Research Article

A Dependent Hidden Markov Model of Credit Quality

Małgorzata Wiktoria Korolkiewicz

Centre for Industrial and Applied Mathematics, School of Mathematics and Statistics,
University of South Australia, City West Campus, GPO Box 2471, Adelaide, SA 5001, Australia

Correspondence should be addressed to
Małgorzata Wiktoria Korolkiewicz, malgorzata.korolkiewicz@unisa.edu.au

We propose a dependent hidden Markov model of credit quality. We suppose that the "true" credit quality is not observed directly but only through noisy observations given by posted credit ratings. The model is formulated in discrete time with a Markov chain observed in martingale noise, where "noise" terms of the state and observation processes are possibly dependent. The model provides estimates for the state of the Markov chain governing the evolution of the credit rating process and the parameters of the model, where the latter are estimated using the EM algorithm. The dependent dynamics allow for the so-called "rating momentum" discussed in the credit literature and also provide a convenient test of independence between the state and observation dynamics.

1. Introduction

Credit ratings summarise a range of qualitative and quantitative information about the credit worthiness of debt issuers and are therefore a convenient signal for the credit quality of the debtor. The estimation of credit quality transition matrices is at the core of credit risk measures with applications to pricing and portfolio risk management. In view of pending regulations regarding the calculation of capital requirements for banks, there is renewed interest in efficiency of credit ratings as indicators of credit quality and models of their dynamics (Basel Committee on Banking Supervision [1]).

In the study of credit quality dynamics, it is convenient to assume that the credit rating process is a time-homogeneous Markov chain, with past changes in credit quality characterised by a transition matrix. The assumptions of time homogeneity and Markovian behaviour of the rating process have been challenged by some empirical studies; see, for example, Bangia et al. [2] or Lando and Skødeberg [3]. In particular, it has been proposed that ratings exhibit "rating momentum" or "drift," where a rating change in response to a

change in credit quality does not fully reflect that change in credit quality. As pointed out by Löffler in [4, 5], these violations of information efficiency could be the result of some of the agencies' rating policies, namely, rating through the cycle and avoiding rating reversals.

In recent years, a number of modelling alternatives were suggested to address departures from the Markov assumption. In Frydman and Schuermann [6], a mixture of two independent continuous time homogeneous Markov chains is proposed for the ratings migration process, so that the future distribution of a firm's ratings depends not only its current rating but also on the past history of ratings. Wendin and McNeil [7] suppose that credit ratings are subject to both observed and unobserved systematic risk. Rating transition patters (e.g., rating momentum) are captured within the context of a generalised linear mixed model (GLMM) that is estimated using Bayesian techniques. Stefanescu et al. [8] propose a Bayesian hierarchical framework, based on Markov Chain Monte Carlo (MCMC) techniques, to model non-Markovian dynamics in ratings migrations. In Wozabal and Hochreiter [9], a coupled Markov chain model is introduced to model dependency among rating migrations of issuers.

In this paper we follow the hidden Markov model (HMM) approach taken in Korolkiewicz and Elliott [10] and assume that the "true" credit quality evolution can be described by a Markov chain but we do not observe this Markov chain directly. Rather, it is hidden in "noisy" observations represented by posted credit ratings. The model is formulated in discrete time, with a Markov chain of "true" credit quality observed in martingale noise. However, we suppose that noise terms of the signal and observation processes are not independent, which allows for the presence of "rating momentum" in posted credit ratings. Application of such dependent hidden Markov model dynamics to modelling credit quality appears to be new. We employ hidden Markov filtering and estimation techniques described in Elliott et al. [11] and use the filter-based EM (Expectation Maximization) algorithm to estimate the parameters of the model. By construction parameters are revised as new information is obtained and so the resulting filters are adaptive and "self-tuning."

The paper is organized as follows. In Section 2 we describe a hidden Markov model (HMM) of credit quality and in Section 3 the dependent dynamics. Recursive filters are given in Section 4 and the parameter estimation procedure is described in Section 5. Section 6 provides an implementation example.

2. Dynamics of the Markov Chain and Observations

Here we briefly describe a hidden Markov model as given in Chapter 2 of Elliott et al. [11]. Formally, a discrete-time, finite-state, time homogeneous Markov chain is a stochastic process $\{X_k\}$ with the state space $S = \{1, 2, \ldots, N\}$ and a transition matrix $A = (a_{ji})_{1 \leq i, j \leq N}$. Without loss of generality, we can assume that the elements of S are identified with the standard unit vectors $\{e_1, e_2, \ldots, e_N\}$, $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)' \in \mathbb{R}^N$.

Write $\mathcal{F}_k = \sigma\{X_0, X_1, \ldots, X_k\}$ for a *filtration* $\{\mathcal{F}_k\}$ models all possible histories of X. The relationship between the state process at time k and the state of the process at time $k + 1$ is then given by $E[X_{k+1}|X_k] = AX_k$.

Define $V_{k+1} = X_{k+1} - AX_k \in \mathbb{R}^N$. Then, the semimartingale representation of the chain X is

$$X_{k+1} = AX_k + V_{k+1}, \quad k = 0, 1, \ldots, \tag{2.1}$$

where V_{k+1} is a martingale increment with $E[V_{k+1}|\mathcal{F}_k] = 0 \in \mathbb{R}^N$.

Suppose we do not observe X directly. Rather, we observe a process Y such that

$$Y_{k+1} = c(X_k, \omega_{k+1}), \quad k = 0, 1, \ldots, \tag{2.2}$$

where c is a function with values in a finite set and $\{\omega_k\}$ is a sequence of i.i.d. random variables independent of X. Random variables $\{\omega_k\}$ represent the noise present in the system. Suppose the range of c consists of M points which are identified with unit vectors $\{f_1, f_2, \ldots, f_M\}$, $f_j = (0, \ldots, 0, 1, 0, \ldots, 0)' \in \mathbb{R}^M$.

Write

$$\mathcal{Y}_k = \sigma\{Y_0, Y_1, \ldots, Y_k\},$$
$$\mathcal{G}_k = \sigma\{X_0, \ldots, X_k, Y_0, \ldots, Y_k\}. \tag{2.3}$$

These increasing families of σ-fields are filtrations representing possible histories of the state process X, the observation process Y, and both processes (X, Y). Write $c_{ji} = P(Y_{k+1} = f_j \mid X_k = e_i)$, $1 \leq i \leq N$, $1 \leq j \leq M$, for the probability of observing a state f_j when the signal process is in fact in state e_i. Then, it can be shown that $E[Y_{k+1}|X_k] = CX_k$, where $C = (c_{ji})_{1 \leq i,j \leq M}$ is a matrix with $c_{ji} \geq 0$ and $\sum_{j=1}^{M} c_{ji} = 1$.

Define $W_{k+1} = Y_{k+1} - CX_k$. The semimartingale representation of the process Y is

$$Y_{k+1} = CX_k + W_{k+1}, \quad k = 0, 1, \ldots, \tag{2.4}$$

where W is a martingale increment with $E[W_{k+1}|\mathcal{G}_k] = 0 \in \mathbb{R}^M$. In our context, the process Y represents posted credit ratings and X "true" credit quality. For reasons which will become apparent in the next section, we assume one-period delay between X and Y.

In summary, the model for the Markov chain X hidden in martingale noise is as follows.

Hidden Markov Model (HMM)

Under a probability measure P,

$$X_{k+1} = AX_k + V_{k+1} \quad \text{(signal equation, true credit quality)},$$
$$Y_{k+1} = CX_k + W_{k+1} \quad \text{(observation equation, posted rating)}. \tag{2.5}$$

A and C are matrices of transition probabilities whose entries satisfy

$$\sum_{j=1}^{N} a_{ji} = 1, \quad a_{ji} \geq 0; \qquad \sum_{j=1}^{M} c_{ji} = 1, \quad c_{ji} \geq 0. \tag{2.6}$$

V_k and W_k are martingale increments satisfying

$$E[V_{k+1} \mid \mathcal{F}_k] = 0, \qquad E[W_{k+1} \mid \mathcal{G}_k] = 0. \tag{2.7}$$

Parameters of this model are (a_{ji}), $1 \leq i, j \leq N$ and (c_{ji}), $1 \leq j \leq M$, $1 \leq i \leq 1$.

3. Dependent Dynamics

The situation considered in this section is that of a hidden Markov model (HMM) for which the "noise" terms in the state and observation processes are possibly dependent.

The dynamics of the state process X and the observation process Y are as given in Section 2. However, the noise terms V_k and W_k are not independent. Instead, we suppose that the joint distribution of Y_k and X_k is given by

$$Y_{k+1}X'_{k+1} = SX_k + \Gamma_{k+1}, \quad k = 0, 1, \ldots, \tag{3.1}$$

where $S = (s_{rji})$ denotes a $(M \times N) \times N$ matrix mapping $\mathbb{R}^N$ into $\mathbb{R}^M \times \mathbb{R}^N$ and

$$s_{rji} = P(Y_{k+1} = f_r, X_{k+1} = e_j \mid X_k = e_i), \quad 1 \le r \le M, \ 1 \le i, \ j \le N. \tag{3.2}$$

Γ_{k+1} is a $\{\mathcal{G}_k\}$-martingale increment with $E[\Gamma_{k+1}|\mathcal{G}_k] = 0$. Write $\mathbf{1} = (1, 1, \ldots, 1)'$ for the vector in $\mathbb{R}^N$ or $\mathbb{R}^M$ depending on the context. Then, for $\mathbf{1} \in \mathbb{R}^M$, $\langle \mathbf{1}, SX_k \rangle = AX_k$ and for $\mathbf{1} \in \mathbb{R}^N$, $\langle SX_k, \mathbf{1} \rangle = CX_k$, where $\langle , \rangle$ denotes the scalar product in $\mathbb{R}^M$ and $\mathbb{R}^N$, respectively.

Write $\gamma_{rji} = P(Y_{k+1} = f_r \mid X_{k+1} = e_j, X_k = e_i) = s_{rji}/a_{ji}$, and let $\tilde{C}$ be the $M \times (N \times N)$ matrix (γ_{rji}), $1 \le r \le M$, $1 \le i, \ j \le N$. Then it can be shown that $Y_{k+1} = \tilde{C}(X_{k+1}X'_k) + \widetilde{W}_{k+1}$, where $E[\widetilde{W}_{k+1}|\mathcal{G}_k] = 0$.

In summary, the model is now as follows.

Dependent Hidden Markov Model (Dependent HMM)

Under a probability measure P,

$$X_{k+1} = AX_k + V_{k+1},$$
$$Y_{k+1} = \tilde{C}(X_{k+1}X'_k) + \widetilde{W}_{k+1}. \tag{3.3}$$

A and $\tilde{C}$ are matrices of transition probabilities whose entries satisfy

$$\sum_{j=1}^{N} a_{ji} = 1, \quad \sum_{r=1}^{M} \gamma_{rji} = 1. \tag{3.4}$$

V_k and $\widetilde{W}_k$ are martingale increments satisfying

$$E[V_{k+1} \mid \mathcal{F}_k] = 0, \quad E\left[\widetilde{W}_{k+1} \mid \mathcal{G}_k\right] = 0. \tag{3.5}$$

Parameters of this model are (a_{ji}), $1 \le i, \ j \le N$, (c_{ji}), $1 \le j \le M$, $1 \le i \le 1$, and (s_{rji}), $1 \le i, \ j \le N$, $1 \le r \le M$.

We are in a situation analogous to the dependent hidden Markov model case discussed in Chapter 2, Section 10 of Elliott et al. [11]. The difference is that we are assuming dynamics where the observation Y_k depends on both X_k and X_{k-1}. In other words, we suppose

that the current credit rating contains information about both current and previous credit quality, thus allowing for the situation where a rating does not immediately reflect all available information about credit quality, as indicated by a number of empirical studies (see, e.g., Lando and Skødeberg [3]). Put differently, in this model X_k and observation Y_k jointly depend on X_{k-1}, which means that, in addition to previous period's credit quality, knowledge of current credit rating carries information about current credit quality. Moreover, probabilities γ_{rji} provide the distribution of the next period's credit rating given both current and next period's credit quality, thus allowing us to capture "rating momentum" or "rating drift."

In the following sections we will presents estimates for the state of the Markov chain X, the number of jumps from one state to another, the occupation time of X in any state, the number of transitions of the observation process Y into a particular state of X, and the number of joint transitions of X and Y. We will then use the filter-based EM (expectation maximization) algorithm as described in Elliott et al. [11], to obtain optimal estimates of the model, making it adaptive or "self-tuning."

Note that if the noise terms in the state X and observation Y are independent, we have

$$
\begin{aligned}
P\big(Y_{k+1} = f_r, X_{k+1} = e_j \mid X_k = e_i\big) \\
= P\big(Y_{k+1} = f_r \mid X_k = e_i\big) P\big(X_{k+1} = e_j \mid X_k = e_i\big).
\end{aligned}
\tag{3.6}
$$

Hence if the noise terms are independent,

$$
s_{rji} = c_{rj} a_{ji}
\tag{3.7}
$$

for $1 \leq r \leq M$, $1 \leq i, j \leq N$. Consequently, a test of independence is to check whether parameter estimates satisfy

$$
\widehat{s}_{rji} = \widehat{c}_{rj} \widehat{a}_{ji}.
\tag{3.8}
$$

4. Recursive Filter

Following Elliott et al. [11], suppose that under some probability measure $\overline{P}$ on $(\Omega, \mathcal{F})$, $\{Y_k\}$ is a sequence of i.i.d. uniform variables, that is, $\overline{P}(Y_{k+1} = f_j \mid \mathcal{G}_k) = \overline{P}(Y_{k+1} = f_j) = 1/M$. Further, under $\overline{P}$, X is Markov chain independent of Y, with state space $S = \{e_1, \ldots, e_N\}$ and transition matrix $A = (a_{ji})$. That is, $X_{k+1} = A X_k + V_{k+1}$, where $\overline{E}[V_{k+1} \mid \mathcal{G}_k] = \overline{E}[V_{k+1} \mid \mathcal{F}_k] = 0 \in \mathbb{R}^N$. Suppose $\widetilde{C} = (\gamma_{rji})$, $1 \leq r \leq M$, $1 \leq i$, $j \leq N$, is a matrix with $\gamma_{ji} \geq 0$, and $\sum_{j=1}^{M} \gamma_{rji} = 1$.

Define $\overline{\lambda}_l = M \sum_{j=1}^{M} \langle \widetilde{C}(X_l X_{l-1}'), f_j \rangle \langle Y_l, f_j \rangle$ and $\overline{\Lambda}_k = \prod_{l=1}^{k} \overline{\lambda}_l$. Define a new probability measure P by putting $(dP/d\overline{P})|_{\mathcal{G}_k} = \overline{\Lambda}_k$. Then, under P, X remains a Markov chain with transition matrix A and $P(Y_{k+1} = f_r \mid X_{k+1} = e_j, X_k = e_i) = \gamma_{rji}$. That is, under P, $X_{k+1} = A X_k + V_{k+1}$ and $Y_{k+1} = \widetilde{C}(X_{k+1} X_k') + \widetilde{W}_{k+1}$.

Suppose we observe $Y_0, \ldots, Y_k$, and we wish to estimate $X_0, \ldots, X_k$. The best (mean square) estimate of X_k given $\mathcal{Y}_k = \sigma\{Y_0, \ldots, Y_k\}$ is $E[X_k \mid \mathcal{Y}_k] \in \mathbb{R}^N$. However, $\overline{P}$ is a much

easier measure under which to work. Using Bayes' Theorem as described in Elliott et al. [11], we have

$$E[X_k \mid \mathcal{Y}_k] = \frac{\overline{E}\left[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k \mid \mathcal{Y}_k\right]}. \tag{4.1}$$

Write $\tilde{q}_k := \overline{E}[\overline{\Lambda}_k X_k \mid \mathcal{Y}_k] \in \mathbb{R}^N$. $\tilde{q}_k$ is then an unnormalized conditional expectation of X_k given the observations $\mathcal{Y}_k$. Note that $\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k] = \langle \tilde{q}_k, \mathbf{1} \rangle$, where $\mathbf{1} = (1, 1, \dots, 1)' \in \mathbb{R}^N$. It then follows that

$$E[X_k \mid \mathcal{Y}_k] = \frac{\tilde{q}_k}{\langle \tilde{q}_k, \mathbf{1} \rangle}. \tag{4.2}$$

Hence, to estimate $E[X_k \mid \mathcal{Y}_k]$ we need to know the dynamics of $\tilde{q}$. Using the methods of Elliott et al. [11], the following recursive formula for $\tilde{q}_{k+1}$ is obtained:

$$\tilde{q}_{k+1} = M Y'_{k+1} S \tilde{q}_k. \tag{4.3}$$

5. Parameter Estimates

To estimate parameters of the model, matrices A, C, and S, we need estimates of the following processes:

$$\mathcal{J}_k^{ij} = \sum_{n=1}^{k} \langle X_{n-1}, e_i \rangle \langle X_n, e_j \rangle, \quad 1 \le i, j \le N,$$

$$\mathcal{O}_k^i = \sum_{n=1}^{k} \langle X_{n-1}, e_i \rangle, \quad 1 \le i \le N,$$

$$\mathcal{T}_k^{ir} = \sum_{n=1}^{k} \langle X_{n-1}, e_i \rangle \langle Y_n, f_r \rangle, \quad 1 \le i \le N, \ 1 \le r \le M,$$

$$\mathcal{L}_k^{ijr} = \sum_{n=1}^{k} \langle X_{n-1}, e_i \rangle \langle X_n, e_j \rangle \langle Y_n, f_r \rangle, \quad 1 \le r \le M, \ 1 \le i, j \le N. \tag{5.1}$$

The above processes are interpreted as follows:

$\mathcal{J}_k^{ij}$ is the number of jumps of X from state e_i to state e_j up to time k. $\mathcal{O}_k^i$ is the amount of time, up to time $k-1$, X has spent in state e_i. $\mathcal{T}_k^{ir}$ is the number of transitions, up to time k, from state e_i to observation f_r. $\mathcal{L}_k^{ijr}$ is the number of jumps of X from state e_i to state e_j while Y was in state f_r up to time k.

Note that $\sum_{j=1}^{N} \mathcal{J}_k^{ij} = \sum_{j=1}^{M} \mathcal{T}_k^{ij} = \sum_{r=1}^{M} \sum_{j=1}^{N} \mathcal{L}_k^{ijr} = \mathcal{O}_k^i$.

Consider first the jump process $\{\mathcal{J}_k^{ij}\}$. We wish to estimate $\mathcal{J}_k^{ij}$ given the observations $Y_0, \ldots, Y_k$. As in the case of a filter for the state X described in Section 4, the best (mean-square) estimate is

$$E\left[\mathcal{J}_k^{ij} \mid \mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k \mathcal{J}_k^{ij} \mid \mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k \mid \mathcal{Y}_k\right]} := \frac{\sigma(\mathcal{J}^{ij})_k}{\langle \tilde{q}_k, \mathbf{1} \rangle}. \tag{5.2}$$

We wish to know how $\sigma(\mathcal{J}^{ij})_k$ is updated as time passes and new information arrives. However, as noted in Elliott et al. [11], we work with $\sigma(J^{ij}X)_k = \overline{E}[\overline{\Lambda}_k J_k^{ij} X_k \mid \mathcal{Y}_k]$ rather than $\sigma(J^{ij})_k = \overline{E}[\overline{\Lambda}_k J_k^{ij} \mid \mathcal{Y}_k]$, in order to obtain closed-form recursions. The quantity of interest, namely, $\sigma(\mathcal{J}^{ij})_k$, is then readily obtained as $\sigma(\mathcal{J}^{ij})_k = \langle \sigma(\mathcal{J}^{ij}X)_k, \mathbf{1} \rangle$. We have

$$\sigma\left(\mathcal{J}^{ij}X\right)_{k+1} = MY_{k+1}'S\sigma\left(\mathcal{J}^{ij}X\right)_k + \langle \tilde{q}_k, e_i \rangle \left(M \sum_{r=1}^{M} s_{rji} \langle Y_{k+1}, f_r \rangle \right) e_j. \tag{5.3}$$

Similarly, we consider the best (mean square) estimates of $\mathcal{O}_k^i$, $\mathcal{T}_k^{jr}$, and $\mathcal{L}^{rji}$ given $\mathcal{Y}_k$:

$$E\left[\mathcal{O}_k^i \mid \mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k \mathcal{O}_k^i \mid \mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k \mid \mathcal{Y}_k\right]} := \frac{\sigma(\mathcal{O}^i)_k}{\langle \tilde{q}_k, \mathbf{1} \rangle},$$

$$E\left[\mathcal{T}_k^{jir} \mid \mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k \mathcal{T}_k^{jir} \mid \mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k \mid \mathcal{Y}_k\right]} := \frac{\sigma(\mathcal{T}^{ir})_k}{\langle \tilde{q}_k, \mathbf{1} \rangle}, \tag{5.4}$$

$$E\left[\mathcal{L}_k^{ijr} \mid \mathcal{Y}_k\right] = \frac{\overline{E}\left[\overline{\Lambda}_k \mathcal{L}_k^{ijr} \mid \mathcal{Y}_k\right]}{\overline{E}\left[\overline{\Lambda}_k \mid \mathcal{Y}_k\right]} := \frac{\sigma(\mathcal{L}^{ijr})_k}{\langle \tilde{q}_k, \mathbf{1} \rangle}.$$

Recursive formulae for the processes

$$\sigma\left(\mathcal{O}^i X\right)_k := \overline{E}\left[\overline{\Lambda}_k \mathcal{O}_k^i X_k \mid \mathcal{Y}_k\right],$$

$$\sigma\left(\mathcal{T}^{ir} X\right)_k := \overline{E}\left[\overline{\Lambda}_k \mathcal{T}_k^{ir} X_k \mid \mathcal{Y}_k\right], \tag{5.5}$$

$$\sigma\left(\mathcal{L}^{ijr} X\right)_k := \overline{E}\left[\overline{\Lambda}_k \mathcal{L}_k^{ijr} X_k \mid \mathcal{Y}_k\right]$$

are as follows:

$$\sigma\left(\mathcal{O}^i X\right)_{k+1} = MY'_{k+1} S\sigma\left(\mathcal{O}^i X\right)_k + \langle \tilde{q}_k, e_i \rangle \sum_{j=1}^{M} \left(M \sum_{r=1}^{M} s_{rji} \langle Y_{k+1}, f_r \rangle \right) e_j,$$

$$\sigma\left(\mathcal{T}^{ir} X\right)_{k+1} = MY'_{k+1} S\sigma\left(\mathcal{T}^{ir} X\right)_k + M\langle \tilde{q}_k, e_i \rangle \left(\sum_{j=1}^{N} s_{rji} e_j \right) \langle Y_{k+1}, f_r \rangle, \tag{5.6}$$

$$\sigma\left(\mathcal{L}^{ijr} X\right)_{k+1} = MY'_{k+1} S\sigma\left(\mathcal{L}^{ijr} X\right)_k + \langle \tilde{q}_k, e_i \rangle M s_{rji} \langle Y_{k+1}, f_r \rangle e_j.$$

As in the case of the number of jumps of the state process X, quantities of interest $\sigma(\mathcal{O}^i)_k$, $\sigma(\mathcal{T}^{ir})_k$, and $\sigma(\mathcal{L}^{ijr})_k$ are obtained by taking inner products with $\mathbf{1} = (1, 1, \ldots, 1)'$:

$$\sigma\left(\mathcal{O}^i\right)_k = \left\langle \sigma\left(\mathcal{O}^i X\right)_k, \mathbf{1} \right\rangle,$$

$$\sigma\left(\mathcal{T}^{ir}\right)_k = \left\langle \sigma\left(\mathcal{T}^{ir} X\right)_k, \mathbf{1} \right\rangle, \tag{5.7}$$

$$\sigma\left(\mathcal{L}^{ijr}\right)_k = \left\langle \sigma\left(\mathcal{L}^{ijr} X\right)_k, \mathbf{1} \right\rangle.$$

The model is determined by parameters $\theta = \{a_{ji}, 1 \le i, j \le N; c_{ji}, 1 \le i \le N, 1 \le j \le M; s_{rji}, 1 \le r \le M, 1 \le i, j \le N\}$. These satisfy

$$a_{ji} \ge 0, \quad \sum_{j=1}^{N} a_{ji} = 1, \qquad c_{ji} \ge 0, \quad \sum_{j=1}^{M} c_{ji} = 1, \qquad s_{rji} \ge 0, \quad \sum_{r=1}^{M} \sum_{j=1}^{N} s_{rji} = 1. \tag{5.8}$$

We want to determine a new set of parameters $\hat{\theta} = \{\hat{a}_{ji}, 1 \le i, j \le N; \hat{c}_{ji}, 1 \le i \le N, 1 \le j \le M\}; \hat{s}_{rji}, 1 \le r \le M, 1 \le i, j \le N\}$ given the arrival of new information embedded in the values of the observation process Y. This requires maximum likelihood estimation. As in [11], we proceed by using the filter-based EM (Expectation Maximization) algorithm, which retains the well-established statistical properties of the EM algorithm while reducing memory costs and thus allowing for faster computation (see, e.g., Krishnamurthy and Chung [12]).

Consider first the parameter a_{ji}. Suppose that, under measure P_θ, X is a Markov chain with transition matrix $A = (a_{ji})$. We define a new probability measure $P_{\hat{\theta}}$ such that, under $P_{\hat{\theta}}$, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$, that is,

$$P_{\hat{\theta}}(X_{k+1} = e_j \mid X_k = e_i) = \hat{a}_{ji}, \tag{5.9}$$

$\hat{a}_{ji} \ge 0$, $\sum_{j=1}^{N} \hat{a}_{ji} = 1$. Define

$$\Lambda_0 = 1,$$

$$\Lambda_k = \prod_{l=1}^{k} \left(\sum_{r,s=1}^{N} \left(\frac{\hat{a}_{sr}}{a_{sr}} \right) \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \right). \tag{5.10}$$

In case $a_{ji} = 0$ take $\hat{a}_{ji} = 0$ and $\hat{a}_{sr}/a_{sr} = 1$.

Define $P_{\widehat{\theta}}$ by setting $(dP_{\widehat{\theta}}/dP_\theta) \mid \mathcal{F}_k = \Lambda_k$. It can then be shown that, under $P_{\widehat{\theta}}$, X is a Markov chain with transition matrix $\widehat{A} = (\widehat{a}_{ji})$. Moreover, given the observations up to time k, $\{Y_0, Y_1, \ldots, Y_k\}$, and given the parameter set $\theta = \{a_{ji}, 1 \le i, \ j \le N; \ c_{ji}, 1 \le i \le N, \ 1 \le j \le M\}$, the EM estimates $\widehat{a}_{ji}$ are given by

$$\widehat{a}_{ji} = \frac{\sigma(\mathcal{J}^{ij})_k}{\sigma(\mathcal{O}^i)_k}. \tag{5.11}$$

Consider now the parameter c_{ji}. Suppose that, under measure P_θ,

$$Y_{k+1} = CX_k + W_{k+1}, \tag{5.12}$$

where $C = (c_{ji})$. We define a new probability measure $P_{\widehat{\theta}}$ as follows. Put

$$\Lambda_0 = 1,$$

$$\Lambda_k = \prod_{l=1}^{k} \left(\sum_{r,s=1}^{N} \left(\frac{\widehat{c}_{sr}}{c_{sr}} \right) \langle X_{l-1}, e_r \rangle \langle Y_l, f_s \rangle \right). \tag{5.13}$$

In case $c_{ji} = 0$ take $\widehat{c}_{ji} = 0$ and $\widehat{c}_{sr}/c_{sr} = 1$.

Define $P_{\widehat{\theta}}$ by setting $(dP_{\widehat{\theta}}/dP_\theta) \mid \mathcal{G}_k = \Lambda_k$. Again it can be shown that, under $P_{\widehat{\theta}}$,

$$Y_{k+1} = \widehat{C}X_k + \widehat{W}_{k+1}, \tag{5.14}$$

that is, $P_{\widehat{\theta}}(Y_{k+1} = f_s \mid X_k = e_r) = \widehat{c}_{sr}$. Moreover, given the observations up to time k, $\{Y_0, Y_1, \ldots, Y_k\}$, and given the parameter set $\theta = \{a_{ji}, 1 \le i, \ j \le N; \ c_{ji}, 1 \le i \le N, \ 1 \le j \le M\}$, the EM estimates $\widehat{c}_{ji}$ are given by

$$\widehat{c}_{ji} = \frac{\sigma(\mathcal{T}^{ij})_k}{\sigma(\mathcal{O}^i)_k}. \tag{5.15}$$

Finally, consider the parameter s_{rji}. A new probability measure $P_{\widehat{\theta}}$ is defined by putting

$$\Lambda_0 = 1,$$

$$\Lambda_k = \prod_{l=1}^{k} \left(\sum_{r=1}^{M} \sum_{i,j=1}^{N} \left(\frac{\widehat{s}_{rji}}{s_{rji}} \right) \langle Y_l, f_r \rangle \langle X_l, e_j \rangle \langle X_{l-1}, e_i \rangle \right). \tag{5.16}$$

In case $s_{rji} = 0$ take $\widehat{s}_{rji} = 0$ and $\widehat{s}_{rji}/s_{rji} = 1$. Define $P_{\widehat{\theta}}$ by setting $(dP_{\widehat{\theta}}/dP_\theta) \mid \mathcal{G}_k = \Lambda_k$. Then, under $P_{\widehat{\theta}}$, $Y_{k+1}X'_{k+1} = \widehat{S}X_k + \widehat{\Gamma}_{k+1}$, that is,

$$P_{\widehat{\theta}}(Y_{k+1} = f_r, X_{k+1} = e_j \mid X_k = e_i) = \widehat{s}_{rji}. \tag{5.17}$$

Given the observations up to time k, $\{Y_0, Y_1, \ldots, Y_k\}$, and given the parameter set $\theta = \{a_{ji}, 1 \leq i, j \leq N; c_{ji}, 1 \leq i \leq N, 1 \leq j \leq M; s_{rji}, 1 \leq r \leq M, 1 \leq i, j \leq N\}$, the EM estimates $\hat{s}_{rji}$ are then given by

$$\hat{s}_{rji} = \frac{\sigma(\mathcal{L}^{ijr})_k}{\sigma(\mathcal{O}^i)_k}. \tag{5.18}$$

6. Implementation Example

The dependent hidden Markov model (Dependent HMM) described in previous sections was applied to a dataset of Standard & Poor's credit ratings. Description of the data and implementation results are given below.

6.1. Data Description

Our analysis takes advantage of the Standard & Poor's COMPUSTAT database, which contains rating histories for 1,301 obligors over the period 1985–1999 (Standard & Poor's [13]). The universe of obligors is mainly large US and Canadian corporate institutions. The obligors include industrials, utilities, insurance companies, banks and other financial institutions, and real-estate companies. The COMPUSTAT database provides annual ratings. Every year each of the rated obligors is assigned to one of the Standard and Poor's 7 rating categories, ranging from *AAA* (highest rating) to *CCC* (lowest rating) as well as *D* (payment in default) and the NR (not rated) state.

We have a total of 19,515 firm-years in our sample. However, only 34% of those observations correspond to one of the eight Standard & Poor's rating labels in a given year. The remaining 66% of observations represent the so-called NR (not rated) status. As discussed in the literature, transitions to NR may be due to several reasons, such as expiration of the debt, calling of the debt, or the issuer deciding to bypass an agency rating (see, e.g., Bangia et al. [2]). Unfortunately, details of individual transitions to NR are not known.

Excluding NR, approximately 85% of the remaining ratings are in categories *A* down to *B*. The median rating is *BB*, the highest non-investment-grade rating. Approximately 1% of the observed ratings are *AAA* and 2% are defaults. The most common rating is *B*, two rating categories above default, which accounts for 25.5% of the observations.

6.2. Implementation Results

Since individual firms generally experience few rating changes and changes that do occur are to neighbouring categories, we apply the Dependent HMM algorithm to an aggregate of firms in the dataset rather to allow for more observed transitions between rating categories and make inferences possible. Specifically, we follow the *filter-based cohort approach* adopted in Korolkiewicz and Elliott [10], and instead of estimating the distribution and parameters for the Markov chain X_k^l for each firm l, we estimate the distribution and parameters for $\sum_{l=1}^{L} X_k^l$ given the additivity of all stochastic processes discussed in Sections 4 and 5.

Given the fairly large number of parameters to be estimated compared to the number of rating transitions in the dataset, we have reclassified all firms in the sample as IG (investment grade), SG (speculative grade), *D*, or NR and then applied the Dependent HMM

Table 1: Estimates of matrices A, C, and S.

(a)

	Estimated matrix A					Estimated matrix C			
	IG	SG	D	NR		IG	SG	D	NR
IG	0.408	0.018	0.000	0.000	IG	0.126	0.038	0.000	0.000
SG	0.068	0.249	0.000	0.000	SG	0.094	0.118	0.010	0.000
D	0.000	0.017	1.000	0.000	D	0.001	0.004	0.000	0.000
NR	0.524	0.715	0.000	0.999	NR	0.780	0.840	0.990	1.000

(b)

	Estimated matrix S								
	IG category					D category			
	IG	SG	D	NR		IG	SG	D	NR
IG	0.119	0.007	0.000	0.000	IG	0.000	0.000	0.000	0.000
SG	0.076	0.018	0.000	0.000	SG	0.000	0.000	0.010	0.000
D	0.000	0.001	0.000	0.000	D	0.000	0.000	0.000	0.000
NR	0.211	0.043	0.000	0.524	NR	0.000	0.000	0.990	0.000
	SG category					NR category			
	IG	SG	D	NR		IG	SG	D	NR
IG	0.007	0.032	0.000	0.000	IG	0.000	0.000	0.000	0.000
SG	0.006	0.105	0.008	0.000	SG	0.000	0.000	0.000	0.000
D	0.000	0.003	0.000	0.000	D	0.000	0.000	0.000	0.000
NR	0.008	0.108	0.009	0.715	NR	0.000	0.000	0.000	0.999

algorithm to the new dataset. This classification is motivated by the fact that a corporation which can issue higher rated debt usually receives better financing terms. Further, as a matter of policy or law, some institutional investors can only purchase investment-grade bonds. Hence it is often crucial for a borrower to maintain an investment-grade rating and so it is interesting to see if rating transition data reflects this.

Each modified credit rating category IG, SG, as well as default D and NR, was identified with a unit vector in $\mathbb{R}^4$. Given the relatively short time period, parameter estimates were updated with the arrival of every new observation for the 1,301 firms in the dataset. Repetition of the estimation procedures ensures that the model and estimates improve with each iteration. Estimated parameters of the model, namely, matrices $\widehat{A}$, $\widehat{C}$, and S, are given in Table 1.

Considering the estimated transition matrix $\widehat{A}$, note that entries above the diagonal correspond to rating upgrades and those below the diagonal to rating downgrades. Nonzero transition probabilities are concentrated and highest on the diagonal and the second largest probability is in the last row, indicating that obligors generally either maintain their rating or enter the NR (not rated) category. Our results show that investment-grade firms generally hold on to their status. The probability of downgrade to speculative-grade status is estimated as 6.8%. However, for speculative-grade firms, the probability of upgrade to investment-grade status is lower (estimated probability of 1.8%). Speculative-grade firms tend to maintain their status or disappear from the dataset because of either default or withdrawn rating. The probability of transition to the NR status is higher for speculative-grade obligors (71.5%) than for investment-grade obligors (52.4%).

Table 2: Linear regression results.

(a) Summary output

Regression statistics	
Multiple R	0.9935
R square	0.9871
Adjusted R square	0.9868
Standard error	0.0236
Observations	64

(b) ANOVA

	df	SS	MS	F	Significance F
Regression	1	2.6345	2.6345	4728.10	0.0000
Residual	62	0.0345	0.0006		
Total	63	2.6690			

(c)

	Coeff	Std err	t stat	P value	Lower 95%	Upper 95%
Intercept	−0.0003	0.0031	−0.1109	0.9121	−0.0065	0.0058
$s(rji)$	1.0058	0.0146	68.7612	0.0000	0.9765	1.0350

Recall from Section 3 that, given estimates of matrices A and C, our Dependent HMM also provides the distribution of posted credit ratings at time $k+1$ given "true" credit quality at times k and $k+1$, namely, estimates of conditional probabilities $\gamma_{rji} = P(Y_{k+1} = f_r \mid X_{k+1} = e_j,\ X_k = e_i)$. To illustrate, consider a borrower with investment-grade "true" credit quality at times k and $k+1$. The probability that this borrower is assigned to a speculative-grade rating class is $P(Y_{k+1} = \mathrm{SG} \mid X_{k+1} = \mathrm{IG}, X_k = \mathrm{IG})$, which, given our model parameter estimates, is given by $\hat{s}_{\mathrm{SG,IG,IG}}/\hat{a}_{\mathrm{IG,IG}} = 0.007/0.408 = 0.017$. Similarly, for a borrower whose "true" credit quality improves from SG to IG, the probability of being assigned to an IG rating class is given by $P(Y_{k+1} = \mathrm{IG} \mid X_{k+1} = \mathrm{IG}, X_k = \mathrm{SG})$, which we would estimate to be $0.007/0.018 = 0.389$. These estimates again suggest that rating agencies may be somewhat reluctant to downgrade (upgrade) borrowers from (to) investment grade, thus introducing a degree of "rating momentum."

6.3. Test of Independence

Recall that the Dependent HMM allows the "noise" terms in the state and observation processes to be possibly dependent. As indicated in Section 3, a convenient test of independence is to check whether the estimated parameters of the model satisfy $\hat{s}_{rji} = \hat{c}_{rj}\hat{a}_{ji}$.

Given our estimates of matrices $\hat{A}$ and $\hat{C}$, products $\hat{c}_{rj}\hat{a}_{ji}$ were calculated and then compared to corresponding entries of the estimated matrix $\hat{S}$ using linear regression. The regression results are given in Table 2. As indicated by the high F-statistic (4728.10) and high R^2 value (98.71%), the fitted regression model is significant. The slope estimate is very close to one with low standard error and P value of 0.000, while the intercept estimate is very close zero and not significant (P value of 0.91). These regression results suggest no major departures from independence, which seems to agree with findings in Kiefer and Larson [14]

that indicate the Markov assumption, implicit in most credit risk models, does not seem to be "too wrong" for typical forecast horizons. However, longer rating histories may be necessary to verify these results.

7. Conclusion

We have proposed a Dependent Hidden Markov Model for the evolution of credit quality in discrete time with a Markov chain observed in martingale noise. We have applied the estimation techniques of hidden Markov models from Elliott et al. [11] to obtain the best estimate of the Markov chain representing "true" credit quality and estimates of the parameters. The estimation procedure was repeated to ensure that the model and estimates improved with each iteration. The model was applied to a dataset of Standard & Poor's issuer ratings and our preliminary results agree with some qualitative observations made in the literature regarding credit rating systems but also indicate no significant dependence in the dynamics of the "state" (credit quality) and "observation" (credit rating) processes.

References

[1] Basel Committee on Banking Supervision, "Studies on validation of of internal rating systems," BCBS Publications 14, Bank for International Settlements, 2005.

[2] A. Bangia, F. X. Diebold, A. Kronimus, C. Schagen, and T. Schuermann, "Ratings migration and the business cycle, with application to credit portfolio stress testing," *Journal of Banking and Finance*, vol. 26, no. 2-3, pp. 445–474, 2002.

[3] D. Lando and T. M. Skødeberg, "Analyzing rating transitions and rating drift with continuous observations," *Journal of Banking and Finance*, vol. 26, no. 2-3, pp. 423–444, 2002.

[4] G. Löffler, "An anatomy of rating through the cycle," *Journal of Banking and Finance*, vol. 28, no. 3, pp. 695–720, 2004.

[5] G. Löffler, "Avoiding the rating bounce: Why rating agencies are slow to react to new information," *Journal of Economic Behavior and Organization*, vol. 56, no. 3, pp. 365–381, 2005.

[6] H. Frydman and T. Schuermann, "Credit rating dynamics and Markov mixture models," *Journal of Banking and Finance*, vol. 32, no. 6, pp. 1062–1075, 2008.

[7] J. Wendin and A. J. McNeil, "Dependent credit migrations," *Journal of Credit Risk*, vol. 2, pp. 87–114, 2006.

[8] C. Stefanescu, R. Tunaru, and S. Turnbull, "The credit rating process and estimation of transition probabilities: A Bayesian approach," *Journal of Empirical Finance*, vol. 16, no. 2, pp. 216–234, 2009.

[9] D. Wozabal and R. Hochreiter, "A coupled Markov chain approach to credit risk modeling," *Journal of Economic Dynamics and Control*, vol. 36, no. 3, pp. 403–415, 2012.

[10] M. W. Korolkiewicz and R. J. Elliott, "A hidden Markov model of credit quality," *Journal of Economic Dynamics & Control*, vol. 32, no. 12, pp. 3807–3819, 2008.

[11] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, vol. 29, Springer, New York, NY, USA, 1995.

[12] V. Krishnamurthy and S. H. Chung, "Signal processing based on hidden Markov models for extracting small channel currents," in *Handbook of Ion Channels: Dynamics, Structure, and Applications*, S. H. Chung, S. H. Andersen, and V. Krishnamurthy, Eds., Springer, Berlin, Germany, 2006.

[13] Standard & Poor's, Standard & Poor's COMPUSTAT (North America) User's Guide, Standard & Poor's, Englewood Colorado, 2000.

[14] N. M. Kiefer and C. E. Larson, "A simulation estimator for testing the time homogeneity of credit rating transitions," *Journal of Empirical Finance*, vol. 14, no. 5, pp. 818–835, 2007.