



**HAL**  
open science

## Rating transitions forecasting: a filtering approach

Areski Cousin, Jérôme Lelong, Tom Picard

► **To cite this version:**

Areski Cousin, Jérôme Lelong, Tom Picard. Rating transitions forecasting: a filtering approach. 2022.  
hal-03347521v3

**HAL Id: hal-03347521**

**<https://hal.science/hal-03347521v3>**

Preprint submitted on 21 Mar 2022 (v3), last revised 31 May 2023 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rating transitions forecasting: a filtering approach

Areski Cousin\*      J. Lelong†      T. Picard‡

March 21, 2022

## Abstract

Analyzing the effect of business cycle on rating transitions has been a subject of great interest these last fifteen years, particularly due to the increasing pressure coming from regulators for stress testing. In this paper, we consider that the dynamics of rating migrations is governed by an unobserved latent factor. Under a point process filtering framework, we explain how the current state of the hidden factor can be efficiently inferred from observations of rating histories. We then adapt the classical Baum-Welch algorithm to our setting and show how to estimate the latent factor parameters. Once calibrated, we may reveal and detect economic changes affecting the dynamics of rating migration, in real-time. To this end we adapt a filtering formula which can then be used for predicting future transition probabilities according to economic regimes without using any external covariates. We propose two filtering frameworks: a discrete and a continuous version. We demonstrate and compare the efficiency of both approaches on fictive data and on a corporate credit rating database. The methods could also be applied to retail credit loans.

## 1 Introduction

Credit risk research has been on the rise over the last 20 years. In particular, the challenges that arose from the previous financial crisis prompted researchers to develop credit risk valuation models that take into account the evolution of the business cycle. The evolution of the banking supervisor regulations and accounting rules follow this trend: official guidelines of IFRS 9 as [40] recommend the use of point-in-time estimation of credit risk, i.e., the use of macro-economic factors in the credit risk assessment process. Moreover, the EBA guidelines [2] on LGD downturn, require to identify economic downturn periods to adjust the initial LGD estimations. In addition, EBA stress testing methodology described in [14] strongly relies on past economical scenarios.

A credit rating system evaluates the confidence in the ability of the borrower to comply with the credit's terms. A default probability is associated to each rating, which under the capital rule, Basel regulations (see [20]), impacts the amount of capital required for a credit.

Such ratings may be generated by internal rating systems (IRB) or issued when it is available, by external rating agencies (like Moody's, Standard and Poor's, Fitch Ratings...). After the assignment of the initial credit rating, reviews are performed either periodically or based on market events. In that way an entity's rating may evolve through time according to its health and the economic cycle. Therefore, predicting the evolution of rating migrations is of primary importance for every financial institution. The migrations of a group of credit entities can be described by

---

\*Institut de Recherche en Mathématique Avancée, Université de Strasbourg, 7 rue René Descartes, 67084 Strasbourg, cedex, France. (a.cousin@unistra.fr).

†Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France. (jerome.lelong@univ-grenoble-alpes.fr).

‡Univ. Grenoble Alpes, LJK, 38000 Grenoble, France and Nexialog, 75011 Paris, France. (tom.picard@grenoble-inp.org).

transition matrices, defining the probabilities to move from one rating state to another in a given period of time. Given recent evolution in banking supervisory and accounting rules, the challenge is to explain changes in transition probabilities due to changes in the business cycle.

Factor-based migration models provide a nice framework for capturing migration sensitivities to macro-economic changes. Factor migration models allow transition probabilities to depend on dynamic factors. Two main families of models are usually considered in the credit risk literature : the “ordered Probit” (or structural approach) introduced by [44], popularized by [34] and studied for credit ratings , e.g., in [3], [17], [22], [35] and the “multi-state latent factor intensity model” (or intensity approach) studied, e.g., in [18], [28], [29]. This paper focuses on the second approach.

In the basic reduced intensity form model, a credit event corresponds to the first jump time of a Poisson process with a constant intensity. The reduced form approach has been widely studied in the credit risk literature, see, e.g., [13], [26]... Nevertheless we can point out a limit of this naive modeling: many studies show that rating migrations’ dynamics first exhibit the non-Markovian behavior (migration data exhibit correlation among rating change dates, known as “rating drift”, time-dependent default and transition probabilities, contagion effect ...) that cannot be captured by this model. Other papers such as [1], [19] and [25] highlighted that migration intensities vary over time. In their research, [25] and [30] came to the conclusion that the rating transition probabilities depend on whether the bond entered its current rating by an upgrade or a downgrade. [30] also noticed that the probability to leave a rating category tends to decrease with the time spent at that rating. Above all, [3], [35] gave strong evidence that credit risk exposure is considerably affected by the macroeconomic conditions and differs across different economic regimes. The first paper highlighted quantified rating migrations’ dependency on the industry, domicile of the obligor and on the stage of the business cycle.

In the factor intensity approach, the migration dynamics of each credit entity is described by a stochastic intensity matrix (or generator matrix) whose components depend on a pool of common factors.

In both models the factors may be considered observable or not. The second approach has emerged in response to criticisms made against the first. As [22] pointed out, the risk in selecting covariates lies in excluding other ones which could be more relevant. [10] provided an overview of usual modelling and estimation approaches and compared the estimation and predictive performance of each approach on real data. When the underlying factors are unobservable, they adapted a method given in [22] to represent the considered factor migration model as a linear Gaussian model, and apply a Kalman filter to predict the state of the underlying latent factor. This approximation works on the hypothesis that the transition probabilities’ estimator relies on asymptotic normality. This assumption may be too restrictive and typically does not hold when the data set is too small. In this respect, [10] obtained poor calibration and predictive results using this approximation compared to the estimation performed under the observable factor approach.

Our approach consists in directly filtering the hidden factor given rating transitions’ past history. Assuming that the counting process has a latent intensity determined by the hidden factor’s state, this project will look into estimating this intensity only using observations of the counting process. For a bond portfolio, the dynamics of rating migrations can be mathematically represented as a multidimensional counting process, each component representing the cumulative number of transitions from one rating category to another. This framework has been already used in the credit risk literature for other applications: [21] used a counting process filtering approach to estimate the latent factor given default observation and use this estimation to price credit derivatives. Filtering of point processes has also been applied to credit risk in, e.g., [24] and [31].

Rating transitions are usually registered on a daily basis, so that transitions of different entities may be observed at the exact same date. [6] derives a general multivariate counting process filtering formula. Nevertheless, its formula holds in continuous time and its application relies on

the assumption of no simultaneous jumps between migration counting processes. The proposed framework cannot directly be applied to the context of rating migrations.

A discrete-time approach would be more relevant given the daily format of the data and frequent observations of simultaneous migrations. Behind every model mentioned, choosing a continuous or discrete approach is crucial and is a matter of debate. This paper aims at participating to this debate by presenting different results: we provide a general filtering formula in discrete-time and show how to adapt the continuous-time filtering framework to handle discrete-time data and simultaneous jumps. Then, the discrete-time and the continuous-time general filtering formulas are applied to credit rating migrations and compared.

For this application, we assume that the unobserved driving factor is given as a finite state Markov chain and that the rating transition process follows a Hidden Markov model (HMM). We refer to [15] for a detailed analysis of discrete-time filtering under special HMM assumptions. Hidden Markov Chain modeling (HMM) remains a popular approach in credit risk analysis ([9], [15], [16], [43],...). Among others, [23], uses an EM algorithm, the classical Baum-Welch algorithm (introduced in [4]), for estimation of a two-state hidden factor driving occurrence of defaults. They obtain estimates for the model parameters and are able to reconstruct the most likely past sequence of the risk state. Their approach only holds for an unique transition and is not suitable for providing online estimations of the risk state.

In the same vein [12] and [37] identify two hidden states as one of expansion and the other of contraction. In particular, [37] uses an extension of the Baum-Welch algorithm adapted to "regime switching hidden Markov model" (RSMC) to propose a forecasting of sovereign credit rating transitions. In a different scope, they also assumed that every rating processes are governed by an unique discrete hidden Markov chain but issued from different trajectories of this Markov chain. Those studies are theoretically different from our models in which a realization of the observable factor is common to every firm. We believe that our approach which rather keeps the dependencies within the observations sample, is more reliable and realistic. Indeed rating entities should be affected by the same realization of the economic factor. This different consideration changes the way to calibrate and to filter: our filtering framework uses the whole history of aggregated number of jumps. Furthermore for the sake of interpretability, our approach is more reliable if we want to consider the hidden factor as the universal and shared factor describing the economic cycle.

The contributions of this paper are both theoretical and practical. First, we derive a general discrete version of the filtering equation. Then continuous-time and discrete-time formulae are adapted and applied to credit rating migrations. In particular, we show how the continuous-time version can be adapted to cope with simultaneous jumps occurring in discrete-time observations. An EM algorithm is adapted for both settings to estimate the parameters involved. Applying the recursive equations, we update the filtered factor to forecast rating transitions. Our approach may be considered as a new Point-in-time (PIT) rating transitions modeling which does not use any macro-economic factors. We assess and compare both approaches on a fictive data set and on a Moody's ratings history [01/2000-05/2021] of a diversified portfolio of 5030 corporate entities. To the best of our knowledge, this study is the first attempt to apply and compare both discrete and continuous multivariate counting process filtering to predict future rating transitions.

The paper is organised as follows. First, we present the derivation of the discrete-time point process filtering methods and its application to credit migration processes throughout Section 2. Section 3 presents the continuous-time filtering version and its adaptations to discrete-time credit rating environment. Section 4 shows and validates the two filtering approaches on fictive data. Finally, in Section 5, we test and compare the two filters on real data sets.

## 2 Discrete version of the filter

We present in this section a discrete version of the filter. To apply filtering to credit rating migrations, discrete time environment tends to be more compliant with daily format of rating data and releases the strong hypothesis of jumps simultaneity: jumps can occur at the same time.

### 2.1 Univariate Form

Let  $\Gamma \in \mathbb{N}$  be the discrete time horizon. We work with the filtered probability space  $(\Omega, \mathbf{A}, \mathbf{F} = (\mathcal{F}_n)_{n \in \{0, \dots, \Gamma\}}, \mathbb{P})$ . Let  $N$  be a discrete-time  $\mathbf{F}$ -adapted counting process starting from 0. Let  $\mathbf{F}^N$  be the natural filtration of  $N$ , augmented with  $\mathbb{P}$ -null sets. We write

$$\forall n \in \{1, \dots, \Gamma\}, \Delta N_n = N_n - N_{n-1}. \quad (1)$$

We denote by  $\mathbb{J}$ , the support of the jumps of  $N$  ( $\mathbb{J} = \mathbb{N}$  for a Poisson process). The support could vary over time but as this would not impact our results, it is assumed constant for the sake of clearness. To account for the times when  $N$  does not jump, we assume that  $0 \in \mathbb{J}$  and we define  $\bar{\mathbb{J}} = \mathbb{J} \setminus \{0\}$  as the support of the true jumps. Let  $\Theta$  be a square integrable  $\mathbf{F}$ -adapted process such that conditionally on the  $\sigma$ -field  $\mathcal{F}_{n-1}$ ,  $\Theta_n$  and  $\Delta N_n$  are independent. Let  $\mathbf{F}^\Theta$  be the natural filtration of  $\Theta$ , augmented with  $\mathbb{P}$ -null sets.  $\Theta$  has a natural decomposition of the form

$$\Theta_n = A_n + M_n, \quad (2)$$

where  $A$  is a  $\mathbf{F}$ -predictable square integrable process and  $M$  is a square integrable  $\mathbf{F}$ -martingale. We define the filtration  $\mathbf{G} = (\mathcal{G}_n)_{n \in \{0, \dots, \Gamma\}}$  by  $\mathcal{G}_n = \mathcal{F}_n^\Theta \vee \mathcal{F}_{n-1}^N$ . For  $n \in \{1, \dots, \Gamma\}$ , we introduce

$$\hat{O}_n = \mathbb{E}[O_n | \mathcal{F}_n^N], \tilde{O}_{n-1} = \mathbb{E}[O_n | \mathcal{F}_{n-1}]. \quad (3)$$

For  $n \in \{1, \dots, \Gamma\}$  and  $j \in \mathbb{J}$ , we define

$$\epsilon_n^j = \mathbb{1}_{[\Delta N_n = j]}, \lambda_{n-1}^j = \mathbb{E}[\epsilon_n^j | \mathcal{F}_{n-1}], \hat{\lambda}_{n-1}^j = \mathbb{E}[\lambda_{n-1}^j | \mathcal{F}_{n-1}^N] = \mathbb{E}[\mathbb{E}[\epsilon_n^j | \mathcal{F}_{n-1}] | \mathcal{F}_{n-1}^N]. \quad (4)$$

Note that, for all  $n$ , we have  $\sum_{j \in \mathbb{J}} \epsilon_n^j = \sum_{j \in \mathbb{J}} \lambda_n^j = \sum_{j \in \mathbb{J}} \hat{\lambda}_n^j = 1$ . Under this setting, we present a discrete time adaptation of the univariate filtering equation presented in [6].

**Proposition 1.** *The filtered process  $\hat{\Theta}$  satisfies the following equation*

$$\hat{\Theta}_n = \sum_{j \in \mathbb{J}} \frac{\widehat{(\tilde{\Theta} \lambda^j)}_{n-1}}{\hat{\lambda}_{n-1}^j} \mathbb{1}_{[\Delta N_n = j]}, \quad \forall n = 1, \dots, \Gamma. \quad (5)$$

The proof of Proposition 1 heavily relies on the following lemma.

**Lemma 2.** *Let  $K$  be a square integrable and  $\mathbf{G}$ -adapted process such that  $\hat{K}$  is a  $\mathbf{F}^N$ -martingale. Then,  $K$  is solution of the following recursive equation*

$$\hat{K}_n = K_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \frac{\widehat{(\tilde{K} \lambda^j)}_{k-1}}{\hat{\lambda}_{k-1}^j} (\epsilon_k^j - \hat{\lambda}_{k-1}^j), \quad \forall n = 1, \dots, \Gamma. \quad (6)$$

**Remark 3.** *If  $K$  is also a  $\mathbf{F}$ -martingale then the previous equation also holds true with  $\tilde{K} = K$ .*

*Proof (of Lemma 2).* Let  $P$  be a square integrable  $\mathbf{F}^N$ -martingale. Then, there exists a measurable function  $g$  such that  $P_n = g(n, N_0, \dots, N_n)$ . It can equivalently be written  $P_n = h(n, \Delta N_1, \dots, \Delta N_n)$ . We can write

$$P_n = \sum_{j \in \mathbb{J}} h(n, \Delta N_1, \dots, \Delta N_{n-1}, j) \epsilon_n^j.$$

Since  $P$  is a  $\mathbf{F}^N$ -martingale,

$$P_n - P_{n-1} = P_n - \mathbb{E}[P_n | \mathcal{F}_{n-1}^N] = \sum_{j \in \mathbb{J}} h(n, \Delta N_1, \dots, \Delta N_{n-1}, j) (\epsilon_n^j - \hat{\lambda}_{n-1}^j).$$

Then,  $P$  has the following martingale representation

$$P_n = P_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} H_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j), \quad (7)$$

with  $H_{k-1}^j$  being  $\mathcal{F}_{k-1}^N$ -measurable. Conversely, it is easy to check that a process written like that is a  $\mathbf{F}^N$ -martingale.

Since  $\epsilon_k^0 = 1 - \sum_{j \in \mathbb{J}} \epsilon_k^j$  and  $\hat{\lambda}_k^0 = 1 - \sum_{j \in \mathbb{J}} \hat{\lambda}_k^j$ , we can rewrite  $P$  with the following martingale representation

$$P_n = P_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} W_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j), \quad (8)$$

with  $W_{k-1}^j = H_{k-1}^j - H_{k-1}^0$ .

Let  $K$  be a square integrable  $\mathbf{G}$ -adapted process such that  $\hat{K}$  is a  $\mathbf{F}^N$ -martingale. According to (8), we can write  $\hat{K}_n = \zeta + \sum_{k=1}^n \sum_{j \in \mathbb{J}} W_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j)$ , where the sequence  $(W_k)_k$  is  $\mathcal{F}^N$ -adapted. For any square integrable  $\mathbf{F}^N$ -adapted process  $X$ , we have  $\mathbb{E}[(K_n - \hat{K}_n)X_n] = 0$ . Choosing  $X$  to be a  $\mathbf{F}^N$ -martingale with the decomposition  $g + \sum_{k=1}^n \sum_{j \in \mathbb{J}} G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j)$  with  $(G_k)_k$   $\mathbf{F}^N$ -adapted, we obtain

$$\mathbb{E} \left[ \left( K_n - \zeta - \sum_{k=1}^n \sum_{j \in \mathbb{J}} W_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right) \left( g + \sum_{k=1}^n \sum_{j \in \mathbb{J}} G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right) \right] = 0. \quad (9)$$

Choosing  $G_{k-1}^j = 0$  for all  $j \in \mathbb{J}$  leads to  $\zeta = \mathbb{E}[\hat{K}_n] = \mathbb{E}[K_n] = K_0$ .

With no loss of generality, we can consider that  $K_0 = 0$ . By choosing  $g = 0$  we obtain:

$$\mathbb{E} \left[ K_n \sum_{k=1}^n \sum_{j \in \mathbb{J}} G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right] - \mathbb{E} \left[ \sum_{j_1, j_2 \in \mathbb{J}} \sum_{k_1, k_2 \geq 1} W_{k_1-1}^{j_1} (\epsilon_{k_1}^{j_1} - \hat{\lambda}_{k_1-1}^{j_1}) G_{k_2-1}^{j_2} (\epsilon_{k_2}^{j_2} - \hat{\lambda}_{k_2-1}^{j_2}) \right] = 0. \quad (10)$$

For  $k_1 < k_2$ ,

$$\mathbb{E} \left[ \mathbb{E} \left[ W_{k_1-1}^{j_1} (\epsilon_{k_1}^{j_1} - \hat{\lambda}_{k_1-1}^{j_1}) G_{k_2-1}^{j_2} (\epsilon_{k_2}^{j_2} - \hat{\lambda}_{k_2-1}^{j_2}) | \mathcal{F}_{k_2-1}^N \right] \right] = 0$$

Noticing that  $\mathbb{E} [K_n | \mathcal{F}_k^N] = \mathbb{E} [\mathbb{E} [K_n | \mathcal{F}_n^N] | \mathcal{F}_k^N] = \hat{K}_k$ , we compute the term

$$\begin{aligned} S_{n,k} &= \mathbb{E} \left[ K_n G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right] = \mathbb{E} \left[ \mathbb{E} [K_n | \mathcal{F}_k^N] G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right] \\ &= \mathbb{E} \left[ \hat{K}_k G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right] = \mathbb{E} \left[ K_k G_{k-1}^j (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \right]. \end{aligned}$$

Now, let us compute

$$\begin{aligned} \mathbb{E} [K_k (\epsilon_k^j - \hat{\lambda}_{k-1}^j) | \mathcal{F}_{k-1}^N] &= \mathbb{E} [K_k \epsilon_k^j | \mathcal{F}_{k-1}^N] - \mathbb{E} [K_k \hat{\lambda}_{k-1}^j | \mathcal{F}_{k-1}^N] \\ &= \mathbb{E} \left[ \mathbb{E} [K_k \epsilon_k^j | \mathcal{F}_{k-1}^N] | \mathcal{F}_{k-1}^N \right] - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j. \end{aligned}$$

Remember that  $K$  is  $\mathbf{G}$ -measurable and that  $\mathcal{G}_k = \mathcal{F}_k^\theta \vee \mathcal{F}_{k-1}^N$ . The conditional independence of  $\Theta_k$  and  $\Delta N_k$  knowing  $\mathcal{F}_{k-1}$  yields the conditional independence of  $K_k$  and  $\epsilon_k$ . Hence, we obtain

$$\begin{aligned} \mathbb{E} \left[ \mathbb{E}[K_k \epsilon_k^j | \mathcal{F}_{k-1}] | \mathcal{F}_{k-1}^N \right] - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j &= \mathbb{E} \left[ \mathbb{E}[K_k | \mathcal{F}_{k-1}] \mathbb{E}[\epsilon_k^j | \mathcal{F}_{k-1}] | \mathcal{F}_{k-1}^N \right] - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j \\ &= (\widehat{\tilde{K} \lambda^j})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j. \end{aligned}$$

From 10, we obtain

$$\mathbb{E} \left[ \sum_{k=1}^n \sum_{j \in \mathbb{J}} G_{k-1}^j \left( (\widehat{\tilde{K} \lambda^j})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j - \mathbb{E} \left[ \sum_{i \in \mathbb{J}} W_{k-1}^i (\epsilon_k^i - \hat{\lambda}_{k-1}^i) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) | \mathcal{F}_{k-1}^N \right] \right) \right] = 0$$

As this holds true for any  $\mathbf{F}^N$ -adapted process  $(G_k)_k$ , we can choose

$$G_{k-1}^j = (\widehat{\tilde{K} \lambda^j})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j - \mathbb{E} \left[ \sum_{i \in \mathbb{J}} W_{k-1}^i (\epsilon_k^i - \hat{\lambda}_{k-1}^i) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) | \mathcal{F}_{k-1}^N \right].$$

Then, we deduce the following system of linear equations,

$$\forall j \in \mathbb{J}, \sum_{i \in \mathbb{J}} W_{k-1}^i \mathbb{E} \left[ (\epsilon_k^i - \hat{\lambda}_{k-1}^i) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) | \mathcal{F}_{k-1}^N \right] = (\widehat{\tilde{K} \lambda^j})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j \quad a.s. \quad (11)$$

Let  $c_j^{(k-1)} = (\widehat{\tilde{K} \lambda^j})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j$ . Note that

$$\mathbb{E} \left[ (\epsilon_k^i - \hat{\lambda}_{k-1}^i) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) | \mathcal{F}_{k-1}^N \right] = \begin{cases} -\hat{\lambda}_{k-1}^i \hat{\lambda}_{k-1}^j & \text{for } i \neq j \\ \hat{\lambda}_{k-1}^j (1 - \hat{\lambda}_{k-1}^j) & \text{for } i = j \end{cases}$$

The equations in (11) can be written for all  $j \in \mathbb{J}$

$$\hat{\lambda}_{k-1}^j (1 - \hat{\lambda}_{k-1}^j) W_{k-1}^j - \sum_{\substack{i \in \mathbb{J} \\ i \neq j}} \hat{\lambda}_{k-1}^i \hat{\lambda}_{k-1}^j W_{k-1}^i = c_j^{(k-1)}.$$

Then, we obtain for all  $j \in \mathbb{J}$

$$\hat{\lambda}_{k-1}^j W_{k-1}^j - \hat{\lambda}_{k-1}^j \sum_{i \in \mathbb{J}} \hat{\lambda}_{k-1}^i W_{k-1}^i = c_j^{(k-1)}. \quad (12)$$

By summing for  $j \in \mathbb{J}$ , we deduce from (12),

$$\sum_{j \in \mathbb{J}} \hat{\lambda}_{k-1}^j W_{k-1}^j - \sum_{j \in \mathbb{J}} \hat{\lambda}_{k-1}^j \sum_{i \in \mathbb{J}} \hat{\lambda}_{k-1}^i W_{k-1}^i = \sum_{j \in \mathbb{J}} c_j^{(k-1)}.$$

Then

$$\hat{\lambda}_{k-1}^0 \sum_{j \in \mathbb{J}} \hat{\lambda}_{k-1}^j W_{k-1}^j = \sum_{j \in \mathbb{J}} c_j^{(k-1)}.$$

Inserting the expression of  $\sum_{i \in \mathbb{J}} \hat{\lambda}_{k-1}^i W_{k-1}^i$  in (12) gives

$$W_{k-1}^j = \frac{c_j^{(k-1)}}{\hat{\lambda}_{k-1}^j} + \frac{\sum_{i \in \mathbb{J}} c_i^{(k-1)}}{\hat{\lambda}_{k-1}^0}.$$

Then, we obtain

$$W_{k-1}^j = \frac{1}{\hat{\lambda}_{k-1}^j} (\widehat{(\tilde{K}\lambda^j)})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^j + \frac{1}{\hat{\lambda}_{k-1}^0} \sum_{i \in \mathbb{J}} (\widehat{(\tilde{K}\lambda^i)})_{k-1} - \hat{K}_{k-1} \hat{\lambda}_{k-1}^i.$$

Finally, by replacing  $\sum_{j \in \mathbb{J}} \epsilon_k^j$ ,  $\sum_{j \in \mathbb{J}} \hat{\lambda}_{k-1}^j$  by  $1 - \epsilon_k^0$  and  $1 - \hat{\lambda}_{k-1}^0$ , we derive the general filtering formula

$$\begin{aligned} \hat{K}_n &= K_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \left( \frac{(\widehat{(\tilde{K}\lambda^j)})_{k-1}}{\hat{\lambda}_{k-1}^j} - \frac{(\widehat{(\tilde{K}\lambda^0)})_{k-1}}{\hat{\lambda}_{k-1}^0} \right) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \\ &= K_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{K}\lambda^j)})_{k-1}}{\hat{\lambda}_{k-1}^j} (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \end{aligned}$$

Finally,  $K$  is solution of the following recursion formula

$$\hat{K}_n = K_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{K}\lambda^j)})_{k-1}}{\hat{\lambda}_{k-1}^j} (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \quad \blacksquare$$

*Proof (of Proposition 1).* We define

$$a_n = A_n - A_{n-1} = \mathbb{E}[\Theta_n | \mathcal{F}_{n-1}] - \Theta_{n-1}, \quad (13)$$

$$B_n = \sum_{k=1}^n \mathbb{E}[a_k | \mathcal{F}_{k-1}^N]$$

$$L_n = \sum_{k=1}^n a_k - \mathbb{E}[a_k | \mathcal{F}_{k-1}^N].$$

Note that

$$\Theta_n - B_n = M_n + L_n.$$

We know that  $\hat{M}$  is a  $\mathbf{F}^N$ -martingale and  $\hat{L}$  is clearly a  $\mathbf{F}^N$ -martingale too. Then, we can apply Lemma 2 to  $\Theta - B$ . Note that  $B$  is  $\mathbf{F}^N$  predictable, then  $\tilde{B}_{k-1} = B_k$  and so  $\widehat{\tilde{B}\lambda^j}_{k-1} = B_k \hat{\lambda}_{k-1}^j$ . Then,

$$\begin{aligned} \hat{\Theta}_n - \hat{B}_n &= \hat{\Theta}_0 - \hat{B}_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \left( \frac{(\widehat{((\tilde{\Theta} - \tilde{B})\lambda^j)})_{k-1}}{\hat{\lambda}_{k-1}^j} \right) (\epsilon_k^j - \hat{\lambda}_{k-1}^j) \\ &= \hat{\Theta}_0 - \hat{B}_0 + \sum_{k=1}^n \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{\Theta}\lambda^j)})_{k-1}}{\hat{\lambda}_{k-1}^j} (\epsilon_k^j - \hat{\lambda}_{k-1}^j). \end{aligned}$$

We compute  $\hat{\Theta}_n - \hat{\Theta}_{n-1} = \hat{B}_n - \hat{B}_{n-1} + f(\lambda_{n-1}, \epsilon_n, \tilde{\Theta}_{n-1}) = \mathbb{E}[a_n | \mathcal{F}_{n-1}^N] + f(\lambda_{n-1}, \epsilon_n, \tilde{\Theta}_{n-1})$ . From (13),  $\mathbb{E}[a_n | \mathcal{F}_{n-1}^N] = \mathbb{E}[\Theta_n | \mathcal{F}_{n-1}^N] - \hat{\Theta}_{n-1}$  and using that

$$f(\lambda_{n-1}, \epsilon_n, \tilde{\Theta}_{n-1}) = \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{\Theta}\lambda^j)})_{n-1}}{\hat{\lambda}_{n-1}^j} (\epsilon_n^j - \hat{\lambda}_{n-1}^j) = \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{\Theta}\lambda^j)})_{n-1}}{\hat{\lambda}_{n-1}^j} \mathbb{1}_{[\Delta N_n = j]} - \mathbb{E}[\Theta_n | \mathcal{F}_{n-1}^N],$$

we deduce the final form of the filtering formula

$$\hat{\Theta}_n = \sum_{j \in \mathbb{J}} \frac{(\widehat{(\tilde{\Theta}\lambda^j)})_{n-1}}{\hat{\lambda}_{n-1}^j} \mathbb{1}_{[\Delta N_n = j]}. \quad (14)$$



Note that, from (13), the previous formula can also be stated as

$$\hat{\Theta}_n = \sum_{j \in \mathbb{J}} \left( \frac{\widehat{(a\lambda^j)}_{n-1}}{\hat{\lambda}_{n-1}^j} + \frac{\widehat{(\Theta\lambda^j)}_{n-1}}{\hat{\lambda}_{n-1}^j} \right) \mathbb{1}_{[\Delta N_n = j]}. \quad (15)$$

where  $\widehat{(a\lambda^j)}_{n-1} = \mathbb{E}[a_n \lambda_{n-1}^j | \mathcal{F}_{n-1}^N]$ . ■

This formula can be extended to a multivariate setting.

## 2.2 Multivariate form

Let  $\Gamma \in \mathbb{N}$  be the discrete time horizon. We work with the filtered probability space  $(\Omega, \mathbf{A}, \mathbf{F} = (\mathcal{F}_n)_{n \in \{0, \dots, \Gamma\}}, \mathbb{P})$ . Let  $N = (N^1, \dots, N^\rho)$  be a discrete-time multivariate counting process where for  $i = 1, \dots, \rho$ ,  $N^i = (N_n^i)_{n \in \{0, \dots, \Gamma\}}$ , is a simple  $\mathbf{F}$ -adapted counting processes starting from 0. Let  $\mathbf{F}^N$  be the natural filtration of  $N$ , augmented with  $\mathbb{P}$ -null sets. We write

$$\forall i \in \{1, \dots, \rho\}, \forall n \in \{1, \dots, \Gamma\}, \Delta N_n^i = N_n^i - N_{n-1}^i. \quad (16)$$

We denote by  $\mathbb{J}_i$  the support of the jumps of  $N^i$ , for  $i = 1, \dots, \rho$ . The support may vary over time but as this would not impact our results, it is assumed constant for the sake of clearness. To account for the times when  $N^i$  does not jump, we assume that  $0 \in \mathbb{J}_i$  and we define  $\bar{\mathbb{J}}_i = \mathbb{J}_i \setminus \{0\}$  as the support of the true jumps. Let us define the product spaces  $\mathbb{J}^\otimes = \prod_{i=1}^\rho \mathbb{J}_i$  and  $\bar{\mathbb{J}}^\otimes = \prod_{i=1}^\rho \bar{\mathbb{J}}_i$ . Let  $\Theta$  be a square integrable  $\mathbf{F}$ -adapted process such that conditionally on the  $\sigma$ -field  $\mathcal{F}_{n-1}$ ,  $\Theta_n$  and  $(\Delta N_n^i)_{i=1, \dots, \rho}$  are independent. Let  $\mathbf{F}^\Theta$  be the natural filtration of  $\Theta$ , augmented with  $\mathbb{P}$ -null sets. We define the filtration  $\mathbf{G} = (\mathcal{G}_n)_{n \in \{0, \dots, \Gamma\}}$  by  $\mathcal{G}_n = \mathcal{F}_n^\Theta \vee \mathcal{F}_{n-1}^N$ .

We extend the previous setting to multivariate case,

$$\forall \delta \in \mathbb{J}^\otimes, \epsilon_n^\delta = \mathbb{1}_{[\Delta N_n = \delta]} = \mathbb{1}_{[\cap_{i=1}^\rho \Delta N_n^i = \delta_i]}, \lambda_{n-1}^\delta = \mathbb{E}[\epsilon_n^\delta | \mathcal{F}_{n-1}]. \quad (17)$$

**Proposition 4.** *The filtered process  $\hat{\Theta}$  satisfies for all  $n = 1, \dots, \Gamma$ ,*

$$\hat{\Theta}_n = \sum_{\delta \in \mathbb{J}^\otimes} \frac{\widehat{(\hat{\Theta}\lambda^\delta)}_{n-1}}{\hat{\lambda}_{n-1}^\delta} \mathbb{1}_{[\Delta N_n = \delta]}. \quad (18)$$

*Proof.* We leave the proof to the reader as it goes along the same lines as the proof of Prop. 1. ■

**Remark 5.** *This formula extends the univariate filtering formula derived in [15, Chapter 2] when the underlying factor  $\Theta$  follows a finite state space Markov chain. Under a similar Markovian setting, using ((18)), we extend the result stated in [15] to a multivariate setting and we apply this result to the context of rating migration processes. The next section is dedicated to describe this application.*

## 2.3 Application to rating migrations

As for the continuous framework, we apply this general formula to credit rating transitions. We first present the formula in the context of a single pair of rating categories; thus a single transition from one given rating to another (the transition to default for instance). Then we extend the approach to multiple rating transitions.

### 2.3.1 Unique rating transition

In this framework  $\Theta$  is assumed to be a Markov chain with finite number of states in  $\mathbb{T} = \{1, \dots, m\}$ . We present here the recursive equation verified by  $\hat{I}_n^h = \mathbb{E}[\mathbb{1}_{[\Theta_n=h]} | \mathcal{F}_n^N]$ , where  $N$  represents the cumulative number of the considered transition. The process  $Z^q$  takes the value 1 when the entity  $q$  realizes the transition and 0 otherwise. We define  $\forall n \in \{1, \dots, \Gamma\}$ :

$$K^{sh} = \mathbb{P}(\Theta_n = h | \Theta_{n-1} = s), \quad \Pi^h = \mathbb{P}(\Theta_0 = h), \quad L^s = \mathbb{P}(Z_n^q = 1 | Z_{n-1}^q = 0, \Theta_{n-1} = s),$$

the hidden factor's transition probabilities, its initial state's probabilities and the common conditional transition probabilities of entities.

The process  $Y$  represents the number of entities that may jump. This process is assumed to be  $\mathbf{F}^N$ -predictable. According to the previous notations, the support of  $\Delta N_n$  is  $\{0, \dots, Y_n\}$ .

Knowing the event  $\{\Theta_{n-1} = h, Y_n = y_n\}$ , we assume that the conditional distribution of the random variable  $\Delta N_n$  is binomial with parameters  $(y_n, L^h)$ . Similar settings can be found in [7] and [23].

**Proposition 6.** *With these assumptions, the filtered process  $\hat{I}_n^h$  solves the following recursive equation for  $n = 1, \dots, \Gamma$*

$$\hat{I}_n^h = \sum_{j \in \mathbb{J}} \frac{\sum_{s=1}^m K^{sh} (L^s)^j (1-L^s)^{Y_n-j} \hat{I}_{n-1}^s}{\sum_{s=1}^m (L^s)^j (1-L^s)^{Y_n-j} \hat{I}_{n-1}^s} \mathbb{1}_{[\Delta N_n=j]}. \quad (19)$$

*Proof.* We have

$$\mathbb{E}(\mathbb{1}_{[\Delta N_n=j]} | \mathcal{F}_{n-1}) = \mathbb{E}(\mathbb{1}_{[\Delta N_n=j]} | N_{n-1}, \Theta_{n-1}, Y_n), \quad (20)$$

and

$$\mathbb{P}(\Delta N_n = j | \Theta_{n-1} = h, Y_n = y_n) = \binom{y_n}{j} (L^h)^j (1-L^h)^{y_n-j}. \quad (21)$$

Then we compute the following expressions

$$\hat{\lambda}_{n-1}^j = \mathbb{E}[\mathbb{1}_{[\Delta N_n=j]} | \mathcal{F}_{n-1}^N] = \sum_{h=1}^m \binom{Y_n}{j} (L^h)^j (1-L^h)^{Y_n-j} \hat{I}_{n-1}^h, \quad (22)$$

$$\widehat{(I^h \lambda^j)}_{n-1} = \binom{Y_n}{j} (L^h)^j (1-L^h)^{Y_n-j} \hat{I}_{n-1}^h \quad (23)$$

and

$$\begin{aligned} \widehat{(\tilde{I}^h \lambda^j)}_{n-1} &= \mathbb{E}[\mathbb{E}[I_n^h | \mathcal{F}_{n-1}] \mathbb{E}[\mathbb{1}_{[\Delta N_n=j]} | \mathcal{F}_{n-1}] | \mathcal{F}_{n-1}^N] \\ &= \mathbb{E} \left[ \left( \sum_{\mu=1}^m K^{\mu h} I_{n-1}^\mu \right) \left( \sum_{s=1}^m \binom{Y_n}{j} (L^s)^j (1-L^s)^{Y_n-j} I_{n-1}^s \right) | \mathcal{F}_{n-1}^N \right] \\ &= \sum_{s=1}^m K^{sh} \binom{Y_n}{j} (L^s)^j (1-L^s)^{Y_n-j} \hat{I}_{n-1}^s. \end{aligned}$$

Finally using (5) we obtain

$$\hat{I}_n^h = \sum_{j \in \mathbb{J}} \frac{\sum_{s=1}^m K^{sh} (L^s)^j (1-L^s)^{Y_n-j} \hat{I}_{n-1}^s}{\sum_{s=1}^m (L^s)^j (1-L^s)^{Y_n-j} \hat{I}_{n-1}^s} \mathbb{1}_{[\Delta N_n=j]}. \quad (24) \quad \blacksquare$$

In order to apply our framework to rating migrations, the previous setting is extended using the multivariate filtering formula (18).

### 2.3.2 Multiple Rating transitions

In practice the number of entities monitored over time may vary: either because some names appear or disappear or simply because of missing data. This happens when the data is missing, censored or when it is not appeared yet. This consideration is deeply discussed in Section 5.1. We attribute the rating 0 to an entity in this case. Then it is clear that a transition involving the rating of censure 0, is assumed to be independent with the states of the hidden factor. Let consider the list of ratings  $\tilde{\Upsilon} = \{1, \dots, p\}$  and  $\Upsilon = \{0, \dots, p\}$ , the completed list of ratings. Note that the number of entities observed on  $\Upsilon$  is constant equal to  $Q$ . Let  $Z_n^q$  be the random variable  $\in \Upsilon$ , describing the state of bond  $q$ ,  $q \leq Q$ , at time  $n \in \{0, \dots, \Gamma\}$  and  $Z^q = (Z_n^q)_{n \in \{0, \dots, \Gamma\}}$  be the migration process that describes its evolution. The counting process associated, which counts the total number of jumps of the entities, from rating  $i$  to  $r$ ,  $\in \Upsilon^2$ , is denoted by  $N^{ir}$  and is such that,  $\forall n \in \{1, \dots, \Gamma\}$ ,

$$\Delta N_n^{ir} = \sum_{q \leq Q} \mathbb{1}_{[Z_{n-1}^q=i, Z_n^q=r]}. \quad (25)$$

We define the transition probabilities of  $\Theta$  as

$$\forall (s, h) \in \mathbb{T}^2, \forall n \in \{1, \dots, \Gamma\}, K^{sh} = \mathbb{P}(\Theta_n = h | \Theta_{n-1} = s).$$

We also define the conditional transition probabilities of  $(Z^q)_q$

$$\forall (i, r) \in \Upsilon^2, \forall s \in \mathbb{T}, \forall n \in \{1, \dots, \Gamma\}, L^{s,ir} = \mathbb{P}(Z_n^q = r | Z_{n-1}^q = i, \Theta_{n-1} = s).$$

Note that if  $i = 0$  or  $r = 0$  then  $\forall s \in \mathbb{T}$ ,  $L^{s,ir} = \mathbb{P}(Z_n^q = r | Z_{n-1}^q = i)$  as the transitions from or to the rating 0 are assumed to be independent of the hidden factor.

We can notice that the processes  $(Z^q)_{q \leq Q}$  are not independent. Indeed, only one realisation of trajectory of  $\Theta$  governs observed rating processes.

The process  $Y^i$  represents the number of entities that belong to rating  $i$ . This process is assumed to be  $\mathbf{F}^N$ -predictable. According to the previous notations, the support  $\mathbb{J}_i$  of the counting processes  $\Delta N_n^{ir}$ ,  $r \in \tilde{\Upsilon}$ , is  $\{0, \dots, Y_n^i\}$ . The conditional distribution of the multivariate random variable  $\Delta N_n^{ir}$ , knowing the event  $\{\Theta_{n-1} = s, Y_n^i = y_n^i\}$ , is multinomial with parameters  $(y_n^i, (L^{s,ir})_r)$ .

**Proposition 7.** *With such assumptions, the filtered process  $\hat{I}_n^h$  is solution of the following recursive equation*

$$\hat{I}_n^h = \sum_{\delta \in \mathbb{J}^\otimes} \frac{\sum_s K^{sh} \prod_{i,r=1}^p (L^{s,ir})^{\delta_{ir}} \hat{I}_{n-1}^s}{\sum_s \prod_{i,r=1}^p (L^{s,ir})^{\delta_{ir}} \hat{I}_{n-1}^s} \mathbb{1}_{[\Delta N_n = \delta]} \quad (26)$$

*Proof.* We leave the proof to the reader as it goes along the same lines as the proof of Proposition 6. ■

Once the hidden factor filtered state is obtained, we are able to predict the future migration probabilities.

### 2.3.3 Transition probability prediction

We define for  $(i, r) \in \tilde{\Upsilon}^2$ , the process  $\nu^{ir}$ , which forecasts for the next time step, according to the hidden factor, the transition probability from rating  $i$  to rating  $r$ .

$$\forall (i, r) \in \tilde{\Upsilon}^2, \forall n \in \{1, \dots, \Gamma\} : \nu_{n-1}^{ir} = \mathbb{E} [\mathbb{1}_{[Z_{n-1}^q=r]} | Z_{n-1}^q = i, \mathcal{F}_{n-1}] = \sum_{h \in \mathbb{T}} L^{h,ir} \hat{I}_{n-1}^h \quad (27)$$

With the filtered current hidden factor, we can forecast the future transition probabilities

$$\forall (i, r) \in \tilde{\Upsilon}^2, \forall n \in \{1, \dots, \Gamma\} : \hat{\nu}_{n-1}^{ir} = \sum_{h \in \mathbb{T}} L^{h,ir} \hat{I}_{n-1}^h \quad (28)$$

## 2.4 Calibration

In this section, we explain how to estimate model parameters involved in the filtering equation (26). We apply the so-called Baum-Welch algorithm to our discrete-time framework.

### 2.4.1 A Baum-Welch algorithm adapted for a discrete framework

The proposed method is a maximisation expectation (EM) algorithm for hidden Markov chains (HMM), adapted to the model. We can find studies on the classical model in [5], [38], [39] and [42].

However the classical algorithm is not totally suitable for calibration of the discrete filtering equation (26). We highlight two inconsistencies between the classical algorithm and our model:

- In HMM classical framework, rating status law is only governed by the state of the hidden factor. In the provided framework, it also depends on the previous rating.
- Rating process trajectories of each entity must be independent whereas in our framework, they are dependent through the common factor  $\Theta$ .

The first step of the algorithm assigns initial values to the parameters we want to estimate. Then the algorithm replaces the missing data (states of  $\Theta$ ) with Bayesian estimators using the observations and the current parameters estimated values.

The second one consists in improving a conditional likelihood. Better parameters are estimated. Then these new estimates are used to repeat the first step. We iterate this process to converge to a local maximum.

Let  $Z = (Z^q)_{q \leq Q}$  be the multivariate rating process and we call for  $(n_1, n_2) \in \{0, \dots, \Gamma\}^2$ ,  $(Z)_{n \in \{n_1, \dots, n_2\}} = Z_{n_1|n_2}$ , the rating trajectories between time  $n_1$  and  $n_2$ . As the new rating does not only depend on the economic cycle (state of  $\Theta$ ) but also on the previous rating, we apply the Baum-Welch algorithm by considering that

$$\forall n \in \{1, \dots, \Gamma\}, \mathbb{P}(Z_n | Z_0, \dots, Z_{n-1}, \Theta_0, \dots, \Theta_{n-1}) = \mathbb{P}(Z_n | Z_{n-1}, \Theta_{n-1}). \quad (29)$$

Furthermore, as rating history of all entities are dependent on a same realization of  $\Theta$ , we must adapt our algorithm differently from [37] who considered that each rating process is governed by its own and independent trajectory of  $\Theta$ .

### 2.4.2 Initialization

The calibration algorithm presented is based on iterative improvement of a likelihood. This expectation maximization algorithm (EM) (see [11]), as most of iterative maximisation algorithms, might be trapped in a local maximum. Obtained parameters may not be relevant when the global maximum is not found. This success is deeply dependant on the initialization. Several empirical and analytical methods have been proposed to deal with this matter. In [32], transition probabilities are initiated using empirical frequencies. They succeed to considerably reduce the number of iterations to find their local maximum. By noticing that the transition matrices have strong diagonals, [37] initialized their model by adding small perturbations to identity matrix or to uniform distributions. In our study we choose a third option which seems to be more reliable: we test a high number of initial values (picked at random) in order to find the global maximum. In order to guarantee almost surely convergence to the global maximum, initial values are chosen according to a uniform distribution on the parameters space.

### 2.4.3 Bayesian estimators

This part only presents the main results of the algorithm. One can find more details of the computations in Appendix A.

We define the forward probability as denote,

$$\forall s \in \mathbb{T}, \forall n \in \{1, \dots, \Gamma\} : \alpha_n(s) = \mathbb{P}(Z_{0|n} = z_{0|n}, \Theta_{n-1} = s) \quad (30)$$

and the backward probability as

$$\forall s \in \mathbb{T}, \forall n \in \{1, \dots, \Gamma - 1\} : \beta_n(s) = \mathbb{P}(Z_{n+1|\Gamma} = z_{n+1|\Gamma} | Z_n = z_n, \Theta_{n-1} = s). \quad (31)$$

We use the following recursive formulas in order to compute the two previous probabilities

$$\alpha_n(s) = \sum_{l=1}^m \alpha_{n-1}(l) K^{ls} \prod_{i,r \in \Upsilon} (L^{s,ir})^{\Delta N_n^{ir}}, \quad (32)$$

$$\beta_n(s) = \sum_{l=1}^m \beta_{n+1}(l) K^{sl} \prod_{i,r \in \Upsilon} (L^{l,ir})^{\Delta N_{n+1}^{ir}}. \quad (33)$$

For  $n \in \{1, \dots, \Gamma\}$ , we introduce two random variables useful to describe  $\Theta$

$$u_n(h) = \mathbb{1}_{\{\Theta_n=h\}}, \quad (34)$$

$$v_n(s, h) = \mathbb{1}_{\{\Theta_n=h, \Theta_{n-1}=s\}}. \quad (35)$$

The forward and backward probabilities are helpful to compute the following Bayesian estimators

$$\check{u}_n(h) = \mathbb{P}(\Theta_n = h | Z_{0|\Gamma} = z_{0|\Gamma}) = \frac{\beta_{n+1}(h) \alpha_{n+1}(h)}{L_\Gamma}, \quad (36)$$

and

$$\check{v}_n(s, h) = \mathbb{P}(\Theta_n = h, \Theta_{n-1} = s | Z_{0|\Gamma} = z_{0|\Gamma}) = \frac{\beta_{n+1}(h) K^{sh} \alpha_n(s) \prod_{i,r \in \Upsilon} (L^{h,ir})^{\Delta N_n^{ir}}}{L_\Gamma}, \quad (37)$$

where  $L_\Gamma$  is the likelihood of the whole sample,

$$L_\Gamma = \mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma}) = \sum_{j=1}^m \alpha_\Gamma(j). \quad (38)$$

### 2.4.4 Parameters estimation

The maximization phase consists in finding better parameters than those of the previous iteration. We call  $M^{(\gamma)} = (\Pi^{(\gamma)}, L^{(\gamma)}, K^{(\gamma)})$ , the parameters obtained at the iteration  $(\gamma)$ .

The new parameters are deemed to improve the likelihood according to:

$$\mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma+1)}) \geq \mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma)}). \quad (39)$$

To achieve that, we are looking for maximizing  $\log \frac{\mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma+1)})}{\mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma)})}$ , which is equivalent to maximize

$$Q(M^{(\gamma)}, M^{(\gamma+1)}) = \sum_{\theta \in \{1, \dots, m\}^\Gamma} \mathbb{P}(\Theta_{0|\Gamma} = \theta, Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma)}) \log \mathbb{P}(\Theta_{0|\Gamma} = \theta, Z_{0|\Gamma} = z_{0|\Gamma} | M^{(\gamma+1)}). \quad (40)$$

After optimization, we obtain the following estimators

$$\Pi^h = \check{u}_0(h); \quad L^{s,ir} = \frac{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s) \Delta N_n^{ir}}{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s) Y_n^i}; \quad K^{sh} = \frac{\sum_{n=1}^{\Gamma} \check{v}_n(s, h)}{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s)}. \quad (41)$$

### 3 Continuous-time version of the filter

In this section, we present the different results of point process filtering in continuous time. We explain how we can adapt the continuous time framework to discrete-time data with simultaneous jumps in the framework of rating transitions.

#### 3.1 Framework and statements

Let  $(\Omega, \mathbf{A}, \mathbf{F} = (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ , be a filtered probability space satisfying the ‘‘usual conditions’’ of right-continuity and completeness needed to justify all operations to be made. All stochastic processes encountered are assumed to be adapted to the filtration  $\mathbf{F}$  and integrable on  $[0, T]$ . In particular, we have  $\mathbf{A} = \mathcal{F}_T$ . The time horizon  $T$  is supposed to be finite.

Let  $N = (N^1, \dots, N^\rho)$  be a multivariate counting process where  $N^i = (N_t^i)_{t \in [0, T]}$ ,  $i = 1, \dots, \rho$ , is a set of simple counting processes, such that,  $N_t^i = \sum_{0 < s \leq t} \Delta N_s^i < \infty$  and  $\Delta N_s^i \in \{0, 1\}$ , for any  $i = 1, \dots, \rho$ . It is assumed that these processes admit a predictable  $\mathbf{F}$ -intensity  $\nu^i = (\nu_t^i)_{t \in [0, T]}$ , and that they do not have any common jumps, i.e.,  $\Delta N_t^i \Delta N_t^j = \delta_{ij} \Delta N_t^i$  ie  $[N^i, N^j]_t = 0$  (the continuous martingale part of a counting process being null). We introduce  $\mathbf{F}^N = (\mathcal{F}_t^N)_{t \in [0, T]}$  the natural filtration of the multivariate counting process  $N = (N^1, \dots, N^\rho)$ .

Let  $\Theta$  be a square integrable process of the form

$$\Theta_t = \int_0^t a_s ds + M_t, \quad (42)$$

where  $a$  is a  $\mathbf{F}$ -adapted process and  $M$  is a square integrable  $\mathbf{F}$ -martingale. We assume that  $\Theta$  and  $\Delta N^i$  have no common jumps. Let  $\mathbf{F}^\Theta$  be the natural filtration of  $\Theta$ , augmented with  $\mathbb{P}$ -null sets.. The problem is to estimate the states of the unobserved process  $\Theta$  using only the information  $\mathbf{F}^N$ , resulting from the observation of the multivariate counting process  $N$ . By definition of the conditional expectation,

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

is the  $L^2$  approximation of  $\Theta$  knowing  $N$ . With the same notation, all the processes  $O$  filtered by  $\mathcal{F}_t^N$  is written

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

The main result on univariate point process filtering can be stated in the following way (see [6], [27], [31], [45]).

**Proposition 8.** *The process  $\hat{\Theta}$  is solution of the SDE*

$$d\hat{\Theta}_t = \hat{a}_t dt + \sum_{j=1}^{\rho} \eta_t^j (dN_t^j - \hat{\nu}_{t-}^j dt), \quad (43)$$

with

$$\eta_t^j = \frac{\widehat{(\Theta \nu^j)}_{t-}}{\hat{\nu}_{t-}^j} - \hat{\Theta}_{t-}, \quad (44)$$

and initial condition

$$\hat{\Theta}_0 = \mathbb{E}[\Theta_0]. \quad (45)$$

The following proposition can be found in [6] but the different are expressed in terms of change measure and are not explicit. Although his result is valid with simultaneous jumps between a counting process  $N^i$  and the hidden factor  $\Theta$ , this leads to an extra term which cannot be computed in practice. We aimed at obtaining an implementable formula so we had to forbid simultaneous jumps. For the sake of completeness, we provide in Appendix B.1 a self contained proof yielding the explicit filtering formula with no simultaneous jumps.

### 3.2 Finite latent factor model and a credit risk application

We present here how the previous filtering framework can be applied to credit rating migrations. In order to remain realistic and to fix the terminology, a bond market containing a finite number of individual bonds is considered. All bonds are affected by variable and random market conditions represented by the same latent process  $\Theta$ . A bond  $q$  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 bond can only be in a state belonging to a finite set of states,  $\mathbb{Y} = \{1, \dots, p\}$ . This space represents different credit risk scores or ratings in descending order,  $p$  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 us define  $\tilde{\mathbb{Y}} = \{(i, j) \in \mathbb{Y}^2, i \neq j\}$ , the space of possible migrations. Let  $Z_t^q \in \mathbb{Y}$  be the state of bond  $q$  at time  $t$  and  $Z^q = (Z_t^q)_{t \in [s^q, u^q]}$  be the rating process describing its evolution. The migration counting process associated with  $Z^q$ , which counts the number of jumps of the entity  $q$  from rating  $i$  to  $j$ , is denoted by  $N^{q,ij}$  and is such that,  $\forall t \in [s^q, u^q]$ ,

$$\Delta N_t^{q,ij} = \mathbf{1}_{[Z_{t-}^q=i, Z_t^q=j]}. \quad (46)$$

In this study, we assume that  $(Z_t^q, t \in [0, T])_q$  are described within a factor migration model. More specifically, knowing  $\mathcal{F}_T^\Theta$ , the rating processes  $Z^q$  are assumed to be conditionally independent Markov chains with the same generator matrix. In reality the change of rating of a bond may also induce the change of state of other bonds but this contagion effect is not considered in this paper. Moreover, the censorship mechanism governing  $(s^q, u^q)$  is assumed to be non-informative and can therefore be considered deterministic and belonging to  $\mathcal{F}_0^N$ . Under this exchangeable setting, to infer information on the underlying hidden factor  $\Theta$ , it is sufficient to observe the aggregated counting processes  $N^{ij}$ ,  $(i, j) \in \tilde{\mathbb{Y}}$ , defined by

$$N_t^{ij} = \sum_{q; s^q \leq t < u^q} N_t^{q,ij}. \quad (47)$$

and the exposure processes  $Y^i$ ,  $i \in \mathbb{Y}$  defined by

$$Y_t^i = \sum_{q; s^q \leq t < u^q} \mathbf{1}_{[Z_{t-}^q=i]}. \quad (48)$$

Note that the exposure process  $Y^i$  is left continuous. It increases by 1 when  $N^{ji}$  jumps for any  $j \neq i$ , or when a new bond enters the pool with rating  $i$ . It decreases by 1 when a bond jumps outside rating  $i$ , i.e., whenever  $N^{ij}$  jumps for  $j \neq i$  or when a bond expires with rating  $i$ .

We will now consider the filtering problem of  $\Theta$  given observation of the multivariate processes  $N := (N^{ij}, (i, j) \in \tilde{\mathbb{Y}})$ , using Prop. 8. We denote by  $(\nu_t^{ij})_{(i,j) \in \tilde{\mathbb{Y}}}$  the  $\mathbf{F}$  intensity of  $N$ . The dynamic (43) suggests a recursive algorithm to update the process  $\hat{\Theta}$ . A consequence of (43) is that the dynamics of  $\hat{\Theta}$  depends on  $(\widehat{\Theta\nu^j})$  which in turn depends on the term  $(\widehat{\Theta(\nu^j)^2})$ , and so on... This filtering equation induces an infinite nesting problem.

To solve the issue of infinite imbrication, we assume that the intensities  $(\nu^{ij})_{(i,j) \in \tilde{\mathbb{Y}}}$  are governed by a finite state Markov chain. Consequently the hidden factor driving process  $\Theta$  is assumed to be a Markov chain with finite number of states in  $\mathbb{T} = \{1, \dots, m\}$  and with constant transition intensities  $k^{rh}$ ,  $r \neq h$ , such that  $k^{rr} = -\sum_{l; l \neq r} k^{rl}$  and for small enough  $dt$

$$\forall i \neq h \in \mathbb{T}^2 : \mathbb{P}(\Theta_{t+dt} = h | \Theta_t = r) \approx k^{rh} dt. \quad (49)$$

The initial distribution of  $\Theta$  is defined

$$\forall h \in \mathbb{T} : \Pi_h = \mathbb{P}(\Theta_0 = h). \quad (50)$$

Let us introduce the processes  $I^h$ ,  $h \in \mathbb{T}$ , defined by

$$I_t^h = \mathbb{1}_{[\Theta_t=h]}, \quad h \in \mathbb{T}. \quad (51)$$

The process  $\Theta$  can be represented as a finite sum of indicator processes

$$\Theta_t = \sum_{h \in \mathbb{T}} h I_t^h. \quad (52)$$

With these assumptions, the processes  $(Z^q)_q$  are governed as described before by their common intensity matrices  $(l^h)_{h \in \mathbb{T}}$ , such as for small enough  $dt$

$$\mathbb{P}[Z_{t+dt}^q = j \mid Z_t^q = i, \Theta_t = h] \approx \ell^{h,ij} dt. \quad (53)$$

Then the counting processes  $N^{jk}$  are governed by the  $\mathbf{F}$  intensities

$$\nu_t^{ij} = Y_t^i \sum_{r \in \mathbb{T}} \ell^{r,ij} I_{t-}^r. \quad (54)$$

As  $Y^i$  is  $\mathbf{F}^N$ -predictable, the  $\mathbf{F}^N$ -intensity of  $N$ , the counting process which counts the cumulative number of transitions, may be written

$$\hat{\nu}_{t-}^{ij} = Y_t^i \sum_{r \in \mathbb{T}} \ell^{r,ij} \hat{I}_{t-}^r, \quad (55)$$

By applying the previous multivariate filtering formula (43) to the process  $(I_t^h)_{t \in [0, T]}$ , we obtain the following proposition.

**Proposition 9.** *With the previous assumptions, the unobserved indicator process filtered with rating jumps satisfies the following recursive equation*

$$d\hat{I}_t^h = \sum_{r=1}^m k^{rh} \hat{I}_{t-}^r dt + \sum_{i \neq j} \left( \frac{l^{h,ij} \hat{I}_{t-}^h}{\sum_r l^{r,ij} \hat{I}_{t-}^r} - \hat{I}_{t-}^h \right) \left( dN_t^{ij} - Y_t^i \sum_{r=1}^m l^{r,ij} \hat{I}_{t-}^r dt \right) \quad (56)$$

See Appendix B.2 for a proof of this results.

Usually, information on rating migrations are only available to public on a daily basis. For large credit portfolios, it is then frequent to observe multiple transitions (of several entities) occurring at the same day. In addition, clustering of rating migrations may also happen following the disclosure of a major economic events. Then, the presented continuous-time filtering approach is not fully compliant with migration data since it precludes simultaneous jumps between counting processes. This has lead us to preprocess the data and adapt the calibration algorithm.

### 3.3 Continuous adaptation: compliance with discrete data

This part aims to explain how to adapt continuous filtering to discrete rating migration framework. We propose an adaptation of the calibration algorithm in order to be compliant with the continuous filtering formula (56). Previous adaptations done for the discrete framework are still required but are not sufficient: Baum-Welch algorithm is an estimation in discrete time which is not compliant with the continuous version of the filter. Furthermore the continuous-time filtering framework assumes the absence of simultaneous jumps.

For the first deviation, we propose to realize the calibration in discrete time anyway. Then it is



necessary to switch to continuous time dimension for filtering. The transition between probabilities to intensities turns out to be easy when the time interval chosen is small enough according to (49) and (53).

The second deviation is also essential. There is no common jumps among the rating processes. However, ratings are not natural processes. Human decisions and algorithm appreciations are reported at the same moment in a day. Therefore we have to deal with simultaneous daily observations.

Our solution consists in considering a different time grid. Each day is cut into small intervals and jumps are randomly spread on these time intervals. We insure to cut enough finely to have a maximum of one jump per interval. This manipulation has two drawbacks. First, the conditional independence between rating entities might be lost if the non-simultaneity of jumps is enforced. Then, distributing simultaneous jumps on a finer time grid may ultimately modify the original information.

The second effect is studied in a testing benchmark at Section 4.1.1.

Once the data have been modified, we adapt our calibration algorithm to respect continuous structure. To this end, we use a prior law of jumps which respects the constraint of no simultaneity (that the number of observed jumps is only 0 or 1). The main idea consists on assuming that one entity is randomly chosen to be allowed to jump. Then, the entity may jump according to the common migration matrices  $(L^{h,ij})_h, (i, j) \in \{1, \dots, p\}$ .

More details of this adaptation can be found in Appendix B.3.

## 4 Filtering on simulated data

The purpose of this section is to test and validate the continuous-time and discrete-time versions of the filter using simulated data: we build two rating migration databases from the two underlying credit migration models: the discrete-time model (as described in Section 2.3) and the continuous-time factor migration model (as described in Section 3.2). Since the structure of the two models are different, different data sample are used for testing the two approaches.

### 4.1 Discrete time filtering approach

To build the discrete-time database, we assume that the hidden factor is described by a finite state space Markov chain with 7 states. We consider a given set of model parameters. Each state of the hidden factor is associated to a specific rating migration matrix. We try to choose matrices compliant with filtering: we must have sufficient variability among conditional transition probabilities  $(L^{h,ij})_h, (i, j) \in \{1, \dots, p\}$ . We work with 3 ratings categories  $\{A, B, C\}$  and initialize our sample with 1000 entities per rating. Then the hidden Markov chain is simulated on 300 time steps. According to the hidden factor's sample path, we simulate transitions using conditional transition matrices.

At this stage, we can only use the cumulative number of transitions as inputs. We use the first 200 time-steps and the remainder to test the model.

We perform the calibration as described in Section 2.4. We recover the parameters used for data generation.

Figure 1 shows the real trajectory and the filtered trajectory of the hidden factor computed on the testing sample (of 100 time-steps).

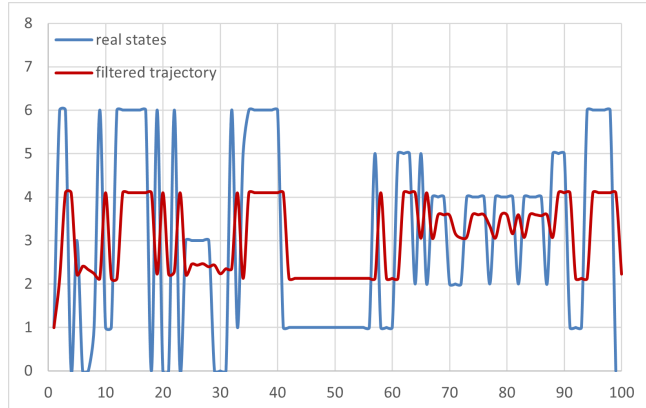


Figure 1: Simulated and filtered trajectories of the hidden factor on the last 100 testing dates

We can notice that the filter is able to detect the changes of states. It faithfully follows the real trajectory and respects the different phases and trends. However it does not exactly mimic the true value, since the filtering formula is a weighted average of the state values. We can also observe a small delay in the estimation. The explanation is theoretical: it is caused by the effect of delay in the filtering model: the impact of the hidden factor at time  $t - 1$  is observable on ratings at time  $t$ . Therefore, when the hidden factor at time  $t$  is filtered, the freshest observations available at this time, is the rating jumps at time  $t$  which have been governed by the hidden factor at time  $t - 1$ . Consequently we infer the current hidden factor state with information generated by its previous value.

We can easily understand that the calibration plays a crucial role to make the filtering efficient. In order to forecast in time, states need to be strongly linked at least to another state. Let's imagine a rare and very unstable state. Since it is hardly visited from other states, it will never influence the direction of the filter and will be difficult to predict. Once the filter realizes that the hidden factor jumps to this state, it is too late, the hidden factor has already returned to another state. Finally, the filter is unable to capture rare events to unstable states. This remains acceptable since our main purpose is to detect transitions to stable regimes. Visiting a state for a brief period of time does not represent useful information for long term forecasting.

The following Figures 2, 3, 4, 5, 6, 7, represent real ratios of observed transitions with the predicted transitions dynamics, obtained from (26) and (28), between the three considered rating categories  $\{A, B, C\}$ .

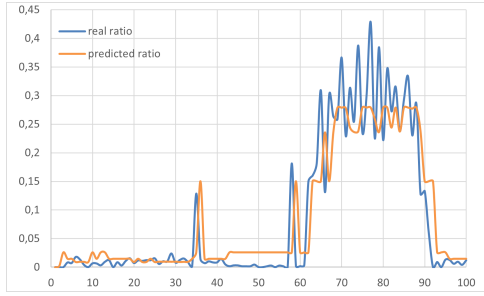


Figure 2: Real and predicted ratios from A to B

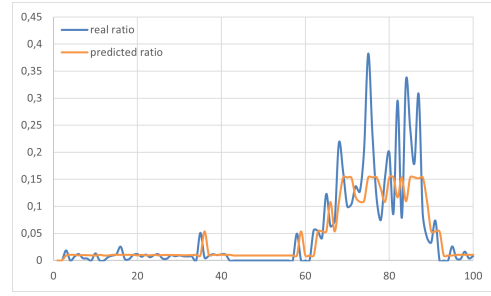


Figure 3: Real and predicted ratios from A to C

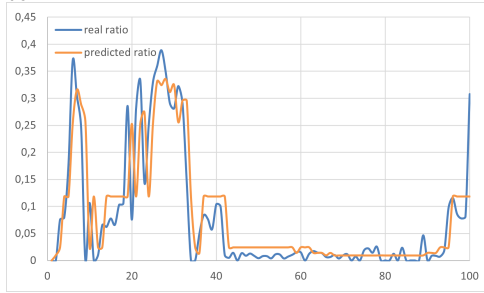


Figure 4: Real and predicted ratios from B to A

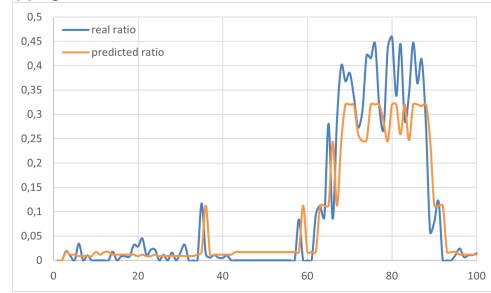


Figure 5: Real and predicted ratios from B to C

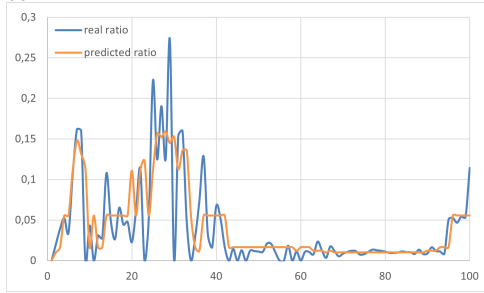


Figure 6: Real and predicted ratios from C to A

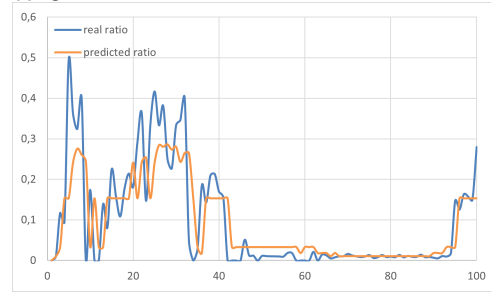


Figure 7: Real and predicted ratios from C to B

The results are very encouraging. The filter provides good predictions of future jumps. The predictions vary as a function of the regime cycle. Even when the real ratios sharply increase or decrease, the prediction are immediately corrected. Although, one of our previous intuitions is confirmed, the filtering approach can not capture extreme variations since our approach forecasts an average of the rating transition probabilities.

#### 4.1.1 Continuous framework

In order to validate the continuous-time filtering approach, we generate a data set using the migration model described in Section 3.2. The simulated rating processes exhibit no simultaneous jumps. We consider 3 rating categories and 1000 entities per class. We build a continuous Markov chain with 5 states. Note that we reduce the number of states compared to the discrete-time framework as the continuous framework is much more computationally demanding. Since the data is fictive and specific to the continuous-time model, this choice has no impact on our validation experiment. We directly applied the continuous-time filtering approach on the simulated data set, which does not contain simultaneous jumps. Then, in order to challenge the relevance of the

use of the continuous model on discrete data, we transform the data set. Jumps are aggregated and randomly spread before filtering as described in Section 3.3. We apply the continuous-time filtering approach and compare the two predicted ratios dynamics. This comparison highlights the effect of the random re-distribution of jumps.

Figures 8 and 9 show the dynamics of the proportion of transitions predicted against the real observed ratios from rating A to rating B, respectively without and with redistribution.

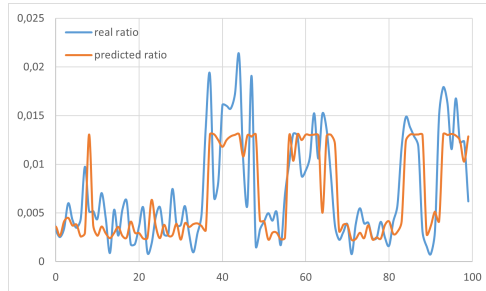


Figure 8: Real and predicted ratios from A to B without redistribution of common jumps

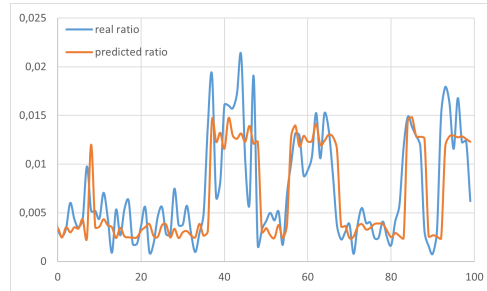


Figure 9: Real and predicted ratios from A to B with redistribution of common jumps

The predicted ratios dynamics in Figure 8, validate the use of the continuous-time filtering approach: the predicted ratios follow the real trajectory of ratios.

By comparing with Figure 9, we deduce that spreading information (to avoid the simultaneity of jumps) does not alter the predictions. Thanks to this comparison exercise, we can apply continuous framework to real data without concern that the results are altered by this action.

Even if the data samples used are different, we can notice that the changes in both predicted ratios dynamics are less brutal than in the discrete-time filtering framework applied in Section 4.1. The continuous filter is updated with progressive information (due to the absence of simultaneous jumps) and is more flexible than the discrete filter to anticipate regime changes. Assimilating jumps one by one, seems to improve the quality of predictions. Nevertheless, the effect of delay is still observable.

## 5 Application on real data

This section compares the results of our different models on a real rating database. We consider two discrete-time versions of the filter (one univariate and one multivariate) and a continuous-time multivariate alternative approach.

### 5.1 Data Description

Credit ratings are forward-looking opinions about the creditworthiness of an obligor with respect to a specific financial obligation. We build a transitions rating database from Moody's credit rating disclosure. We only use aggregated data (number of transitions). The considered sample contains 7791 days from January 2000 to May 2021. We study the evolution of Long Term ratings of 5030 corporate entities during this period without sector consideration. For specific experiments (analyses, validation, comparison), we consider the whole sample to calibrate the models. For others, such as testing the predictive power of model, we proceed to a cross validation. We choose a 5 states hidden factor for each experiment.

Moody's rating system relates 21 ratings categories. Keeping this granularity means estimating more than 420 transitions. Therefore, many studies ([16], [29]) reduce the number of rating

categories. In the same way, we decide to aggregate the 22 ratings to 6 : A, Baa, Ba, B, C and W. An obligation is rated W when it has no rating. We will also rate W the entity whose rating is not observed. This happens when the data is missing, censored or when it is not appeared yet. There exists many ways to manage not rated status (W). It can be considered as bad information, good information, no information for the credit or not considering them at all. According to [8], only few (roughly 13 percent) of the migration to the not rated category are related to changes in credit quality. This argument motivated [35] to use the last method, consisting in removing from the sample all the entities that experiences a not rated status. But this approach is dubious in regard of the loss of information. In this study, we will consider no rated status as censorship. This is achieved by progressively eliminating companies whose rating is not known or withdrawn and adding them when a new rating is provided.

A reference time-step is chosen for each experiment. The daily data are aggregated in order to observe and to predict rating transitions on a larger time window.

## 5.2 Discrete-time filtering in sample

In order to observe and interpret the effect of the discrete-time framework on a real credit rating database, we present in this section, the main results of a univariate and a multivariate filters, calibrated on the whole period.

### 5.2.1 Univariate discrete-time filtering

In this part, we assume that each transition is governed by its own hidden factor. Under this assumption, each transition evolves according to the evolution of its own latent factor, independently from the others. This modeling is meaningful to integrate rating specificities in the predictions.

On the data set described above, we focus on a single transition: from rating B to C. We choose this transition because it could be identified as “transition to default” and witness of crisis. This will entail the use and calibration of the univariate form of the discrete-time filter (19).

A first step consists in calibrating the models with the past history of the involved transition. The reference time step, at stake in every transition, is 30 days. We highlight the efficiency of our approach without cross validation: all past transition history available (from January 2000 to May 2021) is used to calibrate the model.

We obtain in Table 1, the calibrated 30 days transition matrix of the hidden factor  $\Theta$ . Table 2 presents the conditional transition probabilities from rating B to C in each state.

Table 1:  $\Theta$ 's transition matrix

	$\Theta = 0$	$\Theta = 1$	$\Theta = 2$	$\Theta = 3$	$\Theta = 4$
$\Theta = 0$	0.90598	0.074109	0.018316	0	0.001595
$\Theta = 1$	0.230415	0.715919	0.040626	0	0.013040
$\Theta = 2$	0.000304	0.381375	0.540412	0.077909	0
$\Theta = 3$	0	0	0.740452	0.259548	0
$\Theta = 4$	0.491597	0	0	0.508403	0

Table 2: 30 days transition probabilities from B to C

	$\Theta = 0$	$\Theta = 1$	$\Theta = 2$	$\Theta = 3$	$\Theta = 4$
$B \rightarrow C$	0.001814	0.0050001	0.0158818	0.0451715	0.085771

Table 1 highlights two stable states, 0 and 1 and an unstable and rare state, state 4. By analysing Table 2, we notice a hierarchy of risk between the states of  $\Theta$ . State 4 is clearly

identified as the riskiest state with a downgrade probability fifty time greater than in state 0, the most favourable state. State 3 is also a state of crisis which is more stable. State 2 can be interpreted as an intermediate state between favourable and unfavourable situation. Consequently we can expect that the economy often remains in a calm and favourable situation and experiences sometimes brief transitions to stressed states when downgrade probability B to C increases a lot. Figure 10 presents the filtered indicator function trajectories of the own hidden factor of the transition B to C, without cross validation. Figure 11 shows the dynamics of 30 days forecasted ratios from rating B to C, according to (19).

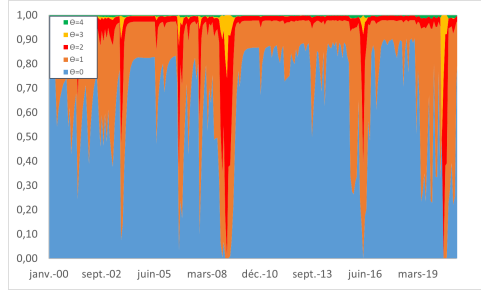


Figure 10: Filtered trajectories of the hidden factor indicator functions

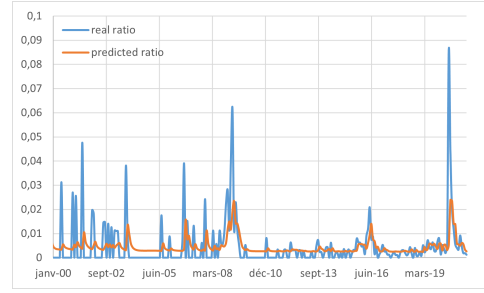


Figure 11: Real and predicted ratios for transition B to C

Figure 10 shows that the dominant state changes across time and highlights regime switching. Our intuitions are confirmed, the filter is often "closed" to favourable states 0 and 1. The dominant state is sometimes, for a brief moment, state 2, an intermediate state, where the downgrade probability from B to C increases. After periods when state 2 is dominant, the filter sometimes indicates that a state of true crisis, state 3, becomes dominant. Transitions from periods where state 0 or 1 are dominant to periods where state 4 is dominant may be sudden but remain rare. Fortunately this state of extreme "crisis" is only dominant for very brief periods. By analysing Figure 11, it can be noted that the predicted ratios from B to C reflect the general trend of real ratios with the same "lag" effect observed than on fictive data. The filter is able to detect regimes and transition phases but cannot capture brutal and short transitions. Finally the filter infers that the economic cycle experiences long periods of favourable situations and brief transitions to stress states.

Note that the hidden factor is specific to the involved transition. It may cover systemic risk but also the risk which might be specific to the ratings at stake.

We now consider the multivariate case where the hidden factor is shared by several transitions.

### 5.2.2 Multivariate discrete-time filtering

Using multiple transitions to infer the hidden factor assumes that the later is shared by those transitions. This approach should bring more information to forecast the dynamics of these transitions but presents several difficulties. The calibration algorithm finds centroids in the parameters space which might be far from each other due to the high dimension of the parameters space. Consequently the predicted number of transitions may be very different from the realized one. Furthermore rating transition events may not be sufficiently correlated. Indeed certain transitions are weakly correlated and might bring noise. We must only consider the most correlated transitions to extract the global factor dynamics. Therefore we decide to only focus on adjacent downgrade transitions (the upper diagonal). Indeed empirical results from [10] show that the upgrades are more subject to idiosyncratic shocks than downgrades. To remove the impact of the remaining transitions on the model, we assign them the same probability for each state of the hidden factor: we use the time-homogeneous intensity estimators to compute these probabilities (see, e.g., [10], [13], [26], [29], [30]). Consequently we reduce the number of transitions to calibrate to four.

We achieve two experiments. First we consider a time step reference of 30 days. We calibrate on whole period of the data set to observe the behaviour of the multivariate model. Then, along a second experiment, we will proceed to a cross validation to faithfully assess the predictive power of the model. For this experiment which is computationally more expensive, we will choose a larger time window, with a time step of 50 days.

For the first experiment, as in Section 5.2.1, we again consider 5 states for the hidden factor, a time step of 30 days and we do not proceed to cross validation.

Table 3 gives the calibrated transition matrix of the hidden factor. Table 4 presents the conditional downgrade probabilities for a time step of 30 days.

Table 3:  $\Theta$ 's transition matrix

	$\Theta = 0$	$\Theta = 1$	$\Theta = 2$	$\Theta = 3$	$\Theta = 4$
$\Theta = 0$	0.9499	0.0418	0.0010	0	0.0073
$\Theta = 1$	0.1075	0.7661	0.1264	0	0
$\Theta = 2$	0.0004	0.2685	0.6340	0.0503	0.0469
$\Theta = 3$	0	0	0.5133	0.4867	0
$\Theta = 4$	0	0	1	0	0

Table 4: Adjacent 30 days downgrade probabilities

	$\Theta = 0$	$\Theta = 1$	$\Theta = 2$	$\Theta = 3$	$\Theta = 4$
$A \rightarrow Baa$	0.00297589	0.00224944	0.00838262	0.00801885	0.0194804
$Baa \rightarrow Ba$	0.00125687	0.00146192	0.00492593	0.00985172	0.031583
$Ba \rightarrow B$	0.00326413	0.00633207	0.0150595	0.0282736	0.0228716
$B \rightarrow C$	0.00189228	0.00492691	0.0128149	0.0641203	0.0114155

By analysing the tables, it is noteworthy that states 0 and 1 are stable states which induce a “favourable” situation, where downgrade probabilities are quite low. States 3 and 4 can be interpreted as a stressed economy, where downgrade probabilities are higher. Note that state 4 is totally unstable and transitory. The transition between favourable periods (state 0 and 1) and stable stressed periods (state 3) is exclusively achieved through state 2.

Figure 12 shows the filtered trajectories of state probabilities according to (26). Figure 13 presents the dynamics of the predicted ratio from rating B to C, within a multivariate framework, without cross validation. We focus on transition B to C to compare with Section 5.2.1.

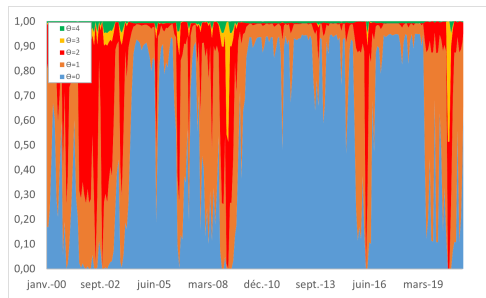


Figure 12: Filtered trajectories of hidden factor indicator functions

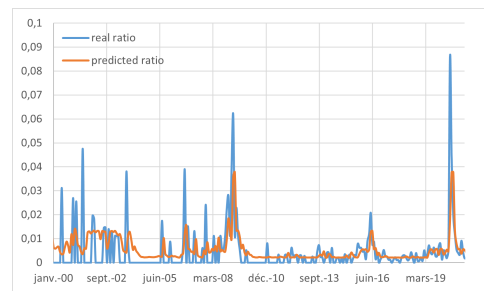


Figure 13: Real and predicted ratios for 30 days transition from B to C

Figure 12 brings us new information on the evolution of the predicted hidden state. Periods of crisis when state 3 and 4 dominant, are pretty rare and brief. By analyzing Figure 13, we can first notice that the multivariate framework is also a good predictor. The forecasted transition ratios

follow the trend of observed ratios and fit with different regimes. Comparing with the univariate case (see Figure 10), the multivariate model seems to be more sensitive to events: the multivariate model better captures the crisis of sep-2000 compared to the univariate model. The forecasted rating transition B to C is not only based on its own past evolution but also stem from the history of others.

### 5.3 Comparison of the filters out of sample: annual recalibration

We use a cross-validation approach to assess the predictive power of the multivariate models both in the continuous-time and discrete-time frameworks. To this end, we use data from 2000 to 2008 to perform a first calibration and to initialize our parameters. Then, from January 2008 to may 2021, we predict the dynamics of the 50 days transition rates. The model is re-calibrated every year, integrating the new observations of the last year. Note that we changed the reference time step to 50 days for a sake of computational speed.

Note also that since we re-calibrate the model yearly, parameters and states structure vary over time.

#### 5.3.1 Multivariate continuous-time filtering

In this section, we apply the continuous filtering framework, presented in Section 3.2 and its adaptations, described in Section 3.3, to real data. We choose a reference time step equal to 50 days. The real and predicted 50 days rating transition ratios are presented in Figures 14, 15, 16 and 17.

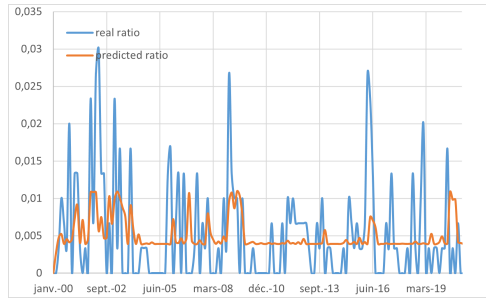


Figure 14: Real and predicted ratios 50 days transition from A to Baa

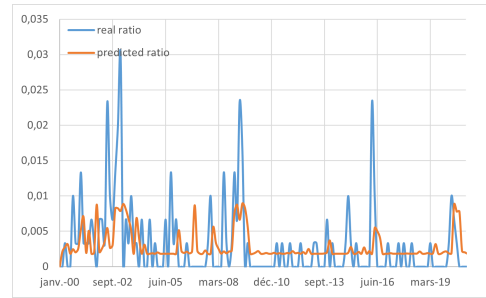


Figure 15: Real and predicted ratios 50 days transition from Baa to Ba

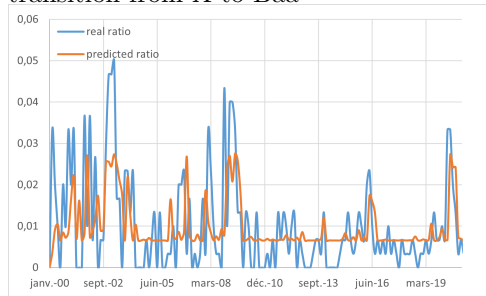


Figure 16: Real and predicted ratios 50 days transition from Ba to B

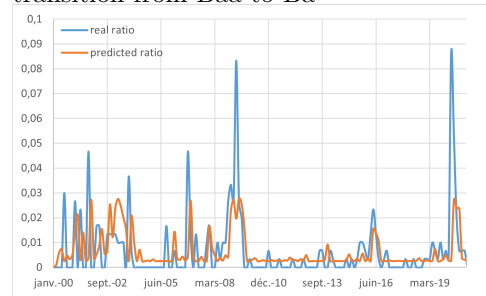


Figure 17: Real and predicted ratios 50 days transition from B to C



### 5.3.2 Multivariate discrete-time filtering

Here, we apply the discrete-time filtering framework, presented in Section 2.3 to real data. The model is applied on the same sample used for the continuous-time filtering approach, with annual recalibration as in Section 5.3.1. We keep a reference time step equal to 50 days. Figures 18, 19, 20, 21 compare the dynamics of predicted transition ratios to observed one.

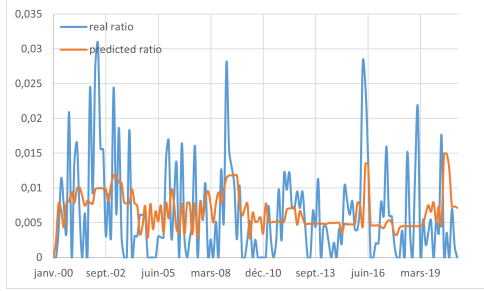


Figure 18: Real and predicted 50 days-transition ratios from A to Baa

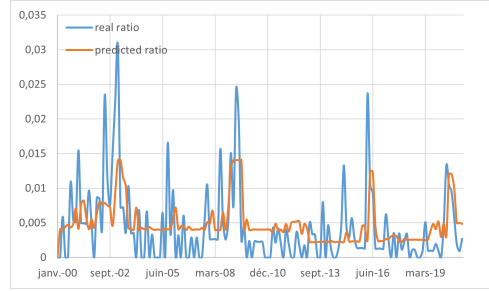


Figure 19: Real and predicted 50 days-transition ratios from Baa to Ba

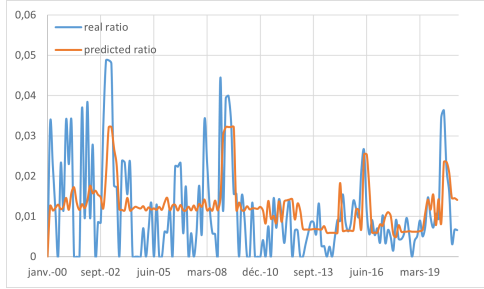


Figure 20: Real and predicted 50 days-transition ratios from Ba to B

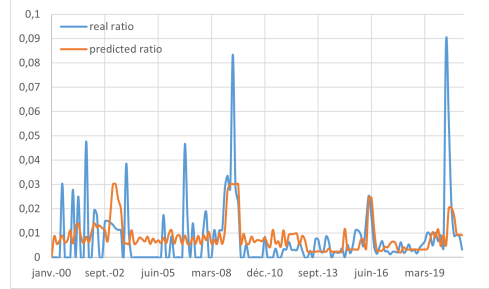


Figure 21: Real and predicted 50 days-transition ratios from B to C

### 5.3.3 Comparisons and analyses

The results looks almost similar in both approaches. The dynamics of predicted ratios follow the trend of realized ratios. The forecasts also evolve when noteworthy crisis occurs. We notice that transitions are more correlated during specific periods like crisis. Four crisis periods can be identified: a first small one around 2002, a moderated one in 2016 and two significant in 2008 and 2020. These latter are clearly identified as the subprime crisis and the health crisis caused by the Covid 19. The two others, moderated, would be respectively the consequences of the dot-com bubble in 2000 and the China stock market crash in June, 2015. During these periods the downgrades probabilities increase.

Both models are able to detect the evolution of the economic cycle from observations of rating migrations. The forecasts are adapted to the inferred economic state. During crisis periods, the models are able to predict adapted and higher downgrade probabilities.

We can underline three advantages of the continuous-time version compared to the discrete-time one.

- The effect of delay (or lag effect) is less significant than in the discrete-time framework. By spreading simultaneous jumps in small time intervals, we make last information used for filtering fresher than it actually is. This fictive operation, however, improves the

predictions.

- We note that the discrete-time model struggles to capture brief and brutal variations. As we observed in Section 4.1.1, continuous-time filtering approach has the advantage of assimilating jumps one by one and of being more flexible and suitable to anticipate sudden transitions. Since information is spread and distributed in fictive intervals, the filter progressively assimilates information and is therefore quicker to adapt its predictions.

Nevertheless we can see that this framework is not fully adapted to rating migrations. The discrete version is easier and faster to compute: manipulations described in Section 3.3 increases considerably the number of time intervals to consider, the complexity and remains laborious. Above all, the discrete model is more consistent with the data and finally, provides predictions of a better quality.

The continuous approach should be improved by using a true continuous EM algorithm. The effectiveness of this method would rather be highlighted by filtering a continuous phenomenon, where observations cannot occur simultaneously and exact occurrence dates are known.

This intuition is confirmed by the following experiment. We compute the  $R^2$  coefficient in the sample, to compare the forecasting power of the considered predictive models. We keep a reference time step of 50 days. We respectively compare the  $R^2$  of the constant generator intensity model, the univariate models, and the multivariate discrete models and the continuous model in Table 5.

Table 5:  $R^2$  in the sample

	$A \rightarrow Baa$	$Baa \rightarrow Ba$	$Ba \rightarrow B$	$B \rightarrow C$
Constant	0.463012	0.250668	0.483855	0.250514
Univ.Discrete	0.431184	0.346886	0.608684	0.22263515
Mult.Discrete	0.494395	0.479324	0.644975	0.367094
Mult.Continuous	0.2022	0.331062	0.49396	0.279736

We can directly notice that the multivariate discrete-time filter is the most accurate whatever the transition. The  $R^2$  of the continuous filter is closed to discrete filter for transitions Baa to Ba and B to C but is lower for the transitions A to Ba, even lower than the  $R^2$  from the constant generator model. This phenomenon can be explained by the poorer calibration achieved for the continuous-time filtering (for a sake of rapidity) and the inconsistency with the format of the data. The adapted continuous-time version can be applied to rating transitions framework and provides satisfactory predictions but can not reach the performance of the discrete-time version. Note that, in the univariate case, each transition has its specific model.

## 6 Conclusion

In this paper, we propose a discrete-time counting process filtering version, usually studied in continuous time. Because discrete time data are not compliant with the continuous time framework, we also propose several adaptations of the continuous time approach. The two alternative point-process filtering models are studied and compared in the context of credit rating migrations. For both approaches, we assume that rating transitions are driven by the same systemic hidden factor. We discussed calibration issues and compared the predicted future rating transition probabilities on fictive and real data.

As illustrated in Sections 4 and 5, our methodology provides predictors adapted to the evolution of the economical cycle. We believe that our approaches can be used for PIT-estimations of transitions and detection of regimes. During crisis periods, our models are able to predict adapted and higher downgrade probabilities. Compared to other PIT-estimation models, our approaches base their predictions on the business cycle without concern of macro economic factors. From a practical point of view, our approaches also have the advantage to be interpretable. Observing the

risk profile of each state and their filtered trajectories allows us to better understand the dynamics of the economic cycle as well as its systemic effect on rating migrations.

However, both approaches cannot capture idiosyncratic information: indeed [41] found that only 18% to 26% of global default risk variation is systematic while the remainder is idiosyncratic. The share of systematic default risk is higher (39% to 51%) if industry-specific variation is counted as systematic.

Applying the continuous framework to discrete time data is tedious and presents a risk of altering information and the quality of predictions. Since its complexity is much more important, the continuous-time algorithm is very slow to apply. Therefore, it suffers from a poorer calibration compared to the discrete-time version. Thanks to the adaptations realized, we saw in Section 5.3 that the continuous-time version is able to provide satisfactory predictions but can not reach the performances of the discrete-time model. For these reasons, the discrete-time approach turns out to be more adapted and efficient for the context of rating migrations.

Several improvements could be made to our framework, such as, using a continuous-time version of the Baum-Welch algorithm for the continuous-time model (as, e.g., in [33] and [36]) or considering additional idiosyncratic observable factors as in [29]. These considerations are left for future research.

## Acknowledgments

The authors are immensely grateful to Ragnar Norberg who significantly contributes to early stages of this work. We also want to acknowledge Baye Matar Kandji who greatly improves the Baum-Welch type calibration algorithm. This work was carried out as part of a PhD thesis ‘CIFRE’. The authors thank Nexialog Consulting for its support.

## References

- [1] Edward I Altman and Duen Li Kao. The implications of corporate bond ratings drift. *Financial Analysts Journal*, 48(3):64–75, 1992.
- [2] European Banking Authority. Guidelines on pd estimation, lgd estimation and the treatment of defaulted exposures. 2017.
- [3] Anil Bangia, Francis X Diebold, André Kronimus, Christian Schagen, and Til Schuermann. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of banking & finance*, 26(2-3):445–474, 2002.
- [4] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.
- [5] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [6] Pierre Brémaud. *Point processes and queues: martingale dynamics*, volume 50. Springer, 1981.
- [7] Anisa Caja, Quentin Guibert, and Frédéric Planchet. Influence of economic factors on the credit rating transitions and defaults of credit insurance business. 2015.

- [8] Lea Carty. Moody’s rating migration and credit quality correlation. *Moody’s Sepcial Report July*, 1997.
- [9] Wai-Ki Ching, Tak Kuen Siu, Li-min Li, Tang Li, and Wai-Keung Li. Modeling default data via an interactive hidden markov model. *Computational Economics*, 34(1):1–19, 2009.
- [10] A. Cousin and M. R. Kheliouen. A comparative study on the estimation of factor migration models. *Bulletin français d’actuariat*, 2015.
- [11] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [12] Servaas Deroose, Werner Roeger, and Sven Langedijk. Reviewing adjustment dynamics in emu: from overheating to overcooling. *European Economy Economic Paper*, (198), 2008.
- [13] Darrell Duffie, Leandro Saita, and Ke Wang. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3):635–665, 2007.
- [14] European Banking Authorities EBA. Guidelines on institutions’ stress testing. 2018.
- [15] Robert J Elliott, Lakhdar Aggoun, and John B Moore. *Hidden Markov models: estimation and control*, volume 29. Springer Science & Business Media, 2008.
- [16] Robert J Elliott, Tak Kuen Siu, and Eric S Fung. A double hmm approach to altman z-scores and credit ratings. *Expert Systems with Applications*, 41(4):1553–1560, 2014.
- [17] Dingan Feng, Christian Gouriéroux, and Joann Jasiak. The ordered qualitative model for credit rating transitions. *Journal of Empirical Finance*, 15(1):111–130, 2008.
- [18] Stephen Figlewski, Halina Frydman, and Weijian Liang. Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1):87–105, 2012.
- [19] Peter Fledelius, David Lando, and Jens Perch Nielsen. Non-parametric analysis of rating transition and default data. *Journal of Investment Management*, 2(2), 2004.
- [20] Bank for International Settlements. Basel committee on banking supervision basel iii: Finalising post-crisis reforms. *Official Journal of the European Union*, 2017.
- [21] Rüdiger Frey and Thorsten Schmidt. Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering. *Finance and Stochastics*, 16(1):105–133, 2012.
- [22] Patrick Gagliardini and Christian Gouriéroux. Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics*, 3(2):188–226, 2005.
- [23] Giacomo Giampieri, Mark Davis, and Martin Crowder. Analysis of default data using hidden markov models. *Quantitative Finance*, 5(1):27–34, 2005.
- [24] Kay Giesecke and Gustavo Schwenkler. Filtered likelihood for point processes. *Journal of Econometrics*, 204(1):33–53, 2018.
- [25] David T Hamilton. Rating transitions and defaults conditional on watchlist, outlook and rating history. *Outlook and Rating History (February 2004)*, 2004.
- [26] Robert A Jarrow, David Lando, and Stuart M Turnbull. A markov model for the term structure of credit risk spreads. *The review of financial studies*, 10(2):481–523, 1997.
- [27] Alan Karr. *Point processes and their statistical inference*. Routledge, 2017.

- [28] Dimitrios Kavvathas. Estimating credit rating transition probabilities for corporate bonds. 2001.
- [29] Siem Jan Koopman, André Lucas, and André Monteiro. The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142(1):399–424, 2008.
- [30] David Lando and Torben M Skødeberg. Analyzing rating transitions and rating drift with continuous observations. *Journal of banking & finance*, 26(2-3):423–444, 2002.
- [31] Vincent Leijdekker and Peter Spreij. Explicit computations for a filtering problem with point process observations with applications to credit risk. *Probability in the Engineering and Informational Sciences*, 25(3):393–418, 2011.
- [32] Tingting Liu, Jan Lemeire, and Lixin Yang. Proper initialization of hidden markov models for industrial applications. In *2014 IEEE China summit & international conference on signal and information processing (ChinaSIP)*, pages 490–494. IEEE, 2014.
- [33] Yu-Ying Liu, Alexander Moreno, Shuang Li, Fuxin Li, Le Song, and James M Rehg. Learning continuous-time hidden markov models for event data. In *Mobile Health*, pages 361–387. Springer, 2017.
- [34] Robert C Merton. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2):449–470, 1974.
- [35] Pamela Nickell, William Perraudin, and Simone Varotto. Stability of rating transitions. *Journal of Banking & Finance*, 24(1-2):203–227, 2000.
- [36] Uri Nodelman, Christian R Shelton, and Daphne Koller. Expectation maximization and complex duration distributions for continuous time bayesian networks. *arXiv preprint arXiv:1207.1402*, 2012.
- [37] Sung Youl Oh, Jae Wook Song, Woojin Chang, and Minhyuk Lee. Estimation and forecasting of sovereign credit rating migration based on regime switching markov chain. *IEEE Access*, 7:115317–115330, 2019.
- [38] Emre Özkan, Fredrik Lindsten, Carsten Fritsche, and Fredrik Gustafsson. Recursive maximum likelihood identification of jump markov nonlinear systems. *IEEE Transactions on Signal Processing*, 63(3):754–765, 2014.
- [39] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [40] European Commission Regulation. Amending regulation (ec) no 1126/2008 adopting certain international accounting standards in accordance with regulation (ec) no 1606/2002 of the european parliament and of the council as regards international financial reporting standard 9. *Official Journal of the European Union*, 2016.
- [41] Bernd Schwaab, Siem Jan Koopman, and André Lucas. Global credit risk: World, country and industry factors. *Journal of Applied Econometrics*, 32(2):296–317, 2017.
- [42] Anton Tenyakov. Estimation of hidden markov models and their applications in finance. 2014.
- [43] Lyn C Thomas, David E Allen, and Nigel Morkel-Kingsbury. A hidden markov chain model for the term structure of bond credit risk spreads. *International Review of Financial Analysis*, 11(3):311–329, 2002.
- [44] James Tobin. Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, pages 24–36, 1958.
- [45] JH Van Schuppen. Filtering, prediction and smoothing for counting process observations, a martingale approach. *SIAM Journal on Applied Mathematics*, 32(3):552–570, 1977.

## A Calibration of the discrete Version

This part describes the derivation for the Baum-Welch algorithm adaptation presented in Section 2.4.

We compute  $\forall s \in \{1, \dots, m\}, \forall n \in \{1, \dots, \Gamma\}$ , the forward probability  $\alpha_n(s) = \mathbb{P}(Z_{0|n} = z_{0|n}, \Theta_{n-1} = s)$  and  $\forall s \in \{1, \dots, m\}, \forall n \in \{1, \dots, \Gamma - 1\}$ , the backward probability  $\beta_n(s) = \mathbb{P}(Z_{n+1|\Gamma} = z_{n+1|\Gamma} | Z_n = z_n, \Theta_{n-1} = s)$ .

We derive  $\forall n \in \{2, \dots, \Gamma - 1\}$  and  $\forall s \in \{1, \dots, m\}$ ,

$$\begin{aligned}
\alpha_n(s) &= \mathbb{P}(Z_{0|n} = z_{0|n}, \Theta_{n-1} = s) = \sum_{l=1}^m \mathbb{P}(Z_{0|n} = z_{0|n}, \Theta_{n-1} = s, \Theta_{n-2} = l) \\
&= \sum_{l=1}^m \prod_{d=1}^Q \mathbb{P}(Z_n^d = z_n^d | Z_{0|n-1} = z_{0|n-1}, \Theta_{n-1} = s, \Theta_{n-2} = l) \mathbb{P}(Z_{0|n-1} = z_{0|n-1}, \Theta_{n-1} = s, \Theta_{n-2} = l) \\
&= \sum_{l=1}^m \prod_{d=1}^Q \mathbb{P}(Z_n^d = z_n^d | Z_{n-1} = z_{n-1}, \Theta_{n-1} = s) \mathbb{P}(\Theta_{n-1} = s | \Theta_{n-2} = l, Z_{0|n-1} = z_{0|n-1}) \alpha_{n-1}(l) \\
&= \sum_{l=1}^m \prod_{d=1}^Q \mathbb{P}(Z_n^d = z_n^d | Z_{n-1}^d = z_{n-1}^d, \Theta_{n-1} = s) \mathbb{P}(\Theta_{n-1} = s | \Theta_{n-2} = l) \alpha_{n-1}(l) \\
&= \sum_{l=1}^m \alpha_{n-1}(l) K^{ls} \prod_{d=1}^Q L^{s, z_{n-1}^d z_n^d} = \sum_{l=1}^m \alpha_{n-1}(l) K^{ls} \prod_{i,r \in \Upsilon} (L^{s, ir})^{\Delta N_n^{ir}}.
\end{aligned}$$

For  $n=1$ , we do not know the state of the individuals before the simulation. To tackle this issue, we use the initial proportion of the ratings. We have

$$\begin{aligned}
\alpha_1(s) &= \mathbb{P}(Z_1 = z_1, \Theta_0 = s) \\
&= \prod_{d=1}^Q \mathbb{P}(Z_1^d = z_1^d | \Theta_0 = s) \Pi(s) \\
&= \prod_{d=1}^Q \sum_{j \in \Upsilon} \mathbb{P}(Z_1^d = z_1^d | Z_0^d = j, \Theta_0 = s) \mathbb{P}(Z_0^d = j | \Theta_0 = s) \Pi(s) \\
&= \prod_{d=1}^Q \sum_{j \in \Upsilon} \mathbb{P}(Z_1^d = z_1^d | Z_0^d = j, \Theta_0 = s) \mathbb{P}(Z_0^d = j) \Pi(s) \\
&= \prod_{d=1}^Q \sum_{j \in \Upsilon} L^{s, j z_1^d} \mathbb{P}(Z_0^d = j) \Pi(s).
\end{aligned}$$

Similarly, we recursively derive the backward probability for all  $n \in \{1, \dots, \Gamma - 2\}$  and  $s \in$

$\{1, \dots, m\}$ ,

$$\begin{aligned}
\beta_n(s) &= \mathbb{P}(Z_{n+1|\Gamma} = z_{n+1|\Gamma} | \Theta_{n-1} = s, Z_n = z_n) \\
&= \sum_{l=1}^m \mathbb{P}(Z_{n+1|\Gamma} = z_{n+1|\Gamma}, \Theta_n = l | \Theta_{n-1} = s, Z_n = z_n) \\
&= \sum_{l=1}^m \mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | \Theta_n = l, Z_{n+1} = z_{n+1}, Z_n = z_n, \Theta_{n-1} = s) \mathbb{P}(Z_{n+1} = z_{n+1}, \Theta_n = l | \Theta_{n-1} = s, Z_n = z_n) \\
&= \sum_{l=1}^m \mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | \Theta_n = l, Z_{n+1} = z_{n+1}) \prod_{d=1}^Q (\mathbb{P}(Z_{n+1}^d = z_{n+1}^d | \Theta_n = l, Z_n^d = z_n^d)) \mathbb{P}(\Theta_n = l | \Theta_{n-1} = s) \\
&= \sum_{l=1}^m \beta_{n+1}(l) K^{sl} \prod_{d=1}^Q L^{l, z_n^d z_{n+1}^d} = \sum_{l=1}^m \beta_{n+1}(l) K^{sl} \prod_{i,r \in \Upsilon} (L^{l,ir})^{\Delta N_n^{ir}}.
\end{aligned}$$

For  $n = \Gamma - 1$ , we take  $\forall s \in \{1, \dots, m\}$ ,  $\beta_{\Gamma-1}(s) = 1$ .

Both estimators will be used to replace the missing data during the maximization phase. The missing data describing the hidden factor are defined  $\forall h \in \{1, \dots, m\}, \forall n \in \{1, \dots, \Gamma\}$ ,  $u_n(h) = \mathbb{1}_{[\Theta_n = j]}$ , and  $v_n(s, h) = \mathbb{1}_{[\Theta_n = h, \Theta_{n-1} = s]}$ . We define the associated Bayesian estimators  $\check{u}_n(h) = \mathbb{P}(\Theta_n = h | Z_{0|\Gamma} = z_{0|\Gamma})$ , and  $\check{v}_n(s, h) = \mathbb{P}(\Theta_n = h, \Theta_{n-1} = s | Z_{0|\Gamma} = z_{0|\Gamma})$ .

We derive expression of these Bayesian estimators with the forward and the backward probabilities. For all  $h \in \{1, \dots, m\}$  and all  $n \in \{1, \dots, \Gamma - 2\}$ ,

$$\begin{aligned}
\check{u}_n(h) &= \mathbb{P}(\Theta_n = h | Z_{0|\Gamma} = z_{0|\Gamma}) \\
&= \frac{\mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | \Theta_n = h, Z_{0|n+1} = z_{0|n+1}) \alpha_{n+1}(h)}{\mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma})} \\
&= \frac{\mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | \Theta_n = h, Z_{n+1} = z_{n+1}) \alpha_{n+1}(h)}{L_\Gamma} \\
&= \frac{\beta_{n+1}(h) \alpha_{n+1}(h)}{L_\Gamma}.
\end{aligned}$$

With  $L_\Gamma$  being the likelihood on the whole sample,  $L_\Gamma = \mathbb{P}(Z_{0|\Gamma} = z_{0|\Gamma}) = \sum_j \alpha_\Gamma(j)$ .

$$\begin{aligned}
\check{v}_n(s, h) &= \mathbb{P}(\Theta_n = h, \Theta_{n-1} = s | Z_{0|\Gamma} = z_{0|\Gamma}) \\
&= \frac{\mathbb{P}(\Theta_n = h, Z_{n+1|\Gamma} = z_{n+1|\Gamma} | Z_{0|n} = z_{0|n}, \Theta_{n-1} = s) \alpha_n(s)}{L_\Gamma} \\
&= \frac{\mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | Z_{0|n+1} = z_{0|n+1}, \Theta_{n-1} = s, \Theta_n = h) \mathbb{P}(\Theta_n = h, Z_{n+1} = z_{n+1} | Z_{0|n} = z_{0|n}, \Theta_{n-1} = s) \alpha_n(s)}{L_\Gamma} \\
&= \frac{\mathbb{P}(Z_{n+2|\Gamma} = z_{n+2|\Gamma} | Z_{n+1} = z_{n+1}, \Theta_n = h) \prod_{d=1}^Q (\mathbb{P}(Z_{n+1}^d = z_{n+1}^d | \Theta_n = h, Z_n^d = z_n^d)) \mathbb{P}(\Theta_n = h | \Theta_{n-1} = s) \alpha_n(s)}{L_\Gamma} \\
&= \frac{\beta_{n+1}(h) K^{sh} \alpha_n(s) \prod_{d=1}^Q L^{h, z_n^d z_{n+1}^d}}{L_\Gamma} = \frac{\beta_{n+1}(h) K^{sh} \alpha_n(s) \prod_{i,r \in \Upsilon} (L^{h,ir})^{\Delta N_n^{ir}}}{L_\Gamma}.
\end{aligned}$$

Using the concavity of the log function, we establish a useful inequality for the next step of the derivations. For two strictly positive sequences  $w$  and  $w'$ ,

$$\begin{aligned} \log\left(\frac{\sum_i w'_i}{\sum_k w_k}\right) &= \log\left(\sum_i \frac{w_i w'_i}{\sum_k w_k w_i}\right) \\ &\geq \sum_i \frac{w_i}{\sum_k w_k} \log(w'_i) - \frac{w_i}{\sum_k w_k} \log(w_i) \\ &= \frac{1}{\sum_k w_k} \left(\sum_i (w_i \log(w'_i) - w_i \log(w_i))\right). \end{aligned}$$

The maximization step consists in finding better parameters than those of the previous iteration. We call  $M^{(\gamma)} = (\Pi^{(\gamma)}, L^{(\gamma)}, K^{(\gamma)})$ , the parameters of the current iteration ( $\gamma$ ).

We are seeking new parameters  $M^{(\gamma+1)} = (\Pi^{(\gamma+1)}, L^{(\gamma+1)}, K^{(\gamma+1)})$ .

Let consider the finite spaces  $\Xi_k = \{1, \dots, m\}^k, k \in \{1, \dots, \Gamma\}$ . We call a possible trajectory of  $\Theta, \theta$ , belonging to the finite space  $\Xi_\Gamma$ . Using  $w_\theta = \mathbb{P}(Z = z, \Theta = \theta | M^{(\gamma)})$  and  $w'_\theta = \mathbb{P}(Z = z, \Theta = \theta | M^{(\gamma+1)})$  in the previous inequality and defining  $Q(M^{(\gamma)}, M^{(\gamma+1)}) = \sum_{\theta \in \Xi_\Gamma} w_\theta \log(w'_\theta)$  and  $Q(M^{(\gamma)}, M^{(\gamma)}) = \sum_{\theta \in \Xi_\Gamma} w_\theta \log(w_\theta)$ , we obtain

$$\begin{aligned} \log\left(\frac{\sum_{\theta \in \Xi_\Gamma} w'_\theta}{\sum_{\theta \in \Xi_\Gamma} w_\theta}\right) &= \log\left(\frac{\mathbb{P}(Z = z | M^{(\gamma+1)})}{\mathbb{P}(Z = z | M^{(\gamma)})}\right) \\ &\geq \frac{1}{\mathbb{P}(Z = z | M^{(\gamma)})} (Q(M^{(\gamma)}, M^{(\gamma+1)}) - Q(M^{(\gamma)}, M^{(\gamma)})). \end{aligned}$$

This last inequality shows that we obtain  $\mathbb{P}(Z = z | M^{(\gamma+1)}) \geq \mathbb{P}(Z = z | M^{(\gamma)})$  by maximizing

$$Q(M^{(\gamma)}, M^{(\gamma+1)}) = \sum_{\theta \in \Xi_\Gamma} \mathbb{P}(\Theta = \theta, Z = z | M^{(\gamma)}) \log \mathbb{P}(\Theta = \theta, Z = z | M^{(\gamma+1)}).$$

We cut  $\log(\mathbb{P}(Z = z, \Theta = \theta | M^{(\gamma+1)})) = \log(\mathbb{P}(Z = z | \Theta = \theta, M^{(\gamma+1)})) + \log(\mathbb{P}(\Theta = \theta | M^{(\gamma+1)}))$ . Since the processes  $(Z^d)_d$  are independent knowing the unobserved factor, we have

$$\begin{aligned} \log(\mathbb{P}(Z = z, \Theta = \theta | M^{(\gamma+1)})) &= \log(\mathbb{P}(\theta_0)) + \sum_{n=1}^{\Gamma} \log(\mathbb{P}(\theta_n | \theta_{n-1})) + \sum_{d=1}^Q \log(\mathbb{P}(z_{0\Gamma}^d | \theta)) \\ &= \log(\mathbb{P}(\theta_0)) + \sum_{n=1}^{\Gamma} \log(\mathbb{P}(\theta_n | \theta_{n-1})) + \sum_{d=1}^Q \sum_{n=1}^{\Gamma} \log \mathbb{P}(z_n^d | z_{n-1}^d, \theta_{n-1}). \end{aligned}$$



So,

$$\begin{aligned}
Q(M^{(\gamma)}, M^{(\gamma+1)}) &= \sum_{\theta \in \Xi_\Gamma} \log(\mathbb{P}(\theta_0)) \mathbb{P}(\theta, z | M^{(\gamma)}) + \sum_{\theta \in \Xi_\Gamma} \sum_{n=1}^{\Gamma} \log(\mathbb{P}(\theta_n | \theta_{n-1})) \mathbb{P}(\theta, z | M^{(\gamma)}) \\
&+ \sum_{\theta \in \Xi_\Gamma} \sum_{d=1}^Q \sum_{n=1}^{\Gamma} \log \mathbb{P}(z_n^d | z_{n-1}^d, \theta_{n-1}) \mathbb{P}(\theta, z | M^{(\gamma)}) \\
&= \sum_{\substack{\theta \setminus \theta_0 \\ \in \Xi_{\Gamma-1}}} \sum_{h=1}^m \log(\mathbb{P}(\Theta_0 = h)) \mathbb{P}(\Theta_0 = h, \theta \setminus \theta_0, z | M^{(\gamma)}) \\
&+ \sum_{n=1}^{\Gamma} \sum_{\theta \setminus (\theta_n, \theta_{n-1})} \sum_{h, s=1}^m \log(\mathbb{P}(\Theta_n = h | \Theta_{n-1} = s)) \mathbb{P}(\theta \setminus (\theta_n, \theta_{n-1}), \Theta_n = h, \Theta_{n-1} = s, z | M^{(\gamma)}) \\
&+ \sum_{n=1}^{\Gamma} \sum_{\theta \setminus \theta_{n-1}} \sum_{s=1}^m \sum_{d=1}^Q \log(\mathbb{P}(z_n^d | z_{n-1}^d, \Theta_{n-1} = s)) \mathbb{P}(\theta \setminus \theta_{n-1}, \Theta_{n-1} = s, z | M^{(\gamma)}) \\
&= \sum_{h=1}^m \log(\mathbb{P}(\Theta_0 = h, z | M^{(\gamma)})) \Pi^h + \sum_{n=1}^{\Gamma} \sum_{h, s=1}^m \log(K^{sh}) \mathbb{P}(\Theta_n = h, \Theta_{n-1} = s, z | M^{(\gamma)}) \\
&+ \sum_{d=1}^Q \sum_{n=1}^{\Gamma} \sum_{s=1}^m \sum_{i, r \in \Upsilon} \log(L^{h, ir}) \mathbb{P}(\Theta_{n-1} = s, z | M^{(\gamma)}) \mathbb{1}_{[Z_n^d=r, Z_{n-1}^d=i]}.
\end{aligned}$$

Then, we can maximize by considering the three terms independently. We obtain

$$\Pi^h = \frac{\mathbb{P}(\Theta_0 = h, Z | M^{(\gamma)})}{\sum_{j=1}^m \mathbb{P}(\Theta_0 = j, Z | M^{(\gamma)})} = \mathbb{P}(\Theta_1 = h | Z, M^{(\gamma)}) = \check{u}_0(h),$$

$$L^{s, ir} = \frac{\sum_{d=1}^Q \sum_{n=1}^{\Gamma} \mathbb{P}(\Theta_{n-1} = s, Z | M^{(\gamma)}) \mathbb{1}_{[Z_n^d=r, Z_{n-1}^d=i]}}{\sum_{d=1}^Q \sum_{n=1}^{\Gamma} \sum_{j \in \Upsilon} \mathbb{P}(\Theta_{n-1} = s, Z | M^{(\gamma)}) \mathbb{1}_{[Z_n^d=j, Z_{n-1}^d=i]}} = \frac{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s) \Delta N_n^{ir}}{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s) Y_n^i},$$

$$K^{sh} = \frac{\sum_{n=1}^{\Gamma} \mathbb{P}(\Theta_{n-1} = s, \Theta_n = h, Z | M^{(\gamma)})}{\sum_{n=1}^{\Gamma} \sum_{j=1}^m \mathbb{P}(\Theta_{n-1} = s, \Theta_n = j, Z | M^{(\gamma)})} = \frac{\sum_{n=1}^{\Gamma} \check{v}_n(s, h)}{\sum_{n=1}^{\Gamma} \check{u}_{n-1}(s)}.$$

## B Continuous-time version of the filter

### B.1 General case

*Proof (of Prop. 8).* Let  $g$  and  $h$  be two  $\mathbf{F}$  predictable processes such that  $\mathbb{E} \left[ \int_0^T (g_s^2 + h_s^2) \nu_s^j ds \right] \leq \infty$ . We introduce the processes  $X$  and  $Y$  defined by  $X_t = \int_0^t g_r (dN_r^j - \nu_r^j dr)$  and  $Y_t = \int_0^t h_s (dN_s^k - \nu_s^k ds)$  for all  $t \leq T$ .  $X$  and  $Y$  are two  $\mathbf{F}$ -martingales. The Itô formula applied to  $XY$  yields

$$d(X_t Y_t) = X_{t-} dY_t + Y_{t-} dX_t + \Delta X_t \Delta Y_t.$$

Since  $N^j$  and  $N^k$  have no common jumps,  $\Delta X_t \Delta Y_t = g_t h_t \Delta N_t^j \delta_{jk}$ . Then, we obtain

$$\mathbb{E}[X_T Y_T - X_t Y_t | \mathcal{F}_t] = \delta_{jk} \mathbb{E} \left[ \int_t^T g_s h_s dN_s^j | \mathcal{F}_t \right]. \quad (57)$$

Note that  $X_T Y_T - X_t Y_t = (X_T - X_t)(Y_T - Y_t) - 2X_t Y_t + X_t Y_T + X_T Y_t$ . Then,

$$\mathbb{E}[X_t Y_T - X_t Y_t | \mathcal{F}_t] = \mathbb{E}[(X_T - X_t)(Y_T - Y_t) | \mathcal{F}_t]. \quad (58)$$

Note that the process  $Z = \left( \int_0^t g_s h_s (dN_s^j - \nu_s^j ds) \right)_{0 \leq t \leq T}$  is a  $\mathbf{F}$ -martingale.

So,  $\mathbb{E}[Z_T - Z_t | \mathcal{F}_t] = 0$ . Combining this remark with Equations (57) and (58), we finally obtain

$$\mathbb{E} \left[ \int_t^T g_r (dN_r^j - \nu_r^j dr) \int_t^T h_s (dN_s^k - \nu_s^k ds) \middle| \mathcal{F}_t \right] = \delta_{jk} \mathbb{E} \left[ \int_t^T g_s h_s \nu_s^j ds \middle| \mathcal{F}_t \right]. \quad (59)$$

The innovation theorem says that the  $\mathbf{F}^N$ -intensities of the counting processes  $N^j$  exist and are

$$\hat{\nu}_{s-}^j = \mathbb{E} \left[ \nu_{s-}^j | \mathcal{F}_{s-}^N \right] = \mathbb{E} \left[ \nu_s^j | \mathcal{F}_{s-}^N \right]. \quad (60)$$

For any  $\mathbf{F}^N$  predictable process  $h$  satisfying  $\mathbb{E} \left[ \int_0^T |h_s \nu_s^j| ds \right] < \infty$ , we have

$$\mathbb{E} \left[ \int_0^\infty h_s dN_s^j \right] = \mathbb{E} \left[ \int_0^\infty h_s \nu_s^j ds \right] = \mathbb{E} \left[ \int_0^\infty h_s \mathbb{E} \left[ \nu_s^j | \mathcal{F}_{s-}^N \right] ds \right] = \mathbb{E} \left[ \int_0^\infty h_s \hat{\nu}_{s-}^j ds \right].$$

Now, rewrite (42) as

$$\Theta_t = \int_0^t \hat{a}_s ds + L_t + M_t, \quad (61)$$

with

$$L_t = \int_0^t (a_s - \hat{a}_s) ds. \quad (62)$$

Taking conditional expectation, w.r.t.  $\mathcal{F}_t^N$ , in (61) yields

$$\hat{\Theta}_t = \int_0^t \hat{a}_s ds + \hat{L}_t + \hat{M}_t. \quad (63)$$

While  $L$  need not be a  $\mathbf{F}$ -martingale, it is clear that  $\hat{L}$  is an  $\mathbf{F}^N$ -martingale. For  $t_1 < t_2$ ,

$$\mathbb{E}[\hat{L}_{t_2} - \hat{L}_{t_1} | \mathcal{F}_{t_1}^N] = \mathbb{E} \left[ \int_{t_1}^{t_2} (a_s - \mathbb{E}[a_s | \mathcal{F}_s^N]) ds \middle| \mathcal{F}_{t_1}^N \right] = 0.$$

From the tower property, we deduce that  $\hat{M}$  is also a  $\mathbf{F}^N$ -martingale. Introduce

$$K_t = L_t + M_t = \Theta_t - \int_0^t \hat{a}_s ds. \quad (64)$$

Since  $\Theta$  and  $N$  have no common jumps, we can deduce that  $K$  and  $N$  have any either. It has been shown that  $\hat{K} = \hat{L} + \hat{M}$  is a  $\mathbf{F}^N$ -martingale. Therefore it has a predictable representation,

$$\hat{K}_t = \gamma + \sum_j \int_0^t \eta_s^j (dN_s^j - \hat{\nu}_{s-}^j ds), \quad (65)$$

where  $\gamma = \hat{K}_0$  is  $\mathcal{F}_0^N$ -measurable and the  $\eta^j$  are  $\mathbf{F}^N$ -predictable processes (see [6]). Note that  $\hat{K}_0 = \mathbb{E}[\Theta_0]$ . Now, any integrable  $\mathbf{F}^N$ -measurable random variable has a representation  $g + \sum_j \int_0^t h_s^j (dN_s^j - \hat{\nu}_{s-}^j ds)$ , with  $g$  constant and the  $h^j$  are  $\mathbf{F}^N$ -predictable. Therefore, since  $\hat{K}_t$  is the  $L^2$  projection of  $K_t$  onto the space of square integrable  $\mathcal{F}_t^N$ -measurable random variables, the coefficients in the representation (65) are uniquely determined by the normal equations

$$\mathbb{E} \left[ \left( K_t - \gamma - \sum_j \int_0^t \eta_s^j (dN_s^j - \hat{\nu}_{s-}^j ds) \right) \left( g + \sum_j \int_0^t h_s^j (dN_s^j - \hat{\nu}_{s-}^j ds) \right) \right] = 0,$$

for all constants  $g$  and all  $\mathbf{F}^N$ -predictable processes  $h^j$ . Setting  $g = 0$  and using (59) give

$$\mathbb{E} \left[ K_t \sum_j \int_0^t h_s^j (dN_s^j - \hat{\nu}_{s-}^j ds) - \sum_j \int_0^t h_s^j \eta_s^j \hat{\nu}_{s-}^j ds \right] = 0. \quad (66)$$

For  $j \in \{1, \dots, \rho\}$ , we compute  $\mathbb{E} \left[ K_t \int_0^t h_s^j dN_s^j \right]$ . Using that  $\hat{K}$  is a  $\mathbf{F}^N$ -martingale and that  $K$  and  $N^j$  have no common jumps, we have

$$\begin{aligned} \mathbb{E} \left[ K_t \int_0^t h_s^j dN_s^j \right] &= \mathbb{E} \left[ K_t \sum_{s \leq t} h_s^j \Delta N_s^j \right] = \sum_{s \leq t} \mathbb{E} \left[ \mathbb{E} [K_t | \mathcal{F}_s^N] h_s^j \Delta N_s^j \right] \\ &= \sum_{s \leq t} \mathbb{E} \left[ \hat{K}_s h_s^j \Delta N_s^j \right] = \mathbb{E} \left[ \int_0^t K_{s-} h_s^j dN_s^j \right] \\ &= \mathbb{E} \left[ \int_0^t K_{s-} h_s^j \nu_s^j ds \right] = \int_0^t \mathbb{E} [h_s^j \mathbb{E} [K_{s-} \nu_s^j | \mathcal{F}_s^N]] ds \\ &= \mathbb{E} \left[ \int_0^t h_s^j \widehat{\Theta \nu^j}_{s-} ds \right] - \mathbb{E} \left[ \int_0^t h_s^j \hat{\nu}_{s-}^j - \int_0^s \hat{a}_u du ds \right]. \end{aligned}$$

Using similar arguments, we compute the second term

$$\begin{aligned} \mathbb{E} \left[ K_t \int_0^t h_s^j \hat{\nu}_{s-}^j ds \right] &= \int_0^t \mathbb{E} [h_s^j \mathbb{E} [K_t | \mathcal{F}_{s-}^N] \hat{\nu}_{s-}^j] ds = \int_0^t \mathbb{E} [h_s^j \hat{K}_{s-} \hat{\nu}_{s-}^j] ds \\ &= \mathbb{E} \left[ \int_0^t h_s^j \hat{\Theta}_{s-} \hat{\nu}_{s-}^j ds \right] - \mathbb{E} \left[ \int_0^t h_s^j \hat{\nu}_{s-}^j - \int_0^s \hat{a}_u du ds \right]. \end{aligned}$$

Inserting these expressions into (66), gives

$$\sum_j \mathbb{E} \left[ \int_0^t h_s^j \left( \widehat{\Theta \nu^j}_{s-} - \hat{\Theta}_{s-} \hat{\nu}_{s-}^j - \eta_s^j \hat{\nu}_{s-}^j \right) ds \right] = 0.$$

Choosing  $h_s^j$  equal to the expression in the parentheses, gives  $\sum_j \mathbb{E} \left[ \int_0^t (h_s^j)^2 ds \right] = 0$  hence all  $h^j$  vanish and  $\forall j = 1, \dots, \rho$ :

$$\eta_s^j = \frac{\widehat{\Theta_{s-} \nu_{s-}^j}}{\hat{\nu}_{s-}^j} - \hat{\Theta}_{s-}. \quad (67)$$

From (65), (63), and the equality  $\hat{K} = \hat{L} + \hat{M}$ , it follows that

$$\hat{\Theta}_t = \mathbb{E}[\Theta_0] + \int_0^t \hat{a}_s ds + \sum_j \int_0^t \eta_s^j (dN_s^j - \hat{\nu}_{s-}^j ds) \quad (68) \quad \blacksquare$$

with the  $\eta^j$  are given by (67). This finishes the proof of the proposition.

## B.2 Finite latent factor model and a credit risk application

*Proof (of Prop. 9).* In order to apply Prop. 8, one needs to find the representation (42) for  $I_t^h$ . Let  $\Psi^{rh}$ ,  $r \neq h$ ,  $r, h \in \mathbb{T}$ , be the counting processes defined by

$$\Psi_t^{rh} = \#\{s \in (0, t]; \Theta_{s-} = r, \Theta_s = h\}.$$

The starting point is the expression

$$I_t^h = I_0^h + \sum_{r; r \neq h} (\Psi_t^{rh} - \Psi_t^{hr}),$$

which comes from the obvious dynamics

$$dI_t^h = \sum_{r; r \neq h} (d\Psi_t^{rh} - d\Psi_t^{hr})$$

The counting processes  $\Psi^{rh}$  have intensities of the form  $I_{t-}^r \kappa^{rh}$ . Reshaping the last expression as

$$\begin{aligned} I_t^h &= I_0^h + \int_0^t \sum_{r; r \neq h} (I_{s-}^r \kappa^{rh} - I_{s-}^h \kappa^{hr}) ds \\ &+ \int_0^t \sum_{r; r \neq h} [(d\Psi_s^{rh} - I_{s-}^r \kappa^{rh} ds) - (d\Psi_s^{hr} - I_{s-}^h \kappa^{hr} ds)], \end{aligned}$$

shows that  $I^h$  is of the form

$$I_t^h = \int_0^t a_s^h ds + M_t^h,$$

with

$$a_t^h = \sum_{r; r \neq h} (I_{t-}^r \kappa^{rh} - I_{t-}^h \kappa^{hr}) = \sum_r \kappa^{rh} I_{t-}^r \quad (69)$$

and  $M^h$  is a martingale commencing at  $M_0^h = I_0^h$ .

Then the role of  $a_t$  is taken by  $a_t^h$  in (69), the role of  $(\Theta \nu^j)_{t-}$  is taken by

$$(I^h \nu^{ij})_{t-} = I_{t-}^h Y_t^i \sum_r \ell^{r,ij} I_{t-}^r = Y_t^i \ell^{r,ij} I_{t-}^h,$$

and the  $F^N$ -intensities of  $N$  are given in (55). Inserting these expressions into (43), gives

$$d\hat{I}_t^h = \sum_{r=1}^m k^{rh} \hat{I}_{t-}^r dt + \sum_{i \neq j} \left( \frac{l^{h,ij} \hat{I}_{t-}^h}{\sum_r l^{r,ij} \hat{I}_{t-}^r} - \hat{I}_{t-}^h \right) \left( dN_t^{ij} - Y_t^i \sum_{r=1}^m l^{r,ij} \hat{I}_{t-}^r dt \right) \quad (70) \quad \blacksquare$$

This result may look similar to R4 of [6, Sec.IV.1] but we actually consider a more general framework. On the hand, we deal with an aggregated version over the entire portfolio of the multivariate process and on the other hand we take censorship into account though the processes of risk exposure  $Y^i$ .

## B.3 Calibration of the continuous version

Here, we present the detailed computations of the adaptations of the calibration for the continuous filtering framework, presented in Section 3.3. In practice the number of entities monitored over time may vary: either because some names appear or disappear or simply because of missing data. This happens when the data is missing, censored or when it is not appeared yet. We attribute

the rating 0 to an entity in this case. Then it is clear that a transition involving the rating of censure 0, is assumed to be independent with the states of the hidden factor. Let consider the list of ratings  $\tilde{\Upsilon} = \{1, \dots, p\}$  and  $\Upsilon = \{0, \dots, p\}$ , the completed list of ratings. Note that the total number of entities observed on  $\Upsilon$  is constant equal to  $Q$ . Let  $Q_t$ , be the number of entities which have their rating in  $\tilde{\Upsilon}$  (have a real rating) at time  $t$ .

We propose a calibration algorithm which assumes that no more than one entity may jump at a given time step. In order to make the model identifiable while considering the impact of the size of the sample (which may evolve), we define an independent process  $I$ , with values in  $\{0, \dots, Q\}$ , which uniformly picks the entity that may jump. If  $I$  picks an entity which is rated 0, (because not already rated or censored), we do not observe jumps. Otherwise the entity jumps according to the transition matrices  $(L^h)_h$ .

We have  $\forall t \in \{0, \dots, \Gamma\}$ ,  $(i, j) \in \{1, \dots, p\}^2$ ,  $h \in \mathbb{T}$ ,  $q \in \{0, \dots, Q\}$

$$\mathbb{P}(Z_t^q = j | Z_{t-1}^q = i, I_{t-1} = q, \Theta_{t-1} = h) = L^{h,ij}.$$

For  $z_t, z_{t-1} \in \Upsilon^Q$ , we define  $W_{t-1}^h = \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, \Theta_{t-1} = h)$ , where  $Z_t = (Z_t^q)_{q \leq Q_t}$ . We compute

$$\begin{aligned} W_{t-1}^h &= \sum_{d=1}^{Q_t} \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) \mathbb{P}(I_{t-1} = d) \\ &= \sum_{d=1}^{Q_{t-1}} \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) \mathbb{P}(I_{t-1} = d) \\ &\quad + \sum_{d=Q_{t-1}+1}^{Q_t} \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) \mathbb{P}(I_{t-1} = d) \end{aligned}$$

Let focus on the first sum, describing the situation when the chosen entity has a rating at current time.

$$\begin{aligned} \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) &= \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) \mathbb{1}_{[z_t^d = z_{t-1}^d, \forall l \neq d: z_t^l = z_{t-1}^l]} \\ &\quad + \mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) \mathbb{1}_{[z_t^d \neq z_{t-1}^d, \forall l \neq d: z_t^l = z_{t-1}^l]} \\ &= L^{h, z_{t-1}^d z_t^d} \mathbb{1}_{[z_t = z_{t-1}]} + L^{h, z_{t-1}^d z_t^d} \mathbb{1}_{[z_t^d \neq z_{t-1}^d, \forall l \neq d: z_t^l = z_{t-1}^l]} \end{aligned}$$

For the second sum, we have:  $\mathbb{P}(Z_t = z_t | Z_{t-1} = z_{t-1}, I_{t-1} = d, \Theta_{t-1} = h) = \mathbb{1}_{[z_t = z_{t-1}]}$ .

So finally,

$$W_{t-1}^h = (1 - \frac{Q_{t-1}}{Q}) \mathbb{1}_{[z_t = z_{t-1}]} + \sum_{d=1}^{Q_{t-1}} \frac{1}{Q} \mathbb{1}_{[|z_t - z_{t-1}|_0 \leq 1]} \mathbb{1}_{[\forall l \neq d: z_t^l = z_{t-1}^l]} L^{h, z_{t-1}^d z_t^d},$$

where  $|x|_0 = \#\{x_i \neq 0\}$ .

Then, it is easy to check that the previous algorithm can be adapted to the new framework

$$\alpha_t(h) = \sum_{s=1}^m \alpha_{t-1}(s) K^{sh} W_{t-1}^h,$$

$$\beta_t(h) = \sum_{l=1}^m \beta_{t+1}(l) K^{hl} W_t^h,$$

$$\check{u}_t(h) = \mathbb{P}(\Theta_t = h | Z_{0|\Gamma} = z_{0|\Gamma}) = \frac{\beta_{t+1}(h) \alpha_{t+1}(h)}{L_\Gamma},$$

$$\check{v}_t(s, h) = \mathbb{P}(\Theta_t = h, \Theta_{t-1} = s | Z_{0|\Gamma} = z_{0|\Gamma}) = \frac{\beta_{t+1}(h) K^{sh} \alpha_t(s) W_t^h}{L_\Gamma}.$$

The forms of the transitions matrices  $(L^h)_h$ , are a lot impacted by this adaptation. The maximisation does not run as simply as it does for the discrete setting. Explicit forms are heavy to derive. Then, these parameters are directly estimated with optimization algorithms.