Measurement, estimation and comparison of credit migration matrices

Yusuf Jafry [a], Til Schuermann [b,*]

[a] *Risk Integrated Group, Third Floor, Central Chambers, 6 Victoria Street, Douglas, IM1 2LH Isle of Man, UK*

[b] *Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, USA*

Abstract

Credit migration matrices are cardinal inputs to many risk management applications; their accurate estimation is therefore critical. We explore two approaches: cohort and two variants of duration – one imposing, the other relaxing time homogeneity – and the resulting differences, both statistically through matrix norms and economically using a credit portfolio model. We propose a new metric for comparing these matrices based on singular values and apply it to credit rating histories of S&P rated US firms from 1981–2002. We show that the migration matrices have been increasing in "size" since the mid-1990s, with 2002 being the "largest" in the sense of being the most dynamic. We develop a testing procedure using bootstrap techniques to assess statistically the differences between migration matrices as represented by our metric. We demonstrate that it can matter substantially which estimation method is chosen: economic credit risk capital differences implied by different estimation techniques can be as large as differences between economic regimes, recession vs. expansion. Ignoring the efficiency gain inherent in the duration methods by using the cohort method instead is more damaging than imposing a (possibly false) assumption of time homogeneity.
© 2004 Published by Elsevier B.V.

---

[*] Corresponding author. Tel.: +1-212-720-5968; fax: +1-212-720-8363.

E-mail addresses: yusuf.jafry@riskintegrated.com (Y. Jafry), til.schuermann@ny.frb.org (T. Schuermann).

1. Introduction

Credit migration or transition matrices, which characterize past changes in credit quality of obligors (typically firms), are cardinal inputs to many risk management applications, including portfolio risk assessment, modeling the term structure of credit risk premia, pricing of credit derivatives and assessment of regulatory capital. For example, in the New Basel Accord (BIS, 2001), capital requirements are driven in part by ratings migration. Their accurate estimation is therefore critical. In this paper we present several methods for measuring, estimating and comparing credit migration matrices. Specifically we explore two approaches, cohort and two variants of duration (or hazard) – parametric (imposing time homogeneity or invariance) and non-parametric (relaxing time homogeneity). [1] We ask three questions: (1) how would one measure the scalar difference between these matrices; (2) how can one assess whether those differences are statistically significant; and (3) even if the differences are statistically significant, are they economically significant?

We use these different estimation methods to compute credit migration matrices from firm credit rating migration histories from Standard and Poors (S&P) covering 1981–2002. We then compare the resulting differences, both statistically through matrix norms, eigenvalue and -vector analysis, and economically with a credit portfolio model. Along the way we develop a convenient scalar metric which captures the overall dynamic size of a given matrix and contains sufficient information to facilitate meaningful comparisons between different credit migration matrices. We show that these migration matrices have been getting ''larger'' since the mid-1990s and that they tend to increase during recessions. The most recent year available, 2002, has generated the ''largest'' migration matrix. Moreover, the matrices estimated with the cohort method tend to be ''smaller'' than the duration matrices, and this difference seems to be increasing recently.

We propose a bootstrap test to assess the differences between Markov matrices as represented by our metric. We demonstrate that it can matter substantially which method is chosen: for example, economic credit risk capital differences implied by different estimation techniques can be as large as differences between economic regimes, recession vs. expansion. Viewed through the lens of credit risk capital, ignoring the efficiency gain inherent in the duration methods is more damaging than making a (possibly false) assumption of time homogeneity, a significant result given that the cohort method is the method of choice for most practitioners.

---

[1] In simple terms, the cohort approach just takes the observed proportions from the beginning of the year to the end (for the case of annual migration matrices) as estimates of migration probabilities. For example, if two firms out of 100 went from grade 'AA' to 'A', then the $P_{AA \to A} = 2\%$. Any movements within the year are not accounted for. The duration approach counts all rating changes over the course of the year and divides by the time spent in the starting state or rating to obtain the migration probability estimate.

Table 1
Value change of a BBB bond 1 year hence

Credit grade at year-end	Marrison (2002, Chapter 18)	Saunders and Allen (2002, Chapter 6)
AAA	101.7%	101.7%
AA	101.4%	101.5%
A	100.9%	101.0%
BBB	100.0%	100.0%
BB	96.9%	94.9%
B	92.6%	91.2%
CCC	89.3%	77.8%
D	70.0%	47.5%

Comparison is relative to no rating change.

Perhaps the simplest use of a transition or migration matrix is the valuation of a bond or loan portfolio. Given a credit grade today, say BBB, [2] the value of that credit asset one year hence will depend, among other things, on the probability that it will remain BBB, migrate to a better or worse credit grade, or even default at year-end. This can range from an increase in value of 1–2% in case of upgrade to a decline in value of 30–50% in case of default, as illustrated in Table 1. [3] More sophisticated examples of risky bond pricing methods, such as outlined by Jarrow and Turnbull (1995) and Jarrow et al. (1997), require these matrices as an input, as do credit derivatives such as models by Kijima and Komoribayashi (1998) and Acharya et al. (2002). In risk management, credit portfolio models such as CreditMetrics® (Gupton et al., 1997) make use of this matrix to simulate the value distribution of a portfolio of credit assets.

To our knowledge there has been little work in establishing formal comparisons between credit migration matrices. In the literature such metrics applied to the transition matrices for general Markov chains, which measure the amount of migration (mobility), are sometimes called *mobility indices*. Shorrocks (1978), looking at income mobility, propose indices for Markov matrices using eigenvalues and determinants, a line of inquiry extended in Geweke et al. (1986), hereafter GMZ. They present a set of criteria by which the performance of a proposed metric (for arbitrary transition matrices) should be judged. We propose an additional criterion, *distribution discriminatory*, which is particularly relevant for credit migration matrices: the metric should be sensitive to the distribution of off-diagonal probability mass. This is important since far migrations have different economic and financial meaning than near migrations. The most obvious example is migration to the Default state (typically the last column of the migration matrix) which clearly has a different impact than migration of just one grade down (i.e. one off the diagonal).

Credit migration matrices are said to be diagonally dominant, meaning that most of the probability mass resides along the diagonal; most of the time there is no migration. Bangia et al. (2002) estimate coefficients of variation of the elements or

---

[2] For no reason other than convenience and expediency, we will make use of the S&P nomenclature for the remainder of the paper.

[3] Default rarely results in total loss.

parameters of the migration matrix as a characterization of estimation noise or uncertainty. Unsurprisingly they find that by and large only the diagonal elements are estimated with high precision. The further one moves away from the diagonal, the lower the degree of estimation precision. They also conduct *t*-tests to analyze cell-by-cell differences between different migration matrices; again, because of the low number of observations for far-off-diagonal elements these *t*-tests were rarely significant. [4] Christensen et al. (2004) develop bootstrap methods to estimate confidence sets for transition probabilities, focusing on the default probabilities in particular, which are superior to traditional multinomial estimates; specifically, they are tighter. Arvanitis et al. (1999) (hereafter referred to as AGL) propose a metric to compare migration matrices of different horizons and test the first-order Markov assumption. They suggest a cut-off value of 0.08 but do not tell us why 0.08 is sufficiently small, nor what would be sufficiently large to reject similarity. Moreover, they ignore estimation noise and concomitant parameter uncertainty.

In a separate but related line of research, Israel et al. (2001) show conditions under which generator matrices exist for an empirically observed Markov transition matrix and propose adjustments to guarantee existence. They use the L^1 norm (average absolute difference) to examine differences in pre- and post-adjustment migration matrices. However, they do so without recognizing that the matrices are estimated with error, making it difficult to judge whether a computed distance is in fact large enough to overcome estimation noise. We are the first to propose a formal scalar metric suitable for credit migration matrices and to devise a procedure for evaluating their statistical significance in the presence of estimation noise.

The outline of the paper is as follows. In Section 2 we establish notation and definitions, review some well-known dynamic properties of Markov matrices using eigenvalues (and eigenvectors), and summarize the existing techniques for comparing matrices. In Section 3 we propose our new metric and new performance criterion. Here we motivate the subtraction of the identity matrix and develop a metric based on singular values. Section 4 gives a brief presentation of how to estimate the migration matrices and of the bootstrap method which we use to assess statistical significance of our matrix metric presented in Section 3. In Section 5 we apply our metric to credit rating histories of US firms from 1981 to 2002, and explore different methods for estimating credit migration matrices. We examine whether the empirical estimates are statistically distinguishable and whether they make material economic difference. Section 6 provides some concluding remarks.

2. Credit migration matrices

In this section we briefly introduce definitions and notation and set up the analytic framework. The basic question we want to answer is: how does one compare two

---

migration matrices? Typically these are large objects. For instance, with seven whole grades or ratings plus the Default state, we have an 8×8 matrix with $(7 \times 7) + 7 = 56$ unique cells to compare. To be sure, these matrices have a lot of structure, and in this and the next section we explore and exploit that structure to arrive at a scalar metric of comparison.

2.1. Definitions

Consider the state vector, $\mathbf{x}(t)$, defined as a row-vector containing the discrete probability distribution of the credit rating (say, for a given firm) at time t. For example, this could be a portfolio of loans apportioned by rating. The number of elements of $\mathbf{x}$, denoted N, corresponds to the number of different possible credit ratings (typically arranged in the order from best to worst, with the ''Default'' rating as its last element). [5] Usually such matrices are evaluated at discrete points in time separated by the sample period, Δt, in which case $\mathbf{x}(k)$ is then taken to represent the state at time $k\,\Delta t$ (i.e. the time epoch with index k). The row-vector $\mathbf{x}(k + 1)$ describes the discrete probability distribution of the credit rating (for the same firm or portfolio) at the next discrete point in time, $(k + 1)\Delta t$. We assume that the discrete evolution of the state vector is governed by a Markov process [6] such that $\mathbf{x}(k + 1) = \mathbf{x}(k)\mathbf{P}$, where $\mathbf{P}$ represents the migration matrix defining the transition of the state vector from one epoch to the next. Each row of $\mathbf{P}$ defines a discrete probability distribution describing the probability of transitioning from a given credit rating at time $k\Delta t$ to any of the possible credit ratings at time $(k + 1)\Delta t$.

In order to understand the dynamics of the process and therefore any mobility metrics, it will be key to look at its time evolution. We turn to this next.

2.2. Time evolution: Eigenvalue decomposition

Assuming that $\mathbf{P}$ is time homogeneous (or time invariant), i.e. constant over time (perhaps unrealistic in practice, but a useful mathematical simplification), the solution for the state vector at any future time can be expressed in terms of the initial state vector, $\mathbf{x}(0)$, for any k as

$$\mathbf{x}(k) = \mathbf{x}(0)\mathbf{P}^k. \tag{2.1}$$

Since the eigenvalues (and eigenvectors) of the transition matrix are intimately related to the time evolution of the state vector, we often express $\mathbf{P}$ in terms of its eigenvalue decomposition, i.e.

$$\mathbf{P} = \mathbf{S}\mathit{\Lambda}\mathbf{S}^{-1},$$

where $\mathit{\Lambda}$ represents the diagonal matrix containing the eigenvalues of $\mathbf{P}$, and $\mathbf{S}$ contains the corresponding eigenvectors (one per column). Eq. (2.1) becomes

---

[5] Purely as a matter of convenience, we will follow the notation from Standard and Poors (S&P) which, from best to worst, is AAA, AA, A, BBB, BB, B, and CCC. Thus typically $N = 8$ (including default D).

[6] See Israel et al. (2001) for conditions under which credit migration matrices are Markov.

$$\mathbf{x}(k) = \mathbf{x}(0)\mathbf{S}\Lambda^k\mathbf{S}^{-1}. \tag{2.2}$$

Eigenvalues and -vectors will play a central role in several mobility indices discussed below in Section 2.5.

2.3. Steady-state behavior

An important consequence of the fact that each row of $\mathbf{P}$ sums to unity is that the dynamic system (describing the evolution of $\mathbf{x}(k)$) is *neutrally stable*. In other words, the solution for $\mathbf{x}(k)$ never dies away to zero (as it would for a *stable* system) nor does it explode to infinity (as it would for an *unstable* system): instead it reaches a steady-state solution, $\mathbf{x}_\infty \triangleq \mathbf{x}(k \to \infty)$. This implies that (at least) one of the eigenvalues (elements of Λ) must be equal to 1, such that when raised to the kth power (in Eq. (2.2)) it persists indefinitely, and that all the other (non-unity) eigenvalues have magnitudes less than one such that when raised to the kth power, they eventually decay away. The steady-state probability distribution, $\mathbf{x}_\infty$, is given by

$$\mathbf{x}(\infty) = \mathbf{x}(0)\mathbf{P}^\infty = \mathbf{x}(0)\mathbf{S}\Lambda^\infty\mathbf{S}^{-1},$$

where $\mathbf{P}^\infty$ represents $\mathbf{P}$ raised to the infinite power, resulting in the limiting transition matrix (denoted $\mathbf{P}^\dagger$ in GMZ).

In the most general case, only one eigenvalue will be equal to unity with all others less than unity. In that case, $\mathbf{x}_\infty$ is given by (some multiple of) the eigenvector (of the transpose of $\mathbf{P}$) corresponding to the unity eigenvalue. [7] The rate at which the system decays towards $\mathbf{x}_\infty$ is governed by the slowest-decaying term or the second-largest eigenvalue (denoted λ_2).

Consider the following 2-d example to make things clear. Assuming a two-ratings system, say "A" and "B", with no Default absorbing state for now, the most general migration matrix is given by

$$\mathbf{P} = \begin{pmatrix} 1 - p_1 & p_1 \\ p_2 & 1 - p_2 \end{pmatrix}; \quad 0 \leqslant p_1 \leqslant 1; \ \ 0 \leqslant p_2 \leqslant 1. \tag{2.3}$$

The rate of decay in the solution is governed by the magnitude of λ_2. [8] For example, the time taken for the system to decay to, say, within 10% of the steady state, is given by

$$k_{10\%} = \frac{\log(10\%)}{\log(\lambda_2)} = \frac{\log(0.1)}{\log(|\ 1 - p_1 - p_2\ |)}. \tag{2.4}$$

If we consider typical values of, say, 0.05 for p_1 and p_2 in a yearly transition matrix, Eq. (2.4) gives

---

[7] These particular properties associated with the unity eigenvalue (and its eigenvector) of a Markov matrix are a consequence of the Perron–Frobenius theorem in matrix theory.

[8] See the appendix in Jafry and Schuermann (2003) for details and an illustration.

$$k_{10\%} = \frac{\log(0.1)}{\log(0.9)} \approx 22 \text{ years.} \tag{2.5}$$

In economic terms, this is a long time, especially when we consider that records of credit ratings are available only over a few decades. Moreover, the validity of the linear, time invariant Markov assumptions become more questionable over such long time periods.

2.4. The absorbing "Default" state

The "Default" state is usually considered as absorbing, implying that any firm which has reached this state can never return to another credit rating. [9] An important mathematical consequence of the inclusion of the absorbing state is that the steady-state solution (i.e. the first eigenvector of (the transpose of) $\mathbf{P}$) is identically equal to the absorbing row of $\mathbf{P}$. In other words, for a general migration matrix which exhibits non-trivial probabilities of default (i.e. with some non-zero elements in the absorbing column), the probability distribution $\mathbf{x}(k)$ will always settle to the default state. Given sufficient time, all firms will eventually sink to the Default state. [10] This behavior is clearly a mathematical artifact, stemming from the idealized linear, time invariant assumptions inherent in the simple Markov model. In reality the economy (and hence the migration matrix) will change on time-scales far shorter than required to reach the idealized Default steady-state proscribed by an assumed constant migration matrix.

2.5. Existing techniques for comparing matrices

2.5.1. Cell-by-cell distance metrics

Two common approaches for comparing two matrices (say $\mathbf{P}_A$ and $\mathbf{P}_B$, each with dimension $N \times N$) are the L^1 and L^2 (Euclidean) distance metrics. [11] The L^1 metric computes the average absolute difference while the L^2 (Euclidean) distance metric computes the average root-mean-square difference between corresponding elements of the matrices. Specifically,

$$
\begin{aligned}
\Delta M_{L1}(\mathbf{P}_A, \mathbf{P}_B) &\triangleq \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{P}_{A,i,j} - \mathbf{P}_{B,i,j}|}{N^2}, \\
\Delta M_{L2}(\mathbf{P}_A, \mathbf{P}_B) &\triangleq \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{P}_{A,i,j} - \mathbf{P}_{B,i,j})^2}}{N^2}.
\end{aligned}
\tag{2.6}
$$

---

[9] There are exceptions, namely if a firm re-emerges from bankruptcy and then obtains a credit rating on a debt instrument.

[10] See the appendix in Jafry and Schuermann (2003) for details and an illustration.

[11] The L^1 metric is used in Israel et al. (2001) for comparing migration matrices while the L^2 metric is used in Bangia et al. (2002).

Although appealing in their simplicity, these methods offer no absolute measure for an individual matrix: they only provide a relative comparison between two matrices. For example, if the Euclidean distance between two matrices turns out to be, say, 0.1, it is not clear if this is a "large" or a "small" distance, nor is it possible to infer which matrix is the "larger" of the two.

2.5.2. Eigenvalue-based metrics

The mobility indices presented in GMZ for general transition matrices are all, in essence, based on the eigenvalues of **P**. They can be summarized as follows:

$$M_\mathrm{P}(\mathbf{P}) = \frac{1}{N-1}(N - \mathrm{tr}(\mathbf{P})),$$

$$M_\mathrm{D}(\mathbf{P}) = 1 - |\det(\mathbf{P})|,$$

$$M_\mathrm{E}(\mathbf{P}) = \frac{1}{N-1}\left(N - \sum_{i=1}^{N} |\lambda_i(\mathbf{P})|\right), \tag{2.7}$$

$$M_2(\mathbf{P}) = 1 - |\lambda_2(\mathbf{P})|,$$

where tr(...) denotes the *trace* of the matrix (i.e. the sum of its diagonal elements), det(...) denotes the *determinant*, and λ_i(...) denotes the ith eigenvalue (arranged in the sequence from largest to smallest absolute value, with λ_2 denoting the largest less than unity). Note that when all the eigenvalues of **P** are real and non-negative, M_P is identical to M_E since the trace equals the sum of the eigenvalues.

2.5.3. Eigenvector distance metric

Since credit migration matrices incorporate an absorbing state, they will have identical steady-state solutions (i.e. given by the absorbing state vector itself), meaning that the steady-state solution, or, equivalently, the first eigenvector of (the transpose of) the migration matrix is ineffective as a basis for comparing matrices. The remaining eigenvectors do, however, contain useful information which can be used to construct a relative metric. AGL propose to assess the similarity of all eigenvectors between two matrices by computing a (scalar) ratio of matrix norms. Specifically, their approach is motivated by the need to compare migration matrices with different horizons (i.e. sample intervals, Δt) and test the first-order Markov assumption, but the mathematics are equally valid for comparing two different transition matrices over the same horizon:

$$\Delta M_\mathrm{AGL}(\mathbf{P}_\mathrm{A}, \mathbf{P}_\mathrm{B}) \triangleq \frac{\|\mathbf{P}_\mathrm{A}\mathbf{P}_\mathrm{B} - \mathbf{P}_\mathrm{B}\mathbf{P}_\mathrm{A}\|}{\|\mathbf{P}_\mathrm{A}\| \cdot \|\mathbf{P}_\mathrm{B}\|}, \tag{2.8}$$

where for a vector **x**, $\|\mathbf{x}\|$ denotes the length of the vector.

This quantity is bounded between zero and two; it is equal to zero if $\mathbf{P}_\mathrm{A}$ and $\mathbf{P}_\mathrm{B}$ have exactly the same eigenvectors (regardless of their eigenvalues), and it increases the more dissimilar the eigenvectors become. AGL conclude that values for the metric ΔM_AGL of around 0.08 at an annual frequency are sufficiently small to suggest that the eigenvectors vary by only small amounts and can thus assumed to be similar.

However, they do not tell us why 0.08 is sufficiently small, nor do they account for estimation noise.

3. Devising a new metric

3.1. Performance criteria

GMZ (elaborating on the work of Shorrocks, 1978) present a set of criteria by which the performance of a proposed metric (for arbitrary transition matrices) should be judged. These are grouped in three distinct areas: *persistence criteria* which stipulate that a metric should be consistent with some simple, intuitively appealing interpretations of the transition matrix $\mathbf{P}$; *convergence criteria* which stipulate that a metric ought to establish an ordering among transition matrices $\mathbf{P}$ that is consistent with the rate at which the multiperiod transition matrices $\mathbf{P}^k$ converge to the limiting transition matrix $\mathbf{P}^\dagger$; and *temporal aggregation criteria* which remove the influence of the length of the basic time period (Δt) on comparisons of mobility.

Of the persistence criteria, *monotonicity* (M) stipulates that $M(\mathbf{P}) > M(\mathbf{P}^*)$ if $p_{ij} \geqslant p_{ij}^*$ for all $i \neq j$ and $p_{ij} > p_{ij}^*$ for some $i \neq j$. Imposing, without loss of generality, that $M(\mathbf{I}) = 0$ (i.e. metric is zero if the matrix implies zero mobility), then the criterion of immobility ($\mathbf{I}$) stipulates $M(\mathbf{P}) \geqslant 0$, and under *strong immobility* (SI) $M(\mathbf{P}) \geqslant 0$ unless $\mathbf{P} = \mathbf{I}$.

Of the convergence criteria, *perfect mobility* (PM) requires that $M(\mathbf{P}^\dagger) \geqslant M(\mathbf{P})$ for all $\mathbf{P}$, and strong perfect mobility (SPM) requires that the inequality be strict unless $\mathbf{P} = \mathbf{P}^\dagger$.

Of the temporal aggregation criteria, *period consistency* (PC) is based on the idea that the comparisons of rates of convergence should not be reversed by changes in Δt, which implies that if $\mathbf{P}$ and $\mathbf{P}^*$ are two transition matrices and $M(\mathbf{P}) \geqslant M(\mathbf{P}^*)$, then $M(\mathbf{P}^k) \geqslant M(\mathbf{P}^{*k})$ for all integers $k > 0$.

For the class of transition matrices with real non-negative eigenvalues, GMZ show that all the criteria are logically consistent, implying that it should be possible, in principle, to construct a class of mobility indices which satisfy all criteria. That said, the authors do not claim that the particular mobility indices which they then proceed to present (summarized in Eq. (2.7)), do, in fact, satisfy all the criteria, even for the limited class of matrices considered; see GMZ for examples. Moreover, for general matrices (which may violate the real non-negative eigenvalue restriction), the criteria are logically inconsistent across the persistence and convergence categories.

Credit migration matrices are generally diagonally dominant, and so the non-unity eigenvalues are rather close to unity in magnitude, which in turn implies that the decay rates towards steady-state are generally rather slow (see also Section 2.3, in particular, Eq. (2.5)). We therefore contend that it is appropriate to ignore the requirements of meeting the convergence criteria. We likewise neglect the temporal aggregation criteria because we are most often concerned with comparing credit migration matrices evaluated for a fixed Δt (e.g. one year) considerably shorter than

the natural decay-time of the system. We therefore focus on the persistence criteria only, [12] thereby removing the logical inconsistencies of attempting to satisfy all categories.

As an additional persistence criterion, our metric is required to be *distribution discriminatory* (**DD**), such that it can discriminate matrices having the same rowwise probabilities of change but different distributions across each row. Specifically, for $p_{ii} = p_{ii}^*$, if $p_{ij} \neq p_{ij}^*$ for all $j \neq i$, then $M(\mathbf{P}) \neq M(\mathbf{P}^*)$. For example, consider the following two matrices:

$$\mathbf{P}_1 = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{pmatrix}, \qquad \mathbf{P}_2 = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.7 & 0 \\ 0.4 & 0 & 0.6 \end{pmatrix}. \tag{3.1}$$

A "good" metric should satisfy **DD**, i.e. it should yield different values for these two matrices. By contrast, as demonstrated in Appendix A.3, none of the eigenvalue-based metrics in Eq. (2.7) makes a distinction between these matrices. To be able to make such a distinction is important in the context of credit migration matrices (and may be important more broadly) since far migrations have different economic and financial meaning than near migrations. The most obvious example is migration to the Default state (typically the last column) which clearly has a different impact than migration of just one grade down (i.e. one off the diagonal).

3.2. Subtraction of the identity matrix

Since the migration matrix, by definition, determines quantitatively how a given state vector (or probability distribution) will migrate from one epoch to the next, a central characteristic of the matrix is the amount of migration (or "mobility") imposed on the state vector from one epoch to the next. We can highlight this characteristic by simply subtracting the identity matrix before proceeding with further manipulations. This apparently trivial observation turns out to be key. The identity matrix (of the same order as the state vector) corresponds to a *static* migration matrix, i.e. the state vector is unchanged by the action of the matrix from one epoch to the next. Subtracting the identity matrix from the migration matrix leaves only the *dynamic* part of the original matrix, which reflects the "magnitude" of the matrix in terms of the implied mobility. By definition, the metric will satisfy **SI**. Note that the metrics in GMZ (Eq. (2.7)) have been constructed such that they also satisfy **SI**. However, we will go a step further by subtracting the identity matrix before computing the metric, in contrast to GMZ where all the metrics are computed directly on **P**.

We will henceforth refer to the *mobility matrix* (denoted $\widetilde{\mathbf{P}}$) defined as the original (hereafter referred to as the *raw*) matrix minus the identity matrix (of the same dimension), i.e.

$$\widetilde{\mathbf{P}} \triangleq \mathbf{P} - \mathbf{I}. \tag{3.2}$$

---

[12] This does not mean to say that our proposed metric will not satisfy the convergence and temporal aggregation criteria, but simply that we do not explicitly impose these criteria when devising the metric.

Our task now is to devise a metric based on manipulations of the mobility matrix $\widetilde{\mathbf{P}}$ which satisfies the persistence criteria of *M* and *DD* with respect to the raw migration matrix **P**.

3.3. A metric based on singular values

Recalling the state equation, $\mathbf{x}(k + 1) = \mathbf{x}(k)\mathbf{P}$, and substituting Eq. (3.2) for **P**, we obtain

$$\mathbf{x}(k + 1) = \mathbf{x}(k)\widetilde{\mathbf{P}} + \mathbf{x}(k)\mathbf{I}.$$

Clearly, the greater the "magnitude" of $\widetilde{\mathbf{P}}$, the greater the degree of migration applied to the state vector (and likewise, a zero value for $\widetilde{\mathbf{P}}$ implies zero migration). Therefore, for our metric to satisfy *M* and *DD* with respect to **P**, we need to capture the "magnitude" of $\widetilde{\mathbf{P}}$ with regard to its "amplifying power" on **x**. This is precisely the question posed when defining the norm for a given matrix, as described in Strang (1988, p. 366), whereby the norm of given matrix **A** (or equivalently of **A**′) is the scalar quantity defined by

$$\|\mathbf{A}\| = \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \tag{3.3}$$

such that $\|\mathbf{A}\|$ bounds the "amplifying power" of the matrix: $\|Ax\| \leqslant \|\mathbf{A}\|\|\mathbf{x}\|$ for all column-vectors **x** (or $\|\mathbf{x}'\mathbf{A}'\| \leqslant \|\mathbf{A}'\|\|\mathbf{x}'\|$ for all row-vectors **x**′).

The equality holds for at least one non-zero **x** (representing the specific direction(s) in which the amplification is maximized). Again following Strang (1988, p. 368), the solution for the norm is the largest singular value of **A**, which, in turn, is identically given by the square-root of the largest eigenvalue of **A**′**A**. The vector which is amplified the most is given by the eigenvector of **A**′**A** corresponding to the maximum eigenvalue. However, this maximally amplified vector is generally *not* representative of a feasible state vector.

So, rather than using just the maximum singular value (as prescribed by the matrix norm), we now propose to use the *average* of all of the singular values of $\widetilde{\mathbf{P}}$. By incorporating all the singular values, we can hope to capture the general characteristics of $\widetilde{\mathbf{P}}$ acting on a feasible state vector.

In summary we propose a metric defined as the *average of the singular values of the mobility matrix*, i.e.

$$M_{\mathrm{SVD}}(\mathbf{P}) \triangleq \frac{\sum_{i=1}^{N} \sqrt{\lambda_i(\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}})}}{N}. \tag{3.4}$$

Note that we have not proven that this metric satisfies *M* and *DD*; such proofs are beyond the scope of this paper. We will rather verify by example in the following sections.

3.4. Calibrating the metric

In this section, we provide an intuitively appealing "calibration" for the magnitude of the metric which is independent of N, the dimension (number of rating categories) of the migration matrix. In this way we can provide meaningful answers to questions such as "supposing the metric for a given matrix has a value of, say 0.1, what does this tell us about the matrix?"

In order to calibrate the metric, we introduce a hypothetical "average" migration matrix, denoted $\mathbf{P}_{avg}$, which has been devised such that all diagonal elements are equal to $(1 - p)$ and all off-diagonal elements are equal to $p/(N - 1)$, where p represents the probability that a given state will undergo a migration (to any of the others) under the action of $\mathbf{P}_{avg}$:

$$\mathbf{P}_{avg} \triangleq \begin{pmatrix} 1-p & p/(N-1) & \ddots & & & \\ p/(N-1) & 1-p & p/(N-1) & \ddots & & \\ \ddots & p/(N-1) & 1-p & p/(N-1) & \ddots & \\ \ddots & & p/(N-1) & 1-p & \ddots & \\ \ddots & & & \ddots & & \ddots \end{pmatrix}. \tag{3.5}$$

Although not representative of a real migration matrix, $\mathbf{P}_{avg}$ captures the qualitative attributes of an "average" migration matrix in that a given state has probability p of undergoing a migration.

The question of calibration now reduces to the following: "how does the value of our metric for an arbitrary test matrix relate to the value for a hypothetical $\mathbf{P}_{avg}$ of the same dimension?" thereby indicating intuitively the "average amount of migration" contained in the test matrix.

As derived in Appendix A.2, the M_{SVD} metric applied to $\mathbf{P}_{avg}$ yields the following exact result:

$$M_{SVD}(\mathbf{P}_{avg}) = p, \tag{3.6}$$

which states that our average singular value metric is *numerically identical* to the average probability of migration, p, for the hypothetical average matrix $\mathbf{P}_{avg}$, irrespective of dimension N.

To return to our previous hypothetical question, if the value of the singular value metric for a given arbitrary test matrix is, say 0.1, then we can now say that the matrix has an effective average probability of migration of 0.1. [13]

---

[13] Note that this will generally be numerically different to the average probability of migration computed directly from the diagonal elements of a given empirical matrix. This latter quantity is, in fact, exactly equal to M_{dev} in Eq. (3.8).

3.5. Mobility metrics compared

Recall that our singular value metric has been developed intuitively from the principles underlying the idea of the matrix norm which captures the essence of the "amplifying power" of the dynamic part of the migration matrix. The development rests on the key idea of first subtracting the identity matrix (to yield the dynamic part).

For comparison, we can also evaluate the absolute deviation and Euclidean distance metrics between $\mathbf{P}$ and $\mathbf{I}$, yielding respectively (from Eq. (2.6)):

$$
\begin{aligned}
M_{L1}(\mathbf{P},\mathbf{I}) &\triangleq \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{P}_{i,j} - \mathbf{I}_{i,j}|}{N^2}, \\
M_{L2}(\mathbf{P},\mathbf{I}) &\triangleq \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{P}_{i,j} - \mathbf{I}_{i,j})^2}}{N^2}.
\end{aligned}
\tag{3.7}
$$

We can "calibrate" these metrics against $\mathbf{P}_{\text{avg}}$ in order to provide meaningful comparisons of magnitude with our singular value metric. Specifically, evaluating the metrics between $\mathbf{P}_{\text{avg}}$ and $\mathbf{I}$, yields the following exact results:

$$
\begin{aligned}
M_{L1}(\mathbf{P}_{\text{avg}},\mathbf{I}) &\triangleq \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{P}_{\text{avg},i,j} - \mathbf{I}_{i,j}|}{N^2} \equiv \frac{2p}{N}, \\
M_{L2}(\mathbf{P}_{\text{avg}},\mathbf{I}) &\triangleq \frac{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{P}_{\text{avg},i,j} - \mathbf{I}_{i,j})^2}}{N^2} \equiv \frac{p}{N\sqrt{N-1}}.
\end{aligned}
$$

We can use the denominators of these results as modifying scale-factors on the respective metrics (relative to $\mathbf{I}$) in order to yield the same numerical results as with the singular value metric applied to $\mathbf{P}_{\text{avg}}$. In other words, we propose modified absolute deviation (L^1) and Euclidean (L^2) metrics, defined as follows:

$$
\begin{aligned}
M_{\text{dev}}(\mathbf{P}) &\triangleq \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{P}_{i,j} - \mathbf{I}_{i,j}|}{2N}, \\
M_{\text{euc}}(\mathbf{P}) &\triangleq \frac{\sqrt{N-1}}{N}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{P}_{i,j} - \mathbf{I}_{i,j})^2}.
\end{aligned}
\tag{3.8}
$$

It is straightforward to show that since each row of $\mathbf{P}$ must sum to unity, the quantity represented by M_{dev} is exactly equal to the average (across all rows) of the sum of the off-diagonal elements (per row) of $\mathbf{P}$. Equivalently, M_{dev} is exactly equal to $(N-1)$ multiplied by the average of all off-diagonal elements of $\mathbf{P}$. This averaging effect of M_{dev} smoothes out the differences in the off-diagonal content, thereby potentially violating $\boldsymbol{DD}$ and diminishing the usefulness of M_{dev} as a metric for comparing matrices with generally different off-diagonal distributions. The same applies for the metrics based on the eigenvalues of $\mathbf{P}$, as we will now demonstrate.

Let us compare M_{dev} and M_{euc} with M_{SVD} and with the eigenvalue-based metrics in Eq. (2.7). For example, for the general 2-d matrix in Eq. (2.3) the closed-form expressions for the respective metrics are given by

$$M_{\mathrm{SVD}} = \frac{1}{\sqrt{2}}\sqrt{p_1^2 + p_2^2}, \qquad M_{\mathrm{euc}} = \frac{1}{\sqrt{2}}\sqrt{p_1^2 + p_2^2}, \qquad M_{\mathrm{dev}} = \frac{1}{2}(p_1 + p_2),$$

$$M_{\mathrm{P}} = M_{\mathrm{D}} = M_{\mathrm{E}} = M_2 = p_1 + p_2.$$

It is clear that all metrics satisfy **M** and **DD** for this general 2-d case since any increase p_1 or p_2 will yield a larger M, thereby satisfying **M**, and any change in p_1 or p_2 will yield a different M, thereby satisfying **DD**, except for the special circumstance when either of p_1 or p_2 increases by the same amount as the other decreases.

Note that M_{SVD} and M_{euc} are identical for this general 2-d case, and, moreover, are sensitive to the *square* of each off-diagonal element, such that the largest off-diagonal element will dominate the result. By contrast, M_{P}, M_{D}, M_{E} and M_2 which are all identical for this general 2-d case, and differ from M_{dev} by only a constant scale-factor, are sensitive to the *linear* sum of the off-diagonal elements, which, in effect, means they are sensitive to the sum of the diagonal terms.

Consider now a (not completely general) 3-d example:

$$\mathbf{P} = \begin{pmatrix} 1 - p_1 & p_1 & 0 \\ 0 & 1 - p_2 & p_2 \\ 0 & p_3 & 1 - p_3 \end{pmatrix}.$$

The exact expressions for the respective metrics are given by (recalling the M_{SVD} result from Eq. (A.3) in Appendix A)

$$M_{\mathrm{SVD}} = \frac{\sqrt{2}}{3}\sqrt{p_1^2 + p_2^2 + p_3^2 + p_1\sqrt{3(p_2^2 + p_3^2)}}, \qquad M_{\mathrm{euc}} = \frac{2}{3}\sqrt{p_1^2 + p_2^2 + p_3^2},$$

$$M_{\mathrm{dev}} = \frac{1}{3}(p_1 + p_2 + p_3), \qquad M_{\mathrm{P}} = \frac{1}{2}(p_1 + p_2 + p_3),$$

$$M_{\mathrm{D}} = 1 - |(1 - p_1)(1 - (p_2 + p_3))|$$
$$= (p_1 + p_2 + p_3) - p_1(p_2 + p_3) \quad \text{for small } p_1, p_2, p_3,$$

$$M_{\mathrm{E}} = \frac{1}{2}(1 + p_1 - |1 - (p_2 + p_3)|)$$
$$= \frac{1}{2}(p_1 + p_2 + p_3) \quad \text{for small } p_1, p_2, p_3,$$

$$M_2 = 1 - \max\{|1 - (p_2 + p_3)|, |1 - p_1|\}$$
$$= \min\{(p_2 + p_3), p_1\} \quad \text{for small } p_1, p_2, p_3.$$

$$(3.9)$$

M_{SVD} and M_{euc} are sensitive to the squares of each off-diagonal term and are thereby dominated by the largest off-diagonal terms. Although M_{SVD} and M_{euc} are not generally identical, they are structurally similar. Notice that M_{SVD} includes the familiar

"sum of squares" of M_{euc} but has additional "cross-coupling" between p_1, p_2 and p_1, p_3. Without these additional terms, M_{euc} cannot satisfy the **DD** criterion.

Likewise, M_{dev}, M_P, M_D, M_E and M_2, though no longer identical, are structurally similar to one another – especially for small p_1, p_2, p_3 in which case the absolute-value delimiters in M_D, M_E, and M_2 can be eliminated, yielding simpler expressions. Moreover, to first order, M_{dev}, M_P, M_D and M_E are again sensitive to the linear sum of the off-diagonal terms, i.e. in effect, to the sum of the diagonal terms. M_2 is an exception, since, by definition, it uses only one of the eigenvalues and hence encapsulates limited information compared with the others.

The sensitivity of M_{SVD} and M_{euc} to the squares of the off-diagonal terms (and M_{dev}, M_P, M_D, M_E and M_2 to their linear sum) naturally extends to higher orders. An important consequence is that M_{SVD} and M_{euc} generally satisfy **DD**, since by design, they "seek out" and amplify the off-diagonal terms of **P**. By contrast, M_{dev}, M_P, M_D, M_E and M_2 "seek out" and amplify the diagonal terms of **P** which are generally of less interest for credit migration matrices.

We therefore expect to favor M_{SVD} (and M_{euc}) for comparing credit migration matrices since they are particularly sensitive to off-diagonal concentrations – which are of economic interest, particularly if they occur in the Default column. This expectation is borne out in the next section when we consider 8-d migration matrices from real data. Also, see Appendix A.3, where we present some numerical comparisons on low-order matrices.

4. Estimating migration matrices

Now that we have established how one might compare a set of migration matrices, we proceed with a discussion of how to estimate these matrices using firm credit rating histories. We then go on to apply the metrics to actual migration matrices.

4.1. Cohort approach

The method which has become the industry standard is the straight forward cohort approach. Let $p_{ij}(\Delta t)$ be the probability of migrating from grade i to j over horizon (or sampling interval) Δt. E.g. for $\Delta t = 1$ year, there are n_i firms in rating category i at the beginning of the year, and n_{ij} migrated to grade j by year-end. An estimate of the transition probability p_{ij} ($\Delta t = 1$ year) is $p_{ij} = \frac{n_{ij}}{n_i}$. Typically firms whose ratings were withdrawn or migrated to Not Rated (NR) status are removed from the sample. [14] The probability estimate is the simple proportion of firms at the end of the period, say at the end of the year for an annual matrix, with rating j having started out with rating i.

---

[14] The method which has emerged as an industry standard treats transitions to NR as non-informative. The probability of transitions to NR is distributed among all states in proportion to their values. This is achieved by gradually eliminating companies whose ratings are withdrawn. We use this method, which appears sensible and allows for easy comparisons to other studies.

Any rating change activity which occurs within the period is ignored, unfortunately. As we show below, there are other reasons to be skeptical of the cohort method providing an accurate and efficient estimate of the migration matrix. Since it is an industry standard, a statistical assessment seems crucial.

4.2. Duration or hazard rate approach (transition intensities)

One may draw parallels between ratings histories of firms and other time-to-event data such (un)employment histories and clinical trials involving treatment and response. In all cases one follows ''patients'' (be they people or firms) over time as they move from one state (e.g. ''sick'') to another (e.g. ''healthy''). Two other key aspects are found in credit rating histories: (right) *censoring* where we do not know what happens to the firm after the sample window closes (e.g. does it default right away or does it live on until the present) and (left) *truncation* where firms only enter sample if they have either survived long enough or have received a rating. Both of these issues are ignored in the cohort method.

A rich literature and set of tools exist to address these issues, commonly grouped under the heading of survival analysis. The classic text is Kalbfleisch and Prentice (1980) with more recent treatments covered in Klein and Moeschberger (1997) emphasizing applications in biology and medicine, and Lancaster (1990) who looks at applications in economics, especially unemployment spells. For applications to estimating credit migration matrices, see Kavvathas (2001) and Lando and Skodeberg (2002).

The formal construct is a N-state homogeneous Markov chain where state 1 refers to the highest rating, 'AAA', and state N is the worst, denoting default. For a time homogeneous Markov chain, the transition probability matrix is a function of the distance between dates (time) but not the dates themselves (i.e. where you are in time). Accepting or relaxing the time homogeneity assumption will dictate the specific estimation method.

With the assumption of time homogeneity in place, transition probabilities can be described via a $N \times N$ generator or intensity matrix $\boldsymbol{\Gamma}$. [15] Following Lando and Skodeberg (2002), the $N \times N$ transition probability matrix $\mathbf{P}(t)$ can be written as

$$\mathbf{P}(t) = \exp(\boldsymbol{\Gamma}t), \quad t \geqslant 0, \tag{4.1}$$

where the exponential is a matrix exponential, and the entries of $\boldsymbol{\Gamma}$ satisfy

$$\begin{aligned} \gamma_{ij} &\geqslant 0 \quad \text{for } i \neq j, \\ \gamma_{ii} &= -\sum_{j \neq i} \gamma_{ij}. \end{aligned} \tag{4.2}$$

The second expression merely states that the diagonal elements are such to ensure that the rows sum to zero.

---

[15] Lando and Skodeberg (2002) point out that it is only for the case of time homogeneity that one gets a simple formulaic mapping from intensities to transition probabilities.

We are left with the task of obtaining estimates of the elements of the generator matrix $\mathbf{\Gamma}$. The maximum likelihood estimate of γ_{ij} is given by

$$\widehat{\gamma}_{ij} = \frac{n_{ij}(T)}{\int_0^T Y_i(s)\,\mathrm{d}s}, \tag{4.3}$$

where $Y_i(s)$ is the number of firms with rating i at time s, and $n_{ij}(T)$ is the total number of transitions over the period from i to j where $i \neq j$. The denominator effectively is the number of "firm years" spent in state i. Thus for a horizon of one year, even if a firm spent only some of that time in transit, say from 'AA' to 'A' before ending the year in 'BBB', that portion of time spent in 'A' will contribute to the estimation of the transition probability $P_{\mathrm{AA} \to \mathrm{A}}$. In the cohort approach this information would have been ignored. Moreover, firms which ended the period in an 'NR' status are still counted in the denominator, at least the portion of the time which they spent in state i.

A common assumption for credit modeling (either at the instrument or portfolio level) is for the system to be first-order Markov. However, Carty and Fons (1993), Altman and Kao (1992), Altman (1998), Nickell et al. (2000), Bangia et al. (2002), Lando and Skodeberg (2002) and others have shown the presence of non-Markovian behavior such as ratings drift, and time non-homogeneity such as sensitivity to the business cycle. Realistically the economy (and hence the migration matrix) will change on time-scales far shorter than required to reach the idealized Default steady-state proscribed by an assumed constant migration matrix.

The migration matrix can also be estimated using non-parametric methods such as the Aalen–Johansen estimator which imposes fewer assumptions on the data generating process by allowing for time non-homogeneity while fully accounting for all movements within the sample period (or estimation horizon). [16] It is unclear, however, whether relaxing the assumption of time homogeneity results in estimated migration matrices which are different in any meaningful way, either statistical or economic.

We can get an early taste of the consequences of working with transition intensities by looking at estimates of the probability of default for a particular rating j (PD_j) using ratings histories for US firms from S&P. For the sample range we examine, 1981–2002, no defaults within one year were observed for rating class 'AAA.' The duration approach may still yield a positive probability of default for highly rated obligors even though no default was observed during the sampling period. It suffices that an obligor migrated from, say, 'AAA' to 'AA' to 'A−', and that a default occurred from 'A−' to contribute probability mass to $\mathrm{PD}_{\mathrm{AAA}}$. This can be seen by comparing the empirical PDs (in basis points – bp) in the first three columns of Table 2. For example, the estimated annual probability of default for an 'AA' company, $\mathrm{PD}_{\mathrm{AA}}$, is exactly zero for the cohort approach, 0.71 bp for the (parametric) time

---

[16] For details, see Aalen and Johansen (1978) and Lando and Skodeberg (2002).

Table 2
Estimated annual probabilities of default (PDs) in basis points (1981–2002), across methods

Rating categories	Cohort	Homogeneous	Non-homogeneous	$\%\frac{\text{Cohort}}{\text{Homog.}}$	$\%\frac{\text{Non-homog.}}{\text{Homog.}}$
AAA	0.000	0.020	0.003	0.00%	12.88%
AA+	0.000	0.049	0.035	0.00%	71.22%
AA	0.000	0.706	0.100	0.00%	14.16%
AA−	2.558	0.317	0.459	805.56%	144.43%
A+	5.942	0.380	0.545	1562.18%	143.39%
A	5.576	1.024	0.952	544.54%	92.99%
A−	4.403	0.854	0.545	515.76%	63.81%
BBB+	36.049	3.952	4.464	912.09%	112.96%
BBB	36.017	12.122	12.998	297.12%	107.22%
BBB−	45.519	17.517	22.509	259.86%	128.50%
BB+	51.378	28.817	30.910	178.29%	107.26%
BB	126.206	49.963	50.775	252.60%	101.63%
BB−	228.637	98.704	101.291	231.64%	102.62%
B+	363.529	198.279	211.548	183.34%	106.69%
B	1030.928	801.539	802.320	128.62%	100.10%
B	1460.674	1356.182	1371.660	107.70%	101.14%
CCC[a]	3092.243	4401.658	4224.314	70.25%	95.97%

S&P rated US obligors.

[a] Includes CC and C rated obligors.

homogeneous and 0.10 bp for the (non-parametric) non-homogeneous duration approach.[17] For an 'A+' rated firm, PD_{A+} is 5.9 bp for the cohort approach but a much smaller 0.38 and 0.55 bp for time homogeneous and non-homogeneous duration approach respectively, meaning that the less efficient cohort method *over* estimates default risk by about 10-fold. We would obtain these lower duration based probability estimates if firms spend time in the 'A+' state during the year on their way up (down) to a higher (lower) grade from a lower (higher) grade. This would reduce the default intensity, thereby reducing the default probability.

At the riskiest end of the spectrum, CCC-rated companies, the differences are striking: 30.92% for the cohort method but 44.02% and 42.24% for homogeneous and non-homogeneous duration respectively. Thus using the more popular but less efficient cohort method would *under* estimate default risk by over 10% points. One way we might see such differences is if firms spend rather little time in the 'CCC' state which would yield a small denominator in the rating intensity expression (for either homogeneous or non-homogeneous duration) and hence a higher default probability.[18]

---

[17] Brand and Bahar (2001) report non-zero PDs for high grades by smoothing and extrapolating S&P default experiences.

[18] Frydman (2003) estimates Markov mixture models and finds evidence of a two-regime process. The results for 'CCC' are especially interesting as they suggest that firms starting in 'CCC' are much less likely to default than firms downgraded to 'CCC'.

4.3. Bootstrapping

The estimates of the transition matrices are just that: estimates with error (or noise). Let $\widehat{\mathbf{P}}_a$ be an estimate of the migration matrix $\mathbf{P}$ using method a; consequently we may denote $\widetilde{\mathbf{P}}_a = \widehat{\mathbf{P}}_a - \mathbf{I}$ to be an estimate of the mobility matrix (see (3.2)). Then distance metrics such as SVD-based $\Delta M_{\text{SVD}}(\widetilde{\mathbf{P}}_a, \widetilde{\mathbf{P}}_b) \triangleq M_{\text{SVD}}(\widetilde{\mathbf{P}}_a) - M_{\text{SVD}}(\widetilde{\mathbf{P}}_b)$ are also noisy. In order to help us answer the question "how large is large" for a distance metric such as $\Delta M_{\text{SVD}}(\widetilde{\mathbf{P}}_a, \widetilde{\mathbf{P}}_b)$, we need its distributional properties. In the absence of any asymptotic theory a straightforward and efficient way is through the resampling technique of bootstrapping.

Consider, for example, $\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}(t)$ and $\widehat{\widetilde{\mathbf{P}}}_{\text{hom}}(t)$ as the cohort and homogeneous duration estimates at time t respectively, obtained using n_t observations. [19] Suppose we create k bootstrap samples [20] of size n_t each so that we can compute a set of R differences based on singular values, $\{\Delta M_{\text{SVD}}^{(j)}(\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}(t), \widehat{\widetilde{\mathbf{P}}}_{\text{hom}}(t))\}_{j=1}^{R}$ where $j = 1, \ldots, R$ denotes the number of bootstrap replications. This will give us a bootstrap distribution of singular value based distances. For a chosen critical value α (say $\alpha = 5\%$), we see if 0 falls within the $1 - \alpha$ range of $\{\Delta M_{\text{SVD}}^{(j)}(\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}(t), \widehat{\widetilde{\mathbf{P}}}_{\text{hom}}(t))\}_{j=1}^{R}$ for some relatively large R (≈ 1000). [21]

Ideal conditions for the bootstrap require that the underlying data is a random sample from a given population. Specifically the data should be independently and identically distributed (iid). Broadly one may think of at least two sources of heterogeneity: cross-sectional and temporal. It is difficult to impose temporal independence across multiple years, but easier at shorter horizons such as one year. We will still be subject to the effects of a common (macro-economic) factor, but this problem can be mitigated by focusing the analysis on either expansion or recession years only. [22] We will come back to this particular issue in Section 5.2. Sources of cross-sectional heterogeneity may be country, type of entity (corporation, government), and for corporations, industry. Nickell et al. (2000) document that for

---

[19] To be sure, with the presence of transitions to NR, the number of observations is not identical for the two methods: the cohort method drops them, the duration methods do not.

[20] A bootstrap sample is created by sampling *with replacement* from the original sample. For an excellent exposition of bootstrap methods, see Efron and Tibshirani (1993).

[21] Efron and Tibshirani (1993) suggest that for obtaining standard errors of bootstrapped statistics, bootstrap replications of 200 are sufficient. For confidence intervals, they suggest bootstrap replications of 1000 which we employ. Andrews and Buchinsky (1997) explore the impact of non-normality on the number of bootstraps. With multimodality and fat tails the number of bootstrap replications often must increases 2- or 3-fold relative to the Efron and Tibshirani benchmarks. For several cases we increased R to 3000 and found no meaningful evidence of non-normality: all densities were unimodal with average kurtosis around 3.2, ranging from 2.8 to 6.4. In fact the bootstrap distributions of ΔM_{SVD} were surprisingly close to normal.

[22] Similarly Christensen et al. (2004) perform their bootstrap simulations by dividing their sample into multiyear "stable" and "volatile" periods. See also Lopez and Saidenberg (2000) for a related discussion on evaluating credit models.

corporations, country of domicile and industry influence ratings migration persistence (momentum). We are able to control for some but not all of these factors. We restrict our analysis to US firms, i.e. no government entities (municipal, state or sovereign), and no non-US entities, but do not perform separate analysis by industry largely for reasons of sample size. By mixing industries together, the resulting bootstrap samples will likely be noisier than they would be otherwise, biasing the analysis against finding differences. [23]

5. Empirical results

Our data set of S&P ratings histories, CreditPro V. 6.2, is very similar to the data used in Bangia et al. (2002) but covers an additional four years: the total sample ranges from January 1, 1981 to December 31, 2002. The universe of obligors is mainly large corporate institutions around the world. Ratings for sovereigns and municipals are not included, leaving the total number of unique obligors to be 9929. The share of the most dominant region in the data set, North America, has steadily decreased from 98% to 60%, as a result of increased coverage of companies domiciled outside US. The database has a total of 60,133 obligor years of data, excluding withdrawn ratings, of which 1059 ended in default yielding an average default rate of 1.76% for the entire sample. For our analysis we will restrict ourselves to US obligors only; there are 6776 unique US domiciled obligors in the sample.

5.1. Comparing the metrics

To illustrate the empirical applications of the metrics proposed here, we first compare the annual ($\Delta t = 1$ year) migration matrices estimated via the parametric (time homogeneous) duration method for the years 1981–2002, and then compare each annual matrix to the average matrix.

We obtain 22 matrices for each migration estimation method. Fig. 1 contains the various metrics computed for the homogeneous matrices. It is apparent that all the metrics except for M_2 are highly correlated, consistent with them satisfying $\boldsymbol{M}$ for the matrices in question. Actual correlations bear this out, with M_{dev}, M_P and M_E being perfectly correlated since the metrics differ only by constant scale-factors. It is also clear from Fig. 1 that M_{SVD}, M_{dev}, M_P and M_E have roughly the same magnitudes whereas M_2 is consistently smaller (because the matrices are diagonally dominant with second eigenvalues close to unity); M_{euc}, is consistently larger; and M_D is consistently significantly larger.

Focusing our attention just on M_{SVD} we compare the "size" of the migration matrices across the different estimation methods. In Fig. 2 we can clearly see that matrices have been getting "larger" since the mid-1990s and that they tend to

---

[23] To be sure, the methodology could be used to actually test whether mixing across industries matters, once country of domicile and economic regime are controlled for.
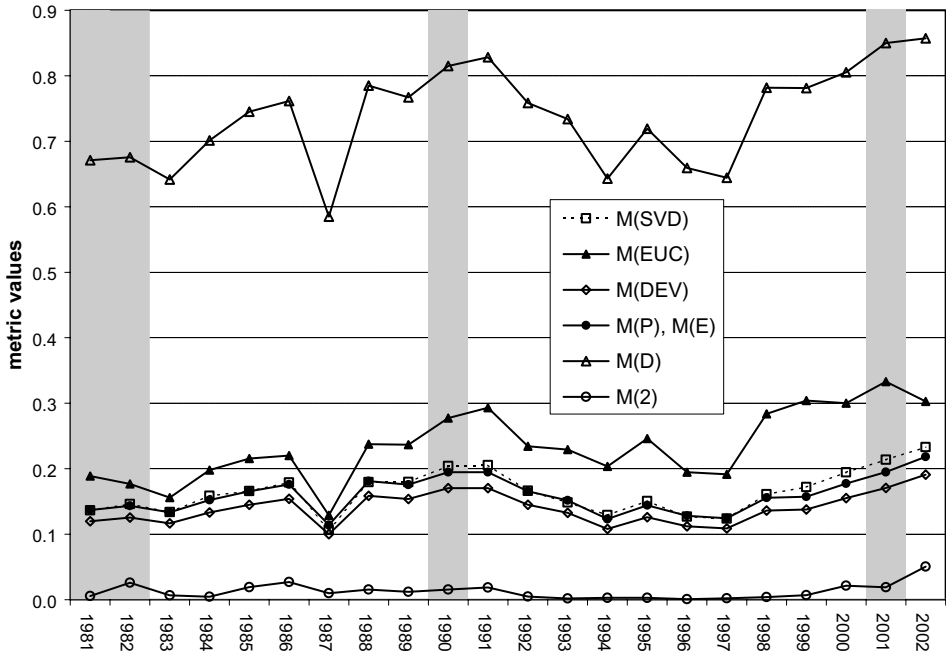
Fig. 1. Comparison of matrix metrics: S&P, US, 1981–2002, using the duration based (time homogeneous) method. Metrics are defined in Eq. (2.7) for M_D, M_E, M_P and M_2, Eq. (3.4) for M_{SVD}, and Eq. (3.8) for M_{euc} and M_{dev}. Shaded areas correspond to NBER recession periods.

increase leading up to and during recessions. The most recent year available, 2002, has generated the "largest" migration matrix. Moreover, the matrices estimated with the cohort method tend to be "smaller" than the duration matrices, and this difference seems to be increasing recently. Relaxing time homogeneity appears to have only a very small effect. These casual observations are borne out using bootstrap methods in Section 5.2 below.

Despite having 22 years of migration data with nearly 6800 unique US obligors, not all cells of a single-year migration matrix are estimated with high precision; the further away from the diagonal, the fewer observations. Most of the migrations are of one to two grades. As a result, practical applications of migration matrices tend to make use of longer time spans or averages over the entire sample range. [24] But how different are particular years from that long-run average? Fig. 3 depicts the deviation of each annual migration matrix (P_{hom}) from the full-sample (long-run average) matrix ($\overline{\mathbf{P}}_{hom}$) using the more efficient homogeneous duration estimation method. Specifically, the quantity plotted is given by (for each year) $\Delta M_{SVD}(\widetilde{\mathbf{P}}_{hom}(t), \widetilde{\overline{\mathbf{P}}}_{hom})$. The metric reveals that the amount of variation over time is substantial, with migration

---

[24] This is sometimes called the unconditional migration matrix in that it does not condition on, say, a particular year or point in the business cycle.
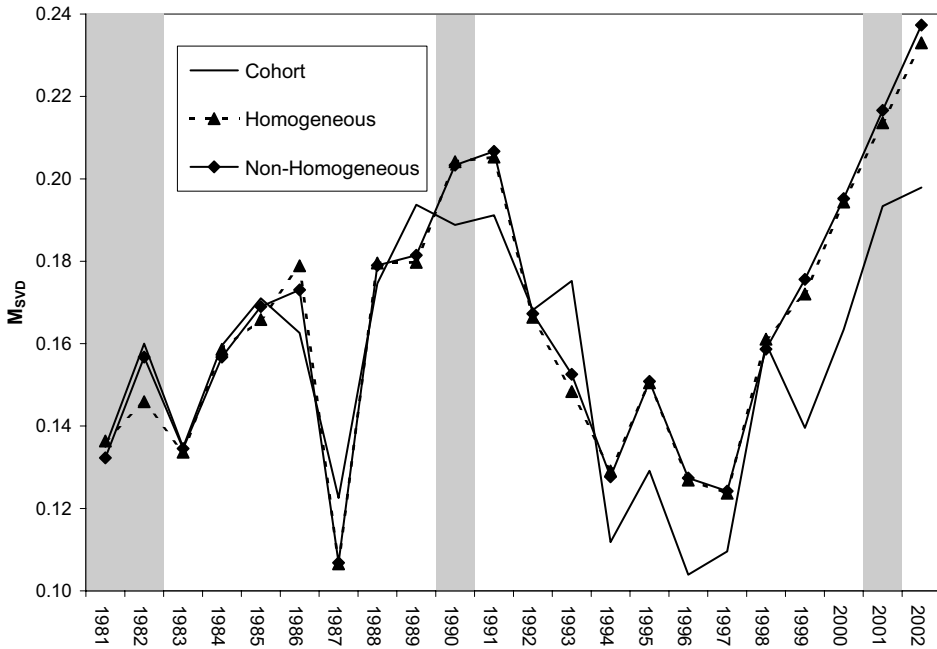
Fig. 2. M_{SVD} of annual mobility matrices, comparing three estimation methods: cohort and two duration based methods, parametric (imposing time homogeneity) and non-parametric (relaxing time homogeneity). S&P, US, 1981–2002. Shaded areas correspond to NBER recession periods.

matrices in the last three years being consistently "larger" than the average, and "smaller" for most of the 1990s. The largest positive deviation occurred in 2002, the largest negative deviation in 1987, the latter being actually a little larger in absolute value: -0.0650 vs. 0.0615. If nothing else it suggests that the underlying Markov process is unlikely to be time homogeneous.

5.2. Statistical differentiation

In this section we compare the different estimation methods using the SVD metric M_{SVD} on migration matrices estimated for a one-year horizon which is typical for many risk management applications. We show that the method matters in often dramatic ways. The difference between the duration methods are much smaller than between cohort and duration methods, implying that using the two efficient duration method, even with the (possibly false) assumption of time homogeneity over the cohort method has a far greater impact than relaxing the time homogeneity assumption.

In Fig. 4 we display ΔM_{SVD} of the cohort and non-homogeneous duration relative to the homogeneous duration. It is apparent that the difference between cohort-based and homogeneous duration-based matrices (dashed line) are larger than the
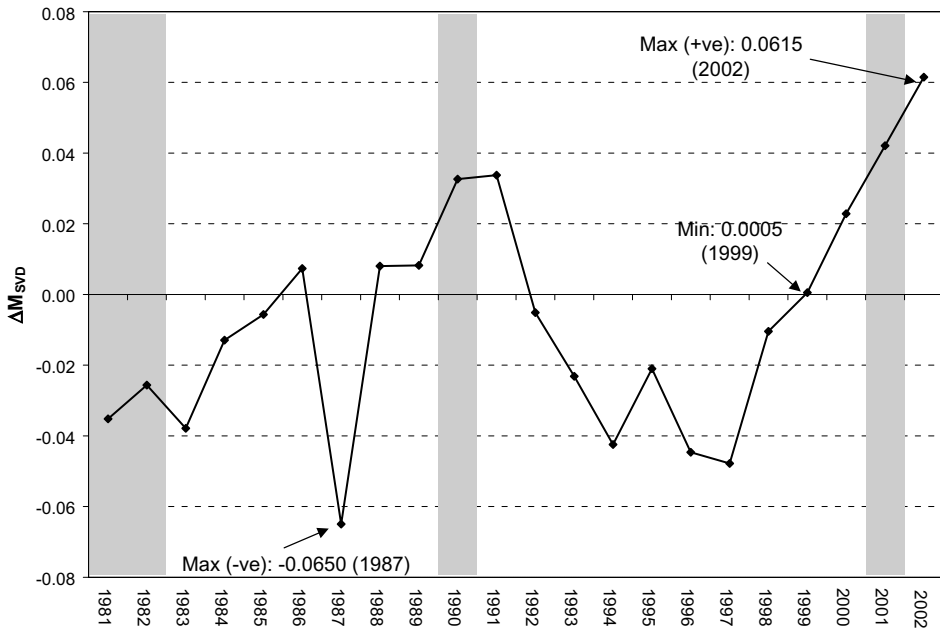
Fig. 3. Using differences in M_{SVD} to assess the time evolution of the deviation of the annual migration matrix from the full-sample (long-run average) estimated matrix, 1981–2002. S&P, US, homogeneous duration. Shaded areas correspond to NBER recession periods.

differences of the two variations of the duration method (solid line). Moreover, the migration matrices estimated non-parametrically (i.e. allowing for time non-homo-geneity) are typically larger than the parametric duration estimates (the solid line is typically in positive territory), while the cohort matrices appear to be the smallest (the dashed line is typically in negative territory, especially in the last 7–8 years of the sample). Duration-based migration matrices estimated non-parametrically exhibit more mobility than cohort-based matrices.

The largest difference between cohort and homogeneous duration methods occurs in 2002 while the smallest difference is found in 1984. In absolute value, its minimum occurs in 1984 (0.00094) and its maximum in 2002 (−0.03516). Is either different from zero, meaning do the two methods generate statistically indistinguishable tran-sition matrices? Table 3 (left column) provides some summary statistics of the boot-strap, including several quantiles. Indeed we are unable to reject that the 1984 matrices are different (0 is near the median) but can do so for the 2002 matrices: the 98% confidence interval from the 1st to the 99th percentile is (−0.05144, −0.01931).

Moving on to the comparison of the parametric vs. non-parametric estimation of the duration-based matrices, we see that the largest difference, in absolute value, occurs in 1982 (0.01082), and the smallest in 1987 (0.00026). The minimum difference between cohort and homogeneous duration occurs in 1984, a year where
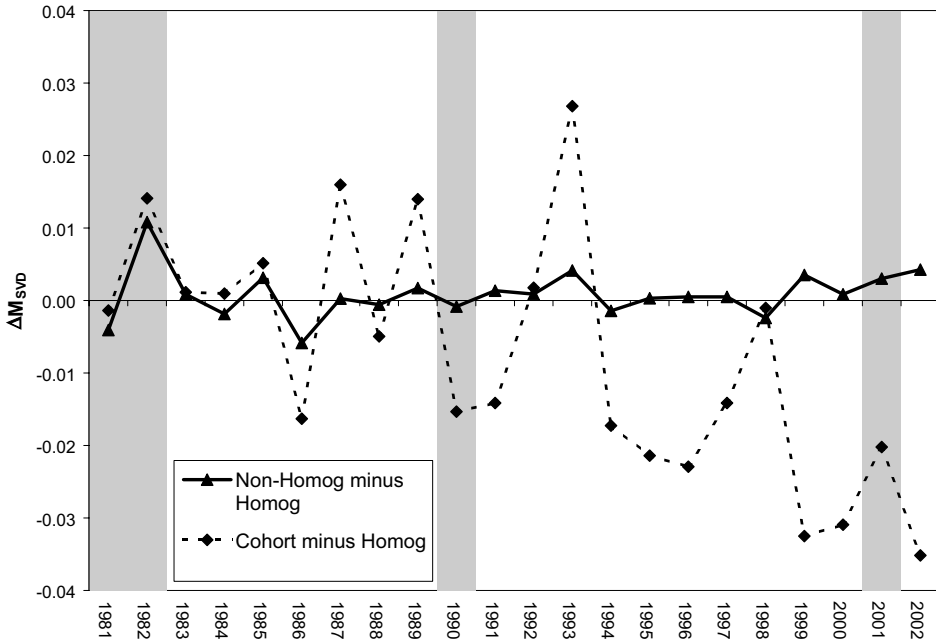
Fig. 4. ΔM_{SVD} assessing differences in migration matrices estimated using different estimation methods, always relative to the time homogeneous (parametric) duration estimator. S&P, US, 1981–2002. Shaded areas correspond to NBER recession periods.

the difference between the two duration methods is also quite small: -0.00186. Table 3 (right panel) shows the bootstrap results for 1987 (min difference) and 1982 (max difference). Even for the year of maximum difference between duration methods, namely 1982, we are unable to reject the hypothesis that the difference is zero. The 98% confidence interval from the 1st to the 99th percentile is $(-0.00174, 0.03892)$. For 1987 the zero is contained already in the 90% confidence interval $(-0.00059, 0.00120)$.

Bootstrapping requires that the underlying data not exhibit dependence and come from the same distribution. An obvious source of heterogeneity for credit migrations are business cycle regimes. Thus years where the economy moved from expansion to recession or vice versa will result in a mixture of regimes and cast some doubt on the bootstrap results. This is the case for 1982, the maximum ΔM_{SVD} (at 0.01082) between parametric and non-parametric estimates of migration matrices. To assess the robustness of those results, we took the next largest (in absolute value) ΔM_{SVD} year, 1986 (at -0.00585) which was not a transition year (it was, in fact, an expansion year) and repeated the bootstrap exercise. The results did not change. We cannot reject the null of no difference; zero is easily contained in the 90% confidence interval which is $(-0.01563, 0.00618)$.

The bootstraps for the non-parametric method are extremely computationally intensive, but not so for the other two methods. Thus in Fig. 5 we can display the 95% confidence band for ΔM_{SVD} between the cohort and parametric duration meth-

Table 3
Bootstrapped SVDs: $\Delta M_{\text{SVD}}(\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}, \widehat{\widetilde{\mathbf{P}}}_{\text{hom}})$, $\Delta M_{\text{SVD}}(\widehat{\widetilde{\mathbf{P}}}_{\text{non-h}}, \widehat{\widetilde{\mathbf{P}}}_{\text{hom}})$, $(R = 1000)$

Stats	Cohort − homogeneous duration		Non-homogeneous − homogeneous duration		
	1984 (min)	2002 (max)	1987 (min)	1982 (max)	1986 (second largest)
$\widehat{M}_{\text{SVD}}$	0.00094	−0.03516	0.00026	0.01082	−0.00585
Mean	0.00103	−0.03491	0.00030	0.01071	−0.00535
Std. dev.	0.00910	0.00701	0.00059	0.00834	0.00686
Q1	−0.02546	−0.05144	−0.00121	−0.00174	−0.02082
Q5	−0.01500	−0.04648	−0.00059	0.00037	−0.01563
Q50	0.00225	−0.03491	0.00029	0.00916	−0.00580
Q95	0.01329	−0.02338	0.00120	0.02599	0.00618
Q99	0.01679	−0.01931	0.00164	0.03829	0.01294

Quantiles from bootstrapping the SVD-based distance metric using S&P rated US obligors, 1981–2002. Zero is contained in each confidence interval except for cohort minus homogeneous duration in 2002 which is the maximum difference between the two methods.

od for each year over the entire sample period. Zero lies inside the band for much of the 1980s and outside it for the last four years.

What becomes clear is that the differences (in SVD terms) between the duration methods are much smaller than between cohort and duration methods, [25] implying that using the efficient duration method, even with the (possibly false) assumption of time homogeneity, over the cohort method has a far greater impact than relaxing the time homogeneity assumption. The smoothing imposed by the parametric maximum likelihood estimator (4.3) appears to be modest. In Table 4 we make a formal comparison of means between the different methods. The mean absolute difference between the cohort and the homogeneous duration, $\Delta M_{\text{SVD}}(\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}, \widehat{\widetilde{\mathbf{P}}}_{\text{hom}})$, is 0.0149, the mean absolute difference between cohort and non-homogeneous duration, $\Delta M_{\text{SVD}}(\widehat{\widetilde{\mathbf{P}}}_{\text{coh}}, \widehat{\widetilde{\mathbf{P}}}_{\text{n-hom}})$, is 0.0143, while the mean difference of the two duration methods $\Delta M_{\text{SVD}}(\widehat{\widetilde{\mathbf{P}}}_{\text{n-hom}}, \widehat{\widetilde{\mathbf{P}}}_{\text{hom}})$ is a much smaller 0.0024. Indeed we cannot reject that the mean absolute difference between the cohort and either duration method is different (from zero) with a *p*-value of 0.33, but we can do so for the difference between cohort and either duration method and the average difference between the two duration methods. We show one of them in Table 4 (the other test yields the same result), where the *p*-value is <0.001, allowing us to strongly reject that the two mean absolute differences are the same.

The degree of divergence between the cohort and either duration method is obviously a function of the time horizon over which the migration matrices are estimated. The longer that horizon, the more migration potential there is. Hence we would expect these differences to be smaller for shorter horizons such as semi-annually or quarterly. Our focus is on the one-year horizon as that is typical for many credit applications.

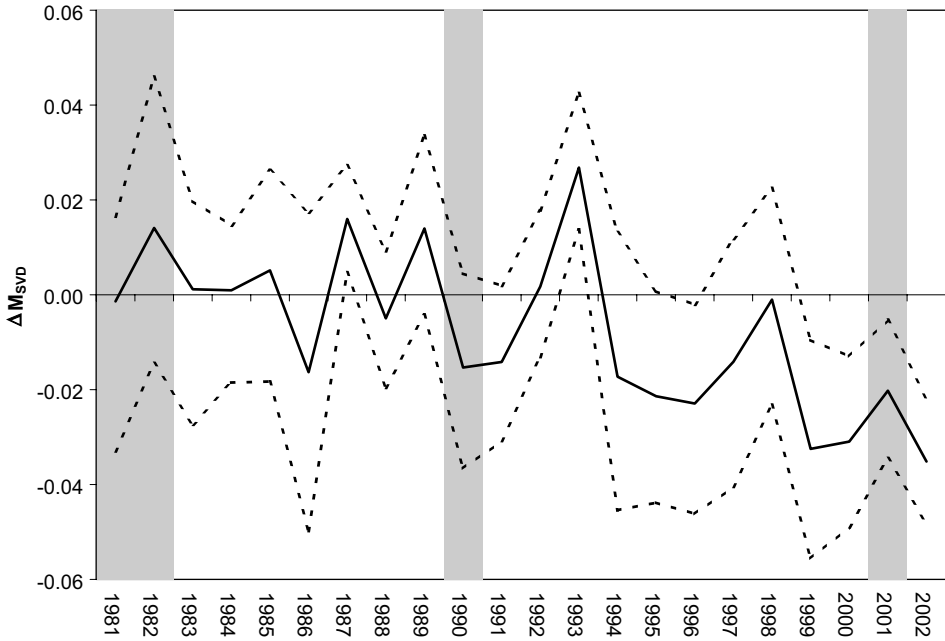[25] These results confirm a conjecture in Lando and Skodeberg (2002).

Fig. 5. Bootstrapped 95% confidence bands of $\Delta M_{\mathrm{SVD}}(\widehat{\widehat{\mathbf{P}}}_{\mathrm{coh}}, \widehat{\widehat{\mathbf{P}}}_{\mathrm{hom}})$, assessing the difference between cohort and homogeneous duration estimation methods for each yearly migration matrix. S&P, US, 1981–2002. Shaded areas correspond to NBER recession periods.

Table 4
Difference in mean absolute differences of M_{SVD} between methods: $\Delta M_{\mathrm{SVD}}(\widehat{\widehat{\mathbf{P}}}_{\mathrm{coh}}, \widehat{\widehat{\mathbf{P}}}_{\mathrm{hom}})$, $\Delta M_{\mathrm{SVD}}(\widehat{\mathbf{P}}_{\mathrm{coh}}, \widetilde{\mathbf{P}}_{\mathrm{n-hom}})$, $\Delta M_{\mathrm{SVD}}(\widetilde{\mathbf{P}}_{\mathrm{n-hom}}, \widetilde{\mathbf{P}}_{\mathrm{hom}})$

	Cohort − homog. duration	Cohort − non-homog. duration	Cohort − homog. duration	Non-homog. − homog. duration
Mean	0.0149	0.0143	0.0149	0.0024
Std. dev.	0.0107	0.0118	0.0107	0.0025
Δmean		0.0125		0.0006
$F(\Delta\text{mean} = 0)$		0.824		18.938
$\Pr(F \leqslant f)$		0.330		<0.001

F-test assumes different variances. M_{SVD} estimated for each method for S&P rated US obligors, 1981–2002.

5.3. Economic relevance: Credit risk capital

The statistical difference between two empirically estimated transition matrices may not translate to economic significance. As an illustration we look at credit risk

capital levels implied by credit portfolio models which are used to generate value distributions of a portfolio of credit assets such as loans or bonds. [26]

The purpose of capital for a financial institution is to provide a cushion against losses. The amount of economic capital is commensurate with the risk appetite of the financial institution. This boils down to choosing a confidence level in the loss (or value change) distribution of the institution with which senior management is comfortable. For instance, if the bank wishes to have an annual survival probability of 99%, this will require less capital than a survival probability of 99.9%, the latter being typical for a regional bank (commensurate with a rating of about A−/BBB+). The loss (or value change) distribution is arrived at through internal credit portfolio models.

There are a variety of models which can be used to compute economic risk capital for a given portfolio of credit assets. [27] Consider now an example using one of the popular credit portfolio models, CreditMetrics®, where a cardinal input is the grade migration matrix as it describes the evolution of the portfolio's credit quality.

In an exercise similar to Bangia et al. (2002), we constructed a fictitious bond portfolio with 400 exposures with a current value of $415.9 MM. We did so by taking a random sample of rated US corporates that mimics the ratings distribution of the S&P US universe as of December 2002 in such a way that we have at least one obligor for each major industry group. Maturity ranges from 1 to 30 years, and interest is paid semi-annually or annually. We use preset mean recovery rates and their standard deviations from Altman and Kishore (1996) and take the yield curves and credit spreads as of August 1, 2003. We then ask the question: what is the portfolio value distribution one year hence using different transition matrices but leaving all other parameters [28] unchanged?

We summarize our findings in Tables 5 and 6. Three sets of numbers are displayed for each experiment: the standard deviation of horizon value (i.e. portfolio value one year hence) and VaR (value-at-risk) at 99% and 99.9%, the former being an oft-seen standard and the latter roughly corresponding to the default probability commensurate with an A−/BBB+ rating. The top panel of Table 5 compares the impact of business cycles, namely recession to expansion, which was shown in Bangia et al. (2002) to generate significant differences in risk capital. [29] We estimate the homogeneous duration matrices over the relevant sub-sample periods.

---

[26] For an application to credit derivatives, see Schuermann and Jafry (2003).

[27] For a review and comparison of many of these models, see Koyluoglu and Hickman (1998), Gordy (2000) and Saunders and Allen (2002).

[28] Parameters such as those governing the recovery process. For each scenario we generated 5000 trials using importance sampling.

[29] We use the NBER dates for delineating expansions and recessions.

Table 5
Credit risk capital: recession vs. expansion

	Recession	Expansion	$\% \frac{\text{Recession}}{\text{Expansion}}$		
Time homogeneous duration: recession vs. expansion ($\Delta m_{svd}(P_R, P_E) = 0.0434$)					
Mean horizon value	$392,853,876	$397,900,019	98.73%		
Std. dev. of value	$8,194,719	$6,706,755	122.19%		
VaR (99%)	$27,801,133	$22,959,290	121.09%		
VaR (99.9%)	$45,005,850	$38,965,277	115.50%		
	Cohort	Homog.	Non-homog.	$\% \frac{\text{Cohort}}{\text{Homog.}}$	$\% \frac{\text{Non-homog.}}{\text{Homog.}}$
1981–2002 average migration matrix by estimation method					
Mean horizon value	$397,694,458	$397,340,664	$396,982,381	100.09%	99.91%
Std. dev. of value	$7,971,401	$6,892,581	$7,038,120	115.65%	102.11%
VaR (99%)	$27,949,382	$23,646,358	$24,187,380	118.20%	102.29%
VaR (99.9%)	$48,415,318	$39,927,619	$40,508,559	121.26%	101.45%

Credit risk capital as computed by CreditMetrics™ using a 1-year horizon, 5000 replications (using their importance sampling option). All input parameters save migration matrices the same across runs. The sample portfolio is as described in Section 5.3. Recession and expansion matrices were estimated using monthly NBER business cycle classifications ("peak" and "trough" months were partitioned at the 15th of day of the month).

Here we broadly confirm results in Bangia et al. (2002) where capital held during a recession should be about 20–30% higher than during an expansion (21.09% at the 99% level, 15.5% at the 99.9% level). Moreover, the portfolio volatility is about 22% higher during a recession than an expansion. [30] This difference is about the same as the difference between the cohort and parametric duration method applied to the overall sample annual migration matrix, as can be seen from the bottom panel of Table 5. The capital difference for 99% VaR is about 18%, and at 99.9% VaR it is about 21%. By contrast, whether duration method is applied parametrically or non-parametrically makes little difference from the point of view of risk capital as can be seen in the last column of the bottom panel of Table 5. These differences are small, between 1.5% and 2.3%.

Although the duration method better captures the migration dynamics, and indeed typically has higher M_{SVD} values, the levels of capital implied by the cohort method is typically higher: most of the ratios (cohort to duration) are larger than 100%. The reason is simple: the cohort method actually tends to overestimate default probabilities, the last column of the migration matrix, relative to the duration methods. This is clearly seen in Table 2, where the last two columns takes the ratios of PDs by different methods. Ignoring the first three ratings ('AAA', 'AA+' and

---

[30] Including 2002 changes these results somewhat since 2002, while being classified as an expansion year by the NBER, looked more like a recession year considering the default experience in debt markets. Indeed, if we exclude 2002, then the differences are more pronounced: 24% at 99% VaR and 17% at 99.9% VaR.

Table 6
Credit risk capital: comparing estimation methods

	Cohort	Homog.	Non-homog.	$\%\frac{\text{Cohort}}{\text{Homog.}}$	$\%\frac{\text{Non}-\text{Homog.}}{\text{Homog.}}$
1982 (max ΔM_{SVD} for non-homog. − homog.)					
Mean horizon value	$396,416,966	$396,640,220	$396,301,440	99.94%	99.91%
Std. dev. of value	$8,700,887	$7,915,465	$7,923,430	109.92%	100.10%
VaR (99%)	$31,564,197	$28,569,182	$28,622,911	110.48%	100.19%
VaR (99.9%)	$53,324,594	$48,591,410	$48,108,281	109.74%	99.01%
1984 (min ΔM_{SVD} for cohort − homog.)					
Mean horizon value	$400,505,388	$400,731,919	$400,745,284	99.94%	100.00%
Std. dev. of value	$7,063,373	$6,632,943	$6,639,106	106.49%	100.09%
VaR (99%)	$24,966,449	$22,964,864	$22,982,175	108.72%	100.08%
VaR (99.9%)	$42,336,742	$38,677,713	$38,939,280	109.46%	100.68%
1987 (min ΔM_{SVD} for non-homog − homog.)					
Mean horizon value	$400,374,321	$401,102,917	$401,125,266	99.82%	100.01%
Std. dev. of value	$7,067,727	$6,434,677	$6,456,697	109.84%	100.34%
VaR (99%)	$24,535,519	$23,008,569	$22,907,149	106.64%	99.56%
VaR (99.9%)	$39,377,615	$38,735,753	$38,582,377	101.66%	99.60%
2002 (max ΔM_{SVD} for cohort − homog.)					
Mean horizon value	$393,734,801	$392,943,756	$392,866,927	100.20%	99.98%
Std. dev. of value	$9,829,981	$8,346,925	$8,236,980	117.77%	98.68%
VaR (99%)	$35,968,475	$28,671,849	$28,406,497	125.45%	99.07%
VaR (99.9%)	$60,070,064	$45,711,447	$45,180,273	131.41%	98.84%

Credit risk capital as computed by CreditMetrics™ using a 1-year horizon, 5000 replications (using their importance sampling option). All input parameters save migration matrices the same across runs. The sample portfolio is as described in Section 5.3.

'AA'), where the cohort estimate is identically equal to zero, the only rating category where cohort underestimates the PD is in the last one: 'CCC' (and below). [31]

The previous section highlighted several years in the sample which are of particular interest. They are 1982 (max ΔM_{SVD} for the two duration methods), 1984 (min ΔM_{SVD} for cohort and parametric (time homogeneous) duration), 1987 (min ΔM_{SVD} for the two duration methods) and finally 2002 (max ΔM_{SVD} for cohort and parametric (time homogeneous) duration). Table 6 summarizes the differences in economic capital for these years across methods. Without exception, the differences between the cohort and more efficient duration methods are larger than between the different duration methods, with differences of 10–30% for the former, and always less than 2% for the latter. Viewed through the lens of credit risk capital,

---

[31] The economic impact of a default is severe, much more so than a downgrade to some other rating. This suggests the desirability for devising a metric which somehow captures more specific locational aspects of the elements of the matrix, e.g. amplifying the effects of elements the closer they are to the Default column. This is a topic of current research by the authors whereby we adopt an information-theoretic approach to construct an alternative metric encompassing spatial gradients across the matrix.

ignoring the efficiency gain inherent in the duration methods is more damaging that making a (possibly false) assumption of time homogeneity.

Looking at 1982, the year where the two duration methods were most divergent, the difference in VaR capital is larger between the cohort and the homogeneous duration method (9–10%) than between the two duration methods themselves (<1%). This pattern persists when we move to 1984, where the divergence between cohort and homogeneous duration is the smallest. Even here the VaR differences are larger for cohort and homogeneous duration (6–9%) than between duration methods (<1%). The year 1987 is no exception.

The difference is startling when we look at the 2002, the year where we experienced the largest divergence between cohort and homogeneous duration methods. VaR differences are more than when comparing recession to expansion: 25% to over 30%.

Finally note that the differences in portfolio mean horizon value (i.e. the expected value of the portfolio one year hence) changes little across methods. So for example, in 2002 the difference in expected value of the portfolio between the cohort and homogeneous duration methods is essentially nil, but the difference in risk is substantial.

6. Conclusions

In this paper we presented several methods for measuring, estimating and comparing credit migration matrices. We look at three estimation methods for credit migration matrices: a popular but inefficient approach called cohort, and two efficient duration approaches, with and without the assumption of time homogeneity. We ask three questions: (1) how would one measure the scalar difference between these matrices; (2) how can one assess whether those differences are statistically significant; and (3) even if the differences are statistically significant, are they economically significant?

To help answer the first question, we develop a new metric based on singular values and show that this metric approximates the average probability of migration. The question of statistical significance is addressed using resampling methods, and economic relevance is addressed by simulating the credit risk capital levels implied by the credit portfolio model in CreditMetrics®. We find that indeed, the method matters, both statistically and economically, when analyzing migration matrices estimated for a one-year horizon which is typical for many risk management applications. For years where the singular value decomposition (SVD) metric is small we cannot reject the null that they are not different; for years where the SVD metric is large we are able to reject the null of no difference. Relaxing the time homogeneity assumption has little impact; even at its maximum, the two methods yield statistically indistinguishable migration matrices. Looking at the credit risk capital implied by the credit portfolio model we find that the differences between the

cohort and more efficient duration methods are larger than between the different duration methods, with differences of 15–30% for the former, and never more than 2% for the latter. Thus ignoring the efficiency gain inherent in the duration methods is more damaging than making a (possibly false) assumption of time homogeneity.

Which estimation method is the preferred one? It seems clear to us that the cohort is certainly not preferred over the duration approach. Although there is a lot of circumstantial evidence of time non-homogeneity of the underlying process, allowing for this in the duration-based estimation using the non-parametric Aalen–Johansen estimator seems to have very little impact. Computationally this non-parametric estimator is quite intensive, taking on average more than 100 times longer to compute than either the parametric duration or the cohort method. Thus our bottom line is a preference for the parametric duration estimator.

Appendix A

A.1. Matrix norm: 3-d example

For example, consider a 3-d migration matrix given by

$$\mathbf{P} = \begin{pmatrix} 1 - p_1 & p_1 & 0 \\ 0 & 1 - p_2 & p_2 \\ 0 & p_3 & 1 - p_3 \end{pmatrix}. \tag{A.1}$$

Note that this is not a completely general 3-d example (otherwise the algebra would become unwieldy). The corresponding $\widetilde{\mathbf{P}}$ and $\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}}$ matrices are given by

$$\widetilde{\mathbf{P}} = \begin{pmatrix} -p_1 & p_1 & 0 \\ 0 & -p_2 & p_2 \\ 0 & p_3 & -p_3 \end{pmatrix}, \qquad \widetilde{\mathbf{P}}'\widetilde{\mathbf{P}} = \begin{pmatrix} p_1^2 & -p_1^2 & 0 \\ -p_1^2 & p_1^2 + p_2^2 + p_3^2 & -(p_2^2 + p_3^2) \\ 0 & -(p_2^2 + p_3^2) & p_2^2 + p_3^2 \end{pmatrix}.$$

The eigenvalues of $\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}}$ are given by

$$
\begin{pmatrix}
p_1^2 + p_2^2 + p_3^2 + \sqrt{p_1^4 + p_2^4 + p_3^4 - p_1^2(p_2^2 + p_3^2) + 2p_2^2 p_3^2} \\
p_1^2 + p_2^2 + p_3^2 - \sqrt{p_1^4 + p_2^4 + p_3^4 - p_1^2(p_2^2 + p_3^2) + 2p_2^2 p_3^2} \\
0
\end{pmatrix}. \tag{A.2}
$$

Hence the norm of $\widetilde{\mathbf{P}}$ is given by (the square-root of the largest eigenvalue of $\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}}$)

$$
\|\widetilde{\mathbf{P}}\| = \sqrt{p_1^2 + p_2^2 + p_3^2 + \sqrt{p_1^4 + p_2^4 + p_3^4 - p_1^2(p_2^2 + p_3^2) + 2p_2^2 p_3^2}}.
$$

Now, the specific vector that is maximally amplified by $\widetilde{\mathbf{P}}$ (i.e. such that $\|\widetilde{\mathbf{P}}\mathbf{x}_{\max}\| = \|\widetilde{\mathbf{P}}\|\|\mathbf{x}_{\max}\|$, temporarily reverting to the equivalent but more familiar columnwise form) is in the direction (i.e. some multiple of) the eigenvector of $\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}}$ corresponding to the largest eigenvalue, i.e.

$$
\mathbf{x}_{\max} =
\begin{pmatrix}
p_1^2 - p_2^2 - p_3^2 + \sqrt{p_1^4 + p_2^4 + p_3^4 - p_1^2(p_2^2 + p_3^2) + 2p_2^2 p_3^2} \\
p_2^2 + p_3^2 - \sqrt{p_1^4 + p_2^4 + p_3^4 - p_1^2(p_2^2 + p_3^2) + 2p_2^2 p_3^2} \\
1
\end{pmatrix}.
$$

This maximally amplified vector is generally *not* representative of a feasible state vector. For example, if $p_1 = p_2 = p_3 = 0.1$, we obtain

$$
\mathbf{x}'_{\max} \approx (-0.01 \quad 0.02 \quad 1),
$$

which cannot correspond to a feasible direction (i.e. multiple of) a state vector since all state vector elements must be non-negative (in line with the probability definition).

Completing our 3-d example from (A.1), by taking the average of the square-roots of the eigenvalues in (A.2), the closed-form expression for the metric is given by [32]

$$
M_{\text{SVD}} = \frac{\sqrt{2}}{3} \sqrt{p_1^2 + p_2^2 + p_3^2 + p_1 \sqrt{3(p_2^2 + p_3^2)}}. \tag{A.3}
$$

A.2. Calibration of M_{SVD} against $\mathbf{P}_{avg}$

We present here the proof of the result stated in Eq. (3.6).

From the definition of $\mathbf{P}_{\text{avg}}$, (Eq. (3.5)) the corresponding mobility matrix, $\widetilde{\mathbf{P}}_{\text{avg}}$, can be expressed as

---

[32] Note that for general credit migration matrices, closed-form solutions are completely intractable and the metric must be computed numerically. This can be achieved with a single line of MATLAB® code, as follows: `m = mean(svd(P − eye(size(P))))`.

$$\widetilde{\mathbf{P}}_{\text{avg}} \triangleq \mathbf{P}_{\text{avg}} - \mathbf{I} \equiv \frac{p}{N-1} \begin{pmatrix} -(N-1) & 1 & \ddots & & & \\ 1 & -(N-1) & 1 & \ddots & & \\ \ddots & 1 & -(N-1) & 1 & \ddots & \\ & \ddots & & 1 & -(N-1) & \ddots \\ & & \ddots & & \ddots & \ddots \end{pmatrix}.$$

The matrix product $\widetilde{\mathbf{P}}'\widetilde{\mathbf{P}}$ used in the definition of the singular values is therefore given by

$$\widetilde{\mathbf{P}}'_{\text{avg}}\widetilde{\mathbf{P}}_{\text{avg}} \equiv \frac{p^2 N}{(N-1)^2} \begin{pmatrix} (N-1) & -1 & \ddots & & \\ -1 & (N-1) & -1 & \ddots & \\ \ddots & -1 & (N-1) & -1 & \ddots \\ & \ddots & & -1 & (N-1) & \ddots \\ & & \ddots & & \ddots & \ddots \end{pmatrix} \triangleq \frac{p^2 N}{(N-1)^2} \mathbf{A}_{\text{avg}}.$$

We must now compute the eigenvalues of the matrix $\mathbf{A}_{\text{avg}}$ above. To start, note that each column is linearly independent of any other. However, by inspection it can be seen that the first column is equal to the sum of all the others, for all N. This means that the rank of the matrix $\mathbf{A}_{\text{avg}}$ is equal to $(N-1)$, implying that exactly one of the N eigenvalues is equal to zero (see Strang, 1988, p. 250). Now, from the properties of eigenvalues we know that the sum of the eigenvalues of any matrix is equal to the trace of the matrix (i.e. the sum of the diagonal elements). Hence, for the matrix $\mathbf{A}_{\text{avg}}$:

$$\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_N \equiv \text{tr}(\mathbf{A}_{\text{avg}}) = N(N-1)$$

and incorporating the fact that exactly one eigenvalue equals zero (say $\lambda_1 = 0$, without loss of generality), we obtain

$$\lambda_2 + \lambda_3 + \cdots + \lambda_N = N(N-1).$$

This equation is satisfied identically if all the $(N-1)$ remaining eigenvalues are equal to N. Consider first the 2-d case. The characteristic equation (from the determinant) and the corresponding eigenvalue solutions are given by

$$\lambda^2 - 2\lambda = 0; \quad \lambda = [0, 2].$$

Likewise, for the 3-d and 4-d cases, respectively:

$$\lambda^3 - 6\lambda^2 + 9\lambda = 0; \quad \lambda = [0, 3, 3],$$
$$\lambda^4 - 12\lambda^3 + 48\lambda^2 - 64\lambda = 0; \quad \lambda = [0, 4, 4, 4].$$

Extending these findings to arbitrary order, we conclude that the remaining non-zero eigenvalues are all equal to N. The M_{SVD} metric applied to $\mathbf{P}_{avg}$ can therefore be expressed as

$$M_{SVD}(\mathbf{P}_{avg}) \triangleq \frac{\sum_{i=1}^{N} \sqrt{\lambda_i(\widetilde{\mathbf{P}}'_{avg} \widetilde{\mathbf{P}}_{avg})}}{N} = \frac{p}{\sqrt{N}(N-1)} \sum_{i=1}^{N} \sqrt{\lambda_i(\mathbf{A}'_{avg} \mathbf{A}_{avg})}$$

$$= \frac{p\sqrt{N}}{(N-1)} \left(\frac{0 + (N-1)\sqrt{N}}{N} \right) = p,$$

thereby proving the result in Eq. (3.6).

A.3. Comparisons of metrics applied to simple test matrices

As appreciated from Section 3.5, closed-form expressions for the various metrics rapidly become unwieldy, even for third-order matrices. We will therefore consider some numerical trials to compare the metrics.

Recall first the 3-d examples in Eq. (3.1), whereby the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are contrived to have the same diagonal elements, but different off-diagonal distributions. We demand that any proposed metric should be able to distinguish between these matrices (i.e. should satisfy **DD**). The corresponding results for M_{SVD}, M_{dev}, M_{euc}, M_P, M_D, M_E and M_2 are given by

$M_{SVD}(\mathbf{P}_1) = 0.3164;$ $M_{SVD}(\mathbf{P}_2) = 0.3463;$

$M_{dev}(\mathbf{P}_1) = 0.3;$ $M_{dev}(\mathbf{P}_2) = 0.3;$

$M_{euc}(\mathbf{P}_1) = 0.3197;$ $M_{euc}(\mathbf{P}_2) = 0.3590;$

$M_P(\mathbf{P}_1) = 0.45;$ $M_P(\mathbf{P}_2) = 0.45;$

$M_D(\mathbf{P}_1) = 0.7;$ $M_D(\mathbf{P}_2) = 0.7;$

$M_E(\mathbf{P}_1) = 0.45;$ $M_E(\mathbf{P}_2) = 0.45;$

$M_2(\mathbf{P}_1) = 0.4;$ $M_2(\mathbf{P}_2) = 0.4.$

Clearly M_{SVD} (and M_{euc}) both exhibit the desirable **DD** behavior: namely a difference in value for $\mathbf{P}_1$ and $\mathbf{P}_2$. Also, both vary in the same direction, and show an increase when the off-diagonal probability is concentrated (as in $\mathbf{P}_2$) rather than diluted (as in $\mathbf{P}_1$). This observation is consistent with M_{SVD} (and M_{euc}) being sensitive to the squares of the off-diagonal elements (discussed above). By contrast M_{dev}, M_P, M_D, M_E and M_2 give identical values for both matrices, thereby violating **DD** and making them less desirable metrics (at least for this example).

Likewise for a more extreme 5-d example

$$\mathbf{P}_1 = \begin{pmatrix} 0.5 & 0.2 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.5 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.5 \end{pmatrix}, \qquad \mathbf{P}_2 = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \end{pmatrix}.$$

The corresponding metric values are given by

$$M_{\text{SVD}}(\mathbf{P}_1) = 0.5028; \qquad M_{\text{SVD}}(\mathbf{P}_2) = 0.5785;$$
$$M_{\text{dev}}(\mathbf{P}_1) = 0.5; \qquad M_{\text{dev}}(\mathbf{P}_2) = 0.5;$$
$$M_{\text{euc}}(\mathbf{P}_1) = 0.5060; \qquad M_{\text{euc}}(\mathbf{P}_2) = 0.6325;$$
$$M_{\text{P}}(\mathbf{P}_1) = 0.625; \qquad M_{\text{P}}(\mathbf{P}_2) = 0.625;$$
$$M_{\text{D}}(\mathbf{P}_1) = 0.9808; \qquad M_{\text{D}}(\mathbf{P}_2) = 1;$$
$$M_{\text{E}}(\mathbf{P}_1) = 0.625; \qquad M_{\text{E}}(\mathbf{P}_2) = 0.625;$$
$$M_2(\mathbf{P}_1) = 0.6; \qquad M_2(\mathbf{P}_2) = 0.5.$$

Again, M_{SVD} and M_{euc} discriminate between $\mathbf{P}_1$ and $\mathbf{P}_2$ (with a larger value for the more extreme matrix, $\mathbf{P}_2$), whereas M_{dev}, M_{P} and M_{E} are "blind" to the variations in the distribution of the off-diagonal "mass". M_{D} and M_2 do discriminate between the two matrices in this example, though they did not in the last. Also, M_2 yields a larger value for $\mathbf{P}_1$ than for $\mathbf{P}_2$, inconsistent with the other metrics (M_{SVD}, M_{euc} and M_{D}) which yield higher values for the more extreme matrix, $\mathbf{P}_2$.

From the above examples, it is clear that M_{SVD} and M_{euc} are preferable to the others from the **DD** point of view. However, it is not immediately apparent which is preferable between M_{SVD} and M_{euc}. To answer this, consider the following two matrices which differ only in the permutation of the non-diagonal entries within each row:

$$\mathbf{P}_1 = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.7 & 0 \\ 0 & 0.4 & 0.6 \end{pmatrix}, \qquad \mathbf{P}_2 = \begin{pmatrix} 0.8 & 0 & 0.2 \\ 0 & 0.7 & 0.3 \\ 0.4 & 0 & 0.6 \end{pmatrix}. \tag{A.4}$$

The corresponding M_{SVD} and M_{euc} metric values are given by

$$M_{\text{SVD}}(\mathbf{P}_1) = 0.3463; \qquad M_{\text{SVD}}(\mathbf{P}_2) = 0.3407;$$
$$M_{\text{euc}}(\mathbf{P}_1) = 0.3590; \qquad M_{\text{euc}}(\mathbf{P}_2) = 0.3590.$$

Since M_{SVD} distinguishes between these matrices (i.e. satisfies **DD**) whereas M_{euc} does not, we therefore prefer M_{SVD} over M_{euc} on the grounds that it satisfies **DD** more generally than does M_{euc}. Note that we would ideally prefer that our metric yielded larger values for matrices whose off-diagonal content is distributed further from the diagonal since these intuitively represent greater mobility from one time step to the next (we may refer to this criterion as strong-distribution-discriminatory, **SDD**). For example, a metric satisfying **SDD** would yield a larger value for $\mathbf{P}_2$ than for $\mathbf{P}_1$ (from Eq. (A.4)), rather than the opposite behavior as exhibited by M_{SVD}. We have hitherto been unable to devise a suitable metric which satisfies **SDD** without introducing ad hoc pre-weighting of matrix elements. [33] This remains a topic of current research by the authors whereby we are attempting to adopt an information-theoretic approach to construct an alternative metric which satisfies **SDD**.

---

[33] For example, by pre-weighting each matrix element by $|i - j|$ (where i and j represent the row and column index, respectively) representing the distance of the given element from the diagonal.

References

Aalen, O.O., Johansen, S., 1978. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics 5, 141–150.

Acharya, V.V., Das, S.R., Sundaram, R.K., 2002. Arbitrage-free pricing of credit derivatives with rating transitions. Financial Analysts Journal 58 (3), 28–44.

Altman, E.I., 1998. The importance and subtlety of credit rating migration. Journal of Banking & Finance 22, 1231–1247.

Altman, E.I., Kao, D.L., 1992. Rating drift of high yield bonds. Journal of Fixed Income (March), 15–20.

Altman, E.I., Kishore, V.M., 1996. Almost everything you wanted to know about recoveries on defaulted bonds. Financial Analysts Journal (November/December), 57–64.

Andrews, D.W.K., Buchinsky, M., 1997. On the number of bootstrap repetitions for bootstrap standard errors, confidence intervals, and tests. Cowles Foundation Paper 1141R.

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

Bangia, A., Diebold, F.X., Kronimus, A., Schagen, C., Schuermann, T., 2002. Ratings migration and the business cycle, with applications to credit portfolio stress testing. Journal of Banking & Finance 26 (2/3), 445–474.

Bank for International Settlements, 2001. The New Basel Capital Accord. Available from <http://www.bis.org/publ/bcbsca.htm>.

Brand, L., Bahar, R., 2001. Corporate defaults: Will things get worse before they get better? S&P Special Report, January, pp. 5–40.

Carty, L.V., Fons, J.S., 1993. Measuring changes in corporate credit quality. Moody's Special Report, Moody's Investors Service, Inc., New York.

Christensen, J., Hansen, E., Lando, D., 2004. Confidence sets for continuous-time rating transition probabilities. Journal of Banking & Finance 28 (11), this issue.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Frydman, H., 2003. Estimation in the mixture of Markov chains moving with different speeds. Working paper, NYU Stern School of Business.

Geweke, J., Marshall, R.C., Zarkin, G.A., 1986. Mobility indices in continuous time Markov chains. Econometrica 54, 1407–1423.

Gordy, M.B., 2000. A comparative anatomy of credit risk models. Journal of Banking & Finance 24, 119–149.

Gupton, G.M., Finger, C.C., Bhatia, M., 1997. CreditMetrics™ – Technical Document his version: April 2, 1997. J.P. Morgan, New York.

Israel Robert, 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 (2), 245–265.

Jafry, Y., Schuermann, T., 2003. Metrics for comparing credit migration matrices. Wharton Financial Institutions Center Working Paper #03-09.

Jarrow, R.A., Turnbull, S.M., 1995. Pricing derivatives on financial securities subject to credit risk. Journal of Finance 50, 53–86.

Jarrow, R.A., Lando, D., Turnbull, S.M., 1997. A Markov model for the term structure of credit risk spreads. The Review of Financial Studies 10 (2), 481–523.

Kalbfleisch, J.D., Prentice, R.L., 1980. The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York.

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

Kijima, M., Komoribayashi, K., 1998. A Markov chain model for valuing credit risk derivatives. The Journal of Derivatives (Fall), 97–108.

Klein, J.P., Moeschberger, M.L., 1997. Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York.

Koyluoglu, H.U., Hickman, A., 1998. Credit risk: Reconcilable differences. Risk 11 (10), 56–62.

Lancaster, T., 1990. The Econometric Analysis of Transition Data. Cambridge University Press, Cambridge, UK.

Lando, D., Skodeberg, T., 2002. Analyzing ratings transitions and rating drift with continuous observations. Journal of Banking & Finance 26 (2/3), 423–444.

Lopez, J.A., Saidenberg, M., 2000. Evaluating credit risk models. Journal of Banking & Finance 24 (1/2), 151–167.

Marrison, C., 2002. The Fundamentals of Risk Measurement. McGraw-Hill, New York.

Nickell, P., Perraudin, W., Varotto, S., 2000. Stability of rating transitions. Journal of Banking & Finance 24, 203–227.

Saunders, A., Allen, L., 2002. Credit Risk Measurement, second ed. John Wiley & Sons., New York, NY.

Schuermann, T., Jafry, Y., 2003. Measurement and estimation of credit migration matrices. Wharton Financial Institutions Center Working Paper #03-09.

Shorrocks, A.F., 1978. The measurement of mobility. Econometrica 46, 1013–1024.

Strang, G., 1988. Linear Algebra and its Applications, third ed. Saunders HBJ, Orlando, FL.