Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings

by

Robert B. Israel*, Jeffrey S. Rosenthal**, and Jason Z. Wei***

(December 1999; revised October 2000.)

The authors thank Melanie Cao, Ed Perkins, Jeremy Quastel, Peter Rosenthal, Tom Salisbury, two anonymous referees, and Stanley Pliska (the editor) for helpful comments and suggestions. They also thank David Lando for his communications with Jason Wei on the topic. Jeffrey Rosenthal acknowledges financial support by NSERC, and Jason Wei acknowledges financial support by SSHRC and the Connaught Fund of the University of Toronto.
* Department of Mathematics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2. Internet: `israel@math.ubc.ca`.
** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: `jeff@math.toronto.edu`.
*** Division of Management, University of Toronto at Scarborough, Scarborough, Ontario, Canada M1C 1A4. Internet: `wei@scar.utoronto.ca`.

Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings

Abstract.

In this paper we identify conditions under which a true generator does or does not exist for an empirically observed Markov transition matrix. We show how to search for valid generators and choose the "correct" one that is the most compatible with bond rating behaviors. We also show how to obtain an approximate generator when a true generator does not exist. We give illustrations using credit rating transition matrices published by Moody's and by Standard and Poor's.

1. Introduction.

Since the seminal work of Jarrow, Lando and Turnbull (1997), the use of credit rating transition matrices in credit risk modeling has received increasing attention. For example, Kijima and Komoribayashi (1998) provide an improvement on the estimation procedure in Jarrow et al. (1997); Belkin, Suchower and Forest Jr. (1998) propose a one-factor Markov process to model credit rating transitions; Kijima (1998), from a technical perspective, explains how a Markov chain model can lead to the known empirical regularities such as memory in rating changes and long term reversion of ratings; and Arvanitis, Gregory and Laurent (1999) develop a framework within which credit rating migrations can exhibit the usual empirical regularities. In a useful note, Lando (2000) shows how a transition matrix can be used to value credit derivatives such as a default swap.

Rating transition matrices have also received increasing attention in the financial industry. Two major bond rating services in the U.S., Moody's and Standard and Poor's, now both publish annual update of historical transition matrices, together with a wealth of other related information.

The shortest time interval within which a transition matrix is estimated is typically one year. The number of transition observations within a shorter period would be too small for a reliable transition matrix to be estimated. However, for valuation purposes (e.g. valuing a default swap), we frequently need a transition matrix for a period shorter than one year. If one can obtain a generator for a transition matrix P, meaning a matrix Q having row-sums 0 and non-negative off-diagonal entries such that $\exp(Q) = P$, then one can set $P(t) = \exp(tQ)$ to obtain matrices for any time $t \geq 0$. Unfortunately, there is a lack of guidance in the finance literature as to when such a generator Q exists; or how to find such a matrix if indeed it exists.

Authors such as Lando (2000) and Arvanitis et al. (1999) simply assume the existence of such generators. The only authors in finance who have addressed the issue of estimating a generator are Jarrow et al. (1997) who obtain an approximate generator by assuming that the probability for one rating to make more than one transition in one year is small. To our knowledge, no one in the finance literature has addressed the issues of existence and identification of transition matrix generators.

The objective of this paper is to identify conditions under which a true generator does or does not exist. We will show how to search for valid generators, and how to choose a "correct" generator which is the most compatible with bond rating behaviors. When a true generator does not exist, we will show how to obtain an approximate one.

Toward the later stage of our research, our attention was drawn to a study by Singer and Spilerman (1976) who examine similar issues in the context of social sciences. As they have succinctly summarized, to determine if an empirical transition matrix is compatible with a true generator or an underlying Markov process is an *embeddability* problem; to seek for the true generator once its existence is established is an *identification* problem. Some of our results turn out to be re-discoveries of those by Singer and Spilerman (1976) and earlier authors such as Elfving (1937), Kingman (1962), Chung (1967), and Johansen (1973, 1974); others are of our own. We attempt to attribute the results to the earlier authors whenever called for.

The rest of the paper is presented in seven sections. Section 2 presents a simple method of finding a generator and identify conditions under which this method works. Section 3 provides a "fix" to the method in Section 2 when negative off-diagonal entries are present in a candidate generator matrix. Section 4 discusses some examples from finance. Section 5 presents some further results on the existence and uniqueness of generators. In Section 6, an algorithm is outlined for searching a valid generator when the simple series method in Section 2 fails. Further discussions are given in Section 7. The paper is concluded in Section 8.

Proofs of theorems are relegated to the appendix.

2. Finding a Candidate Generator.

Let P be a time-homogeneous Markov transition matrix, i.e. an $N \times N$ real matrix with non-negative entries and with row-sums 1. We are interested in finding a generator Q, i.e. an $N \times N$ real matrix with non-negative off-diagonal entries and with row-sums 0, such that $\exp(Q) = P$. (Here and throughout,

$$\exp(tQ) = I + tQ + (tQ)^2/2! + (tQ)^3/3! + \ldots ,$$

where I is the $N \times N$ identity matrix. Without loss of generality, we will assume that P is for one year, i.e. $t = 1$.)

Our computational starting point is given by the following theorem. To state it, let

$$S \;=\; \max\left\{(a-1)^2 + b^2\,;\; a + bi \text{ is an eigenvalue of } P,\; a, b \in \mathbf{R}\right\} .$$

That is, S is computed by examining all of the (possibly complex) eigenvalues of P, of the form $a + bi$ where a and b are real, and computing the absolute-square of the eigenvalue minus 1, i.e. $(a-1)^2 + b^2$, and then taking the maximum of these absolute-squares over all of the eigenvalues of P.

Theorem 1. Let P be an $N \times N$ Markov transition matrix, and suppose that $S < 1$. Then the series

$$\widetilde{Q} \;=\; (P - I) - (P - I)^2/2 + (P - I)^3/3 - (P - I)^4/4 + \ldots \qquad (1)$$

converges geometrically quickly, and gives rise to an $N \times N$ matrix $\widetilde{Q}$ having row-sums 0, such that $\exp(\widetilde{Q}) = P$ exactly.

This theorem is proved in the Appendix (as are all the other theorems). The series (1) has been considered in this context by others. Indeed, a proof of Theorem 1 can be found in Zahl (1955, p. 96), using a result of Wedderburn (1934, pp. 122–123), and is also discussed by Singer and Spilerman (1976, p. 8). It provides a simple method (in terms of summing a series of matrices) for obtaining a matrix $\widetilde{Q}$ that automatically has most of the desired properties of a generator.

Remark. We wish to emphasise that, even if the series (1) fails to converge, or converges to a matrix $\widetilde{Q}$ which is not a valid generator, this does not preclude the possibility that a valid generator for P still exists; see Section 5.

We note that in practice it is not too important to check the condition $S < 1$. Indeed, as long as the series (1) converges absolutely, the conclusions that $\widetilde{Q}$ has row-sums 0 and that $\exp(\widetilde{Q}) = P$ are automatically satisfied, so that the condition $S < 1$ is no longer necessary. Furthermore, for many transition matrices arising in the credit risk literature, the condition that $S < 1$ is satisfied automatically via the following result.

Theorem 2. Suppose the diagonal entries of a transition matrix P are all greater than $\frac{1}{2}$ (i.e., $p_{ii} > 0.5$ for all i). Then $S < 1$, i.e. the convergence of the series (1) is guaranteed.

Remark. In matrix language, the condition of Theorem 2 is equivalent to the transition matrix P being *strictly diagonally dominant* (see e.g. Horn and Johnson (1985), p. 302). Furthermore, it is proved by Cuthbert (1972, 1973) that, under this condition, P can have at most one generator, i.e. if a generator exists then it is unique.

We emphasise that the condition is only a *sufficient* one, in that the series (1) may well converge, and we may well have $S < 1$, even if some of the diagonal entries of P are less than 0.5.

Remark. Elfving (1937) is the first to pose the problem of identifying test criteria on the entries of a transition matrix P so that there is a valid generator Q with $\exp(Q) = P$. This problem became known as the embedding problem.

Remark. In the case where P is diagonalisable with all eigenvalues real and positive, summing the series (1) (if it converges) is equivalent to first diagonalising P, then replacing the diagonal entries by their logarithms, and then converting back to the original basis. This is analogous to the approach of finding the n^{th} root of P by diagonalising and then replacing diagonal entries by their n^{th} roots (see e.g. Chapter 11 of Press et al., 1988). However, we note that this method does not guarantee that the resulting n^{th} root matrix will have non-negative entries. Moreover, it is not clear as to which root to choose when more than one real root exists.

3. The Non-Negativity Condition.

The main drawback of Theorem 1 is that the matrix $\widetilde{Q}$ is not guaranteed to have non-negative off-diagonal entries. (Several examples from finance are given in Section 4 below; for an example from sociology see Singer and Spilerman, 1976.) This is problematic since if $\widetilde{Q}$ has a negative off-diagonal entry, then so will $P_t = \exp(t\widetilde{Q})$ for sufficiently small $t > 0$. This means that P_t will not be a proper Markov transition matrix, which is unacceptable.

However, any negative off-diagonal entries of $\widetilde{Q}$ will usually be quite small. Therefore, it is possible to correct the problem simply by replacing these negative entries with 0, and adding the appropriate value back into the corresponding diagonal entry to preserve the property of having row-sums 0. That is, once we have obtained $\widetilde{Q}$, we can obtain a new matrix Q by setting

$$q_{ij} = \max(\widetilde{q}_{ij}, 0)\,, \quad j \neq i\,; \qquad\qquad q_{ii} = \widetilde{q}_{ii} + \sum_{j \neq i} \min(\widetilde{q}_{ij}, 0)\,. \tag{2}$$

(This modification technique is also considered in Zahl, 1955.) The new matrix Q will still have row-sums 0, and will have non-negative off-diagonal entries. However, it will no longer satisfy $\exp(Q) = P$ exactly.

A different version of Q can be obtained by adding the negative values back into *all* the entries of the same row (not just the diagonal one) which have the correct sign, proportional to their absolute values. That is, we could instead let

$$G_i \;=\; |\widetilde{q}_{ii}| \;+\; \sum_{j \neq i} \max(\widetilde{q}_{ij},\, 0)\,; \qquad\qquad B_i \;=\; \sum_{j \neq i} \max(-\widetilde{q}_{ij},\, 0)$$

be the "good" and "bad" totals for each row i, and then set

$$q_{ij} \;=\; \begin{cases} 0, & i \neq j \text{ and } \widetilde{q}_{ij} < 0 \\ \widetilde{q}_{ij} - B_i |\widetilde{q}_{ij}| / G_i & \text{otherwise if } G_i > 0 \\ \widetilde{q}_{ij}, & \text{otherwise if } G_i = 0 \end{cases} \tag{2$'$}$$

(Note that since $\sum_j \widetilde{q}_{ij} = 0$, we always have $G_i \geq B_i$; hence, (2$'$) guarantees that $q_{ij} \geq 0$ for $i \neq j$.) When $\widetilde{Q}$ has only slightly negative off-diagonal elements, and the values $-q_{ii}$ are reasonably large, (2) and (2$'$) will usually be fairly similar.

Remark. It may be possible to improve still further the choice of where to add the extra back in for the modification Q, by optimising this choice as a multivariate function. However, it appears that this further improvement would rarely make a substantial difference to the distance of $\exp(Q)$ to P, so we do not pursue it here.

Is it possible that a valid generator still exists even if the $\widetilde{Q}$ computed by Theorem 1 is not a valid one? The answer is yes. Furthermore, it may be possible that there exist more than one valid generators. for a given matrix P. Such issues and related examples are discussed in Section 5.

Is it possible to conclude the non-existence of a generator for a given transition matrix P under certain conditions? The answer is again yes. For example, we have the following theorem. (Part (a) below is also stated in Kingman, 1962; part (b) below is proved as Theorem 6.1 in Goodman, 1970; and part (c) below follows from the standard *Lévy Dichotomy*, see e.g. Chung, 1992 and Grimmett and Stirzaker, 1992.) To state it, recall that a state j is *accessible* from a state i if there is a sequence of states $k_0 = i, k_1, k_2, \ldots, k_m = j$ such that $p_{k_\ell \, k_{\ell+1}} > 0$ for each ℓ.

Theorem 3. Let P be a transition matrix, and suppose that either

(a) $\det(P) \leq 0$; or

(b) $\det(P) > \prod_i p_{ii}$; or

(c) there are states i and j such that j is accessible from i, but $p_{ij} = 0$.

Then there does not exist an exact generator for P.

Remark. We note that, if $p_{ii} > 0.5$ for all i as in Theorem 2, then we necessarily have $\det(P) > 0$, so that Theorem 3(a) never applies. To see this, we can write $P = mI + (1-m)R$ for a stochastic matrix R, where $m = \min_i p_{ii} > 0.5$. Hence, setting $A_s = sP + (1-s)I$, we have $A_s = (1 - s + sm)I + s(1-m)R$, so that A_s is invertible for all $0 \leq s \leq 1$ (since $1 - s + sm > s(1-m)$). But then $\det(A_s)$ is a real continuous non-zero function of $s \in [0,1]$, which is positive at $s = 0$, hence it is also positive at $s = 1$.

Some further results about the existence and uniqueness of generators are presented in Section 5. We first pause to consider some numerical examples.

4. Some Examples from Finance.

In finance, the finite state space $\Omega = \{1, 2, \ldots, N\}$ covers possible bond ratings, with state 1 being the highest rating and state N being Default. Typically, $\Omega = \{$AAA, AA, A, BBB, BB, B, CCC, Default$\}$. Insofar as a lower rating presents a higher credit risk, a rating transition matrix should satisfy one of the two conditions as outlined by Jarrow et al. (1997):

Lemma 1. Let Q be a valid generator for P. Then the following two conditions are equivalent:

(a) $\sum_{j \geq k} p_{ij}$ is a nondecreasing function of i for every fixed k.

(b) $\sum_{j \geq k} q_{ij} \geq \sum_{j \geq k} q_{i+1,j}$ for all i and k such that $k \neq i + 1$.

Proof. See Jarrow et al. (1997, p. 495). ∎

Remark. In the language of Markov chains, the above conditions are equivalent to requiring that the underlying Markov chain be "stochastically monotonic".

Let's first examine the annual transition matrix considered by Jarrow et al. (1997), which is based on empirical observations from Standard and Poor's Credit Review (1993), for the years 1981 through 1991. After distributing the "Not Rated" weights to the other entries via their equation (31), they obtain the following average transition matrix (their Table 3):

$$P = \begin{pmatrix} 0.8910 & 0.0963 & 0.0078 & 0.0019 & 0.0030 & 0.0000 & 0.0000 & 0.0000 \\ 0.0086 & 0.9010 & 0.0747 & 0.0099 & 0.0029 & 0.0029 & 0.0000 & 0.0000 \\ 0.0009 & 0.0291 & 0.8894 & 0.0649 & 0.0101 & 0.0045 & 0.0000 & 0.0009 \\ 0.0006 & 0.0043 & 0.0656 & 0.8427 & 0.0644 & 0.0160 & 0.0018 & 0.0045 \\ 0.0004 & 0.0022 & 0.0079 & 0.0719 & 0.7764 & 0.1043 & 0.0127 & 0.0241 \\ 0.0000 & 0.0019 & 0.0031 & 0.0066 & 0.0517 & 0.8246 & 0.0435 & 0.0685 \\ 0.0000 & 0.0000 & 0.0116 & 0.0116 & 0.0203 & 0.0754 & 0.6493 & 0.2319 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{pmatrix}.$$

Here the eight columns and rows represent, in order, the credit ratings AAA, AA, A, BBB, BB, B, CCC, and Default.

Remark. In this paper, we are implicitly assuming (as have Jarrow et al., 1997, and other authors) that the credit ratings process is time-homogeneous and Markovian. In reality, both assumptions are open to debate, and likely neither one is precisely true for real credit ratings. In fact, the empirical question of whether time-homogeneous Markov model is suitable for bond rating transitions has yet to be addressed in the literature. We do not consider such issues here, but rather implicitly assume a time-homogeneous Markov model throughout.

It can be shown that condition (a) in Lemma 1 is violated for this transition matrix. For example, when $k = 4$ and $i = 6$, we compute that $\sum_{j \geq 4} p_{6j} = 0.9949$ while $\sum_{j \geq 4} p_{7j} = 0.9885$. It turns out that this condition (a) is also violated in the other matrices that we will examine in this section. Since whether a transition matrix is stochastically monotonic has very little bearing on the existence of a generator, and since any "fixing" would be necessarily guided by arbitrary, *ad hoc* rules, we will leave the matrices as they are. In cases where the monotone condition is crucial (as in e.g. estimation of risk premiums), researchers may make some adjustments of the transition matrix based on their best judgements.

By Theorem 3(c), the transition matrix P does not have an exact generator. For example, $p_{12} > 0$ and $p_{26} > 0$, but $p_{16} = 0$. Thus, it will be useful to find an approximate generator Q, such that $\exp(Q)$ is approximately equal to P, and that a transition matrix for any time t can be approximated by $\exp(Qt)$.

Remark. It may be considered disappointing that this transition matrix P does not have an exact generator. It is possible that an empirical matrix P estimated from many more years of data would have no zero entries and would in fact be embeddable, though we lack sufficient data to test this. In the absence of such data, it may be tempting to use the "trick" of simply raising the given matrix P to a high power, as a stand-in for an empirical transition matrix for many more years. However, this trick does not help with the embeddability problem. Indeed, clearly if P is not embeddable then neither is any power of P (and moreover, if P is embeddable, then any power of P has the same set of valid generators as P does).

Jarrow et al. obtain an approximate generator Q_{JLT} for this P by assuming that there is never more than one transition per year. Their method leads to the general formula

$$q_{ii} = \log(p_{ii}); \qquad q_{ij} = p_{ij} \log(p_{ii})/(p_{ii} - 1) \qquad (i \neq j). \tag{3}$$

To continue the above example, their approximate generator is

$Q_{JLT} =$

$$\begin{pmatrix} -0.1154 & 0.1019 & 0.0083 & 0.0020 & 0.0031 & 0.0000 & 0.0000 & 0.0000 \\ 0.0091 & -0.1043 & 0.0787 & 0.0105 & 0.0030 & 0.0030 & 0.0000 & 0.0000 \\ 0.0010 & 0.0309 & -0.1172 & 0.0688 & 0.0107 & 0.0048 & 0.0000 & 0.0010 \\ 0.0007 & 0.0047 & 0.0713 & -0.1711 & 0.0701 & 0.0174 & 0.0020 & 0.0049 \\ 0.0005 & 0.0025 & 0.0089 & 0.0813 & -0.2530 & 0.1181 & 0.0144 & 0.0273 \\ 0.0000 & 0.0021 & 0.0034 & 0.0073 & 0.0568 & -0.1929 & 0.0479 & 0.0753 \\ 0.0000 & 0.0000 & 0.0142 & 0.0142 & 0.0250 & 0.0928 & -0.4318 & 0.2856 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.000 & 0.0000 \end{pmatrix}.$$

This approximate generator Q_{JLT} does indeed have row-sums 0 and non-negative off-diagonal entries. Furthermore we calculate that

$$\exp(Q_{JLT}) =$$

$$\begin{pmatrix} 0.891431 & 0.091476 & 0.011075 & 0.002650 & 0.002847 & 0.000339 & 0.000025 & 0.000062 \\ 0.008198 & 0.902501 & 0.070932 & 0.011685 & 0.003332 & 0.003028 & 0.000093 & 0.000231 \\ 0.001042 & 0.027907 & 0.892718 & 0.060232 & 0.011116 & 0.005288 & 0.000231 & 0.001465 \\ 0.000676 & 0.005202 & 0.062348 & 0.847319 & 0.057651 & 0.018185 & 0.002246 & 0.006373 \\ 0.000457 & 0.002533 & 0.010214 & 0.066701 & 0.781647 & 0.095894 & 0.012458 & 0.030092 \\ 0.000025 & 0.001947 & 0.003751 & 0.008390 & 0.046273 & 0.829090 & 0.035518 & 0.074915 \\ 0.000016 & 0.000317 & 0.011507 & 0.012042 & 0.020256 & 0.069455 & 0.651072 & 0.235331 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 1.000000 \end{pmatrix}.$$

We thus see that $\exp(Q_{JLT})$ is close to, but not exactly equal to, the original P.

Let us now instead use the series method presented in Theorem 1. By Theorem 2, we automatically have $S < 1$ in this case, so that Theorem 1 does apply. Indeed, we find that the series (1) for $\widetilde{Q}$ indeed converges very quickly. In fact, summing just the first 16 terms of the series, we compute that the effect of subsequent terms is less than 10^{-8}, thus giving extremely high accuracy. Retaining six digits, we obtain that

$$\widetilde{Q} =$$

$$\begin{pmatrix} -0.115931 & 0.107466 & 0.004208 & 0.001334 & 0.003372 & -0.000409 & -0.000014 & -0.000025 \\ 0.009566 & -0.106131 & 0.083233 & 0.008115 & 0.002567 & 0.002942 & -0.000114 & -0.000168 \\ 0.000831 & 0.032362 & -0.121382 & 0.074626 & 0.009028 & 0.004013 & -0.000274 & 0.000589 \\ 0.000623 & 0.003572 & 0.075556 & -0.177517 & 0.079050 & 0.013991 & 0.001350 & 0.003277 \\ 0.000440 & 0.002181 & 0.005768 & 0.088535 & -0.261178 & 0.129535 & 0.013816 & 0.020800 \\ -0.000027 & 0.002086 & 0.002711 & 0.004655 & 0.063962 & -0.199781 & 0.059016 & 0.067268 \\ -0.000015 & -0.000420 & 0.014447 & 0.013639 & 0.024547 & 0.101298 & -0.435344 & 0.281980 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \end{pmatrix}.$$

We then compute that

$$\exp(\widetilde{Q}) = P$$

exactly, to at least six decimal places (i.e., to within 10^{-6}). Furthermore, our $\widetilde{Q}$ has row-sums 0, as it must.

Unfortunately, our $\widetilde{Q}$ has some (small) negative off-diagonal entries. We thus replace $\widetilde{Q}$ by Q as in equation (2), to obtain

$$Q = \begin{pmatrix} -0.116380 & 0.107466 & 0.004208 & 0.001334 & 0.003372 & 0.000000 & 0.000000 & 0.000000 \\ 0.009566 & -0.106414 & 0.083233 & 0.008115 & 0.002567 & 0.002942 & 0.000000 & 0.000000 \\ 0.000831 & 0.032362 & -0.121656 & 0.074626 & 0.009028 & 0.004013 & 0.000000 & 0.000589 \\ 0.000623 & 0.003572 & 0.075556 & -0.177517 & 0.079050 & 0.013991 & 0.001350 & 0.003277 \\ 0.000440 & 0.002181 & 0.005768 & 0.088535 & -0.261178 & 0.129535 & 0.013816 & 0.020800 \\ 0.000000 & 0.002086 & 0.002711 & 0.004655 & 0.063962 & -0.199808 & 0.059016 & 0.067268 \\ 0.000000 & 0.000000 & 0.014447 & 0.013639 & 0.024547 & 0.101298 & -0.435779 & 0.281980 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \end{pmatrix}.$$

This Q now has non-negative off-diagonal entries and still has row-sums 0. However, it no longer exactly satisfies that $\exp(Q) = P$. In fact,

$$\exp(Q) = \begin{pmatrix} 0.890600 & 0.096265 & 0.007798 & 0.001901 & 0.003010 & 0.000351 & 0.000026 & 0.000048 \\ 0.008597 & 0.900745 & 0.074680 & 0.009899 & 0.002901 & 0.002906 & 0.000097 & 0.000175 \\ 0.000900 & 0.029092 & 0.889158 & 0.064893 & 0.010102 & 0.004511 & 0.000211 & 0.000935 \\ 0.000600 & 0.004300 & 0.065591 & 0.842700 & 0.064400 & 0.016000 & 0.001808 & 0.004501 \\ 0.000401 & 0.002202 & 0.007899 & 0.071900 & 0.776400 & 0.104299 & 0.012698 & 0.024100 \\ 0.000024 & 0.001911 & 0.003100 & 0.006600 & 0.051699 & 0.824577 & 0.043491 & 0.068498 \\ 0.000014 & 0.000323 & 0.011610 & 0.011599 & 0.020296 & 0.075384 & 0.649019 & 0.231854 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 1.000000 \end{pmatrix},$$

which is very close to, but not exactly equal to, the original transition matrix P.

It is useful to know which approximation is more accurate. One method of comparison is to compute the L^1-norm (i.e., the sum of the absolute values of the 64 matrix entries) of $P - \exp(Q_{JLT})$ and of $P - \exp(Q)$, to see which is smaller. We compute that

$$\text{norm}[P - \exp(Q_{JLT})] = 0.116900\,; \qquad \text{norm}[P - \exp(Q)] = 0.002736\,.$$

We thus see that, in terms of closeness of $\exp(Q)$ to P, the choice of Q is better than Q_{JLT} by a factor of $0.116900 \,/\, 0.002736 \approx 42.7$.

In fact, if we instead use the alternative approximation ($2'$), we then compute that

$$\text{norm}[P - \exp(Q)] = 0.002686\,,$$

which is a slight additional improvement.

We thus tentatively conclude that, while our method requires the awkward transformation of $\widetilde{Q}$ to Q, it can be substantially better than the method of Jarrow et al. whereby we must assume away multiple transitions in a single year.

Remark. As an aside, we note that the row entries of this example matrix P do not add exactly to 1, presumably due to round-off errors. As a check, we slightly modified the diagonal entries of P so that the row entries sum to exactly 1; but this had negligible effect on any of the resulting matrices or distances.

We now consider some additional examples.

We first consider the average one-year transition matrix from Moody's which covers the period of 1980–1998 (Moody's, 1999). After re-assigning the "Not Rated" weights to the other entries, we obtain the following transition matrix:

$$P = \begin{pmatrix} 0.886583 & 0.102937 & 0.010169 & 0.000000 & 0.000311 & 0.000000 & 0.000000 & 0.000000 \\ 0.010787 & 0.887045 & 0.095530 & 0.003423 & 0.001452 & 0.001452 & 0.000000 & 0.000311 \\ 0.000625 & 0.028759 & 0.902053 & 0.059185 & 0.007398 & 0.001771 & 0.000104 & 0.000104 \\ 0.000529 & 0.003386 & 0.070680 & 0.852291 & 0.060523 & 0.010052 & 0.000846 & 0.001587 \\ 0.000328 & 0.000765 & 0.005571 & 0.056806 & 0.835809 & 0.080839 & 0.005353 & 0.014638 \\ 0.000109 & 0.000435 & 0.001738 & 0.006519 & 0.065950 & 0.827032 & 0.027597 & 0.070621 \\ 0.000000 & 0.000000 & 0.006600 & 0.010500 & 0.030500 & 0.061100 & 0.629700 & 0.261600 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 1.000000 \end{pmatrix}$$

We again see from Theorem 3(c) that there does not exist an exact generator for this transition matrix. Hence, we search for an approximate generator.

Using the method of Jarrow et al. in equation (3) to compute the approximate generator Q_{JLT}, we can then calculate the L^1-norm as

$$\text{norm}[P - \exp(Q_{JLT})] = 0.100047 \, .$$

Using the series method (1), together with the modification (2), yields a different approximate generator Q, which gives an L^1 distance of $\exp(Q)$ to P of

$$\text{norm}[P - \exp(Q)] = 0.001401 \, .$$

Once again, using the modification (2′) instead gives a slightly better approximation, with

$$\text{norm}[P - \exp(Q)] = 0.001371 \, .$$

Compared with Q_{JLT}, the generators obtained above represent an improvement on the order of 71.4 and 73.0 respectively.

As another example, we consider the transition matrix from Standard and Poor's which covers a longer period than the one considered by Jarrow et al. (see Standard and Poor's, 1999). After re-assigning the "Not

Rated" weights, we obtain the following annual transition matrix:

$$P = \begin{pmatrix} 0.919347 & 0.074592 & 0.004829 & 0.000822 & 0.000411 & 0.000000 & 0.000000 & 0.000000 \\ 0.006396 & 0.918085 & 0.067575 & 0.005984 & 0.000619 & 0.001135 & 0.000310 & 0.000000 \\ 0.000730 & 0.022725 & 0.916814 & 0.051183 & 0.005629 & 0.002502 & 0.000104 & 0.000417 \\ 0.000425 & 0.002658 & 0.055609 & 0.878894 & 0.048272 & 0.010207 & 0.001701 & 0.002339 \\ 0.000441 & 0.000991 & 0.006058 & 0.077542 & 0.814847 & 0.078973 & 0.011125 & 0.010133 \\ 0.000000 & 0.001019 & 0.002830 & 0.004641 & 0.069496 & 0.827957 & 0.039615 & 0.054556 \\ 0.001860 & 0.000000 & 0.003720 & 0.007439 & 0.024294 & 0.121237 & 0.604557 & 0.237010 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 1.000000 \end{pmatrix}$$

Once again, we see by Theorem 3(c) that an exact generator for this P does not exist.

Following the same procedures as above, we compute the L^1-norm for the Jarrow et al. method and the series method (based on (2) and (2′)) as

$$\text{norm}[P - \exp(Q_{JLT})] = 0.103516\,; \qquad \text{norm}[P - \exp(Q)] = 0.001096 \text{ (based on (2))}\,;$$

$$\text{and} \qquad \text{norm}[P - \exp(Q)] = 0.001088 \text{ (based on (2′))}\,.$$

It can be seen that the improvement factor is over 90 for both adjustment methods.

Remark. It is possible that our method has applications in other fields as well. For an example of Alzheimer's Disease therapy evaluation, see Stewart et al. (1998).

5. Further Results on the Existence and Uniqueness of Generators.

According to Kingman (1962), the embeddability problem is completely solved for the case of 2×2 matrices by D.G. Kendall, who proves that a 2×2 transition matrix is compatible with a continuous Markov process if and only if the sum of the two diagonal entries is larger than 1. For matrices with higher dimensions, it is no longer clear cut, as we have seen in Sections 2 and 3. In this section, we add some additional results applicable to higher-dimensional transition matrices. Again, many of our results have appeared previously in other papers (especially Singer and Spilerman, 1976), as we shall indicate.

One might hope that, if the series (1) converges to a matrix $\widetilde{Q}$ which does *not* have all its off-diagonal entries non-negative, then perhaps this proves that no generator for P can possibly exist. Unfortunately, this is not true in general, as the following proposition shows. Intuitively, the series (1) only represents the principal branch of $\log(P)$; other branches of $\log(P)$ may still provide valid generators.

Proposition 1. For a given stochastic matrix P, the fact that a real matrix Q_1 with row-sums 0 and some negative off-diagonal entries satisfies $\exp(Q_1) = P$ does not preclude the possibility that there is a valid generator Q_2 for P. This is true even if Q_1 is equal to the matrix $\widetilde{Q}$ obtained from the series (1).

Proof. We prove the proposition by citing an example. Let

$$P = \begin{pmatrix} .284779445 & .284035268 & .283826586 & .147358701 \\ .284191780 & .284779445 & .284035268 & .146993507 \\ .283487477 & .284191780 & .284779445 & .147541298 \\ .284543931 & .283487477 & .284191780 & .147776812 \end{pmatrix}$$

Then the eigenvalues of P are 1, 0.0010572 (with multiplicity two), and 7.46905×10^{-7}. It follows that $S = (1 - 7.46905 \times 10^{-7})^2$, so that $S < 1$.

We then compute the series (1), which converges very slowly since S is very close to 1, and obtain

$$Q_1 \equiv \widetilde{Q} = \begin{pmatrix} -6.194496074 & 2.807322994 & 1.374197570 & 2.012975514 \\ 3.882167062 & -6.194496074 & 2.807322994 & -.494993979 \\ -.954631245 & 3.882167062 & -6.194496074 & 3.266960260 \\ 6.300566216 & -.954631245 & 3.882167062 & -9.228102034 \end{pmatrix}$$

a matrix with some negative off-diagonal entries. (Of course, we still have $\exp(Q_1) = P$, as we must.)

However, the matrix

$$Q_2 = \begin{pmatrix} -5.642931358 & 5.642931358 & 0.000000000 & 0.000000000 \\ 0.000000000 & -5.642931358 & 5.642931358 & 0.000000000 \\ 0.000000000 & 0.000000000 & -5.642931358 & 5.642931358 \\ 61.410871840 & 0.000000000 & 0.000000000 & -61.410871840 \end{pmatrix}$$

is indeed a valid generator for P, since $\exp(Q_2) = P$ and Q_2 has row-sums 0 and positive off-diagonal entries. ∎

This proposition shows that even if the $\widetilde{Q}$ from Theorem 1 does not satisfy the non-negativity condition, a valid generator for P may still exist. (Similarly, an example where the series (1) fails to converge, but where a valid generator still exists, is given by Singer and Spilerman, 1976, Example 4.)

On the other hand, even if the existence of a generator for P is established, the generator may not be *unique*, as the following example shows. (A similar example is given in Singer and Spilerman, 1976, Example 12.)

Proposition 2. There exist transition matrices P which have more than one valid generators.

Proof. Let $b = e^{-4\pi}$, and let

$$P = \begin{pmatrix} (2+3b)/5 & (2-2b)/5 & (1-b)/5 \\ (2-2b)/5 & (2+3b)/5 & (1-b)/5 \\ (2-2b)/5 & (2-2b)/5 & (1+4b)/5 \end{pmatrix}$$

$$Q_1 = \begin{pmatrix} -2\pi & 2\pi & 0 \\ 0 & -2\pi & 2\pi \\ 4\pi & 0 & -4\pi \end{pmatrix}$$

and

$$Q_2 = \begin{pmatrix} -12\pi/5 & 8\pi/5 & 4\pi/5 \\ 8\pi/5 & -12\pi/5 & 4\pi/5 \\ 8\pi/5 & 8\pi/5 & -16\pi/5 \end{pmatrix}$$

It can be verified that Q_1 and Q_2 both qualify as generators, and $\exp(Q_1) = \exp(Q_2) = P$. Thus, Q_1 and Q_2 are both valid generators for P. ∎

Remark. This last result raises this question: when a transition matrix P has multiple valid generators Q, which generator is the "best"? This question could be important, since different generators lead to different values of $P_t = \exp(tQ)$ for most t (e.g. for all sufficiently small t). Mathematically, the notion of "desirability" of a generator is not well defined, since any valid generator is as good as any other. However, in practice, it is indeed meaningful to ask which valid generator will represent the empirical transition matrices the best. To shed some light on this issue, realise that with most empirical transition matrices (such as the ones presented in Section 4), it is generally unlikely for a rating to migrate to a remote rating in a short time period. Thus, one method of choosing among valid generators is to take the one with the smallest value of

$$J \;=\; \sum_{i,j} |j-i|\,|q_{ij}|\,,$$

which ensures that the chance of jumping too far too quickly is minimised. In the example from the above proposition, we have

$$J_1 = 2\pi + 2\pi + 2(4\pi) = 12\pi \doteq 37.70\,;$$

while

$$J_2 = 8\pi/5 + 2(4\pi/5) + 8\pi/5 + 4\pi/5 + 2(8\pi/5) + 8\pi/5 = 52\pi/5 \doteq 32.67\,;$$

we therefore might conclude that Q_2 is a better choice than Q_1. For discussions of somewhat related issues in rather different contexts, see Cuthbert (1972) and Singer and Spilerman (1976, Section 4).

Despite the difficulties raised in Proposition 2, it is sometimes possible to prove the uniqueness of generators for P, as the following result shows. (Part (c) below also appears in Cuthbert, 1972, 1973.) To state it, we will use $\mathrm{Log}(P)$ to denote the principal branch of the logarithm of P. This will be equal to the sum of the series (1) whenever the series converges. It always satisfies $\exp(\mathrm{Log}(P)) = P$, and if P has row-sums 1 then $\mathrm{Log}(P)$ has row-sums 0.

Theorem 4. Let P be a transition matrix.

(a) If $\det(P) > 1/2$, then P has at most one generator.

(b) If $\det(P) > 1/2$ and $\|P - I\| < 1/2$ (using any operator norm), then the only possible generator for P is $\mathrm{Log}(P)$.

(c) If P has distinct eigenvalues and $\det(P) > e^{-\pi}$, then the only possible generator for P is $\mathrm{Log}(P)$.

In addition, we have the following result, which is also observed by Singer and Spilerman (1976, pp. 29–30).

Theorem 5. Let P be a transition matrix which has real distinct eigenvalues.

(a) If all eigenvalues of P are positive, then $\mathrm{Log}(P)$ is the only real matrix Q such that $\exp(Q) = P$.

(b) If P has any negative eigenvalues, then there is no real matrix Q such that $\exp(Q) = P$.

The last two theorems provide some further validity to our matrix $\widetilde{Q}$ from Theorem 1. In particular, they immediately show the following.

Corollary. Let P be a transition matrix such that at least one of the following three conditions hold:

(i) $\det(P) > \frac{1}{2}$ and $\|P - I\| < \frac{1}{2}$; or

(ii) P has distinct eigenvalues and $\det(P) > e^{-\pi}$; or

(iii) P has distinct real eigenvalues.

Suppose further that the series (1) converges to a matrix $\widetilde{Q}$ with negative off-diagonal entries. Then there does not exist a valid generator for P.

There are other known conditions on the eigenvalues of P which either preclude or establish the possibility of a valid generator Q. For example, Elfving (1937) proves that if P has a (complex) eigenvalue other than 1 of absolute value 1, or if P has a negative (real) eigenvalue of odd multiplicity, then no valid generator exists for P. A more refined result is obtained by Runnenberg (1962), who proves that if an $N \times N$ matrix P has a valid generator, then each (complex) eigenvalue of P must lie within the region of the complex plane having boundary curve

$$\Big\{ \Big(\exp\left(-s + s\cos(2\pi/N)\right) \, \cos\left(s\sin(2\pi/N)\right) , \; \exp\left(-s + s\cos(2\pi/N)\right) \, \sin\left(s\sin(2\pi/N)\right) \Big) ; \quad 0 \le s \le \frac{\pi}{\sin(2\pi/N)} \Big\}$$

together with its reflection about the real axis. Each of these conditions is a further necessary condition for P to have a valid generator, in the spirit of Theorem 3 or Theorem 5(b).

Finally, we note that it is possible to provide a more quantitative version of Theorem 3(c) (the Lévy Dichotomy), as follows.

Theorem 6. If P does have a valid generator, then its entries must satisfy that

$$p_{ik} \ge m^m r^r (m+r)^{-m-r} \sum_j (p_{ij} - b_m)(p_{jk} - b_r) \; \mathbf{1}_{p_{ij} > b_m, \; p_{jk} > b_r} \, ,$$

for any positive integers m and r. Here $b_m = \sum_{\ell=m+1}^{\infty} e^{-\lambda} \lambda^{\ell} / \ell! = 1 - \sum_{\ell=0}^{m} e^{-\lambda} \lambda^{\ell} / \ell!$ which equals the probability that $N' > m$, where N' is a Poisson random variable with mean $\lambda \equiv \max_i(-Q_{ii})$. Furthermore $\mathbf{1}_B$ is the indicator function of the boolean event B.

For example, for the first matrix in Section 4, we have from equation $(A1)$ that

$$\lambda \equiv \max_i(-Q_{ii}) \le -\mathrm{trace}(Q) = -\log\det(P) \doteq 1.417263 \, , \tag{4}$$

and also that, for example, $p_{7,3} = 0.0116$ and $p_{3,2} = 0.0291$. Choosing $r = 5$ and $m = 4$, we compute that $b_5 \le 0.00339622$ and $b_4 \le 0.0149457$. The theorem then tells us that, for a valid generator to exist, we would require $p_{7,2} \ge 4^4 5^5 9^{-9} (p_{7,3} - b_4)(p_{3,2} - b_5) \ge 2.39779 \times 10^{-7}$. (This bound could perhaps be tightened somewhat by choosing different values of r and m, and summing over all possible intermediate states j, rather than just $j = 3$.) We thus see that this quantitative bound is not sufficiently large to rule out, say, the possibility that $P_{7,2} = 10^{-5}$ but only appears to be 0 because of rounding off.

Remark. An improved bound could be obtained if we had a way of bounding $\max_i(-Q_{ii})$ better than the one presented in (4).

Remark. We note that, even if a generator for P does not exist, it is still possible that, for example, there is a valid transition matrix $P_{1/2}$ such that $\left(P_{1/2}\right)^2 = P$. On the other hand, it is shown by Kingman (1962, Proposition 7) that if P is non-singular, and if for *all* positive integers n there is a matrix $P_{1/n}$ such that $\left(P_{1/n}\right)^n = P$, then there *does* exist a generator for P.

6. An Algorithm to Search for a Valid Generator.

If the series (1) fails to converge, or converges to a matrix which has some negative off-diagonal terms, then the results of the previous section indicate that it is still possible that a generator exists. In theory, it is possible to find this generator by checking all branches of the logarithm function and computing $\log(P)$ for each one, each time checking to see if the non-negativity condition is satisfied.

If P has distinct eigenvalues, then this search is finite, as the following result shows.

Theorem 7. If P has distinct eigenvalues, and if Q is a generator for P, then each eigenvalue λ of Q satisfies that $|\mathcal{I}m\ \lambda| \leq |\ln(\det(P))|$. In particular, there are only a finite number of possible branch values of $\log(P)$ which could possibly be generators of P. (Note that, by Theorem 3(a), if $\det(P) \leq 0$ then no generator exists.)

In light of this Theorem, whenever P has distinct eigenvalues, it is possible to construct a finite algorithm to search for all possible generators of P, as follows.[1] (Ideas which are very similar in spirit to ours, but somewhat different in precise detail especially regarding the eigenvalue ranges considered, are discussed by Singer and Spilerman, 1976, Section 3.3a.)

The idea is to use Lagrange interpolation. If P is $n \times n$ with distinct eigenvalues $r_1, r_2, \ldots, r_n$, and f is any function that is analytic in a neighbourhood of each eigenvalue, then $f(P) = g(P)$ where g is the

[1] We have written a Maple program to carry out this search, available on the web at `http://www.math.ubc.ca/~israel/irw/`

polynomial of degree $n - 1$ such that $g(r_j) = f(r_j)$ for each j. The Lagrange interpolation formula (or, Sylvester's formula, cf. Singer and Spilerman, 1976, p. 18) says that

$$g(x) = \sum_j \prod_{k \neq j} \frac{x - r_k}{r_j - r_k} f(r_j); \tag{5}$$

here the sum is over all eigenvalues r_j, and the product is over all eigenvalues r_k except r_j. (If P has repeated eigenvalues, this must be modified to require some derivatives to agree: $g^{(m)}(r_j) = f^{(m)}(r_j)$ for $m <$ (multiplicity of r_j).)

To search for generators, each of the unknown values $f(r_j)$ can be equal to any branch of the logarithm of r_j. In general r_j may be a complex number, and the values $f(r_j)$ must satisfy

$$f(r_j) = \text{Log}|r_j| + i\left(\text{Arg}(r_j) + 2\pi k_j\right),$$

for any integers k_j subject to the condition of Theorem 7. (Here $|r| = \sqrt{\mathcal{R}e(r)^2 + \mathcal{I}m(r)^2}$ and $\text{Arg}(r) = \arctan(\mathcal{I}m(r) / \mathcal{R}e(r))$.) The search would be carried out by using Lagrange interpolation to compute $f(P)$ for each possible choice of $k_1, k_2, \ldots, k_n$, and checking the non-negative off-diagonal condition for each possible $f(P)$ which arises.

We note that many symbolic calculation packages contain functions for computing such matrices $g(P)$, e.g. "`linalg/matfunc(P, g(z), z)`" in Maple, or "`(funm(P,'g'))`" in Matlab.

We summarise the above algorithm as follows:

Step 1. Compute the eigenvalues $r_1, r_2, \ldots, r_n$ of P, and verify that they are all distinct.

Step 2. For each eigenvalue r_j, choose an integer k_j such that $\lambda \equiv \text{Log}|r_j| + i(\text{Arg}(r_j) + 2\pi k_j)$ satisfies the restriction of Theorem 7, i.e. such that $|\mathcal{I}m\ \lambda| = |\text{Arg}(r_j) + 2\pi k_j| \leq |\ln(\det(P))|$.

Step 3. For the collection of integers $k_1, k_2, \ldots, k_n$, set $f(r_j) = \text{Log}|r_j| + i(\text{Arg}(r_j) + 2\pi k_j)$, and let $g(x)$ be the function given in (5).

Step 4. For this function $g(x)$, compute the matrix $Q = g(P)$, and see if it is a valid generator.

Step 5. Return to Step 2, modifying one or more of the integers k_j. Continue until all allowable collections $k_1, k_2, \ldots, k_n$ have been considered.

In practice, the credit rating transition matrices would most likely have distinct eigenvalues, and the above procedure is adequate. In the rare cases where repeated eigenvalues are present, the search would be much more involved. For some guidance in this case, please see e.g. Singer and Spilerman (1976), Section 3.3b.

7. Discussions.

As we have seen in Section 4, most of the empirical annual transition matrices are "sparse" in that the probabilities of remote migrations are mostly zero. As a result, almost all matrices would fail the test of Theorem 3(c), which renders a simple resolution: a true generator does not exist, i.e. the empirical transition matrix is not compatible with a continuous Markov process. In such a case, a finance researcher has two choices. One is to opt for the simplest way out: to compute the series in (1) (since by Theorem 2, the convergence is guaranteed for most matrices in practice) and make adjustments (when necessary) as outlined in Section 3. The other is to make certain adjustment to the transition matrix first so that there are no obvious violations of e.g. Lemma 1 and Theorem 3(c), and then proceed to search for a valid generator. If multiple generators are obtained via the procedures in Section 6, then the "quick and dirty" selection rule in Section 5 can be used to choose the "correct" generator. Based on our limited experiments in Section 3, it appears that the series method works well for all practical purposes.

It should be pointed out that our research focuses on the compatibility of an empirical transition matrix with a continuous Markov process. When the underlying process is discrete, the research problem becomes identifying the one-step transition matrix given a multi-step matrix, i.e., it becomes a problem of taking the root of a matrix. Unfortunately, as pointed out by Singer and Spilerman (1976), none of the difficulties encountered in the continuous time setting would go away. Just as a logarithm is multi-valued, so is a root operation. Therefore, switching the modeling framework to a discrete setting would not avoid the fundamental difficulties.

Lastly, as pointed out by Singer and Spilerman (1976) and readily intuitive to most researchers, if we could empirically estimate two transition matrices $P(t_1)$ and $P(t_2)$ such that $t_2 > t_1 > 0$ but t_2/t_1 is not rational, then the generator can be uniquely identified if it does exist. Alternatively, one could directly estimate the generator using transition data as outlined by Jarrow et al. (1997, p. 504). To make these approaches empirically feasible, we need time-stamped transitions for a universe of bonds. This type of data does appear to be available from Moody's or Standard and Poor's. Our future research endeavor will be to develop estimation procedures to directly extract a generator from such rating migration data.

8. Conclusions.

Credit risk modeling and credit derivatives pricing have become a mainstream research subject in finance. A key building block for this line of research is the rating transition matrix. Given that empirically estimated matrices are mostly for a one-year period, there is a need to recover a matrix generator so that one can obtain a transition matrix for any arbitrary period of time, as often dictated by a valuation problem such as pricing a default swap. This paper identifies conditions under which a true generator does or does not exist.

It is shown that most of the observed average annual transition matrices would not be compatible with a continuous Markov process, i.e., they do not admit a valid generator. Therefore a researcher is left with two options: either to obtain an approximate generator for the observed transition matrix, or to modify the transition matrix first (to make it compatible with an underlying Markov process) and then search for true generators. Our limited experiments point to the choice of the first option.

We have presented a collection of theorems and propositions to help determine if and when a generator exists. The results, by their nature of generality, will have applications beyond research in finance.

Appendix: Theorem Proofs.

In this Appendix, we provide proofs of all the theorems in the text (including, for completeness, those theorems which have been previously proved elsewhere as indicated).

We begin with some standard facts from the functional calculus, which we shall use (explicitly or implicitly) in many of our proofs.

Standard Results. Let A be any $n \times n$ matrix. Let $H(A)$ be the algebra of functions that are analytic in a neighbourhood of the set $\sigma(A)$ of eigenvalues of A. There is a mapping $f \mapsto f(A)$ from H to $n \times n$ complex matrices such that

(1) $f \mapsto f(A)$ is an algebra homomorphism.

(2) If $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ is a Taylor series converging in a neighbourhood of $\sigma(A)$, then $f(A) = \sum_{n=0}^{\infty} a_n (A - z_0 I)^n$. In particular, $f(A) = \exp(A)$ (as previously defined) for $f = \exp$.

(3) $\sigma(f(A)) = f(\sigma(A))$.

(4) $f(A)$ commutes with every $n \times n$ matrix that commutes with A.

(5) $f(A) = g(A)$ if $f - g$ is divisible (as a member of $H(A)$) by the characteristic polynomial of A.

(6) If $g \in H(f(A))$ then $g(f(A)) = (g \circ f)(A)$.

(7) If A has distinct eigenvalues then $f(A) = g(A)$ if and only if $f(r) = g(r)$ for each eigenvalue r of A.

(8) If A has distinct eigenvalues then every $n \times n$ matrix that commutes with A is $f(A)$ for some f.

(9) If A is a real matrix and $f(\bar{z}) = \overline{f(z)}$ in a neighbourhood of $\sigma(A)$, then $f(A)$ is real.

(10) If $\mathbf{v}$ is an eigenvector of A for eigenvalue r, then $f(A)\mathbf{v} = f(r)\mathbf{v}$.

In particular, if f is any branch of the logarithm which is analytic in a neighbourhood of $\sigma(P)$, then by (6), $\exp(f(P)) = P$. On the other hand, if P has distinct eigenvalues then by (4) every matrix Q such that $\exp(Q) = P$ must commute with P, and therefore by (8) $Q = g(P)$ for some $g \in H(P)$. By (6), $P = \exp(g(P)) = (\exp \circ g)(P)$, so by (7) $r = \exp(g(r))$ for each eigenvalue r of P, i.e. $g(r)$ is one of the values of $\log r$. There is a branch of the logarithm, f, which is analytic in a neighbourhood of $\sigma(P)$ with $f(r) = g(r)$ for each r, and so by (7) again, $Q = f(P)$.

Suppose P has row-sums 1, i.e. the vector $\mathbf{1}$ whose entries are all 1 is an eigenvector of P with eigenvalue 1. Then by (10), $f(P)$ has row-sums $f(1)$.

Recall that we will use Log to denote the principal branch of the logarithm, i.e. for any nonzero complex number z, $\mathrm{Log}(z)$ is the unique complex number w such that $\exp(w) = z$ and $-\pi < \mathcal{I}m\ w \le \pi$. This is analytic in the complex plane except for the branch cut $(-\infty, 0]$. Thus $\mathrm{Log}(P)$ is defined whenever P has no eigenvalues in $(-\infty, 0]$, and is a real matrix.

We now proceed to the proofs of the theorems from the text.

Proof of Theorem 1.

Note first that the quantity S is simply the square of the numerical radius of $P - I$. Hence, by the *spectral radius formula* (see e.g. Rudin, 1991, Theorem 10.13), the norm of $(P-I)^k$ is asymptotic (as $k \to \infty$) to $S^{k/2}$. Hence, if $S < 1$, then the series (1) converges geometrically quickly as claimed, and in fact converges absolutely.

Furthermore, recall the expansion

$$\log(x) = (x-1) - (x-1)^2/2 + (x-1)^3/3 - (x-1)^4/4 + \ldots$$

for real numbers x, provided that the series converges absolutely. This also applies to matrices, with 1 replaced by I. It follows that the definition (1) ensures that $\widetilde{Q} = \log(P)$, i.e. that $P = \exp(\widetilde{Q})$.

All that remains to prove is that $\widetilde{Q}$ has row-sums 0. For this we use a lemma:

Lemma 2. Let A and B be $N \times N$ matrices. Suppose that A has row-sums α, and B has row-sums β. Set $C = AB$. Then C has row-sums $\alpha\beta$.

Proof of Lemma 2. We compute that

$$\sum_j c_{ij} = \sum_j \sum_k a_{ik} b_{kj} = \sum_k a_{ik} \left(\sum_j b_{kj} \right) = \sum_k a_{ik} (\beta) = \beta \left(\sum_k a_{ik} \right) = \beta (\alpha) ,$$

thus proving Lemma 2. ∎

Now, since $P - I$ has row-sums 0, the lemma proves that $(P-I)^k$ has row-sums 0 for all positive integers k. Since the series (1) for $\widetilde{Q}$ is a limit of linear combinations of different $(P-I)^k$, it follows that $\widetilde{Q}$ also has row-sums 0. This completes the proof of Theorem 1. ∎

Remark. If instead $S > 1$ then the series (1) will never converge. This is not surprising. For example, if P is self-adjoint (e.g. symmetric), then $S > 1$ if and only if the matrix P has negative eigenvalues; and it is well-known that in this case there does not exist a generator for P.

Proof of Theorem 2.

Let $m = \min\{p_{ii}\}$. Then $m > \frac{1}{2}$ by assumption. Write

$$P = mI + (1-m)R$$

where $R = \frac{1}{1-m}(P - mI)$. Then R is also a Markov transition matrix. Now, we compute that

$$P - I = (1-m)(R - I)\,.$$

Since R is a transition matrix, we have $\|R\| \leq 1$ (where $\|\cdots\|$ is the usual operator norm), so that $\|R-I\| \leq 2$ by the triangle inequality, and $\|(R-I)^k\| \leq 2^k$. Hence, $\|(P-I)^k\| \leq (2-2m)^k$. But then by the spectral radius formula again,

$$\sqrt{S} = \lim_{k\to\infty} \|(P-I)^k\|^{1/k} \leq \lim_{k\to\infty} \left((2-2m)^k\right)^{1/k} = 2 - 2m\,.$$

Finally, since $m > \frac{1}{2}$, we have $2-2m < 1$, so that $\sqrt{S} < 1$, and hence also $S < 1$. This completes the proof. ■

Proof of Theorem 3.

For part (a), we recall the well-known formula that

$$\det\left(\exp(Q)\right) = \exp\left(\operatorname{trace}(Q)\right)\,. \qquad (A1)$$

(To see this, first transform into a basis where Q is upper-triangular.) Hence, if $P = \exp(Q)$ for some matrix Q (whether a generator or not), then we must have $\det(P) = \det\left(\exp(Q)\right) > 0$.

For part (b), suppose to the contrary that P had a generator Q. Let $R(t) = \exp(tQ)$. Then $R'_{ii}(t) \geq Q_{ii}R_{ii}(t)$ and $R_{ii}(0) = 1$, so $R_{ii}(t) \geq \exp(tQ_{ii})$. Hence, $p_{ii} = R_{ii}(1) \geq \exp(Q_{ii})$. Hence, using $(A1)$, we have

$$\prod_i p_{ii} \geq \prod_i \exp(Q_{ii}) = \exp\Big(\sum_i Q_{ii}\Big) = \exp\left(\operatorname{trace}(Q)\right) = \det(P)\,,$$

contradicting condition (b). Hence, assuming condition (b), there is no such generator Q.

Part (c) follows from the *Lévy Dichotomy*, which states that if P has a proper generator Q, then for each pair (i,j) of states, we must have either $p_{ij}(t) > 0$ for all $t > 0$ or $p_{ij}(t) = 0$ for all $t > 0$ (where $p_{ij}(t)$ is the ij entry of P_t).

To prove the Lévy Dichotomy, suppose that P did have a generator. Then for each state k we would have $p_{kk}(s) \to 1$ as $s \searrow 0$, so that for sufficiently small s we have $p_{kk}(s) > 0$ for all states k. (In fact, it then follows that this condition holds for all $s > 0$.) Hence, if $p_{ij}(t) = 0$ for some $t > 0$, then we must have $p_{ij}(t/n) = 0$ for all sufficiently large integers n (otherwise we would have $p_{ij}(t) \geq p_{ij}(t/n)\,(p_{jj}(t/n))^{n-1} > 0$). That is, the set of zeros of the function $p_{ij}(s)$ would have a limit point (i.e. 0). But $p_{ij}(s)$ is an analytic function of s. Hence, it must be that $p_{ij}(s) = 0$ for all $s > 0$, contradicting our assumption that $p_{ij}(t) = 0$ for some $t > 0$. This proves the Lévy Dichotomy.

To complete the proof of Theorem 3(c), suppose that j is accessible from i. Then we have $p_{ij}(m) > 0$ for appropriate positive integer m. Hence, we must have $p_{ij}(t) > 0$ for all t, and in particular $p_{ij}(1) = p_{ij} > 0$, as claimed. ∎

Proof of Theorem 4.

We claim that exp is one-to-one on $\{Q : \|Q\| < \ln 2\}$. To see this, note that if $\|Q_1\| \leq r$ and $\|Q_2\| \leq r$ with $0 < r < \ln 2$,

$$\begin{aligned}
\|\exp(Q_1) - \exp(Q_2) - (Q_1 - Q_2)\| &\leq \sum_{n=2}^{\infty} \|Q_1^n - Q_2^n\|/n! \\
&\leq \sum_{n=2}^{\infty} \sum_{j=0}^{n-1} \|Q_2^j (Q_1 - Q_2) Q_1^{n-1-j}\|/n! \\
&\leq \|Q_1 - Q_2\| \sum_{n=2}^{\infty} n r^{n-1}/n! \\
&= \|Q_1 - Q_2\|(e^r - 1) < \|Q_1 - Q_2\|
\end{aligned}$$

so $\|\exp(Q_1) - \exp(Q_2)\| > 0$.

Since $\det(\exp(Q)) = \exp(\operatorname{trace} Q)$, if Q is a generator for a transition matrix P with $\det(P) > 1/2$, we have $\operatorname{trace} Q > -\ln 2$. Now if we use the L^1-norm on vectors, $\|Q\| \leq \max_j \sum_i |Q_{ij}|$. But for a generator, $|Q_{ij}| \leq -Q_{ii}$ so $\|Q\| \leq -\sum_i Q_{ii} = -\operatorname{trace} Q < \ln 2$. Therefore there can be at most one generator. This proves (a).

For (b), note that if $\|P - I\| = r < 1/2$ then the series (1) for $\operatorname{Log} P$ converges, and $\|\operatorname{Log} P\| \leq -\ln(1 - r) < \ln 2$. Thus $Q = \operatorname{Log} P$.

For (c), note that if Q is a generator for P and $\det(P) > \exp(-\pi)$, then $0 \geq \operatorname{trace} Q > -\pi$. Let

$r = -\mathrm{trace}\, Q$. Now $Q + rI$ is a matrix with nonnegative entries, and its maximum nonnegative eigenvalue is r. By the Perron-Frobenius Theorem, all eigenvalues of this matrix have absolute value at most r, and in particular the imaginary part of any eigenvalue of Q is in $[-r, r]$. But Q must be $f(P)$ for some branch of the logarithm, and the only one whose imaginary parts are all in $(-\pi, \pi)$ is Log. ∎

Proof of Theorem 5.

Suppose Q is a generator for P. Since $\sigma(\exp(Q)) = \exp(\sigma(Q))$, exp must be one-to-one on $\sigma(Q)$. Since Q is a real matrix, any non-real eigenvalues come in complex-conjugate pairs. If r and $\bar{r}$ form such a pair, then $\exp(r)$ and $\exp(\bar{r})$ are eigenvalues of $P = \exp(Q)$. But $\exp(\bar{r}) = \overline{\exp(r)} = \exp(r)$ since the eigenvalues of P are real, so exp would not be one-to-one on $\sigma(Q)$. Thus Q must have all real eigenvalues. As remarked earlier, if P has distinct eigenvalues then $Q = f(P)$ where f is some branch of the logarithm. Thus for each eigenvalue r of P, $f(r)$ is real and thus must be $\mathrm{Log}(r)$. Therefore $Q = \mathrm{Log}(P)$. Moreover, P can't have any negative eigenvalues, because negative numbers have no real logarithms. ∎

Proof of Theorem 6.

The key is to condition on the number of jumps N which the chain makes in a time interval of length 1. Now $N \leq N'$, where N' is a Poisson random variable with mean $\lambda \equiv \max_i(-Q_{ii})$. Let $R(t) = \exp(tQ)$ and $R_{i,j}(n,t) = \mathbf{Pr}(X(t) = j | X(0) = i, N = n)$. Then $R_{i,j}(n, t/2) \geq 2^{-n} R_{i,j}(n,t)$ (as can be seen by writing the probability of a given sequence of n transitions as an n-fold integral). So

$$R_{i,j}(s) \geq s^m \sum_{n \leq m} R_{i,j}(n,1) \geq s^m (P_{i,j} - \mathbf{Pr}(N' > m))$$

So if $\mathbf{Pr}(N' > m) = b_m$, $P_{i,j} > b_m$ and $P_{j,k} > b_r$, then we must have

$$P_{i,k} \geq R_{i,j}(s) R_{j,k}(1-s) \geq s^m (1-s)^r (P_{i,j} - b_m)(P_{j,k} - b_r)$$

for any $0 < s < 1$. This bound is maximised when $s = m/(m+r)$. Summing over all j with $P_{i,j} > b_m$ and $P_{j,k} > b_r$ then gives the result. ∎

Proof of Theorem 7.

If P has all distinct eigenvalues, then any Q will be $f(P)$ where f is some branch of the logarithm, and thus will be determined by the values $f(r)$ for eigenvalues r of P, where $f(r) = \mathrm{Log}(r) + 2\pi k(r)i$ for some integers $k(r)$. If Q can't have eigenvalues with |imaginary part| $\geq \pi$ then uniqueness is established. Now if $\det(P) > \exp(-\pi)$, then $\mathrm{trace}(Q) > -\pi$, and (for some $\epsilon > 0$), $Q + (\pi - \epsilon)I$ is a matrix with all nonnegative entries and its largest real eigenvalue is $\pi - \epsilon$. By Perron-Frobenius, all eigenvalues of $Q + (\pi - \epsilon)I$ have absolute value $\leq \pi - \epsilon$, so the imaginary part of any eigenvalue of Q has absolute value $< \pi$. ∎

REFERENCES

A. Arvanitis, J. Gregory, and J.-P. Laurent (1999), Building models for credit spreads. *Journal of Derivatives*, Spring 1999, 27–43.

B. Belkin, S. Suchower, and L. Forest Jr. (1998), A one-parameter representation of credit risk and transition matrices. CreditMetrics Monitor, Third Quarter, 1998.

K.L. Chung (1967), Markov chains with stationary transition probabilities. Springer-Verlag, New York.

J.R. Cuthbert (1972), On uniqueness of the logarithm for Markov semi-groups. J. London Mathematical Society **4**, 623–630.

J.R. Cuthbert (1973), The logarithmic function for finite-state Markov semi-groups. J. London Mathematical Society **6**, 524–532.

G. Elfving (1937), Zur Theorie der Markoffschen Ketten. Acta Social Science Fennicae n., series A.2, no. 8, 1–17.

G.S. Goodman (1970), An intrinsic time for non-stationary finite Markov chains. Zeitschrift für Wahrscheinlichkeitstheorie **16**, 165–180.

G.R. Grimmett and D.R. Stirzaker (1992), Probability and random processes (2^{nd} ed.). Oxford University

Press.

R.A. Horn and C.R. Johnson (1985), Matrix Analysis. Cambridge University Press.

R.A. Jarrow, D. Lando, and S.M. Turnbull (1997), A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**, 481–523.

S. Johansen (1973), A central limit theorem for finite semi-groups and its application to the imbedding problem for finite-state Markov chains. Zeitschrift für Wahrscheinlichkeitstheorie **26**, 171–190.

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

M. Kijima (1998), Monotonicities in a Markov chain model for valuing corporate bonds subject to credit risk. Mathematical Finance, Vol. **8 (3)**.

M. Kijima and K. Komoribayashi (1998), A Markov chain model for valuing credit risk derivatives. Journal of Derivatives, Fall, 1998.

J.F.C. Kingman (1962), The imbedding problem for finite Markov chains. Zeitschrift für Wahrscheinlichkeitstheorie **1**, 14–24.

D. Lando (2000), Some Elements of Rating-Based Credit Risk Modeling. In: N. Jegadeesh and B. Tuckman (eds): Advanced Fixed-Income Valuation Tools, Wiley, 193–215.

Moody's Investors Services (1999), Historical default rates of corporate bond issuers, 1920–1998. Moody's, New York.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1988), Numerical recipes in C: the art of scientific computing. Cambridge University Press.

W. Rudin (1991), Functional Analysis, 2nd ed. McGraw-Hill, New York.

J. Th. Runnenberg (1962), On Elfving's problem of imbedding a time-discrete Markov chain in a continuous time one for finitely many states. Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, series A, Mathematical Sciences **65 (5)**, 536–541.

B. Singer and S. Spilerman (1976), The representation of social processes by Markov models. American Journal of Sociology **82**, 1–54.

Standard & Poor's Credit Review (1993), Corporate default, rating transition study updated. McGraw-Hill, New York.

Standard & Poor's (1999), Rating performance 1998: Stability and transition. Standard and Poor's, New York.

A. Stewart, R. Philips, and G. Dempsey (1998), A Markov-cycle evaluation of five years' therapy using Donepezil. *International Journal of Geriatric Psychiatry* **13**, 445–453.

J.H.M. Wedderburn (1934), Lectures on Matrices. American Mathematical Society Colloquium Publications **17**. (Republished by Dover Publications, New York, 1964.)

S. Zahl (1955), A Markov process model for follow-up studies. Human Biology **27**, 90–120.