economics*, Kluwer Academic Publishers, 2004.

[3] Cappé, O., Moulines, E. and Rydén, T., *Inference in Hidden Markov Models*, Springer, 2005.

[4] Duffie, D. and Singleton, K.J., *Credit Risk: Pricing, Measurement and Management*, Princeton University Press: Princeton and Oxford, 2003.

[5] Giampieri, G., Davis, M., and Crowder, M., "Analysis of Default Data Using Hidden Markov Models" *Quantitative Finance*, Vol. 5, No. 1, pp. 27-34, February 2005.

[6] Mamon, R.S. and Elliot, R.J., *Hidden Markov Models in Finance*, Springer US, 2007.

[7] McCallum, A., *Hidden Markov Models Baum Welch Algorithm Lecture Notes*, March 2004.

[8] McCulloch, C.E. and Searle, S.R., *Generalized, Linear, and Mixed Models*, Wiley-Interscience Publication, 2001.

[9] McNeil, A.J., Frey, R., and Embrechts, P., *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance, 2005.

[10] Tu, S. *Derivation of Baum-Welch Algorithm for Hidden Markov Models*, 2006.

[11] Van Handel, R., *Hidden Markov Models Lecture Notes*, July 2008.

[12] Venables, W.N. and Smith, D.M., *An Introduction to R*, R Core Team, 2015.

[13] Zucchini, W. and MacDonald, I.L., *Hidden Markov Models for Time Series, An Introduction Using R*, Chapman & Hall, 2009.

Appendix A Standard & Poor's Data

	rating	**firms**	**defaults**		**rating**	**firms**	**defaults**
31-12-1981	A	484	0	31-12-1981	A	602	0
31-12-1981	BBB	267	0	31-12-1981	BBB	376	2
31-12-1981	BB	217	0	31-12-1981	BB	241	6
31-12-1981	B	81	0	31-12-1981	B	287	39
31-12-1981	C	11	0	31-12-1981	C	61	19
31-12-1982	A	478	2	31-12-1982	A	678	0
31-12-1982	BBB	292	1	31-12-1982	BBB	399	0
31-12-1982	BB	167	7	31-12-1982	BB	243	0
31-12-1982	B	162	5	31-12-1982	B	225	16
31-12-1982	C	14	3	31-12-1982	C	51	12
31-12-1983	A	455	0	31-12-1983	A	762	0
31-12-1983	BBB	305	1	31-12-1983	BBB	458	0
31-12-1983	BB	171	2	31-12-1983	BB	286	1
31-12-1983	B	157	7	31-12-1983	B	236	5
31-12-1983	C	16	0	31-12-1983	C	50	6
31-12-1984	A	457	0	31-12-1984	A	845	1
31-12-1984	BBB	295	2	31-12-1984	BBB	528	0
31-12-1984	BB	172	2	31-12-1984	BB	374	1
31-12-1984	B	181	6	31-12-1984	B	346	9
31-12-1984	C	19	3	31-12-1984	C	26	4
31-12-1985	A	514	0	31-12-1985	A	1024	0
31-12-1985	BBB	282	0	31-12-1985	BBB	639	2
31-12-1985	BB	204	3	31-12-1985	BB	428	3
31-12-1985	B	204	11	31-12-1985	B	405	17
31-12-1985	C	19	2	31-12-1985	C	29	8
31-12-1986	A	551	1	31-12-1986	A	1087	0
31-12-1986	BBB	295	1	31-12-1986	BBB	718	0
31-12-1986	BB	232	3	31-12-1986	BB	471	3
31-12-1986	B	291	25	31-12-1986	B	438	11
31-12-1986	C	17	3	31-12-1986	C	28	1
31-12-1987	A	505	0	31-12-1987	A	1144	0
31-12-1987	BBB	317	0	31-12-1987	BBB	834	1
31-12-1987	BB	268	1	31-12-1987	BB	551	1
31-12-1987	B	358	12	31-12-1987	B	476	15
31-12-1987	C	63	6	31-12-1987	C	27	3
31-12-1988	A	520	0	31-12-1988	A	1183	0
31-12-1988	BBB	333	0	31-12-1988	BBB	997	3
31-12-1988	BB	291	3	31-12-1988	BB	662	5
31-12-1988	B	418	16	31-12-1988	B	700	32
31-12-1988	C	59	13	31-12-1988	C	32	11
31-12-1989	A	561	0	31-12-1989	A	1208	1
31-12-1989	BBB	334	2	31-12-1989	BBB	1085	2
31-12-1989	BB	282	2	31-12-1989	BB	793	8
31-12-1989	B	416	14	31-12-1989	B	899	63
31-12-1989	C	55	16	31-12-1989	C	73	22
31-12-1990	A	584	0	31-12-1990	A	1215	1
31-12-1990	BBB	347	2	31-12-1990	BBB	1157	4
31-12-1990	BB	286	10	31-12-1990	BB	887	10
31-12-1990	B	365	31	31-12-1990	B	961	69
31-12-1990	C	48	15	31-12-1990	C	86	25

Appendix B R Code

In this Appendix the R code used to perform the analyses presented in this dissertation,developed supported on the examples found in Zucchini and MacDonald, 2009 [13], is introduced.

It may be important to refer that it is recognized that the code presented in this appendix is susceptible to improvements, having been produced with the aim of maximum clarity, with possible disregard for efficiency and presentation.

B.1 Baum-Welch versus Maximum Likelihood Estimation

```
# Installation of the library
library(HiddenMarkov)
# Definition of transition matrix
riskStates <- c("1", "2")
m <- 2
gamma <- matrix(data = c(0.8, 0.2, 0.25, 0.75),
byrow = TRUE, nrow = m,dimnames = list(riskStates, riskStates))
# Initial distribution of the Markov Chain
dist_ini <- c(0,1,0)
Lambda<-c(1,23)
lambda0<-c(0.9,18)
gamma0<-matrix(data = c(0.75, 0.25, 0.9, 0.1), byrow = TRUE, nrow = m,
    dimnames = list(riskStates, riskStates))
```

1. As a first step, it is necessary to install the library `HiddenMarkov` for R, containing a package of Hidden Markov Models simulation functions that will be used in the estimation of the model. It is then necessary to define the number of risk states to consider, `m`, the $m \times m$ transition probability matrix `gamma`, the initial distribution of the Markov Chain `dist_ini`, which, for simplification, is assumed to be deterministic, i.e. the state variable is assumed to be in state 2. The next step is to define a vector `Lambda` of expected values and to provide the initial values `lambda0` and `gamma0` for the λ and Γ parameters that will be used as an input for the estimation algorithm

```
# Definition of the object "Discrete Time Hidden Markov Model"
x<-dthmm(NULL,gamma,dist_ini,"pois",list(lambda=Lambda),discrete = TRUE)
# simulation of HMM using the library HiddenMarkov
nsample = 500
s <- simulate(x,nsim = nsample)
sample<-s[[1]]
```

2. Definition of the object `dthmm` using the parameters defined above, the `NULL` in the first entry corresponds to an array of observations that could consist on real data or simulated data and simulation of `nsample` (500) observations of a Hidden Markov Model using the library `HiddenMarkov`.

```
# Redifinition of the object dthmm
x<-dthmm(sample,gamma0,dist_ini,"pois",list(lambda=lambda0),discrete =
    TRUE)
# Estimation of the parameters of the model using BAUMWELCH
y<-BaumWelch(x)
```

3. Redefinition of the object `dthmm` to consider the array of simple observations of the Hidden Markov Model simulated above and the initial guesses `lambda0` and `gamma0`, proceeding with the estimation of the parameters using the Baum-Welch method.

The next step will be to proceed with the Maximum Likelihood Estimation method:

```
#Converting natural parameters into working parameters
pois.HMM.pn2pw <- function(m,lambda ,gamma) {
  tlambda <- log(lambda)
  tgamma <- NULL
  if(m>1) {
    foo <- log(gamma/diag(gamma))
    tgamma <- as.vector(foo[! diag(m)])
  }
  parvect <- c(tlambda ,tgamma)
  parvect
}
parvect <- pois.HMM.pn2pw(m,lambda ,gamma)
```

4. Before applying the Maximum Likelihood algorithm it is necessary to guarantee that the outputs will be subject to the necessary constraints. Particularly, it is mandatory that the rows of the maximum likelihood estimator $\hat{\Gamma}$, being a transition probability matrix, sum up to 1, while the values found in the expected value vector λ should be non-negative, i.e. $\lambda_i \geqslant 0$ for $i = 1, ..., m$. This function transforms the inputs of the Maximum Likelihood Estimation in order to guarantee that the results will comply with these restrictions.

```
#Converting working parameters to natural parameters
pois.HMM.pw2pn <- function(m,parvect) {
  epar <- exp(parvect)
  lambda <- epar[1:m]
  gamma <- diag(m)
  if(m>1) {
```

```
    gamma [! gamma] <- epar[(m+1):(m*m)]
    gamma <- gamma/apply(gamma ,1,sum)
  }
  delta <- solve(t(diag(m)-gamma +1),rep(1,m))
  list(lambda=lambda ,gamma=gamma ,delta=delta)
}
```

5. This function works as opposed to the previous, transforming the working parameters into natural in order to grant a clear interpretation of the outputs of the algorithm, establishing the formulation of the estimators.

```
#Log-likelihood of a stationary Poisson-HMM
pois.HMM.mllk <- function(parvect, X, m) {
  if(m==1) return(-sum(dpois(X, exp(parvect), log=TRUE)))
  n <- length(X)
  pn <- pois.HMM.pw2pn(m, parvect)
  allprobs <- outer(X, pn$lambda, dpois)
  allprobs <- ifelse (!is.na(allprobs), allprobs, 1)
  lscale <- 0
  foo <- pn$delta
  for (i in 1:n) {
    foo <- foo %*% pn $gamma *allprobs[i,]
    sumfoo <- sum(foo)
    lscale <- lscale+log(sumfoo)
    foo <- foo/sumfoo
  }
  mllk <- -lscale
  mllk
}
```

6. Computation of minus the log-likelihood of the m-state Hidden Markov Model for a given (previously computed) vector `parvect` of working parameters and for vector `X` of simulated Hidden Markov Model observations.

```
#Maximum Likelihood Estimation of a stationary Poisson-HMM
pois.HMM.mle <- function(X, m, lambda0, gamma0) {
  parvect0 <- pois.HMM.pn2pw(m, lambda0, gamma0)
  mod <- nlm(pois.HMM.mllk,parvect0, X, m)
  pn <- pois.HMM.pw2pn(m, mod$estimate)
  mllk <- mod$minimum
```

```
  np <- length(parvect0)
  AIC <- 2*( mllk+np)
  n <- sum(!is.na(X))
  BIC <- 2* mllk+np*log(n)
  list(lambda=pn$lambda, gamma=pn$gamma, delta=pn$delta ,
  code=pn$code ,mllk=mllk ,AIC=AIC ,BIC=BIC)
}
# Estimation of the parameters of the model using MLE
w<-pois.HMM.mle(sample, m, lambda0, gamma0)
```

7. Application of the MLE algorithm through the minimization of the log-likelihood, considering the initial guesses `lambda0` and `gamma0` and storage of the results of the application of this function to the simulated Hidden Markov Model observations in object `w`.

```
# original lambda and gamma
print(Lambda)
print(gamma)
# estimated lambda and gamma using BAUMWELCH
print(y[5])
print(y[2])
# estimated lambda and gamma using MLE
print(w[1])
print(w[2])
```

8. Finally, a comparison between the two methods can be made by printing the results obtained for the parameters of the Hidden Markov Model using each of the methods.

B.2 Maximum Likelihood of a 2-state Hidden Markov Model

```
library(HiddenMarkov)
library(QRM)
# S&P data manipulation
defaults <- as.numeric(as.vector(spdata[,3]))
firms <- as.numeric(as.vector(spdata[,2]))
defaultsA <- round(1000*defaults[seq(1,length(defaults),5)]/firms[seq(1,
    length(defaults),5)])
defaultsBBB <- round(1000*defaults[seq(2,length(defaults),5)]/firms[seq
    (2,length(defaults),5)])
defaultsBB <- round(1000*defaults[seq(3,length(defaults),5)]/firms[seq(3,
    length(defaults),5)])
defaultsB <- round(1000*defaults[seq(4,length(defaults),5)]/firms[seq(4,
    length(defaults),5)])
```

```
10 defaultsC <- round(1000*defaults[seq(5,length(defaults),5)]/firms[seq(5,
     length(defaults),5)])
11 total_defaults<-defaultsA + defaultsBBB + defaultsBB + defaultsB +
     defaultsC
12 date <- rownames(spdata)
13 date <- date[seq(1,length(date),5)]
14 
15 def <- total_defaults
16 type_def <- "TOTAL"
```

1. First of all, the data to be used needs to be prepared with the installation of the `QRM` library in order to access the S&P data and performing calculations in order to have a more comparable data base. In parallel, it is also necessary to install the `HiddenMarkov` library, that will allow the estimation and manipulation of the model without major difficulties. In lines 15 and 16 of this code the group of data to be addressed can be chosen by choosing one of the names defined in lines 6 to 11. This detail will be relevant in the last phase of comparison between real data and simulated data.

```
# Definition of transition matrix
riskStates <- c("low", "high")
m <- 2
# Initial distribution of the Markov Chain
dist_ini <- c(1,0)
# Initial guess for the lambda and gamma parameters
epsilon <- 0.5
lambda0<-c(mean(def)*epsilon,mean(def)*(1+epsilon))
gamma0<-matrix(data = c(0.8, 0.2, 0.2,0.8), byrow = TRUE, nrow = m,
    dimnames = list(riskStates, riskStates))
```

2. Next, it is necessary to define the number and name of the risk states being considered, the initial risk state and the initial guesses `lambda0` and `gamma0` for the parameters of the model. In this case, each λ was defined as being given by the mean of the sample multiplied by a “deregulating” factor `epsilon`.

Next, steps 2 and 3 of the previous section need to be applied in order to provide a series of discrete time Hidden Markov Models, taking care to define, in line 3 of step 2, `sample` to be equal to our defined `def` and to ignore lines 3 and 4 of step 3.

After this, the Maximum Likelihood Estimation method, already explained in full detail in the previous section (steps 4 to 7) is applied.

```
# Initial lambda and gamma
print(lambda0)
```

```
print(gamma0)
# estimated lambda and gamma using MLE
print(w[1])
print(w[2])
# ReDefinition of the object "Discrete Time Hidden Markov Model" (dthmm)
    using estimated parameters.
gamma_est <- matrix(data=unlist(w[2]),byrow = TRUE, nrow = m,dimnames =
    list(riskStates, riskStates))
lambda_est <- as.vector(unlist(w[1]))
x<-dthmm(NULL,gamma_est,dist_ini,"pois",list(lambda=lambda_est),discrete
    = TRUE)
# simulation of estimated HMM using the library HiddenMarkov.
plot(def, type="o", col="black", axes=FALSE, ann=FALSE)
n<-1 #number of simulations
axis(1, at=1:20, lab=date)
max_y <- max(def)
axis(2, las=1, at=round(max_y/5)*0:max_y)
plot_colors <- rainbow(n, s = 1, v = 1, start = 0, end = max(1, n - 1)/n,
     alpha = 1)
nsample <- length(def) #number of years
for (i in 1:n) {
  s <- simulate(x,nsim = nsample)
  sample<-s[[1]]
  print(sample)
  lines(sample, type="o", pch=22, lty=2, col=plot_colors[i])
}
```

9. Finally, the estimated values for λ and Γ are presented and a new Discrete Time Hidden Markov Model is estimated, this time using the obtained Maximum Likelihood Estimators as the parameters of the model, finishing with the simultaneous plot of the real series and the estimated series of defaults, allowing for the formation of conclusions concerning the results.

B.3 Maximum Likelihood of a 3-state Hidden Markov Model

The single existing difference in computation between the Maximum Likelihood Estimation of a 2-state and a 3-state Hidden Markov Model can be found in the definition of the number of states, initial distribution and initial guesses for the model parameters. The difference, therefore, can be observed in step 2 of the previous computation procedure, that can be altered to:

```
# Definition of transition matrix
riskStates <- c("low", "medium", "high")
m <- 3
# Initial distribution of the Markov Chain
dist_ini <- c(1,0,0)
# Initial guess for the lambda and gamma parameters. This will be the input for the estimation algorithm.
epsilon <- 0.5
lambda0<-c(mean(def)*epsilon,mean(def),mean(def)*(1+epsilon))
gamma0<-matrix(data = c(1/3, 1/3, 1/3,1/3,1/3,1/3,1/3,1/3,1/3), byrow = TRUE, nrow = m,dimnames = list(riskStates, riskStates))
```

Appendix C Results

```
> # original lambda and gamma
> print(Lambda)
[1]  1 23
> print(gamma)
     1    2
1 0.80 0.20
2 0.25 0.75
```

```
> # estimated lambda and gamma using BAUMWELCH
> print(y[5])
$pm
$pm$lambda
[1]  0.9456851 22.7005122


> print(y[2])
$Pi
          [,1]      [,2]
[1,] 0.8429487 0.1570513
[2,] 0.2673792 0.7326208
```

```
> # estimated lambda and gamma using MLE
> print(w[1])
$lambda
[1]  0.9456847 22.7004764

> print(w[2])
$gamma
          [,1]      [,2]
[1,] 0.8412628 0.1587372
[2,] 0.2649210 0.7350790
```

Figure 10: Results of the application of the Baum-Welch and Maximum Likelihood Estimation methods

Results for the 2 risk states case

$$\hat{\lambda}(0)_{Total} = (125.25 \quad 375.75)$$
$$\hat{\lambda}_{Total} = (124.22 \quad 353.82)$$
$$\hat{\boldsymbol{\Gamma}}_{Total} = \begin{pmatrix} 0.4734 & 0.5266 \\ 0.4313 & 0.5687 \end{pmatrix}$$

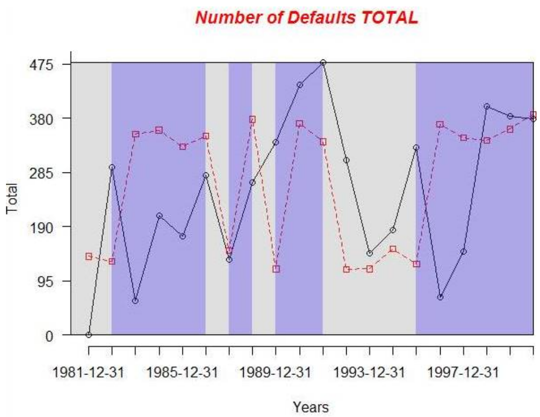


Figure 11: Real data series (black) and estimated series (red) for the total S&P data series

$$\hat{\lambda}(0)_A = (0.23 \quad 0.68)$$
$$\hat{\lambda}_A = (0.14 \quad 2.15)$$
$$\hat{\boldsymbol{\Gamma}}_A = \begin{pmatrix} 0.8174 & 0.1826 \\ 1.0000 & 4.32 \times 10^{-9} \end{pmatrix}$$

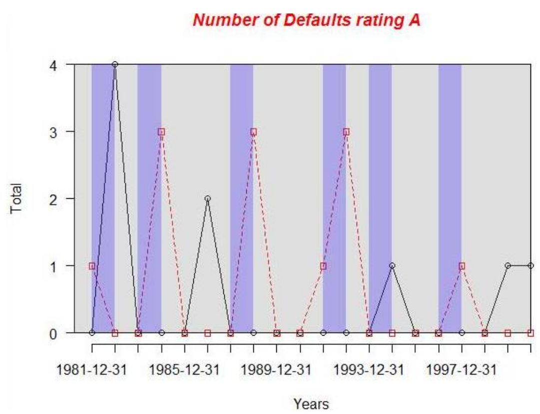


Figure 12: Real data series (black) and estimated series (red) for rating A

$$\hat{\lambda}(0)_{BBB} = (1.13 \quad 3.38)$$
$$\hat{\lambda}_{BBB} = (2.67 \times 10^{-9} \quad 3.65)$$
$$\hat{\mathbf{\Gamma}}_{BBB} = \begin{pmatrix} 0.3898 & 0.6101 \\ 0.3806 & 0.6194 \end{pmatrix}$$

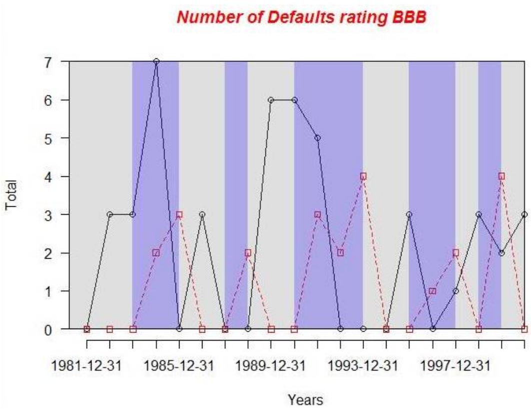


Figure 13: Real data series (black) and estimated series (red) for rating BBB

$$\hat{\lambda}(0)_{BB} = (5.63 \quad 16.88)$$
$$\hat{\lambda}_{BB} = (7.23 \quad 33.98)$$
$$\hat{\mathbf{\Gamma}}_{BB} = \begin{pmatrix} 0.8829 & 0.1171 \\ 0.6825 & 0.3175 \end{pmatrix}$$

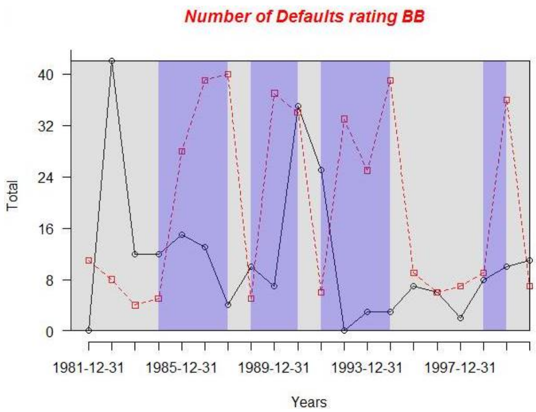


Figure 14: Real data series (black) and estimated series (red) for rating BB

$$\hat{\lambda}(0)_{B} = (24.53 \quad 73.58)$$
$$\hat{\lambda}_{B} = (32.52 \quad 85.28)$$
$$\hat{\mathbf{\Gamma}}_{B} = \begin{pmatrix} 0.8004 & 0.1995 \\ 0.4157 & 0.58423 \end{pmatrix}$$

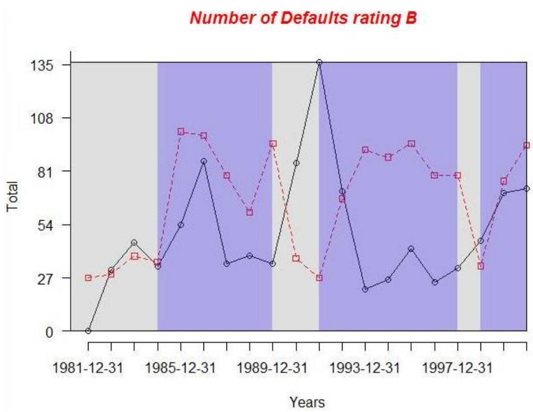


Figure 15: Real data series (black) and estimated series (red) for rating B

$$\hat{\lambda}(0)_C = (93.75 \quad 281.25)$$
$$\hat{\lambda}_C = (66.71 \quad 252.54)$$
$$\hat{\mathbf{\Gamma}}_C = \begin{pmatrix} 0.1576 & 0.8424 \\ 0.4468 & 0.5532 \end{pmatrix}$$

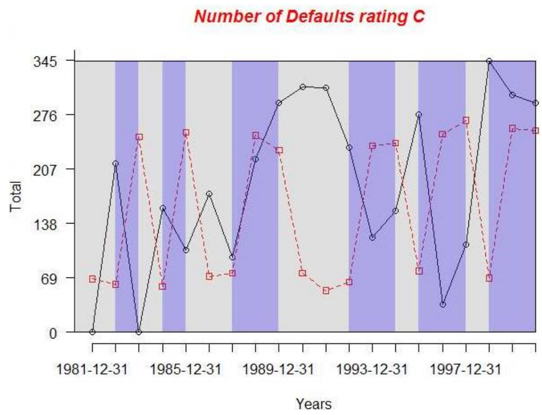


Figure 16: Real data series (black) and estimated series (red) for rating C

Results for the 3 risk states case

$$\hat{\lambda}(0)_{Total} = (125.25 \quad 250.50 \quad 375.75)$$
$$\hat{\lambda}_{Total} = (42.33 \quad 165.17 \quad 353.82)$$
$$\hat{\mathbf{\Gamma}}_{Total} = \begin{pmatrix} 2.02 \times 10^{-10} & 0.6504 & 0.3496 \\ 5.43 \times 10^{-8} & 0.3182 & 0.6818 \\ 0.2569 & 0.1754 & 0.5677 \end{pmatrix}$$

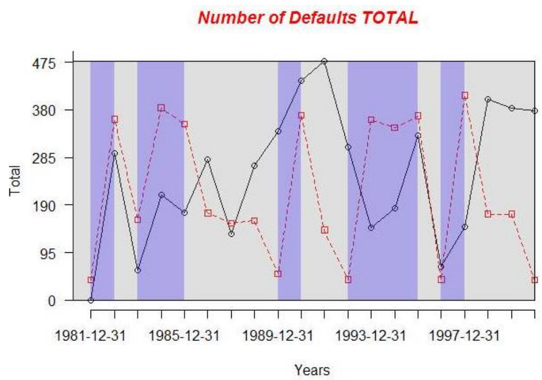


Figure 17: Real data series (black) and estimated series (red) for the total S&P data series

$$\hat{\lambda}(0)_A = (0.23 \quad 0.45 \quad 0.68)$$
$$\hat{\lambda}_A = (0.21 \quad 7.54 \times 10^{-10} \quad 1.78)$$
$$\hat{\mathbf{\Gamma}}_A = \begin{pmatrix} 0.2829 & 0.7171 & 2.03 \times 10^{-23} \\ 2.79 \times 10^{-17} & 0.5851 & 0.4149 \\ 1.0000 & 0.0000 & 0.0000 \end{pmatrix}$$

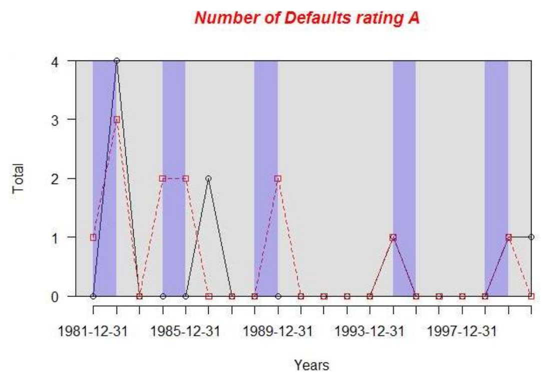


Figure 18: Real data series (black) and estimated series (red) for rating A

$$\hat{\lambda}(0)_{BBB} = (1.13 \quad 2.25 \quad 3.38)$$
$$\hat{\lambda}_{BBB} = (5.48 \times 10^{-25} \quad 3.07 \quad 5.38)$$
$$\hat{\mathbf{\Gamma}}_{BBB} = \begin{pmatrix} 0.3772 & 0.6227 & 2.15 \times 10^{-7} \\ 0.2745 & 0.4955 & 0.2300 \\ 0.6723 & 4.23 \times 10^{-9} & 0.3277 \end{pmatrix}$$

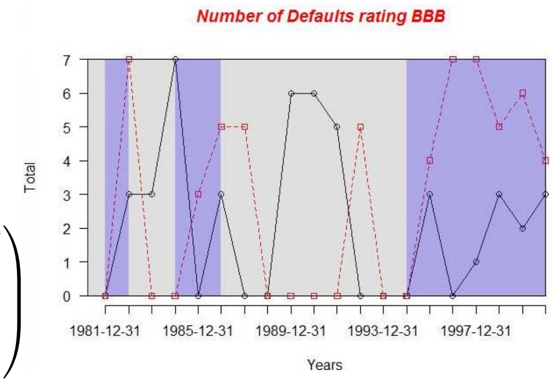


Figure 19: Real data series (black) and estimated series (red) for rating BBB

$$\hat{\lambda}(0)_{BB} = (5.63 \quad 11.25 \quad 16.88)$$
$$\hat{\lambda}_{BB} = (1.67 \quad 9.50 \quad 34.00)$$
$$\hat{\mathbf{\Gamma}}_{BB} = \begin{pmatrix} 0.4522 & 0.3544 & 0.1934 \\ 0.1290 & 0.7753 & 0.0956 \\ 0.3956 & 0.2875 & 0.3169 \end{pmatrix}$$

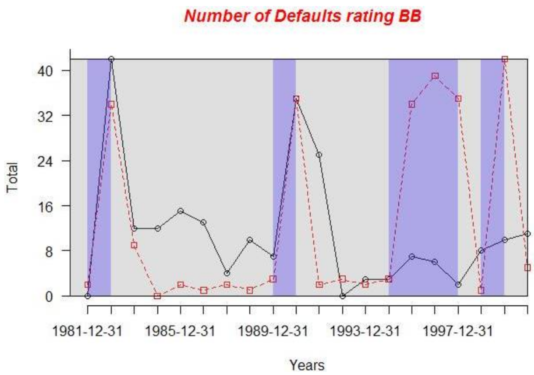


Figure 20: Real data series (black) and estimated series (red) for rating BB

$$\hat{\lambda}(0)_{B} = (24.53 \quad 49.05 \quad 73.58)$$
$$\hat{\lambda}_{B} = (5.30 \times 10^{-30} \quad 35.54 \quad 86.43)$$
$$\hat{\mathbf{\Gamma}}_{B} = \begin{pmatrix} 2.39 \times 10^{-6} & 0.9999 & 1.51 \times 10^{-12} \\ 1.79 \times 10^{-8} & 0.7349 & 0.2651 \\ 0.1406 & 0.2961 & 0.5634 \end{pmatrix}$$

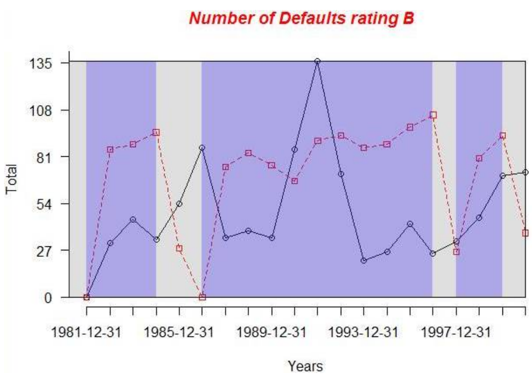


Figure 21: Real data series (black) and estimated series (red) for rating B

$$\hat{\lambda}(0)_C = (93.75 \quad 187.50 \quad 281.25)$$

$$\hat{\lambda}_C = (12.00 \quad 131.29 \quad 290.50)$$

$$\hat{\mathbf{\Gamma}}_C = \begin{pmatrix} 2.52 \times 10^{-8} & 0.6417 & 0.3583 \\ 1.45 \times 10^{-14} & 0.5465 & 0.4534 \\ 0.2772 & 0.09338 & 0.6294 \end{pmatrix}$$

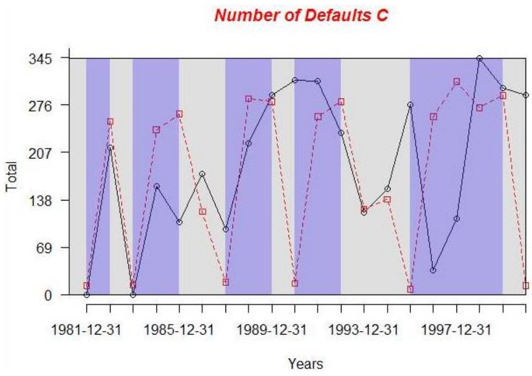


Figure 22: Real data series (black) and estimated series (red) for rating C