A comparative study on the estimation of factor migration models

Areski Cousin, Mohamed Reda Kheliouen

A comparative study on the estimation of factor migration models

Areski Cousin
E-mail: areski.cousin@univ-lyon1.fr

Mohamed Reda Kheliouen
E-mail: mohamed-reda.kheliouen@etu.univ-lyon1.fr

Address: Institut de Science Financière et d'Assurances,
50 Avenue Tony Garnier, F-69007 Lyon, France.

May 15, 2015

Abstract

In this paper, we study the statistical estimation of some factor credit migration models, that is, multivariate migration models for which the transition matrix of each obligor is driven by the same dynamic factors. In particular, we compare the statistical estimation of the ordered Probit model as described for instance in Gagliardini and Gourieroux (2005) and of the multi-state latent factor intensity model used in Koopman et al. (2008). For these two approaches, we also distinguish the case where the underlying factors are observable and the case where they are assumed to be unobservable. The paper is supplied with an empirical study where the estimation is made on a set of historical Standard & Poor's rating data on the period $[01/2006 - 01/2014]$. We find that the intensity model with observable factors is the one that has the best fit with respect to empirical transition probabilities. In line with Kavvathas (2001), this study shows that short migrations of investment grade firms are significantly correlated to the business cycle whereas, because of lack of observations, it is not possible to state any relation between long migrations (more than two grades) and the business cycle. Concerning non investment grade firms, downgrade migrations are negatively related to business cycle whatever the amplitude of the migration.

Keywords: Factor migration models, ordered Probit model, multi-state latent intensity model, mobility index, Kalman filter.

1 Introduction

Analyzing the effect of business cycle on rating transition probabilities has been a subject of great interest these last fifteen years, particularly due to the increasing pressure coming from regulators for stress testing.

We consider the statistical estimation of credit migration models where, for each firm, the transition matrix is stochastic and may depend on a pool of common dynamic factors. The underlying factors aim at representing the evolution of the business cycle. The estimation procedure of such models differs given that the factors are considered to be observable or unobservable. In the first approach, one selects observable factors (macroeconomic variables for example) via economic analysis and then estimate the transition probabilities' sensitivities with respect to these covariates. Kavvathas (2001) calibrates a multi-state extension of a Cox proportional hazard model with respect to the 3-months and 10-years interest rates, the equity return and the equity return volatility. He finds that an increase in the interest rates, a lower equity return and a higher equity return volatility are associated with higher downgrade intensities. Nickell et al. (2000) employ an ordered Probit model and prove the dependence of transition probabilities on the industry of the obligor, the business country of the obligor and the stage of the business cycle. Bangia et al. (2002) separate the economy into two regimes, expansion and contraction, estimate an ordered Probit model and show that the loss distribution of credit portfolios can differ greatly, as can the concomitant level of economic capital to be assigned. The second approach has emerged in response to criticisms made against the first approach. As Gourieroux and Tiomo (2007) point out, the risk in selecting covariates lies in excluding other ones which would be more relevant. Specifications with latent covariates appear in articles such as Gagliardini and Gourieroux (2005) who consider an ordered Probit model with three unobservable factors and perform its estimation using a Kalman filter on ratings data of french corporates. They find that the two first factors are related to the change in french GDP (considered as a proxy of the business cycle). Koopman et al. (2008) proceed on a parametric intensity model by conditioning the migration intensity on both observable factors and latent dynamic factors. The estimation shows the existence of a common risk factor for all migrations. The impact of this risk factor is higher for downgrades than for upgrades, this empirical result suggests that upgrades are more subject to idiosyncratic shocks than downgrades.

The aim of this paper is to assess and compare two alternative stochastic migration models on their ability to link the transition probabilities to either observable or unobservable dynamic risk factors. In this respect, we use the same data set to compare the the multi-state latent factor intensity model used in Koopman et al. (2008) and the ordered Probit model as described for instance in Gagliardini and Gourieroux (2005). We use the S&P credit ratings history $[01/2006 - 01/2014]$ of a diversified portfolio composed of 2875 obligors and distributed across several regions and sectors. When the underlying factors are assumed to be unobservable, we adapt a method introduced in Gagliardini and Gourieroux (2005) to represent the considered factor migration model as a linear Gaussian models. The unobservable factors are then filtered by a standard Kalman filter.

The paper is organized as follows. In Section 2, we present the class of factor migration models. Then, we introduce intensity models and structural models as particular factor migration models. Section 3 describes the estimation procedure for a multi-state factor intensity model in the two cases where the factor is assumed to be observable or unobservable. In Section 4, we present the estimation procedure associated with a structural ordered Probit model and we also carry the case of observable and unobserbable factors. In Section 5, we perform the estimation procedures on S&P credit ratings historical data and provide our main results and findings.

2 Factor migration models

In this paper, we consider Markovian model consisting of a multivariate factor process X and a vector R representing the rating migration process in a pool of n obligors. More specifically, for any time $t \geq 0$, $R_t = (R_t^1, \dots, R_t^n)$ is a vector in $\{1, \cdots, d\}^n$, where d corresponds to the default state and 1 corresponds to the state with the best credit quality. The l-th entry of R describes the rating migration dynamics of obligor l in the set $S = \{1, \cdots, d\}$. The transition probabilities are driven by a factor process X. We consider that all sources of risk are defined with respect to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with some reference filtration $\mathbb{F} = (\mathcal{F}_u)_{0 \leq u \leq t}$ satisfying the usual conditions. For any $l = 1, \dots, n$, let $\mathcal{H}^l$ be the filtration associated with the migration process R^l and let $\mathcal{H}$ be the global filtration such that $\mathcal{H}_u^l = \mathcal{H}_u^1 \vee \cdots \vee \mathcal{H}_u^n$, $u \geq 0$. Let $\mathcal{G}$ be the filtration associated with the factor process X. The credit migration models we consider in this paper are in the class of factor migration models.

Definition 2.1. *A factor credit migration model is a Markov process (X, R) such that*

- *The factor X is a Markov process (in its own filtration).*
- *Given the history of X, i.e. given $\mathcal{G}_\infty$, the marginal rating migration processes $R^1, \dots, R^n$ are independent (time-inhomogeneous) Markov chains with the same transition matrices.*

Contrary to the definition of stochastic migration models in Gagliardini and Gourieroux (2005), here the factor process X is not necessarily identified as a common stochastic transition matrix. Note that the factor process X plays two important roles : it introduces dynamic dependence between the obligor rating migration processes and it allows for non-Markovian serial dependence in the migration dynamics: the rating migration process R is not a Markov process. Note also that the conditional Markov chain $(R^1, \ldots, R^n)$ has d^n transition states. The conditional independence property allows to significantly reduce the problem dimension when it turns to compute the joint distribution of rating migration events. In that sense, conditionally to the filtration $\mathcal{G}_\infty$, the $R_t^1, \ldots, R_t^n$ are *iid* for every date t. Moreover, there is no contagion mechanism in this framework. A migration event in the pool has no impact on the obligors migration probabilities since the latter are driven by a process X which is Markov in its own filtration. In Koopman et al. (2008), the factor process X can contain obligor-specific informations (microeconomic variables) as well as common observable factors (macroeconomic variables) or common unobserved latent factors. The model presented in this paper does not distinguish between obligors as the process X stands only for common observable or latent factors.

2.1 Intensity models

In this section, we define multi-state factor intensity models as particular factor migration models. Given the history of X (given $\mathcal{G}_\infty$), the rating migration processes $(R_t^l)_{t\geq 0}$, $l = 1, \ldots, n$ are conditionally independent continuous-time Markov chains. Moreover, for any time t, they are assumed to have a common generator matrix $\Lambda_X(t)$ defined by

$$\Lambda_X(t) = \begin{pmatrix} -\lambda_1(X_t) & \lambda_{12}(X_t) & \cdot & \lambda_{1,d-1}(X_t) & \lambda_{1,d}(X_t) \\ \lambda_{21}(X_t) & -\lambda_2(X_t) & \cdot & \lambda_{2,d-1}(X_t) & \lambda_{2,d}(X_t) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_{d-1,1}(X_t) & \lambda_{d-1,2}(X_t) & \cdot & -\lambda_{d-1}(X_t) & \lambda_{d-1,d}(X_t) \\ 0 & 0 & \cdot & 0 & 0 \end{pmatrix},$$

where, for any $j \neq i$, λ_{ij} are positive functions and

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

for $i = 1, \cdots, d-1$. For $j \neq i$, the product $\lambda_{ij}(X_t)dt$ corresponds to the conditional probability of going to rating state j in the small time intervalle $(t, t+dt]$, given $\mathcal{G}_t$ and the fact that the obligor is in rating state i at time t. For small time length dt, the following first-order approximation holds

$$\lambda_{ij}(X_t)\, dt \approx \mathbb{P}\left(R_{t+dt}^l = j \mid R_t^l = i, \mathcal{G}_t\right), \tag{1}$$

for any obligor $l = 1, \ldots, n$. In turns out that $\lambda_i(X_t)\, dt$ is the conditional probability to depart from state i in the small interval $(t, t+dt]$ given $\mathcal{G}_t$.

A classical specification of migration intensities is an exponential-affine transformation of the common factor. More specifically, for any $i \neq j$,

$$\lambda_{ij}(X_t) = \exp\left(\alpha_{ij} + \langle \beta_{ij}, X_t\rangle\right), \tag{2}$$

where α_{ij} is a constant parameter and β_{ij} accounts for the sensitivity of migration intensity λ_{ij} to the common factor X. This specification garanties the positivity of migration intensities. This specification corresponds to a multi-state extension of the Cox proportional hazard model.

In this framework, the conditional probability transition matrix P (given $\mathcal{G}_\infty$), can be computed by solving the forward Kolmogorov equation

$$\frac{\partial P(t,s)}{\partial s} = P(t,s)\Lambda_X(s), \quad P(t,t) = \mathrm{Id}. \tag{3}$$

The unconditional transition matrix is then given by

$$\Pi(t,s) = \mathbb{E}\left[P(t,s)\right], \tag{4}$$

where the expectation is taken over the distribution of $(X_u)_{t<u\leq s}$. In practice, the transition matrix $\Pi(t,s)$ can be approximated by Monte Carlo simulations, based on independent simulations of the path of X between t and s. For each realization of $(X_u)_{t<u\leq s}$, a simple numerical scheme can be used to solve (3).

If the factor process X only changes at discrete times $t_1 < \ldots < t_N$, the conditional transition matrix can be expressed as a product involving the following matrix exponential terms

$$P(t_k, t_{k+1}) = e^{\Lambda_X(t_k)(t_{k+1}-t_k)}. \tag{5}$$

Then, under this assumption, the forward Kolmogorov equation (3) has an explicit solution. Note that, if the generator matrix Λ_X is time-inhomogeneous, it is generally not true that one can extend (5) to get

$$P\left(t,s\right) = e^{\int_t^s \Lambda_X(u)\,du}. \tag{6}$$

However, this relation holds for some specification of Λ_X for which the matrix $\Lambda_X(u)$ and $\Lambda_X(u')$ for all u, u' in $[t,s]$.

2.2 Structural models

We consider a discrete-time structural models where any firm l jumps to a new rating category when a quantitative latent process S^l crosses some pre-specified levels or barriers. In the classical structural Merton model, S^l is defined as the ratio of asset value and liabilities. The rating of name l at time t is given by the position of the latent variable S_t^l inside a pre-specified partition of the real line $-\infty = C_{d+1} < C_d < \cdots < C_{i+1} < C_i < \cdots < C_1 = +\infty$. More formally,

$$R_t^l = \sum_{i=1}^{d} i \mathbf{1}_{\{C_{i+1} \leq S_t^l < C_i\}}. \tag{7}$$

Several models exist in literature for the specification of S^l (see, e.g., Nickell et al. (2000), Bangia et al. (2002), Albanese et al. (2003), Feng et al. (2008)). In this study, we choose Gagliardini and Gourieroux (2005)'s approach since it is a generalization of previously cited models, it is also investigated in Gourieroux and Tiomo (2007) and Feng et al. (2008). For any obligor $l = 1, \ldots, n$ and any time t, the latent process S_t^l is expressed as a deterministic affine transformation of a common factor X_t and of an independent idiosyncratic factor ε_t^l. The characteristics of this affine transformation may depend on the rating state at the preceding date $t-1$. The latent process S^l is then described by

$$S_t^l = \sum_{i=1}^{d} \left(\alpha_i + \langle \beta_i, X_t \rangle + \sigma_i \varepsilon_t^l\right) \mathbf{1}_{\{R_{t-1}^l = i\}}, \tag{8}$$

where α_i is the level of S^l at rating i, β_i represents the sensitivity in rating i of the latent factor S^l to the common factor X, σ_i corresponds to the volatility of residuals and ε_t^l are iid random variables independent of X. When the idiosyncratic residual processes ε^l, $l = 1, \ldots, n$ are not specified, this corresponds to an ordered polytomous model. The most common version is the ordered Probit model where ε_t^l are independent standard Gaussian variables (see Gagliardini and Gourieroux (2005) or Feng et al. (2008) for more details). In what follows, we consider the ordered Probit model and we denote by $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\ldots,d-1}$ the set of unknown parameters.

In this framework, the conditional transition probabilities p_{ij} are given by[1],

$$
\begin{aligned}
p_{ij} &= \mathbb{P}\left[R_t^l = j \mid R_{t-1}^l = i, \mathcal{G}_\infty\right], \\
&= \mathbb{P}\left[C_{j+1} \leq S_t^l < C_j \mid R_{t-1}^l = i, \mathcal{G}_\infty\right], \\
&= \mathbb{P}\left[C_{j+1} \leq \alpha_i + \langle \beta_i, X_t \rangle + \sigma_i \varepsilon_t^l < C_j \mid R_{t-1}^l = i, X_t\right], \\
&= \mathbb{P}\left[\frac{C_{j+1} - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i} \leq \varepsilon_t^l < \frac{C_j - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i} \mid X_t\right].
\end{aligned}
$$

Then, if $\Phi(\cdot)$ is the cumulative distribution function of a standard Gaussian variable, we obtain, for any $i = 1, \cdots, d-1$,

$$
p_{ij} = \Phi\left(\frac{C_j - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i}\right) - \Phi\left(\frac{C_{j+1} - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i}\right), \ j = 2, \cdots, d-1, \tag{9}
$$

$$
p_{i1} = 1 - \Phi\left(\frac{C_2 - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i}\right), \tag{10}
$$

$$
p_{id} = \Phi\left(\frac{C_d - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i}\right). \tag{11}
$$

3 Statistical estimation of intensity models

In this section, we consider a multi-state intensity model as described in Section 2.1 and we explain how to estimate the parameters of the generator matrix given the sample history of the obligors rating migrations. In a general setting, we first give the expression of the conditional likelihood function. For the ease of the presentation, we distinguish the simple case where migration intensities are constant over time and the case where, as in a (multi-state) proportional hazard model, the intensities are given as exponential-affine transformation of an observable factor X. In this setting, the estimation of model parameters is made by maximizing the conditional likelihood function. When the factor X is assumed to be unobservable, the estimation requires the computation of the unconditional likelihood function. This task may be computationally intensive as explained in Koopman et al. (2008). In this paper, we choose to adapt the approach of Gagliardini and Gourieroux (2005) (estimation of an ordered Probit model with unobervable factors) to the multi-state factor intensity model.

We assume that, for any obligor $l = 1, \cdots, n$, the observed number of ratings visited during the period $[0, t)$ is denoted by N_l ($N_l \geq 1$). For any $k = 1, \cdots, N_l$, the time intervals $\left[t_{k-1}^l, t_k^l\right)$ correspond to the visiting of new state r_k^l where $t_0^l = 0$ and $t_{N_l}^l = t$. Then, the observed path of ratings of obligor l during the period $[0, t)$ is described by

$$
r_u^l = \sum_{k=1}^{N_l} r_k^l \mathbf{1}_{\{t_{k-1}^l \leq u < t_k^l\}}, \tag{12}
$$

There are $N_l - 1$ migration events observed for obligor l in the time interval $[0, t)$, each of them has took place at time t_k^l for $k = 1, \cdots, N_l - 1$.

3.1 Conditional likelihood function

Proposition 3.1. *Let θ be the set of parameters which characterizes the functional link between the generator matrix Λ_X and the factor process X. The conditional likelihood function given the observed*

[1] See Gagliardini and Gourieroux (2005) for more details.

path of rating migration histories $(r_u^l)_{0\le u<t}$ *and the path of the risk factor* $(X_u)_{0\le u\le t}$ *can be expressed as*

$$\mathcal{L}(\theta \mid \mathcal{G}_t) = \prod_{l=1}^{n}\prod_{k=1}^{N_l} \lambda_{r_k^l, r_{k+1}^l}(X_{t_k^l})\; e^{-\int_{t_{k-1}^l}^{t_k^l} \lambda_{r_k^l}(X_u)\,du}, \tag{13}$$

with the convention $\lambda_{r_{N_l}^l, r_{N_l+1}^l} = 1$.

Proof. See proof in appendix A. □

Note that the conditional likelihood function can be expressed as a product of (conditional) marginal likelihood function, that is

$$\mathcal{L}(\theta \mid \mathcal{G}_t) = \prod_{i\neq j} \mathcal{L}_{ij}(\theta \mid \mathcal{G}_t) \tag{14}$$

where the product is taken over all transition types and $\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)$ is given in the following proposition.

Proposition 3.2. *The marginal likelihood function* $\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)$ *associated with migration from state* i *to state* j $(i \neq j)$ *is given by,*

$$\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t) = \prod_{l=1}^{n}\prod_{k=1}^{N_l} \exp\left(Y_{ij}^l(t_k^l)\log \lambda_{ij}(X_{t_k^l}) - S_i^l(t_k^l)\int_{t_{k-1}^l}^{t_k^l} \lambda_{ij}(X_u)\,du \right) \tag{15}$$

where

$$Y_{ij}^l(t_k^l) = \begin{cases} 1 & \text{if } i = r_{t_k^l}^l \text{ and } j = r_{t_{k+1}^l}^l, \\ 0 & \text{else,} \end{cases} \tag{16}$$

for $k = 1, \dots, N_l - 1$, $Y_{ij}^l(t_k^l) = 0$ *for* $k = N_l$ *and*

$$S_i^l(t_k^l) = \begin{cases} 1 & \text{if } i = r_{t_k^l}^l \\ 0 & \text{else,} \end{cases} \tag{17}$$

for $k = 1, \dots, N_l$.

According to (12), the quantity $Y_{ij}^l(t_k^l)$ is equal to one (and zero otherwise) if the couple of indices (i,j) corresponds to the observed migration event of obligor l at time t_k^l, i.e., if $i = r_{t_k^l}^l$ and $j = r_{t_{k+1}^l}^l$. Moreover, $Y_{ij}^l(t_k^l) = 0$ for $k = N_l$ since, by construction, no migration occurs after time $t_{N_l-1}^l$. The quantity $S_i^l(t_k^l)$ is equal to one (and zero otherwise) if index i corresponds to the rating state visited by obligor l just before migration date t_k^l, i.e., if $i = r_{t_k^l}^l$.

Proof. Proposition (3.2) is a direct consequence of Proposition (3.1). □

If each migration intensity function λ_{ij} is described by specific parameters, the maximization of the conditional likelihood function (13) can be done by maximizing each marginal likelihood independently with respect to its own set of parameters. Numerically, it is more convenient to work with the log-likelihood function, we then transform (15) to

$$\log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right) = \sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)\log\left(\lambda_{ij}(X_{t_k^l})\right) - \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\int_{t_{k-1}^l}^{t_k^l} \lambda_{ij}(X_u)\,du, \tag{18}$$

Likelihood function in the multi-state intensity model is classical from the point of view of point processes (Hougaard (2000)). We compute here the likelihood function for the conditional Markov chain assumption. The link between the two approaches is clear as soon as we recall that Markov chains are particular marked point processes. The likelihood expression (18) is the same one obtained in Koopman et al. (2008) except that in Koopman et al. (2008) the time is discrete and divided according to the observed dates of change in value of the common process X.

3.2 Time-homogeneous intensities

Let us assume that the transition intensity functions λ_{ij} are constant, so that, $\lambda_{ij}(X_t) = \lambda_{ij,0}$ for any time t. As explained above, the max-likelihood estimation of $\lambda_{ij,0}$, $i \neq j$ is obtained by maximizing each corresponding marginal likelihood functions.

Proposition 3.3. *The maximum likelihood estimate of the transition intensities is given by,*

$$\hat{\lambda}_{ij,0} = \frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\left(t_k^l - t_{k-1}^l\right)}. \tag{19}$$

Note that the estimator $\hat{\lambda}_{ij,0}$ is expressed as the ratio of the total number of observed transitions from state i to state j over the cumulated time spent by the obligors in state i.

Proof. Given (18), $\hat{\lambda}_{ij,0}$ is the solution of $\dfrac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \lambda_{ij,0}} = 0$. And, we have

$$\begin{aligned}
\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \lambda_{ij,0}} &= \sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)\frac{\partial \log\left(\lambda_{ij}\right)}{\partial \lambda_{ij,0}} - \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\frac{\partial \lambda_{ij}}{\partial \lambda_{ij,0}}\left(t_k^l - t_{k-1}^l\right) \\
&= \frac{1}{\lambda_{ij,0}}\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l - \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\left(t_k^l - t_{k-1}^l\right) \\
&= 0
\end{aligned}$$

Moreover, the marginal likelihood function is concave since it is easily seen that $\dfrac{\partial^2 \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial^2 \lambda_{ij,0}} = \sum_{l=1}^{n}\sum_{k=1}^{N_l} \dfrac{-Y_{ij}^l(t_k^l)}{\left(\lambda_{ij,0}\right)^2} < 0$. □

Under this specification, the dynamics of rating migrations does not depend on the business cycle. In the numerical part (see Section 5), this estimation procedure will be used to construct Through The Cycle[2] (TTC) transition matrix.

3.3 Intensities depending on observable factors

We now assume that migration intensities depend on a K-dimensional observable factor $X = (X^1, \dots, X^K)$ through an exponential-affine relation:

$$\lambda_{ij}(X_t) = \lambda_{ij,0}\ \exp\left(\langle \beta_{ij}, X_t\rangle\right), \quad i \neq j,\ i < d. \tag{20}$$

for all. The vector $\beta_{ij} = \left(\beta_{ij}^1, \cdots, \beta_{ij}^K\right)$ contains the sensitivities of λ_{ij} with respect to each component of factor X. The term $\langle \beta_{ij}, X_t\rangle$ denotes the inner product between vectors β_{ij} and X_t. The baseline intensity $\lambda_{ij,0}$ is assumed to be constant. This specification corresponds to a multi-state version of the so-called Cox proportional hazards regression model, where here the baseline function is assumed to be constant. This setting has been investigated by among others Kavvathas (2001), Lando and Skodeberg (2002), Koopman et al. (2008), Naldi et al. (2011).

Proposition 3.4. *Under specification (20), for any transition type (i, j), the maximum likelihood estimate $\hat{\beta}_{ij} = \left(\hat{\beta}_{ij}^1, \cdots, \hat{\beta}_{ij}^K\right)$ is solution to the following non-linear system:*

$$\frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)\ X_{t_k^l}^s}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)} = \frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\ \int_{t_{k-1}^l}^{t_k^l} X_{s,u}\ e^{\langle \hat{\beta}_{ij}, X_u\rangle}\, du}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\int_{t_{k-1}^l}^{t_k^l} e^{\langle \hat{\beta}_{ij}, X_u\rangle}\, du}, \quad s = 1, \cdots, K. \tag{21}$$

[2] Through The Cycle transition matrix corresponds to the long term time-invariant transition matrix of the homogeneous Markov chain. The transition probabilities are almost unaffected by the economical conditions.

The maximum likelihood estimate $\hat{\lambda}_{ij,0}$ is given by

$$\hat{\lambda}_{ij,0} = \frac{\sum_{l=1}^{n} \sum_{k=1}^{N_l} Y_{ij}^{l}(t_k^l)}{\sum_{l=1}^{n} \sum_{k=1}^{N_l} S_i^l(t_k^l) \int_{t_{k-1}^l}^{t_k^l} e^{\langle \hat{\beta}_{ij}, X_u \rangle} \, du}. \tag{22}$$

where $\hat{\beta}_{ij}$ is the solution of system (21).

Proof. See proof in appendix B. □

We notice from (22) that, when $\hat{\beta}_{ij}$ is fixed to 0, the maximum likelihood estimate of the baseline intensity falls back to the one of the homogeneous case (see (19)).

3.4 Intensities depending on unobservable factors

In the previous setting, we implicitly assume that the dynamics of migration intensities are fully explained by pre-identified observable factors. However, it may be the case that the dynamics of the intensities are driven by other (un-identified) factors or even by totally unobservable factors. We consider here that the migration intensities follow specification (20) but the vector of factors X is now assumed to be unobservable. Koopman et al. (2008) use a similar latent factor specification. We assume that the K-dimensional dynamic factor X only changes at times $t_1, \cdots, t_N$ and that its dynamics follows an auto-regressive (AR) process

$$X_t = AX_{t-1} + \zeta_t, \quad t = t_1, \ldots, t_N \tag{23}$$

where the matrix A characterizes the auto-regression coefficients and ζ_t are iid and standard Gaussian variables. The maximum likelihood estimation of $\lambda_{ij,0}$ and β_{ij} involves the computation of the unconditional likelihood function $\tilde{\mathcal{L}}(\theta)$ given by,

$$\tilde{\mathcal{L}}(\theta) = \mathbb{E}\left[\mathcal{L}\left(\theta \mid \mathcal{G}_t\right)\right], \tag{24}$$

where expectation in (24) is taken over the joint distribution of $(X_{t_1}, \cdots, X_{t_N})$. Consequently, maximizing the likelihood function is computationally very intensive[3] (see Koopman et al. (2008)) as soon as N is larger than a few units. We choose to follow another route by adapting the approach of Gagliardini and Gourieroux (2005) (initially proposed for a structural model) to the multi-state factor intensity model. The idea is to construct an approximation of model (20) as a linear Gaussian model which can be dealt with a Kalman filter. As usual, the estimation of model parameters is made by maximizing the likelihood of the filtered model.

Representation as a linear Gaussian model

Let us first remark that equation (20) is equivalent to the following linear relation

$$y_{ij,t} = \alpha_{ij} + \langle \beta_{ij}, X_t \rangle, \quad t = t_1, \ldots, t_N \tag{25}$$

where $y_{ij,t} = \log(\lambda_{ij,t})$ and $\alpha_{ij} = \log(\lambda_{ij,0})$. Note that, even if the transition intensities $\lambda_{ij,t}$, $t = t_1, \ldots, t_N$ (and then $y_{ij,t}$) are not directly observed, they can be estimated from the panel data. We assume that, for any time indices $k = 1, \ldots, N$ and for any transition type (i, j), λ_{ij,t_k} can be estimated from the migration dynamics observed in time interval (t_{k-1}, t_k) using the max-likelihood estimate (19). We then construct a time series of estimated migration intensities $\hat{\lambda}_{ij,t}$ and log-intensities $\hat{y}_{ij,t} := \log(\hat{\lambda}_{ij,t})$ $t = t_1, \ldots, t_N$. The asymptotic normality of the maximum likelihood estimate (19) writes $\sqrt{n}\left(\hat{\lambda}_{ij,t} - \lambda_{ij,t}\right) \to \mathcal{N}(0, \sigma_{ij,t}^2)$ as the number n of obligors goes to infinity. The asymptotic

[3] Koopman et al. (2008) propose to use a suitable important sampling technique to improve the efficiency of the Monte Carlo estimator.

variance $\sigma^2_{ij,t}$ can be approximated easily[4] (see Hougaard (2000)). Following the Delta method, we also know that $\sqrt{n}\,(\hat{y}_{ij,t} - y_{ij,t}) \to \mathcal{N}\left(0, \chi^2_{ij,t}\right)$ where $\chi^2_{ij,t} = \dfrac{\sigma^2_{ij,t}}{(\lambda_{ij,t})^2}$. As a result, if the panel data is sufficiently large, by using (25), we obtain the following linear Gaussian system

$$\begin{cases} \hat{y}_{ij,t} \simeq \alpha_{ij} + \langle \beta_{ij}, X_t \rangle + \tilde{\chi}_{ij,t}\kappa_{ij,t}, \quad \forall i \neq j,\ t = t_1, \ldots, t_N \\ X_t = AX_{t-1} + \zeta_t, \end{cases} \tag{26}$$

where for any $i \neq j$, $\tilde{\chi}_{ij,t} := \frac{\chi_{ij,t}}{\sqrt{n}}$, $\kappa_{ij,t}$ and ζ_t are independent error terms, distributed as standard Gaussian variables. For any $t = t_1, \ldots, t_N$, we denote by $\hat{y}_t$ the column vector composed of the elements $\hat{y}_{ij,t}$, $i \neq j$, $i < d$. Let p be the number of rows of $\hat{y}_t$ and $Y_t = (\hat{y}_{t_1}, \ldots, \hat{y}_t)$ be the information collected up to time t. Using the Kalman filter, the unobserved factor X can be filtered recursively given the estimated process $\hat{y}$. For any $t = t_1, \ldots, t_N$, let us denote by $\bar{X}_t := \mathbb{E}(X_t \mid Y_t)$ the filtered version of the factor process and by $\bar{X}_{t|t-1} := \mathbb{E}(X_t \mid Y_{t-1})$ the best prediction of X_t given information up to time $t-1$. In the Kalman filter terminology, the first equation in (26) corresponds to the measurement equation, the second to the transition equation.

Maximum likelihood of the filtered model

Let $\theta = (\alpha_{ij}, \beta_{ij})_{i\neq j, i<d}$ be the vector of unknown model parameters. The log-likelihood function associated with the filtered Gaussian model is given by (see, e.g., Durbin and Koopman (2012)),

$$\log\left(\mathcal{L}(\theta, A)\right) = -\frac{Np}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}\left(\log|F_t| + e_t^T F_t^{-1} e_t\right), \tag{27}$$

For any $t = t_1, \ldots, t_N$, the column vector e_t is the prediction error and is equal to $e_t := \hat{y}_t - \mathbb{E}[\hat{y}_t \mid Y_{t-1}]$.

The matrix F_t is the conditional variance of e_t given Y_{t-1}, i.e. $F_t = \mathrm{Var}(e_t \mid Y_{t-1})$, and $|F_t|$ is its determinant. Given the vector of "observations" Y_{t_N}, the prediction error e_t and the matrix F_t can be obtained at each time t as outputs of the Kalman routine.

The parameters $\hat{\alpha}_{ij}$, $\hat{\beta}_{ij}$ are chosen as the ones that maximize the log-likelihood function (27). We then define the estimated migration intensities as

$$\hat{\lambda}_{ij}(\bar{X}_t) = \exp\left(\hat{\alpha}_{ij} + \left\langle \hat{\beta}_{ij}, \bar{X}_t \right\rangle\right), \tag{28}$$

where $\bar{X}$ the filtered factor process associated with max-likelihood parameters $\hat{\theta}, \hat{A}$. The relation (28) will be used in the application on S&P data (see Subsection 5.4).

4 Statistical estimation of structural models

In this section, we consider the structural model described in Section 2.2 and we explain how to estimate the parameters of the conditional transition probability p_{ij} (see (9-11)) given the sample history of the obligors rating migrations. As for the intensity model, we distinguish the case where the underlying factor is observable and the case where it is unobservable. When the factor X is observable, the estimation of model parameters is made by maximizing the conditional likelihood function. When the factor X is assumed to be unobservable, the estimation requires the computation of the unconditional likelihood function. This task can be computationally very intensive. As for the intensity model, we consider as an alternative route the estimation approach used in Gagliardini and Gourieroux (2005).

[4] Following Hougaard (2000), the estimate $\hat{\sigma}_{ij,t}$ of $\sigma_{ij,t}$ is given by $\hat{\sigma}_{ij,t} = \dfrac{N_{ij,t}}{(T_{i,t})^2}$, where $N_{ij,t}$ is the total number of transitions from i to j (numerator of (19) for $t \in [t_{k-1}, t_k)$), and $T_{i,t}$ is the total time spent in state i (denominator of (19) for $t \in [t_{k-1}, t_k)$).

4.1 Observable factors

Assuming that the conditional transition probability p_{ij} depends on a K-dimensional observable factor $X = (X^1, \ldots, X^K)$ and considering $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,d-1}$ as the set of parameters which characterize the functional link between p_{ij} and the factor process X. For any $k = 1, \cdots, N$, the time intervals $[t_{k-1}, t_k)$ correspond to the visiting of the state r_k^l where $t_0 = 0$ and $t_N = t$. The observed path of ratings of obligor l during the period $[0, t)$ is described by

$$r_t^l = \sum_{k=0}^{N} r_k^l \mathbf{1}_{\{t=t_k\}}, \tag{29}$$

Proposition 4.1. *The conditional likelihood function associated with the observed path of rating migration histories* $\left(r_t^l\right)$, $t = t_0, \ldots, t_N$, $l = 1, \ldots, n$ *and given the initial ratings and the path of the risk factor* (X_t), $t = t_1, \ldots, t_N$ *can be expressed as,*

$$\mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0}) = \prod_{l=1}^{n} \prod_{k=1}^{N} \prod_{i \neq j} \left(p_{ij}\left(X_{t_k}\right)\right)^{Y_{ij}^l(t_k)}, \tag{30}$$

where, for $k = 1, \ldots, N$,

$$Y_{ij}^l(t_k) = \begin{cases} 1 & \text{if } i = r_{t_{k-1}}^l \text{ and } j = r_{t_k}^l, \\ 0 & \text{else.} \end{cases}$$

Proof.

$$\begin{aligned} \mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0}) &= \prod_{l=1}^{n} \mathbb{P}\left(R_u^l = r_u^l, u = t_1, \ldots, t_N \mid \mathcal{G}_{t_N}, R_{t_0}^l = r_{t_0}^l\right) \\ &= \prod_{l=1}^{n} \prod_{k=1}^{N} \mathbb{P}\left(R_{t_k}^l = r_{t_k}^l \mid R_{t_{k-1}}^l = r_{t_{k-1}}^l, \mathcal{G}_{t_N}\right) \end{aligned}$$

□

$p_{ij}\left(X_{t_k}\right)$ is defined by (9), (10) and (11) and $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,d-1}$.

We deduce that,

$$\hat{\theta} = \arg\max_{\theta} \left\{ \sum_{l=1}^{n} \sum_{k=1}^{N} \sum_{i \neq j} Y_{ij}^l(t_k) \log\left(p_{ij}\left(X_{t_k}\right)\right) \right\}, \tag{31}$$

solving (31) is done numerically by considering constraints on the positivity of $\sigma_i > 0$ and the order of threshold C_i $(-\infty = C_{d+1} < C_d < \cdots < C_2 < C_1 = +\infty)$ (see appendix C to have more details).

4.2 Unobservable factors

We now consider that the conditional transition probability p_{ij} depends on a K-dimensional unobservable factor $X = (X^1, \ldots, X^K)$. The dynamic factor X is assumed to change at times $t_1, \cdots, t_N$ with the following auto-regressive dynamics,

$$X_t = AX_{t-1} + \eta_t, \tag{32}$$

where the matrix A characterizes the covariates dynamics and η_t are iid and standard Gaussian.

The maximum likelihood estimation of $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,d-1}$ involves the computation of the unconditional likelihood function $\tilde{\mathcal{L}}(\theta)$ given by,

$$\tilde{\mathcal{L}}(\theta) = \mathbb{E}\left(\prod_{l=1}^{n} \prod_{k=1}^{N} \prod_{j \neq i} \left(p_{ij}\left(X_{t_k}\right)\right)^{Y_{ij}^l(t_k)}\right). \tag{33}$$

Estimating θ using (33) raises the issue of the potentially high number of integrals to compute. Indeed, the distribution must be integrated with respect to covariates values $X_{t_1}, \cdots, X_{t_N}$, which represents $[K \times N]$ integrals. Similarly to the intensity model, we use the method presented in Gagliardini and Gourieroux (2005) which consists in transforming the probability of migrating towards a rating J or below, denoted $p^*_{iJ,t}$ (see (34)), as a linear function of latent covariates (see (36) below) and thus construct a linear Gaussian model.

Representation as a linear Gaussian model

First, we formally define $p^*_{iJ,t}$ for $t = t_1, \ldots, t_N$ as,

$$p^*_{iJ,t_k} = \sum_{j=d}^{J} p_{ij,t_k} = \mathbb{P}\left[R^l_{t_k} < J | R^l_{t_{k-1}} = i, X_{t_k}\right], \tag{34}$$

Note that $p^*_{i1,t_k} = 1$ and thus the associated threshold $C_1 = +\infty$. The thresholds to be estimated are C_J for $J = 2, \cdots, d$, or equivalently C_{i+1} for $i = 1, \cdots, d-1$. We use the formulation C_{i+1} instead of C_J in the parameters vector $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,d-1}$ for the ease of notation.

Assuming an ordered Probit model, (34) becomes,

$$\begin{aligned} p^*_{iJ,t} &= \Phi\left(\frac{C_J - \alpha_i - \langle \beta_i, X_t \rangle}{\sigma_i}\right), \\ \Rightarrow \Phi^{-1}\left(p^*_{iJ,t}\right) &= \frac{C_J - \alpha_i}{\sigma_i} - \frac{1}{\sigma_i}\langle \beta_i, X_t \rangle, \end{aligned} \tag{35}$$

or equivalently,

$$\psi_{iJ,t} = \frac{C_J - \alpha_i}{\sigma_i} - \frac{1}{\sigma_i}\langle \beta_i, X_t \rangle, \tag{36}$$

Note that $p^*_{iJ,t}$ and $\psi_{iJ,t}$, $t = t_1, \ldots, t_N$ are not observed but can be estimated from the panel data. We assume that, for any time indices $k = 1, \ldots, N$ and for any arrival state J, $p^*_{iJ,t}$ (and then $\psi_{iJ,t}$) can be estimated from the migration dynamics observed in time interval (t_{k-1}, t_k) using the max-likelihood estimate (19) and the exponential matrix term expressed in (5). We then construct a time series of estimated $\hat{p}^*_{iJ,t}$ and $\hat{\psi}_{iJ,t}$, $t = t_1, \ldots, t_N$. The asymptotic normality of the estimator $\hat{p}^*_{iJ,t}$ writes $\sqrt{n}\left(\hat{p}^*_{iJ,t} - p^*_{iJ,t}\right) \hookrightarrow \mathcal{N}(0, \Omega^2_{iJ,t})$ as the number n of obligors goes to infinity[5]. Following the Delta method, we also know that $\sqrt{n}\left(\hat{\psi}_{iJ,t} - \psi_{iJ,t}\right) \hookrightarrow \mathcal{N}\left(0, \Xi^2_{iJ,t}\right)$ where $\Xi_{iJ,t} = \left(\Phi^{-1}\right)'(\hat{p}^*_{iJ,t})\Omega_{iJ,t}$. As a result, if the panel data is sufficiently large, by using (36), we obtain the following linear Gaussian representation

$$\begin{cases} \hat{\psi}_{iJ,t} \simeq \dfrac{C_J - \alpha_i}{\sigma_i} - \dfrac{1}{\sigma_i}\langle \beta_i, X_t \rangle + \tilde{\Omega}_{iJ,t} \upsilon_{iJ,t} & , \forall i, J,\ t = t_1, \ldots, t_N, \\ X_t = A X_{t-1} + \eta_t, \end{cases} \tag{37}$$

where $\tilde{\Omega}_{iJ,t} := \frac{\Omega_{iJ,t}}{\sqrt{n}}$, $\upsilon_{iJ,t}$ and η_t are independent error terms, distributed as standard Gaussian variables. For any $t = t_1, \ldots, t_N$, we denote by $\hat{\psi}_t$ the column vector composed of the elements $\hat{\psi}_{iJ,t}$, $i < d$. Let p be the number of rows of $\hat{\psi}_t$ and $\Psi_t = (\hat{\psi}_{t_1}, \ldots, \hat{\psi}_t)$ be the information collected up to time t. Using the Kalman filter, the unobserved factor X can be filtered recursively given the estimated process $\hat{\psi}_t$. For any $t = t_1, \ldots, t_N$, let us denote $\bar{X}_t := \mathbb{E}(X_t \mid \Psi_t)$ the filtered version of the factor process and by $\bar{X}_{t|t-1} := \mathbb{E}(X_t \mid \Psi_{t-1})$ the best prediction of X_t given information up to time $t-1$. In the Kalman filter terminology, the first equation in (37) corresponds to the measurement

[5] $\Omega^2_{iJ,t}$ is approximated by $\hat{\Omega}^2_{iJ,t} = \hat{p}^*_{iJ,t}(1 - \hat{p}^*_{iJ,t})$. Indeed, if we consider the binomial variable : starting from i, either go to J or below, it is clear that the standard error for $\hat{p}^*_{iJ,t}$ can be calculated as a binomial standard error. see Nickell et al. (2000).

equation, the second to the transition equation.

Maximum likelihood of the filtered model

Let $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,d-1}$ be the vector of unknown model parameters. The log-likelihood function associated with the filtered Gaussian model is given by Durbin and Koopman (2012) as,

$$\log\left(\mathcal{L}(\theta, A)\right) = -\frac{Np}{2}\log(2\pi) - \frac{1}{2}\sum_{t=0}^{N}\left(\log|F_t| + e_t^T F_t^{-1} e_t\right), \tag{38}$$

For any $t = t_1, \ldots, t_N$, the column vector e_t is the prediction error and is equal to $e_t := \hat{\psi}_t - \mathbb{E}[\hat{\psi}_t \mid \Psi_{t-1}]$. The matrix F_t is the conditional variance of e_t given Ψ_{t-1}, i.e. $F_t = \mathrm{Var}(e_t \mid \Psi_{t-1})$, and $|F_t|$ is its determinant. Given the vector of "observations" Ψ_{t_N}, the prediction error e_t and the matrix F_t can be obtained at each time t as outputs of the Kalman routine.

The parameters $\hat{\theta} = \left(\hat{C}_{i+1}, \hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i\right)_{i=1,\cdots,d-1}$ are chosen as the one that maximize the log-likelihood function (38) of the filtered model. We then define the estimated probability of migrating towards a rating J or below as

$$\psi_{iJ,t}(\bar{X}_t) = \frac{C_J - \alpha_i}{\sigma_i} - \frac{1}{\sigma_i}\left\langle \beta_i, \bar{X}_t \right\rangle, \tag{39}$$

where $\bar{X}$ the filtered factor process associated with max-likelihood parameters $\hat{\theta}, \hat{A}$. The relation (39) will be used in the application on S&P data (see Subsection 5.4).

5 Empirical study

In order to perform the estimation procedures developed in sections 3 and 4, we have built a worldwide bonds portfolio, composed with a mixture of investment and non investment grade positions. The diversification of the portfolio in terms of sectors and regions will allow to compare between the two estimation approaches; estimation with observable factors and estimation with latent factors.

In this section, we describe the data and explain how we proceed to compute the transition and generator matrices in time homogeneous Markov chain (no covariates). In the second part, i.e. non-homogeneous Markov chain (with observable or latent covariates), we perform the parameters estimation on both the intensity and the structural approaches. The third part of the application deals with the comparison between models and assesses the ability of both models to adjust the migration dynamics in terms of mobility (see section 5.5.2 for more details on mobility).

5.1 Data description

Our data stands for S&P credit ratings history, it covers the period January 2006 to January 2014 with a monthly frequency observation on a portfolio of 2875 obligors. The database has a total of 16168 obligor years excluding withdrawn ratings. The overall shares of the most dominant regions in the dataset, i.e. North America, Western Europe and Asia are respectively 48.4%, 24.5% and 17.2% (see Table (1)). The split by sector is less concentrated as the top five dominant sectors are Finance, Energy, Industry, Utilities and Telecom with respectively 17.5%, 9.1%, 7%, 6% and 5.1% of the total shares (see Table (2)). We notice also that the portfolio is dynamic, as the composition is changing through time but the overall shares per rating remain stable (see Table (3)) .

The S&P ratings contain 8 classes of risk, AAA (lowest risk), AA, A, BBB, BB, B, CCC and D which stands for Default. For convenience we replace these qualitative notations with numeric equivalent, as 1 denote the AAA, 2 the AA,... until 8 for D.

region	percentage	region	percentage
North America	48.4	Central America	2.2
Western Europe	24.5	South America	1.8
Asia	17.2	Central Europe	1.7
Australia and Pacific	2.8	Africa and Middle East	1.4

Table 1: Portfolio split by regions

sector	percentage	sector	percentage	sector	percentage
Finance	17.5	Cable Media	3.7	Food Bev	3
Energy	9.1	Chemicals	3.6	Consumer	2.9
Industry	7.0	Auto-mobile	3.5	Health	2.9
Utilities	6.0	Building	3.4	Leisure	2.7
Telecom	5.1	Insurance	3.1	Transport	2.7
Technology	4.5	Retail	3.1	Others	16.2

Table 2: Portfolio split by sectors

Year	1	2	3	4	5	6	7	8	Total
2006	3.3	8.1	31.0	31.0	15.5	9.9	0.8	0.4	100
2007	3.5	9.2	29.0	31.6	16.6	8.8	0.9	0.4	100
2008	3.4	9.8	27.4	31.0	15.8	11.5	0.8	0.3	100
2009	3.6	8.7	28.7	31.4	14.3	11.6	1.2	0.5	100
2010	2.9	8.0	28.5	31.9	13.4	12.5	1.9	0.9	100
2011	3.4	7.5	28.2	32.5	13.8	13.3	1.0	0.3	100
2012	2.4	6.5	26.4	32.6	16.5	13.8	1.4	0.4	100
2013	2.5	6.2	25.7	34.0	15.9	14.1	1.4	0.2	100

Table 3: Percentage of obligors per ratings and per years

5.2 Time-homogeneous intensity model (no covariates)

In the case of homogeneous Markov chain, the transition probabilities are time-invariant. This means that both the cohort approach (see Löffler and Posch (2007)) and the intensity approach with a constant generator matrix (see (19) in section 3.2) can be used to compute the transition matrix. We have investigated both of them. The cohort approach turns out to produce a high number of null transition probabilities, which makes the calculation of quantities like $\Phi^{-1}\left(\hat{p}^*_{ij}\right)$ impossible[6]. This is not the case of the intensity approach where almost all transition probabilities were not null[7]. For instance, using a constant generator matrix on time interval $[01/2013 - 01/2014]$ leads to the following results on the 2013 one year transition matrix[8], see Table (4).

	1	2	3	4	5	6	7	8
1	87.13	12.77	0.10	0	0	0	0	0
2	0	98.57	1.39	0.04	0	0	0	0
3	0	1.57	93.59	4.78	0.06	0	0	0
4	0	0.03	3.17	94.29	2.36	0.15	0	0
5	0	0	0.13	7.56	89.50	2.71	0.08	0.02
6	0	0	0.01	0.68	9.31	83.68	4.84	1.48
7	0	0	0	0.05	0.82	14.13	59.60	25.40
8	0	0	0	0	0	0	0	100

[6] See (35) in 4.2 to get the definition of $\hat{p}^*_{ij}$.

[7] This difference is due to the fact that the cohort approach does not make full use of the available data. Specifically, the estimates of the cohort approach are not affected by the timing and sequencing of transitions within the period where the intensity approach captures within-period transitions.

[8] The probabilities lower than 0.001 are reported as null in the matrix.

Table 4: One year transition matrix on 2013

The transition matrix in Table (4) is typical of the migration dynamic behaviour. The probabilities on the diagonal are close to one which means that the ratings are quite stable during a short time period (which is the case for 1 year). The probabilities around the main diagonal (above and below) are also significant, while the other ones tend generally to zero, indeed rating migrations are usually by one grade. However, for the non investment grades, i.e. ratings 5, 6 and 7 it is common to have more heterogeneity than the investment grades. This is due in one hand to the financial situation of the firms standing at these ratings and in the other hand to the *correction effect* of agencies. When a default which has not been predicted by the rating agencies occurs, several adjustments are performed to correct for this prediction error, some downgrades are then applied to the firms which are in the same situation as the defaulted one.

If we consider a constant matrix generator on the overall study period, i.e. $[01/2006 - 01/2014]$ we get the following transition matrix that we can consider as a Through The Cycle transition matrix(see Table (5)).

	1	2	3	4	5	6	7	8
1	92.69	6.47	0.48	0.03	0.31	0.02	0	0
2	0.64	89.93	8.96	0.46	0.01	0	0	0
3	0.02	1.47	92.37	5.91	0.20	0.03	0	0
4	0	0.08	3.12	92.13	4.16	0.41	0.06	0.04
5	0.06	0.09	0.25	6.89	83.99	8.16	0.41	0.15
6	0	0	0.13	0.52	7.96	82.77	6.10	2.52
7	0	0	0.03	0.61	1.24	22.92	43.27	31.93
8	0	0	0	0	0	0	0	100

Table 5: One year Through The Cycle transition matrix

Comparing to the 2013 matrix, the TTC matrix is less concentrated on the diagonal and almost all the transition probabilities are not null. This is due to computation with respect to all the duration of data history, all the migrations are likely to happen during this time.

5.3 Estimation with observable factors

Many studies have addressed the question of selecting the appropriate covariates to explain time variation in the behaviour of ratings. This question is a large topic of investigation, one can refer to Kavvathas (2001), Bangia et al. (2002), Couderc and Renault (2005) and Koopman et al. (2009) to have more details. In this study, we have followed the idea of Couderc and Renault (2005) by selecting variables according to three covariates "families", i.e. macroeconomic information, financial and commodity markets information and credit market information. As our portfolio is mainly composed of US and Euro positions we focused on US and Euro indicators. The entire list of covariates is given below[9].

Macroeconomic variables : US and Euro Real Gross Domestic Product, US and Euro Consumer Price Indices, US and Euro Civilian Unemployment rates, US Effective Federal Funds rate (short term interest rate), Euro short term interest rates, US 10 years treasury constant maturity rate, long-term Government bond yields 10-years, US and Euro Industrial Production Indices, US and Euro Purchasing Manager Indices, Case-Shiller Index (real estate price Index for US).

Financial and commodity markets variables : Crude Oil price, Gold price, NASDAQ 100 Index, Dow Jones Index, Euro STOXX Index and Chicago Board Options Exchange Market Volatility Index.

[9]The macroeconomic indicators time series are in free access on the website of the FEDERAL RESERVE BANK OF ST. LOUIS, where the financial, commodity and credit indicators time series can be retrieved from Bloomberg.

Credit market variables : Bank of America Merrill Lynch US Corp BBB total return Index, Bank of America Merrill Lynch US Corp AA total return Index, Markit iBoxx AA and Markit iBoxx BBB.

The data time series go from 01/2006 to 01/2014 which encompasses our study period. Each variable consists in a year to year growth without overlapping between the sub-periods, this gives 8 sub-periods. For example, the US Real GDP growth for the date 01/2014 (last sub-period) is calculated as the relative variation during the period [31/12/2012 − 31/12/2013], which gives,

$$\begin{aligned}\text{Real GDP Growth}_{01/2014} &= \frac{\text{Real GDP}_{31/12/2013} - \quad \text{Real GDP}_{31/12/2012}}{\text{Real GDP}_{31/12/2012}} \\ &= \frac{15916\ b\$ - 15433\ b\$}{15433\ b\$} = 3.13\%\end{aligned}$$

As the number of covariates is important (25 series), we have performed a Principal Component Analysis (PCA) in order to reduce the number of explicative factors. Table (6) summarizes the 8 first eigenvalues with cumulated variance explained in percentage.

Eigenvalues	Percentage of variance	Cumulated percentage
8.79	43.99	43.99
3.39	16.98	60.98
2.43	12.19	73.18
1.61	8.09	81.27
0.95	4.77	86.04
0.72	3.64	89.69
0.56	2.83	92.53
0.36	1.83	94.36

Table 6: Observable factors eigenvalues

The first factor explains almost 44% of the total variance, where the rest of factors (2 until 8) explain 50% cumulated variance. As there is not a strict rule to choose the number of covariates, we have investigated the estimation procedure for different number of covariates. Above five factors the log-likelihood is no more significantly improved, we have then chosen the five first factors with a cumulated explained variance of 86% to realise the estimation procedure. The number of selected factors can be challenged in the light of a deeper investigation, but it is not the purpose of this paper.

Parameters estimation: we recall that for the intensity model, the parameters to be estimated are the baseline intensity Λ_0 and the matrices of factor sensitivities β_s, $s = 1, \cdots, 5$. We use proposition 3.4 to estimate these parameters from the panel data composed of 8 sub-periods. For the structural model, the parameters are $\theta = (C_{i+1}, \alpha_i, \beta_i, \sigma_i)_{i=1,\cdots,7}$ (see (9 - 11)).

The estimated parameters are reported in the following tables. See Tables (7), (8) below and Tables (17 - 20) in the appendix for the intensity model and Tables (9), (10) for the structural model.

	1	2	3	4	5	6	7	8
1	-0.03	0.03	0	0	0	0	0	0
2	0	-0.05	0.05	0	0	0	0	0
3	0	0.01	-0.07	0.06	0	0	0	0
4	0	0	0.03	-0.06	0.03	0	0	0
5	0	0	0	0.06	-0.13	0.07	0	0
6	0	0	0	0	0.10	-0.14	0.04	0
7	0	0	0	0	0	0.45	-0.72	0.27
8	0	0	0	0	0	0	0	0

Table 7: Baseline intensity generator for observable factors intensity model

	1	2	3	4	5	6	7	8
1	0	-0.06	0	0	0	0	0	0
2	0.34	0	-0.11	1.58	0	0	0	0
3	-4.93	0.22	0	-0.03	0.28	2.77	0	0
4	0	12.38	0.02	0	-0.06	0.07	0.80	-2.47
5	-2.42	-2.42	-0.17	0.04	0	-0.04	-2.92	-0.17
6	0	0	-7.67	1.63	0.05	0	-0.04	-0.08
7	0	0	0	3.52	0	0.09	0	-0.11
8	0	0	0	0	0	0	0	0

Table 8: Beta for the first factor intensity model

The migrations are expected to be correlated with the business cycle, negatively when the migration is a downgrade and positively when the migration is an upgrade. According the Table (8), where the first factor can be considered as a proxy of the business cycle, this behaviour is confirmed for the *short migrations* (i.e. one grade migration, for instance $\hat{\beta}_{12} = -0.06, \hat{\beta}_{23} = -0.11, \hat{\beta}_{32} = 0.22, \hat{\beta}_{21} = 0.34$). When the migration is farther (i.e. more than two grades), due to the lack of observations, it is no more possible to state any significant relation with the business cycle. When we focus on the non investment grade obligors (ratings 5, 6 and 7), we can observe that, all downgrades are negatively correlated to the business cycle, whatever the amplitude of the migration.

$\hat{C}$	$\hat{\alpha}$	$\hat{\sigma}$
$\hat{C}_2$ = 5.16	$\hat{\alpha}_1$= 5.08	$\hat{\sigma}_1$= 0.17
$\hat{C}_3$ = 4.35	$\hat{\alpha}_2$=3.83	$\hat{\sigma}_2$= 0.38
$\hat{C}_4$ = 3.47	$\hat{\alpha}_3$= 3.14	$\hat{\sigma}_3$= 0.74
$\hat{C}_5$ = 2.16	$\hat{\alpha}_4$= 2.47	$\hat{\sigma}_4$= 1.18
$\hat{C}_6$ = 1.05	$\hat{\alpha}_5$= 1.52	$\hat{\sigma}_5$= 1.17
$\hat{C}_7$ = -0.19	$\hat{\alpha}_6$=0.63	$\hat{\sigma}_6$= 1.17
$\hat{C}_8$ = -1.55	$\hat{\alpha}_7$=-1.01	$\hat{\sigma}_7$= 1.72

Table 9: Parameters for observable factors structural model

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$	$\hat{\beta}_5$
$\hat{\beta}_{11}$=0.01	$\hat{\beta}_{21}$= 0.03	$\hat{\beta}_{31}$=0.05	$\hat{\beta}_{41}$=-0.06	$\hat{\beta}_{51}$=0.27
$\hat{\beta}_{12}$=0.03	$\hat{\beta}_{22}$=0.02	$\hat{\beta}_{32}$=-0.02	$\hat{\beta}_{42}$=0.09	$\hat{\beta}_{52}$=-0.25
$\hat{\beta}_{13}$=0.06	$\hat{\beta}_{23}$= 0.06	$\hat{\beta}_{33}$=-0.06	$\hat{\beta}_{43}$=0.08	$\hat{\beta}_{53}$=-0.17
$\hat{\beta}_{14}$=0.06	$\hat{\beta}_{24}$=0.04	$\hat{\beta}_{34}$=0.00	$\hat{\beta}_{44}$=-0.01	$\hat{\beta}_{54}$=0.00
$\hat{\beta}_{15}$=0.04	$\hat{\beta}_{25}$=0.05	$\hat{\beta}_{35}$=-0.06	$\hat{\beta}_{45}$=-0.03	$\hat{\beta}_{55}$=0.01
$\hat{\beta}_{16}$=0.07	$\hat{\beta}_{26}$=0.13	$\hat{\beta}_{36}$=0.14	$\hat{\beta}_{46}$=-0.19	$\hat{\beta}_{56}$=0.40
$\hat{\beta}_{17}$=0.14	$\hat{\beta}_{27}$=0.23	$\hat{\beta}_{37}$=0.22	$\hat{\beta}_{47}$=0.07	$\hat{\beta}_{57}$=0.38

Table 10: Beta for observable factors structural model

The results of the Tables (9) and (10) are consistent with those obtained by Gagliardini and Gourieroux (2005) as the estimated barrier levels $\hat{C}$ increase with respect to rating quality. The same conclusion holds for the intercepts $\hat{\alpha}_i$ which confirms that the downgrade risk is higher for low quality credit. The estimated rating volatilities are consistent with the fact that the non investment grade firms are likely to move in a larger set of ratings. Indeed, $\hat{\sigma}_7 = 1.72$ and $\hat{\sigma}_6 = 1.17$, whereas $\hat{\sigma}_1 = 0.17$ and $\hat{\sigma}_2 = 0.38$. One should also notice that the $\hat{\beta}$ coefficients for the non investment grade ratings 6 and 7 are significantly higher than for the investment grade firms. This confirms that the weakest firms are more subject to the uncertainty of the economy.

5.4 Estimation with unobservable latent factors

The use of latent factors to explain time variation of the credit ratings requires to compute a relatively high number of empirical transition intensities $\hat{\lambda}_{ij,t}$, $i \in \{1, \cdots, 7\}$, $j \in \{1, \cdots, 8\}$, $i \neq j$. Indeed, $\hat{\lambda}_{ij,t}$ are used as inputs for both models, they are transformed into $\hat{\psi}_{iJ,t}$ for the structural model (see next section) and kept as such for the intensity model. In order to increase the number of observations, we have considered an overlapping year to year sub-period with one month slope. With 8 years of monthly frequency data, this allows to have 86 observations instead of 8 observations if the sub-periods were not overlapping[10]. For instance, $\hat{\lambda}_{ij,t_1}$ is calculated on the sub-period $[01/2006 - 01/2007]$, $\hat{\lambda}_{ij,t_2}$ is calculated on the sub-period $[02/2006 - 02/2007]$, etc. We recall that $\hat{\lambda}_{ij,t}$ are computed using the proposition 3.3.

5.4.1 Determining the number of unobservable factors

Since the covariates are unobserved we proceed by using the method proposed in Gagliardini and Gourieroux (2005) to find the number of covariates in both structural and intensity models. The idea is to build time series of $\hat{\lambda}_{ij,t}$ and $\hat{\psi}_{iJ,t}$ for $t = t_1, \cdots, t_{86}$, to consider them as the realisations of random variables and then carry on a PCA to find the number of common factors.

For the intensity model, we count 49 variables $\hat{\lambda}_{12,t}, \hat{\lambda}_{13,t}, \cdots, \hat{\lambda}_{78,t}$, each one going from t_1 to t_{86} which makes 49 time series. The PCA is then applied on this set to compute the common factors. Concerning the structural model, the $\hat{\lambda}_{ij,t}$ are transformed into empirical transition probabilities $\hat{p}_{ij,t}$ using (5), then the series $\hat{\psi}_{iJ,t}$ such as $\hat{\psi}_{iJ,t} = \Phi^{-1}\left(\hat{p}^*_{iJ,t}\right)$ for $t = t_1, \cdots, t_{86}$ are computed (see (35) in 4.2 to get the definition of $\hat{p}^*_{iJ,t}$). We get also 49 times series on which we apply the PCA.

The corresponding eigenvalues are given in decreasing order in the following table.

intensity eigenvalues	Cumulated percentage	structural eigenvalues	Cumulated percentage
1.83	79.50	1853.46	48.89
0.43	98.28	735.23	72.29
0.02	99.27	515.81	81.52
0.007	99.58	311.41	89.04
0.003	99.71	306.01	93.81
0.002	99.82	247.94	96.50

Table 11: Eigenvalues

For the $\hat{\lambda}_{ij,t}$ time series, the first two factors explain 98% of the total variance. The variance of $\hat{\psi}_{iJ,t}$ time series are explained at 72% by the two first factors. We have applied the Kalman filter for one common factor and two common factors, we found that the filtered process was the same in both cases. We then consider that the first factor represent the business cycle. Once the number of covariates known, we have done the parameters estimation using the two steps procedure, filtering of the process $\bar{X}_t$ and maximisation of the log-likelihood (see sections 3.3 and 4.2). We present and discuss the results for the one factor models below.

5.4.2 Results and discussion

We report in the Figures (1) and (2) the filtered processes $\mathbb{E}(X_t \mid Y_t)$ and $\mathbb{E}(X_t \mid \Psi_t)$ respectively[11], for $t = t_1, \cdots, t_{86}$. Recall that $\mathbb{E}(X_t \mid Y_t)$ and $\mathbb{E}(X_t \mid \Psi_t)$ correspond to the filtered factor process associated with max-likelihood models (see (3.4) and (4.2)). The figures below clearly show periods of

[10] We need a high number of observations in order to perform the Kalman filter, we have performed the Kalman filter on 8 observations but we didn't get a satisfying result.

[11] Note that the process $\mathbb{E}(X_t \mid \Psi_t)$ has been centered in the Figure (2) in order to be easily interpreted.

peak, *normal* times and *trough* which are properties of the business cycle. Our period of study encompasses what it is known as the *Liquidity Crisis* and *Sovereign Debt Crisis*, with some famous events like the Lehman Brothers bankruptcy on September 2008 and the falls of market indices like CAC40, DAX and FTSE100 during the summer 2011. These crisis are perfectly captured by the structural model while the intensity model fails to capture the sovereign debt crisis. The study of the default probability of non investment grade ratings shows two peaks during the periods [01/2009 – 01/2010] and [01/2012 – 01/2013] which correspond to the *trough* periods of the business cycle (see Figures (1), (2), (3) and (4)).

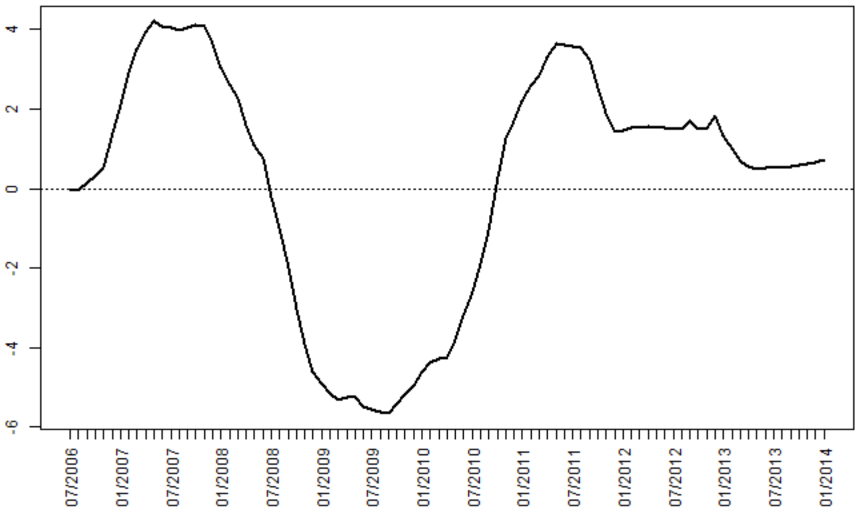

Figure 1: Filtered common factor in the intensity model

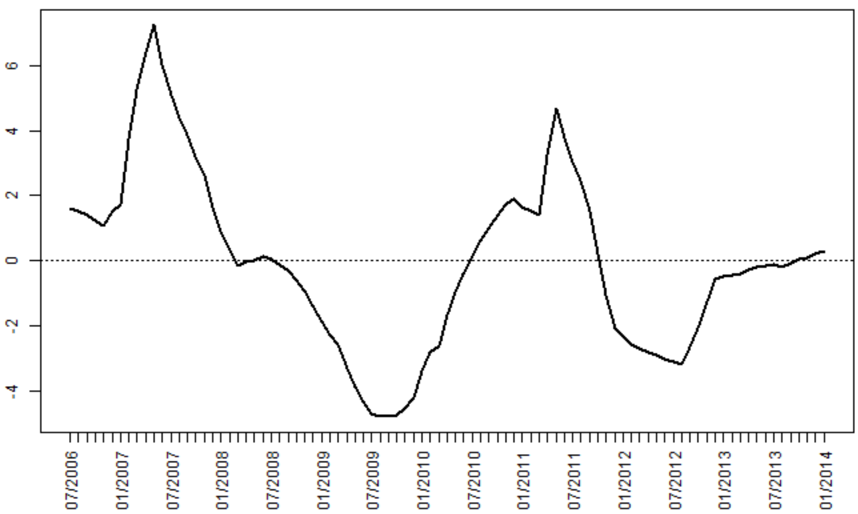

Figure 2: Filtered common Factor in structural Model

Parameters estimation: the estimated parameters for the one factor models are reported in the following tables. See Tables (12) and (13) for the one factor intensity parameters and Table (14) for the one factor structural parameters.

	1	2	3	4	5	6	7	8
1	-0.03	0.03	0	0	0	0	0	0
2	0	-0.08	0.08	0	0	0	0	0
3	0	0	-0.07	0.07	0	0	0	0
4	0	0	0.03	-0.07	0.04	0	0	0
5	0	0		0.07	-0.17	0.1	0	0
6	0	0	0	0	0.07	-0.15	0.08	0.
7	0	0	0	0	0	0.27	-0.67	0.42
8	0	0	0	0	0	0	0	0

Table 12: Baseline intensity for one factor intensity model

	1	2	3	4	5	6	7	8
1	0	-0.25	0	0	0	0	0	0
2	0.43	0	-0.12	0.20	0	0	0	0
3	0.02	0.19	0	-0.05	0.16	0.05	0	0
4	0	0.07	0.03	0	-0.06	0.06	0.05	-0.23
5	-0.16	-0.27	-0.23	0.03	0	-0.07	-0.46	-0.18
6	0	0	-0.25	0.12	0.12	0	-0.15	-0.28
7	0	0	0	-0.19	0	-0.03	0	-0.11
8	0	0	0	0	0	0	0	0

Table 13: Beta for one factor intensity model

$\hat{C}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\beta}$
$\hat{C}_2$ = 6.82	$\hat{\alpha}_1$= 5.01	$\hat{\sigma}_1$= 1.71	$\hat{\beta}_1$= 0.43
$\hat{C}_3$ = 4.50	$\hat{\alpha}_2$= 3.12	$\hat{\sigma}_2$= 0.76	$\hat{\beta}_2$= 0.10
$\hat{C}_4$ = 3.12	$\hat{\alpha}_3$= 2.56	$\hat{\sigma}_3$= 0.69	$\hat{\beta}_3$= -0.06
$\hat{C}_5$ = 1.84	$\hat{\alpha}_4$= 1.83	$\hat{\sigma}_4$= 0.69	$\hat{\beta}_4$= -0.05
$\hat{C}_6$ = 0.96	$\hat{\alpha}_5$= 1.23	$\hat{\sigma}_5$= 0.75	$\hat{\beta}_5$= 0.03
$\hat{C}_7$ = 0.001	$\hat{\alpha}_6$= 0.12	$\hat{\sigma}_6$= 0.87	$\hat{\beta}_6$= 0.03
$\hat{C}_8$ = -0.43	$\hat{\alpha}_7$= -0.56	$\hat{\sigma}_7$= 1.09	$\hat{\beta}_7$= -0.03

Table 14: Parameters for 1 factor structural model

The parameters obtained for both models are consistent with those obtained in the case of observable factors. However, the values $\hat{\beta}_1 = 0.43$ and $\hat{\sigma}_1 = 1.71$ are not in line with the previous results. This is more a numerical "issue" than a interpretable result.

5.5 Comparison between models

As stated previously, the aim of this paper is to assess and compare the migration models on their ability to link the transition probabilities to dynamic risk factors. In order to achieve this objective, we proceed in two steps. First, we compare the one year default probabilities implied by each model during the study period with the empirical default probabilities. Second, we compare transition matrices of each model with respect to a metric called SVD mobility index and introduced in Jafry and Schuermann (2004). The mobility index is a function that measures the ability of a transition matrix to generate migration events (the mobility index is detailed in the section 5.5.2). The model parameters used for the computation of one-year default probabilities and mobility indices are those obtained from the estimation procedures of the two previous sections (observable and latent covariates). The comparison is made with respect to a credit migration model with no covariate (called the empirical model) where, on each sub-period, the transition matrix is estimated using the procedure described in proposition 3.3 (see also section 5.2). Recall that we have 8 non overlapping sub-periods for the observable factors and 86 overlapping sub-periods for the unobservable factors.

5.5.1 Comparison using the implied one year default probabilities

We compute the one year matrices implied on each sub-period by both intensity and structural models. We present graphically the evolution of the implied default probabilities of the ratings 4, 5, 6 and 7 (the default probabilities of the ratings 1, 2, and 3 are almost null).

Observable factors: Figures (3) and (4) represent the estimated one year default probabilities in the empirical model (solid line), in the intensity model (dashed line) and in the structural model (dotted line). One can see on the Figure (3) that for ratings 4 and 5 the intensity model adjusts perfectly the empirical default probabilities, the curves are almost identical. The structural model over-estimates the default probabilities for the rating 5, when the curve of the default probabilities for the rating 4 is almost flat.

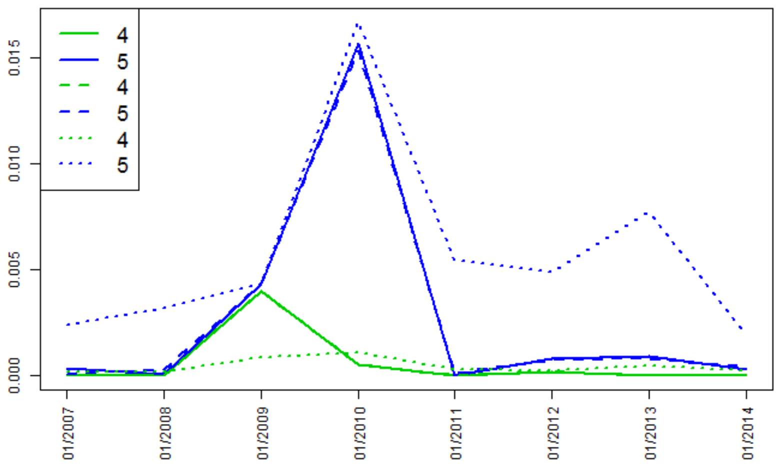

Figure 3: Ratings 4 and 5: implied one year default probabilities. Model estimation performed in the case of observable factors. Empirical model in solid line, intensity model in dashed line, structural model in dotted line.

On the Figure (4) the same observation is made with regards to ratings 6 and 7. The intensity model fits very well the empirical default probabilities when the structural model over-estimates them.

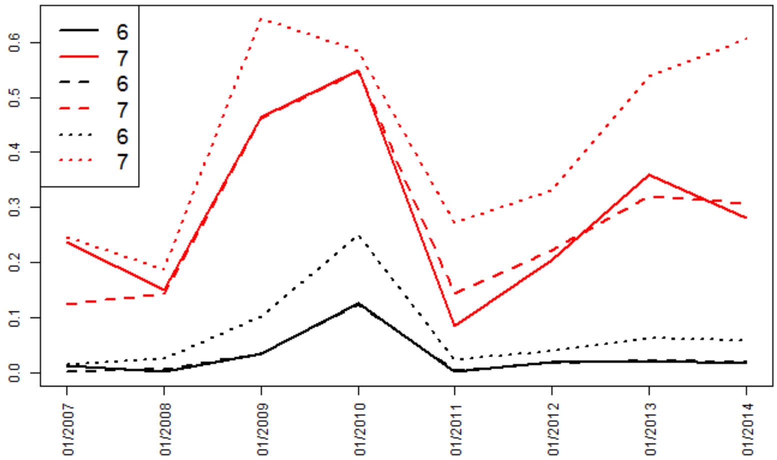

Figure 4: Ratings 6 and 7: implied one year default probabilities. Model estimation performed in the case of observable factors. Empirical model in solid line, intensity model in dashed line, structural model in dotted line.

Unobservable factors: Figure (5) shows that both models are not able to reproduce the one year empirical PD during the *trough* period, see the period $[01/2009 - 07/2010]$. Concerning the rest of the study period, the structural model over-estimates the PD for the rating 5 where it shows an almost flat curve for the PD rating 4. The PD curves of the intensity model (for ratings 4 and 5) are most of the time under the empirical PD.

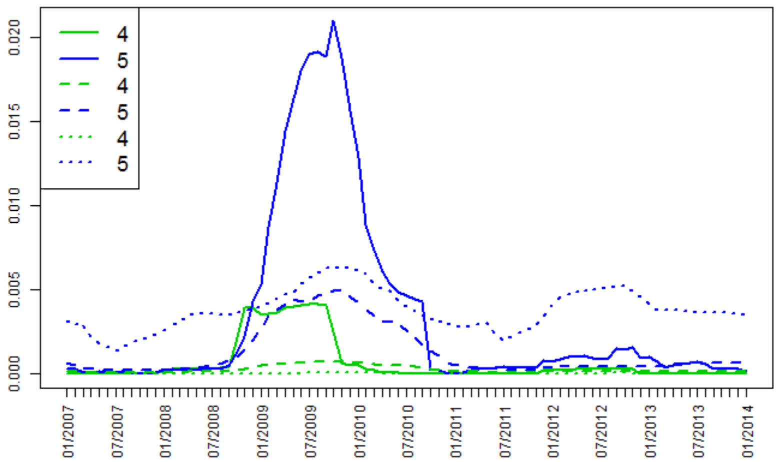

Figure 5: Ratings 4 and 5: implied one year default probabilities. Model estimation performed in the case of unobservable factors. Empirical model in solid line, intensity model in dashed line, structural model in dotted line.

Concerning the one year PD for the ratings 6 and 7 (see Figure (6)), the structural model shows a quasi non sensitivity to the business cycle. Indeed, the curves are high and almost flat comparing to the empirical one year PD. The intensity model shows a better adjustment in the sense that it is more reactive than the structural model to the business cycle but it is still unable to reproduce the empirical PD during the *trough* period.

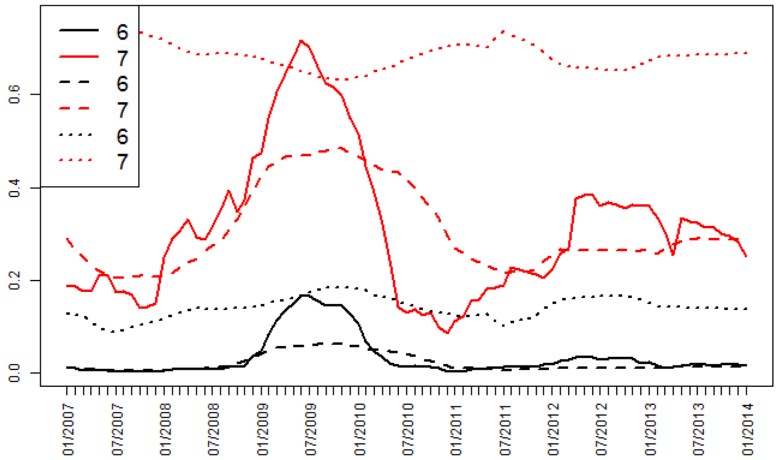

Figure 6: Ratings 6 and 7: implied one year default probabilities. Model estimation performed in the case of unobservable factors. Empirical model in solid line, intensity model in dashed line, structural model in dotted line.

5.5.2 Comparison using the SVD mobility index

In literature, comparing two transition matrices is usually done using three methods, the euclidean distance (see Israel et al. (2001)), statistic testing, t-test or χ^2 (see Nickell et al. (2000), Foulcher et al. (2004)) or with a mobility index (see Geweke et al. (1986), Jafry and Schuermann (2004)). The mobility index was introduced to compare two transition matrices on their ability to generate migration events. It is defined as a function $M(\Pi)$ of $\mathbb{R}^{d\times d} \longrightarrow \mathbb{R}$ where by convention $M(I) = 0$, with I the identity matrix. A good mobility index must be assessed on three criteria, persistence, convergence and temporal aggregation (see the well documented papers of Shorrocks (1978), Geweke et al. (1986) and Jafry and Schuermann (2004) to have a complete overview of this topic). The interest of the mobility index, is that instead of comparing the transition probabilities individually it compares the matrices according to their overall dynamic. We present hereafter the SVD mobility index introduced in Jafry and Schuermann (2004), we will use it as a metric to compare the matrices implied by the intensity and the structural models.

For a d-dimensional square matrix Π, the SVD mobility index is given by,

$$M_{SVD}(\Pi) = \frac{\sum_{i=1}^{d} \sqrt{\lambda_i\left(\tilde{\Pi}'\tilde{\Pi}\right)}}{d}, \tag{40}$$

where $\tilde{\Pi} = \Pi - I$ and $\tilde{\Pi}'$ is its transpose. $\lambda_i\left(\tilde{\Pi}'\tilde{\Pi}\right)$ is the i^{th} eigenvalue of $\tilde{\Pi}'\tilde{\Pi}$, sorted in decreasing order, i.e. $\lambda_1(\tilde{\Pi}'\tilde{\Pi}) > \cdots > \lambda_d(\tilde{\Pi}'\tilde{\Pi})$.

We have applied the SVD mobility index on each matrix calculated according to the intensity and structural models. Figures (7) and (8) show the evolution of the SVD mobility index of the empirical model (black curve), of the on intensity model (red curve) and of the structural model (blue curve). Figure (7) corresponds to estimations with observable factors whereas Figure (8) corresponds to estimations with unobservable factors. The two figures confirm the statements made previously : the intensity model with observable factors fits very well the empirical transition matrices whereas the other models, i.e., the structural with observable factors, the intensity with latent covariates and the structural with latent covariates, over-estimates, under-estimates and over-estimates respectively the empirical transition matrices.

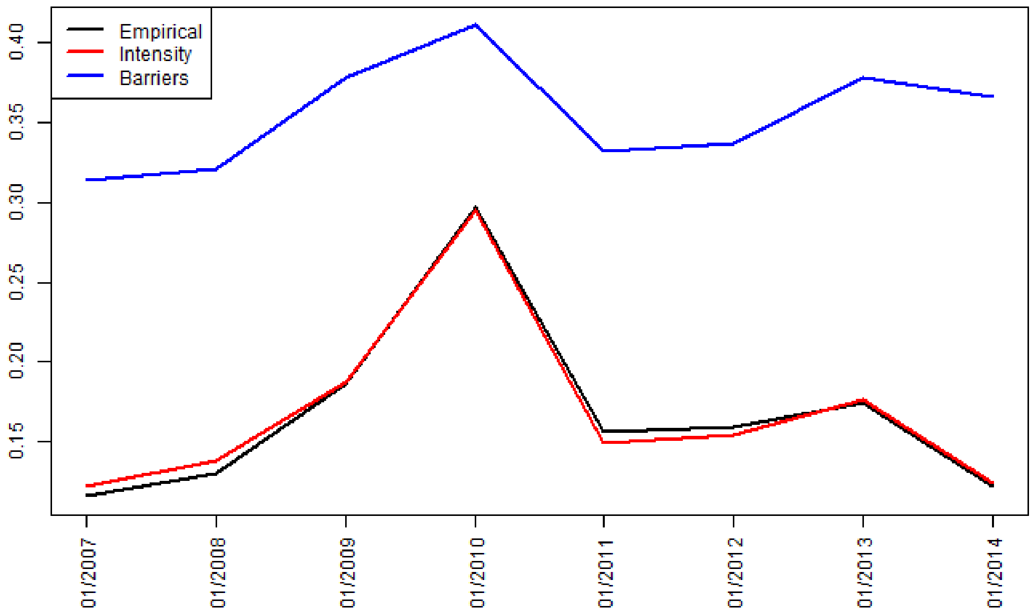

Figure 7: Implied SVD mobility index implied when model estimation is performed in the case of observable factors. Empirical model in black, intensity model in red, structural model in blue.

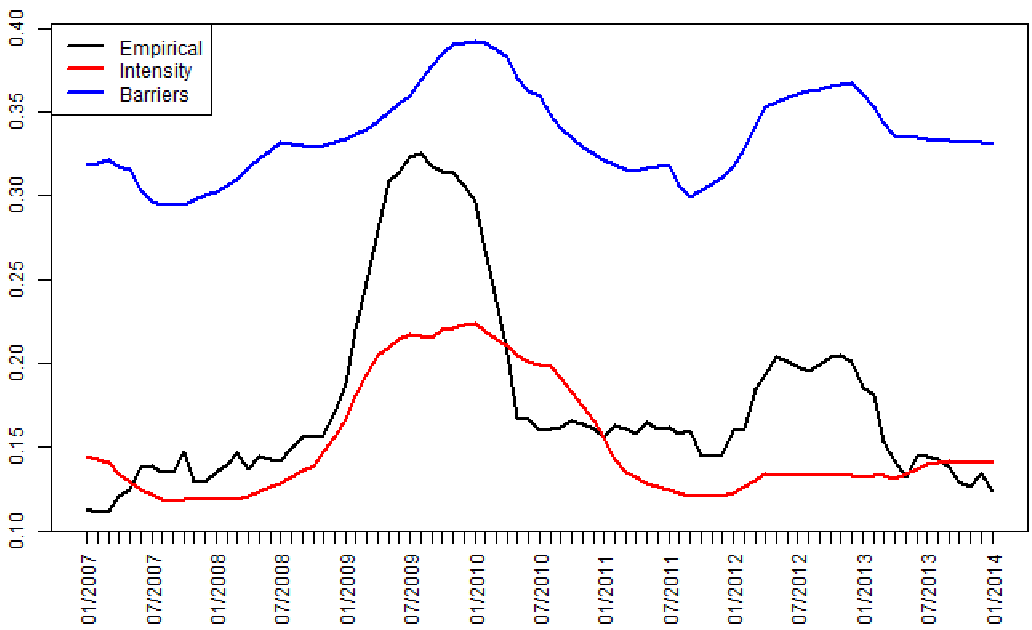

Figure 8: Implied SVD mobility index implied when model estimation is performed in the case of unobservable factors. Empirical model in black, intensity model in red, structural model in blue.

As a conclusion for the section 5.5 we can state that the intensity model with observable factors is the one that best fits the empirical probabilities. This is mainly due to following points:

- for the intensity model with observable factors, the parameters $\lambda_{ij,0}$ and β_{ij} are estimated through the maximization of the marginal likelihood $\mathcal{L}_{ij}$. This means that among the set of explaining factors, it is most likely to have at least one factor that explains correctly the intensity migration λ_{ij}. The estimation of the structural model parameters is done through the maximization of the global likelihood function, this is less tractable than the intensity model in the sens that it is not possible to explain separately the transition probabilities.

- for the intensity model with observable factors, the maximum likelihood problem is solved using a semi-analytical solution which is not the case of the other models. Indeed the baseline intensity estimator $\hat{\lambda}_{ij,0}$ is given analytically when the regression coefficients $\hat{\beta}$ are obtained through the resolution of the gradient of the log-likelihood function which represents a simple non-linear equations system. The parameters of the structural model with observable factors are also estimated through the resolution of the gradient of the log-likelihood function but the non-linear equations system is far more difficult to resolve.

- for the statistical estimation using unobservable factors, the solution of Gagliardini and Gourieroux (2005) to skip the issue of high dimension integrals requires to build a linear Gaussian model to apply the Kalman filter. The assumption of an infinite number of obligors n and the asymptotic normality of the maximum likelihood estimator at each time t and for every migration (i,j) is not satisfied. Indeed, even with a very high value of n, some migrations are unlikely to happen frequently like the long migrations (three grades or more). The intensity model with observable factors use directly the observations without making some approximations based on very restrictive assumptions.

- the latent factors models have more parameters to estimate (the filtered process $\bar{X}_t$ and the autoregression matrix A). This constraint added to the limited quantity of data plays an important role in the final results.

6 Conclusion

The aim of this paper is to assess and compare two alternative stochastic migration models on their ability to link the transition probabilities to either observable or unobservable dynamic risk factors. In this respect, we use the same data set (S&P credit ratings on the period $[01/2006 - 01/2014]$) to estimate model parameters in the multi-state latent factor intensity model and in the ordered Probit model. The estimation procedure is detailed in the two cases where the underlying dynamic factors are assumed to be either observable or unobervable. When the underlying factors are unobservable, we adapt a method given in Gagliardini and Gourieroux (2005) to represent the considered factor migration model as a linear Gaussian model. In that case, we identify the business cycle as the factor process filtered by a standard Kalman filter. The estimation methods are compared on their ability to fit the one-year empirical default probabilities and their ability to reproduce the empirical SVD mobility index. We conclude that the intensity model with observable factors was the one which best fits the S&P rating history. In line with Kavvathas (2001), this study shows that short migrations of investment grade firms are significantly correlated to the business cycle whereas, because of lack of observations, it is not possible to state any relation between long migrations (more than two grades) and the business cycle. Concerning non investment grade firms, downgrade migrations are negatively related to business cycle whatever the amplitude of the migration.

References

Albanese, C., Campolieti, G., Chen, O., and Zavidonov, A. (2003). Credit barrier models. *Risk Magazine*, 16:109–113.

Bangia, A., Diebold, F., and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance*, 26:445–474.

Couderc, F. and Renault, O. (2005). Time-to-default: Life cycle, global and industry cycle impacts. *FAME Research Paper*, 142.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods, Second Edition*. Oxford University Press.

Feng, D., Gourieroux, C., and Jasiak, J. (2008). The ordered qualitative model for credit rating transition. *Journal of Empirical Finance*, 15:111–130.

Foulcher, S., Gourieroux, C., and Tiomo, A. (2004). Term structure of defaults and ratings. *Insurance and Risk Management*, 72.

Gagliardini, P. and Gourieroux, C. (2005). Stochastic migration models. *Journal of Financial Econometrics*, 3:188–226.

Geweke, J., Marshall, R., and Zarkin, G. (1986). Mobility indices in continuous time Markov chains. *Econometrica*, 54:1407–1423.

Gourieroux, C. and Tiomo, A. (2007). *Risque de crédit : une approche avancée*. Economica.

Hougaard, P. (2000). *Analysis of Multivaiate Survival Data*. Springer.

Israel, R., Rosenthal, J., and Wei, J. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11:245–265.

Jafry, Y. and Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking and Finance*, 28:2603–2639.

Kavvathas, D. (2001). Estimating credit rating transition probabilities for corporate bonds. *University of Chicago, Working Paper.*

Koopman, S., Lucas, A., and Monteiro, A. (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142:399–424.

Koopman, S., R.Krausell, Lucas, A., and Monteiro, A. (2009). Credit cycles and macro fundamentals. *Journal of Empirical Finance*, 16:42–54.

Lando, D. and Skodeberg, T. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance*, 26:423–444.

Löffler, G. and Posch, P. (2007). *Credit Risk Modelling using VBA and Excel.* Wiley Finance.

Naldi, M., Owen, A., Pederson, C., Rees, S., and Tchangue, S. (2011). Quantitative credit quarterly : Transition matrix estimation. *Barclays Capital.*

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking and Finance*, 24:203–227.

Shorrocks, A. (1978). The measurement of mobility. *Econometrica*, 46:1013–1024.

Appendices

A Proof of Proposition 3.1

We recall the model specification, for each obligor $l = 1, \cdots, n$;

- we denote R_u^l the random variable representing the rating of obligor l at time u, $R_u^l \in \{1, \cdots, d\}$, where d is the default state and 1 the best credit quality state;
- we fix a finite horizon date $t > 0$. The underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is endowed with some reference filtration $\mathbb{F} = (\mathcal{F}_u)_{0\leq u\leq t}$ and is sufficiently rich to support the stochastic processes $\left\{R_u^l, 0 \leq u < t\right\}$ and $\{X_u, 0 \leq u \leq t\}$;
- let $\mathcal{H}^l_{u(0\leq u<t)}$ be the filtration associated to the process $\left\{R_u^l, 0 \leq u < t\right\}$ for every value of $l = 1, \cdots, n$. We consider a global filtration $\mathcal{H}_{u(0\leq u<t)}$ which is the union of all the filtrations $\mathcal{H}^l_u$, that is, $\mathcal{H}^l_u = \mathcal{H}^1_u \vee \cdots \vee \mathcal{H}^n_u$;
- let $\mathcal{G}_{u(0\leq u\leq t)}$ be the filtration associated to the process $\{X_u, 0 \leq u \leq t\}$. We assume that conditional on $\mathcal{G}_t$, the processes $\left\{R_u^l, 0 \leq u < t\right\}$ are independent across obligors;
- we assume that intensities are adapted to the filtration $\mathcal{G}_{u(0\leq u\leq t)}$ and independent from the filtration $\mathcal{H}_{u(0\leq u<t)}$.
- we denote N_l the number of ratings visited during the period $[0, t)$ by the obligor l, with the condition $N_l \geq 1$. $[0, t)$ is thus divided into a sequence of intervals $\left[t^l_{k-1}, t^l_k\right)$, $k = 1, \cdots, N_l$, where $t^l_0 = 0$ and $t^l_{N_l} = t$;
- each period $\left[t^l_{k-1}, t^l_k\right)$, $k = 1, \cdots, N_l$, corresponds to the visiting of the state r^l_k, we then have the visited states $r^l_1, \cdots, r^l_{N_l}$. Note that each consecutive states are obligatory different, thus $r^l_{k-1} \neq r^l_k$;
- there are $N_l - 1$ migration events for each obligor, each migration event takes place at the time t^l_k for $k = 1, \cdots, N_l - 1$;

- we denote $\{r_u^l, 0 \le u < t\}$ the observed path of ratings of obligor l during the period $[0,t)$;

Proof. We know that,

$$\begin{aligned}\mathcal{L}(\theta \mid \mathcal{G}_t) &= P\left(R_u^l = r_u^l, u \in [0,t), l = 1, \cdots, n \mid \mathcal{G}_t\right) \\ &= \prod_{l=1}^{n} P\Big(R_u^l = r_u^l, u \in [0,t) | \mathcal{G}_t\Big)\end{aligned} \tag{41}$$

by conditional independence of $(R_u^l)_{0<u\le t}$, $l = 1, \ldots, n$ given $\mathcal{G}_t$.

$$\begin{aligned}P\left(R_u^l = r_u^l, u \in [0,t) | \mathcal{G}_t\right) = P\Big(&\bigcap_{k=1}^{N_l-1} \left\{\left(R_u^l = r_k^l\right) \cap \left(R_{t_k}^l = r_{k+1}^l\right), u \in \left[t_{k-1}^l, t_k^l\right)\right\}, \\ &\cap \left\{R_u^l = r_{N_l}^l, u \in \left[t_{N_l-1}^l, t_{N_l}^l\right)\right\} \mid \mathcal{G}_t\Big)\end{aligned} \tag{42}$$

The equation (42) describes the likelihood of the entire path of the obligor l. The obligor starts in the rating r_k^l at time t_{k-1}^l, it spends the period $\left[t_{k-1}^l, t_k^l\right)$ in this state until it jumps to the rating r_{k+1}^l at time t_k. The last period $\left[t_{N_l-1}^l, t_{N_l}^l\right)$ is particular because we only observe the rating R_u^l until time $t_{N_l}^{l-}$ (which corresponds to time t^-). We can only conclude on the stay of the obligor l in the state $r_{N_l}^l$.

$$\begin{aligned}&P\left(R_u^l = r_u^l, u \in [0,t) | \mathcal{G}_t\right) \\ &= \mathbb{E}\left[\mathbb{1}_{\left(\bigcap_{k=1}^{N_l-1}\left\{\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right\}\cap\left\{R_u^l = r_{N_l}^l, u\in\left[t_{N_l-1}^l, t_{N_l}^l\right)\right\}\right)} \mid \mathcal{G}_t\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\left(\bigcap_{k=1}^{N_l-1}\left\{\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right\}\cap\left\{R_u^l = r_{N_l}^l, u\in\left[t_{N_l-1}^l, t_{N_l}^l\right)\right\}\right)} \mid \mathcal{G}_t \vee \mathcal{H}_{t_{N_l-1}}^l\right] \mid \mathcal{G}_t\right] \\ &= \mathbb{E}\left[\prod_{k=1}^{N_l-1} \mathbb{1}_{\left(\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right)} \times P\left(R_u^l = r_{N_l}^l, u \in \left[t_{N_l-1}^l, t_{N_l}^l\right) \mid \mathcal{G}_t, R_{t_{N_l-1}}^l = r_{N_l}^l\right) \mid \mathcal{G}_t\right] \\ &= \mathbb{E}\left[\prod_{k=1}^{N_l-1} \mathbb{1}_{\left(\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right)} \times\ e^{-\int_{t_{N_l-1}^l}^{t_{N_l}^l} \lambda_{r_{N_l}^l}(X_u)\, du} \mid \mathcal{G}_t\right].\end{aligned}$$

The Tower Property is used with respect to $\mathcal{G}_t \vee \mathcal{H}_{t_{N_l-1}}^l$ in order to exhibit the probability of staying in the rating $r_{N_l}^l$ during the period $\left[t_{N_l-1}^l, t_{N_l}^l\right)$. The rest of the path is also obtained with a recursive use of the Tower Property by stepping back each time with one period on $\mathcal{H}_u^l$. We show the case of $\mathcal{G}_t \vee \mathcal{H}_{t_{N_l-2}}^l$ and deduce the rest to get (13).

$$\begin{aligned}&P\left(R_u^l = r_u^l, u \in [0,t) | \mathcal{G}_t\right) \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{\left(\bigcap_{k=1}^{N_l-1}\left\{\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right\}\right)} \mid \mathcal{G}_t \vee \mathcal{H}_{t_{N_l-2}}\right] e^{-\int_{t_{N_l-1}^l}^{t_{N_l}^l} \lambda_{r_{N_l}^l}(X_u)\, du} \mid \mathcal{G}_t\right] \\ &= \mathbb{E}\Bigg[\prod_{k=1}^{N_l-2} \mathbb{1}_{\left(\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right)} \times \\ &\quad P\left(\left(R_u^l = r_{N_l-1}^l\right) \cap \left(R_{t_{N_l-1}}^l = r_{N_l}^l\right), u \in \left[t_{N_l-2}^l, t_{N_l-1}^l\right) \mid \mathcal{G}_t, R_{t_{N_l-2}}^l = r_{N_l-1}^l\right) e^{-\int_{t_{N_l-1}^l}^{t_{N_l}^l} \lambda_{r_{N_l}^l}(X_u)\, du} \mid \mathcal{G}_t\Bigg] \\ &= \mathbb{E}\left[\prod_{k=1}^{N_l-2} \mathbb{1}_{\left(\left(R_u^l = r_k^l\right)\cap\left(R_{t_k}^l = r_{k+1}^l\right), u\in\left[t_{k-1}^l, t_k^l\right)\right)} \times \lambda_{r_{N_l-1}^l, r_{N_l}^l}\left(X_{t_{N_l-1}}\right) e^{-\int_{t_{N_l-2}^l}^{t_{N_l-1}^l} \lambda_{r_{N_l-1}^l}(X_u) du}\ e^{-\int_{t_{N_l-1}^l}^{t_{N_l}^l} \lambda_{r_{N_l}^l}(X_u)\, du} \mid \mathcal{G}_t\right]\end{aligned}$$

Stepping back recursively on $\mathcal{H}_u^l$ leads to get,

$$P\left(R_u^l = r_u^l, u \in [0,t) | \mathcal{G}_t\right) = \prod_{k=1}^{N_l} \lambda_{r_k^l, r_{k+1}^l}(X_{t_k^l})\ e^{-\int_{t_{k-1}^l}^{t_k^l} \lambda_{r_k^l}(X_u)\,du}, \tag{43}$$

with the convention $\lambda_{r_k^l, r_{k+1}^l} = 1$ for $k = N_l$. □

B Proof of Proposition 3.4

Proof. Contrary to the homogeneous case, the maximum likelihood estimator is not only related to $\lambda_{ij,0}$ but to $\lambda_{ij,0}$ and the vector β_{ij}, we denote our vector parameters $\theta = (\lambda_{ij,0}, \beta_{1,ij}, \beta_{2,ij}, \cdots, \beta_{K,ij})$ and look to estimate θ such as,

$$\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \theta} = 0 \Rightarrow \left\{ \frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \lambda_{ij,0}}, \frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{ij}} \right\} = 0.$$

$$\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \lambda_{ij,0}} = \sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l) \frac{\partial \log\left(\lambda_{ij}(X_{t_k^l})\right)}{\partial \lambda_{ij,0}} - \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l) \int_{t_{k-1}^l}^{t_k^l} \frac{\partial \lambda_{ij}\left(X_u\right)}{\partial \lambda_{ij,0}}\,du.$$

We recall that,

$$\lambda_{ij}\left(X_t\right) = \lambda_{ij,0}\ e^{\langle \beta_{ij}, X_t\rangle} \quad \text{and} \quad \frac{\partial \lambda_{ij}\left(X_t\right)}{\partial \lambda_{ij,0}} = e^{\langle \beta_{ij}, X_t\rangle}.$$

$$\begin{aligned}\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \lambda_{ij,0}} &= \sum_{l=1}^{n}\sum_{k=1}^{N_l} \frac{Y_{ij}^l(t_k^l)}{\lambda_{ij,0}} - \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l) \int_{t_{k-1}^l}^{t_k^l} e^{\langle \beta_{ij}, X_u\rangle}\,du \\ &= 0\end{aligned}$$

we get,

$$\lambda_{ij,0} = \frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l) \int_{t_{k-1}^l}^{t_k^l} e^{\langle \beta_{ij}, X_u\rangle}\,du} \tag{44}$$

Getting vector $\hat{\beta}_{ij}$ goes through solving equation system $\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{ij}} = 0$, every partial derivative will result in a relation between vector β_{ij} and past realised covariates values (i.e. $X_{s,t}$) as well as migration observations.

$$\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{ij}} = 0 \Leftrightarrow \begin{pmatrix} \frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{1,ij}} = 0 \\ \vdots \\ \frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{2,ij}} = 0 \\ \vdots \\ \vdots \\ \frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{K,ij}} = 0 \end{pmatrix}, \tag{45}$$

$$\begin{aligned}\frac{\partial \log\left(\mathcal{L}_{ij}(\theta \mid \mathcal{G}_t)\right)}{\partial \beta_{s,ij}} &= \sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)\ X_{s,t_k^l} - \lambda_{ij,0} \sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\ \int_{t_{k-1}^l}^{t_k^l} X_{s,u}\ e^{\langle \beta_{ij}, X_u\rangle}\,du \\ &= 0\end{aligned}$$

By replacing with $\lambda_{ij,0}$'s expression we get the equation system,

$$\frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)\ X_{s,t_k^l}}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} Y_{ij}^l(t_k^l)} = \frac{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\ \int_{t_{k-1}^l}^{t_k^l} X_{s,u}\ e^{\langle \beta_{ij}, X_u\rangle}\, du}{\sum_{l=1}^{n}\sum_{k=1}^{N_l} S_i^l(t_k^l)\int_{t_{k-1}^l}^{t_k^l} e^{\langle \beta_{ij}, X_u\rangle}\, du}.$$

for $s = 1, \cdots, K$

This non linear equation system has no analytical solution, it can be solved by numerical optimisation (multi-dimensional Newton-Raphson for instance). The solution provides the values of $\hat{\beta}_{ij}$. Once we have $\hat{\beta}_{ij}$ we go back to $\lambda_{ij,0}$ to get its estimator $\hat{\lambda}_{ij,0}$.

□

C Maximisation of equation (31) using gradient

One can notice that the gradient of (31) with respect to α_i, β_i and σ_i can easily be derived. This means that it is possible to transform the maximization problem to a system of non-linear equations and then to resolve this system.

$$\frac{\partial \log\left(\mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0})\right)}{\partial \theta} = 0 \Rightarrow \left\{\frac{\partial \log\left(\mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0})\right)}{\partial \alpha_i}, \frac{\partial \log\left(\mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0})\right)}{\partial \beta_i}, \frac{\partial \log\left(\mathcal{L}(\theta \mid \mathcal{G}_{t_N} \vee \mathcal{H}_{t_0})\right)}{\partial \sigma_i}\right\} = 0.$$

For $i \in \{1, \cdots, d-1\}$, with a fixed values of threshold C_j ($-\infty = C_{d+1} < C_d < \cdots < C_2 < C_1 = +\infty$), every vector $\hat{\theta}_i = \left(\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i\right)$ is obtained through the resolution of the following non-linear equations system,

$$\begin{cases} \displaystyle\sum_{l=1}^{n}\sum_{k=1}^{N}\sum_{j\neq i} \frac{Y_{ij,t_k}^l}{p_{ij}\left(X_{t_k}\right)} \left[\phi\left(\frac{C_j - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) - \phi\left(\frac{C_{j+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right)\right] = 0 \\ \\ \displaystyle\sum_{l=1}^{n}\sum_{k=1}^{N}\sum_{i\neq j} \frac{Y_{ij,t_k}^l}{p_{ij}\left(X_{t_k}\right)} \left[\left(C_j - \alpha_i - \langle \beta_i, X_{t_k}\rangle\right)\phi\left(\frac{C_j - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) - \right. \\ \displaystyle\left.\left(C_{j+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle\right)\phi\left(\frac{C_{j+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right)\right] = 0 \\ \\ \displaystyle\sum_{l=1}^{n}\sum_{k=1}^{N}\sum_{i\neq j} \frac{Y_{ij,t_k}^l\ X_{t_k,s}}{p_{ij}\left(X_{t_k}\right)} \left[\phi\left(\frac{C_j - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) - \phi\left(\frac{C_{j+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right)\right] = 0 \\ \text{for } s = 1, \cdots, K \end{cases} \tag{46}$$

Where, $\phi(\cdot)$ is the density function of the standard normal distribution. By convention, we have,

$$\begin{cases} \phi\left(\dfrac{C_1 - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) = 0 \\ \phi\left(\dfrac{C_{d+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) = 0 \\ \left(C_1 - \alpha_i - \langle \beta_i, X_{t_k}\rangle\right)\phi\left(\dfrac{C_1 - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) = 0 \\ \left(C_{d+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle\right)\phi\left(\dfrac{C_{d+1} - \alpha_i - \langle \beta_i, X_{t_k}\rangle}{\sigma_i}\right) = 0. \end{cases} \tag{47}$$

7 Results

Year	1	2	3	4	5	6	7	8	Total
2006	52	128	487	486	244	155	12	6	1570
2007	58	151	474	517	272	144	15	6	1637
2008	59	168	470	531	271	198	14	6	1717
2009	58	142	468	513	234	190	20	8	1633
2010	49	136	487	545	230	213	34	16	1710
2011	59	132	494	569	241	232	18	5	1750
2012	52	140	564	697	354	296	29	8	2140
2013	53	133	548	725	338	300	30	5	2132

Table 15: Portfolio cohorts by years and ratings

ratings	1	2	3	4	5	6	7	8
1	-0.069	0.069	0	0	0	0	0	0
2	0.005	-0.104	0.096	0.002	0	0	0	0
3	0.0002	0.016	-0.084	0.066	0.0006	0.00025	0	0
4	0	0.0005	0.031	-0.078	0.043	0.002	0.0003	0.0003
5	0.0004	0.001	0.001	0.07	-0.181	0.100	0.001	0.0004
6	0	0	0.001	0.002	0.088	-0.200	0.099	0.008
7	0	0	0	0.009	0	0.358	-0.837	0.469
8	0	0	0	0	0	0	0	0

Table 16: Through The Cycle Generator

	1	2	3	4	5	6	7	8
1	0.000	-0.339	0.000	0.000	0.000	0.000	0.000	0.000
2	-6.787	0.000	-0.098	-0.717	0.000	0.000	0.000	0.000
3	0.524	0.279	0.000	-0.072	3.437	1.409	0.000	0.000
4	0.000	10.349	0.071	0.000	-0.046	0.157	1.646	0.595
5	0.586	0.586	-2.542	0.049	0.000	-0.036	-1.002	-2.542
6	0.000	0.000	1.289	-30.021	0.128	0.000	-0.159	-0.259
7	0.000	0.000	0.000	-2.249	0.000	0.083	0.000	-0.155
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 17: Beta for the second factor intensity model

	1	2	3	4	5	6	7	8
1	0.000	-0.100	0.000	0.000	0.000	0.000	0.000	0.000
2	-0.607	0.000	0.174	0.743	0.000	0.000	0.000	0.000
3	13.442	-0.137	0.000	0.022	-3.643	-2.497	0.000	0.000
4	0.000	0.775	-0.005	0.000	-0.047	0.049	-3.421	0.002
5	0.002	0.002	0.058	-0.091	0.000	-0.015	-0.952	0.058
6	0.000	0.000	-0.994	0.395	0.185	0.000	-0.263	-0.347
7	0.000	0.000	0.000	-8.445	0.000	0.296	0.000	-0.139
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 18: Beta for the third factor intensity model

	1	2	3	4	5	6	7	8
1	0.000	-0.008	0.000	0.000	0.000	0.000	0.000	0.000
2	15.156	0.000	0.162	-0.830	0.000	0.000	0.000	0.000
3	-10.443	0.113	0.000	0.027	10.258	4.557	0.000	0.000
4	0.000	-10.219	0.091	0.000	0.154	0.243	3.762	-1.008
5	-0.988	-0.988	2.695	0.153	0.000	0.250	2.484	2.695
6	0.000	0.000	1.786	53.178	-0.312	0.000	0.387	0.461
7	0.000	0.000	0.000	14.605	0.000	-0.016	0.000	0.130
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 19: Beta for the forth factor intensity model

	1	2	3	4	5	6	7	8
1	0.000	-0.660	0.000	0.000	0.000	0.000	0.000	0.000
2	-55.417	0.000	-0.461	-13.762	0.000	0.000	0.000	0.000
3	36.762	-0.574	0.000	-0.132	-7.647	-26.454	0.000	0.000
4	0.000	8.184	-0.055	0.000	-0.161	-0.202	-24.512	3.355
5	3.289	3.289	-1.168	-0.232	0.000	-0.386	1.462	-1.168
6	0.000	0.000	49.685	-222.853	0.374	0.000	-0.848	-1.666
7	0.000	0.000	0.000	-45.191	0.000	0.811	0.000	-0.384
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 20: Beta for the fifth factor intensity model