Bank of Japan Working Paper Series

Estimating Continuous Time Transition Matrices From Discretely Observed Data

Yasunari Inamura[*]
yasunari.inamura@boj.or.jp

Bank of Japan
2-1-1 Nihonbashi Hongoku-cho, Chuo-ku, Tokyo 103-8660

No.06 – E07-
April 2006

[*] Financial Systems and Bank Examination Department, Bank of Japan

Estimating Continuous Time Transition Matrices From Discretely Observed Data

Yasunari Inamura[1]

Financial Systems and Bank Examination Department, Bank of Japan

Version: April 2006

Abstract

A common problem in credit risk management is the estimation of probabilities of rare default events in high investment grades, when sufficient default data are not available. In addressing this issue, increasing attention has been paid to the use of continuous time Markov chains for modeling transition matrices. This approach incorporates the possibility of successive downgrades leading to defaulting in such a way that a very slight probability of default can be captured. In banking applications, however, the approach faces a problem with data limitations, since it requires continuously observed rating data to estimate intensities for transition matrices. In reality, the data frequency of internal rating systems for individual banks is either annual or bi-annual. To make the approach more applicable, the estimation methodology based on discretely observed rating data needs to be examined from a practical perspective. Against this background, the paper discusses and compares the small sample performances of the five estimation methods designed for discrete time observations – diagonal adjustment, weighted adjustment, quasi-optimization approach, expectation maximization algorithm and Markov chain Monte Carlo (MCMC) estimation – by measuring differences in default probabilities of investment grades and several matrix norms. Monte Carlo experiments reveal that the MCMC gives the most accurate finite-sample performance, both in terms of the estimated default probabilities and the matrix norms. Moreover, a case study to examine the impact on the loss distribution of a hypothetical investment grade portfolio shows that differences in these estimation methods have the potential to yield significantly different estimates of economic capital.

keywords: Default probability, LDPs, Markov chains, Infinitesimal generator matrix
JEL-codes: C13, G21

1. INTRODUCTION

Default probability is a long-established subject for theoretical and empirical studies in credit risk modeling. Numerous methods have been developed in the last

[1] Any views expressed represent those of the author only and not necessarily those of the Bank of Japan. I would like to thank Mogens Bladt, Til Schuermann and Toshiaki Watanabe for their helpful comments and advice. I am also grateful for comments received from staff members of the Institute for Monetary and Economic Studies, and the Financial Systems and Bank Examination Department at the Bank of Japan. Any remaining errors are solely upon the author.

decade to extract a default probability from historical observations. Most of the past studies, however, focus on corporate debts in low rating grades, where a minimal amount of the required historical observations is available for statistical inferences. The estimation of default probabilities for the higher rating grades, such as large corporations with a rating of Aaa or Aa, is much less explored because their default data are very scarce over short periods. The issue of data limitation becomes more acute for exposure types, such as specialized lending or sovereign debts, where very few defaults have been observed over time. The problem is even more apparent when individual banks attempt to utilize their own internal rating data. In these cases, a straightforward estimation based on the simple average often results in deriving zero default probabilities.

The Financial Sector has started to show serious concern regarding a similar data problem, the so-called issue of Low Default Portfolios (LDPs), because balance sheets of many financial institutions contain a significant proportion of sub-portfolios with few default records. The British Bankers' Association et al. (2005) delineate the issue of LDPs by identifying business types and circumstances where data limitations arise. They also provide a conceptual framework for the assessment of models that cover the LDPs. In response to this industry concern, the Basel Committee on Banking Supervision (2005) present the views of its sub-working group regarding the issue of LDPs in the internal ratings-based (IRB) approaches and provide some general suggestions for the treatment and data-enhancement of LDPs.

Recent studies on the modeling of continuous time rating transition matrices seem to provide a potential solution to these problems. The key idea here is to capture the possibility of successive downgrades of an obligor in the higher rating grades toward the lower rating grades where defaults occur more frequently. For example, if transitions from Aaa to A and transitions from A to default are observed within a period, then one can consider the possibility that an obligor with Aaa might default after successive downgrades, even without observing direct transitions from Aaa to default. In a discrete time context, financial practitioners know that the possibility of these successive defaults can be captured by multiplying transition matrices and picking out relevant elements in the cumulative transition probability matrix. Intuitively speaking, a similar thing happens in the continuous time approach. The approach incorporates the possibility of successive downgrades leading to defaulting in such a way that a very slight probability of defaults can be captured. Lando and Skødeberg (2002) provide two continuous time estimators for credit rating matrices, which differ in terms of their assumption regarding transition intensities. Specifically, one estimator, called the Aalen-Johansen estimator, allows transition intensities to be time inhomogeneous (i.e. time-varying) due to business cycles, while the other estimator assumes that transition intensities are time homogeneous (i.e. time-invariant) and thus estimates a so-called infinitesimal generator matrix in continuous time Markov chains. Their empirical studies show that both of the estimators generate non-zero values for default and migrating probabilities, which ordinary multinomial methods do not capture. Later, Christensen, Hansen and Lando (2004) propose a parametric bootstrapping-based method to derive confidence intervals for default probabilities under the framework of the continuous time approach. Their results demonstrate that the confidence intervals for default probabilities in investment grades are much tighter than multinomial-based confidence sets. Recently, Hansen and Schuermann (2005) propose a non-parametric bootstrap to relax the assumption of time homogeneity. Developing a new distance metric for transition matrices, Jafry and Schuermann (2004) find that the distance of transition matrices between the continuous time approach and the multinomial approach is significantly large. As for the studies explicitly relating to the issue of LDPs, Fuertes and Kalotychou (2005) present empirical studies on the small sample properties of sovereign credit migration data by applying the continuous time estimators and simulated confidence intervals. Their empirical results indicate non-zero default probabilities and reasonably narrow confidence intervals for sovereign credits.

One concern with the continuous time approach is that most of the previous literature is based on continuously observed transition data from rating agencies. Both the intra-year transition records and the length of time that obligors spend in the rating grades are required to apply the methods. In real applications, however, such a high-frequency database is still costly at the individual bank level. The data frequency of internal default data for individual banks is either annual or bi-annual in many cases. In order to make the continuous time approach more applicable in banking practice, financial practitioners will need an estimation methodology based on discretely observed rating data.

Several important works have emerged in recent years. Israel et al. (2001) propose the logarithmic expansion of an empirical transition probability matrix and the post-adjustment of its elements to obtain a valid generator matrix. The authors also provide some conditions to indicate whether the given transition data will allow the existence and uniqueness of a valid generator matrix from the logarithm of an empirical transition probability matrix. Following the post-adjustment approach, Kreinin and Sidelnikova (2001) propose a quasi-optimization methodology to adjust the matrix logarithm into a valid generator matrix, together with a fast computational algorithm to achieve it. The authors also present empirical studies comparing the fitting performance of several post-adjustment methods, in the context of calculating the root of the transition matrices. Recently, Bladt and Sørensen (2005) present the use of the (penalized) expectation-maximization algorithm (EM algorithm) or the Markov chain Monte Carlo (MCMC) estimation method to obtain a maximum likelihood estimator of an empirical generator matrix from discretely sampled data. The authors also present theoretical evidence for the existence and uniqueness of the maximum likelihood estimator for given empirical transition data.

Based on these recent advances, this paper discusses and compares the small sample performances of the five competing estimation methods: diagonal adjustment, weighted adjustment, quasi-optimization approach, EM algorithm and MCMC, by measuring differences in the estimates of default probabilities and several matrix norms. The results of Monte Carlo experiments show that the diagonal and weighted adjustment methods may generate a substantial deviation in both the estimated default probabilities and the mobility of a transition probability matrix. In contrast, the MCMC method gives the most accurate finite-sample performance. To illustrate the practical relevance of the Monte Carlo experiments, a case study examining the impact on the loss distribution of a hypothetical investment grade portfolio shows that differences in these estimation methods have the potential to yield significantly different estimates of economic capital. Finally, an empirical generator matrix based on the annual transition data of Japanese corporations is provided to show that the method examined in this paper gives reasonable non-zero estimates for investment grade default probabilities, which ordinary multinomial approaches do not provide.

The paper is organized as follows. Section 2 provides some preliminaries regarding Markov chains, relevant to the arguments in later sections. Section 3 discusses the five estimation methods designed to obtain an empirical generator matrix from credit rating data. Section 4 deals with the Monte Carlo experiments designed to compare the small sample performances of these estimation methods. The section also explores their impact on the economic capital of a hypothetical investment grade portfolio. An empirical study regarding a generator matrix for Japanese corporations during the 1990s is also provided. Section 5 concludes.

2. PRELIMINARIES

There is much extant literature and many texts on Markov chains and related issues mentioned in this section. The preparatory discussion here is brief and intuitive, so for further details on Markov chains, see Norris (1998).

2.1. Discrete Time Markov Chains

Let us start with a discrete time Markov chain (DTMC) in the context of credit risk modeling. In rating-based credit risk models, an individual transition between credit rating grades is modeled as a random process, characterized by a finite state space and transition probabilities. More formally:

- A set of K credit rating grades is denoted by a finite state space $S = \{s_1, ..., s_K\}$, which is usually indexed with integers in the order of credit quality such as $\{1, ..., K\}$. Normally, the last state K corresponds to "Default".

- Rating grade for an obligor at an arbitrary time $t_n, n = 1, \dots, T$ is denoted by $X(t_n)$, which has a countable number of possible values defined by S.

The time series behavior of $X = \{X(t_n)|n = 1, \dots, T\}$ is governed by its conditional probability distribution, which is a function of the past rating history. The Markov property is an assumption on the conditional probability distribution that allows the future rating to be independent of the past rating history. A finite state stochastic process with this property is called a Markov chain. In a discrete time setting, the Markov property is

$$\begin{aligned}&\mathbb{P}(X(t_{n+1})=j|X(t_0)=i_0,X(t_1)=i_1,\ldots,X(t_n)=i)\\=\;&\mathbb{P}(X(t_{n+1})=j|X(t_n)=i)=p_{ij}(t_n,t_{n+1}).\end{aligned}$$

Transition probabilities are normally assembled into the matrix form called a 'transition probability matrix'. Since the elements in each row are the mass of mutually exclusive conditional probabilities, each row of the transition probability matrix must add up to 1 for the conservation of probability. The transition probability matrix is convenient for describing the behavior of a Markov chain because multi-step transition probabilities are easily obtained. Consider the following matrix multiplication over m-periods

$$\mathbf{P}(t_n,t_{n+m})=\mathbf{P}(t_n,t_{n+1})\times\mathbf{P}(t_{n+1},t_{n+2})\cdots\times\mathbf{P}(t_{n+m-1},t_{n+m}) \qquad (2.1)$$

where the m-step transition probability from i to j is the ij th element of $\mathbf{P}(t_n,t_{n+m})$. Note that the m-step transition probability incorporates more than one path of the rating process with successive migrations, in addition to the direct transition from state i to state j. This is very suggestive when it comes to considering default probabilities. In the real world, when investing in highly-rated assets, the risk of direct defaults is slight. Rather, the major risk lies in the possibility of downgrading with a subsequent increase in the likelihood of defaulting.

For tractability, industry standards often make the assumption on transition probabilities that one-step transition probabilities remain constant over time. If the assumption holds, a Markov chain is said to be time homogeneous. An important consequence of the time-homogeneous assumption is that the m-step transition probability matrix is a function of the time distance m between the observations, not the calender time. For example, the m-step transition matrix given in (2.1) is calculated just by raising the one-period transition matrix to the power m as

$$\mathbf{P}(t_n,t_{n+m})=\bar{\mathbf{P}}^m$$

where each element of $\bar{\mathbf{P}}$ is a constant value p_{ij}.

A common practice in credit risk modeling is to ignore the possibility that an obligor recovers from the default state even if it is an unrealistic assumption. Therefore, once an obligor reaches the default state K, it is assumed to remain there for ever. The state K is said to be an absorbing state and the following specification ensures this assumption

$$p_{KK}=1 \; and \; p_{Kj}=0,\ \forall\, j\in S\ \backslash\{K\}.$$

With an absorbing default state, ordinary credit rating transition matrices satisfy the following,

$$\lim_{m\to\infty}\bar{\mathbf{P}}^m\to\mathbf{D}$$

$$\text{where} \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{2.2}$$

Thus, the default state is assumed to occur in the long run, regardless of the initial rating grades, and as such is accessible from every rating grades.

2.2. Continuous Time Markov Chains

As with the discrete time setting, the time series behavior in a continuous time Markov chain (CTMC) is defined by a stochastic process $X = \{X(t)|0 \leq t\}$, which satisfies the following for all $t \geq 0$, $s \geq 0$, and $i, j \in S$.

$$\begin{aligned} & \mathbb{P}(X(s+t) = j|X(s) = i, \{X(u) : 0 \leq u < s\}) \\ = \quad & \mathbb{P}(X(s+t) = j|X(s) = i). \end{aligned}$$

The assumption of time homogeneity can be understood with an analogy to a DTMC. If a transition probability satisfies the following

$$\mathbb{P}(X(s+t) = j|X(s) = i) = \mathbb{P}(X(t) = j|X(0) = i)$$

then, a CTMC is said to be time homogeneous.

Arguments in line with discrete time analogues face the problem of the 'time step' in continuous time. In discrete time, the time interval between the transitions is basically a unit, regardless of the frequency of data. In continuous time, however, there is no notion of 'time step' since the time index parameter t is continuous. In other words, one needs to consider the distribution of the holding time S_i, which is in our context the time that an obligor spends in rating grade i before migrating from it. In this regard, it is well known that the holding time follows an exponential distribution because of the Markov assumption (See, Norris (1998)). This means that for each rating grade i, there exists a positive constant rate q_i such that an obligor, when entering rating grade i, remains there for an amount of time which is a random draw $S_i \sim \exp(-q_i t)$, independent of its past rating history.

Since the holding time is exponentially distributed (which means that the number of jump events follows a Poisson distribution), the probability that one transition occurs during a short interval is given by

$$\mathbb{P}(X(t) \neq i|X(t - \Delta t) = i) = q_i \Delta t + o(\Delta t). \tag{2.3}$$

On the other hand, the probability of a transition from i to j for $(t - \Delta t, t]$ is $\mathbb{P}(X(t) = j|X(t - \Delta t) = i)$. From the Markov property, we have

$$\mathbb{P}(X(t) = j|X(t - \Delta t) = i) = \mathbb{P}(X(\Delta t) = j|X(0) = i).$$

Now, let us introduce q_{ij} defined by

$$q_{ij} = \lim_{\Delta t \to 0} \frac{\mathbb{P}(X(\Delta t) = j | X(0) = i)}{\Delta t} \tag{2.4}$$

assuming that the limit on the right-hand side exists in $[0, \infty)$. By definition, q_{ij} must be non-negative. To build a trajectory for the rating process of an obligor in a time-homogeneous CTMC, consider the conditional probability of migrating from i to j, given a jump from rating grade i, defined by

$$\mathbb{P}(X(t) = j | X(t - \Delta t) = i, X(t) \neq i) = \frac{\mathbb{P}(X(t) = j | X(t - \Delta t) = i)}{\mathbb{P}(X(t) \neq i | X(t - \Delta t) = i)}.$$

Using (2.3) and (2.4), we have

$$\mathbb{P}(X(t) = j | X(t - \Delta t) = i, X(t) \neq i) = \frac{q_{ij}\Delta t + o(\Delta t)}{q_i \Delta t + o(\Delta t)}.$$

Taking the limit $\Delta t \to 0$, we have

$$\mathbb{P}(X(t) = j | X(t-) = i, X(t) \neq i) = \frac{q_{ij}}{q_i}$$

which is the conditional probability that an obligor enters a new rating grade j, given a jump from i.

Because there are $K-1$ possible grades for the next rating grade, the conservation of probability requires

$$q_i = \textstyle\sum_{j=1, j\neq i}^{K} q_{ij}.$$

Hence, all we need to construct sample paths of the rating process is the parameters q_{ij}. We can summarize the rating process of an obligor in a time-homogeneous CTMC as follows:

- *The holding time of an obligor in rating grade i is exponentially distributed with a parameter q_i.*

- *Given a transition in rating grade i, the conditional probability of an obligor migrating to a new rating grade j is multinomially distributed with $\frac{q_{ij}}{q_i}$.*

Now we can introduce an infinitesimal generator matrix $\mathbf{Q}$, defined by

$$\mathbf{Q} = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1K} \\ q_{21} & -q_2 & \cdots & q_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{K1} & q_{K2} & \cdots & -q_K \end{bmatrix}$$

which satisfies the following properties:

- $\sum_{j=1}^{K} q_{ij} = 0$, for $1 \leq i \leq K$
- $0 \leq -q_{ii} = q_i \leq \infty$, for $1 \leq i \leq K$
- $q_{ij} \geq 0$ for $1 \leq i, j \leq K$ with $i \neq j$.

(2.5)

Given this, a transition probability matrix for the interval Δt can be expressed as

$$\mathbf{P}(t, t+\Delta t) = \mathbf{I} + \Delta t\mathbf{Q} + o(\Delta t)$$

where $\mathbf{I}$ is an identity matrix. An m-period (not m-step) transition probability matrix can be obtained in a similar manner. Let s denote $t + m\Delta t$. Then, we have

$$\begin{aligned} \mathbf{P}(t, s) &\approx (\mathbf{I} + \Delta t\mathbf{Q})^m \\ &= \left(\mathbf{I} + \frac{(s-t)}{m}\mathbf{Q}\right)^m. \end{aligned}$$

Taking the limit $m \to \infty$, we have

$$\mathbf{P}(t, s) = \exp((s-t)\mathbf{Q}) \tag{2.6}$$

$$\text{where} \quad \exp(h\mathbf{Q}) = \sum_{n=0}^{\infty} \frac{(h\mathbf{Q})^n}{n!}.$$

Thus, by calculating matrix exponentials of a generator matrix $\mathbf{Q}$, one can obtain a transition probability matrix for an arbitrary period.

Making the default state K absorbing in a CTMC is identical to the condition that $q_K = 0$ and $p_{Kj} = 0$, $\forall\, j \in S$, $j \neq K$. This absorbing assumption leads to the following result

$$\lim_{h \to \infty} \exp(h\mathbf{Q}) \to \mathbf{D} \tag{2.7}$$

where $\mathbf{D}$ is defined by (2.2).

2.3. Estimation of Markov Chains

For a DTMC with discrete observations: The conventional approach for estimating a transition probability matrix for a time-homogeneous DTMC with discrete observations for each obligor $\mathbf{x} = \{x(t_n) | n = 1, \ldots, T\}$ is the cohort method. With this method, the likelihood function is derived from K independent multinomial distributions. Hence, the likelihood is

$$L(\mathbf{P}) = \prod_{i=1}^{K}\prod_{j=1}^{K} p_{ij}^{N_{ij}(m)}$$

where $N_{ij}(m)$ is the number of obligors migrating from grade i to grade j during the period of m observations. The maximum likelihood estimator for a transition probability is given by

$$\hat{p}_{ij} = \frac{N_{ij}(m)}{N_i(m)} \tag{2.8}$$

where $N_i(m) = \sum_{j=1}^{K} N_{ij}(m)$.

For a CTMC with continuous observations: In estimating an empirical generator matrix, the maximum likelihood estimation is equally tractable if continuous observations (i.e. fully time-stamped observations) for each obligor $\mathbf{x} = \{x(t)|0 \leq t \leq T\}$ are available. Consider the likelihood of observations with a transition from i to j at time τ_1, followed by a subsequent transition from j to k at time τ_2, and etc., for each obligor. Assuming that an initial state probability is known, the likelihood can be expressed as

$$\begin{aligned} L(\mathbf{Q}) &= \exp(-q_i(\tau_2 - \tau_1))q_{ij}\exp(-q_j(\tau_3 - \tau_2))q_{jk}\cdots \\ &= \prod_{i=1}^{K}\prod_{i \neq j}(q_{ij})^{N_{ij}(T)}\exp(-q_i R_i(T)) \qquad (2.9) \end{aligned}$$

where $R_i(t) = \int_0^t 1_{\{x(s)=i\}} ds$, which is the total value of the holding time at rating grade i by the time t. $N_{ij}(t)$ is the number of times for ij transition by the time t.

The log-likelihood is

$$\log L(\mathbf{Q}) = \sum_{i=1}^{K}\sum_{j \neq i}\log(q_{ij})N_{ij}(T) - \sum_{i=1}^{K}\sum_{j \neq i} q_{ij}R_i(T). \tag{2.10}$$

Hence, the maximum likelihood estimator for the elements of an infinitesimal generator matrix is explicitly given as

$$\hat{q}_{ij} = \frac{N_{ij}(T)}{R_i(T)}. \tag{2.11}$$

For a DTMC with continuous observations, one can easily estimate a transition probability matrix by calculating the equation (2.6). Hence, our final problem, and our topic of interest, is whether one can estimate an empirical generator matrix from discrete time observations.

2.4. The Embeddability Problem

The equation (2.6) provides a connection between the generator matrix and the transition probability matrix. One may consider the empirical transition probability matrix $\tilde{\mathbf{P}}$ a sampling of a time-homogeneous CTMC at regular intervals Δt. Given the transition data, $\tilde{\mathbf{P}}$ can be obtained via the cohort method given in (2.8). Then, a natural idea to obtain an empirical generator matrix $\tilde{\mathbf{Q}}$ is to solve the equation

$$\tilde{\mathbf{P}} = \exp(\tilde{\mathbf{Q}}\Delta t). \tag{2.12}$$

Indeed, if $\tilde{\mathbf{P}}$ belongs to $\mathcal{P} = \{\exp(\mathbf{Q})|\mathbf{Q} \in Q\}$, where Q is the set of valid generator matrices as given in (2.5), then the solution to (2.12) is identical to the maximum likelihood estimator, whose likelihood is defined as

$$L(\mathbf{Q}) = \prod_{i=1}^{K} \prod_{j=1}^{K} p(\mathbf{Q})_{ij}^{N_{ij}(m)} \tag{2.13}$$

where $p(\mathbf{Q})_{ij}$ is the ij th element of a matrix exponential of $\mathbf{Q}$. In this case, $\tilde{\mathbf{P}}$ is called embeddable. In many cases, however, the solution $\tilde{\mathbf{Q}}$ is not necessarily a valid generator matrix because it may have complex or negative off-diagonal elements. This issue is called the 'embeddability problem'. Necessary and sufficient conditions where $\tilde{\mathbf{P}}$ is embeddable are already for the case where $K = 2$. However, for greater dimensions of K, only some partial conditions are known, which do not provide a complete characterization of uniquely embeddable $\tilde{\mathbf{P}}$. Unfortunately, it is known that the embeddability problem is nearly unavoidable in credit risk modeling. Israel et al. (2001) provide several necessary conditions for the non-existence of an exact valid generator. One of these conditions poses a serious challenge for obtaining a generator matrix by solving the equation (2.12). Specifically, if the following condition is satisfied with respect to $\tilde{\mathbf{P}}$, an exact generator matrix does not exist:

- *There are states i and j such that j is accessible from i, but $p_{ij} = 0$.*

This condition is likely to hold for the majority of empirical rating transition matrices. For example, high investment grades tend to exhibit zero default probability in the empirical transition probability matrix, even if the true probability is not zero. However, default state is accessible from the same high investment grades if successive downgrades are considered. Hence, the above condition is almost unavoidable and a simple matrix logarithm of an empirical transition matrix is very likely to contain negative off-diagonal elements. Following Israel et al. (2001), we provide a typical example of this condition, using the average annual transition matrix of corporations over the period 1981-2003, published by Standard & Poor's.

	AAA	AA	A	BBB	BB	B	CCC/C	D
AAA	92.07	7.09	0.63	0.15	0.06	0.00	0.00	0.00
AA	0.62	90.84	7.76	0.59	0.06	0.10	0.02	0.01
A	0.05	2.09	91.38	5.79	0.44	0.16	0.04	0.05
BBB	0.03	0.21	4.10	89.37	4.82	0.86	0.24	0.37
BB	0.03	0.08	0.40	5.53	83.25	8.15	1.11	1.45
B	0.00	0.08	0.27	0.34	5.39	82.41	4.92	6.59
CCC/C	0.10	0.00	0.29	0.58	1.55	10.54	52.80	34.14
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100

Table 1: Adjusted transition probability matrix $\tilde{\mathbf{P}}$ (%)
of S&P corporate data, 1981-2003[2]

Note that $\tilde{p}_{13} > 0$ and $\tilde{p}_{37} > 0$, but $\tilde{p}_{17} = 0$. Hence, $\tilde{\mathbf{P}}$ satisfies the condition provided above and therefore an exact valid generator matrix does not exist. If the solution $\tilde{\mathbf{Q}}$ is not a valid generator matrix, the resulting transition probability matrix will not preclude the possibility of containing negative or complex elements because $\mathbf{P} \approx \mathbf{I}+\Delta t\mathbf{Q}$ holds.

There are two possible explanations why the solution $\tilde{\mathbf{P}}$ does not belong to $\mathcal{P}$. The first is that the true transition probability matrix $\mathbf{P}$ is not embeddable either and hence is not generated by $\mathbf{Q} \in Q$. Admittedly, this possibility is difficult to dismiss as some previous empirical studies have confirmed time inhomogeneity and non-Markov properties in the historical movements of credit rating matrices. The assumption of a time homogeneous CTMC may be especially dubious over a long time horizon (say, greater than 20 years). A time-inhomogenous and non-Markov model would be preferable if possible. However, internal data at individual banks and data for LDP-related assets are limited in both time series and cross-sectional dimensions. Modeling time-inhomogenous and non-Markov movements seems difficult in situations where only limited discretely observed data are available. Indeed, the validity of time-homogeneous Markov assumption seems to depend on the asset type and on the time horizon of historical data. For example, empirical studies by Kiefer et al. (2004) show that time-homogeneous Markov assumption holds for the transitions of municipal bond ratings for up to five years, while the transitions of sovereign bonds are well described under the same assumption even in the long run.

Therefore, knowing that time-homogeneous Markov assumption is still open to question, the paper proceeds by considering the second explanation, where the true $\mathbf{P}$ can be considered embeddable, driven by some $\mathbf{Q} \in Q$, but the observed $\tilde{\mathbf{P}}$ is not embeddable due to the variability and finiteness of the observable rating migration data.

[2] To preclude the effect of rounding, the diagonal elements in $\tilde{\mathbf{P}}$ are adjusted such that each of the rows adds up to one.

There are several ways to cope with the embeddability problem. One way is to adjust the matrix logarithm of $\tilde{\mathbf{P}} \notin \mathcal{P}$ such that the adjusted $\tilde{\mathbf{Q}}$ satisfies the properties of (2.5). Specifically, the approach first sets the negative off-diagonals to zero and then adds the extra value to the other elements to compensate. There are a variety of numerical procedures, from ad hoc adjustment to optimization-based adjustment, as shown in Israeli et al. (2001) and in Kreinin and Sidelnikova (2002). Another way to find the empirical generator matrix is to consider the maximization of (2.13) directly over the space Q, instead of solving (2.12). Kalbfleisch and Lawless (1985) argue that the maximization of (2.13) is preferable even when $\tilde{\mathbf{P}}$ is embeddable because solving (2.12) does not provide the standard errors of the parameters. However, even if one resorts to the maximization of (2.13), there is a possibility that the maximum likelihood estimator does not exist. In this respect, Bladt and Sørensen (2005) provide theoretical evidence for the existence and non-existence of the maximum likelihood estimator. According to their results, the maximum likelihood estimator does not exist on Q, when $\det(\tilde{\mathbf{P}}) = 0$. To clarify what this condition means, assume that $\tilde{\mathbf{P}}$ is a Δ-year transition probability matrix. If one remembers that $\tilde{\mathbf{P}} \approx \exp(\Delta\mathbf{Q})$, this condition implies that the determinant of $\exp(\Delta\mathbf{Q})$ is likely to be near zero. As the authors explain, if the underlying Markov process is ergodic, $\exp(\Delta\mathbf{Q})$ converges to the singular matrix as $\Delta \to \infty$ (as given in (2.7)). Hence, the non-existence of a generator is more likely to occur as the observation interval increases. Also, the non-existence case becomes more likely as the speed of migration, given the observational interval Δ, increases (hence, the relative size of Δ increases). This increase in the relative size of Δ can be interpreted as a situation where the observations become more discrete and partial (hence, the data become less informative). Although a due care should be taken, the majority of annual (or bi-annual) ordinary credit rating data seems to have empirical properties in which $\det(\tilde{\mathbf{P}}) \neq 0$ and its eigenvalues are all distinct. Kreinin and Sidelnikova (2001) indicate these empirical properties as the common characteristics of most credit rating matrices. According to the result by Bladt and Sørensen (2005), if $\tilde{\mathbf{P}} \notin \mathcal{P}$ but $\det(\tilde{\mathbf{P}}) \neq 0$ and all the eigenvalues are distinct, the maximum likelihood estimator, given the empirical credit rating transition data, is likely to be on the boundary of Q, where some infinitesimal parameters are zero. Finally, one can use a Bayesian inference to obtain a valid empirical generator by specifying a suitable prior distribution to ensure the non-negativity of q_{ij}. The key issue in the Bayesian inference is how to obtain the posterior distribution for given observations. A recent trend is to employ MCMC methods for the calculation of integrals with respect to the posterior distribution, as shown by Bladt and Sørensen (2005) in the case of a time-homogeneous Markov chain.

The next section provides more detail of methods for estimating a valid empirical generator matrix against the embeddability problem.

3. METHODOLOGIES FOR DISCRETE TIME OBSERVATIONS

This section details the five competing methods for obtaining a valid generator matrix given empirical credit rating data, namely, diagonal adjustment, weighted adjustment, quasi-optimization approach, expectation-maximization algorithm and Markov chain Monte Carlo estimation.

3.1. Diagonal Adjustment and Weighted Adjustment

Israel et al. (2001) suggest the use of a matrix logarithm to solve the equation (2.12) as

$$\tilde{\mathbf{Q}} = \frac{1}{(s-t)} \log \tilde{\mathbf{P}}(t,s). \tag{3.1}$$

The logarithm of a matrix is defined by its power series as

$$\log(\tilde{\mathbf{P}}) = (\tilde{\mathbf{P}} - \mathbf{I}) - \frac{(\tilde{\mathbf{P}} - \mathbf{I})^2}{2} + \frac{(\tilde{\mathbf{P}} - \mathbf{I})^3}{3} - \frac{(\tilde{\mathbf{P}} - \mathbf{I})^4}{4} + \cdots$$

where $\mathbf{I}$ is an identity matrix. Israel et al. (2001) also present some necessary conditions for the existence of a real matrix logarithm, given $\tilde{\mathbf{P}}$. However, the equation (3.1) generates $\tilde{\mathbf{Q}}$ with negative off-diagonal elements, due to the embeddability problem. To cope with this, the authors provide two adjustment methods for $\log(\tilde{\mathbf{P}})$, based on simple numerical procedures, called diagonal adjustment (DA) and weighted adjustment (WA).

With the DA method, all negative off-diagonal elements in each row are replaced with zero and the diagonal elements are re-calculated as the negative sum of the non-diagonal elements, in order to satisfy the properties of a generator matrix given in (2.5). In contrast, with the WA method, any negative off-diagonals are, as with the DA, set to zero, but all the other non-zero elements, together with the diagonal elements, are modified to ensure that the matrix has rows that sum to zero.

Let $\tilde{q}_{ij}^{DA}$ and $\tilde{q}_{ij}^{WA}$ denote the elements of an empirical generator matrix obtained by the DA and the WA method, respectively. The computational procedures for the DA and the WA can be summarized as follows:

1. let $\tilde{q}_{ij}$ denote an adjusted elements of the matrix logarithm of $\tilde{\mathbf{P}}$ and q_{ij} denote the elements of the matrix logarithm $\tilde{\mathbf{P}}$ before the adjustment. Obtain $\tilde{q}_{ij}$ as follows,

$$\tilde{q}_{ij} = \begin{cases} 0 & \textit{if } (i \neq j) \textit{ and } q_{ij} < 0 \\ q_{ij} & \textit{otherwise} \end{cases} .$$

2. for the DA, set the diagonal elements to the negative sum of the non-diagonal elements as below,

$$\tilde{q}_{ii}^{DA} = -\sum_{j=1, j\neq i}^{K} \tilde{q}_{ij} \qquad \text{for } i = 1, 2, \ldots, K$$

for the WA, adjust non-zero elements according to their relative magnitudes as below,

$$\tilde{q}_{ij}^{WA} = \tilde{q}_{ij} - |\tilde{q}_{ij}| \frac{\sum_{j=1}^{K} \tilde{q}_{ij}}{\sum_{j=1}^{K} |\tilde{q}_{ij}|} \qquad \text{for } i, j = 1, 2, \ldots, K.$$

3.2. Quasi-optimization of the Generator

The DA and the WA method are very parsimonious approaches as they both do not involve complex calculations. However, these adjustments are not based on any norm or any optimality. In this respect, Kreinin and Sidelnikova (2001) have extended the post-adjustment method by incorporating a distance-minimizing optimization, called quasi-optimization of the generator (QOG). In the QOG, approximating the generator matrix is reformulated as the problem of minimizing the sum of squared deviations between $\log(\tilde{\mathbf{P}})$ and $\mathbf{Q} \in Q$. Thus, the problem setting is defined by

$$\min_{\mathbf{Q} \in Q} \left\| \mathbf{Q} - \log(\tilde{\mathbf{P}}) \right\| \tag{3.2}$$

for finding the optimal $\mathbf{Q}^* \in Q$. Note that the above problem can be solved on a row by row basis because the conditions of a generator matrix (2.5) are closed on each row. Hence, (3.2) can be reduced to K independent problems of minimizing the Euclidean distance between a row of $\log(\tilde{\mathbf{P}})$ and a row of $\mathbf{Q} \in Q$. Let $\mathbf{z} = (z_1, z_2, \ldots, z_K)$ denote a row of $\mathbf{Q} \in Q$, which is permutated such that z_1 is a diagonal element in the row. Note that the permutation does not affect the solution in this problem setting. The feasible set of $\mathbf{z}$ can be expressed as a standard cone $\mathcal{C}(K)$ as

$$\mathcal{C}(K) = \left\{ \mathbf{z} \in R^K \mid \sum_{i=1}^{K} z_i = 0, z_1 \leq 0, z_i \geq 0 \text{ for } i \geq 2 \right\}.$$

A row of $\log(\tilde{\mathbf{P}})$ is considered a point $\mathbf{a} =(a_1, a_2, \ldots, a_K) \in R^K$. Then, the problem (3.2) is reduced to the nonlinear programming problem for finding the optimal $\mathbf{z}^* \in \mathcal{C}(K)$ as defined by

$$\min_{\mathbf{z} \in \mathcal{C}(K)} \sum_{i=1}^{K} (a_i - z_i)^2 . \tag{3.3}$$

Kreinin and Sidelnikova (2001) also provide a fast computational algorithm for solving (3.3). The trick of their algorithm is to transform the multivariate nonlinear constrained problem (3.3) into a univariate problem. The computational algorithm is as follows:

1. Construct $\mathbf{b} = (b_1, b_2, \ldots, b_K \mid b_i = a_i + \lambda)$ for $i = 1, 2, \ldots K$, where $\lambda = -\frac{1}{K} \sum_{i=1}^{K} a_i$.
2. Compute the vector $\hat{\mathbf{a}} = \pi(\mathbf{b})$, where π is a permutation that orders $\mathbf{b}$ such that $b_i \leq b_{i+1}$.

3. Find the smallest number m^*, for $2 \leq m \leq K - 1$, which satisfies $(K - m + 1)\hat{a}_{m+1} - \left(\hat{a}_1 + \sum_{i=0}^{K-m-1} \hat{a}_{K-i}\right) \geq 0$.
4. Construct the vector $\mathbf{z}^* \in \mathcal{C}(K)$, whose element is defined by

$$z_i = \begin{cases} 0 & if\ 2 \leq i \leq m^* \\ \hat{a}_i - \frac{1}{K-n^*+1}\left(\hat{a}_1 + \sum_{j=m^*+1}^{K} \hat{a}_j\right) & otherwise \end{cases} .$$

5. Compute the inverse permutation $\pi^{-1}(\mathbf{z}^*)$, which is the solution of QOG.

Appendix A provides an exposition, based on Tuenter (2000), on the derivation of the QOG algorithm.

The great advantage of the post-adjustment methods, including the DA and the WA, is that an approximated generator matrix can be obtained even with a single transition probability matrix, as long as the real matrix logarithm exists. One may find it especially helpful when only an average annual rating matrix is available, which is often the case for assets belonging to LDPs.

3.3. Expectation Maximization Algorithm

Another way to cope with the embeddability problem is to apply a maximum likelihood estimation directly to (2.13) as is shown by Kalbfleisch and Lawless (1985). However, when the estimation involves incomplete data either in the form of missing data or latent variables, an ordinary maximum likelihood estimation may be difficult to apply. The estimation of a generator matrix from empirical rating matrices is a typical example for this incomplete-data problem. In these cases, an iterative scheme called an expectation maximization (EM algorithm) is often used to obtain the maximum likelihood estimator, as shown in Asmussen et al. (1996) and in Bladt and Sørensen (2005). The basic idea of the EM algorithm is simple: replace missing values with estimated values and then estimate parameters. The algorithm is characterized by iterations of the Expectation-step (E-step) and the Maximization-step (M-step). The E-step is to construct a complete data by replacing unobserved parts with their respective expected values conditional on the observed data, assuming some initial set of parameters. The M-step is to implement the maximum likelihood estimation using the constructed complete data. New estimates obtained in the M-step are then used as the parameters in the next E-step. These two steps are iterated until convergence of the likelihood function is achieved. For details of the EM algorithm, see McLachan & Krishnan (1997).

Let us review the procedures for the EM-algorithm to determine an empirical generator matrix for rating grades. Assume that we have a set of discretely observed migrating data $\mathbf{x}^{obs}$ for a portfolio of N obligors with T observations. The observational interval Δt is assumed to be equidistant (say, annual). The set of the data $\mathbf{x}^{obs}$ contains a discrete time observation for each obligor h, denoted by $\mathbf{x}^h = \left\{x^h(t_n) \in S | n = 1, \ldots, T\right\}$, for $1 \leq h \leq N$. Let $\mathbb{E}\left[N_{ij}(T)|\mathbf{x}^{obs}\right]$ and $\mathbb{E}\left[R_i(T)|\mathbf{x}^{obs}\right]$ denote the conditional expectation of the number of transitions from rating grade i to rating grade j and the conditional expectation of the total amount of time in rating grade i, given a discrete time observation $\mathbf{x}^{obs}$, respectively. Then, we can rewrite the log likelihood (2.10) as

$$\begin{aligned}
&\mathbb{E}\left[\log L(\mathbf{Q})|\mathbf{x}^{obs}\right] \\
= \quad &\sum_{i=1}^{K}\sum_{j\neq i}\log(q_{ij})\mathbb{E}\left[N_{ij}(T)|\mathbf{x}^{obs}\right] - \sum_{i=1}^{K}\sum_{j\neq i} q_{ij}\mathbb{E}\left[R_i(T)|\mathbf{x}^{obs}\right].
\end{aligned}$$

Then, in the M-step, the maximum likelihood estimator for q_{ij} can be obtained explicitly as

$$\tilde{q}_{ij} = \frac{\mathbb{E}\left[N_{ij}(T)|\mathbf{x}^{obs}\right]}{\mathbb{E}\left[R_j(T)|\mathbf{x}^{obs}\right]}. \tag{3.4}$$

Note that the conditional expectations in (3.4) can be expressed as

$$\mathbb{E}\left[N_{ij}(T)|\mathbf{x}^{obs}\right] = \sum_{h}^{N}\mathbb{E}\left[N_{ij}^{h}(T)|\mathbf{x}^{h}\right], \quad \mathbb{E}\left[R_i(T)|\mathbf{x}^{obs}\right] = \sum_{h}^{N}\mathbb{E}\left[R_i^{h}(T)|\mathbf{x}^{h}\right] \tag{3.5}$$

where $\mathbb{E}\left[N_{ij}^{h}(T)|\mathbf{x}^{h}\right]$ and $\mathbb{E}\left[R_i^{h}(T)|\mathbf{x}^{h}\right]$ are the conditional expectation of the number of transitions from rating grade i to rating grade j and the conditional expectation of the total amount of time in rating grade i, given a discrete time observation $\mathbf{x}^{h}$, respectively. Hence, the intractable part of the algorithm is the computation of $\mathbb{E}\left[N_{ij}^{h}(T)|\mathbf{x}^{h}\right]$and $\mathbb{E}\left[R_i^{h}(T)|\mathbf{x}^{h}\right]$ in (3.5) in the E-step. Let $x^h(t_{n+1})$ and $x^h(t_n)$ denote the rating grade observed at t_{n+1} and t_n for each obligor h, respectively. The Markov property allows us to have

$$\mathbb{E}\left[R_i^{h}(T)|\mathbf{x}^{h}\right] = \sum_{n=1}^{T-1}\mathbb{E}\left[R_i^{h}(\Delta t)|X^h(t_{n+1}) = x^h(t_{n+1}), X^h(t_n) = x^h(t_n)\right]$$

and

$$\mathbb{E}\left[N_{ij}^{h}(T)|\mathbf{x}^{h}\right] = \sum_{n=1}^{T-1}\mathbb{E}\left[N_{ij}^{h}(\Delta t)|X^h(t_{n+1}) = x^h(t_{n+1}), X^h(t_n) = x^h(t_n)\right].$$

Thus, the computation of $\mathbb{E}\left[R_i^{h}(T)|\mathbf{x}^{h}\right]$ and $\mathbb{E}\left[N_{ij}^{h}(T)|\mathbf{x}^{h}\right]$ is reduced to the calculation of the conditional expectation over the interval between t_n and t_{n+1}. Let $\mathbf{e}_i$ be a K vector of zeros with one in position i. Then, for each obligor and for each interval, we have

$$\begin{aligned}
&\mathbb{E}\left[R_i^{h}(\Delta t)|X^h(t_{n+1}) = x^h(t_{n+1}), X^h(t_n) = x^h(t_n)\right] \\
= \quad &\frac{1}{D} \times \mathbf{e}_{x^h(t_n)}^{T}\left(\int_{t_n}^{t_{n+1}} \exp((s-t_n)\mathbf{Q})(\mathbf{e}_i\mathbf{e}_i^{T})\exp((t_{n+1}-s)\mathbf{Q})ds\right)\mathbf{e}_{x^h(t_{n+1})}
\end{aligned} \tag{3.6}$$

and

$$
\begin{aligned}
&\mathbb{E}\left[N_{ij}^h(\Delta t)|X^h(t_{n+1}) = x^h(t_{n+1}), X^h(t_n) = x^h(t_n)\right] \\
=\ &\frac{1}{D} \times \mathbf{e}^T_{x^h(t_n)} \left(q_{ij} \int_{t_n}^{t_{n+1}} \exp((s - t_n)\mathbf{Q})(\mathbf{e}_i \mathbf{e}_j^T) \exp((t_{n+1} - s)\mathbf{Q}) ds \right) \mathbf{e}_{x^h(t_{n+1})}
\end{aligned}
\tag{3.7}
$$

where

$$D = \mathbf{e}^T_{x^h(t_n)} \exp((t_n - t_{n+1})\mathbf{Q}) \mathbf{e}_{x^h(t_{n+1})}.$$

See Appendix B for a brief exposition, based on the result by Asmussen et al. (1996), on the derivation of (3.6) and (3.7). The key numerical procedure here is the calculation of the integrals of matrix exponentials in the expectations. There are several options available for the computation. Asmussen et al. (1996) use the Runge-Kutta algorithm to solve a system of matrix-valued differential equations for (3.6) and (3.7), while Bladt and Sørensen (2005) employ the uniformization approach for computing the integrals. In control theory, the calculation of the augmented matrix exponentials, proposed by van Loan (1978), is usually selected for these computational problems. We applied the uniformization method and the approach provided by van Loan to several simulated samples and confirmed that both of the methods generate the same default probabilities at the level of the displayed figures in this paper. The next empirical section provides empirical results based on van Loan's approach. Appendix C also provides details of the computation of integrals of matrix exponentials.

The procedures of the EM algorithm are summarized as follows:

1. Let $\mathbf{Q}_0$ be the a matrix with the initial generator matrix. Set $\mathbf{Q}_k = \mathbf{Q}_0$ initially.
2. Calculate (3.6) and (3.7) for all the obligors in the portfolio over each interval, up to T.
3. Calculate $\tilde{q}_{ij} = \frac{\mathbb{E}\left[N_{ij}(T)|\mathbf{x}^{obs}\right]}{\mathbb{E}[R_i(T)|\mathbf{x}^{obs}]}$ to obtain a new $\tilde{\mathbf{Q}}$. Then, set $\mathbf{Q}_k = \tilde{\mathbf{Q}}$ and go to 2.
4. Iterate 2.$\sim$ 3. until the convergence of the likelihood function is achieved.

3.4. Markov Chain Monte Carlo Estimation

Markov Chain Monte Carlo (MCMC) estimation approximates the posterior distribution for parameters Θ or latent variables Y, given observations X, through samples obtained by generating a sequence of Markov chain $\left\{\Theta^{(g)}, Y^{(g)}\right\}_{g=1}^{G}$ from the posterior distribution $p(\Theta, Y|X)$. Without any optimization, one can estimate variables of interests by summarizing the statistics of these simulated samples. For example, the posterior mean estimate of $f(\Theta, Y)$ defined by

$$\mathbb{E}\left[f(\Theta, Y) \mid X\right] = \int f(\Theta, Y) p(\Theta, Y|X) d\Theta dY$$

is obtained as

$$\frac{1}{G} \sum_{g=1}^{G} f(\Theta^{(g)}, Y^{(g)}).$$

Note that using the Bayes rule we can factorize the posterior distribution into the components as

$$p(\Theta, Y|X) \propto p(X|\Theta, Y)p(Y|\Theta)p(\Theta)$$

where $p(\Theta)$ is the prior distribution of parameters. Here, we may find one of the advantages of the MCMC in the presence of a prior distribution. The prior distribution allows us to impose economic or statistical constraints on the inferences of the parameters. This means that in the context of estimating a generator matrix, positivity conditions can be easily imposed on q_{ij} by choosing the appropriate $p(\mathbf{Q})$.

Hence, for the estimation, the key issues are the choice of a prior distribution $p(\mathbf{Q})$ and the method used to generate the sequence $\left\{\mathbf{Q}^{(g)}, X^{(g)}\right\}_{g=1}^{G}$, from the high-dimensional joint posterior distribution conditional on partial observations $X = x$. Bladt and Sørensen (2005) propose Gibbs sampling for generating the sequence. Specifically, given an initial $\mathbf{Q}^{(0)}$, the Gibbs sampler proceeds as:

1. Draw $X^{(1)} \sim p(X|\mathbf{Q}^{(0)}, x)$
2. Draw $\mathbf{Q}^{(1)} \sim p(\mathbf{Q}|X^{(1)}, x)$

   $\cdots\cdots$

The iteration of this sampling process generates a sequence $\left\{\mathbf{Q}^{(g)}, X^{(g)}\right\}_{g=1}^{G}$, which converges to $p(\mathbf{Q}, X|x)$. For the choice of $p(\mathbf{Q})$, Bladt and Sørensen (2005) propose a gamma distribution given by

$$p(\mathbf{Q}) \propto \prod_{i=1}^{K} \prod_{j \neq i} q_{ij}^{\alpha_{ij}-1} e^{-q_{ij}\beta_i} \tag{3.8}$$

where $\alpha_{ij} > 0$, $i, j \in S$ and $\beta_i > 0, i \in S$ are the constant values provided exogenously. Thus, we have $q_{ij} \sim \Gamma(1/\beta_i, \alpha_{ij})$, where the mean and variance are given by $\frac{\alpha_{ij}}{\beta_i}$ and $\frac{\alpha_{ij}}{\beta_i^2}$ respectively. Since the likelihood function for the complete data is given as (2.9), the posterior distribution of $\mathbf{Q}$ is written as

$$\begin{aligned} p(\mathbf{Q}|X, x) &= p(\mathbf{Q}|X) \\ &\propto p(\mathbf{Q})p(X|\mathbf{Q}) \\ &= \prod_{i=1}^{K} \prod_{j \neq i} q_{ij}^{N_{ij}(T)+\alpha_{ij}-1} e^{-q_{ij}(R_i(T)+\beta_i)}. \end{aligned}$$

Note that the posterior distribution of $\mathbf{Q}$ also follows a gamma distribution, which makes the drawing $\mathbf{Q}^{(i)} \sim p(\mathbf{Q}|X^{(i)}, x)$ tractable. As for the drawing of the Markov process $X^{(i+1)} \sim p(X|\mathbf{Q}^{(i)}, x)$, Bladt and Sørensen (2005) suggest that a simple rejection sampling can be applied as follows: First, obtain the sample of holding time S_k by drawing from $f_k = q_k \exp(-q_k \Delta t)$, where $\Delta t (= t_n - t_{n-1})$ is the equidistant observation interval. If $t_{n-1} + S_k < t_n$, then let the process make a transition to another rating grade with the probability of q_{kj}/q_k. Continue this procedure until the process reaches an observed rating grade by time $t_{n-1} + \Delta t$ (otherwise the sample is rejected). If the sample is accepted, obtain the records for the holding time and the number of transitions by time $t_{n-1} + \Delta t$. Continue this procedure until all the other observed transitions from time t_{n-1} to t_n are realized. Then repeat the same simulation from time t_n to t_{n+1}. Thus, thanks to the Markov property, we can implement the simulation on an interval-by-interval basis[3].

Following Bladt and Sørensen (2005), we summarize the procedures of the MCMC approach as follows:

1. Construct the initial $\mathbf{Q}$ by drawing q_{ij} from $\Gamma(1/\beta_i, \alpha_{ij})$ for $j \neq i$.
2. Simulate a continuous time Markov chain $X(t)$ with the generator matrix $\mathbf{Q}$ such that all the observations over each interval Δt are realized. Repeat this simulation up to time T.
3. Calculate the statistics $\tilde{N}_{ij}(T)$ and $\tilde{R}_i(T)$ from the accepted samples in 2.
4. Construct a new $\mathbf{Q}$ by drawing q_{ij} from $\Gamma(1/(\tilde{R}_i(T) + \beta_i), \tilde{N}_{ij}(T) + \alpha_{ij})$.
5. Iterate 2. $\sim$ 4. up to G times. Then, summarize the statistics of interests from $\left\{q_{ij}^{(g)}\right\}_{g=1}^{G}$.

4. NUMERICAL STUDIES

This section explores statistical differences of the five competing methods through Monte Carlo experiments and their practical impact on real banking applications. An empirical generator matrix for Japanese corporations is also provided. The following notation is used throughout the section:

- a prespecified true generator matrix: $\mathbf{Q}$
- an estimate of $\mathbf{Q}$ by the DA method: $\tilde{\mathbf{Q}}_{DA}$
- an estimate of $\mathbf{Q}$ by the WA method: $\tilde{\mathbf{Q}}_{WA}$
- an estimate of $\mathbf{Q}$ by the QOG method: $\tilde{\mathbf{Q}}_{QOG}$
- an estimate of $\mathbf{Q}$ by the EM algorithm: $\tilde{\mathbf{Q}}_{EM}$
- an estimate of $\mathbf{Q}$ by the MCMC: $\tilde{\mathbf{Q}}_{MC}$

Also, let $\mathbf{P}(\mathbf{Q})$ denote the matrix exponential of $\mathbf{Q}$ (hence, $\mathbf{P}(\mathbf{Q}) = \exp(\mathbf{Q})$).

[3] The author thanks Mogens Bladt for recognizing this point.

4.1. Monte Carlo Experiments

4.1.1. Procedures

We assume that a true generator matrix $\mathbf{Q}$ is given to us as follows.

	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	-0.071371	0.065881	0.005490	0.000000	0.000000	0.000000	0.000000	0.000000
Aa	0.008506	-0.123337	0.114831	0.000000	0.000000	0.000000	0.000000	0.000000
A	0.000600	0.033012	-0.117043	0.080430	0.003001	0.000000	0.000000	0.000000
Baa	0.001469	0.000734	0.088133	-0.163046	0.067569	0.004407	0.000734	0.000000
Ba	0.000000	0.000000	0.009159	0.184699	-0.293077	0.096166	0.003053	0.000000
B	0.000000	0.000000	0.002280	0.014822	0.093489	-0.246265	0.124273	0.011401
Caa	0.000000	0.000000	0.000000	0.000000	0.000000	0.120209	-0.540939	0.420730
D	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 2: True generator matrix $\mathbf{Q}$ for the Monte Carlo experiments

This generator matrix is taken from the paper by Christensen, Hansen and Lando (2004), which provides an empirical generator matrix based on the continuously observed rating data (hence, the complete data) for senior unsecured debt issues in the United States[4]. The observation period for the above generator matrix is from the 1st of January 1995 to the 31st of December 1999, based on 'Moody's Corporate Bond Default Database'. For a description of the data, see the original paper.

We create a synthetic migration database by simulating sample paths of a Markov jump process, given the true generator matrix in Table 2. Considering the limitedness of internal migration records of individual banks, we set the number of obligors in each rating grade to be 100. Also, a maturity of sample is set to seven years. Hence, for each simulation, seven partially observed migrating matrices for 700 obligors are constructed. The 250 simulations are implemented and the 250 generator matrices are estimated from the simulated seven-year annual migration histories for the DA, the WA, the QOG, the EM algorithm and the MCMC.

For the MCMC, 10,000 intensity matrices, including a burn-in period [5]of 1,000 iterations, are drawn for each estimation. For the choice of the prior parameters given in (3.8), we utilize the results of the EM algorithm. Specifically, we set α_{ij} to zero if the ij th elements of the generator matrix, estimated by the EM algorithm, converge to the value less than 1e-14. Otherwise, α_{ij} is set to 1. Also, β_i is set to 1. Normally, the posterior mean of the distribution is chosen for the point estimate of the parameters. Instead of the mean estimate, we choose the posterior mode estimate from the samples of $\tilde{q}_{ij}$ because the posterior distribution for some parameters is

[4] To preclude the effect of rounding, diagonal elements are adjusted from the original generator matrix so that each row adds up to zero.

[5] The first n samples are usually discarded to allow the Markov chain to approach its stationary distribution. These n values are known as a "burn-in".

found to be heavily skewed. To obtain the posterior mode, we use a kernel smoothing method with the normal kernel function to find the density. To restrict the density to positive values, we apply the logarithmic transform to the samples. The density is evaluated at 100 equally-spaced points.

We provide some examples of the estimated generator matrix before analyzing the results. Table 3 shows a set of the 250th estimates for the generator matrix obtained by the five estimation methods.

$$\tilde{\mathbf{Q}}_{DA} = \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline & \text{Aaa} & \text{Aa} & \text{A} & \text{Baa} & \text{Ba} & \text{B} & \text{Caa} & \text{D} \\ \hline \text{Aaa} & -0.067939 & 0.062851 & 0.005087 & 0.000000 & 0.000000 & 0.000001 & 0.000000 & 0.000000 \\ \hline \text{Aa} & 0.006932 & -0.139647 & 0.132640 & 0.000000 & 0.000051 & 0.000024 & 0.000000 & 0.000000 \\ \hline \text{A} & 0.002164 & 0.031909 & -0.121346 & 0.086827 & 0.000413 & 0.000000 & 0.000032 & 0.000000 \\ \hline \text{Baa} & 0.002630 & 0.001270 & 0.091054 & -0.180665 & 0.082242 & 0.003469 & 0.000000 & 0.000000 \\ \hline \text{Ba} & 0.000000 & 0.000000 & 0.000000 & 0.211340 & -0.315921 & 0.103667 & 0.000000 & 0.000914 \\ \hline \text{B} & 0.000006 & 0.000000 & 0.004982 & 0.000000 & 0.117700 & -0.302074 & 0.179387 & 0.000000 \\ \hline \text{Caa} & 0.000000 & 0.000008 & 0.000000 & 0.000222 & 0.000000 & 0.150542 & -0.611930 & 0.461158 \\ \hline \text{D} & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ \hline \end{array}$$

$$\tilde{\mathbf{Q}}_{WA} = \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline & \text{Aaa} & \text{Aa} & \text{A} & \text{Baa} & \text{Ba} & \text{B} & \text{Caa} & \text{D} \\ \hline \text{Aaa} & -0.067859 & 0.062777 & 0.005081 & 0.000000 & 0.000000 & 0.000001 & 0.000000 & 0.000000 \\ \hline \text{Aa} & 0.006785 & -0.136673 & 0.129815 & 0.000000 & 0.000049 & 0.000024 & 0.000000 & 0.000000 \\ \hline \text{A} & 0.002162 & 0.031870 & -0.121197 & 0.086720 & 0.000413 & 0.000000 & 0.000032 & 0.000000 \\ \hline \text{Baa} & 0.002629 & 0.001270 & 0.091009 & -0.180577 & 0.082202 & 0.003467 & 0.000000 & 0.000000 \\ \hline \text{Ba} & 0.000000 & 0.000000 & 0.000000 & 0.205863 & -0.307734 & 0.100981 & 0.000000 & 0.000890 \\ \hline \text{B} & 0.000006 & 0.000000 & 0.004905 & 0.000000 & 0.115882 & -0.297409 & 0.176616 & 0.000000 \\ \hline \text{Caa} & 0.000000 & 0.000008 & 0.000000 & 0.000221 & 0.000000 & 0.149948 & -0.609515 & 0.459339 \\ \hline \text{D} & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ \hline \end{array}$$

$$\tilde{\mathbf{Q}}_{QOG} = \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline & \text{Aaa} & \text{Aa} & \text{A} & \text{Baa} & \text{Ba} & \text{B} & \text{Caa} & \text{D} \\ \hline \text{Aaa} & -0.067833 & 0.062798 & 0.005034 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ \hline \text{Aa} & 0.005016 & -0.135739 & 0.130723 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ \hline \text{A} & 0.002111 & 0.031856 & -0.121101 & 0.086774 & 0.000360 & 0.000000 & 0.000000 & 0.000000 \\ \hline \text{Baa} & 0.002601 & 0.001241 & 0.091024 & -0.180519 & 0.082213 & 0.003439 & 0.000000 & 0.000000 \\ \hline \text{Ba} & 0.000000 & 0.000000 & 0.000000 & 0.206324 & -0.304976 & 0.098652 & 0.000000 & 0.000000 \\ \hline \text{B} & 0.000000 & 0.000000 & 0.002686 & 0.000000 & 0.115404 & -0.295181 & 0.177091 & 0.000000 \\ \hline \text{Caa} & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.149015 & -0.608646 & 0.459631 \\ \hline \text{D} & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ \hline \end{array}$$

$\tilde{\mathbf{Q}}_{EM} =$

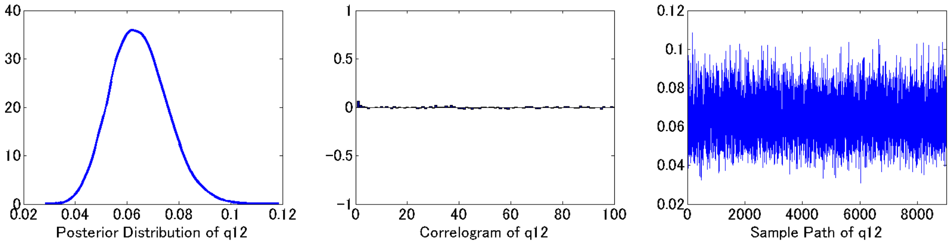
	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	-0.067778	0.062904	0.004874	0.000000	0.000000	0.000000	0.000000	0.000000
Aa	0.006925	-0.133669	0.126744	0.000000	0.000000	0.000000	0.000000	0.000000
A	0.002173	0.031864	-0.116871	0.082627	0.000206	0.000000	0.000000	0.000000
Baa	0.002442	0.001140	0.087266	-0.176067	0.082361	0.002858	0.000000	0.000000
Ba	0.000000	0.000000	0.000000	0.200875	-0.294905	0.094030	0.000000	0.000000
B	0.000000	0.000000	0.003981	0.000000	0.110317	-0.278473	0.164174	0.000000
Caa	0.000000	0.000000	0.000000	0.000000	0.000000	0.143990	-0.598919	0.454929
D	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

$\tilde{\mathbf{Q}}_{MC} =$

	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	-0.066968	0.061980	0.004988	0.000000	0.000000	0.000000	0.000000	0.000000
Aa	0.007334	-0.135485	0.128151	0.000000	0.000000	0.000000	0.000000	0.000000
A	0.002255	0.030827	-0.115493	0.081752	0.000659	0.000000	0.000000	0.000000
Baa	0.002190	0.001284	0.085048	-0.173103	0.081350	0.003231	0.000000	0.000000
Ba	0.000000	0.000000	0.000395	0.199069	-0.294236	0.094772	0.000000	0.000000
B	0.000000	0.000000	0.003209	0.000545	0.105667	-0.271037	0.160036	0.001581
Caa	0.000000	0.000000	0.000000	0.000000	0.001023	0.143992	-0.585474	0.440458
D	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 3: The 250th estimates of a generator matrix based on the DA, the WA, the QOG, the EM and the MCMC

We provide typical example graphs for the estimated posterior density, correlogram and sample path by the MCMC. Figure 1 presents the graphs of $\tilde{q}_{12}$, $\tilde{q}_{13}$, $\tilde{q}_{31}$and $\tilde{q}_{35}$ with respect to the same 250th estimated generator matrix.



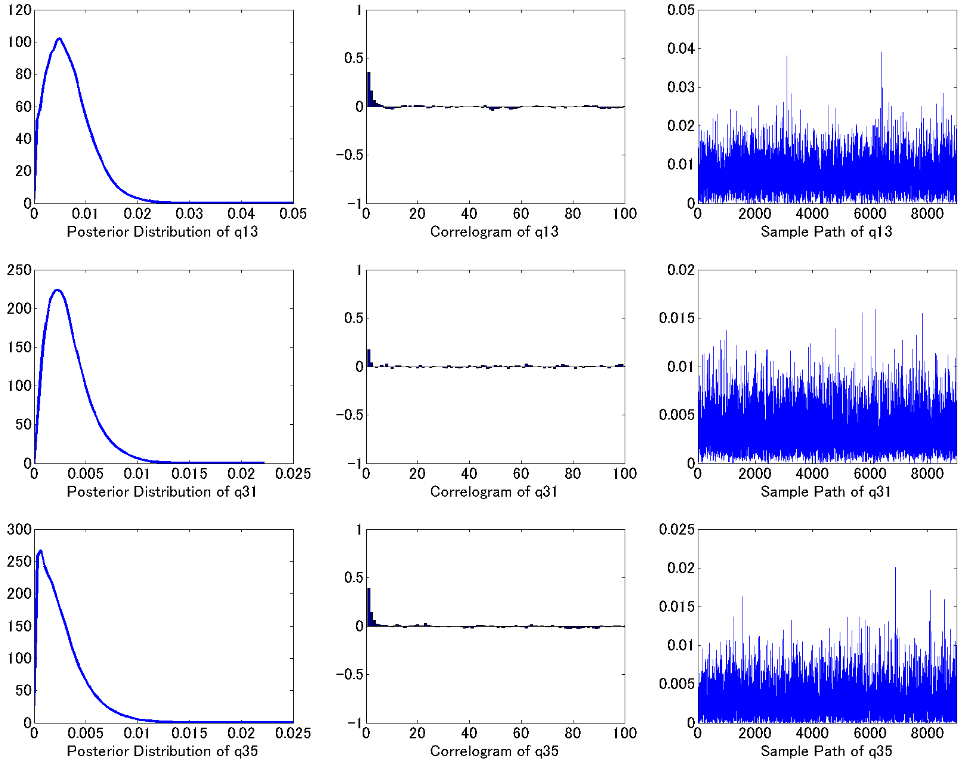


Figure 1: Examples of the posterior distribution, correlogram and sample paths of $\tilde{q}_{12}$, $\tilde{q}_{13}$, $\tilde{q}_{31}$and $\tilde{q}_{35}$

The distributions of some parameters are heavily skewed. This is especially the case with the $\tilde{q}_{35}$, where the peak of the distribution is barely found near zero. This implies that the difference between the posterior mean estimate and the posterior mode estimate is not negligible in this test. Note that none of the figures of correlogram and sample paths show any alarming pathology in the sampling.

4.1.2. Difference by Default Probabilities

Default probabilities can be calculated using the matrix exponentials. We first examine descriptive results of the Monte Carlo experiments. Table 4 shows the average of one-year default probabilities based on the 250 simulations. Table 5 provides the mean difference of the default probabilities based on the full set of 250 estimates.

	DA	WA	QOG	EM	MCMC	$\mathbf{P}(\mathbf{Q})$
Aaa	0.0000200	0.0000196	0.0000028	0.0000022	0.0000017	0.0000011
Aa	0.0003564	0.0003484	0.0000643	0.0000550	0.0000443	0.0000185
A	0.0035477	0.0035090	0.0016666	0.0014105	0.0011234	0.0006722
Baa	0.0394906	0.0390563	0.0298908	0.0260767	0.0206377	0.0208731
Ba	0.2353155	0.2308084	0.1754193	0.1476192	0.1293078	0.1605010
B	2.7628071	2.6962556	2.7124158	2.4275994	2.3806991	3.0429080
Caa	34.0043320	33.8769068	33.9248198	32.9734381	32.5564478	32.6242442

Table 4: Mean estimates of default probabilities (%)

	DA	WA	QOG	EM	MCMC
Aaa	-0.0000189	-0.0000185	-0.0000017	-0.0000011	-0.0000006
Aa	-0.0003380	-0.0003299	-0.0000458	-0.0000365	-0.0000259
A	-0.0028754	-0.0028367	-0.0009943	-0.0007382	-0.0004511
Baa	-0.0186175	-0.0181832	-0.0090178	-0.0052037	0.0002354
Ba	-0.0748145	-0.0703074	-0.0149183	0.0128819	0.0311932
B	0.2801009	0.3466523	0.3304921	0.6153086	0.6622089
Caa	-1.3800878	-1.2526625	-1.3005755	-0.3491938	0.0677965

Table 5: Mean differences of the default probabilities
with respect to $\mathbf{P}(\mathbf{Q})$ (%)

Looking at the difference in Table 5, we find that the MCMC gives the lowest errors in the rating grades except Ba and B. The EM and the QOG rank second and third respectively in the rating grades from Aaa to Baa. The results of the DA and the WA have the worst and the second worst errors respectively in the rating grades from Aaa to Ba. In summary, with respect to investment grades (i.e. from Aaa to Baa), we can rank the five methods in the order of the MCMC, the EM, the QOG, the WA and the DA. However, the descriptive results in the non-investment grade are still mixed.

To make a more clear evaluation of these estimated default probabilities, we then implement a bootstrapping simulation to derive the confidence intervals of default probabilities, given the true $\mathbf{Q}$. We thereby examine whether the default probabilities in Table 4 are statistical distinguishable from the true probabilities of default. Generating a history of the rating process for a continuous time Markov chain is carried out by random sampling from the exponential and the multinomial distributions. The initial distribution of obligors and the simulation horizon for the bootstrapping simulation are the same as those for the creation of the synthetic migration data. The 100,000 bootstrapping simulations are carried out. The procedures for bootstrapping are summarized as follows:

1. Generate the histories of all the obligors for a horizon, which follows a continuous time Markov chain.
2. Calculate the statistics $\tilde{N}_{ij}(T)$ and $\tilde{R}_i(T)$ and estimate $\tilde{\mathbf{Q}}$ from the estimator given in (2.11).
3. Compute $\mathbf{P}(\tilde{\mathbf{Q}})$ to derive the default probabilities of the non-default rating grades.
4. Iterate 2. $\sim$ 4.up to the number of simulations.

Figure 2 shows the bootstrapped distribution of default probabilities for Aaa and Caa as an example. Table 6 provides the resulting confidence interval of default probabilities from the bootstrapped distributions.

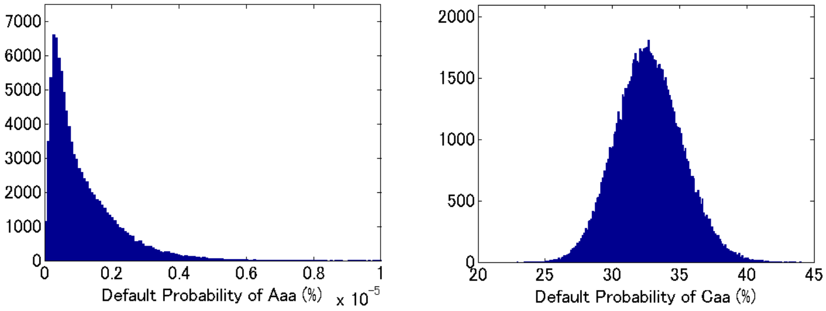


Figure 2: Bootstrapped distribution of default probability for Aaa and Caa

	Critical Value (5%)			Critical Value (1%)	
	lower bound	upper bound		lower bound	upper bound
Aaa	0.00000012	0.0000039	Aaa	0.00000006	0.0000056
Aa	0.0000033	0.0000530	Aa	0.0000022	0.0000704
A	0.0001393	0.0018182	A	0.0000930	0.0023687
Baa	0.0038980	0.0610191	Baa	0.0026707	0.0783694
Ba	0.0833848	0.2684161	Ba	0.0683690	0.3106287
B	2.1646062	4.1133453	B	1.9310066	4.4845520
Caa	28.0858111	37.7123761	Caa	26.7377488	39.4521377

Table 6: Confidence interval of default probabilities (%)

Most striking is the statistical result with respect to the MCMC. Its mean default probabilities of the non-default rating grades are all within the confidence intervals. In contrast, the mean default probabilities of Aaa, Aa and A with respect to the DA and the WA methods are all above the upper bound with a critical value of 1%. In other words, it can be stated with 99% confidence that the default probabilities with respect to the DA and the WA for these rating grades are "too large" from the true ones. As for the EM algorithm and the QOG, the null hypothesis is rejected with regard to Aa at a critical value of 5%.

There are two main points we can take from these results: First, the MCMC may be the best performer to capture investment grade default probabilities in the finite-sample setting. Second, the DA and the WA may create a significant deviation in the estimated default probabilities of investment grades, considering their poor finite-sample performances regarding Aaa, Aa and A. The evaluation of the test result between the EM algorithm and the QOG is still difficult, although the result implies that the two methods are more accurate then the DA and the WA and are less precise than the MCMC in terms of estimating investment grade default probabilities. To further investigate the statistical difference between these methods, we employ another measure for evaluating the competing methods in the next test.

4.1.3. Difference by the L^1 and Singular Value Decomposition Metrics

The difference by default probability is not necessarily a sufficient indicator for the comparison of competing methods because we are dealing with a 'matrix', not a scalar. Special measures are required to examine how the estimated generator matrix differs from the true one. We employ the following two metrics to measure the distance between two different transition matrices $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$, for $i, j \in K$.

$$D_{L^1}(\mathbf{A},\mathbf{B}) = \frac{1}{K^2}\sum_{i,j}|a_{ij} - b_{ij}|\,, \qquad D_{Svd}(\mathbf{A},\mathbf{B}) = M_{Svd}(\mathbf{A}) - M_{Svd}(\mathbf{B}),$$

$$\text{where} \quad M_{Svd}(\tilde{\mathbf{P}}) = \frac{\sum_{i=1}^{K}\sqrt{\lambda_i(\tilde{\mathbf{P}}^T\tilde{\mathbf{P}})}}{K}, \quad \tilde{\mathbf{P}} = \mathbf{P} - \mathbf{I}.$$

D_{L^1} may be intuitively easier to understand since it measures the distance by the mean absolute difference between the elements of the transition matrices. In contrast, D_{Svd}, developed by Jafry and Schuermann (2004), is a singular-value-based metric focusing on the mobility matrix $\tilde{\mathbf{P}}$. The authors explain that $M_{Svd}(\tilde{\mathbf{P}})$ is identical to the average probability of migration, if there is such an average probability constant across all possible states. According to their empirical studies, D_{Svd} is more appropriate in measuring the difference of the transition matrices than other ordinary metrics since it captures the off-diagonal differences better.

In order to make statistical tests with these metrics, we again implement a bootstapping simulation to derive the distribution of the two distance metrics between $\mathbf{P}(\mathbf{Q})$ and $\mathbf{P}(\tilde{\mathbf{Q}})$. The 100,000 simulations are carried out in the similar way as before. Figure 3 shows the bootstrapped distribution for the D_{L^1} and D_{Svd} metrics. Table 7 and Table 8 provide the confidence interval for the D_{L^1} metric and the one for the D_{Svd} metric, respectively

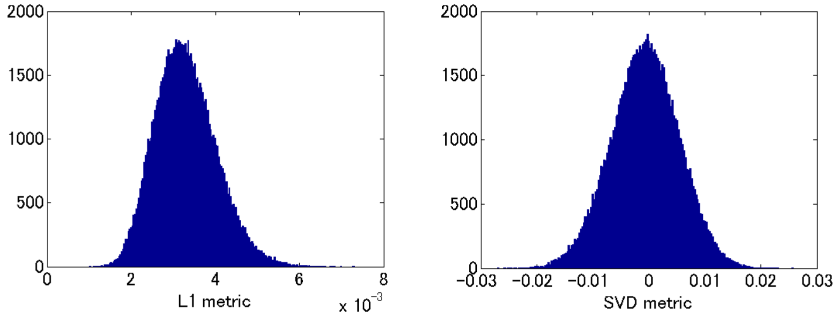


Figure 3: Bootstrapped distribution of the D_{L^1} and the D_{Svd} distance metrics between $\mathbf{P}(\mathbf{Q})$ and $\mathbf{P}(\tilde{\mathbf{Q}})$

	upper bound
Critical Value (0.05)	0.0046
Critical Value (0.01)	0.0052

Table 7: Confidence intervals of D_{L^1} distance between $\mathbf{P}(\mathbf{Q})$ and $\mathbf{P}(\tilde{\mathbf{Q}})$[6]

	lower bound	upper bound
Critical Value (0.05)	-0.0127	0.0110
Critical Value (0.01)	-0.0166	0.0146

Table 8: Confidence intervals of D_{Svd} distance between $\mathbf{P}(\mathbf{Q})$ and $\mathbf{P}(\tilde{\mathbf{Q}})$

Now we can evaluate the mean distance metrics between the true $\mathbf{P}(\mathbf{Q})$ and the five estimation methods. Table 9 provides the distance metrics of the five estimation methods based on the 250 simulations.

	DA	WA	QOG	EM	MCMC
D_{L^1}	0.00493	0.00472	0.00471	0.00422	0.00404
D_{Svd}	-0.01429	-0.01278	-0.01234	-0.00805	-0.00549

Table 9: Averages of the two distance metrics with respect to $\mathbf{P}(\mathbf{Q})$ based on the full set of 250 samples.

[6] The confidence interval for the D_{L^1} metric is expressed in only one direction because it is non-negative by definition.

Remarkably, both of the distance metrics regarding the MCMC and the EM algorithm are well within the confidence levels. This implies that the MCMC and the EM algorithm are likely to generate statistically indistinguishable transition matrices from the true $\mathbf{P}(\mathbf{Q})$. Moving on to the comparison of the rest of the methods, with respect to the DA and the WA method, the null hypothesis is rejected at a critical value of 5% for both the D_{L^1} distance and the D_{Svd} distance. Interestingly, the null hypothesis regarding the QOG is also rejected at a critical value of 5 % for the D_{L^1} distance, implying that the elements in the generator matrix by the QOG may be more biased than those estimated by the EM algorithm or the MCMC.

On balance, the test results almost parallel with those given in the difference by default probability. Thus, the DA and the WA method clearly underperform the other methods, while the MCMC again gives the most accurate performance of all. As for the D_{L^1} distance, the QOG generates a larger deviation than the EM algorithm and the MCMC. Hence, the main point from the statistical results in this section is that we can rank the small-sample performances of the five methods in the order of the MCMC, the EM, the QOG, the WA and the DA. In particular, the statistical results substantially differ between the first three methods and the latter two methods. This is not surprising because the DA and the WA method lack norm or optimality in their algorithm as mentioned earlier. The result that the QOG slightly underperforms the EM algorithm and the MCMC is not necessarily a puzzle as well. Since the nature of the QOG is 'fitting', not 'statistical inference' after all, it can be expected that the possibility of 'over-fitting' makes the QOG more subject to the variability of the finite-sample data than the EM algorithm and the MCMC.

4.1.4. Relevance to Risk Management

To illustrate the economic relevance of the statistical results in this section, we investigate the impact of the biases in the estimation of investment grade default probabilities on the loss calculation with respect to economic capital. Specifically, given a hypothetical credit portfolio and prespecified true parameters, we implement a simulation exercise to examine how much the economic capital based on parameters yielded by the five estimation methods deviate from the economic capital based on the true parameters. The economic capital is the amount of capital that banks and insurance companies set aside for a buffer against potential losses from their business activities and is usually defined as the α-quantile of a portfolio's loss distribution minus its expected loss. In deriving a loss distribution, we employ the so-called one-factor asset value model with the assumption that asset correlation is the same across all the obligors in the portfolio.

For the construction of a hypothetical investment grade portfolio, we utilize the work of Gordy (2000), in which several exemplary banking loan distributions are provided based on internal Federal Reserve Board surveys of large banking organizations. To focus the issue of LDPs, we modify the setting of Gordy (2000) by excluding Ba, B, and Caa from the loan sample so that the hypothetical portfolio is composed of investment grade loans. The resulting distribution of obligors for the portfolio is given in Table 10. The loan size of any single obligor is assumed to be the unit and the loss given default is set to 45% for every exposure.

Aaa	Aa	A	Baa
191	295	1463	1896

Table 10: Sample distribution of obligors for a hypothetical investment grade portfolio

In the asset value model, differences of the default probabilities may have a significant impact on the loss distribution by affecting the value of default thresholds and asset correlation. For the default threshold, note that an ordinary calibration method for a default threshold for rating grade k is given as

$$\beta_k = \Phi^{-1}(p_k) \quad k = 1, ..., K$$

where p_k is an unconditional default probability of rating grade k and $\Phi^{-1}(\cdot)$ is the inverse of the cumulative normal distribution function. Hence, biases in the estimation of a default probability lead to those in the default threshold. Table 11 shows the default thresholds calculated using the default probabilities in Table 4.

	DA	WA	QOG	EM	MCMC	$\mathbf{P}(\mathbf{Q})$
Aaa	-5.0691	-5.0725	-5.4286	-5.4716	-5.5169	-5.5909
Aa	-4.4898	-4.4947	-4.8420	-4.8728	-4.9152	-5.0842
A	-3.9731	-3.9757	-4.1494	-4.1875	-4.2388	-4.3527
Baa	-3.3563	-3.3594	-3.4326	-3.4694	-3.5318	-3.5288

Table 11: Default thresholds based on the default probabilities in Table 4.

For the asset correlation, to be sure, a usual practice is to employ equity return as its proxy. A number of empirical studies, however, report that the equity correlation may be a poor proxy of the asset value model (See De Servigny and Renault (2002) and Zeng and Zhang (2002)). In contrast, recent studies suggest the use of maximum likelihood estimation to back out the asset correlation from the default (panel) data (See Gordy and Heitfeld (2002), Düllmann and Scheule (2003) and Demey et al. (2004)). In the normal maximum likelihood estimation for the asset correlation, the likelihood function contains default thresholds as cardinal inputs. To clarify this point, consider a likelihood function under the assumption of a homogeneous portfolio with K rating grades, given T observations of default data. Let $m_{k,j}$ and $d_{k,j}$ denote the number of obligors and defaults at rating grade k at time j, respectively. Also, let $p^C(\beta_k, \rho)$ denote a conditional default probability for an obligor in rating grade k which is given as

$$p^C(\beta_k, \rho) = \Phi\left(\frac{\beta_k - \sqrt{\rho}Z}{\sqrt{1-\rho}}\right),$$

where Z is a common, standard normally distributed factor Z. Then, the log-likelihood function for the asset correlation is given as

$$L_j(\rho) = \sum_{j}^{T} \log \int_{-\infty}^{\infty} \prod_{k=1}^{K} \binom{m_{k,j}}{d_{k,j}} p^C(\beta_k, \rho)^{d_{k,j}} (1 - p^C(\beta_k, \rho))^{m_{k,j} - d_{k,j}} \phi(z) dz. \quad (4.1)$$

Thus, the likelihood is a function of default thresholds and the difference of their values may also influence the maximum likelihood estimator of the asset correlation to a significant degree[7]. Hence, the following simulation exercises cover two cases. In the first case, the asset correlation is exogenously set to 0.25[8]. Differences of default thresholds only matter in this case. Another case is to employ the maximum likelihood estimator of the asset correlation obtained through (4.1) using different thresholds from the DA, the WA, the QOG, the EM algorithm, and the MCMC. To capture the average estimates, a set of synthetic default data is simulated 250 times by setting the true default thresholds to be those of $\mathbf{P}(\mathbf{Q})$ and the true asset correlation to be 0.25. Unlike the case of estimating default probabilities, here we create a larger synthetic database (consider, for example, external databases from rating agencies mapped to internal default records) because utilizing an internal rating data for estimating the asset correlation seems still inconceivable in real applications. Specifically, we set a sample of maturity to be $T = 20$ years and the total number of observed firms for Aaa, Aa, A and Baa at each year t to be 100, 300, 800 and 1500, respectively. A standard Gaussian quadrature is applied to the numerical integration for the maximization of (4.1). Table 12 shows a summary of the averages of the maximum likelihood estimator of the asset correlation based on the 250 simulations. Not surprisingly, the mean estimates from the DA reveals the largest deviation from the true ρ, while the value from the MCMC is well estimated, closest to 0.25.

	DA	WA	QOG	EM	MCMC
Mean estimates $\tilde{\rho}$	0.384	0.382	0.327	0.301	0.261
$\tilde{\rho}$ / True ρ (0.25)	154%	153%	131%	120%	104%

Table 12: Averages of the estimated asset correlation based on the 250 simulations

[7] It is possible to estimate the threshold β_k and the asset correlation ρ jointly. However, unless it is assumed that all firms in the sample have a single default probability, the increase in the number of parameters to be estimated may discourage practitioners to adopt the joint estimation. In this regard, the past empirical studies assume the existence of a single default probability for the whole portfolio in the case of the joint estimation (See Gordy and Heitfeld (2002) and Demey et al.(2004)).

[8] According to the empirical study by Bluhm and Overbeck (2003) based on the unsmoothed Moody's corporate bond dafault data, the asset correlation in investment grades were found to range from 15.95% to 31.5%. We set the true asset correlation to be 0.25 (25%) based on these results.

We generate a loss distribution of the portfolio with 500,000 simulations for the five methods and the true parameters. The table 13 shows a summary of the computed economic capitals of the first case where the asset correlation is equally set to 0.25. "True E.C." denotes the economic capital using the true default thresholds and the true asset correlation.

	DA	WA	QOG	EM	MCMC
Econ. Capital (99%)	5.04	5.05	3.79	3.37	2.97
E.C. (99%) / True E.C. <2.97>	170%	170%	128%	114%	100%

	DA	WA	QOG	EM	MCMC
Econ. Capital (99.9%)	17.19	17.20	13.69	11.92	10.17
E.C.(99.9%) / True E.C.<10.17>	169%	169%	135%	117%	100%

Table 13: Computed economic capitals of the first case.

A special attention should be paid here to the magnitude of the differences of economic capitals yielded by the five methods. Remarkably, the level of economic capital involving the MCMC is the same as the true economic capital, regardless of the confidence levels. In contrast, the economic capitals involving the DA and the WA are almost 70% larger than the true value at both the 99% and the 99.9% confidence level. Even with respect to the QOG, its magnitude of the difference from the true value amounts to nearly 30% at the 99% confidence level.

	DA	WA	QOG	EM	MCMC
Econ. Capital (99%)	6.39	6.39	4.68	3.82	2.97
E.C. (99%) / True E.C. <2.97>	215%	215%	158%	129%	100%

	DA	WA	QOG	EM	MCMC
Econ. Capital (99.9%)	31.59	31.14	19.53	15.97	11.07
E.C.(99.9%) / True E.C.<10.17>	311%	306%	192%	157%	109%

Table 14: Computed economic capitals of the second case

Table 14 shows a summary of the computed economic capitals of the second case in which the asset correlations in Table 12 are applied. Not surprisingly, the differences become even more significant, due to the biases in the estimation of the asset correlation. The economic capitals based on the DA and the WA are twice larger and three times larger than the true value at the 99% and the 99.9% confidence level respectively. The economic capital involving the QOG almost doubles the true value at the 99.9% confidence level. Even with respect to the EM algorithm, its magnitude of the difference of economic capital amounts to nearly 60%. In contrast, the MCMC shows only the 9% difference from the true economic value at the 99.9% confidence level.

Although the results are still case-specific, it can be safely said that the differences in the estimation methods of a generator matrix have the potential to affect the level of a loss distribution and the resulting economic capital for the investment grade portfolio.

4.2. The Generator Matrix for Japanese Corporations

Finally, we provide an empirical study based on the annual transition data of Japanese corporations. The migration data is taken from Rating and Investment Information, Inc.'s (R&I) public database. The sample period is from 1991 until 2000. The rating information for the categories 'Non-Rated' and 'Lost' are discarded from our database. Because the records of transitions involving 'CCC/C' are not included in the selected sample period, the category 'CCC/C' is omitted in the database (hence, the size of the data matrix here is 7×7). The MCMC is applied to the estimation of the generator matrix, based on the same procedure in the section of the Monte Carlo experiments. Then, point estimates of one-year default probability are determined and a bootstapping simulation is carried out to obtain the upper bounds of the default probability for the given estimated generator matrix.

For the comparison, we first provide one-year default probabilities based on the ordinary cohort-based matrix in Table 14.

	Cohort-based
AAA	0.000000
AA	0.000000
A	0.038895
BBB	0.109449
BB	2.173913
B	17.500000

Table 14: Default probabilities based on the cohort method (%)

Note that the estimates include zero values for default probability in the two highest rating grades. Table 15 shows the estimated generator matrix $\tilde{\mathbf{Q}}_{MC}$ by the MCMC method. Table 16 provides the point estimate of one-year default probabilities, together with their upper bounds based on the 95% and the 99% quantiles from the 100,000 bootstrapping simulations.

	AAA	AA	A	BBB	BB	B	D
AAA	-0.1069178	0.1069178	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
AA	0.0009415	-0.0909755	0.0898465	0.0001875	0.0000000	0.0000000	0.0000000
A	0.0000000	0.0124148	-0.0776680	0.0647725	0.0000811	0.0000000	0.0003996
BBB	0.0000000	0.0001995	0.0335721	-0.0823580	0.0481388	0.0000922	0.0003555
BB	0.0000000	0.0000000	0.0000000	0.0664727	-0.1189581	0.0340593	0.0184261
B	0.0000000	0.0000000	0.0000000	0.0338534	0.0253521	-0.2593532	0.2001477
D	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

Table 15: Estimated empirical generator matrix $\tilde{\mathbf{Q}}_{MC}$ by the MCMC method, based on the R&I database from 1991 to 2000

	Point Estimate	Upper bound (95%)	Upper bound (99%)
AAA	0.000061	0.000185	0.000266
AA	0.001760	0.005092	0.006953
A	0.040626	0.115069	0.153956
BBB	0.082000	0.150236	0.186805
BB	2.040510	2.891271	3.289120
B	17.654014	25.607993	29.555664

Table 16: Point estimate and upper bounds of the default probabilities based on the MCMC method (%)

The results of Table 16 are similar to those of Lando and Skødeberg (2002) and Fuertes and Kalotychou (2005). Non-zero default probabilities are obtained for all the non-default rating grades. The upper bounds of default probabilities in investment grades seem reasonable as compared, for example, with those derived by ordinary binomial methods.

5. CONCLUDING REMARKS

This paper considers the estimation of an empirical generator matrix from discretely observed rating transitions in search for probabilities of rare default events in high investment grades. The five competing estimation methodologies: diagonal adjustment, weighted adjustment, quasi-optimization approach, expectation maximization algorithm and Markov chain Monte Carlo estimation – are investigated in terms of the accuracy and statistical validity of the estimated default probability and various matrix norms. The implications for banking risk management are also explored with a case study regarding economic capital for a hypothetical investment grade portfolio. The paper then presents an empirical generator matrix based on the annual transition data of Japanese corporations.

The results of Monte Carlo experiments suggest that the choice of estimation methodology is likely to significantly affect the resulting default probabilities and the mobility of a transition matrix. In particular, a generator matrix determined by the parsimonious DA or WA methods seems to be strongly affected by a deviation arising from the post-adjustment. The experiments also show that the MCMC is the only method whose estimated default probabilities and matrix norms are all statistically indistinguishable from true parameters in the experiments. Hence, as far as the results here are concerned, we have reached the conclusion that the MCMC method gives the most accurate finite-sample performance of all the five method. A case study regarding the economic capital of a hypothetical investment grade portfolio highlights further differences of these methods. Its result shows that the different estimation methods of a generator matrix have the potential to yield significantly different estimates of a loss distribution and the resulting economic capital of the investment grade portfolio.

The discussions and empirical studies shown in this paper should help practitioners realize the value of having high-frequency observations for rating histories. If a direct estimation based on continuously observed data is possible, one can obtain exact maximum likelihood estimates very easily without concern for the embeddability problem or the existence and uniqueness of the maximum likelihood estimator. In addition, more information can be efficiently taken into account because even censored data, such as 'non-rated' or 'lost', can be used in the direct estimation of the generator matrix. In that sense, the methods investigated in this paper should be regarded as a practical approach for use in the transition period before a more advanced informational environment is realized for real applications.

REFERENCES

[1] Asmussen, S., Nerman, O., Olsson, M., 1996. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics 23, 419-441.

[2] Basel Committee on Banking Supervision, 2005. Validation of low-default portfolios in the Basel II framework. Basel Committee Newsletter No. 6.

[3] Bladt, M., Sørensen, M., 2005. Statistical inference for discretely observed Markov jump processes. Journal of the Royal Statistical Society: Series B 67, 395-410.

[4] Bluhm, C., Overbeck, L., 2003. Systematic risk in uniform credit portfolios. In: Credit Risk; Measurement, Evaluation and Management, Physica-Verlag/Springer.

[5] British Bankers' Association, London Investment Banking Association, International Swaps and Derivatives Association, 2005. Low default portfolios. Joint Industry Working Group Discussion Paper.

[6] Christensen, J., Hansen, E., Lando, D., 2004. Confidence sets for continuous-time rating transition probabilities. Journal of Banking and Finance 28 (5), 2575-2602.

[7] Demey, P., Jouanin, J. F., Roget, C., Roncalli, T., 2004. Maximum likelihood estimate of default correlations. Risk, 104-108.

[8] De Servigny, A., Renault, O., 2002. Default correlation: empirical evidence. Standard and Poor's.

[9] Düllmann, K., Scheule, H., 2003. Asset correlation of German corporate obligors: its estimation, its drivers and implications for regulatory capital. Working paper.

[10] Fuertes, A., Kalotychou, E., 2005. On sovereign credit migration: small sample properties and rating evolution. Working paper, Cass Faculty of Finance.

[11] Gordy, M., 2000. A comparative anatomy of credit risk models. Journal of Banking and Finance 24 (5), 119-149.

[12] Gordy, M., Heitfield, E., Estimating default correlations from short panels of credit rating performance data. Working paper, Federal Reserve Board.

[13] Hansen, S., Schuermann, T., 2005. Confidence intervals for probabilities of default. Working paper, Wharton Financial Institutions Center.

[14] Israel, R. B., Rosenthal, J. S., Wei, J. Z., 2001. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. Mathematical Finance 11, 245-265.

[15] Jafry, Y., Schuermann, T., 2004. Measurement and estimation of credit migration matrices. Journal of Banking and Finance 28 (11), 2603-2639.

[16] Kalbfleisch, J. D., Lawless, J. F., 1985. The analysis of panel data under a Markov assumption. Journal of the American Statistical Association 80, 863-871.

[17] Kiefer, N. M., Larson, C. E., 2004. Testing simple Markov structures for credit rating transitions. Working paper, U.S Treasury Office of the Comptroller of the Currency.

[18] Kreinin, A., Sidelnikova, M., 2001. Regularization algorithms for transition matrices. Technical Paper, Algo Research Quartely 4, 23-40.

[19] Lando, D., Skødeberg, T., 2002. Analyzing rating transitions and rating drift with continuous observations. Journal of Banking and Finance 26 (2-3), 423-444.

[20] McLachlan, G. J., Krishnan, T., 1997. The EM algorithm and extensions. Wiley, New York.

[21] Norris, J. R., 1998. Markov chains. Cambridge University Press.

[22] Tuenter, H., 2000. The minimum L_2-distance projection onto the canonical simplex: a simple algorithm. Working paper, Algorithmics Inc.

[23] Van Loan, C. F., 1978. Computing integrals involving the matrix exponentials. IEEE Transaction Automatic Control. AC-23, 395-404.

[24] Zeng, B., Zhang, J., 2002. Measuring credit correlations: equity correlations are not enough!. KMV working paper.

APPENDIX A
ON THE QOG ALGORITHM

We provide a small exposition on the QOG algorithm along the lines of the argument found in Tuenter (2000). A row of $\log(\tilde{\mathbf{P}})$ is permutated such that $a_1 \leq a_2 \leq \ldots \leq a_K$. Note that the permutation does not affect the solution in this problem setting. Let $\mathbf{z}^*$ denote the optimal solution for (3.3). Let m denote an index for binding elements in z_i^* such that $z_i^* = 0$ for $2 \leq i \leq m$ and otherwise $z_i^* \neq 0$. Also, let $\Im$ denote an index for the set of binding elements in z_i^* such that $\Im = \{i|2 \leq i \leq m\}$. Using the Lagrange multipliers, we have $z_i^* = a_i + \lambda$, for $i \notin \Im$. Then, the problem (3.3) can be rewritten as

$$\min \ (K-m+1)\lambda^2 + \textstyle\sum_{i\in\Im} a_i^2$$

$$\begin{aligned} \text{s.t. } (K-m+1)\lambda + \textstyle\sum_{i\notin\Im} a_i &= 0, \quad \lambda \geq -a_m \\ \text{for } m &\in \{2,\ldots,K\}\,. \end{aligned}$$

Hence, we have

$$\lambda = -\frac{1}{(K-m+1)}\left(\textstyle\sum_{i\notin\Im} a_i\right). \tag{A.1}$$

By substituting for λ, we can further reduce the above problem to a univariate problem $f(m)$ defined by

$$\min \ f(m) = \frac{1}{(K-m+1)}\left(\textstyle\sum_{i\notin\Im} a_i\right)^2 + \textstyle\sum_{i\in\Im} a_i^2 \tag{A.2}$$

$$\begin{aligned} \text{s.t. } (K-m+1)a_m - \left(\textstyle\sum_{i\notin\Im} a_i\right) &\geq 0 \qquad\qquad (A.3) \\ \text{for } m &\in \{2,\ldots,K\}\,. \end{aligned}$$

Note that the sequence $S(m) = (K-m+1)a_m - \left(\sum_{i\notin\Im} a_i\right)$ is non-decreasing over m because

$$S(m) - S(m-1) = (a_m - a_{m-1})(K-m+2).$$

The function $f(m)$ is also non-decreasing over m because

$$f(m) - f(m-1) = \frac{1}{(K-m+1)(K-m+2)}\left(S(m)\right)^2.$$

From these results, one finds that the solution to (A.2) is the smallest index m^*, which satisfies the condition (A.3). Hence, from (A.1), the solution to the QOG is given by

$$z_i^* = \begin{cases} 0 & if \ \ 2 \leq i \leq m^* \\ a_i - \frac{1}{K-m^*+1}\left(\sum_{i\notin\Im} a_i\right) & otherwise \end{cases}.$$

APPENDIX B
CONDITIONAL EXPECTATION FOR THE EM ALGORITHM

This appendix roughly sketches a derivation for (3.6) and (3.7), based on the results by Asmussen et al. (1996). We omit the notation h for simplicity in the following derivation.

B.0.1. Total Amount of Holding Time

Remember that $R_i(t) = \int_{t_n}^{t_{n+1}} 1_{\{X(s)=i\}} ds$ is the total time of staying in i during the observational interval. Let $x(t_n)$ and $x(t_{n+1})$ denote the rating grade observed at t_n and t_{n+1} respectively. Assuming that the observational interval is a constant Δt over the whole period, under the assumption of a CTMC, we have

$$\begin{aligned}
& \mathbb{E}\left[R_i(\Delta t) | X(t_{n+1}) = x(t_{n+1}), X(t_n) = x(t_n)\right] \\
= & \int_{t_n}^{t_{n+1}} \mathbb{E}\left[1_{\{X(s)=i\}} | X(t_{n+1}) = x(t_{n+1}), X(t_n) = x(t_n)\right] ds \\
= & \int_{t_n}^{t_{n+1}} \frac{\mathbb{P}\left(X\left(s\right) = i, X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n)\right)}{\mathbb{P}(X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n))} ds, \qquad \text{(B.1)}
\end{aligned}$$

where

$$\mathbb{P}(X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n)) = \mathbf{e}_{x(t_n)}^T \exp(\mathbf{Q}\Delta t) \mathbf{e}_{x(t_{n+1})}, \qquad \text{(B.2)}$$

which can be moved outside the integral (B.1). From the Markov property, we have

$$\begin{aligned}
& \int_{t_n}^{t_{n+1}} \mathbb{P}\left(X\left(s\right) = i, X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n)\right) ds \\
= & \int_{t_n}^{t_{n+1}} \mathbb{P}\left(X\left(s\right) = i | X(t_n) = x(t_n)\right) \mathbb{P}\left(X\left(t_{n+1}\right) = x(t_{n+1}) | X(s) = i\right) ds \\
= & \mathbf{e}_{x(t_n)}^T \left(\int_{t_n}^{t_{n+1}} \exp(\mathbf{Q}(s - t_n)) \mathbf{e}_i \mathbf{e}_i^T \exp(\mathbf{Q}(t_{n+1} - s)) ds \right) \mathbf{e}_{x(t_{n+1})}.
\end{aligned}$$

Hence, we have reached the expression (3.6).

B.0.2. Number of Transitions

Let us discretize each of the intervals by ϵ such that $\frac{(t_{n+1} - t_n)}{N} = \epsilon$. Consider the discretized approximation of the number of transitions from i to j as

$$N_{ij}(\Delta t) \approx \sum_{g=0}^{N-1} 1_{\{X(t_n + g\epsilon) = i, X(t_n + (g+1)\epsilon) = j\}}.$$

Then, we have

$$\begin{aligned}
& \mathbb{E}\left[\sum_{g=0}^{N-1} 1_{\{X(t_n+g\epsilon)=i, X(t_n+(g+1)\epsilon)=j\}} | X(t_{n+1}) = x(t_{n+1}), X(t_n) = x(t_n)\right] \\
= & \sum_{g=0}^{N-1} \mathbb{P}(X\left(t_n + g\epsilon\right) = i, X(t_n + (g+1)\epsilon) = j | X(t_{n+1}) = x(t_{n+1}), X(t_n) = x(t_n)) \\
= & \sum_{g=0}^{N-1} \frac{\mathbb{P}(X\left(t_n + g\epsilon\right) = i, X(t_n + (g+1)\epsilon) = j, X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n))}{\mathbb{P}(X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n))}.
\end{aligned} \tag{B.3}$$

We already have $\mathbb{P}(X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n))$ from (B.2), which can be moved outside the summation. Let us then consider the rest of the expressions in (B.3). Based on the Markov property, the summation can be expressed as

$$\begin{aligned}
& \sum_{g=0}^{N-1} \mathbb{P}\left(X\left(t_n + g\epsilon\right) = i, X(t_n + (g+1)\epsilon) = j, X(t_{n+1}) = x(t_{n+1}) | X(t_n) = x(t_n)\right) \\
= & \sum_{g=0}^{N-1} \mathbb{P}(X\left(t_n + g\epsilon\right) = i, | X(t_n) = x(t_n)) \mathbb{P}\left(X(t_n + (g+1)\epsilon) = j | X\left(t_n + g\epsilon\right) = i\right) \\
& \times \mathbb{P}\left(X(t_{n+1}) = x(t_{n+1}) | X(t_n + (g+1)\epsilon) = j\right).
\end{aligned} \tag{B.4}$$

For the middle term in last equation of (B.4), by taking the limit as $\epsilon \to 0$, we have

$$\begin{aligned}
& \lim_{\epsilon \to 0} \mathbb{P}\left(X(t_n + (g+1)\epsilon) = j | X\left(t_n + g\epsilon\right) = i\right) \\
= & \; q_{ij} dt.
\end{aligned}$$

Therefore, for (B.4), we have

$$\begin{aligned}
& \lim_{\epsilon \to 0} \sum_{g=0}^{N-1} \mathbf{e}_{x(t_n)}^T \exp(\mathbf{Q} \cdot g\epsilon) \mathbf{e}_i \mathbb{P}\left(J(t_n + (g+1)\epsilon) = j, J\left(t_n + g\epsilon\right) = i\right) \\
& \times \mathbf{e}_j^T \exp(\mathbf{Q} \cdot (t_{n+1} - (t_n + (g+1)\epsilon))) \mathbf{e}_{x(t_{n+1})} \\
= & \; \mathbf{e}_{x(t_n)}^T \left(q_{ij} \int_{t_n}^{t_{n+1}} \exp(\mathbf{Q}(s - t_n))(\mathbf{e}_i \mathbf{e}_j^T) \exp(\mathbf{Q}(t_{n+1} - s)) ds \right) \mathbf{e}_{x(t_{n+1})}.
\end{aligned}$$

Thus, we have reached the expression (3.7).

APPENDIX C
COMPUTING INTEGRALS OF THE MATRIX EXPONENTIAL

The seminal paper by Van Loan (1978) shows that the integral of the matrix exponential can be calculated by considering the augmented matrix as follows

$$\begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{0} & \mathbf{F}_{22} \end{pmatrix} = \exp\left(\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} t \right).$$

By calculating the exponential of the augmented matrix, we obtain

$$\begin{aligned} \mathbf{F}_{11} &= \exp(\mathbf{A}_{11}t), \qquad \mathbf{F}_{22} = \exp(\mathbf{A}_{22}t) \\ \mathbf{F}_{12} &= \int_0^t \exp(\mathbf{A}_{11}(t-s))\mathbf{A}_{12}\exp(\mathbf{A}_{22}s)ds. \end{aligned}$$

Hence, by substituting $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ with $\mathbf{Q}$, and $\mathbf{A}_{12}$ with $(\mathbf{e}_i\mathbf{e}_i^T)$ or $(\mathbf{e}_i\mathbf{e}_j^T)$, we can obtain (3.6) and (3.7). For the proof and the extension, see the original paper.