ENSIMAG
KTH Royal institute

Master Thesis report

Picard Tom

March 9, 2019

Grenoble INP – ENSIMAG
Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées

Master Thesis Report

Realized at Nexialog Consulting

Inference and filtration of a hidden factor in credit risk

Picard Tom
3rd year – Financial engineering

03/09/2018 – 01/03/2019

Nexialog Consulting
110 avenue de la République 75011 Paris
KTH
Brinellvägen 8, 114 28 Stockholm, Suède
Ensimag
681 Rue de la Passerelle, 38400 Saint-Martin-d'Hères

Company person in charge
Ali Behbahani
Internship supervisor
Areski Cousin
School supervisor
Guiol Hervé

Abstract

This Document is a Master Thesis report as part of the cycle of the ENSIMAG and KTH engineer. The internship linked to this report took place at Nexialog Consulting, Paris.
This document deals with the inference and filtration by a counting process of a hidden parameter. The theory developed here is applied to Credit risk and specially to Migration Matrices of Rating.

Acknowledgements

The drafting of such a document requires the participation and collaboration of many people. As a preamble to this report, I would like to acknowledge here the contribution made by those without whom this could not have succeeded.
I would first deeply thank Dr. Areski Cousin to have supervised me during this internship and for his precious help for the realization of this project.
I would also thank Mister Christophe Reigner for his support and his goodwill.
I would like thank the entire Nexialog team for their help and welcome, especially Mister Ali Behbahani and Jérémy Robert for giving me the chance to live by their side in the company.

Contents

1 Introduction

Analyzing the effect of business cycle on rating transition probabilities has been a subject of great interest these last fteen years, particularly due to the increasing pressure coming from regulators for stress testing.
This report presents a new method of estimation, the filtration of an hidden parameter by a counting process applied in credit Risk to ratings migrations.
The study is the continuity of the work of Areski Cousin and Ragnar Norberg on factor-based migration models.

An underlying goal of this thesis is to gain a deeper insight as to what the job of an engineer or researcher consists of. The candidate is expected to work within a company to become more accustomed to the world of business and its organization, structure. This internship aims to challenge technical and human skills alike in a real world scenario.

Besides the personal development, to concretely accomplished a useful project for the host company represents a real motivation of any master thesis. To do it at Nexialog consulting, located in Paris represented the perfect opportunity. This consulting firm specialised in Financial Risk and insurance provides services and advises to the biggest banks and insurance companies in Paris. Nexialog is a specialist in banking and financial risk. It has also developing a strong development and research pole, within which this internship has been conducted. The principal interest consisted in finding an applied project in finance, which would develop my skills in Statistics and in learning more about my future carrier plan.

The first part of this document aims to describe the working conditions. This part will be exempt from technical details.

The second part will be focused on the technical elements of the project: the objectives of the project and the model that has been developed to reach them. This part will close on an objective analysis of the results.

The final part of the report will detail the satisfaction and regrets (that could be gathered after completing this internship).

2 Description of the environment

2.1 The company: Nexialog Consulting

Founded in 2006, Nexialog is an advice firm specialised in bank and insurance which intervenes within the business teams to support them in making strategic decisions and in implementing their projects.
It noteworthy that the firm has had a sharp and stable growth since its creation 12 years ago. The increasing from 11 employees in 2009 up to 68 in 2017 of its manpower, is a sign that attests to its presence and dynamism in the world of baking and insurance.

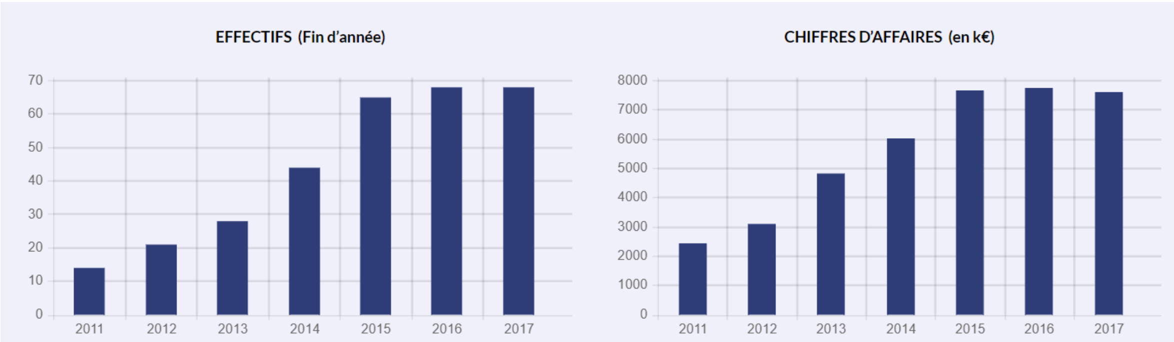
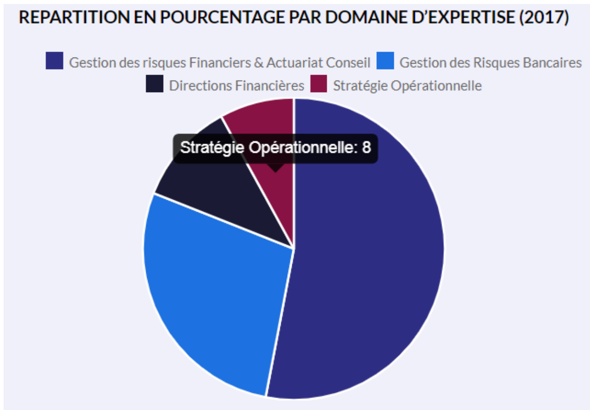


Figure 2.1 Nexialog's evolution

Its areas of activity are articulated around 6 complementary business units (BU):

-Insurance Risk Management
-Bank Risk Management
-Finance Risk Management
-Financial Direction
-Operational Strategy
-Research and development

(a) Nexialog's structure



(b) Business units

Risk Insurance Management

The Actuarial Consulting (the Risk Management and Actuarial Business Unit (RMA)), is in charge of the insurance activities. Regulatory reforms and financial communications requirements have helped in promoting the importance of good risk management in insurance companies.
The main topics covered are actuarial production, data science, asset management, as well as prospective and financial modelling and product actuarial services.
The actuarial consultant is usually required to guarantee the insurance company's coverage while being worried about the competitiveness of the organisation and the offer.

Bank Risk Management

The banking risk management unit provides banking institutions with support for the conduct of their projects by accompanying them both in their capital allocations and in their operational risk management system.
In fact, by anticipating changes of the economic and regulatory environment, this unit provides its clients with expertise in statistical analysis and portfolio management to cover them from the risks associated with their activities.
The areas of expertise can be summarized in risk modelling, prudential reporting and monitoring, regulatory projects and data management.

Financial Risk Management

Present in both banks and insurances, the Business Unit Risk Management and Finance (RMF) is involved in capital market, ALM, asset management and financing topics.
Regulatory reforms (IFRS9, active QRTs S2, EMIR, FRTB, PRIIPS...), modeling and calibration of risk models, complexifications of insurance products and the implementation of innovative reporting represent the daily challenges of the RMF unit's clients.

Financial Direction

This structure works in close collaboration with the financial departments of various banks and insurances on daily tasks or on the implementation of their major transformation projects. It participates in the regulatory compliance, financial management or operational assistance in the management of various accounts.

Operational Strategy

The Operational Strategy structure assists its clients in the management of their own projects.

Research And Development

In order to anticipate the future needs of clients and to propose innovative solutions, the research and development unit acquires knowledge from previous experiences and its own theoretical work. It then diffuses that knowledge throughout the company and to its collaborators and clients.

2.2 Working environment

The working conditions have a vital role in the productivity and well-being of employees.
The term working environment encapsulates the physical location but also the way of working, the organization, the mentalities, the corporate policies, and the human relationship within it...
It is therefore interesting to investigate on the impact of this environment on the productivity and the internship's management.

2.2.1 The physical environment

Nexialog Consulting's headquarters are located in the center of Paris. A centralized and well-connected access reduces the time of transport of the collaborators, an important detail in a big city (the delays of the public transports, the traffic jams involve a waste of time and may affect the morale of the collaborators ...).
There are two open spaces, one for the commercials and for the structure and an another for the consultants. Open space work saves space but also encourages exchange and cohesion among employees. Small private spaces are available to isolate yourself in order to make small meetings or phone calls.

2.2.2 Organization and policies

A crucial policy at Nexialog is to establish a fairly horizontal hierarchical organization. This kind of choice aims to improve the well-being team spirit among all employees, but also to help them to feel more involved in business projects.

Transparency is also be a priority to ensure a good understanding between the commercial unit and the consultants. Trust then plays the role of the corner stone in an environment where everyone must be free to work toward a common interest.

Many events, such as seminars abroad, are essential to create a corporate cohesion within a system where employees are not used to work together.

The communication inside the company is a crucial worry. Every Monday morning, a meeting between the collaborators present at the firm, the commercial structure and the associated founder leads to regular exchanges about new events and opportunities. These meetings, make people aware and concerned about the common project of the company.
Each week, employees have an individual and informal meeting with the HR department to talk about life in the firm.
The many efforts made to ensure the well-being of everyone, testify to the important consideration of the impact of the environment.

2.2.3 Social and environmental impact

Personal computer use of 65W for 35 hours per week over 26 weeks of internship is approximately 216000000J or 60KWh.

Nexialog has implemented various measures to be the most socially and ecologically responsible as possible. Efforts are made in particular on energy consumption (energy saving bulbs, window insulation...), on the use of electronic resources, on paper savings, and on the sorting of waste.

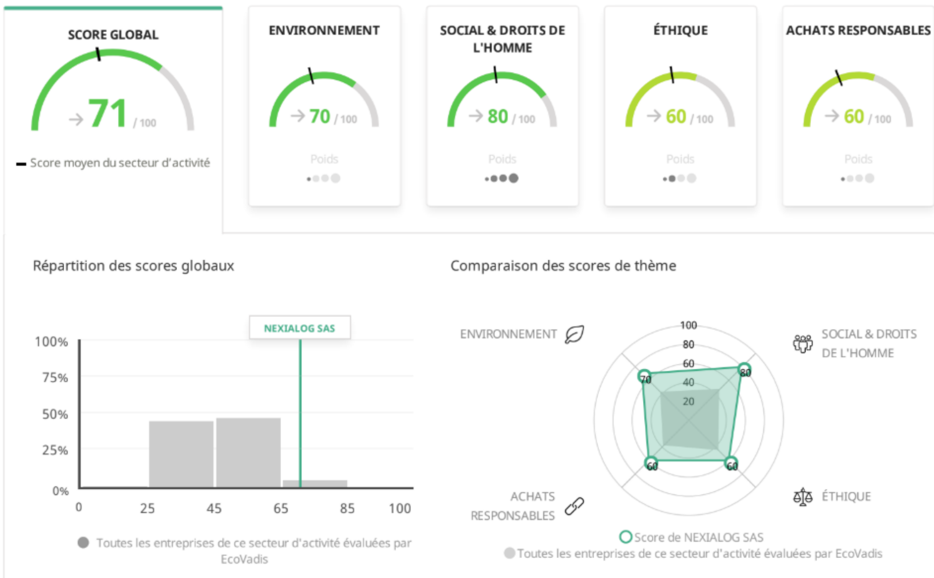


Figure 2.3 EcoVadis score sheet for RSE score from the 10th january 2019

2.3 The planning

The working environment must then influence a provisional schedule for the good management of the internship. The planned schedule was summarized in 4 parts:

-A one-month phase of internal documentation to be immersed in the topic and to improve the skills on the subject.
-A one-month phase of external documentation to gather knowledge on the state of the art.
-A two-month phase of personal research.
-A two-month phase devoted to the implementation and writing.

This schedule was punctuated by the regular visits of the supervisor Dr.Areski Cousin. Our exchanges reorganized the direction of the project.
I met the research and development manager monthly to ensure a regular follow-up during this internship.

3 The model developed

A credit rating evaluates the confidence in the capacity of the borrower to comply with the credit's terms. Predicting the evolution of rating migration matrices has been a crucial concern for all the financial institutions.
In credit risk, the rating of a borrower has an important impact on the terms of the credit. For example, the new capital restrictions imposed by Basel III, use the ratings to fix the amount of money required for the lender to be covered.

3.1 Filtering of counting process

The project of this internship is the continuity of the work lead by Areski Cousin for the thesis of Réda Khéliouen described in [4].
In this last work, different approaches have been compared to estimate factor-based migration models. These models admit probabilities of transitions that depend on dynamic factors. Two types of models were considered: the model "ordered Probit" (structural approach) as described in the example [6] and the "multi-state latent factor intensity model" (approach said of intensity) studied by [8].
We consider the statistical estimation of credit migration models where, the transition matrix is stochastic and may depend on a pool of common dynamic factors. The underlying factors aim at representing the evolution of the business cycle. The estimation procedure of such models differs given that the factors are considered to be observable or unobservable. In the first approach, one selects observable factors (macroeconomic variables for example) via economic analysis and then estimates the transition probabilities' sensitivities by doing a regression.

In the second case, the hidden factor may be inferred by maximizing the unconditional likelihood. The latter being very heavy to compute (see for example, [8]), we adapted an approach proposed by [6] which consists of representing the dynamics of the transition probabilities with the form of a linear Gaussian model. The non-observed factor is then filtered by the Kalman recursive algorithm. The model parameters are estimated by maximizing the unconditional likelihood.
One can not be satisfied by this approach since it is based on estimated observations (transition intensities) and, on the other hand, on the assumption of normality of the estimated intensities, a hypothesis that does not hold for samples of small size.

The method, presented here, consist to observe the hidden process through the observations of a counting process. For a bond portfolio, the dynamics of rating migrations can be modelled by a multidimensional process, each component representing the number of transitions from one rating category to another.
Assuming that the latter admits an intensity that is not observed, is it possible to estimate this intensity only using the observations of the point process ?

We consider a space with a filtration: $(\Omega, \mathcal{F}, \boldsymbol{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$. All the stochastic processes dealt are supposed to be adapted to $\boldsymbol{F}$ and integrable on $[0,T]$.
The problem is to estimate the state Θ_t of an unobserved process using only the information

$\mathcal{F}_t^N$ which results from the observation of a counting process N. By using the least squares criterion, this leads to determining

$$\hat{\Theta}_t = \mathbb{E}[\Theta_t|\mathcal{F}_t^N] \,.$$

With the same notations, all the processes O filtered by $\mathcal{F}_t^N$ are written :

$$\hat{O}_t = \mathbb{E}[O_t|\mathcal{F}_t^N]$$

Unfortunately, the explicit computation of $\hat{\Theta}$ (by the Bayes formula) is only possible in very simple cases, for example, when N is a mixture of Poisson processes. Moreover, numerical approaches can be envisaged but they are extremely heavy to implement as soon as non-trivial models are considered.

The filtering approach is different. It deals with $\hat{\Theta}$ as the solution of a forward EDS (differential stochastic equation) that allows, given $\hat{\Theta}_t$ to compute $\hat{\Theta}_{t+dt}$ per a simple recursive update.
The main result on point process filtering can be stated in the following way (see [2], [7], [10], [14]). Let Θ be a process of the form

$$d\Theta_t = a_t \,\mathrm{d}t + \mathrm{d}M_t \,,$$

where M is a martingale with no jumps in common with N. Then, the $\hat{\Theta}$ process checks the following equation:

$$\mathrm{d}\hat{\Theta}_t = \hat{a}_t \,\mathrm{d}t + \eta_t \,(\mathrm{d}N_t - \hat{\nu}_t \,\mathrm{d}t) \,, \tag{3.1}$$

with

$$\eta_t = \frac{(\Theta\hat{\ }\nu)_{t-}}{\hat{\nu}_{t-}} - \hat{\Theta}_{t-} \,,$$

and where ν is the $\boldsymbol{F}$-intensity of N. The dynamic (3.1) suggests a recursive algorithm to update the $\hat{\Theta}$ process.

A consequence of (3.1) is that the dynamics of $\hat{\Theta}$ implies $(\widehat{\Theta\nu})$ whose dynamic implies the term $(\widehat{\Theta\nu^2})$, and so on ... Therefore this filtering formula induces an infinite nesting problem.

In order to study the dynamics of defaults in a credit portfolio or the dynamics of rating migrations, this formula of filter must be extended to a multivariate case.

3.2 Extension of the filtration formula

In this section, one proposes an adaptation of the filtering formula to multivariate point processes and when the unobserved factor admits a finite number of states. Some proofs are not given in this summary document. They will be detailed in the article being written [3].

3.2.1 Extension of the filter to a multivariate case

In order to study the dynamics of defaults in a credit portfolio or the dynamics of rating migrations, the classical framework filter formula also studied in [2] and [10], must first be extended to a multivariate case.

Let $N^j = (N_t^j)_{t\in[0,T]}$, $j = 1,\ldots,p$, simple counting processes:

$N_t^j = \sum_{0<s\leq t} \Delta N_s^j < \infty$ and $\Delta N_s^j \in \{0,1\}$.
It is assumed that these processes admit a $\boldsymbol{F}$ - intensity $\nu^j = (\nu_t^j)_{t\in[0,T]}$, $j = 1,\ldots,p$, and that they do not have any common jumps, i.e., $\Delta N_t^j \Delta N_t^k = \delta_{jk}$ (1 if $j = k$ and 0 otherwise). In the following, we consider the multivariate counting process $N = (N^1,\ldots,N^p)$ and we write $\nu = (\nu^1,\ldots,\nu^p)$ its intensity from filtration $\boldsymbol{F}$.

Proposition *Let Θ, a square integrable process with the form*

$$\Theta_t = \int_0^t a_s \,\mathrm{d}s + M_t \,, \tag{3.2}$$

where a is $\boldsymbol{F}$-adapted and M is a $\boldsymbol{F}$-square integrable martingale with no jumps in common with N. Therefore the process $\hat{\Theta}$ is solution of the EDS

$$\mathrm{d}\hat{\Theta}_t = \hat{a}_t \,\mathrm{d}t + \sum_j \eta_t^j \,(\mathrm{d}N_t^j - \hat{\nu}_t^j \,\mathrm{d}t) \,, \tag{3.3}$$

with

$$\eta_t^j = \frac{(\Theta \hat{\ } \nu^j)_{t-}}{\hat{\nu}_{t-}^j} - \hat{\Theta}_{t-} \,, \tag{3.4}$$

with initial conditions

$$\hat{\Theta}_0 = \mathbb{E}[\Theta_0] \,. \tag{3.5}$$

3.2.2 Latent factor with finite number of states and application to credit risk

To solve the issue of infinite imbrication mentioned above, we consider that the intensity ν of the multivariate counting can only be in a state which belongs to a space with a finite number of states. Consequently we consider that the hidden factor Θ which governs it is a Markov chain of number of states finite. Let Θ a Markov chain of number of states finite $\mathbb{T} = \{1,\ldots,m\}$ having transition intensities constant k^{hi}, $i \neq h$, which checks $k^{hh} = -\sum_{i;i\neq h} k^{hi}$. Introduce the process I^h, defined by

$$I_t^h = 1_{[\Theta_t = h]},$$

The process Θ can be represented by

$$\Theta_t = \sum_{h\in T} h\, I_t^h \,. \tag{3.6}$$

The filtering theory with a counting process has been applied to credit risk for example in [5] and [10].
In order to remain realistic and fix the terminology, consider a bond market containing a finite

number of individual bonds. All bonds are affected by variable and random market conditions represented by a hidden Θ Markov chain. An obligation q of the sample is observed between the dates s^q and u^q, $0 \leq s^q \leq u^q \leq T$. At all times, the obligation can only be in a single state belonging to a finite set of states, $\mathbb{J} = \{1, \dots, n\}$.
This space represents different credit risk scores in descending order, n being the default state. For example, Standard and Poor's long-term investment ratings can be translated to AAA = 1, AA = 2, A = 3, BBB =4, ..., D (Default) = 10.
Let $Z_t^q \in \mathbb{J}$ the state of the obligation at t and $Z^q = (Z_t^q)_{t\in[s^q,u^q]}$ the process that describes its evolution. The indicator process and the point process associated to Z are

$$J_t^{q,j} = 1_{[Z_t^q = j]}, \qquad N_t^{q,jk} = \sharp\{s \in [0,t]; Z_{s-}^q = j,\, Z_s^q = k\}.$$

The process Z^q is governed by its intensity.

$$\mathbb{P}[Z_{t+dt}^q = k \,|\, \Theta_t = h, Z_t^q = j] = \ell^{h,jk}\, dt + o(dt).$$

The processes of each individual Z^q are conditionally independent, knowing $\mathcal{F}_T^\Theta$. However unconditionally, they are dependent through the effect of $\mathcal{F}_T^\Theta$. Indeed the change of state of an obligation will induce the change of state of other obligations. The censorship mechanism governing (s^q, u^q) is assumed to be non-informative and can therefore be considered deterministic and belonging to $\mathcal{F}_0^N$.

To infer the underlying process Θ, It is sufficient to observe the counting process N^{jk} defined by

$$N_t^{jk} = \sum_q N_t^{q,jk}\,,$$

governed by the $\boldsymbol{F}$ intensities

$$\nu_t^{jk} = Y_t^j \sum_h \ell^{h,jk}\, I_{t-}^h\,,$$

with

$$Y_t^j = \sum_{q;\, s^q \leq t < u^q} J_t^{q,j}\,.$$

The exposure process Y^j is left continuous. It increases by 1 when a N^{kj}, $k \neq j$ has jumped, or when a new bond is entered with the note j, and decreases by 1 during a jump of N^{jk}, $k \neq j$, or when a bond expires with the note j.

I_{t-}^l is the indicator function of our unobserved factor that is equal to 1 in state l.
With such assumptions, we obtain the following form of the filter formula applied to the indicator of our unobserved process:

$$d\hat{I}_t^h = \sum_i k^{ih} \hat{I}_{t-}^i dt + \sum_{j\neq k} \Big(\frac{l^{h,jk}}{\sum_r l^{r,jk} \hat{I}_{t-}^r} - 1\Big) \hat{I}_{t-}^j \Big(dN_t^{jk} - Y_t^j \sum_i l^{i,jk} \hat{I}_{t-}^i dt\Big) \tag{3.7}$$

In this formula, some parameters are unknown: initial distribution and transition intensities of Θ :$(k^{ij})_{i,j}$ and $(\Pi_i)_i$.

$$\forall i \in [1...m] : \Pi_i = \mathbb{P}(\Theta_1 = i)$$

$$\forall i,j \in [1...m] : \mathbb{P}(\Theta_{t+dt} = j \,|\, \Theta_t = i) = k^{ij}\, dt + o(dt)$$

The conditional transition intensities of the observable rating process $((l^{h,ij})_{h,i,j})$ are also unknown.
In order to make this formula usable, it is necessary to estimate these parameters. The major part of the work done during the internship was to seek to make a such estimation.

4 Similarities with the Kalman filter

One tries in this section to find similarities with the classical Kalman filter.

The application of the Kalman filter can be defined in two phases. A first equation, of prediction, provides the image of the hidden state knowing the observations of the previous times. One second, that of correction which makes possible to correct the error made during the prediction phase, calculates the image of the hidden state knowing the observations of the previous and current times (the observation at the present time "corrects" the estimation of the previous step). One can refer to the studies made on the classical Kalman filter applied to probability laws in [11] or in a note written during this internship done in annex.
We identify I_n and N_n, the hidden state and the observation at the n^{th} iteration. With these notations, the Kalman filter is as follows:

Equation of correction

$$E[I_t|F_t^N] = f(E[I_t|F_{t-}^N], N_t)$$

Equation of prediction

$$E[I_t|F_{t-}^N] = g(E[I_{t-}|F_{t-}^N])$$

It is possible to see in the filter formula, (3.7), a similar decomposition to the one in a classical Kalman filter. Using the filter formula we calculate:

$$\begin{aligned}
E[I_t|F_{t-}^N] &= E[E[I_t|F_t^N]|F_{t-}^N] \\
&= E[I_{t-}|F_{t-}^N] + \sum_i k^{ih}\hat{I}_{t-}^i dt + \sum_{j\neq k} \frac{l^{h,jk}\hat{I}_{t-}^h}{\sum_r l^{r,jk}\hat{I}_{t-}^r} E[dN_t^{jk}|F_{t-}^N] - \hat{I}_{t-}^h E[dN_t^{jk}|F_{t-}^N] \\
&- Y_t^j l^{h,jk}\hat{I}_{t-}^h dt + Y_t^j \sum_i l^{i,jk}\hat{I}_{t-}^i \hat{I}_{t-}^h dt.
\end{aligned}$$

Then we are looking to

$$\begin{aligned}
E[dN_t^{jk}|F_{t-}^N] &= E[E[dN_t^{jk}|F_{t-}]|F_{t-}^N] \\
&= E[\nu_t^{jk} dt|F_{t-}^N] \\
&= E[Y_t^j \sum_h l^{h,jk} I_{t-}^h dt|F_{t-}^N] \\
&= \sum_h l^{h,jk}\hat{I}_{t-}^h dt
\end{aligned}$$

Since Y_t^j is F^N predictable. Therefore we obtain:

$$
\begin{aligned}
E[I_t|F_{t-}^N] &= E[I_{t-}|F_{t-}^N] + \sum_i k^{ih}\hat{I}_{t-}^i dt + \sum_{j\neq k} \frac{l^{h,jk}\hat{I}_{t-}^h}{\sum_i l^{i,jk}\hat{I}_{t-}^i} Y_t^j \sum_l l^{l,jk}\hat{I}_{t-}^l dt \\
&- \hat{I}_{t-}^h Y_t^j \sum_l l^{l,jk}\hat{I}_{t-}^l dt - Y_t^j l^{h,jk}\hat{I}_{t-}^h dt + Y_t^j \sum_i l^{i,jk}\hat{I}_{t-}^i \hat{I}_{t-}^h dt \\
&= E[I_{t-}|F_{t-}^N] + \sum_i k^{ih}\hat{I}_{t-}^i dt.
\end{aligned}
$$

What it leads us to distinguish:

$$
E[I_t|F_{t-}^N] = \hat{I}_{t-}^h + \sum_i k^{ih}\hat{I}_{t-}^i dt
$$

$$
E[I_t|F_t^N] = \underbrace{E[I_t|F_{t-}^N]}_{prediction} + \sum_{j\neq k} \left(\frac{l^{h,jk}}{\sum_r l^{r,jk}\hat{I}_{t-}^r} - 1\right)\hat{I}_{t-}^j \underbrace{(dN_t^{jk} - Y_t^j \sum_i l^{i,jk}\hat{I}_{t-}^i dt)}_{new information}
$$

We can identify $Y_t^j \sum_i l^{i,jk}\hat{I}_{t-}^i dt$ as the information predicted from the previous step and $dN_t^{jk} - Y_t^j \sum_i l^{i,jk}\hat{I}_{t-}^i dt$ the new information observed and brought by a jump at t. It would therefore be possible to estimate the parameters using nonlinear and non-Gaussian extensions of the Kalman filter.

5 The solution proposed: an adaptation of an EM algorithm

In this section, one proposes an adaptation of the classical HMM BaumWelh algorithm in order to use the parameters estimations in the filtering formula (3.7).

One notes $z_{ta|tb} = (z_{ta}, ..., z_{tb})$ the sample of processes describing the evolution of the bonds ratings between the times a and b and we note π the initial distribution of θ.
We write $z \setminus z_k = (z_1...z_{k-1}, z_{k+1}...z_T)$, the vector without the k^{th} component.
So one notes $Z = (z^1, ..., z^d)$ the total sample.

In our case, we dispose of a sample of d trajectories of z which are not independent. In deed, it is the same trajectory of θ, the hidden factor, which governs the ones of each element of the sample.
The proposed method is a maximisation expectation (EM) algorithm for hidden Markov chains (HMM), adapted to our model. We can find work on the classical model in [13], [1], [15] and [12].
Indeed, there are some notable differences with the model that we want apply to the filter formula (3.7) page 16, and the classical one:
-The BaumWelsh theorem is an estimation in discrete time (the idea of duration is not considered).
-The rating status depends only on the state of the hidden factor. In our case, it also depends on the previous rating (one is talking about conditional probabilities transitions).
-The elements of the sample of observations (the different ratings of the different obligations) are not independent because they come from the same realisation of the factor Θ not observed.
The first step of the classical algorithm is to replace the missing data (the states of Θ) with estimators computed from the observations and the previous estimation of the parameters (to which a consistent initial value has been assigned).
The second is to maximize a conditional likelihood that leads us to obtain a better estimate of our parameters. These new estimates will then be used to repeat the first step in the next iteration.

5.1 Hypothesis

In hidden Markov chain models, an unobservable chain governs another observable.
We recall the fundamental assumption of a Markov chain: the transition from one state to another depends only on the current state (no memory):

$$P(z_t|z_1, ...z_{t-1}) = P(z_t|z_{t-1})$$

For the classical model HMM, one has a strong hypothesis :

$$P(z_t|z_1, .., z_{t-1}, \theta_1, ..., \theta_t) = P(z_t|\theta_t)$$

Formally, the future observation depends only on the state in which the unobserved chain is located.
This hypothesis is a bit too strong for our model. One proposes an extension of the model, by assuming:

$$P(z_t|z_1, ...z_{t-1}, \theta_1, ..., \theta_t) = P(z_t|\theta_{t-1}, z_{t-1})$$

For the sake of notation, we assume here rather

$$P(z_t|z_1, ..., z_{t-1}, \theta_1, ..., \theta_t) = P(z_t|\theta_t, z_{t-1})$$

During this internship, the proposed solution consists in theoretically adapting the algorithm to the model.
Indeed, the fact that the probability of the observations does not depend only on the state (as it is the case in classical HMM models) but also on the previous observation, leads us to reconsider the main algorithm while being inspired by it.
It was also decided to estimate the parameters of the discrete model in order to use them in our filtration formula, (3.7), which describes a continuous phenomenon. On a time step of subdivision sufficiently small, this approximation seems quite reasonable but does not give a completely trustful description of the model.

5.2 The computations of the estimators

One define the probability forward:

$$\alpha_t(j) = P(Z_{1|t} = Z_{1|t}, \theta_t = j)$$

and the probability backward:

$$\beta_t(j) = P(Z_{t+1|T} = Z_{t+1|T}|\theta_t = j, Z_t = Z_t)$$

We can notice that

$$\begin{aligned}
\alpha_t(j) = P(Z_{1|t} = Z_{1|t}, \theta_t = j) &= \sum_k P(Z_{1|t} = Z_{1|t}, \theta_t = j, \theta_{t-1} = k) \\
&= \sum_k \prod_d P(z_t^d = z_t^d|Z_{1|t-1} = Z_{1|t-1}, \theta_t = j, \theta_{t-1} = k)P(Z_{1|t-1} = Z_{1|t-1}, \theta_t = j, \theta_{t-1} = k) \\
&= \sum_k \prod_d P(z_t^d = z_t^d|Z_{t-1} = Z_{t-1}, \theta_t = j)P(\theta_t = j|\theta_{t-1} = k, Z_{1|t-1} = z_{1|t-1})P(\theta_{t-1} = k, Z_{1|t-1} = z_{1|t-1}) \\
&= \sum_k \prod_d (P(z_t^d = z_t^d|z_{t-1}^d = z_{t-1}^d, \theta_t = j))P(\theta_t = j|\theta_{t-1} = k)\alpha_{t-1}(k)
\end{aligned}$$

So we have with our parameter:

$$\alpha_t(j) = \sum_k K^{kj}\alpha_{t-1}(k) \prod_d L^{j, z_{t-1}^d z_t^d}$$

Similarly, we will recursively derive the backward probability:

$$\beta_t(j) = P(Z_{t+1|T} = Z_{t+1|T}|\theta_t = j, Z_t = Z_t)$$

$$= \sum_l P(Z_{t+1|T} = Z_{t+1|T}, \theta_{t+1} = l|\theta_t = j, Z_t = Z_t)$$
$$= \sum_l P(Z_{t+2|T} = Z_{t+2|T}|\theta_{t+1} = l, Z_{t+1} = Z_{t+1}, Z_t = k, \theta_t = j)P(Z_{t+1} = Z_{t+1}, \theta_{t+1} = l|\theta_t = j, Z_t = Z_t)$$
$$= \sum_l P(Z_{t+2|T} = Z_{t+2|T}|\theta_{t+1} = l, Z_{t+1} = Z_{t+1}) \prod_d (P(z^d_{t+1} = z^d_{t+1}|\theta_{t+1} = l, \theta_t = j, Z_t = Z_t)) * P(\theta_{t+1} = l|\theta_t = j)$$
$$= \sum_l P(Z_{t+2|T} = Z_{t+2|T}|\theta_{t+1} = l, Z_{t+1} = Z_{t+1}) \prod_d (P(z^d_{t+1} = z^d_{t+1}|\theta_{t+1} = l, z^d_t = z^d_t)) * P(\theta_{t+1} = l|\theta_t = j)$$

So

$$\beta_t(j) = \sum_l \beta_{t+1}(l) K^{jl} \prod_d L^{l, z^d_t z^d_{t+1}}$$

Both estimators will be used to replace our missing data in the maximization phase.

We write

$$u_t(j) = 1_{(\theta_t = j)} \tag{5.1}$$

$$v_t(k, j) = 1_{(\theta_t = j, \theta_{t-1} = k)} \tag{5.2}$$

$$\hat{u}_t(j) = P(\theta_t = j|Z_{1|T} = Z_{1|T}) \tag{5.3}$$

$$\hat{v}_t(k, j) = P(\theta_t = j, \theta_{t-1} = k|Z_{1|T} = Z_{1|T}) \tag{5.4}$$

We will use the respective estimators:

$$\hat{u}_t(j) = P(\theta_t = j|Z_{1|T} = Z_{1|T})$$

$$= \frac{P(Z_{t+1|T} = Z_{t+1|T}|\theta_t = j, Z_t = Z_t)\alpha_t(j)}{P(Z_{1|T} = Z_{1|T})}$$
$$= \frac{P(Z_{t+1|T} = Z_{t+1|T}|\theta_t = j, Z_t = Z_t)\alpha_t(j)}{L_T}$$
$$= \frac{\beta_t(j)\alpha_t(j)}{L_T}$$

With

$L_T = P(Z_{1|T} = Z_{1|T}) = \sum_j \alpha_T(j),$
the like-hood on the totality of the sample.

and:

$\hat{v}_t(k,j) = P(\theta_t = j, \theta_{t-1} = k | Z_{1|T} = Z_{1|T})$

$$
\begin{aligned}
&= \frac{P(\theta_t = j, Z_{t|T} = Z_{t|T} | Z_{1|t-1} = Z_{1|t-1}, \theta_{t-1} = k)\alpha_{t-1}(k)}{L_T} \\
&= \frac{P(Z_{t+1|T} = Z_{t+1|T} | Z_{1|t} = Z_{1|t}, \theta_{t-1} = k, \theta_t = j) P(\theta_t = j, Z_t = Z_t | Z_{1|t-1} = Z_{1|t-1})\alpha_{t-1}(k)}{L_T} \\
&= \frac{P(Z_{t+1|T} = Z_{t+1|T} | Z_t = Z_t, \theta_t = j) P(Z_t = Z_t | Z_{1|t-1} = Z_{1|t-1}, \theta_t = j, \theta_{t-1} = k) P(\theta_t = j | \theta_{t-1} = k, Z_{1|t} = Z_{1|t})\alpha_{t-1}(k)}{L_T} \\
&= \frac{P(Z_{t+1|T} = Z_{t+1|T} | Z_t = Z_t, \theta_t = j) \prod_d (P(z_t^d = z_t^d | \theta_t = j, z_{t-1}^d = z_{t-1}^d)) P(\theta_t = j | \theta_{t-1} = k)\alpha_{t-1}(k)}{L_T} \\
&= \frac{\beta_t(j) K^{kj} \alpha_{t-1}(k) \prod_d L^{j, z_{t-1}^d z_t^d}}{L_T}
\end{aligned}
$$

5.3 Estimation of the parameters

Technical notes

A first technical point (1) that will be used several times in the following. It consists to find:

$$
\begin{aligned}
\max_{\pi_j, j=1..m} \quad & \sum_j c_j log(x_j) \\
\text{subject to} \quad & - x_j \leq 0 \\
& \sum_j x_j - 1 = 0.
\end{aligned}
$$

Note that uniform distribution is a feasible solution. KKT can therefore be applied to this system. We recall the notations:

$$
\begin{aligned}
\min_x \quad & f(x) \\
\text{subject to} \quad & h_j(x) \leq 0, \\
& a_i^T x - b = 0.
\end{aligned}
$$

Therefore x is solution iif :

We have a feasible solution and

$$\nabla(f(x)) + A^T \lambda + \sum_j \mu_j \nabla(h_j(x)) = 0$$

With this equation we have

$$-\frac{c_j}{x_j} + - \sum_i \mu_i + \sum_i \lambda_i = 0$$

If the $(\lambda_i)_i$ are equal to 0 then

$$x_j = -\frac{cj}{\sum_i \mu_i}$$

By Summing the last equality and equalising to 1 results in:

$$\sum_i \mu_i = -\sum_j c_j$$

and finally :

$$x_j = \frac{cj}{\sum_i c_i}$$

Our second technical point noted (2) is based on the concavity of the log:

For two suites u_i and v_i:

$$\begin{aligned}
\ln(\frac{\sum_i v_i}{\sum_i u_i}) &= \ln(\sum_i \frac{u_i v_i}{\sum_k u_k u_i}) \\
&\geq \sum_i \frac{u_i}{\sum_k u_k} \ln(v_i) - \frac{u_i}{\sum_k u_k} \ln(u_i) \\
&= \frac{1}{\sum_k u_k} (\sum_i (u_i \ln(v_i) - u_i \ln(u_i)))
\end{aligned}$$

The maximization phase consists of finding better parameters than those of the previous iteration. We notice $M = (\pi^{(q)}, L^{(q)}, K^{(q)})$, the parameters of the current iteration (q).
We also notice $\tilde{M} = (\pi^{(q+1)}, L^{(q+1)}, K^{(q+1)})$, the new parameters to estimate.
With the notations of the technical point (2), we note:

$$\begin{aligned}
u_i &= P(Z, \theta = \theta_i | M) \\
v_i &= P(Z, \theta = \theta_i | \tilde{M})
\end{aligned}$$

where $\theta_i \in N^H$, H the finite number of states of θ.
Then by using (2) we have:

$$\begin{aligned}
\ln(\frac{\sum_{\theta_i} v_i}{\sum_{\theta_i} u_i}) &= \ln \frac{P(Z|\tilde{M})}{P(Z|M)} \\
&\geq \frac{1}{P(Z|M)} (\sum_{\theta_i} u_i \ln(v_i) - \sum_{\theta_i} u_i \ln(u_i)) \\
&= \frac{1}{P(Z|M)} (Q(M, \tilde{M}) - Q(M, M))
\end{aligned}$$

We therefore try to have a higher likelihood with the new parameters:

$$P(Z|\tilde{M}) \geq P(Z|M)$$

With the last equality it amounts to seeking to maximize

$$Q(M, \tilde{M}) = \sum_\theta P(\theta = \theta, Z|M) \ln(P(\theta = \theta, Z|\tilde{M}).$$

So we cut:

$$\begin{aligned}
\ln(P(Z, \theta|\tilde{M})) &= \ln(P(Z|\theta, \tilde{M})) + \ln(P(\theta|\tilde{M})) \\
&= The(\mathrm{z}^d)_d \quad \text{are independent knowing the unobserved factor} \\
&= log(P(\theta_1)) + \sum_t \ln(P(\theta_t|\theta_{t-1})) + \sum_d \ln(P(z^d_{1|T}|\theta)) \\
&= log(P(\theta_1)) + \sum_t \ln(P(\theta_t|\theta_{t-1})) + \sum_d \ln(P(z^d_T|z^d_{1|T-1}, \theta) P(z^d_{1|T-1}|\theta)) \\
&= log(P(\theta_1)) + \sum_t \ln(P(\theta_t|\theta_{t-1})) + \sum_d \sum_t \ln(P(z^d_t|z^d_{t-1}, \theta_t)
\end{aligned}$$

So

$$
\begin{aligned}
Q(M,\tilde{M}) &= \sum_{\theta} log(P(\theta_1))P(\theta,Z|M) \\
&+ \sum_{\theta}\sum_{t} \ln(P(\theta_t|\theta_{t-1}))P(\theta,Z|M) \\
&+ \sum_{\theta}\sum_{d}\sum_{t} \ln(P(z_t^d|z_{t-1}^d,\theta_t)P(\theta,Z|M) \\
\\
&= \sum_{\theta\backslash\theta_1}\sum_{\theta_1} log(P(\theta_1))P(\theta,Z|M) + \sum_{\theta\backslash\theta_t,\theta_{t-1}}\sum_{\theta_{t-1},\theta_t}\sum_{t} \ln(P(\theta_t|\theta_{t-1}))P(\theta,Z|M) \\
&+ \sum_{\theta}\sum_{d}\sum_{t} \ln(P(z_t^d|z_{t-1}^d,\theta_t)P(\theta,Z|M) \\
\\
&= \sum_{\theta\backslash\theta_1}\sum_{i} log(P(\theta_1=i))P(\theta_1=i,\theta\setminus\theta_1,Z|M) \\
&+ \sum_{\theta\backslash(\theta_t,\theta_{t-1})}\sum_{i,j}\sum_{t} \ln(P(\theta_t=i|\theta_{t-1}=j))P(\theta\setminus(\theta_t,\theta_{t-1}),\theta_t=i,\theta_{t-1}=j,Z|M) \\
&+ \sum_{\theta\backslash\theta_t}\sum_{i}\sum_{d}\sum_{t} \ln(P(z_t^d|z_{t-1}^d,\theta_t=i)P(\theta\setminus\theta_t,\theta_t=i,Z|M) \\
\\
&= \sum_{i} log(P(\theta_1=i)) \sum_{\theta\backslash\theta_1} P(\theta_1=i,\theta\setminus\theta_1,Z|M) \\
&+ \sum_{t}\sum_{i,j} \ln(P(\theta_t=i|\theta_{t-1}=j)) \sum_{\theta\backslash(\theta_t,\theta_{t-1})} P(\theta\setminus(\theta_t,\theta_{t-1}),\theta_t=i,\theta_{t-1}=j,Z|M) \\
&+ \sum_{d}\sum_{t}\sum_{i} \ln(P(z_t^d|z_{t-1}^d,\theta_t=i) \sum_{\theta\backslash\theta_t} P(\theta\setminus\theta_t,\theta_t=i,Z|M) \\
&= \sum_{i} log(P(\theta_1=i))P(\theta_1=i,Z|M) \\
&+ \sum_{t}\sum_{i,j} \ln(P(\theta_t=i|\theta_{t-1}=j))P(\theta_t=i,\theta_{t-1}=j,Z|M) \\
&+ \sum_{d}\sum_{t}\sum_{i} \ln(P(z_t^d|z_{t-1}^d,\theta_t=i)P(\theta_t=i,Z|M) \\
\\
&= \sum_{i} log(P(\theta_1=i,Z|M))\pi_i \\
&+ \sum_{t}\sum_{i,j} \ln(K^{ij})P(\theta_t=i,\theta_{t-1}=j,Z|M) \\
&+ \sum_{d}\sum_{t}\sum_{i}\sum_{k,r} \ln(l^{i,rk})P(\theta_t=i,Z|M)1_{(z_t^d=r,z_{t-1}^d=k)}
\end{aligned}
$$

One can then maximize by considering the three terms independently by applying the technical point (1).
We must respect:

$$
\begin{aligned}
\sum_{i}\pi_i &= 1 \\
\sum_{i} K^{ji} &= 1 \\
\sum_{r} L^{i,kr} &= 1
\end{aligned}
$$

One obtains then :

$$\pi_i = \frac{P(\theta_1 = i, Z|M)}{\sum_j P(\theta_1 = j, Z|M)} = P(\theta_1 = i|Z, M) = \hat{u}_1(i) \tag{5.5}$$

$$\begin{aligned} L^{i,kr} &= \frac{\sum_d \sum_t P(\theta_t = i, Z|M) 1_{(z_t^d = r, z_{t-1}^d = k)}}{\sum_d \sum_t \sum_r P(\theta_t = i, Z|M) 1_{(z_t^d = r, z_{t-1}^d = k)}} \\ &= \frac{\sum_d \sum_t P(\theta_t = i, Z|M) 1_{(z_t^d = r, z_{t-1}^d = k)}}{\sum_d \sum_t P(\theta_t = i, Z|M) 1_{(z_{t-1}^d = k)}} \\ &= \frac{\sum_d \sum_t P(\theta_t = i|Z, M) 1_{(z_t^d = r, z_{t-1}^d = k)}}{\sum_d \sum_t P(\theta_t = i|Z, M) 1_{(z_{t-1}^d = k)}} \\ &= \frac{\sum_d \sum_t \hat{u}_t(i) 1_{(z_t^d = r, z_{t-1}^d = k)}}{\sum_d \sum_t \hat{u}_t(i) 1_{(z_{t-1}^d = k)}} \end{aligned} \tag{5.6}$$

$$\begin{aligned} K^{ji} &= \frac{\sum_t P(\theta_{t-1} = j, \theta_t = i, Z|M)}{\sum_t \sum_i P(\theta_{t-1} = j, \theta_t = i, Z|M)} \\ &= \frac{\sum_t P(\theta_{t-1} = j, \theta_t = i|Z, M)}{\sum_t P(\theta_{t-1} = j|Z, M)} \\ &= \frac{\sum_t \hat{v}_t(j, i)}{\sum_t \hat{u}_{t-1}(j)} \end{aligned} \tag{5.7}$$

5.4 The algorithm

Here, one details the different steps of the algorithm.

It is crucial to provide a first image of conditional probabilities of transition. For this, we use the classical density estimator of a Markov chain:

$$\forall i \neq j : \hat{\lambda}_{ij} = \frac{N_{ij}(T)}{\int_0^T Y_i(s) ds}$$

(see the studies in [9] and [4]).

With $N_{ij}(T)$ the number of jumps from i to j over the duration of the sampling and $Y_i(s)$ the number of individuals in state i at date s. To compute the denominator is similar than calculating the aggregate duration of all individuals spent in state i. Subsequently, we compute:

$$\lambda_i = \sum_{i \neq j} \lambda_{ij}$$

To obtain a transition probability for a duration, we use the matrix exponential:

$$P(t) = \exp(\lambda t)$$

Iterations

We do the step (E) :
We derive two estimators on the Θ hidden variable using the recursive calculation of two probabilities: a α probability called forward and an another β called backward.

Computation of α:

In t=0:
We do not know the state of the individuals before the simulation. To remedy this, we consider that the initial state is randomly distributed. So we have :

$$\forall i :$$

$$\begin{aligned}
\alpha_1(i) &= P(Z_1 = Z_1, \theta_1 = i) \\
&= \prod_d P(z_1^d = z_1^d | \theta_1 = i) P(\theta_1) \\
&= \prod_d (\sum_j P(z_1^d = z_1^d | z_0^d = j, \theta_1 = i) P(z_0^d = j | \theta_1 = i)) \pi(i) \\
&= \prod_d (\sum_j P(z_1^d = z_1^d | z_0^d = j, \theta_1 = i) P(z_0^d = j)) \pi(i) \\
&= \prod_d (\sum_j L^{i,j z_1^d} P(z_0^d = j)) \pi(i)
\end{aligned}$$

We estimate then $P(z_0^d = j)$ with the proportion of individuals in the state j just before the beginning of the simulation.

To do for $t = 2...T$:
To do for all j:

$$\begin{aligned}
\alpha_t^{(q)}(j) &= P(Z_{1|t} = Z_{1|t}, \theta_t = j) \\
&= \sum_k K^{kj(q)} \alpha_{t-1}^{(q)}(k) \prod_d L^{j, z_{t-1}^d z_t^d (q)}
\end{aligned} \tag{5.8}$$

Computation of β:

To do for all j,k :

$$\beta_T(j,k) = 1$$

To do for all $t = T-1...1$:
To do for all j:

$$\begin{aligned}
\beta_t^{(q)}(j) &= P(Z_{t+1|T} = Z_{t+1|T} | \theta_t = j, Z_t = Z_t) \\
&= \sum_l \beta_{t+1}(l) K^{jl(q)} \prod_d L^{l, z_t^d z_{t+1}^d (q)}
\end{aligned} \tag{5.9}$$

We compute the like-hood of the observations:

$$L_T^{(q)} = \sum_j \alpha_T^{(q)}(j)$$

Then the estimators of the missing data:

$$\hat{u}_t^{(q)}(j) = P(\theta_t = j | Z_{1|T} = Z_{1|T}) = \frac{\beta_t^{(q)}(j)\alpha_t^{(q)}(j)}{L_T^{(q)}}$$

To initialise $\hat{v}$, we choose:

$$\hat{v}_1^{(q)}(j,k) = P(\theta_1 = k, \theta_0 = j | Z_{1|T})$$

Therefore we use:

$$\begin{aligned}\hat{v}_1^{(q)}(j,k) &= P(\theta_1 = k | \theta_0 = j) P(\theta_0 = j) \\ &= K^{jk}\pi(j)\end{aligned}$$

To do for $t = 2...T$:

$$\hat{v}_t^{(q)}(j,k) = P(\theta_t = j, \theta_{t-1} = k | Z_{1|T} = Z_{1|T}) = \frac{\beta_t^{(q)}(k) K^{jk} \alpha_{t-1}^{(q)}(j) \prod_d L^{k, z_{t-1}^d z_t^d (q)}}{L_T^{(q)}}$$

We then go to the step (M):

It uses the estimators of each individual to finally compute our parameter estimators of the following iteration:

$$\pi_i^{(q+1)} = \hat{u}_1^q(i)$$

$$L^{i,kr(q+1)} = \frac{\sum_d \sum_t \hat{u}_t^{(q)}(i) 1_{(z_t^d = r, z_{t-1}^d = k)}}{\sum_d \sum_t \hat{u}_t^{(q)}(i) 1_{(z_{t-1}^d = k)}}$$

$$K^{ji(q+1)} = \frac{\sum_t \hat{v}_t^{(q)}(j,i)}{\sum_t \hat{u}_{t-1}^{(q)}(j)}$$

<u>End</u>

Depends on the desired precision.

6 Implementation and results

In this section, the different steps and choices for the algorithm's implementation are presented.

6.1 Numerical issue

One issue encountered concerns the numbers manipulated. Indeed, during the derivation of the backward, forward probabilities, and then for the estimators, a high number of probabilities are used. The issue comes from the adaption of the original model over the fact that the observations arise from the same realisation of Θ. With this hypothesis, the probabilities manipulated have to deal with all the sample and therefore a high number of joined experiences.
The computations induce the manipulation of very small numbers (formula dealing with products of thousands of numbers inferior than 1).
The computer is not able to work with so small numbers and attributes them the null value. This approximation make the model dysfunctional.
Since the same kind of techniques have been used to tackle this issue, we will only detail as example the computation of the estimator $\hat{u}_t(j) = P(\theta_t = j | Z_{1|T} = Z_{1|T})$.

Let us remind that:

$$\alpha_t(j) = \sum_k K^{kj} \alpha_{t-1}(k) \prod_d L^{j, z_{t-1}^d z_t^d}$$

In this recursive formula one can notice that the product of the conditional probabilities L and the terms $\alpha_{t-1}(k)$ are tiny. We will work by taking the logarithm of the value. If we note m(t) the mean of $(\log(\alpha_{t-1}(k)))_k$, the formula can be rewritten as:

$$\log(\alpha_t(j)) = \log(\sum_k K^{kj} \alpha_{t-1}(k)) + \sum_d \log(L^{j, z_{t-1}^d z_t^d})$$

The member on the right is therefore easily calculable as sum of log number. For the first part:

$$\begin{aligned}
\log(\sum_k K^{kj} \alpha_{t-1}(k)) &= \log(\sum_k K^{kj} e^{\log(\alpha_{t-1}(k)) - m(t-1) + m(t-1)}) \\
&= \log(e^{m(t-1)} \sum_k K^{kj} e^{\log(\alpha_{t-1}(k)) - m(t-1)}) \\
&= m(t-1) + \log(\sum_k K^{kj} e^{\log(\alpha_{t-1}(k)) - m(t-1)})
\end{aligned}$$

With this form, removing the mean in the exponential allows us to avoid the rounding error. With this trick, we avoid all the approximations. One can note that the mean can be replaced by the maximum or the minimum depending of the data.
We compute with a similar method $log(\beta_t(k))_{t,k}$ and the like-hood L_T.
It is then easy to compute :

$$\log(\hat{u}_t(j)) = \log(\alpha_t(j)) + \log(\beta_t(j)) - \log(L_T)$$

On which one can apply the exponential function to have directly $\hat{u}_t(j)$ since $\hat{u}$ may not have an extreme value.

6.2 The data

The data chosen in this report come from OpenDataSoft. One chooses the base "Credit Rating Agency Ratings History Data" and more precisely the ratings given by Moody's Investor's services. At this moment, it was difficult to know if it is better to have a homogeneous or heterogeneous sample (to compare the ratings of really different instruments might make no sense).
Therefore one chose the data which concerns the financial instruments and one limited our sample at 30000 values for the sake of rapidity.
The main attributes of a value or jump are the following:
-Issuer name
-Rating
-Maturity Date
-Rating action date
These choices were personal and can be different.

6.3 Implementation and parameters

The language used is Python.
The code is composed in 5 parts. one distinguishes the main function and 4 modules:

-*Parameters*: where one can find all the main parameters for the implementation. In particular there are the number of states of Θ, the interval of dicretization for BaumWelsh called *Delta* and the number of cuts per days for the computation of the filter formula,(3.7) page 16, (see the subsection about the discretization).
-*Data*: The module where all the functions for the treatment of the data are implemented. The important parts of the data are stored in a list of dictionaries. Each dictionary is liked to a date and contains the current ratings of each obligation at this date.
One can also find all the functions which computes some information as the number of jumps, the initialisation and perturbation of the conditional transition probabilities of the hidden parameter.
-*BaumWelsh*: All the functions which concern the algorithm are stored in this module. Especially, one locates the different computations of the estimators and parameters with the update mentioned above to avoid the underflow.
-*ModelTheta*: Here one can find the function for the filtration formula ((3.7)).

6.4 The discretization of the model

Two discretizations have been necessary during the implementation.
The first one concerns the BamWelsh algorithm. With our data, the interval of time deals with 2 thousands days. Therefore it seems essential to choose well the length of the pieces of the interval's cutting.
If our discretization is too fine, we will have a lack of rapidity and we will risk to not have enough jumps during these small periods.
Then the algorithm will not be efficient at all. On the other end if the interval of cutting in too large, we will miss too much information. In deed, the algorithm can only consider

an only jump during one period for one obligation. If this last has the time to make several jumps, the algorithm will miss this piece of information.
We had to find a compromise and have chosen 20 days.
The second cutting, done for using the filtration formula (3.7), had to be different.
In deed, this equation describes a continues phenomenon thus we had to choose a cutting as fine as possible in discrete time. Therefore the first idea was to try intervals of 1 day. But each time, the system diverged.
That might be explained by the fact that 1 days was still too large. The interval of times had to be very small to be used with the density.
Plus, the ΔN should not exceed 1 since there is not simultaneous jumps. Although we can sometimes observed more than 10 identical jumps at the same day.
Furthermore the data do not permit us to cut more than that (we don't know precisely when a jump appends in the day). To solve this issue, one decided to cut each days in div=100 and then to attribute to each part, uniformly, what appended during the day (the number of jumps in the day divided by 100).

6.5 Implementation's choices

To make the results clearer and more analysable, one decided to reduce the number of rating to 3. Since the initial number of different ratings was 22 we have joined under the same rating, the ratings with the same first letter.
Therefore we worked with the following ratings: A,B,C (C is almost the state of default).

index	1	2	3
rating	A	B	C

-We reduced the study at 30000 different data for the sake of rapidity.
-For the estimation of the parameters, we chose an interval of dicretization of 20 days.
-For the filtration formula (3.7), we chose an interval of dicretization of $\frac{1}{100}$.
-We modified in the following tests the number of states of Θ from 2 to 3.

6.6 The different experiments

One experience is done in two phases.
Firstly, the first one chooses the initial conditions.
The initial value of L, the conditional probabilities of a rating, is given as detailed in the algorithm (we use the classical estimator of a density of a Markov chain).
Then one can add different perturbations ϵ to each matrices of transition to differentiate the conditioning by the different states of Θ.
For the values of K and π, one works with the uniform distribution to which we might add a perturbation to promote or disadvantage some transitions.
The algorithm is run to estimate the 3 parameters. Then they are used, during a last step, in the filtration formula (3.7).

6.6.1 Uniform initial conditions

One limits our studies for the beginning at 2 states of Θ.
The first idea is to try to run the algorithm with no perturbation. In other words, one takes, as initial conditions, uniform distributions for Π and K, respectively, the initial distribution of the hidden parameter and its transitions probabilities.
Furthermore one keeps the same matrix of transition probabilities (the one got from classical estimation of the density) for all states of Θ. By this way ones considers from the beginning that Θ has no effect on the ratings and the same chance to appear.
Since one brings no information, without any perturbation, the algorithm is not able to differentiate the states of Θ. Therefore it is normal that it sends exactly the same parameters that one has given to him.

	0	1	2
0	-8.13916e-05	8.10729e-05	3.187e-07
1	5.27817e-05	-0.0001039	5.11181e-05
2	3.58489e-06	4.69332e-05	-5.05181e-05

(a) Estimation of the density l

	A	B	C
A	0.998374	0.00161846	7.19316e-06
B	0.00105372	0.997925	0.00102079
C	7.20981e-05	0.000937275	0.998991

(b) Initial L[0] and L[1]

Figure 6.1 Initial transformation of the rating process densities matrices into probabilities

	A	B	C
A	0.998374	0.00161846	7.19316e-06
B	0.00105372	0.997925	0.00102079
C	7.20981e-05	0.000937275	0.998991

(a) outcome of L[0] and L[1]

	0
0	0.5
1	0.5

(b) Outcome of Pi

	0	1
0	0.5	0.5
1	0.5	0.5

(c) Outcome of K

Figure 6.2 Estimation of the parameters

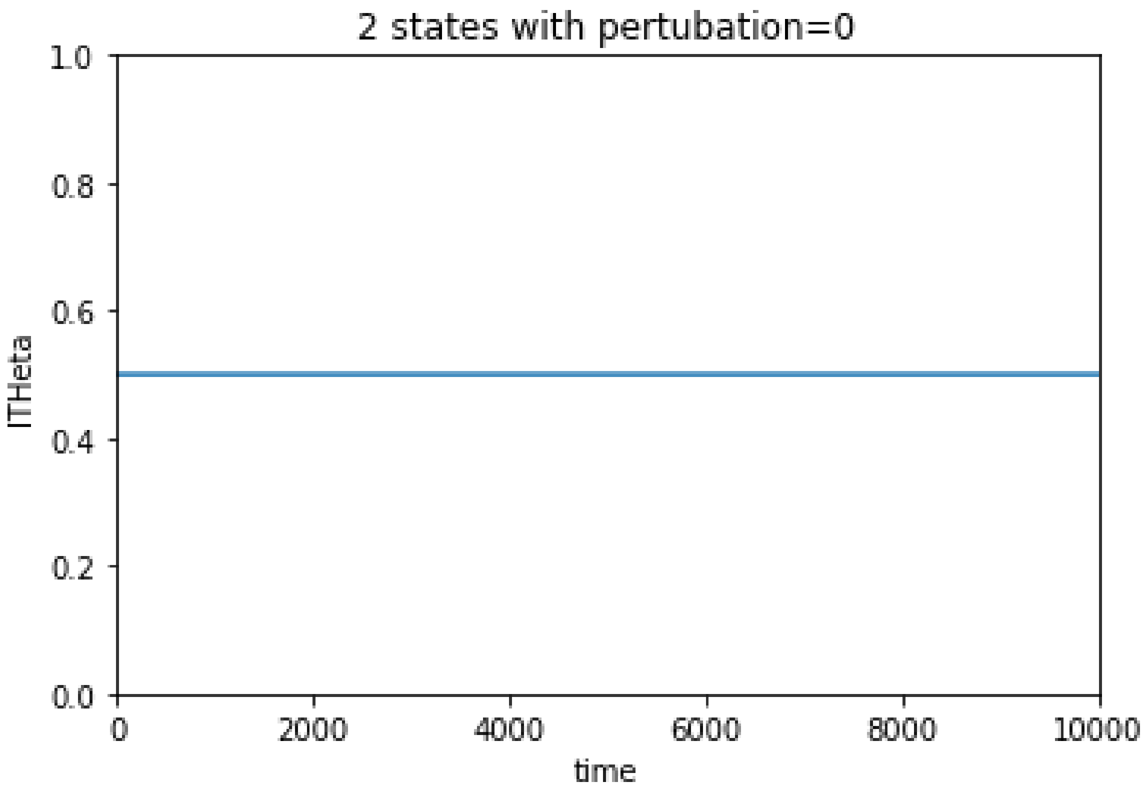


Figure 6.3 $\hat{\Theta}_t$ on 100 days

In that way, the algorithm can't send any information about Θ whatever the data are. Why would it favour some states compared to others if it has no profile, no way to differentiate them from the beginning ?
With such parameters, it is not surprising that the filtration formula (3.7) remains at 0.5, the mean of the possible values of Θ.

6.6.2 Perturbation of L with $\epsilon = 10^{-8}$ and 2 states of Θ

With the previous result, it seemed to be necessary to bring some information about the states of the hidden parameter to help the algorithm to differentiate them.
Therefore, with 2 states, it seemed to be logical to choose one state which tends to influence the obligations to jump to a better grade, the progressive state and the other which tends to influence the obligation to a worst grade, the recessive state.
These two states would describe the health of the global economy. This would help to respond to the following question : Is the economy in a favourable state to provide better confidence to credit ?
To model that difference, one decides to remove ϵ to the recessive transition probabilities and

to add it to the progressives for the progressive state, the state 1. The opposite manipulation is done for the recessive state 0.

	A	B	C
A	0.996606	0.00329374	9.99874e-05
B	0.00362464	0.995316	0.00105912
C	0.00112266	0.0011916	0.997686

(a) Initial L[0]

	A	B	C
A	0.996606	0.00329372	9.99674e-05
B	0.00362466	0.995316	0.0010591
C	0.00112268	0.00119162	0.997686

(b) Initial L[1]

	A	B	C
A	0.982773	0.0172267	1.1124e-54
B	0.000328031	0.997695	0.00197728
C	8.52183e-72	0.00517481	0.994825

(c) Outcome of L[0]

	A	B	C
A	0.999573	0.000419623	7.74565e-06
B	0.00110519	0.997942	0.000952944
C	7.74714e-05	0.00062146	0.999301

(d) Outcome of L[1]

Figure 6.4 Evolution of L

One may observe that 0 remains almost the recessive state and 1 the progressive.
In deed, one can notice that 0 favours the degradation transitions $A \to B$ and $B \to C$ whereas the improvement transition $B \to A$ is more frequent in 1.
It is also clear that 1 is more stable than 0 since 1 favours the most the stationary distributions.
It is also remarkable that 0 is favourable to the recovery distribution $C \to B$. So it is better to be in 0 only if you are in C. The riskier and recessive state 0 is preferable when you have already the worst rate.
The rare events $C \to A$ and $A \to C$ are more frequent in 1 but remain negligible.

	0	1
0	0.282091	0.717909
1	0.0542156	0.945784

(a) Outcome of K

	0
0	0.984975
1	0.015025

(b) Outcome of Pi

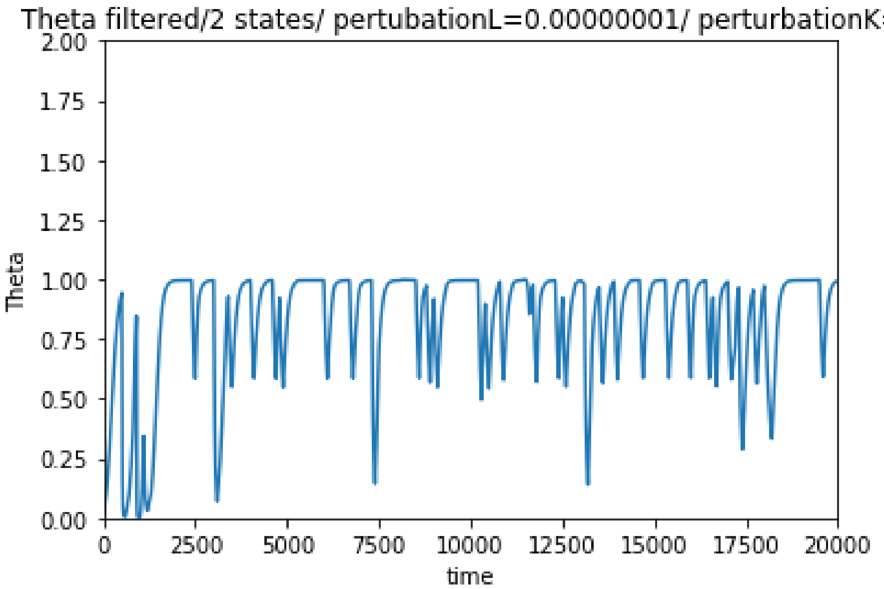
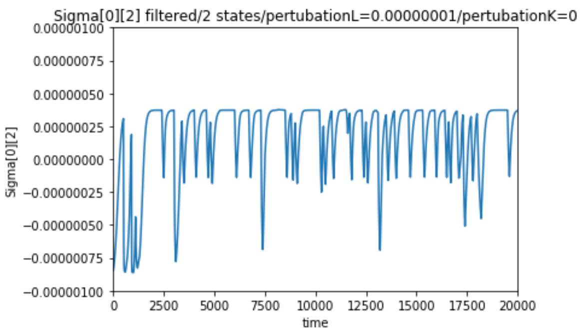
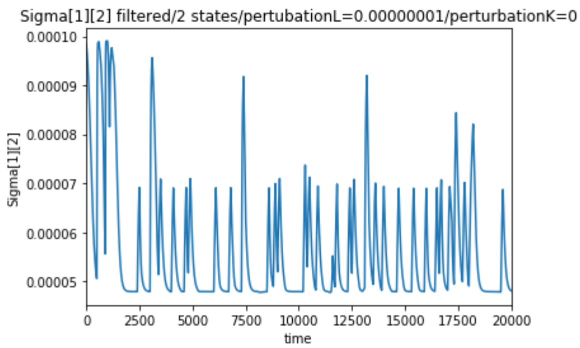


Figure 6.6 $\hat{\Theta}_t$ on 200 days



(a) Outcome of sigma[A][C]



(b) Out of sigma[B][C]

The studies of the the new of value of K and π induce a major part of the time spent in 1, the stable and progressive state, with frequent but short jumps to 0 the recessive state.
The mean of Θ of 0.82464578 confirms that the state 1 is the more frequent.

One can deduce that the obligations were used to be in a state where the chances of improvement and stagnation are higher. Sometimes they move to a riskier state where they tend more to degrade their own grates. So, the obligations live a period of recession but never for a long time (the transition probability from 0 to 1 is 0.71).
This short moment is also beneficial if an obligation is already in poor condition. In the riskier state, the chance to recover from the "default" is higher.

6.6.3 Perturbation of L with $\epsilon = 0.001$ and 3 states of Θ

In this experiment, a perturbation makes the state 0 progressive and 1 recessive as before. The state 2 added will keep the value of the estimation and should be the one closest to the data since one did not modify it.

	A	B	C
A	0.994606	0.00429373	0.00109998
B	0.00262465	0.995316	0.00205911
C	0.000122666	0.000191614	0.999686

(a) Initial L[0]

	A	B	C
A	0.997706	0.00229373	0
B	0.00462465	0.995316	5.9108e-05
C	0.00212267	0.00219161	0.995686

(b) Initial L[1]

	A	B	C
A	0.996606	0.00329373	9.99774e-05
B	0.00362465	0.995316	0.00105911
C	0.00112267	0.00119161	0.997686

(c) Initial L[2]

	A	B	C
A	0.986105	0.0138949	1.95774e-83
B	0.000260049	0.994643	0.00509728
C	1.81416e-130	0.00405952	0.99594

(a) outcome of L[0]

	A	B	C
A	0.998428	0.00157167	1.85536e-77
B	0.0200035	0.97975	0.000246973
C	4.89762e-61	2.15612e-41	1

(b) outcome of L[1]

	A	B	C
A	0.99962	0.000371967	8.26521e-06
B	0.000163867	0.999171	0.000665249
C	8.35177e-05	0.000668145	0.999248

(c) outcome of L[2]

	0	1	2
0	0.393671	9.64524e-16	0.606329
1	1.20531e-34	0.33341	0.66659
2	0.0573416	0.0286621	0.913996

(a) Outcome of K

	0
0	0.999997
1	1.45329e-64
2	2.75331e-06

(b) Outcome of Pi

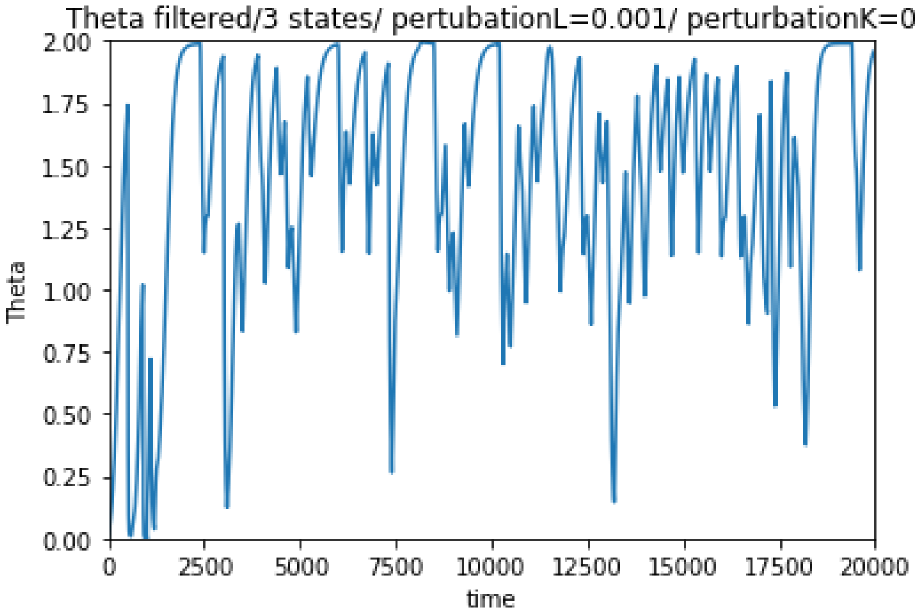
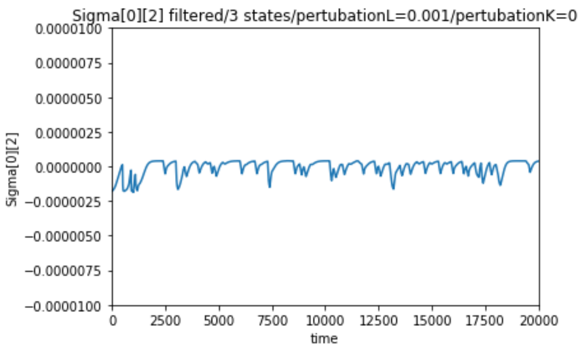
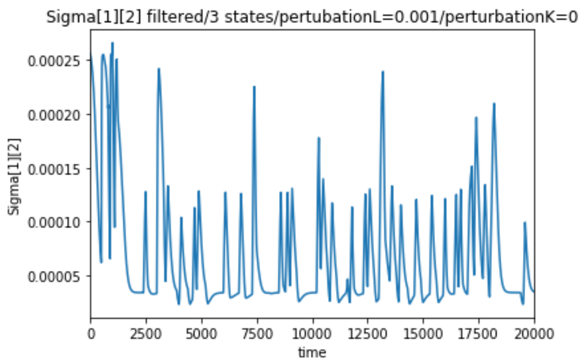


Figure 6.11 $\hat{\Theta}_t$ on 200 days



(a) Outcome of sigma[0][2]



(b) Outcome of sigma[1][2]

The mean of Θ is 1.43832512. That shows that 2 is clearly the most frequent state. It is not surprising since it was the state with the initial conditions the closest to the data.
0 remains the recessive state with the highest probabilities of degradation (A to B and B to C). The recessive one has also the highest chance of recovery (the jump from C to B).
1 remains the progressive since it is conducive to the improvements from B to A.
2 takes position between these two states with the highest probabilities of the rare and big jumps (between C and A).

The values of the parameters and the mean of Θ induce a major part of the time spent in 2 with some short jumps to 1 and 0. One can observe that the direct jumps are very rare. That confirms that the state 2 is the "intermediate" state.

6.6.4 Perturbation of K with $\epsilon = -0.49$ and 2 states of Θ

It was also interesting to perturb the frequencies of the state of Θ. In that way the transition from 0 to 1 is a lot increased, and consequently the chance to be in 1.

	0	1
0	0.01	0.99
1	0.5	0.5

(a) Initial K

	0	1
0	0.91816	0.0818398
1	0.742485	0.257515

(b) Outcome of K

	0
0	0.913485
1	0.0865149

(c) Outcome of Pi

	A	B	C
A	0.999617	0.000374552	7.97611e-06
B	0.000248075	0.998779	0.000973234
C	8.01302e-05	0.000642844	0.999277

(a) Outcome of L0

	A	B	C
A	0.986953	0.0130466	1.17558e-50
B	0.00861482	0.989918	0.0014671
C	1.32778e-57	0.00358013	0.99642

(b) Outcome of L1

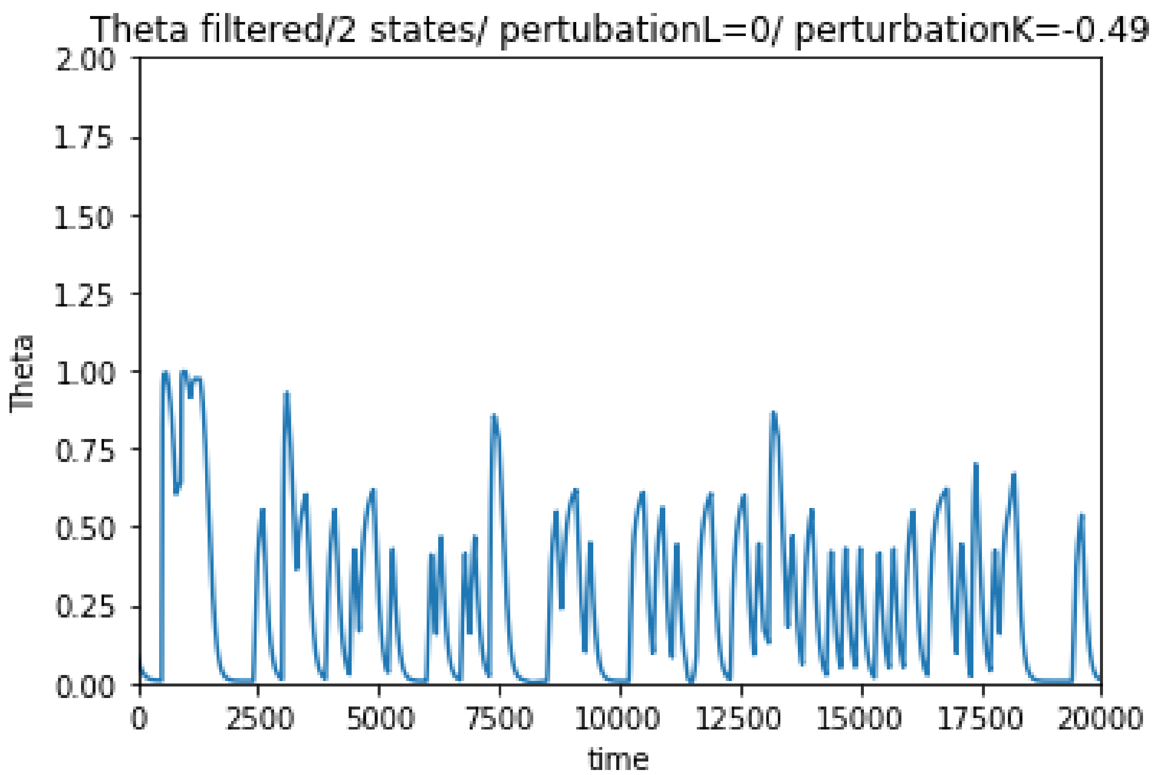


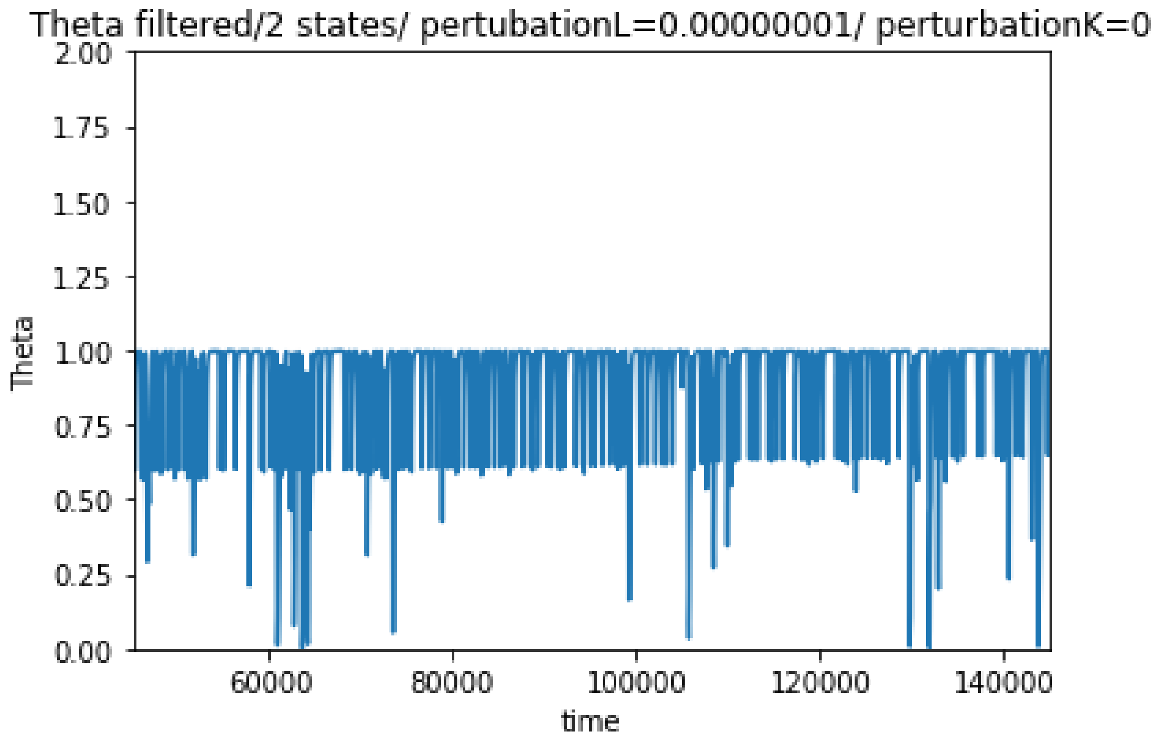
Figure 6.15 $\hat{\Theta}_t$ on 200 days

One finds the same kind of equilibrium than during the experiment with the perturbation on L: while one has increased a lot the transition $0 \to 1$, this last has now a very small probability. The influence has no effect.
0 and 1 exchange their roles, 0 is the most frequent with fast jumps to 1. It is the stable state and 1 is the recessive one since it induces $A \to B$ and $B \to C$. On can therefore make the same interpretations on the economy than during L's perturbations.

The algorithm sends back the same kind of profile for the states of Θ whatever the initial conditions are.
The results might be different because the equilibrium found are local.
The economy is used to be in a stable and positive state for the credit and knows regular and short jumps to a negative state where an obligation has more chance to deteriorate itself but also to recover from the default (a riskier state).
This study may be a tool to detect trouble periods for the credit risk.

6.6.5 Crisis detection

One tries to extend the interval of observations to reaper characteristic moment.
One remains that with this perturbation, Θ is used to stays in 1, the stable and progressive state with short and frequent jumps to 0, the riskier and recessive one.



(a) Out of Θ, from the 8th semptember 2013 to the 15th of september 2017 in hundredths of a day

One can notice that it seems that there were many jumps around 630 days on the scale of

time. This moment corresponds to the summer of 2015. The results show that there were lot of jumps to the recessive state during this period. Therefore the economy was in a riskier and in an unfavourable state for the credit.
That is not surprising since the 12th june 2015, the stock market crashed in China (A third of the value of Shanghai Stock Exchange was lost in a month).

6.7 Summary

The algorithm uses the perturbation to find an equilibrium which fits the best as possible the data. Theses equilibrium are local, therefore the algorithm does not send back exactly the same result with different initial conditions but almost.
The economy is used to be in a stable and progressive state with frequent and short jumps to a recessive and riskier state for the credit.
A raise of the frequency of these jumps may be a witness of a crisis.
It remains difficult to interpret the result when the number of state of Θ increases since one has to give initial perturbations.
Plus it is arduous to detect which probability has the weight in a state of Θ.

7 Extension and criticisms of the solution proposed

7.1 Some critics

One can criticize this solution on many points:

The Filtration formula (3.7) describes a continuous model whereas the algorithm of the parameters estimation considers that the time is discrete.
One looses the information of duration. To be as precise as possible, the passage using the matrix logarithm should also be avoided.
-The cutting of the days (to not have more than one jump each step) to use the filtration formula (3.7) is not really an acceptable approximation. One should cut the line of time at the jumps dates.
-The selection of the data has been done quickly. It would have been interesting to choose different types of data (not financial ones), or by mixing types of instruments rated. Since the financial instruments are used to remain to the highest rating, they might not be the best type in order to try to have a clear image of the economy.

7.2 The possible extensions and improvements

Several perspectives emerge from this work.

7.2.1 The choice of initial conditions

Many kinds of perturbations may be tested to better represent the real economy.
First the perturbation should not be the same for all the transitions. One must have different ϵ for each transitions, depending on the value of the transition probability. Indeed it is not acceptable that a very frequent transition looses the same probability than a rare event.
An other idea would consist in keeping the profile of the conditional transition probabilities matrices. By reducing the number of unknown parameters and in estimating only the fixed perturbation ϵ, we would have an algorithm more efficient. Furthermore we can keep the profile of progressive and recessive matrices and therefore have a better quality analysis.

7.2.2 Estimation of the parameters in continues time

A research track would extend the estimation algorithm proposed to continuous time.
Indeed, a continuous time model keeps the precise date of bonds rating changes whereas a discrete model will confuse the dates of jumps belonging to the same period of subdivision. This loss of information is not negligible and therefore represents an important issue of our model.

A second improvement would use the filter formula (3.7) directly during the algorithm. The estimators used in the algorithm (5.3) and (5.4) (obtained with the backward and forward probabilities) would be directly replaced by the filtered image obtained with the filtration formula.

7.2.3 Detection of regimes by statistical classification

Another perspective would be to analyse the sequences of the observed jumps with different methods of unsupervised learning such as the ones about classification.
By trying to distinguish the similarities among the neighbours, one would like find a cutting of the sequence of jumps for which Θ_t would remain in the same state on each part of the cutting.
For a significant change of values, we will suppose that Θ has jumped to a new state.
Some variations of nearest neighbour classification algorithms (regularly used in DNA sequence studies in biology) or the use of convolutional neural networks (CNN), used in image recognition, would have a real potential for detecting these changes of regime .

7.2.4 Use of the filtration formula

Once the parameters estimated, the filtration formula (3.7) could be used to represent the dynamics of Θ filtered with the observations. Then, once the model validated on the data *in-sample*, different scenarios of Θ could be simulated and projected in order to obtain new simulations of rating changes.

8 Conclusion

The achievement of this internship at Nexialog Consulting provided a clear vision of the integration of a research project into the daily life of an engineer and a company.
I had the chance to experiment the research steps, especially the documentation one, much longer than expected. When conducting such a project, it was very interesting to observe the limits between autonomy and support, or freedom and supervision.
From a technical point of view, the management and the outcome of such a project, although perfectible, represents a real satisfaction. So that an approach to pursue the subject during a PHD has been initiated.

References

[1] C. Bishop. Pattern recognition and machine learning (information science and statistics). 2006.

[2] P. Brémaud. *Point Processes and Queues: Martingale Dynamics.* Springer-Verlag, 1981.

[3] A. Cousin, N. Ragnar, and T. Picard. Filtering with multivariate counting processes and an application to credit risk. 2019.

[4] A. Cousin and M. Reda Khelioouen. A comparative study on the estimation of factor migration models, 2015.

[5] Rüdiger Frey and Wolfgang Runggaldier. Pricing credit derivatives under incomplete information: a nonlinear-filtering approach. *Finance and Stochastics*, 14(4):495–526, 2010.

[6] P. Gagliardini and C. Gourieroux. Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics*, 3(2):188–226, 2005.

[7] A. Karr. *Point Processes and their Statistical Inference.* Marcel Dekker Inc., New York, Basel, Hong Kong, 1991.

[8] S. J. Koopman, A. Lucas, and A. Monteiro. The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142(1):399–424, 07 2007.

[9] D. Lando and M. Skodeberg. Statistical analysis of rating transisitions-a continuous-time approach. 2001.

[10] V. Leijdekker and P. Spreij. Explicit computations for a filtering problem with point process observations with applications to credit risk, 2008.

[11] T. Proietti and A. Luati. Maximum likelihood estimation of time series models: the kalman filter and beyond, 2012.

[12] L. R.Rabinet. A tutorial on hidden markov models and selected applications in speech recognition. 1989.

[13] A. Tenyakov. Estimation of hidden markov models and their applications in finance. 2014.

[14] J.H. van Schuppen. Filtering, prediction, and smoothing for counting process observations – a martingale approach. *SIAM J. Appl. Math.*, 1997.

[15] Lindsten F. Fritsche C. Özkan, E. and Fredrik Gustafsson. Recursive maximum likelihood identification of jump markov nonlinear systems.