Efficient estimation of transition rates between credit ratings from observations at discrete time points

Mogens Bladt
IIMAS–UNAM
A.P. 20-726
01000 Mexico, D.F.
Mexico
bladt@sigma.iimas.unam.mx

Michael Sørensen
Department of Applied Mathematics and Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
michael@math.ku.dk

Abstract

The paper demonstrates how discrete-time credit rating data (e.g. annual observations) can be analysed by means of a continuous-time Markov model. Two methods for estimating the transition intensities are given: the EM-algorithm and an MCMC approach. The estimated transition intensities can be used to estimate the matrix of probabilities of transitions between all credit ratings, including default probabilities, over any time horizon. Thus the advantages of a continuous-time model can be obtained without continuous-time data. Estimates of the variance of estimators as well as confidence and credibility intervals are presented, and a test for equality of two intensity matrices is proposed. The methods are demonstrated in an analysis of a large data set drawn from Moody's Corporate Bond Default Database, where reasonable estimates are obtained from annual observations.

Key words: Continuous-time Markov chains, confidence sets, default probabilities, EM-algorithm, Markov chain Monte Carlo.

1 Introduction

Analysis of credit rating transition data by a continuous-time Markov model is superior to using a discrete time model. This is because a continuous-time model allows meaningful estimation of the probability of rare transitions, even if these transitions have not actually been observed in the data set. An example is the probability of default from a high rating category. The continuous-time approach also hooks up nicely with rating-based term structure modelling in which yield curves are found for different rating classes, see Jarrow, Lando & Turnbull (1997). These facts were pointed out by Lando & Skødeberg (2002), who analysed continuous-time observations of credit transitions.

In this paper we demonstrate that a continuous-time Markov model can also be used to analyse discrete-time observations, where ratings have only been observed at discrete points in time, e.g. monthly or annual observations. We provide two methods for estimating the transition intensities (the generator) of a Markov jump process with finitely many states (the ratings). The estimated transition intensities can be used to calculate estimates of the probabilities of transitions between any two rating classes over any time-horizon for which the Markov model can be assumed to hold. In particular the probability of default from any rating class can be estimated. These transition probabilities are central to modern credit risk management.

The estimation methods proposed in this paper generalize the methods in Bladt & Sørensen (2005) for a single Markov jump process to the case of simultaneous observation of several processes, which is the relevant setting for credit rating data. While the basic algorithms were essentially developed in the previous article, the current paper not only applies these techniques to credit risk data, but further develops methods of particular interest when dealing with this kind of data such as importance sampling for rare transitions, testing the time–homogeneity assumption and issues concerning parameter reduction. Also the results on confidence intervals are new. Furthermore, while Bladt & Sørensen (2005) was a theoretical paper, in this paper the focus is on practical issues concerning the statistical inference for Markov jump processes and a considerable part of the paper is devoted an analysis of a large credit rating data set drawn from Moody's Corporate Bond Default Database.

The main problem we address is that of estimating the matrix of transition intensities (also called the generator) of the Markov process on the basis of discrete time data. The estimation problem is easy if a continuous-time record of all transitions is available so that the exact dates at which companies are up- or downgraded or default are known. In this situation there is an explicit formula for the maximum likelihood estimator. When the data is incomplete in the sense that the only information available is the ratings of all companies at certain points in time, estimation of the intensities is not a trivial task. Under the Markov assumption the discrete time process is a Markov chain, and if the observation times are equidistant, the matrix of transition probabilities, $\boldsymbol{P}$, can be easily estimated by standard cohort or 'multinomial' techniques. The transition matrix $\boldsymbol{P}$ is related to the intensity matrix $\boldsymbol{\Lambda}$ by $\exp(t\boldsymbol{\Lambda}) = \boldsymbol{P}$, where exp denotes the matrix exponential function and t is the time between observations. It would appear that an estimator of $\boldsymbol{\Lambda}$ could be obtained as the matrix–logarithm of $\boldsymbol{P}$. However, the matrix–logarithm is not necessarily unique or, even worse, the matrix–logarithm of $\boldsymbol{P}$ may not be a generator matrix, see Culver (1966). The latter problem is related to the so-called embedding problem for Markov chains first posed by Elfving (1937), and subsequently treated by several authors, e.g. Kingman

(1962) and Johansen (1974). Aspects of the embedding problem of relevance to the problem of estimation of the intensity matrix of a Markov jump process are discussed in Bladt & Sørensen (2005), where more references can be found. In that paper two algorithms for finding the maximum likelihood estimator (or the approximating mean posterior estimator) of the intensity matrix were proposed and investigated. It is well-known that the resulting estimators are effcient.

A few other solutions to the problem have been proposed in the literature. Israel, Rosenthal & Wei (1997) proposed a method of estimating the jump intensities from discrete time observations that is, however, not efficient and does only after an ad hoc modification of the estimator yield an intensity matrix. In Kreinin & Sidelnikova (2001) and references therein methods for obtaining a generator matrix from the matrix–logarithm are described and fall into two categories. Either a generator matrix is obtained directly from the matrix–logarithm by a certain ad–hoc manipulation rule for the elements of the matrix–logarithm or by approximation by some regularization algorithm. Finally, Vanden-Eijnden (2005) used spectral methods to obtain an estimator of the generator from the estimated transition matrix.

The paper is organized as follows. In Section 2 we present the statistical results needed for the analysis of discretely observed credit rating histories. In particular, we review and adapt the maximum likelihood and Markov chain Monte Carlo (MCMC) approaches. In Subsection 2.1 we describe the EM–algorithm and the discuss how to obtain standard errors of the estimators and confidence intervals of the intensities. In Subsection 2.2 we review the basic mechanism of the MCMC sampler and propose an importance sampler to improve the computational efficiency when dealing with the simulation of transitions that happen with small probabilities. Finally approximation of the information matrix as a byproduct of the MCMC algorithm is presented as an alternative way of obtaining confidence intervals of the EM–estimates. In Subsection 2.3 we discuss the parameter reduction problem that in the data analysis section turns out to be important to obtain satisfactory results from the MCMC approach, and in Subsection 2.4 we present a statistical test of the hypothesis that credit rating migrations in disjoint time periods are governed by the same generator matrix. Finally, in Section 3 we analyse credit risk data from the period 1990-2000 provided by Moody's Investor Service. The original data are daily records of the credit ratings, i.e. essentially continuous data. We investigate, in Subsection 3.1, the impact on the estimation of using as our data only the credit ratings at discrete points in time. In Subsection 3.2 we perform maximum likelihood estimation as well as MCMC estimation of transition rates in two different periods 1990-1995 and 1995-2000. Our test for equality of rates in the two periods indicates that the rates are different (Subsection 3.5). We also estimate transition probabilities and default probabilities, and in Subsection 3.3 we find confidence and credibility intervals of intensities and transition probabilities. When applying, in Subsection 3.4, our parameter reduction procedure in connection with the MCMC approach, we obtain similar results by the two approaches.

2 Statistical methods for discretely observed credit rating histories

In this section we generalize the results in Bladt & Sørensen (2005) on estimation of transition rates by the EM-algorithm and by an MCMC approach to observations from several independent Markov jump processes with the same transition rates. This is the setting needed to analyse credit rating data. The theory is extended in a number of directions. Thus we present methods for calculating confidence intervals for the transition rates and an importance sampling method that speeds up the simulation step in the MCMC-algorithm, which is useful for the analysis of credit rating data. We also discuss how to reduce the number of parameters by fixing some at zero and how to test for identical transition rates in different time periods. The theory is presented for the case where all transition rates are free parameters, but the methods can be adapted straightforwardly to the case where some transition rates are fixed (and thus need not be estimated). For applications to credit rating data, the transition rates away from the default state are usually set equal to zero.

Let $\boldsymbol{X}_i = \{X_t^i\}_{t\geq 0}$, $i = 1, ..., N$ be independent Markov jump processes with the same finite state space $E = \{1, \ldots, m\}$ and the same intensity matrix (infinitesimal generator) $\boldsymbol{Q} = \{q_{ij}\}$. In particular, q_{ij} is the rate of transition from state i to state j, and $q_{ii} = -\sum_{j\neq i} q_{ij}$. In the context of credit risk, E is the set of all possible credit ratings, and $\boldsymbol{X}_i$ is the credit rating history of the i'th firm/bond.

If the $\boldsymbol{X}_i$'s have been observed continuously in the time interval $[0, \tau]$, the likelihood function is given by

$$L^{(c)}(\boldsymbol{Q}) = \prod_{k=1}^{N} \prod_{i=1}^{m} \prod_{j\neq i} q_{ij}^{N_{ij}^k(\tau)} e^{-q_{ij}R_i^k(\tau)} = \prod_{i=1}^{m} \prod_{j\neq i} q_{ij}^{N_{ij}(\tau)} e^{-q_{ij}R_i(\tau)}, \tag{2.1}$$

where $N_{ij}^k(t)$ is the number of transitions from state i to state j in the time interval $[0, t]$ of the process $\boldsymbol{X}_k$,

$$R_i^k(t) = \int_0^t I\{X_s^k = i\} ds \tag{2.2}$$

is the time spent in state i before time t by the process $\boldsymbol{X}_k$, and

$$N_{ij}(t) = \sum_{k=1}^{N} N_{ij}^k(t), \quad R_i(t) = \sum_{k=1}^{N} R_i^k(t);$$

see e.g. Jacobsen (1982). It is straightforward to see that the maximum likelihood estimator of $\boldsymbol{Q}$ based on a continuous record of the credit rating histories is given for all $i \neq j$ by

$$\hat{q}_{ij}^{(c)}(\tau) = N_{ij}(\tau)/R_i(\tau), \tag{2.3}$$

provided that $R_i(\tau) > 0$. Obviously, $\hat{q}_{ii}(\tau) = -\sum_{j\neq i} \hat{q}_{ij}(\tau)$ for $i = 1, \ldots, m$.

Credit rating observations may be available only at certain discrete time points, e.g. annually. In the following we develop methods for estimating $\boldsymbol{Q}$ based on discrete time rating histories. Let $0 \leq t_1^i < ... < t_{n_i}^i \leq \tau$ be n_i time points at which the process $\boldsymbol{X}_i$ has been observed. The observed history $Y_i^k = X_{t_i^k}^k$ is a discrete time Markov chain, for which the transition matrix at time i is $P^{\Delta_i^k}(\boldsymbol{Q})$, where $\Delta_i^k = t_{i+1}^k - t_i^k$ and

$$P^t(\boldsymbol{Q}) = \exp(t\boldsymbol{Q}), \quad t > 0, \tag{2.4}$$

with $\exp(\cdot)$ denoting the matrix exponential function. Note that the discrete time process is time-inhomogeneous when the time points t_i^k are not equidistant.

The likelihood function for the discrete time data is then given by

$$L^{(d)}(\boldsymbol{Q}) = \prod_{k=1}^{N} \prod_{i=1}^{n_k-1} P^{\Delta_i^k}(\boldsymbol{Q})_{x_i^k x_{i+1}^k}, \tag{2.5}$$

where $x_1^k, \ldots, x_{n_k}^k$ denote the observed values of $\boldsymbol{X}_k$, and where $P^t(\boldsymbol{Q})_{ij}$ denotes the ijth entry of the transition matrix $P^t(\boldsymbol{Q})$.

An important special case is when $\Delta_i^k = \Delta$. In this case

$$L^{(d)}(\boldsymbol{Q}) = \prod_{i=1}^{m} \prod_{j=1}^{m} P^{\Delta}(\boldsymbol{Q})_{ij}^{K_{ij}},$$

where K_{ij} denotes the number of transitions from i to j observed in all $\boldsymbol{Y}_\ell = \{Y_k^\ell\}_{k=1,\ldots,n_\ell}$, $\ell = 1, \ldots, N$. The maximum likelihood estimator of $\boldsymbol{P} = \exp(\Delta \boldsymbol{Q})$ is hence given by

$$\hat{\boldsymbol{P}}_{ij} = \frac{K_{ij}}{K_{i.}}, \tag{2.6}$$

where $K_{i.} = \sum_{j=1}^{m} K_{ij}$. It can happen that $\hat{\boldsymbol{P}}$ is not the matrix exponential of an intensity matrix, but even in that case the maximum likelihood estimator of $\boldsymbol{Q}$ may well exist. Results on existence and consistency of the maximum likelihood estimator of $\boldsymbol{Q}$ can be found in Theorem 2.1 of Bladt & Sørensen (2005), which also holds in the present more general situation if $\hat{\boldsymbol{P}}$ is defined by (2.6).

In Bladt & Sørensen (2005) two methods of estimating $\boldsymbol{Q}$ were proposed both of which carry over to the present set up with only minor changes: the EM–algorithm and an MCMC approach.

2.1 The EM–algorithm

First we present the EM–algorithm. The essential step in the EM–algorithm is to find

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{Q}_0}(\log L^{(c)}(\boldsymbol{Q}) \,|\, \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N) \\
&= \sum_{i=1}^{m} \sum_{j \neq i} \log(q_{ij}) \mathbb{E}_{\boldsymbol{Q}_0}(N_{ij}(\tau) \,|\, \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N) \\
&\quad - \sum_{i=1}^{m} \sum_{j \neq i} q_{ij} \mathbb{E}_{\boldsymbol{Q}_0}(R_i(\tau) \,|\, \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N),
\end{aligned}$$

where $\boldsymbol{Q}$ and $\boldsymbol{Q}_0$ are typically different. For a given intensity matrix $\boldsymbol{Q}$ we define

$$\begin{aligned}
\tilde{M}_{ij}^{k}(\Delta, \boldsymbol{Q}) &= \mathbb{E}_{\boldsymbol{Q}}\left(R_k^1(\Delta) \,|\, X_\Delta^1 = j, X_0^1 = i\right) \\
\tilde{f}_{ij}^{k\ell}(\Delta, \boldsymbol{Q}) &= \mathbb{E}_{\boldsymbol{Q}}\left(N_{k\ell}^1(\Delta) \,|\, X_\Delta^1 = j, X_0^1 = i\right).
\end{aligned}$$

Since the N processes $\boldsymbol{X}_i$, $i = 1, ..., N$ are i.i.d.,

$$\mathbb{E}_{\boldsymbol{Q}}(N_{ij}(\tau) \,|\, \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N) = \sum_{\ell=1}^{N} \sum_{k=1}^{n_\ell-1} \tilde{f}_{y_k^\ell, y_{k+1}^\ell}^{ij}(t_{k+1}^\ell - t_k^\ell, \boldsymbol{Q}) \tag{2.7}$$

$$\mathbb{E}_{\boldsymbol{Q}}(R_i(\tau) \,|\, \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N) = \sum_{\ell=1}^{N} \sum_{k=1}^{n_\ell-1} \tilde{M}_{y_k^\ell, y_{k+1}^\ell}^{i}(t_{k+1}^\ell - t_k^\ell, \boldsymbol{Q}). \tag{2.8}$$

The quantities $\tilde{M}^k_{ij}$ and $\tilde{f}^{k\ell}_{ij}$ can be found using Theorems 3.1 and 3.2 together with formulas (3.7) and (3.10) in Bladt & Sørensen (2005).

We can now sum up the EM–algorithm for maximum likelihood estimation of the matrix of transition rates $\boldsymbol{Q}$ as follows:

Let $\boldsymbol{Q}_0$ be any intensity matrix for a Markov jump process with state space E. Initially set $\boldsymbol{Q} = \boldsymbol{Q}_0$.

1. Calculate $\tilde{M}^i_{y^\ell_k, y^\ell_{k+1}}(t^\ell_{k+1} - t^\ell_k, \boldsymbol{Q})$ and $\tilde{f}^{ij}_{y^\ell_k, y^\ell_{k+1}}(t^\ell_{k+1} - t^\ell_k, \boldsymbol{Q})$ for all i, j, ℓ, k.

2. Calculate $\mathbb{E}_{\boldsymbol{Q}}\left(R_i(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right)$ and $\mathbb{E}_{\boldsymbol{Q}}\left(N_{ij}(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right)$ by (2.7) and (2.8).

3. Calculate $\hat{\boldsymbol{Q}}$ by

$$\hat{\boldsymbol{Q}}_{ij} = \mathbb{E}_{\boldsymbol{Q}}\left(N_{ij}(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right) / \mathbb{E}_{\boldsymbol{Q}}\left(R_i(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right)$$

   for all $i \neq j$, cf. formula (2.3).

4. $\boldsymbol{Q} := \hat{\boldsymbol{Q}}$. GOTO 1.

Suppose the initial matrix $\boldsymbol{Q}_0$ belongs to the interior of the parameter space, i.e. that $(\boldsymbol{Q}_0)_{ij} > 0$ for all $i \neq j$. Then the sequence $\{\boldsymbol{Q}_k\}$ will either converge to a stationary point of the likelihood function $L^{(d)}$ or $\det(\exp(\boldsymbol{Q}_k)) \to 0$; for details see Bladt & Sørensen (2005). If the latter possibility occurs, it is an indication that the maximum likelihood estimator does not exist. A way around a possible non-existence problem is to use the MCMC-estimator presented below. Before doing so it should be considered whether a case of non-existence of the maximum likelihood estimator is perhaps an indication that the model does not fit the data well.

Estimates of the variances and covariances on the maximum likelihood estimators of the transition rates can be obtained as a by-product of the EM–algorithm by a method proposed by Oakes (1999). More precisely, the observed information matrix (minus the matrix of second order partial derivatives with respect to the parameters of the log-likelihood function), which is an $m(m-1) \times m(m-1)$-matrix, can be calculated as follows. The diagonal element corresponding to q_{ij} is given by

$$\begin{aligned}
J(\boldsymbol{Q})_{ij,ij} &= \frac{1}{q_{ij}^2} \mathbb{E}_{\boldsymbol{Q}}\left(N_{ij}(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right) \\
&\qquad - \frac{1}{q_{ij}} \frac{\partial}{\partial q_{ij}} \mathbb{E}_{\boldsymbol{Q}}\left(N_{ij}(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right) \\
&\qquad\qquad + \frac{\partial}{\partial q_{ij}} \mathbb{E}_{\boldsymbol{Q}}\left(R_i(\tau) | \boldsymbol{Y}_1 = \boldsymbol{y}_1, ..., \boldsymbol{Y}_N = \boldsymbol{y}_N\right) \\
&= \sum_{\ell=1}^{N} \sum_{k=1}^{n_\ell - 1} \left(\frac{1}{q_{ij}^2} \tilde{f}^{ij}_{y^\ell_k, y^\ell_{k+1}}(t^\ell_{k+1} - t^\ell_k, \boldsymbol{Q}) - \frac{1}{q_{ij}} \frac{\partial}{\partial q_{ij}} \tilde{f}^{ij}_{y^\ell_k, y^\ell_{k+1}}(t^\ell_{k+1} - t^\ell_k, \boldsymbol{Q}) \right. \\
&\qquad\qquad \left. + \frac{\partial}{\partial q_{ij}} \tilde{M}^i_{y^\ell_k, y^\ell_{k+1}}(t^\ell_{k+1} - t^\ell_k, \boldsymbol{Q}) \right).
\end{aligned}$$

The off-diagonal element of the observed information matrix corresponding to q_{ij} and $q_{i'j'}$ is given by

$$\begin{aligned} J(\boldsymbol{Q})_{ij,i'j'} &= -\frac{1}{q_{ij}}\frac{\partial}{\partial q_{i'j'}}\mathbb{E}_{\boldsymbol{Q}}\left(N_{ij}(\tau)|\boldsymbol{Y}_1=\boldsymbol{y}_1,...,\boldsymbol{Y}_N=\boldsymbol{y}_N\right) \\ &\qquad\qquad + \frac{\partial}{\partial q_{i'j'}}\mathbb{E}_{\boldsymbol{Q}}\left(R_i(\tau)|\boldsymbol{Y}_1=\boldsymbol{y}_1,...,\boldsymbol{Y}_N=\boldsymbol{y}_N\right) \\ &= -\sum_{\ell=1}^{N}\sum_{k=1}^{n_\ell-1}\left(\frac{1}{q_{ij}}\frac{\partial}{\partial q_{i'j'}}\tilde{f}^{ij}_{y^\ell_k,y^\ell_{k+1}}(t^\ell_{k+1}-t^\ell_k,\boldsymbol{Q}) - \frac{\partial}{\partial q_{i'j'}}\tilde{M}^i_{y^\ell_k,y^\ell_{k+1}}(t^\ell_{k+1}-t^\ell_k,\boldsymbol{Q})\right). \end{aligned}$$

For details see Oakes (1999). The inverse of the information matrix evaluated at maximum likelihood estimator, $J(\hat{\boldsymbol{Q}})^{-1}$, is an estimator of the covariance matrix of the vector of all maximum likelihood estimators of the transition rates $\mathrm{vec}(\hat{q}_{ij}, i \neq j) = (\hat{q}_{12}, \hat{q}_{13}, \dots, \hat{q}_{m,m-1})$. In particular, $(J(\hat{\boldsymbol{Q}})^{-1})_{ij,ij}$ is an estimator of the variance of $\hat{q}_{ij}$. The derivatives of the conditional expectations must be calculated numerically. This must be done with great care, and can in some cases cause problems. We shall return to this issue in the empirical section. An alternative and easier way to determine the observed information matrix will be given in the following discussion of the MCMC-approach.

The information matrix is mainly useful when the estimators are approximately normal distributed. Long time series will imply approximate normality of the estimators, but the time series used in the analysis of credit rating histories are typically not long in order to avoid the problem of time-varying transition rates. However, approximate normality of the estimators can also be a large sample effect stemming from the fact that the data comprise the credit history of a large number of companies.

2.2 Markov chain Monte Carlo

In the MCMC-approach we choose a prior density $\phi(\boldsymbol{Q})$ and are interested in the conditional distribution of $\boldsymbol{Q}$ given the data $\boldsymbol{x} = \{x^k_i | \, i = 1, \dots, n_k, k = 1, \dots, N\}$. In fact, this method provides draws from the conditional distribution of $(\boldsymbol{Q}, \boldsymbol{X})$ given $\boldsymbol{x}$, where $\boldsymbol{X} = \{X^k_t | \, 0 \leq t \leq \tau, k = 1, \dots, N\}$ is the collection of continuous time sample paths of the processes $\boldsymbol{X}_k$, $k = 1, \dots, N$. For this we employ the Gibbs sampler with two sites $\boldsymbol{Q}$ and $\boldsymbol{X}$ and sample by alternately drawing $\boldsymbol{X}$ given $(\boldsymbol{Q}, \boldsymbol{x})$ and $\boldsymbol{Q}$ given $(\boldsymbol{X}, \boldsymbol{x})$ ($\boldsymbol{x}$ is of course of no importance when conditioning on $\boldsymbol{X}$). Iteration of the Gibbs sampler results in a sequence of variables $(\boldsymbol{Q}_n, \boldsymbol{X}^{(n)})$. Under suitable conditions the Gibbs sampler will eventually produce a stationary and ergodic sequence, i.e., after discarding a certain burn–in period, say the first K-1 iterations, the sequence $(\boldsymbol{Q}_n, \boldsymbol{X}^{(n)}))_{n \geq K}$ may be considered stationary, and the stationary distribution is exactly the conditional distribution of $(\boldsymbol{Q}, \boldsymbol{X})$ given $\boldsymbol{x}$. By ergodicity, the empirical average

$$\frac{1}{M}\sum_{i=K}^{M+K}\boldsymbol{Q}_i$$

converges to the true mean of $\boldsymbol{Q}$ conditionally on $\boldsymbol{x}$. Credibility intervals based on the empirical distribution of $(\boldsymbol{Q}_n, \boldsymbol{X}^{(n)})_{n \geq K}$ and quantiles of the empirical distribution can be found.

The intensity matrix $\boldsymbol{Q}$ is not always uniquely determined by the distribution of the discrete time sample $\boldsymbol{x}$ in the sense that two or more intensity matrices may imply the same distribution of the data. In such situations the mean of the posterior distribution may not be a meaningful quantity, and credibility intervals do not make sense for parameters which are not uniquely determined. The set of $\boldsymbol{Q}$s for which this happens is complicated, so it is important to study the posterior distribution carefully for indications that this problem has occurred, for instance by inspecting scatter plots. However, in situations where the distribution of the data is the same for more than one intensity matrix, functionals of $\boldsymbol{Q}$ that are invariant under different representations of the distribution of the data can be estimated using a method similar to that just described. Specifically, let F be some functional which depends on the distribution of the data and is invariant under changes of the representation by $\boldsymbol{Q}$ (i.e. if $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ result in the same distribution of the data, then $F(\boldsymbol{Q}_1) = F(\boldsymbol{Q}_2)$). Then we can estimate $F(\boldsymbol{Q})$ by

$$\frac{1}{M} \sum_{i=K}^{M+K} F(\boldsymbol{Q}_i).$$

For instance, the transition matrix $P^t(\boldsymbol{Q}) = \exp(t\boldsymbol{Q})$ can be estimated in this way if t is the time between two observations in the data (or a multiple of such a time).

In our experience, the prior

$$\phi(\boldsymbol{Q}) \propto \prod_{i=1}^{m} \prod_{j \neq i} q_{ij}^{\alpha_{ij}-1} e^{-q_{ij}\beta_i}, \tag{2.9}$$

works well. Here $\alpha_{ij} > 0, i, j \in E$ and $\beta_i > 0, i \in E$ are constants to be chosen conveniently. Obviously, q_{ij} is gamma distributed with shape parameter α_{ij} and scale parameter $1/\beta_i$, and the q_{ij}s are independent. In this way parameters near the critical boundary, where $\det(\exp(\boldsymbol{Q})) = 0$, are effectively penalized because there at least one of the q_{ij}s must go to infinity. Thus problems of non-existence of the estimator are avoided. This family of priors is conjugate for the model for continuous observation in the time interval $[0, \tau]$, which is an exponential family of processes, see Küchler & Sørensen (1997). Indeed, the posterior based on continuous observation is again independent gamma distributions:

$$p^*(\boldsymbol{Q}) = L^{(c)}(\boldsymbol{Q})\ \phi(\boldsymbol{Q}) \propto \prod_{i=1}^{m} \prod_{j \neq i} q_{ij}^{N_{ij}(\tau)+\alpha_{ij}-1} e^{-q_{ij}(R_i(\tau)+\beta_i)}, \tag{2.10}$$

where the likelihood function $L^{(c)}(\boldsymbol{Q})$ is given by (2.1).

The Gibbs sampler now works as follows.

1. Draw an initial $\boldsymbol{Q}$ from the prior.

2. Simulate the Markov jump processes $\boldsymbol{X}_k$, $k = 1, \ldots, N$ independently with intensity matrix $\boldsymbol{Q}$ up to time τ such that $X^k_{t^k_i} = x^k_i$, $i = 1, \ldots, n_k$, $k = 1, \ldots, N$.

3. Calculate the statistics $N_{ij}(\tau)$ and $R_i(\tau)$ from the simulated sample paths $\{X^k_t | 0 \leq t \leq \tau, k = 1, \ldots, N\}$.

4. Draw a new $\boldsymbol{Q}$ from the posterior distribution (2.10).

5. GO TO 2.

Step two requires a simulation of the Markov jump processes step–by–step through the intervals $[t_i^k, t_{i+1}^k]$ starting from the initial condition $X^k_{t_i^k} = x_i^k$ such that the process will be in the state $X^k_{t_{i+1}^k} = x_{i+1}^k$ at time t_{i+1}^k. This can usually be done by simple rejection sampling where sample paths that do not hit the right state are discarded. This has turned out to be quite efficient, but there may be situations where the transition probability between two states is so low that simple rejection sampling is no longer feasible. To improve the efficiency in such cases, we propose the following *importance sampling* scheme based on the Metropolis–Hastings algorithm.

Suppose that we need to simulate a trajectory from a Markov jump process with intensity matrix $\boldsymbol{Q}$ starting from some state i and ending in state j during a time period of length Δ. Instead of simulating the Markov jump process using the intensity matrix $\boldsymbol{Q}$, we sample instead using a 'neutral' matrix $\tilde{\boldsymbol{Q}}$ given by

$$\tilde{\boldsymbol{Q}} = \frac{1}{\alpha}\begin{pmatrix} -n & 1 & 1 & ... & 1 & 1 \\ 1 & -n & 1 & ... & 1 & 1 \\ . & . & . & ... & . & . \\ 1 & 1 & 1 & ... & -n & 1 \\ 1 & 1 & 1 & ... & 1 & -n \end{pmatrix},$$

where α is a scaling factor to be chosen appropriately to match more or less the overall order of magnitude of the intensities in $\boldsymbol{Q}$. Note that if one of the intensities has been fixed at zero, the same entry in the matrix $\tilde{\boldsymbol{Q}}$ should be modified to equal zero, and one should be added to the corresponding diagonal element of $\tilde{\boldsymbol{Q}}$. For instance, if the default state is the last state, then all entries in the last row of $\tilde{\boldsymbol{Q}}$ should equal zero. When simulating the Markov process with neutral intensity matrix $\tilde{\boldsymbol{Q}}$ the simple rejection sampling described above will produce a trajectory starting from i and ending in j without problems. If $\boldsymbol{X}_k$ is a trajectory in the interval $[0, \Delta]$ of a Markov jump process drawn subject to $\tilde{\boldsymbol{Q}}$ (starting from i and ending in j), then the importance weight is proportional to

$$w(\boldsymbol{X}_k) = \frac{L_k^{(c)}(\boldsymbol{Q}; \boldsymbol{X}_k)\exp(\Delta\tilde{\boldsymbol{Q}})_{ij}}{L_k^{(c)}(\tilde{\boldsymbol{Q}}; \boldsymbol{X}_k)\exp(\Delta\boldsymbol{Q})_{ij}},$$

where

$$L_k^{(c)}(\boldsymbol{Q}; \boldsymbol{X}_k) = \prod_{i=1}^{m}\prod_{j\neq i} q_{ij}^{N_{ij}^k(\Delta)} e^{-q_{ij}R_i^k(\Delta)},$$

and where $\exp(\Delta\boldsymbol{Q})_{ij}$ is the probability of going from i to j in time Δ when the intensity matrix is $\boldsymbol{Q}$. The Metropolis–Hastings acceptance probability equals the minimum of 1 and the ratio between the importance weights of two successive draws $\boldsymbol{X}$ and $\boldsymbol{Y}$, i.e.

$$P_{\text{accept}} = \min\left(1, \frac{w(\boldsymbol{Y})}{w(\boldsymbol{X})}\right).$$

The Metropolis–Hastings algorithm then works by successively drawing Markov jump processes starting at i and ending in j subject to the intensity matrix $\tilde{\boldsymbol{Q}}$ (by simple rejection sampling) and accepting any new draw with probability P_{accept}. This produces an sequence of trajectories that after a burn–in period can be considered stationary with the stationary distribution equal to the distribution of a trajectory of the Markov jump process with intensity matrix $\boldsymbol{Q}$.

Bladt & Sørensen (2005) presented the MCMC algorithm in the more general setting of a process with hidden states. It is therefore possible to apply the MCMC approach to the model with hidden "excited" states proposed by Christensen, Hansen & Lando (2004) in order to capture non-Markovian effects in credit rating data. Models of this type will be consedered in future work.

Let us complete this subsection by discussing how well the parameters are determined. One way to quantify this is to give, for each parameter, a set that has 95 per cent posterior probability. This can be done in many ways. In most cases, we just give the interval between the 2.5 per cent quantile and the 97.5 per cent quantile of the posterior distribution, which is referred to as the credibility interval. In connection with the parameter reduction discussed in the next subsection, we will use the set of parameter values where the posterior density is largest. If $\pi_{ij}(q)$ is the marginal posterior density of q_{ij}, then this set is given by $C_{ij} = \{q \,|\, \pi_{ij}(q) \geq c\}$, where c is such that $\int_{C_{ij}} \pi_{ij}(q)\, dq = 0.95$. This set is sometimes referred to as a 95 per cent Bayesian confidence set.

Another way to quantify how well the parameters are determined is via the observed information matrix, $J(\hat{\boldsymbol{Q}})$; see Subsection 2.1. When the number of observations is large, the posterior distribution is approximated by the normal distribution with mean equal to the maximum likelihood estimator $\hat{\boldsymbol{Q}}$ and covariance matrix equal to the inverse of the observed information matrix $J(\hat{\boldsymbol{Q}})^{-1}$. We can therefore estimate $J(\hat{\boldsymbol{Q}})^{-1}$ from the sequence of intensity matrices $\boldsymbol{Q}_i$ produced by the Gibbs sampler:

$$J(\hat{\boldsymbol{Q}})^{-1} \sim \frac{1}{M} \sum_{i=K+1}^{M+K} \mathrm{vec}(\boldsymbol{Q}_i - \bar{\boldsymbol{Q}})\mathrm{vec}(\boldsymbol{Q}_i - \bar{\boldsymbol{Q}})^*,$$

where $\bar{\boldsymbol{Q}} = \frac{1}{M}\sum_{i=K+1}^{M+K} \boldsymbol{Q}_i$ and where * denotes transposition. Here $\mathrm{vec}(\cdot)$ indicates that we represent intensity matrices and other matrices as column vectors of dimension $m(m-1)$. As previously, the first K draws of the intensity matrix represent a burn-in period. If the maximum likelihood estimator has been determined by the EM–algorithm, we can also estimate the inverse information matrix by

$$J(\hat{\boldsymbol{Q}})^{-1} \sim \frac{1}{M} \sum_{i=K}^{M+K} \mathrm{vec}(\boldsymbol{Q}_i - \hat{\boldsymbol{Q}})\mathrm{vec}(\boldsymbol{Q}_i - \hat{\boldsymbol{Q}})^*. \tag{2.11}$$

This method only makes sense when $\boldsymbol{Q}$ is uniquely determined by the distribution of the data. In the data analysis section below we shall see that this rather straightforward way of estimating the inverse information matrix compares well with the much more complex procedure described in the subsection on the EM–algorithm.

2.3 Parameter reduction

If one or more intensity parameters can be assumed to be zero, we can fix them at zero and thus reduce the number of parameters. Replacing non–significant values by zero is important in the MCMC approach because the prior distribution in such cases has a very considerable influence on the estimates as there is no or very little information available in the data about transitions that have not been observed. In such cases the prior also affects the estimate of the total transition rate out of a state (the diagonal entries of the intensity matrix).

A simple procedure would be to use the maximum likelihood estimator and the observed information matrix (obtained by the EM–algorithm or the MCMC procedure) to make marginal approximate confidence intervals and then fix a parameter at zero whenever its confidence interval contains zero. We will, however, not use this procedure because there is a multiple testing problem, and we cannot be sure that the standard asymptotic theory applies under the null hypothesis that a parameter is equal to zero. Instead we will use the posterior distribution produced by the MCMC approach to reduce the number of parameters. A credibility interval between two quantiles, e.g. the 2.5 and 97.5 per cent quantiles, can never contain zero, and parameters drawn from the posterior distribution will always be strictly positive even in situations where there are no transitions between the corresponding states in the data. We will therefore use the Bayesian confidence sets discussed in the previous subsection. If the posterior density goes continuously down to zero, this set will not contain zero, but otherwise zero can be a boundary point of the confidence set. We will fix those parameters at zero for which the Bayesian confidence interval has zero as a boundary point, provided that the corresponding transition has not been observed in the data set. In this way we exclude the cases where the prior distribution has an undue influence.

To implement this procedure, we need an expression for the marginal posterior densities. One possibility is to estimate the posterior density by a density estimator based on the draws from the posterior distribution produced by the Gibbs sampler. Another way to obtain a smooth approximation to the posterior density is via the continuous time posterior. If we use a gamma prior, the marginal continuous time posterior distributions of the transition rates are gamma distributions too. The MCMC simulation produces a sequence of realizations of the continuous time process. For each of these we calculate the sufficient statistics $N_{ij}(\tau)_\nu$ and $R_i(\tau)_\nu$, $\nu = 1, \ldots, M+K$. Here $N_{ij}(\tau)_\nu$ is the number of jumps from state i to state j of all the processes at the ν'th simulation step, and similarly for $R_i(\tau)_\nu$. As previously, K is the length of a burn-in period. Let $\gamma(t;\alpha,\beta)$ denote the gamma density with parameters α and β. Then an approximation to the marginal posterior density of q_{ij} based on the discrete time data is given by

$$\hat{p}_{ij}(q) = \frac{1}{M} \sum_{\nu=K+1}^{M+K} \gamma(q; \alpha_{ij} + N_{ij}(\tau)_\nu, \beta_i + R_i(\tau)_\nu)), \tag{2.12}$$

where α_{ij} and β_i are the parameters in the prior of q_{ij}. This average approximates the discrete time posterior density because this is the conditional expectation given the data of the continuous time posterior density. The latter method is much more elaborate than the former as we have to keep track of the sufficient statistics from all iterations, but it has the advantage of not depending on an arbitrary choice of smoothness in the density estimator.

After deciding which parameters should be fixed at zero, we need to re–estimate the remaining parameters using the EM and/or the MCMC algorithm on the reduced parameter space. This is in principle straightforward but may be time consuming in practice since it involves invoking the algorithms from scratch.

2.4 Statistical test for identical intensity matrices in different periods

A crucial assumption in the Markov model is the time homogeneity of transition rates. For data observed over longer periods of time one may test whether this assumption is reasonable by splitting the data into two sub–periods and then testing whether the estimated intensity matrices for the two periods are equal in the following way.

Let $\hat{\Lambda}$ and $\hat{\Gamma}$ be maximum likelihood estimators of the intensity matrix in the first and second period, respectively. The two matrices can also be other estimators that are asymptotically equivalent to the maximum likelihood estimators such as mean posterior estimators. Let Λ and Γ be the true values of the intensity matrices in the two periods. In the following the estimated and true intensity matrices are written as $m(m-1)$-dimensional column vectors (shorter vectors in case some parameters have been fixed at zero), not as $m \times m$-matrices. As earlier we indicate this by writing for instance $\text{vec}(\Lambda)$. Let n_{ij} be the number of time points at which the process $\boldsymbol{X}_i$ was observed in sub–period j ($j = 1, 2$), and define $\nu_j = n_{1j} + \cdots + n_{Nj}$. When ν_1 and ν_2 are large,

$$\text{vec}(\hat{\Lambda}) \dot{\sim} N(\text{vec}(\Lambda), i_1(\Lambda)^{-1}) \quad \text{and} \quad \text{vec}(\hat{\Gamma}) \dot{\sim} N(\text{vec}(\Gamma), i_2(\Gamma)^{-1}),$$

where $\dot{\sim}$ denotes approximate distribution, and $i_j(\Gamma)$ is the Fisher information matrix at the parameter value Γ in period j. Under the hypothesis that $\Lambda = \Gamma$, we have that

$$\text{vec}(\hat{\Lambda} - \hat{\Gamma})^*(i_1(\Lambda)^{-1} + i_2(\Lambda)^{-1})^{-1}\text{vec}(\hat{\Lambda} - \hat{\Gamma}) \dot{\sim} \chi^2(m(m-1)),$$

where $*$ denotes transposition. If some transitions have been fixed, the degrees of freedom should be reduced to the number of free transition rates.

In order to be able to use this test statistic, we need to determine the Fisher information matrix $i(\Lambda)$. We will approximate it by the observed information matrix, which can be found by MCMC methods as well as from the EM–algorithm, as previously explained. Here we briefly outline the MCMC approach. The MCMC simulation produces a sequence $\Lambda_1, \ldots, \Lambda_{K+M}$ of draws from the posterior distribution of Λ (conditional on our data), and a sequence $\Gamma_1, \ldots, \Gamma_{K+M}$ of draws from the posterior distribution of Γ (K refers to a burn-in period). When ν_1 and ν_2 are large, the two posterior distributions are approximated by $N(\text{vec}(\hat{\Lambda}), J_1(\hat{\Lambda})^{-1})$ and $N(\text{vec}(\hat{\Gamma}), J_2(\hat{\Gamma})^{-1})$, respectively. Here $J_1(\hat{\Lambda})$ and $J_2(\hat{\Gamma})$ are the observed information matrices in the two periods evaluated at the maximum likelihood estimators. Under the hypothesis of identical intensity matrices we can therefore estimate $(i_1(\Lambda)^{-1} + i_2(\Lambda)^{-1})$ by

$$V = \frac{1}{M} \sum_{j=K+1}^{M+K} \text{vec}(\Lambda_j - \hat{\Lambda} - \Gamma_j + \hat{\Gamma})\text{vec}(\Lambda_j - \hat{\Lambda} - \Gamma_j + \hat{\Gamma})^*.$$

We will use the test statistic

$$\text{vec}(\hat{\Lambda} - \hat{\Gamma})^* V^{-1} \text{vec}(\hat{\Lambda} - \hat{\Gamma}) \dot{\sim} \chi^2(m(m-1)) \tag{2.13}$$

to test the hypothesis that the intensity matrices in the two periods are identical.

When the intensities are small, the distribution of the estimates of the intensities can be very skewed if the sample size is not sufficiently large, as we shall see in the data analysis section. In such cases it is natural to improve the normality by a logarithmic transformation, since the approximate normality of the estimators is crucial to the performance of the test. A test can be performed in exactly the same way as just outlined with estimates of the intensities replaced by their logarithms and with the information matrices based on the logarithm of the draws from the posterior distribution.

3 Analysis of credit risk data

We analyse credit rating data drawn from Moody's Corporate Bond Default Database. The data is a continuous record of the ratings of 2749 issuers/firms in the period of January 1, 1990 to December 31, 1999. The period is split in the two subperiods of 1990-1995 and 1995-2000 where the transition rates are expected to be different because of business cycle effects. This is in accordance with Christensen, Hansen & Lando (2004). A continuous record means that transitions of firms between different ratings are recorded at the dates they happened and that the record is complete in the sense that all changes are registered. Hence continuous record means daily observations. The grades of the ratings are merged into the following groups: Aaa, Aa, A, Baa, Ba, B, C, and D, where all sub–categories with appended number 1,2 or 3 (additional notches introduced in the early 80s) are merged into the corresponding main category. Furthermore, C contains the following main categories Caa, Ca, and C. This reduction of the number of categories was done in order to increase the number of observations per state. The same simplification was done by Christensen, Hansen & Lando (2004). The letter D refers to default and is treated as an absorbing state. Cases in which a transition from the state D to another state is observed are censored at the time of default (absorption). Withdrawals (WR) are observed on numerous occasions. We consider observations of rating histories up to the time of withdrawal as valid data.

3.1 The effect of the sampling frequency

First we study how the estimators of the intensities depend on how often credit ratings are observed. Instead of using the continuous record of ratings, we used only observations at discrete time points separated by $\Delta = 1, 7, 15, 30, 90, 180, 365$ (days), ranging from daily data (continuous data) to annual observations. Specifically, we estimated the transition intensity matrix based on the credit ratings of all bonds as they were on the days numbered $1 + i\Delta$, $i = 1, 2, \ldots$. Assuming that rating grades migrate according to (possibly different) 8–state Markov jump process in the periods 1990-1995 and 1995-2000, we calculated the maximum likelihood estimators of the transition rates by the EM–algorithm. A total of 250 iterations of the EM–algorithm were performed for each value of Δ. Convergence to a six decimal precision could be observed already after about 100 iterations. The estimated rates for selected transitions are displayed for all Δ values in Table A.1 (see the Appendix). The transitions were selected according to whether they are likely to be significantly different from zero. There is no systematic trend for the estimated transition intensities as Δ increases because it is random which transitions in the original data are not included in each of the subsamples. The estimated rates do, of course, change as the sampling frequency changes. Some are, however, remarkably stable, while only in a few cases the order magnitude has changed. In the rest of the paper we will analyse annual observations to investigate how much information can be obtained from such relatively sparse data on transition rates and default probabilities. The analysis could obviously also be done with more frequent observations.

3.2 Estimates of transition rates and transition probabilities

The maximum likelihood estimates of the intensity matrices based on annual observations obtained by the EM–algorithm are displayed in Table B.1 and Table C.1 for the two periods 1990-1995 and 1995-2000 respectively. Tables B.3 and C.3 give the one year transition probabilities for the two periods calculated by (2.4) using the EM estimates of the transition rates. In particular, the last column of the Tables B.3 and C.3 are the one-year default probabilities given the issuer's present rating.

Estimation of the intensity matrix using the Markov chain Monte Carlo approach was done using annual observations from each of the two periods 1990-1995 and 1995-2000. In either case the gamma prior was chosen with parameters $\beta = 5.0$ and $\alpha = 1$. This choice, as opposed to e.g. $\alpha = 1, \beta = 1.0$, enabled a quicker burn–in. Estimates of transition rates that are significantly different from zero are not sensitive to the choice of the parameters in the gamma prior, while the MCMC estimates of rates that are close to zero and perhaps not significantly different from zero will depend on the choice of the prior. An initial burn–in of 1000 iterations was discarded, and the inference was based on the subsequent 4000 iterations. The series settled into a stationary mode after an initial 100-200 iterations, which was concluded on the basis of visual inspection of sample paths and evaluation of auto–correlation and partial auto–correlation functions that were seen to decay very quickly to zero.

From the two series of the remaining 4000 iterations, posterior estimates were obtained by averaging the corresponding 4000 intensity matrices. The results are shown in the tables B.2 and C.2, respectively. Here all transitions are included. If we compare the maximum likelihood estimates and the mean posterior estimates of the intensities where the EM estimates are non-zero, we see a striking resemblance. However, intensities for which the EM estimates are zero, and which therefore are expected to be non–significant, all have mean posterior estimates that are relatively small, but sufficiently large to affect the calculation of transition probabilities and in particular default probabilities. The MCMC estimates of these intensities are strongly influenced by the prior distribution as the data contain little information about such transitions. As a consequence the overall jump rate out of the different states are also over–estimated. In tables B.4 and C.4 we calculate the one–year transition probability matrices by averaging the one–year transition probability matrices calculated by (2.4) from each draw of the intensity matrix. The last column of the Tables B.4 and C.4 are the one-year default probabilities given the issuer's present rating. Comparison to the one–year transition probability matrices obtained by the EM–algorithm reveal that the many non-zero intensities implied by the prior do change the transition probabilities considerably – in particular the small probabilities.

From a statistical point of view it is obviously undesirable that the estimation of certain parameters is so strongly influenced by their priors. One way around this problem is to try to identify such parameters and then fix their value at zero, thus leaving them out of the MCMC-estimation. We shall return to this solution later. On the other hand there are regulations, such as Basle II, that impose certain minimum rates of transition even though they cannot be empirically justified. Such regulatory demands might be built into the priors.

3.3 Confidence intervals and credibility intervals

In this subsection we apply and compare two methods of determining how well the transition rates are determined by the data. We also discuss how well the one-year transition probabilities and the default probabilities are determined.

Tables D.1 and E.1 contain confidence intervals of intensities based on the estimator variances obtained by the method of Oakes (1999) outlined in Subsection 2.1. Confidence intervals are not available for all intensities. There are two reasons for this. Some parameters converged to zero very quickly (in 3–4 iterations). Once a parameter in the EM–algorithm is zero it will remain so in the rest of the iterations. When the maximum likelihood estimator is on the boundary of the parameter set, we cannot be sure that the usual asymptotic theory holds. Therefore we have treated these intensities as fixed at zero. Another problem is that the calculation of the observed information matrix by Oakes' method involves numerical differentiation with respect to the parameters. We applied the simple method of approximating the derivative by a two–sided difference quotient ($f'(x) \approx (f(x+h) - f(x-h))/2h$). This is about the worst method available for calculating a derivative numerically, but since the evaluation of the functions to be differentiated (conditional expectations given data) are computationally very expensive, we have refrained from invoking more sophisticated methods that would have involved several evaluations of the function. Our type of approximation is very sensitive to the choice of h, and in general h must not be chosen too small. One sided difference quotients are even more unstable and were not considered in this paper. For this reason we only considered intensities with estimates greater than the chosen value of h. Also a few of these had to be excluded from the calculation because the numerical derivatives were too unstable. Intensities with estimates smaller than h or otherwise unstable derivatives were assumed to be fixed, and an observed information matrix was calculated for the remaining intensities and inverted to obtain estimates of the asymptotic variances. The calculated variances are thus only approximations in that some intensities have been treated as known.

In general, some of the variance estimates turned out to be very sensitive to the step–size of the approximating difference quotient, while others were more robust. One was even negative and has been omitted. Interestingly the robust variance estimates were well approximated by diagonal elements of the MCMC estimate of the inverse Fisher information discussed in Subsection 2.2, see the comparison in Table D.4. Since the latter estimates of the asymptotic variances are easy and quick to calculate, this is our main reason to abstain from employing more sophisticated methods for the numerical differentiation in the Oakes procedure. We recommend using the MCMC estimates of the asymptotic variances of the maximum likelihood estimators.

Another problem with the EM–algorithm is that estimates of the asymptotic variances of the maximum likelihood estimators of the one–year transition probabilities, and hence confidence intervals for these probabilities, are not readily available because of the transformation of the intensity estimates by a matrix exponential function. However, bootstrap methods could be applied as in Christensen, Hansen & Lando (2004).

In tables D.2 and E.2 we display the 95% credibility intervals for the two periods, which are the intervals between the empirical 0.025 and 0.975 quantiles of the posterior distributions. For the intensities where both a confidence interval and a credibility interval are available, the two intervals are in almost all cases very similar. Credibility intervals for the one–year transition probabilities can be easily obtained and are given in Tables D.3 and E.3. By applying the matrix exponential to each MCMC draw of an intensity matrix, a sample from the posterior distribution of the transition probability matrix is obtained and empirical quantiles can be calculated.

3.4 Parameter reduction

In order to identify intensities for which the prior has an unreasonably strong influence on the posterior distribution, and which can safely be assumed to be zero, we used the simple parameter reduction method described in Subsection 2.3. First we identified the intensities for which zero is a boundary point of the Bayesian confidence interval. It is not an easy numerical task to calculate the exact limits of the Bayesian confidence intervals from the expression (2.12) for the posterior density. Fortunately, this problem could be avoided, as visual inspection of the plots of the posterior density (2.12) for every parameter was enough to determine for which intensities zero is a boundary point of the Bayesian confidence interval. Obviously, we did not want to give the value zero to intensities of transitions that has actually been observed to happen, so among the intensities with zero as a boundary point of the Bayesian confidence interval we only put those equal to zero for which the corresponding transition was not present in the data. These intensities were eliminated from the list of parameters. Then the model with fewer parameters was re–estimated and the parameter reduction procedure was performed again. It might be desirable to modify the procedure by not fixing certain intensities, for instance to ensure a certain structure of the intensity matrix or to ensure that all states (but the absorbing default state) communicate.

The MCMC estimates of the intensities after this procedure are shown in the Tables F.1 and G.1, while credibility intervals of the intensities are given in Tables F.2 and G.2. MCMC estimates and credibility intervals of the one-year transition probabilities, including the default probabilities, are given in the Tables G.3, G.3, G.4, and G.4. The MCMC results obtained after the parameter reduction procedure are in a much better accordance with the results obtained by maximum likelihood estimation. The advantage over the EM-algorithm is that here we have credibility intervals for all non-zero intensities.

3.5 Test for equal intensity matrices of the two periods

Until now we have worked under the assumption that in the periods 1990-1995 and 1995-2000 the transition rates are different, and therefore the data from these two periods we considered separately. In this subsection we test whether the transition pattern is significantly different in the two periods.

To this end we employ the statistical test for equal intensity matrices outlined in Subsection 2.4. One of the key assumptions in the test is the approximate normality of the estimates of the vector of intensities. The MCMC draws from the posterior distribution of the intensities showed strong deviations from normality (for small intensities), which indicated that the intensity estimates are also non-normal. Therefore we applied a normalizing logarithmic transformation to the estimates of the intensities as well as to the draws from the posterior distribution used to estimate the information matrix. Inspection of Q-Q plots showed that the logarithmically transformed draws from the posterior of the intensities were approximately normal. It seems recommendable to employ a logarithmic transformations whenever the intensities are small and their distribution consequently very skew.

Performing a test for equal log–intensities in complete analogy to the test for equal intensity matrices, we obtain a test statistic of 67.1. As it is asymptotically $\chi^2(49)$ distributed, the level of significance is 0.04, so that the hypothesis of equal intensity matrices in the two periods is (just) rejected at the 5% level. Note that the number of degrees of freedom here is different from the number $m(m-1)$ in (2.13) because the final row of the intensity matrix is the zero vector, so the number of free parameters is 49.

Acknowledgements

We are grateful to Moody's Investor Service for permission to use the data set and to David Lando for putting the data at our disposal and for helpful discussions and advice. The research of Michael Sørensen was supported by the Centre for Analytical Finance and the Danish Mathematical Finance Network, both financed by the Danish Social Science Research Council.

References

Bladt, M. & Sørensen, M. (2005). "Statistical inference for discretely observed Markov jump processes". *J. R. Statist. Soc. B*, 67:395 – 410.

Christensen, J. H. E.; Hansen, E. & Lando, D. (2004). "Confidence sets for continuous-time rating transition probabilities". *Journal of Banking and Finance*, 28:2575 – 2602.

Culver, W. J. (1966). "On the existence and uniqueness of the real logarithm of a matrix". *Proc. Am. Math. Soc.*, 17:1146–1151.

Elfving, G. (1937). "Zur Theorie der Markoffschen Ketten". *Acta Soc. Sci. Finnicae A*, 2.

Israel, R. B.; Rosenthal, J. S. & Wei, J. Z. (1997). "Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings". *Rev. Financial Stud.*, 10:481–523.

Jacobsen, M. (1982). *Statistical Analysis of Counting Processes*. Springer-Verlag, New York. Lecture Notes in Statistics 12.

Jarrow, R. A.; Lando, D. & Turnbull, S. M. (1997). "A Markov model for the term structure of credit risk spreads". *Rev. Financial Stud.*, 10:481–523.

Johansen, S. (1974). "Some results on the imbedding problem for finite Markov chains". *J. London Math. Soc.*, 8:345–351.

Kingman, J. F. C. (1962). "The imbeding problem for finite Markov chains". *Z. Wahrscheinlichkeitstheorie*, 1:14–24.

Kreinin, A. & Sidelnikova, M. (2001). "Regularization algorithms for transition matrices". *Algo Research Quarterly*, 4:33 – 40.

Küchler, U. & Sørensen, M. (1997). *Exponential Families of Stochastic Processes.* Springer, New York.

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

Oakes, D. (1999). "Direct calculation of the information matrix via the EM algorithm". *J. R. Statist. Soc. B*, 61:479 – 482.

Vanden-Eijnden, D. T. C. . E. (2005). "Fitting timeseries by continuous-time Markov chains: A quadratic programming approach". Preprint, Courant Institute of Mathematical Sciences.

Appendix

A EM estimates for different sampling frequencies

Transition		Sampling frequency in days						
from	to	1	7	15	30	90	180	365
Aaa	Aa	0.0733	0.0731	0.0739	0.0755	0.0788	0.0912	0.0848
Aa	A	0.1267	0.1267	0.1282	0.1298	0.1364	0.1343	0.1512
A	Aa	0.0119	0.0119	0.0120	0.0123	0.0124	0.0051	0.0022
A	Baa	0.0609	0.0610	0.0618	0.0623	0.0627	0.0614	0.0623
Baa	A	0.0547	0.0543	0.0550	0.0548	0.0571	0.0602	0.0591
Baa	Ba	0.0511	0.0507	0.0515	0.0521	0.0533	0.0543	0.0564
Ba	A	0.0082	0.0082	0.0083	0.0084	0.0089	0.0127	0.0090
Ba	Baa	0.1022	0.1038	0.1055	0.1089	0.1169	0.1346	0.1593
Ba	B	0.1186	0.1204	0.1225	0.1200	0.1203	0.1243	0.1320
Ba	D	0.0041	0.0020	0.0020	0.0041	0.0039	0.0039	0.0458
B	Baa	0.0094	0.0095	0.0096	0.0098	0.0107	0.0153	0.0206
B	Ba	0.0641	0.0654	0.0668	0.0697	0.0808	0.0922	0.1271
B	C	0.0849	0.0830	0.0790	0.0808	0.0773	0.0807	0.0853
B	D	0.0358	0.0399	0.0462	0.0472	0.0655	0.0695	0.1441
C	B	0.0589	0.0603	0.0619	0.0646	0.0771	0.0506	0.1069
C	D	0.5455	0.5267	0.4621	0.4646	0.3750	0.4081	0.5810

Table A.1: Estimates of transition intensities at different sampling frequencies. The sampling frequencies are as follows: 1=daily observations; 7=weekly observations; 15=half–monthly observations; 30=monthly observations; 90=quarterly observations; 180=biannual observations; 365=annual observations.

B Estimates for the period 1990-1995

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.08481	0.08481	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Aa	0.00000	-0.15217	0.15120	0.00097	0.00000	0.00000	0.00000	0.00000
A	0.00000	0.00215	-0.06898	0.06226	0.00387	0.00070	0.00000	0.00000
Baa	0.00000	0.00141	0.05906	-0.12475	0.05638	0.00128	0.00000	0.00663
Ba	0.00000	0.00000	0.00895	0.15929	-0.34630	0.13204	0.00019	0.04582
B	0.00000	0.00000	0.00000	0.02056	0.12711	-0.37714	0.08533	0.14414
C	0.00000	0.00000	0.00000	0.00000	0.04251	0.10698	-0.73046	0.58096
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table B.1: Maximum likelihood estimates of intensities for the period 1990-1995 as obtained by the EM–algorithm. Convergence occurred in less than the 250 iterations performed.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.12081	0.08604	0.00571	0.00587	0.00577	0.00581	0.00590	0.00571
Aa	0.00293	-0.16812	0.14846	0.00544	0.00295	0.00275	0.00283	0.00277
A	0.00105	0.00314	-0.07458	0.06186	0.00477	0.00172	0.00104	0.00100
Baa	0.00139	0.00265	0.05929	-0.13220	0.05611	0.00378	0.00178	0.00719
Ba	0.00398	0.00447	0.01285	0.15705	-0.36021	0.12534	0.00860	0.04792
B	0.00448	0.00460	0.00509	0.02357	0.12650	-0.39227	0.08078	0.14724
C	0.02479	0.02687	0.02726	0.03401	0.06526	0.11201	-0.73514	0.44494
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table B.2: MCMC estimates of intensities for the period 1990-1995.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.91869	0.07535	0.00579	0.00016	0.00001	0.00000	0.00000	0.00000
Aa	0.00000	0.85899	0.13553	0.00506	0.00034	0.00006	0.00000	0.00002
A	0.00000	0.00196	0.93521	0.05691	0.00468	0.00085	0.00003	0.00036
Baa	0.00000	0.00128	0.05403	0.88811	0.04501	0.00384	0.00011	0.00760
Ba	0.00000	0.00010	0.01130	0.12799	0.71672	0.09259	0.00361	0.04768
B	0.00000	0.00002	0.00111	0.02388	0.09062	0.69453	0.04953	0.14031
C	0.00000	0.00000	0.00021	0.00329	0.02935	0.06384	0.48423	0.41907
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table B.3: Annual transition probabilities based on the maximum likelihood estimate for the period 1990-1995.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.88082	0.07794	0.01192	0.00653	0.00558	0.00544	0.00383	0.00793
Aa	0.00275	0.84201	0.13471	0.00937	0.00308	0.00260	0.00181	0.00367
A	0.00102	0.00295	0.92920	0.05697	0.00550	0.00187	0.00079	0.00170
Baa	0.00140	0.00259	0.05473	0.87981	0.04515	0.00591	0.00159	0.00883
Ba	0.00369	0.00440	0.01534	0.12965	0.69778	0.08986	0.00874	0.05053
B	0.00470	0.00489	0.00680	0.02816	0.09239	0.67372	0.04719	0.14214
C	0.01956	0.02085	0.02460	0.03496	0.05384	0.08088	0.36258	0.40272
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table B.4: Annual transition probabilities based on the MCMC estimate for the period 1990-1995.

C Estimates for the period 1995-2000

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.05922	0.05922	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Aa	0.00364	-0.08838	0.08123	0.00000	0.00000	0.00000	0.00000	0.00000
A	0.00000	0.02335	-0.07880	0.05137	0.00407	0.00000	0.00000	0.00000
Baa	0.00102	0.00000	0.04632	-0.09380	0.04455	0.00103	0.00000	0.00088
Ba	0.00000	0.00000	0.00179	0.11820	-0.18089	0.05629	0.00000	0.00461
B	0.00000	0.00000	0.00315	0.00532	0.08396	-0.21929	0.05587	0.07098
C	0.00000	0.00000	0.00000	0.00974	0.00000	0.07557	-0.46511	0.37980
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table C.1: Maximum likelihood estimates of intensities for the period 1995-2000 as obtained by the EM–algorithm. Convergence occurred in less than the 250 iterations performed.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.11858	0.06367	0.00919	0.00916	0.00907	0.00912	0.00911	0.00926
Aa	0.00699	-0.11015	0.08236	0.00371	0.00372	0.00661	0.00349	0.00326
A	0.00104	0.02417	-0.08498	0.05175	0.00510	0.00099	0.00096	0.00097
Baa	0.00201	0.00111	0.04649	-0.09979	0.04487	0.00235	0.00119	0.00177
Ba	0.00212	0.00208	0.00482	0.11680	-0.19125	0.05649	0.00263	0.00630
B	0.00167	0.00171	0.00438	0.00702	0.08397	-0.22763	0.05734	0.07153
C	0.00831	0.00821	0.00936	0.01749	0.00995	0.07927	-0.49068	0.35810
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table C.2: MCMC estimates of intensities for the period 1995-2000.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.94260	0.05503	0.00223	0.00004	0.00001	0.00009	0.00000	0.00000
Aa	0.00339	0.91638	0.07478	0.00193	0.00030	0.00302	0.00008	0.00012
A	0.00006	0.02150	0.92620	0.04741	0.00461	0.00018	0.00000	0.00004
Baa	0.00094	0.00052	0.04260	0.91390	0.03904	0.00195	0.00004	0.00100
Ba	0.00005	0.00004	0.00409	0.10345	0.83875	0.04625	0.00118	0.00618
B	0.00000	0.00003	0.00297	0.00906	0.06901	0.80659	0.03984	0.07249
C	0.00000	0.00000	0.00028	0.00768	0.00256	0.05390	0.62951	0.30607
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table C.3: Annual transition probabilities based on the maximum likelihood estimate for the period 1995-2000.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.88009	0.06070	0.01196	0.00983	0.00922	0.00915	0.00761	0.01144
Aa	0.00656	0.89377	0.07702	0.00575	0.00379	0.00603	0.00287	0.00421
A	0.00108	0.02230	0.91985	0.04798	0.00550	0.00119	0.00083	0.00127
Baa	0.00187	0.00164	0.04309	0.90749	0.03947	0.00324	0.00103	0.00216
Ba	0.00203	0.00209	0.00685	0.10301	0.82761	0.04683	0.00319	0.00839
B	0.00173	0.00187	0.00452	0.01083	0.06945	0.79703	0.04056	0.07401
C	0.00658	0.00697	0.00860	0.01495	0.01080	0.05952	0.58950	0.30308
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table C.4: Annual transition probabilities based on the MCMC estimate for the period 1995-2000.

D Confidence and credibility intervals for the period 1990-1995

from	to	95 % confidence interval
Aaa	Aa	[0.0261,0.1436]
Aa	A	[0.1110,0.1914]
A	Baa	[0.0464,0.0781]
Baa	A	[0.0410,0.0771]
Baa	Ba	[0.0381,0.0747]
Ba	Baa	[0.1038,0.2147]
Ba	B	[0.0827,0.1813]
Ba	D	[0.0171,0.0745]
B	Baa	[-0.005,0.0462]]
B	Ba	[0.0726,0.1816]
B	C	[0.0446,0.1261]
B	D	[0.0872,0.2011]
C	Ba	[-0.035,0.1204]
C	B	[-0.021,0.2350]
C	D	[0.3035,0.8584]

Table D.1: 95 % confidence intervals for transition rates in the period 1990-1995 based on the estimated asymptotic variances as obtained by numerical differentiation via the EM–algorithm.

	Aaa	Aa	A	Baa
Aaa	[-0.18742,-0.07957]	[0.05048,0.14247]	[0.00015,0.02326]	[0.00016 0.02341]
Aa	[0.00007,0.01094]	[-0.21923,-0.13224]	[0.11370,0.19672]	[0.00023,0.01710]
A	[0.00003,0.00387]	[0.00051,0.00794]	[-0.09382,-0.05935]	[0.04766,0.07982]
Baa	[0.00003,0.00526]	[0.00017, 0.00796]	[0.04355, 0.08005]	[-0.16544,-0.10812]
Ba	[0.00010,0.01499]	[0.00012,0.01791]	[0.00086, 0.03362]	[0.11142,0.22654]
B	[0.00013,0.01700]	[0.00011,0.01701]	[0.00012, 0.01840]	[0.00298 0.05611]
C	[0.00077,0.12841]	[0.00086,0.13098]	[0.00085, 0.13298]	[0.00119,0.17505]
	Ba	B	C	D
Aaa	[0.00019,0.02286]	[0.00017,0.02335]	[0.00015,0.02255]	[0.00017,0.02230]
Aa	[0.00007,0.01142]	[0.00006,0.01045]	[0.00007, 0.01037]	[0.00008,0.01070]
A	[0.00063,0.01127]	[0.00008,0.00559]	[0.00003, 0.00407]	[0.00002,0.00370]
Baa	[0.03930,0.07793]	[0.00017,0.01143]	[0.00005, 0.00764]	[0.00128,0.01545]
Ba	[-0.46359,-0.29677]	[0.08467,0.18813]	[0.00029,0.034910]	[0.01698, 0.08655]
B	[0.08192 0.19511]	[-0.50624,-0.32616]	[0.04761, 0.16178]	[0.08825,0.21203]
C	[0.00598,0.28673]	[0.02539,0.40164]	[-1.54734,-0.65951]	[0.33573,1.04245]

Table D.2: 95 % credibility intervals of the posterior distribution of the transition rates for the period 1990-1995.

	Aaa	Aa	A	Baa
Aaa	[0.82946,0.92358]	[0.04445,0.12077]	[0.00502,0.02707]	[0.00103, 0.02166]
Aa	[0.00016,0.00959]	[0.80359,0.87652]	[0.10251, 0.17047]	[0.00435, 0.01938]
A	[0.00008,0.00360]	[0.00062, 0.00713]	[0.91295, 0.94387]	[0.04378,0.07170]
Baa	[0.00015, 0.00479]	[0.00043, 0.00715]	[0.04031, 0.07206]	[0.85451, 0.90212]
Ba	[0.00042, 0.01237]	[0.00068, 0.01436]	[0.00562, 0.03175]	[0.09109 0.17281]
B	[0.00065, 0.01427]	[0.00086, 0.01427]	[0.00209, 0.01794]	[0.01246,0.05154]
C	[0.00081, 0.07275]	[0.00171, 0.07411]	[0.00303, 0.07847]	[0.00617, 0.10478]
	Ba	**B**	**C**	**D**
Aaa	[0.00073,0.01852]	[0.00064, 0.01850]	[0.00031, 0.01336]	[0.00135, 0.02288]
Aa	[0.00065,0.00934]	[0.00035,0.00840]	[0.00018, 0.00601]	[0.00069, 0.01095]
A	[0.00223, 0.01054]	[0.00048, 0.00488]	[0.00011, 0.00259]	[0.00055,0.00437]
Baa	[0.03168, 0.06042]	[0.00269, 0.01167]	[0.00029, 0.00470]	[0.00350, 0.01636]
Ba	[0.64010, 0.75151]	[0.06001, 0.12511]	[0.00294, 0.02032]	[0.02713,0.08116]
B	[0.06042, 0.13102]	[0.61579 0.73020]	[0.02516, 0.07530]	[0.10139 0.18848]
C	[0.01016, 0.14216]	[0.01768, 0.18368]	[0.21807, 0.51997]	[0.24714,0.56403]

Table D.3: 95% credibility bounds for the one–year transition probabilities: 1990-1995

from	to	Oakes variance	MCMC variance
Aaa	Aa	8.98E-04	4.99E-04
Aa	A	4.20E-04	4.07E-04
A	Baa	6.51E-05	6.56E-05
Baa	A	8.47E-05	8.36E-05
Baa	Ba	8.71E-05	9.16E-05
Ba	Baa	8.00E-04	7.77E-04
Ba	B	6.33E-04	6.29E-04
Ba	D	2.15E-04	2.74E-04
B	Baa	1.72E-04	1.66E-04
B	Ba	7.74E-04	7.36E-04
B	C	4.32E-04	5.39E-04
B	D	8.43E-04	8.38E-04
C	Ba	1.58E-03	2.69E-03
C	B	4.27E-03	4.36E-03
C	D	2.00E-02	1.44E-02

Table D.4: Comparison of estimates of the asymptotic variance of the MLE obtained by Oakes' method and by the MCMC approach formula (2.11).

E Confidence and credibility intervals for the period 1995-2000

from	to	95 % confidence interval
Aa	A	[0.0492,0.1133]
A	Aa	[0.0140,0.0327]
A	Baa	[0.0373,0.0654]
Baa	A	[0.0326,0.0600]
Baa	Ba	[0.0307,0.0583]
Ba	Baa	[0.0855,0.1506]
Ba	B	[0.0347,0.0780]
B	Baa	[-0.004,0.0145]
B	Ba	[0.0595,0.1084]
B	C	[0.0366,0.0751]
B	D	[0.0483,0.0937]
C	Baa	[-0.007,0.0263]
C	B	[0.0293,0.1218]
C	D	[0.2667,0.4929]

Table E.1: 95 % confidence intervals for transition rates in the period 1995-2000 based on the estimated asymptotic variances as obtained by numerical differentiation via the EM–algorithm.

	Aaa	Aa	A	Baa
Aaa	[-0.20671,-0.06834]	[0.02818,0.12798]	[0.00027, 0.03802]	[0.00027,0.03828]
Aa	[0.00078 0.02109]	[-0.15675,-0.07623]	[0.05352,0.12330]	[0.00009,0.01462]
A	[0.00002,0.00395]	[0.01582,0.03534]	[-0.10460,-0.06893]	[0.03887,0.06786]
Baa	[0.00020,0.00567]	[0.00003,0.00406]	[0.03429, 0.06200]	[-0.12298,-0.08154]
Ba	[0.00005,0.00821]	[0.00005,0.00806]	[0.00022, 0.01418]	[0.08858,0.15400]
B	[0.00004,0.00624]	[0.00004,0.00635]	[0.00040, 0.01190]	[0.00059,0.01716]
C	[0.00024,0.03349]	[0.00022,0.03324]	[0.00026, 0.03766]	[0.00140,0.05647]
	Ba	B	C	D
Aaa	[0.00028,0.03707]	[0.00023,0.03634]	[0.00024,0.03663]	[0.00027,0.03660]
Aa	[0.00009,0.01384]	[0.00064,0.01976]	[0.00009, 0.01376]	[0.00008,0.01237]
A	[0.00095,0.01072]	[0.00003,0.00389]	[0.00003, 0.00371]	[0.00003,0.00357]
Baa	[0.03225,0.06123]	[0.00013,0.00703]	[0.00003 0.00432]	[0.00008 0.00550]
Ba	[-0.24024,-0.15552]	[0.03615,0.08450]	[0.00007,0.00932]	[0.00057,0.01707]
B	[0.06154,0.11205]	[-0.27427,-0.19362]	[0.03857, 0.08315]	[0.05072,0.09757]
C	[0.00027,0.03983]	[0.03758,0.15663]	[-0.69031,-0.39982]	[0.27607,0.52334]

Table E.2: 95 % credibility intervals of the posterior distribution of the transition rates for the period 1995-2000.

	Aaa	Aa	A	Baa
Aaa	[0.81345,0.93411]	[0.02561,0.11038]	[0.00259, 0.03637]	[0.00107,0.03472]
Aa	[0.00077,0.01856]	[0.85663,0.92737]	[0.04934, 0.10958]	[0.00203,0.01531]
A	[0.00013,0.00370]	[0.01449,0.03187]	[0.90324, 0.93505]	[0.03611,0.06179]
Baa	[0.00026,0.00510]	[0.00057,0.00428]	[0.03171, 0.05631]	[0.88861,0.92449]
Ba	[0.00018,0.00719]	[0.00028,0.00712]	[0.00268, 0.01496]	[0.07780,0.13109]
B	[0.00021,0.00548]	[0.00028,0.00579]	[0.00108, 0.01082]	[0.00535,0.01920]
C	[0.00030,0.02390]	[0.00054,0.02423]	[0.00095, 0.02825]	[0.00225,0.04181]

	Ba	B	C	D
Aaa	[0.00082,0.03171]	[0.00083,0.03125]	[0.00045, 0.02670]	[0.00142,0.03590]
Aa	[0.00061,0.01240]	[0.00084,0.01683]	[0.00025, 0.01018]	[0.00062,0.01282]
A	[0.00199,0.01042]	[0.00028,0.00368]	[0.00009, 0.00281]	[0.00024,0.00373]
Baa	[0.02827,0.05256]	[0.00118,0.00707]	[0.00014, 0.00329]	[0.00050,0.00572]
Ba	[0.79174,0.85934]	[0.02981,0.06767]	[0.00111, 0.00794]	[0.00299,0.01788]
B	[0.05094,0.09002]	[0.76487,0.82746]	[0.02721 0.05639]	[0.05577,0.09521]
C	[0.00272,0.03081]	[0.02680,0.10504]	[0.50397, 0.67183]	[0.22829,0.38540]

Table E.3: 95% credibility bounds for the one–year transition probabilities: 1995-2000

F Re-estimation after parameter reduction for the period 1990-1995

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.08633	0.08633	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Aa	0.00000	-0.15386	0.14858	0.00528	0.00000	0.00000	0.00000	0.00000
A	0.00000	0.00314	-0.07186	0.06233	0.00459	0.00179	0.00000	0.00000
Baa	0.00000	0.00280	0.05941	-0.12916	0.06181	0.00515	0.00000	0.00000
Ba	0.00000	0.00000	0.01323	0.15748	-0.36635	0.12222	0.00848	0.06495
B	0.00000	0.00000	0.00000	0.02461	0.12752	-0.38098	0.07980	0.14905
C	0.00000	0.00000	0.00000	0.00000	0.06342	0.10829	-0.60700	0.43529
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table F.1: MCMC estimates of the transition rates in the period 1990-1995 after fixing some rates at zero

	Aaa	Aa	A	Baa
Aaa	[-0.13515,-0.04862]	[0.04862, 0.13515]		
Aa		[-0.19620,-0.11686]	[0.11251, 0.18987]	[0.00022, 0.01599]
A		[0.00060, 0.00760]	[-0.09003,-0.05590]	[0.04780, 0.07969]
Baa		[0.00029, 0.00792]	[0.04194, 0.07842]	[-0.15838,-0.10249]
Ba			[0.00115, 0.03374]	[0.10739, 0.21699]
B				[0.00384, 0.05497]

	Ba	B	C	D
A	[0.00056, 0.01095]	[0.00009, 0.00544]		
Baa	[0.04331, 0.08318]	[0.00032, 0.01454]		
Ba	[-0.45212,-0.28973]	[0.07776, 0.17489]	[0.00027, 0.02839]	[0.03336, 0.10566]
B	[0.07952, 0.18902]	[-0.47277,-0.30169]	[0.04033, 0.13132]	[0.09534, 0.21010]
C	[0.00433, 0.19792]	[0.01970, 0.27858]	[-0.90821,-0.37279]	[0.23977, 0.69591]

Table F.2: 95 % credibility intervals of transition rates in the period 1990-1995 after fixing some rates at zero. Intervals for transitions with rate zero have been omitted.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.91751	0.07637	0.00577	0.00032	0.00001	0.00000	0.00000	0.00000
Aa	0.00000	0.85779	0.13283	0.00872	0.00050	0.00015	0.00000	0.00002
A	0.00000	0.00289	0.93265	0.05681	0.00541	0.00185	0.00008	0.00032
Baa	0.00000	0.00251	0.05434	0.88466	0.04892	0.00690	0.00040	0.00226
Ba	0.00000	0.00020	0.01461	0.12505	0.70322	0.08503	0.00835	0.06354
B	0.00000	0.00004	0.00142	0.02685	0.09026	0.69212	0.04917	0.14015
C	0.00000	0.00000	0.00042	0.00464	0.04320	0.06796	0.55298	0.33080
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table F.3: MCMC estimates of the one-year transition probabilities in the period 1990-1995 after fixing some transition rates at zero.

	Aaa	Aa	A	Baa
Aaa	[0.87358, 0.95254]	[0.04390, 0.11681]	[0.00307, 0.00964]	[0.00010, 0.00079]
Aa	[0.00000, 0.00000]	[0.82229, 0.88993]	[0.10266, 0.16644]	[0.00394, 0.01792]
A	[0.00000, 0.00000]	[0.00062, 0.00685]	[0.91616, 0.94712]	[0.04408, 0.07190]
Baa	[0.00000, 0.00000]	[0.00033, 0.00693]	[0.03894, 0.07108]	[0.86062, 0.90704]
Ba	[0.00000, 0.00000]	[0.00003, 0.00053]	[0.00487, 0.03084]	[0.08805, 0.16724]
B	[0.00000, 0.00000]	[0.00000, 0.00012]	[0.00057, 0.00269]	[0.01149, 0.04919]
C	[0.00000, 0.00000]	[0.00000, 0.00002]	[0.00005, 0.00142]	[0.00107, 0.01195]
D	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]

	Ba	B	C	D
Aaa	[0.00000, 0.00003]	[0.00000, 0.00001]	[0.00000, 0.00000]	[0.00000, 0.00000]
Aa	[0.00019, 0.00096]	[0.00003, 0.00037]	[0.00000, 0.00001]	[0.00001, 0.00004]
A	[0.00221, 0.01042]	[0.00043, 0.00479]	[0.00002, 0.00020]	[0.00011, 0.00066]
Baa	[0.03476, 0.06490]	[0.00300, 0.01390]	[0.00012, 0.00089]	[0.00130, 0.00355]
Ba	[0.64734, 0.75635]	[0.05545, 0.11831]	[0.00267, 0.02012]	[0.03821, 0.09572]
B	[0.05860, 0.12850]	[0.63258, 0.74609]	[0.02605, 0.07736]	[0.09918, 0.18511]
C	[0.00716, 0.12056]	[0.01490, 0.16016]	[0.40713, 0.69121]	[0.20410, 0.47672]
D	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]	[1.00000, 1.00000]

Table F.4: 95 % credibility intervals for the one–year transition probabilities in the period 1990-1995 after fixing some transition rates at zero.

G Re-estimation after parameter reduction for the period 1995-2000

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.06316	0.06316	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Aa	0.00700	-0.09284	0.08585	0.00000	0.00000	0.00000	0.00000	0.00000
A	0.00000	0.02411	-0.08183	0.05183	0.00590	0.00000	0.00000	0.00000
Baa	0.00201	0.00000	0.04651	-0.09795	0.04483	0.00274	0.00000	0.00186
Ba	0.00000	0.00000	0.00465	0.11737	-0.18675	0.05817	0.00000	0.00657
B	0.00000	0.00000	0.00470	0.00690	0.08429	-0.22448	0.05637	0.07222
C	0.00000	0.00000	0.00000	0.01788	0.00000	0.07846	-0.44921	0.35287
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table G.1: MCMC estimates of the transition rates in the period 1995-2000 after fixing some rates at zero.

	Aaa	Aa	A	Baa
Aaa	[-0.11609,-0.02619]	[0.02619, 0.11609]		
Aa	[0.00089, 0.01983]	[-0.13204,-0.06064]	[0.05472, 0.12451]	
A		[0.01534, 0.03461]	[-0.10023,-0.06519]	[0.03855, 0.06737]
Baa	[0.00024, 0.00571]		[0.03361, 0.06134]	[-0.11898,-0.07908]
Ba			[0.00020, 0.01414]	[0.08810, 0.15073]
B			[0.00065, 0.01177]	[0.00062, 0.01717]
C				[0.00177, 0.05054]

	Ba	B	C	D
A	[0.00168, 0.01240]			
Baa	[0.03196, 0.06055]	[0.00015, 0.00772]		[0.00012, 0.00556]
Ba	[-0.22815,-0.14891]	[0.03767, 0.08393]		[0.00057, 0.01769]
B	[0.06198, 0.11057]	[-0.26459,-0.18729]	[0.03764, 0.07898]	[0.05096, 0.09641]
C		[0.03421, 0.14004]	[-0.58518,-0.33170]	[0.25362, 0.47472]

Table G.2: 95 % credibility intervals of transition rates in the period 1995-2000 after fixing some rates at zero. Intervals for transitions with rate zero have been omitted.

	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0.93926	0.05820	0.00250	0.00004	0.00001	0.00000	0.00000	0.00000
Aa	0.00646	0.91264	0.07859	0.00204	0.00025	0.00001	0.00000	0.00000
A	0.00013	0.02210	0.92354	0.04773	0.00620	0.00023	0.00000	0.00007
Baa	0.00185	0.00057	0.04266	0.91018	0.03918	0.00345	0.00008	0.00203
Ba	0.00011	0.00007	0.00661	0.10223	0.83411	0.04756	0.00123	0.00809
B	0.00001	0.00005	0.00442	0.01057	0.06889	0.80273	0.04035	0.07297
C	0.00001	0.00000	0.00049	0.01397	0.00280	0.05602	0.64100	0.28570
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table G.3: MCMC estimates of the one-year transition probabilities in the period 1995-2000 after fixing some transition rates at zero.

	Aaa	Aa	A	Baa
Aaa	[0.89083, 0.97420]	[0.02466, 0.10436]	[0.00095, 0.00497]	[0.00002, 0.00009]
Aa	[0.00083, 0.01820]	[0.87782, 0.94203]	[0.05096, 0.11206]	[0.00119, 0.00323]
A	[0.00003, 0.00029]	[0.01411, 0.03155]	[0.90724, 0.93865]	[0.03582, 0.06144]
Baa	[0.00022, 0.00528]	[0.00033, 0.00088]	[0.03103, 0.05603]	[0.89173, 0.92660]
Ba	[0.00001, 0.00030]	[0.00002, 0.00018]	[0.00258, 0.01493]	[0.07791, 0.12898]
B	[0.00000, 0.00003]	[0.00001, 0.00014]	[0.00096, 0.01050]	[0.00514, 0.01929]
C	[0.00000, 0.00006]	[0.00000, 0.00001]	[0.00012, 0.00114]	[0.00172, 0.03855]
D	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]

	Ba	B	C	D
Aaa	[0.00000, 0.00001]	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]
Aa	[0.00009, 0.00053]	[0.00000, 0.00001]	[0.00000, 0.00000]	[0.00000, 0.00001]
A	[0.00248, 0.01182]	[0.00009, 0.00044]	[0.00000, 0.00001]	[0.00002, 0.00017]
Baa	[0.02826, 0.05251]	[0.00121, 0.00764]	[0.00002, 0.00020]	[0.00038, 0.00547]
Ba	[0.80109, 0.86495]	[0.03105, 0.06785]	[0.00068, 0.00197]	[0.00268, 0.01812]
B	[0.05128, 0.08920]	[0.77163, 0.83216]	[0.02746, 0.05541]	[0.05485, 0.09349]
C	[0.00131, 0.00501]	[0.02513, 0.09721]	[0.55900, 0.71885]	[0.21639, 0.36513]
D	[0.00000, 0.00000]	[0.00000, 0.00000]	[0.00000, 0.00000]	[1.00000, 1.00000]

Table G.4: 95 % credibility intervals for the one–year transition probabilities in the period 1990-1995 after fixing some transition rates at zero.