Measurement and Estimation of Credit Migration Matrices[1]

Til Schuermann[2]

Federal Reserve Bank of New York
33 Liberty St.
New York, NY 10045
til.schuermann@ny.frb.org

Yusuf Jafry

Suite 322
355 Byres Rd.,
Glasgow G12 8QZ
UK
yjafry@alum.mit.edu

First Draft: January, 2002
This Draft: April 1, 2003
This Print: April 4, 2003

Abstract: Credit migration matrices are cardinal inputs to many risk management applications. Their accurate estimation is therefore critical. We explore three approaches, cohort and two variants of duration – time homogeneous and non-homogeneous -- and the resulting differences, both statistically through matrix norms and economically through credit portfolio and credit derivative models. We develop a testing procedure to assess statistically the differences between migration matrices using bootstrap techniques. The method can have substantial economic import: economic credit risk capital differences between economic regimes, recession vs. expansion, can be as large as difference implied by different estimation techniques. Ignoring the efficiency gain inherent in the duration methods by using the cohort method instead is more damaging that making a (possibly false) assumption of time-homogeneity.

Keywords: Credit risk, risk management, matrix norms, bootstrapping, credit derivatives

JEL Codes: C15, C41, G21, G28

---

[1] We would like to thank Rajiv Bahl and Chris Finger for making available the CreditMetrics™ software, Richard Johnson, Joshua Rosenberg and participants of the Banking Studies brown bag at the Federal Reserve Bank of New York for their insughtful comments, and Mei Kok for her excellent research assistance. All remaining errors are ours.

[2] Corresponding author. Any views expressed represent those of the author only and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

Credit migration or transition matrices, which characterize the 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, and pricing of credit derivatives. For example, in the New Basel Accord (BIS (2001)), capital requirements are driven in part by ratings migration. Their accurate estimation is therefore critical. We explore three approaches, frequentist (cohort) and two variants of duration (hazard) – time homogeneous and non-homogeneous -- using firm credit rating migration data from Standard and Poors (S&P). We compare the resulting differences, both statistically through matrix norms, eigenvalue and -vector analysis and economically through credit portfolio and credit derivative models. We develop a testing procedure based on singular value decomposition to assess statistically the differences (or distances) between Markov matrices using bootstrap techniques. The method can have substantial economic import: difference in economic credit risk capital implied by different estimation techniques can be as large as differences between economic regimes of recession vs. expansion. Viewed through the lens of credit risk capital, ignoring the efficiency gain inherent in the duration methods is more damaging that making a (possibly false) assumption of time homogeneity, a significant result given that the cohort method is the method of choice for most practitioners.

Perhaps the simplest use of a transition or migration matrix is for the valuation of a bond or loan portfolio which might be used by a portfolio or risk manager. Given a credit grade today, say BBB,[3] 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

---

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

illustrated in Table 1.[4] More sophisticated examples of risky bond pricing methods, such as outlined by Jarrow and Turnbull, (1995) and Jarrow, Lando and Turnbull, (1997), require these matrices as a cardinal input, as do credit derivatives such as the model by Kijima and Komoribayashi (1998). As a final example, credit portfolio models such as CreditMetrics™ (Gupton, Finger and Bhatia (1997)) used in risk management 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. Shorrocks (1978), looking at income mobility, proposed indices of mobility for Markov matrices using eigenvalues and determinants, a line of inquiry extended in Geweke, Marshall and Zarkin (1986). They present a set of criteria by which the performance of a proposed metric (for arbitrary transition matrices) should be judged. Jafry and Schuermann (2003) 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 parameters of the migration matrix as a characterization of estimation noise or uncertainty. Unsurprisingly they found that 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

[4] Default very rarely results in total loss.

t-tests were rarely significant.[5] Christensen and Lando (2002) 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. Nickell, Perraudin and Varotto (2000) use an ordered probit model to test for (and find) significance of different exogenous factors such as industry and the business cycle on transition probabilities. Lando and Skodeberg (2002), using duration models, find persistence and momentum in transition intensities. Arvanitis, Gregory and Laurent (1999) (hereafter referred to as AGL) propose to assess the similarity of all eigenvectors between two matrices by computing a (scalar) ratio of matrix norms. Specifically, their approach was motivated by the need to compare migration matrices with different horizons and test the first-order Markov assumption. They propose a cut-off value of 0.08 for their metric below which 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 what would be sufficiently large to reject similarity. Moreover, they ignore estimation noise and concomitant parameter uncertainty.

In a different line of research, Israel, Rosenthal and Wei (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 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 rest of the paper proceeds as follows. Section 2 sets up the context for the data generating mechanism by briefly describing credit ratings and the agencies which provide them. Section 3 describes the ratings data and different methods for estimating credit migration matrices; Section 4 covers different

[5] For this and other reasons, Bangia et al. (2002) found it very difficult to reject the first-order Markov property of credit migration matrices.

approaches to measuring matrix differences. Section 5 examines whether the empirical estimates are statistically distinguishable and whether they make material economic difference; and Section 6 provides some concluding remarks.

2. Credit Ratings and Rating Agencies

The market for credit ratings in the U.S. is dominated by two players: Standard & Poors (hereafter S&P) and Moody's Investor Services (hereafter Moody's); of the smaller rating agencies, only Fitch IBCA plays a significant role in the U.S. (though has a more substantial presence elsewhere) (Cantor and Packer (1995), White (2001)). Because of their broader coverage, S&P and Moody's data have been widely used in published studies. We will follow suit and use, purely out of convenience, the S&P ratings histories.

A credit rating by a credit rating agency represents an overall assessment of an obligor's creditworthiness. There is some difference between the rating agencies about what exactly is assessed.[6] Whereas S&P evaluates an obligor's overall capacity to meet its financial obligation, and is hence best thought of as an estimate of probability of default, Moody's assessment is said to incorporate some judgment of recovery in the event of loss. In the argot of credit risk management, S&P measures *PD* (probability of default) while Moody's measure is somewhat closer to *EL* (expected loss).[7]

Both rating agencies have seven broad rating categories as well as rating modifiers bringing the total to 19 rating classes, plus 'D' (default, an absorbing state[8]) and 'NR' (not rated – S&P) or 'WR'

---

[6] Morgan (2002) provides a measurement of rating difference as a proxy for firm opaqueness.

[7] Specifically, $EL = PD \times LGD$, where *LGD* is loss given default. However, given the paucity of *LGD* data, very little variation in *EL* at the obligor (as opposed to instrument) level can be attributed to variation in *LGD* making the distinction between the agencies modest at best.

[8] One consequence of default being an absorbing state arises when a firm re-emerges from bankruptcy. They are classified as a new firm.

(withdrawn rating – Moody's).[9] Typically ratings below 'CCC', e.g. 'CC' and 'C', are collapsed into 'CCC', reducing the total ratings to 17. Although the rating modifiers provide a finer differentiation between issuers within one letter rating category, they pose two problems: the sample size of issuers per rating class including rating modifiers is not sufficient for low rating categories, causing small sample size concerns that affect statistical inference. Moreover, transition matrices are generally published and applied without rating modifiers, as this format has emerged as an industry standard. Therefore, we exclude the rating modifiers in the course of this paper. So, for example, we consider 'BBB+' and 'BBB-' ratings as 'BBB' ratings. This methodology reduces the data from 17 to 7 rating categories, which ensures sufficient sample sizes for all rating categories.[10]

Ratings are costly: $25,000 for issues up to $500 million, ½ bp for issues greater than $500 million (Kliger and Sarig (2000)).[11] Treacy and Carey (2000) report that the usual fee charged by S&P is 0.0325% of the face amount. But the ratings are informative. Kliger and Sarig (2000) show that bond ratings contain price-relevant information by taking advantage of a natural experiment. On April 26, 1982, Moody's introduced overnight modifiers to their rating system, much like the notching used by S&P and Fitch, effectively introducing finer credit rating information about their issuer base without any change in the firm fundamentals. They find that bond prices indeed adjust to the new information, as do stock prices, and that any gains enjoyed by bondholders are offset by losses suffered by stockholders.

---

[9] The CCC (S&P) and Caa (Moody's) ratings contain all ratings below as well – except default, of course.

[10] For a discussion of ratings dynamics for 17 states, see Bahar and Nagpal (2000).

[11] Until the mid-70s, it was investors, not issuers, who paid fees to the rating agencies. (Partnoy (2002)).

3. Data and Estimation Methods

3.1. Rating Agency Data

Our analysis of S&P data[12] covers the period from January 1, 1981 to December 31, 2001 (Figure 1, left axis). 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 9,178. Summary statistics are provided in Table 2. 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 U.S. (see Figure 1, right axis). The obligors include both U.S. and non-U.S. industrials, utilities, insurance companies, banks and other financial institutions and real estate companies. The databases has a total of 55,010 obligor years of data excluding withdrawn ratings of which 840 ended in default yielding an average default rate of 1.53% for the entire sample. On average, investment grade rated obligors, i.e. having a rating no worse than BBB-, were 71% of the dataset. Figure 2 shows the average distribution by rating over the sample period. For most of the analysis in subsequent sections we will restrict ourselves to U.S. obligors; there are 6398 unique U.S. domiciled obligors in the sample. The principal reason is to allow for the estimation of matrices for U.S. business cycle regimes (recession/expansion).

To capture credit quality dynamics, the creditworthiness of obligors must be assessed, as credit events typically concern a firm as a whole. Unfortunately, published ratings focus on individual bond issues. Therefore, the rating agencies implement a number of transformations:

- Bond ratings are converted to issuer ratings. By convention, all bond ratings are made comparable by considering the implied long-term senior unsecured rating, i.e. the rating a bond would hold if it were senior unsecured. This rating is then considered the issuer rating.[13]

---

[12] We use the data in S&P's CreditPro product.

[13] See, for instance, the documentation for S&P's CreditPro 3.0 (Standard & Poors (2001)).

- Issuers are clustered into economic entities. This promotes correct representation of credit quality dynamics by accounting for parent-subsidiary links, mergers, acquisitions, and contractual agreements about recourse.

Finally the ratings data is censored. A total of 3,605 companies were classified as 'NR' ("not rated") from January 1981 to December 2001. Transitions to 'NR' may be due to any of several reasons, including expiration of the debt, calling of the debt, merger or acquisition, etc. Unfortunately, however, the details of individual transitions to 'NR' are not known. In particular, it is not known whether any given transition to 'NR' is "benign" or "bad." Bad transitions to 'NR' occur, for example, when a deterioration of credit quality known only to the bond issuer (debtor) leads the issuer to decide to bypass an agency rating. Carty (1997), using Moody's data from 1920-96, claims that only 1% of all rating withdrawals may have been due to deteriorating credit quality.

3.2. Estimating Migration Matrices

Conditional upon a given grade at time T, the transition, or migration, matrix **P** is a description of the probabilities of being in any of the various grades at $T+1$. It thus fully describes the probability distribution of grades at $T+1$ given the grade at T. We seek to estimate the (7x7)+7 = 56 unique elements of **P**, a conceptual rendition of which appears in Figure 3.

Jarrow, Lando and Turnbull (1997) make the distinction between implicit and explicit (or historical) estimation of transition matrices, where implicit estimation refers to extracting transition and default information from market prices of risky zero-coupon bonds. We will consider only the different explicit methods in this paper.[14]

[14] See also Kavvathas (2001) for a survey of approaches for estimating credit migration matrices.

3.2.1. Frequentist Approach (Cohort Method)

The method which has become the industry standard is the straight forward frequentist (or 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 Δt = 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 $P_{ij}^{(\Delta t=1yr)}$ is $P_{ij}^{(1)} = \dfrac{N_{ij}}{N_i}$. Typically firms whose ratings were withdrawn or migrated to Not Rated (NR) status are removed from the sample.[15] 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.

Any rating change activity which occurs within the period is ignored, unfortunately. As we show in Section 3.2.2 below, there are many more 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.

3.2.2. 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.

[15] 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.

A rich literature and set of tools exists 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.

The formal construct is a *k*-state homogeneous Markov chain where state 1 refers to the highest rating, 'AAA', and *k* 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).[16] Accepting or relaxing the time homogeneity assumption will dictate the estimation method.

Israel, Rosenthal and Wei (2001) develop conditions under which generator matrices exist for an empirically observed Markov transition matrix. They show that most annual transition matrices would actually not be compatible with a continuous Markov process largely (but not exclusively) as a result of their sparseness; most of the probability mass is on the diagonal, leaving many zero elements on far off-diagonals. They propose a series of simple adjustments and then proceed to compare the adjusted to the unadjusted matrices using the L^1 norm to see which is smaller. However, neither in the checking of conditions for compatibility with a Markov process nor in the matrix comparison do they allow for sampling or estimation noise. No formal hypothesis test is presented making it difficult to decide how small is small (in the matrix comparison) or how close (or far) the estimated matrix was from Markov compatibility.[17] We will return to the sampling and estimation noise issue in Section 4, with particular emphasis on different ways of comparing matrices.

---

[16] 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.

[17] Arvanitis, Gregory and Laurent (1999) propose a metric based on matrix norms to test the 1st order Markov assumption. They provide a cut-off value of 0.08 (for an annual migration matrix) without accounting for estimation noise.

3.2.2.1. Time Homogeneous Case[18]

With the assumption of time homogeneity in place, transition probabilities can be described via a kxk generator or intensity matrix $\mathbf{\Lambda}$. Following Lando and Skodeberg (2002), define $\mathbf{P}$(t) is a kxk matrix of probabilities where the ij^{th} element is the probability of migrating from state i to j in time period t:

$$\mathbf{P}(t) = \exp(\mathbf{\Lambda} t) \quad t \geq 0 \tag{1}$$

where the exponential is a matrix exponential, and the entries of $\mathbf{\Lambda}$ satisfy

$$\begin{aligned} \lambda_{ij} &\geq 0 \ \text{ for } i \neq j \\ \lambda_{ii} &= -\sum_{j \neq i} \lambda_{ij} \end{aligned} \tag{2}$$

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

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

$$\hat{\lambda}_{ij} = \frac{N_{ij}(T)}{\int_0^T Y_i(s)ds} \tag{3}$$

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 $P_{\text{AA}\rightarrow\text{A}}$. 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.

---

[18] This and the next section draws from Lando and Skodeberg (2002); we adopt their notation. See also their excellent examples.

3.2.2.2. Non-homogeneous Case

The duration approach certainly uses the transition information by obligors more efficiently than does the cohort method. We have reason, however, to be skeptical of the time-homogeneity assumption; it may matter where you are in time.

A common assumption for credit modeling (either at the instrument or portfolio level) is for the system to be first-order Markov. This has very convenient implication. For a given period Δt, e.g. one quarter, the k-period transition matrix $\mathbf{P}_{k\Delta t}$ is simply $\mathbf{P}_{\Delta t}^{k}$. For example, the 1-year transition matrix can be computed by taking the quarterly matrix and raising it to the 4th power. However, Carty and Fons (1993), Altman and Kao (1992), Altman (1998), Nickell, Perraudin and Varotto (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.

Again following Lando and Skodeberg (2002), let $\mathbf{P}(s,t)$ be the transition probability matrix from time s to time t. The ij^{th} element of this matrix denotes the probability that the Markov process starting in state i at date s will be in state j at date t. Given a sample of m transitions over the period from s to t, one can consistently estimate $\mathbf{P}(s,t)$ using the nonparametric product-limit, or Aalen-Johansen, estimator (Aalen and Johansen (1978)).

$$\hat{\mathbf{P}}(s,t) = \prod_{i=1}^{m} \left(\mathbf{I} + \Delta \hat{\mathbf{A}}(T_i) \right) \tag{4}$$

T_i is a jump in the time interval from s to t, m is the total number of number of transition days over the relevant horizon (i.e. for an annual matrix, the number of days in the year where at least one rating transition occurs), and

$$
\Delta\hat{\mathbf{A}}(T_i)=\begin{pmatrix}
-\frac{\Delta N_{1\bullet}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{1,2}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{1,3}(T_i)}{Y_1(T_i)} & \cdots & \frac{\Delta N_{1,p}(T_i)}{Y_1(T_i)} \\
\frac{\Delta N_{2,1}(T_i)}{Y_2(T_i)} & -\frac{\Delta N_{2\bullet}(T_i)}{Y_2(T_i)} & \frac{\Delta N_{2,3}(T_i)}{Y_2(T_i)} & \cdots & \frac{\Delta N_{2,p}(T_i)}{Y_2(T_i)} \\
\vdots & \vdots & \ddots & \cdots & \vdots \\
\frac{\Delta N_{p-1,1}(T_i)}{Y_{p-1}(T_i)} & -\frac{\Delta N_{p-1,2}(T_i)}{Y_{p-1}(T_i)} & \cdots & -\frac{\Delta N_{p-1\bullet}(T_i)}{Y_{p-1}(T_i)} & \frac{\Delta N_{p-1,p}(T_i)}{Y_{p-1}(T_i)} \\
0 & 0 & \cdots & \cdots & 0
\end{pmatrix}. \tag{5}
$$

The nonparametric Aalen-Johansen estimator is essentially the frequentist (cohort) method applied to very short intervals. The element $\Delta N_{k,j}(T_i)$ denotes the number of transitions observed from state k to state j at date T_i. The diagonal elements $\Delta N_{k\bullet}(T_i)$ count the total number of transitions *away* from state k at date T_i, and $Y_k(T_i)$ is the number of exposed (or at risk) firms, i.e. the number of firms in state k right before date T_i. The diagonal elements of row k in $\Delta\hat{\mathbf{A}}(T_i)$ count, at any given date T_i, the fraction of exposed firms $Y_k(T_i)$ which leave the state at date T_i. Hence the off-diagonal terms count the specific types of transitions away from the state, normalized by the number of exposed firms. Similar to the homogeneous case, the bottom row of $\Delta\hat{\mathbf{A}}(T_i)$ is zero since default is an absorbing state. Note that the rows of the matrix $\mathbf{I}+\Delta\hat{\mathbf{A}}(T_i)$ automatically sum to one.

Thus the nonparametric Aalen-Johansen estimator imposes the fewest 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). 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 deal with this issue in Sections 5.1 and 5.2.

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).[19] For the sample range we examine, 1981 – 2001, 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 then defaulted to contribute probability mass to PD_{AAA}. This can be seen by comparing the empirical *PD*s in Table 3 which presents *PD* estimates in basis points (bp) using all firms for the entire sample range. For example, the estimated annual probability of default for an 'AA' company, PD_{AA}, is exactly zero for the cohort approach, 0.406bp for the time-homogeneous and 0.105bp for the non-homogeneous duration approach.[20] For an 'A+' rated firm, PD_{A+} is 6.4 bp for the cohort approach but a much smaller 0.5bp and 0.6bp for time-homogeneous and non-homogeneous duration approach respectively, meaning that the less efficient cohort method *over*estimates default risk by more than twelve fold. We would obtain these lower 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: 27.26% for the cohort method but 38.16% and 39.10% for homogeneous and non-homogeneous duration respectively. Thus using the more popular but less efficient cohort method would *under*estimate default risk by over ten percentage 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 (3) or non-homogeneous (5) duration) and hence a higher default probability.

4. Comparing Matrices

We have presented three different methods for estimating the entries in the credit migration matrix; now how would we compare the results? There are several ways of comparing matrices including

---

[19] See also the discussion in Christensen and Lando (2002) on confidence sets for estimated *PD*s.

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

L^1 and L^2 (Euclidean) distance metrics, and eigenvalue and eigenvector analysis such as singular value decomposition. An extensive discussion of the different metrics for comparing credit migration matrices is presented in Jafry and Schuermann (2003). They argue for a metric based on singular values which we will use here.

4.1. 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, we can reasonably posit that the central characteristic of the matrix is the amount of migration (or "mobility") imposed on the state-vector from one epoch to the next. This characteristic can be highlighted by simply subtracting the identity matrix before proceeding with further manipulations. 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. By subtracting the identity matrix from the migration matrix, we are therefore left with only the *dynamic* part of the original matrix. By devising a metric for this dynamic part, we will obtain an intuitively-appealing result which reflects the "magnitude" of the matrix in terms of the implied mobility. We will henceforth refer to the *mobility matrix* (denoted $\tilde{\mathbf{P}}$) defined as the original (or *raw*) matrix minus the identity matrix (of the same dimension), i.e.:

$$\tilde{\mathbf{P}} = \mathbf{P} - \mathbf{I} \tag{6}$$

4.2. A Metric Based on Singular Values

Following Jafry and Schuermann (2003), we utilize the average of the singular values of the mobility matrix $\tilde{\mathbf{P}}$ (i.e. the average of the square-roots of the eigenvalues of $\tilde{\mathbf{P}}'\tilde{\mathbf{P}}$).

$$m_{svd}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b) = \overline{S(\tilde{\mathbf{P}}^a)} - \overline{S(\tilde{\mathbf{P}}^b)} \tag{7}$$

where the overbar denotes "average value" (of the vector of) $S(\tilde{\mathbf{P}})$, which, in turn, denotes the singular values of $\tilde{\mathbf{P}}$, defined as $S(\tilde{\mathbf{P}}) = \sqrt{eig(\tilde{\mathbf{P}}'\tilde{\mathbf{P}})}$. The eigenvalues of $\tilde{\mathbf{P}}$ can be negative making them difficult to interpret as a distance measure, whereas the singular values will always be positive.

This yields a viable metric as it approximates (and in certain cases is identical to) the average probability of migration across all states.[21]

4.3. Bootstrapping

The estimates of the transition matrices are just that: estimates with error (or noise). Consequently the distance metrics such as SVD-based $m_{svd}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)$ is also a noisy estimate. In order to help us answer the question "how large is large" for distance metrics such as $m_{svd}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)$, we need the distributional properties of $\hat{m}_{svd}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)$. In the absence of any theory on the asymptotic properties of estimates of $m_{svd}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)$, a straightforward and efficient way is through the resampling technique of bootstrapping.

Consider, for example, $\tilde{\mathbf{P}}_t^c$ and $\tilde{\mathbf{P}}_t^h$ as the cohort and homogeneous duration estimates at time t respectively, obtained using N_t observations.[22] Suppose we create k bootstrap samples[23] of size N_t each so that we can compute a set of k differences based on singular values, $\left\{m_{svd}^{(j)}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)\right\}_{j=1}^{k}$ where $j = 1,\ldots,K$ 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 α = 5%), we see if 0 falls within the 1-α range of $\left\{m_{svd}^{(j)}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)\right\}_{j=1}^{k}$.

---

[21] See Jafry and Schuermann (2003) for a detailed discussion of the motivation behind the singular value formulation.

[22] 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.

Suppose, in other words, that we run the bootstrap for our sample of N_{1999} firms to obtain $\left\{m_{svd,1999}^{(j)}(\tilde{\mathbf{P}}^a, \tilde{\mathbf{P}}^b)\right\}_{j=1}^{k}$ for some relatively large k (≈ 1000).[24] Suppose further that the 1-α = 95% (say) range turns out to be (-0.055, -0.005). Then we would be able to reject the null hypothesis and state, with 95% confidence, that the two matrices are different with respect to this singular value based metric. If, on the other hand, the range turns out to be something like (-0.055, 0.005), then we would *not* be able to reject the null, and we would have to accept, with 95% confidence, that the two matrices are the same.

Ideal conditions for the bootstrap require that the underlying data is a random sample from a given population. Specifically the data should be *iid*. It is difficult to impose temporal *independence* across multiple years, but easier at shorter horizons such as one year. Moreover, since we restrict our analysis to U.S. firms (i.e. no government entities (municipal, state or sovereign), no non-U.S. entities), we can be comfortable with the assumption of *identically distributed* (no mixing).

5. Empirical Results

5.1. Statistical Differentiation

In this section we compare the three methods (cohort, time homogeneous duration and non-homogeneous duration) 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 efficient duration method, even with the (possibly false)

---

[23] 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).

[24] Efron and Tibshirani (1993) suggest that for obtaining standard errors for bootstrapped statistics, bootstrap replications of 200 are sufficient. For confidence intervals, they suggest bootstrap replications of 1000 which we employ.

assumption of time-homogeneity over the cohort method has a far greater impact than relaxing the time-homogeneity assumption.

By way of illustration using ratings histories of U.S. obligors from 1981-2001, in Figure 4 we compare m_{SVD} of cohort minus homogeneous duration (solid line with square) against non-homogeneous minus homogeneous duration (solid line) and cohort minus non-homogeneous duration (dashed line). Comparing cohort to either of the duration approaches yields very similar SVD distance measures, while the distance between duration approaches (solid line) is very small indeed.[25] Minima and maxima using the SVD metric are summarized in Table 4.

The largest difference between cohort and homogeneous duration methods occurs in 1999 while the smallest difference is found in 1984. In absolute value, its minimum occurs in 1984 (0.0007) and its maximum in 1999 (-0.0525). Is either different from zero, meaning do the two methods generate statistically indistinguishable transition matrices? Table 5 (left column) provides some summary statistics of the bootstrap, including several quantiles, and Figure 5 displays the densities of the bootstraps. Indeed we are unable to reject that the 1984 matrices are different (0 is near the median) but can do so for the 1999 matrices: the 98% confidence interval from the 1st to the 99th percentile is (-0.0904, -0.01280). The density for 1999 is much wider and flatter (and almost entirely to the left of zero) than the density for 1984 which is highly concentrated around zero.

Moving on to the comparison of the two duration methods, we see that the largest difference, in absolute value, occurs in 1982 (0.01004), with 1999 running a close second (0.0079), and the smallest in 1990 (0.00008). 1984, a year where the difference between cohort and homogeneous duration was the smallest, is also rather small when comparing the duration methods: 0.0010. Table 5 (right column) shows the bootstrap results for 1982 (max difference) and 1990 (min difference). Even for the year of maximum difference between duration methods, namely 1982, we are unable to reject the hypothesis that

[25] In Jafry and Schuermann (2003) we find that the cohort matrices are generally "smaller" than the homogeneous matrices, i.e. with generally less mobility. That's because the cohort method ignores within-period movements.

the difference is zero. The 98% confidence interval from the 1st to the 99th percentile is (-0.0014, 0.0515). For 1990 the zero is contained already in the 90% confidence interval (-0.0032, 0.0033). Figure 6 shows density plots of the bootstrap runs for these two years. For 1982, the density is more skewed than it is the case for the SVD differences between cohort and homogeneous duration methods; moreover, less than 1% of the density is to the left of zero (meaning zero is inside the 99% range). The density for 1990 is centered very near zero and is more symmetric than 1982.

Comparing cohort to non-homogeneous duration with the SVD metric, the results are very similar to the previous comparison (cohort to homogeneous duration). The maximum (in absolute value) occurs in the same year, 1999 (-0.06044)) and the minimum in 1982 (0.00035).

What becomes clear is that the differences (in SVD terms) between the duration methods are much smaller than between cohort and duration methods,[26] 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. In Table 6 we make a formal comparison of means between the different methods. The mean difference of the cohort and the homogeneous duration, $m_{svd}(\tilde{\mathbf{P}}^{Cohort}, \tilde{\mathbf{P}}^{Homog.})$, is –0.012, the difference between cohort and non-homogeneous duration, $m_{svd}(\tilde{\mathbf{P}}^{Cohort}, \tilde{\mathbf{P}}^{Non-homog.})$, is –0.014, while the mean difference of $m_{svd}(\tilde{\mathbf{P}}^{Non-Homog.}, \tilde{\mathbf{P}}^{Homog.})$ is a much smaller 0.002. Indeed we cannot reject that the average difference between the cohort and either duration method is different (from zero) with a p-value of 0.456, 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 6 (the other test yields the same result), where the p-value is 0.00001, allowing us to strongly reject that the two average 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.

5.2. Economic Relevance

While we may be able to distinguish statistically between two empirically estimated transition matrices, this may not translate to economic significance. We can think of many ways to measure economic relevance; two come to mind immediately. The first is to 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. The second is the pricing of credit derivatives, a class of new financial instruments that can help banks, financial companies and investors in managing credit risk in their portfolios.

5.2.1. Credit Portfolio Models

The purpose of capital is to provide a cushion against losses for a financial institution. 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,

---

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

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

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 392 exposures with a current value of $427.6 MM. We did so by taking a random sample of rated U.S. corporates that mimics the ratings distribution of the S&P U.S. universe as of January 2002 in such a way that we have at least one obligor for each major industry group. Maturity ranges from one to 29 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 October 16, 2002. 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 some of our findings in Table 7 and Table 8. Four sets of numbers are displayed for each experiment: mean and 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. It is also the confidence level stipulated by the New Basel Accord (BIS (2001)). Table 7 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] Since business cycles are dated at higher frequencies than yearly, we take quarterly estimates using the homogeneous duration method and raise that matrix to the fourth power. While this presumes that the migration process is first-order Markov, this assumption becomes more stringent for longer horizons (multi-year) than shorter ones (quarter to annual). Moreover, this way we are able to maximally concentrate recession and expansion time periods. Moving to an annual frequency would mix in, for recession (expansion) years, expansion (recession) quarters.

---

[28] Parameters such as those governing the recovery process. We generated 5000 trials for each run with importance sampling.

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

Here we confirm results in Bangia et al. (2002) where capital held during a recession should be about 40% higher than during an expansion (39.7% at the 99% level, 28.8% at the 99.9% level).[30] Moreover, the portfolio volatility is over 40% higher during a recession than an expansion.

Table 8 presents results for the average across the entire sample range as well as for five individual years: 1982 (max SVD for non-homogeneous minus homogeneous duration), 1984 (min SVD for cohort minus homogeneous duration), 1990 (min SVD for non-homogeneous minus homogeneous duration), 1999 (max SVD for cohort minus homogeneous duration) and 2001 (most recent). We focus on differences in VaR capital for each experiment. Without exception, the differences between the cohort and more efficient duration methods are larger than between the different duration methods, with differences of 30% to nearly 50% for the former, and never more then 2% for the latter. Viewed through the lens of credit risk capital, 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 (8% to 11%) than between the two duration methods (< 2%). 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 (7% to 9%) than between duration methods (< 1%). The year 1990 is no exception.

The difference is startling when we look at the 1999, the year where we experienced the largest divergence between cohort and homogeneous duration methods. VaR differences are on the same order as when comparing recession to expansion: 36% to nearly 50%.

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 1999 the difference in

30 Bangia et al. (2002) report capital differences of 25-30%. We have the benefit of three additional years of data, one of which (2001) was mostly a recession year.

expected value of the portfolio between the cohort and homogeneous duration methods is essentially nil, but the difference in risk is substantial.

5.2.2. Credit Derivatives

Credit derivatives have been developed in the past few years as new financial instruments that can help banks, financial companies and investors to manage credit risk in their portfolios. For banks in particular these instruments have emerged as a powerful tool for managing the credit exposure of lending portfolios without affecting the balance sheet. The basic idea is to make credit risk transferable while maintaining the ownership of the credit risky asset. This is typically accomplished by periodic or up-front payments of the protection buyer to the protection seller. In case pre-specified credit events take place (e.g. missed interest payments), the protection buyer receives predefined cash flows from the protection seller. A variety of products emerged, including credit default swaps, total return swaps, etc. (see, for instance, Acharya, Das and Sundaram (2002) and references therein). Among the more complicated instruments are yield spread options, as their pay-off not only depends on possible default events but also on credit rating changes.

A yield spread option enables the buyer and seller to speculate on the evolution of the yield spread. The yield spread is defined as the difference between the continuously compounded yield of a risky and a risk-less zero-coupon bond with the same maturity. In the case of a call option, the buyer expects a decreasing, in the case of a put option an increasing credit spread as bond yields typically have an inverse relationship with bond prices. Yield spread options are priced using Markov chain models. An early example is Jarrow and Turnbull (1995) and its extension in Jarrow, Lando and Turnbull (1997). We will consider a more recent refinement as proposed by Kijima and Komoribayashi (1998).

To price a yield spread option we need a variety of inputs. Below we list those inputs along with their value for our example (where relevant):

- Yield curve for default free zero coupon bonds at date T (April 1, 2002)
- Term structure of credit spreads at time T (April 1, 2002)

- Option maturity: 2 years
- Yield spread maturity: 4 years
- Strike price (i.e. a yield spread): 9% (with a current forward yield spread of 7.69%, meaning that we are betting on a widening of credit spreads)
- Current rating of obligor or instrument: CCC
- Recovery rate (% of outstanding recovered in case of default): 50%
- Migration matrix (will vary)

For easy comparison to the previous example of credit portfolio models, we use S&P ratings histories to compare U.S. recession and expansion matrices (derived either by the cohort or homogeneous duration), 1999 U.S. matrix cohort vs. homogeneous and non-homogeneous duration, where 1999 was shown to have the largest deviation in our SVD-based metric, and 1984 U.S. matrices, where 1984 had the smallest deviation. In addition we compare 1982 where the difference in the two duration methods was the largest. For 1984, all three matrices yielded the same result: no options premium. Figure 7 compares the option premiums for 1982 and 1999 (three methods) as well as the premiums implied by the recession/expansion matrices (cohort only). The difference implied by the duration methods, holding year fixed at 1999, is on the same order as the difference between two economic regimes, expansion and recession, holding method fixed. These results are broadly consistent with those obtained from the credit portfolio example. Using the "wrong" matrix can result in substantial mispricing. However, a year where overall differences as measured by SVD are modest, namely 1982, also yields modest differences in the options premium. To be sure, consistent with the SVD metric, differences in pricing between duration methods for this year (recall that 1982 was the year with maximum difference for these two methods) is smaller than between cohort and, say, homogeneous.

6. Concluding Remarks

In this paper we presented 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 best answer these questions, we focus on the mobility matrix, defined as the migration matrix **P** less the identity matrix **I** of same size. The identity matrix corresponds to a *static* migration matrix, i.e. the state vector is unchanged by the action of the matrix from one period to the next. In subtracting the identity matrix, we are left with only the *dynamic* part of the original matrix.

The first question is addressed using a metric based on singular value decomposition, SVD, developed in Jafry and Schuermann (2003). We argue that a metric based on singular values best captures the dynamic properties in the migration matrix. The second question is addressed using resampling methods, and the third using two approaches: the first is to estimate by simulation the credit risk capital levels implied by the credit portfolio model in CreditMetrics™ which is used to generate value distributions of a portfolio of credit assets such as loans or bonds. The second is the pricing of a credit derivative called a credit yield spread using the pricing model of Kijima and Komoribayashi (1998).

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.[31] For years where the 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, but only between the homogeneous duration and cohort methods. Relaxing the time homogeneity assumption

[31] To be sure, we would expect this difference to decrease for shorter horizon matrices, e.g. quarterly, but increase for longer, multi-year horizons.

yields little difference; 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 30% to nearly 50% for the former, and never more then 2% for the latter. Thus ignoring the efficiency gain inherent in the duration methods is more damaging that making a (possibly false) assumption of time homogeneity. These differences in risk capital are on the same order as are implied by business cycle effects which, in turn, are about 40%. These results from the credit derivative pricing exercise are broadly consistent with those obtained from the credit portfolio example: using the "wrong" matrix can result in substantial mispricing, often by more than 50%.

References

Aalen, O.O. and S. Johansen, 1978, "An Empirical Transition Matrix for Nonhomogeneous Markov Chains Based on Censored Observations", *Scandinavian Journal of Statistics* 5, 141-150.

Acharya, V.V., S.R. Das and R. K. Sundaram, 2002, "Pricing Credit Derivatives with Rating Transitions," *Financial Analysts Journal*, May/June, 28-44.

Altman, E.I. and D.L. Kao, 1992, "Rating Drift of High Yield Bonds", *Journal of Fixed Income*, March, 15-20.

Altman, E.I. and V.M Kishore, 1996, "Almost Everything You Wanted to Know about Recoveries on Defaulted Bonds", *Financial Analysts Journal*, November/December, 57-64.

Altman, E.I., 1998, "The Importance and Subtlety of Credit Rating Migration", *Journal of Banking & Finance*, 22, 1231-1247.

Arvanitis, A., J. Gregory and J-P. Laurent, 1999, "Building Models for Credit Spreads", *Journal of Derivatives*, Spring, 27-43.

Bahar, R. Nagpal, K., 2000, "Rating Transitions: Modelling the Dynamics", *Credit*, March, 57-63.

Bangia, A., F.X. Diebold, A. Kronimus and C. Schagen and T. Schuermann, 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*, http://www.bis.org/publ/bcbsca.htm.

Brand, L. and R. Bahar, 2001, "Corporate Defaults: Will Things Get Worse Before They Get Better?", *S&P Special Report*, January, 5-40.

Cantor, R. and F. Packer, 1995, "The Credit Rating Industry", *Journal of Fixed Income*, December, 10-34.

Carty, L.V., 1997, "Moody's Rating Migration and Credit Quality Correlation, 1920 – 1996", *Special Comment*, (Moody's Investor Service, New York).

Carty, L.V. and J.S. Fons, 1993, "Measuring Changes in Corporate Credit Quality", *Moody's Special Report*, (Moody's Investors Service, Inc., New York).

Christensen, J. and D. Lando, 2002, "Confidence Sets for Continuous-Time Rating Transition Probabilities", mimeo, University of Copenhagen; available at http://www.math.ku.dk/~dlando/jcdl1015.pdf.

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

Finger, C., 1999, "Conditional Approaches to for CreditMetrics Portfolio Distributions", *CreditMetrics Monitor*, April.

Geweke, J., R.C. Marshall and G.A. Zarkin, 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., C.C. Finger and M. Bhatia, 1997, *CreditMetrics™ – Technical Document,* this version: April 2, 1997 (J.P. Morgan, New York, NY).

Israel, Robert B., Jeffrey S. Rosenthal and Jason Z. Wei, 2001, "Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings", *Mathematical Finance* 11 (2), 245-265.

Jafry, Y. and T. Schuermann, 2003, "Metrics for Comparing Credit Migration Matrices", Wharton Financial Institutions Working Paper #03-08.

Jarrow, R. A. and S.M. Turnbull, 1995, "Pricing Derivatives on Financial Securities Subject to Credit Risk", *Journal of Finance* 50, 53-86.

Jarrow, R. A., D. Lando, and S.M. Turnbull, 1997, "A Markov Model for the Term Structure of Credit Risk Spreads", *The Review of Financial Studies* 10 (2), 481-523.

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

Kavvathas, D., 2001, "Estimating Credit Transition Probabilities for Corporate Bonds", University of Chicago Working Paper.

Kijima, M., and K. Komoribayashi, 1998, "A Markov Chain Model for Valuing Credit Risk Derivatives", *The Journal of Derivatives*, Fall, 97-108.

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

Kliger, D. and O. Sarig, 2000, "The Information Value of Bond Ratings", *Journal of Finance*, 55 (6), 2879-2902.

Koyluoglu, H.U. and A. Hickman, 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. and T. Skodeberg, 2002, "Analyzing Ratings Transitions and Rating Drift with Continuous Observations", *Journal of Banking & Finance*, 26 (2/3), 423-444.

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

Morgan, D.P., 2002, "Rating Banks: Risk and Uncertainty in an Opaque Industry", *American Economic Review* 92 (4), 874-888.

Nickell, P, W. Perraudin and S. Varotto, 2000, "Stability of Rating Transitions", *Journal of Banking & Finance*, 24, 203-227.

Partnoy, F., 2002, "The Paradox of Credit Ratings", University of San Diego Law and Economics Research Working Paper #20.

Saunders, A. and L. Allen, 2002, *Credit Risk Measurement*, 2nd Edition. John Wiley & Sons, New York, NY.

Shorrocks, A.F., 1978, "The Measurement of Mobility", *Econometrica* 46, 1013-1024.

Standard & Poors, 2001, *CreditPro 3.0 Manual.*

Treacy, W.F. and M. Carey, 2000, "Credit Risk Rating Systems at Large U.S. Banks", *Journal of Banking & Finance* 24, 167-201.

White, L., 2001, "The Credit Rating Industry: An Industrial Organization Analysis," in R.M. Levich, C. Reinhart, and G. Majnoni, eds. *Ratings, Rating Agencies and the Global Financial System*, Kluwer (forthcoming).

Figures

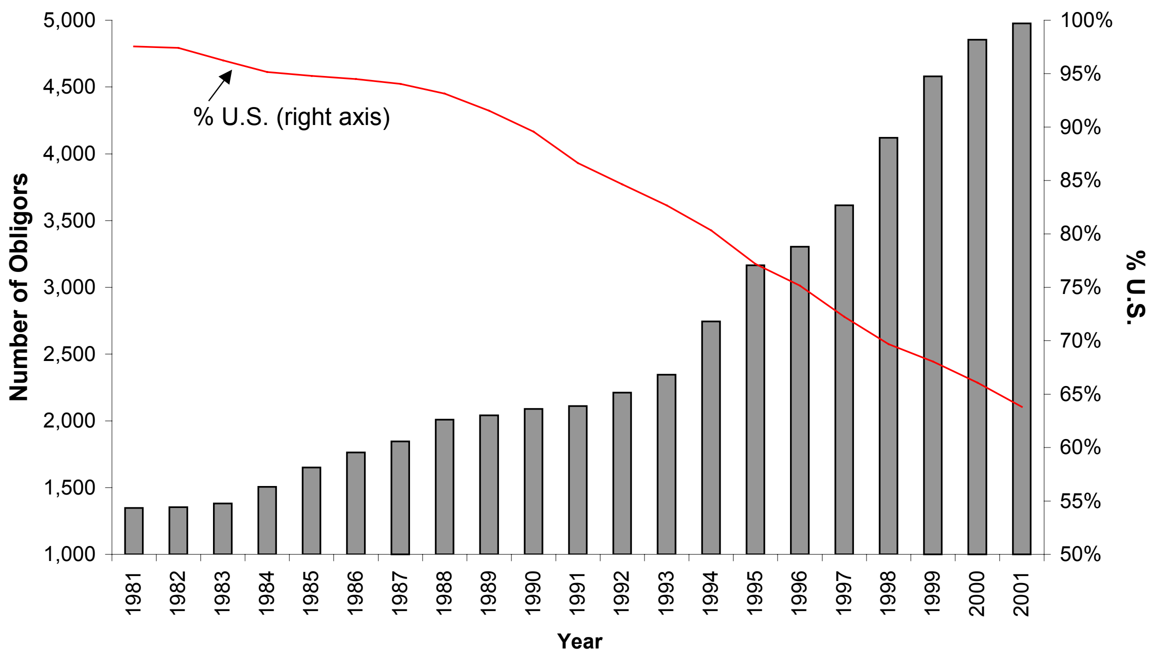

Figure 1: Number of Obligors (Excluding 'NR') at Start of Each Year, S&P, 1981-2001 (left axis); % U.S. domiciled (right axis).

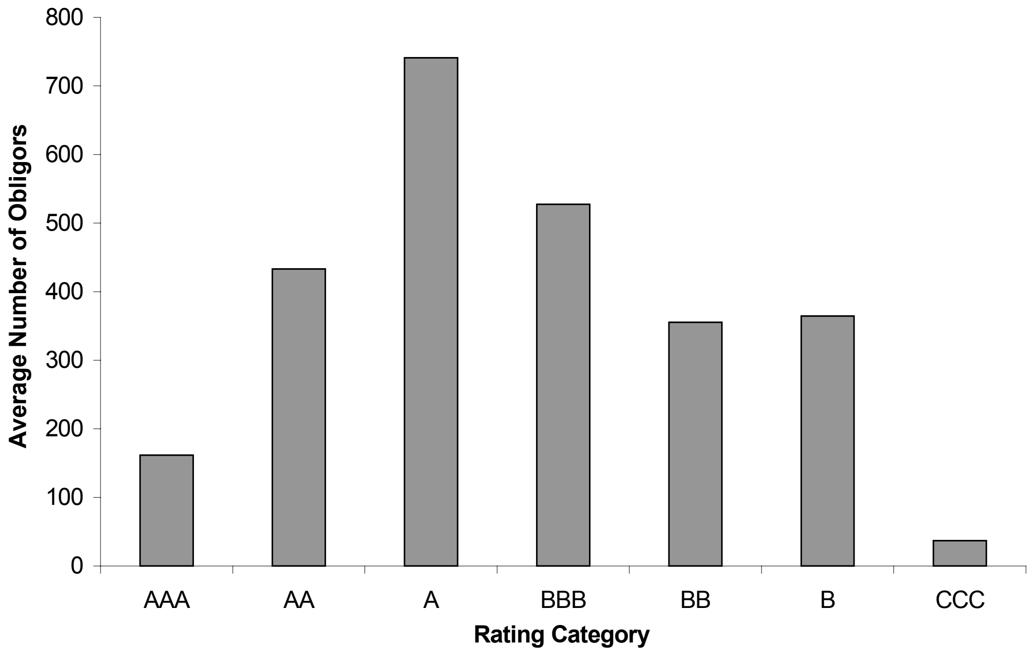

Figure 2: Overall Ratings Distribution, 1981 – 2001, S&P, Global

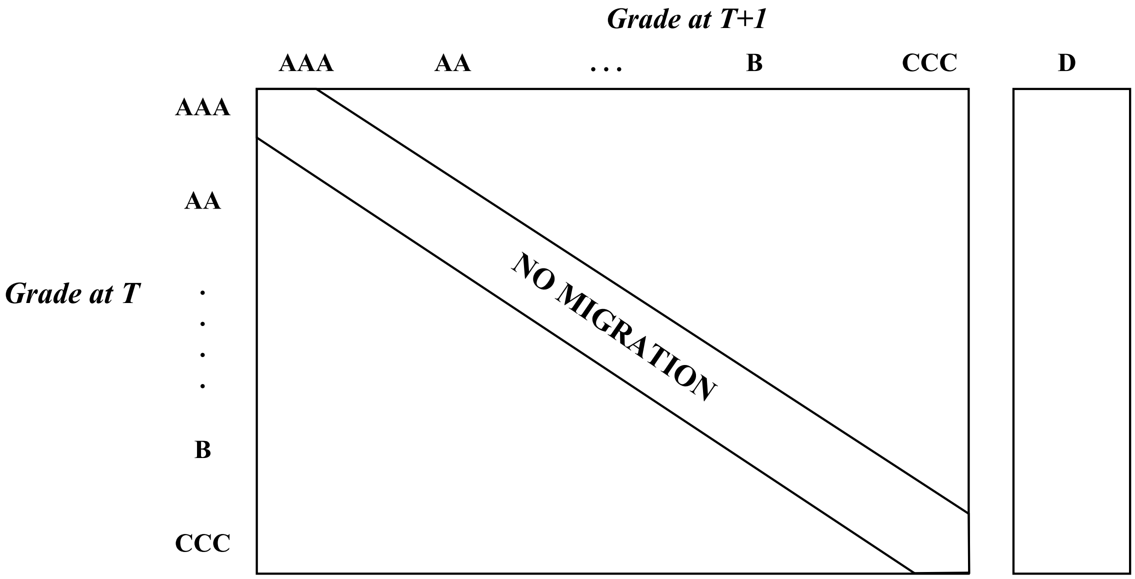

Figure 3: Stylized Migration Matrix

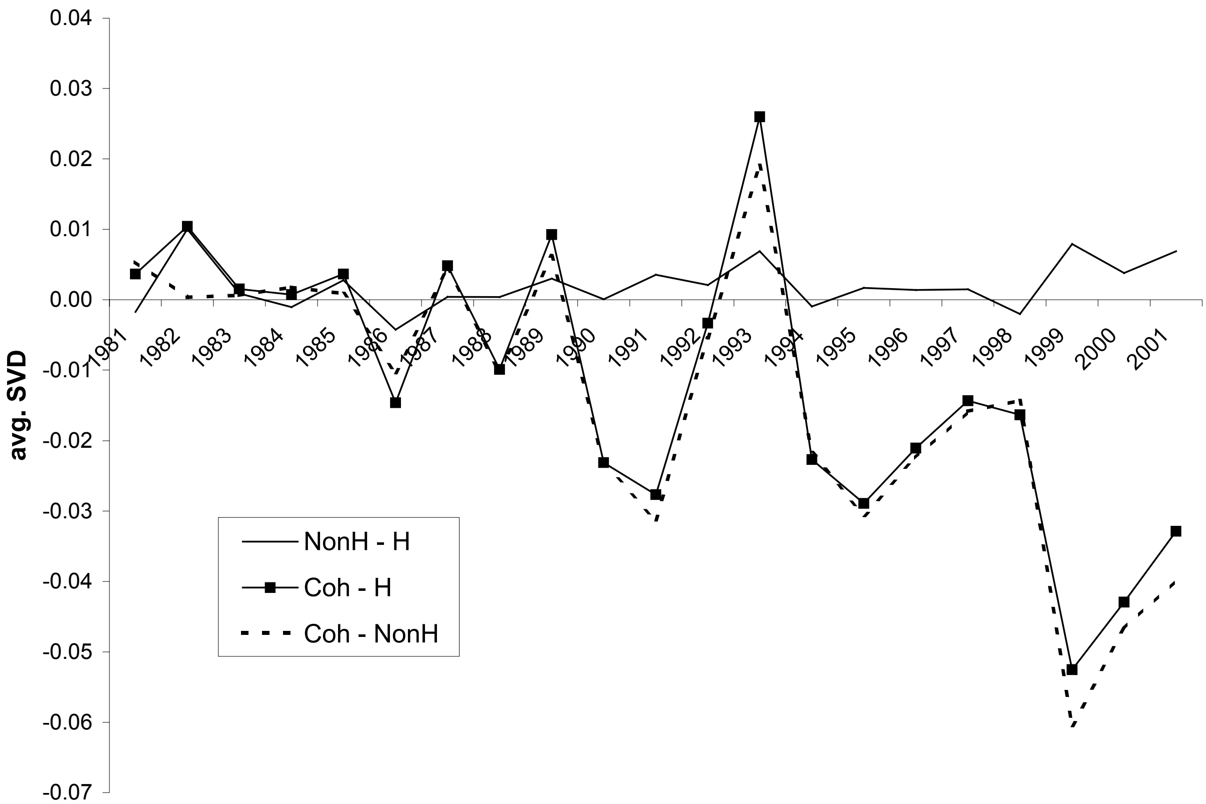

Figure 4: Avg. SVD Distance Metric: Annual Mobility Matrices, Cohort, Homogeneous and Non-Homogeneous Duration. Matrices estimated using S&P rated U.S. obligors, 1981-2001.

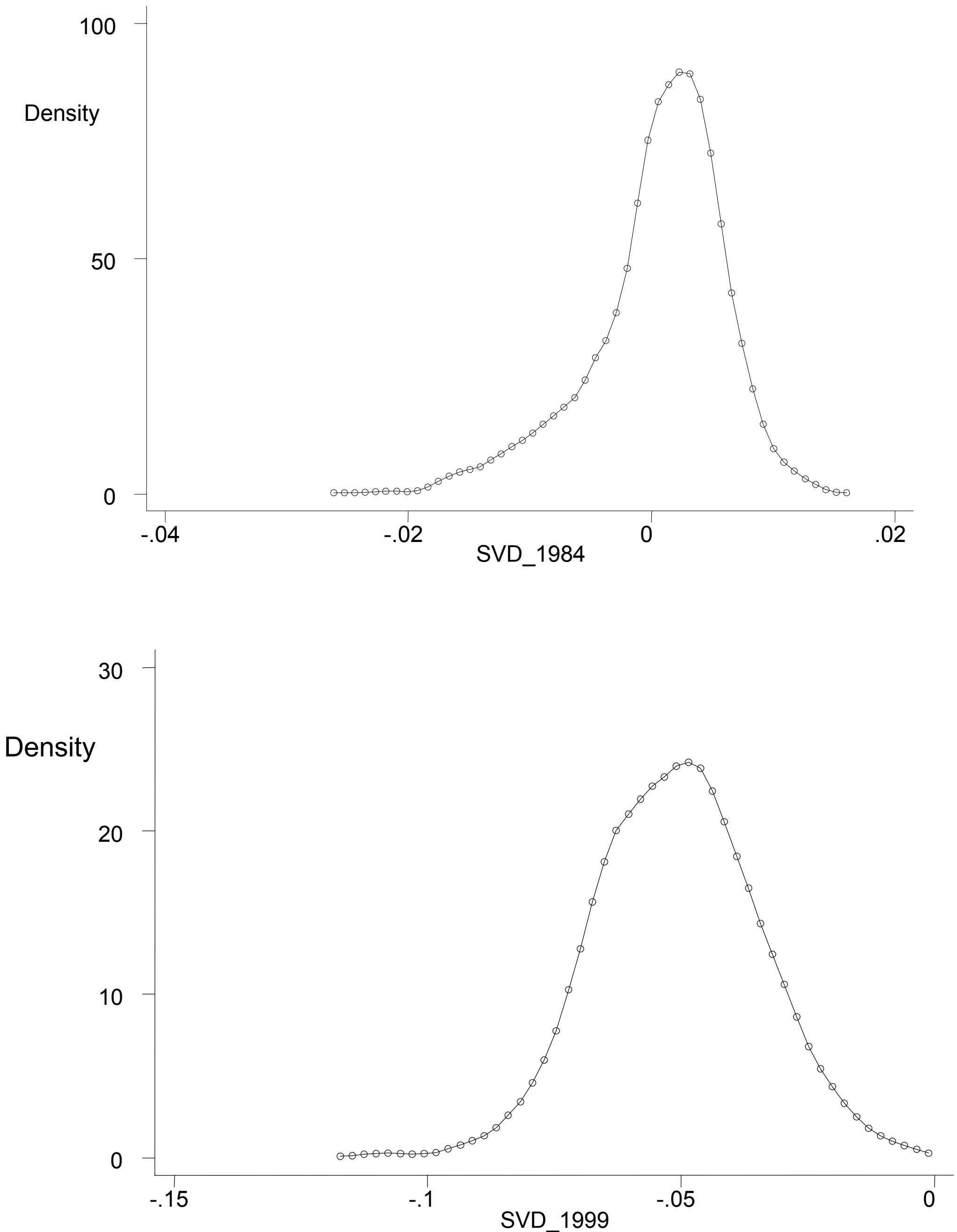

Figure 5: Bootstrapped SVD Distributions, Cohort minus Homogeneous Duration: 1984 and 1999. Matrices estimated using S&P rated U.S. obligors.

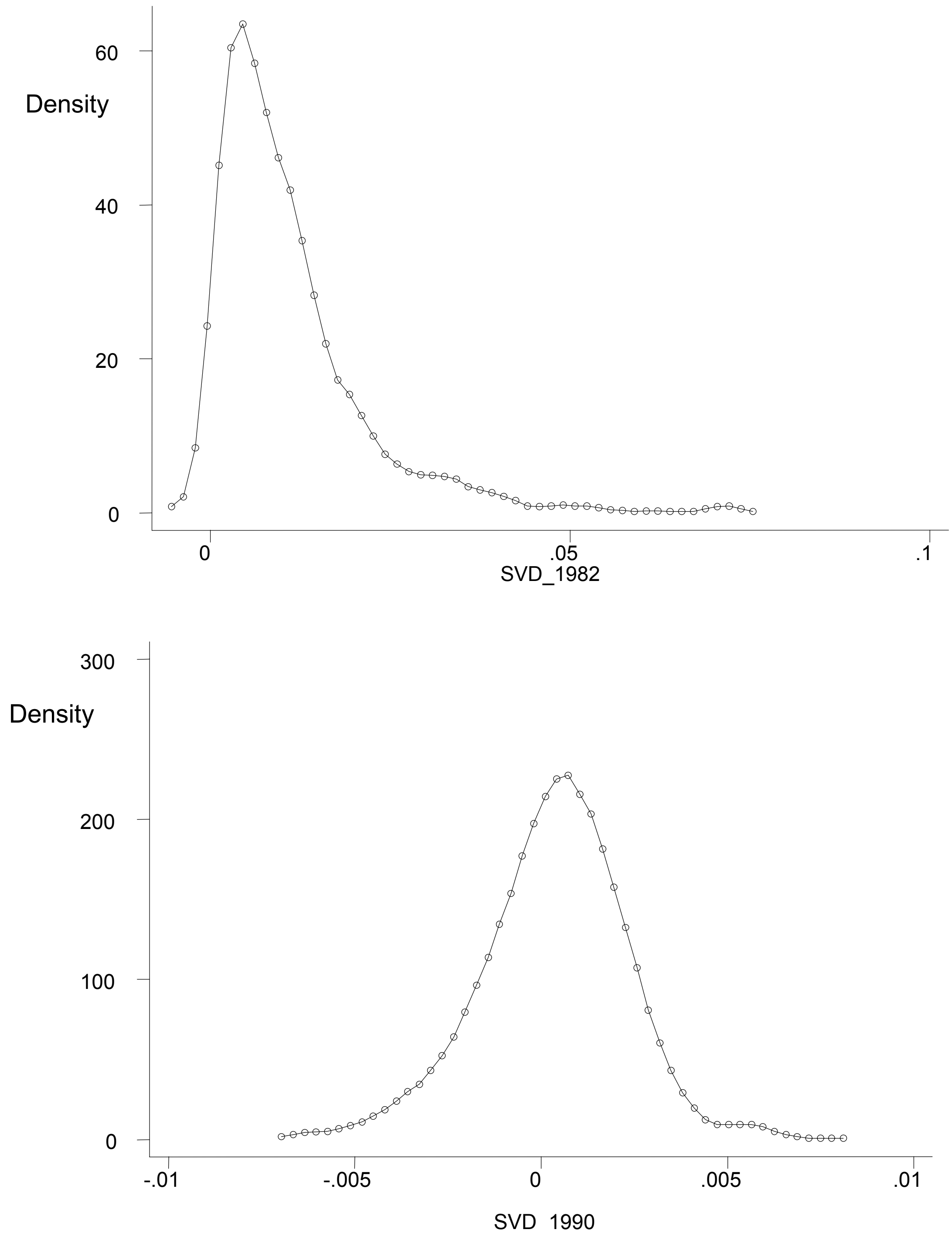

Figure 6: Bootstrapped SVD Distance Metric Distributions, Non-homog minus Homog. Duration: 1982 and 1990. Matrices estimated using S&P rated U.S. obligors.

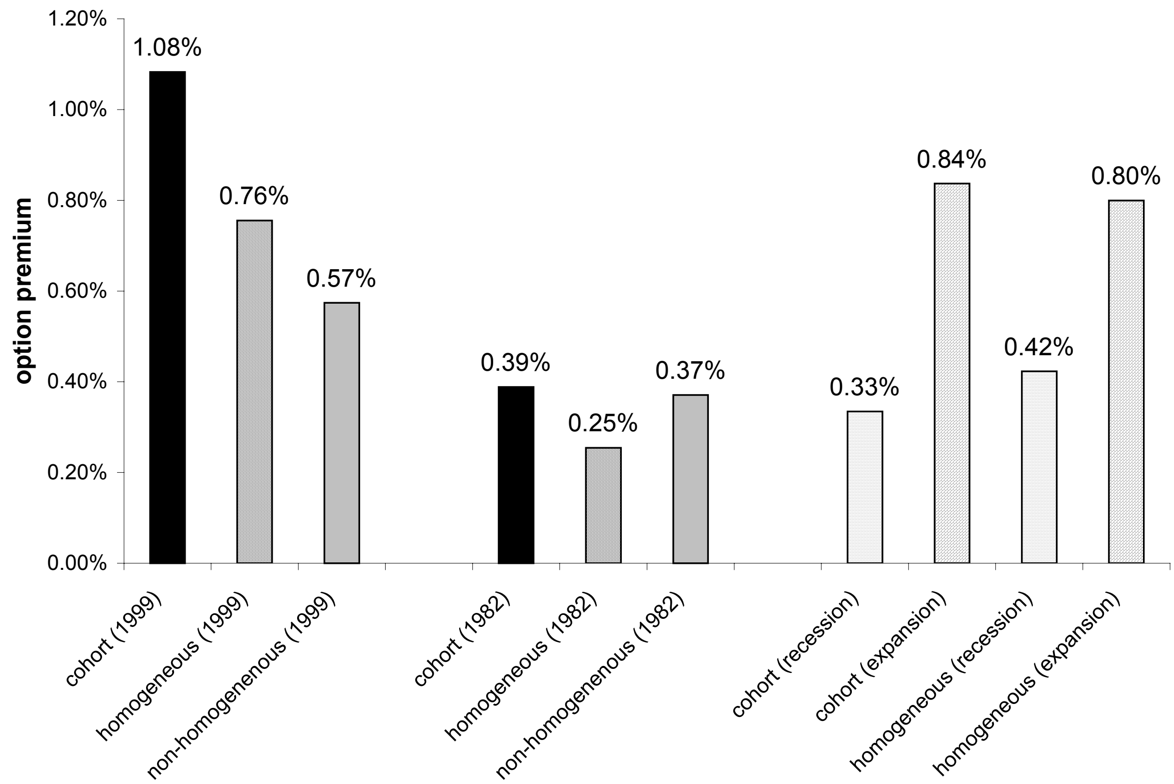

Figure 7: Yield Spread Option Premiums Compared. Yield spread option premium using the risk-free yield curve and term structure of credit spreads as of April 1, 2002, an option maturity of 2 years, a yield spread maturity of 4 years, a strike price of 9% (with a current forward yield spread of 7.69%, meaning that we are betting on a widening of credit spreads), a current obligor rating of CCC, a recovery rate assumption of 50%, and different migration matrices.

Tables

Credit Grade at year-end	Marrison (2002), ch. 18	Saunders and Allen (2002), ch. 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%

Table 1: Value Change of BBB Bond 1 Year Hence. Comparison is relative to no rating change.

Total number of obligors[32]	9,178 (9,769)
Total number of obligor years	55,010
Total number of defaults (annual default rate)	840 (1.53%)
% U.S. in Jan. 1981 (Jun. 2001)[33]	98% (60%)
% investment grade in Jan. 1981 (Dec. 2001)	77% (71%)
Total number of NR / WR (through Dec. 2001)	3,605

Table 2: Descriptive Statistics of S&P CreditPro™ Data, 1981-2001

[32] Through 2001Q4, including NR/WR ratings. The larger number in parentheses includes firms which have re-emerged from either D or NR/WR states.

[33] Includes firms that have re-emerged from either D or NR states.

Rating Categories	Frequentist / Cohort	Duration: time homogeneous	Duration: time non-homogeneous
AAA	0.000	0.008	0.002
AA+	0.000	0.033	0.038
AA	0.000	0.406	0.105
AA-	2.760	0.331	0.458
A+	6.402	0.495	0.593
A	4.449	1.247	1.012
A-	4.838	1.041	0.733
BBB+	18.878	6.344	6.611
BBB	28.309	14.014	13.996
BBB-	35.155	21.323	25.706
BB+	47.071	29.611	28.912
BB	104.910	54.148	47.221
BB-	192.794	46.188	92.845
B+	351.824	454.378	199.043
B	1003.344	788.898	778.295
B-	1313.240	1337.681	1309.674
CCC[34]	2726.098	3816.125	3910.413

Table 3: S&P Annual Estimated *PD*s in Basis Points (1981 – 2001), across methods. All obligors (global).

[34] Includes all grades below CCC.

	Min (year)	Max (year)
Cohort minus Homog. Duration	0.000731 (1984)	-0.05254 (1999)
Cohort minus Non-homog. Duration	0.000354 (1982)	-0.06044 (1999)
Non-homog. minus Homog. Duration	0.000076 (1990)	0.01004 (1982)

Table 4: Min and Max m_{SVD}. Migration matrix distances using SVD metric. Matrices estimated using S&P rated U.S. obligors, 1981-2001.

Stats	Cohort minus Homogeneous Duration 1984 (min)	Cohort minus Homogeneous Duration 1999 (max)	Non-Homogeneous minus Homogeneous Duration 1990 (min)	Non-Homogeneous minus Homogeneous Duration 1982 (max)
$\hat{m}_{SVD}$	0.000731	-0.05254	0.00008	0.01004
mean	0.0005	-0.0508	0.000334	0.0109
st.dev.	0.0056	0.0161	0.001969	0.0106
Q1	-0.0162	-0.0904	-0.00498	-0.0014
Q5	-0.0107	-0.0764	-0.00317	0.0008
Q50	0.0014	-0.0508	0.00043	0.0081
Q95	0.0078	-0.0240	0.00325	0.0325
Q99	0.0112	-0.0128	0.00542	0.0515

Table 5: Bootstrapped SVDs: $m_{svd}(\tilde{\mathbf{P}}^c, \tilde{\mathbf{P}}^h)$, $m_{svd}(\tilde{\mathbf{P}}^{nh}, \tilde{\mathbf{P}}^h)$**.** Quantiles from bootstrapping (k=1000) the SVD-based distance metric using S&P rated U.S. obligors, 1981-2001.

	Cohort – Homog. Duration	*Cohort – Non-homog. Duration*	*Cohort – Homog. Duration*	*Non-homog. – Homog. Duration*
Mean	-0.012	-0.014	-0.012	0.002
Std. Dev.	0.019	0.020	0.019	0.004
F (Δ mean = 0)	0.952		29.717	
Pr (F ≤ f)	0.456		0.00001	

Table 6: Difference in Means of m_{SVD} between Methods. F-test assumes different variances. m_{SVD} estimated for each method for S&P rated U.S. obligors, 1981-2001.

Homogeneous Duration: Recession vs. Expansion ($m_{svd}(\tilde{\mathbf{P}}_R^h, \tilde{\mathbf{P}}_E^h) = 0.04908$)			
	Recession	**Expansion**	**% $\frac{\textbf{Recession}}{\textbf{Expansion}}$**
Mean horizon (1-yr) value	\$406,527,149	\$410,329,633	99.1%
Std. dev. of horizon value	\$6,415,535	\$4,486,968	143.0%
VaR (99.0%)[35]	\$22,917,169	\$16,399,228	139.7%
VaR (99.9%)	\$38,471,154	\$29,871,826	128.8%

Table 7: Credit Risk Capital: Recession vs. Expansion. Credit risk capital as computed by CreditMetrics using a 1-year horizon, 5000 replications (using their importance sampling option). The sample portfolio is as described in Section 5.2.1. Recession and expansion migration matrices (quarterly) were estimated using NBER dating. The quarterly matrices were raised to the 4th power to obtain the annual matrix.

[35] All values ±1%. For details on the derivation of sampling noise for quantiles, see JP Morgan (1997), Appendix B.

Average, 1981 – 2001					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\text{Cohort}}{\text{Homog.}}$	$\% \frac{\text{Non-homog.}}{\text{Homog.}}$
Mean horizon (1-yr) value	\$409,464,723	\$409,643,552	\$409,623,251	99.96%	100.00%
Std. dev. of horizon value	\$5,708,601	\$4,942,358	\$4,960,373	115.50%	100.36%
VaR (99%)	\$21,067,231	\$18,024,322	\$18,040,026	116.88%	100.09%
VaR (99.9%)	\$38,788,885	\$32,470,519	\$32,432,528	119.46%	99.88%
1982 (max m_{SVD} for Non-homog. minus Homog. Duration)					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\text{Cohort}}{\text{Homog.}}$	$\% \frac{\text{Non-homog.}}{\text{Homog.}}$
Mean horizon (1-yr) value	\$408,742,427	\$408,927,127	\$408,778,510	99.95%	99.96%
Std. dev. of horizon value	\$6,472,555	\$5,995,187	\$5,976,406	107.96%	99.69%
VaR (99%)	\$24,424,298	\$22,667,586	\$22,321,742	107.75%	98.47%
VaR (99.9%)	\$44,570,378	\$40,011,622	\$39,603,882	111.39%	98.98%
1984 (min m_{SVD} for Cohort minus Homog. Duration)					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\text{Cohort}}{\text{Homog.}}$	$\% \frac{\text{Non-homog.}}{\text{Homog.}}$
Mean horizon (1-yr) value	\$410,799,680	\$410,963,324	\$410,962,755	99.96%	100.00%
Std. dev. of horizon value	\$5,064,971	\$4,691,610	\$4,718,069	107.96%	100.56%
VaR (99%)	\$19,203,545	\$17,666,536	\$17,715,169	108.70%	100.28%
VaR (99.9%)	\$34,405,042	\$32,066,682	\$32,282,974	107.29%	100.67%

1990 (min m_{SVD} for Non-homog. Minus Homog. Duration)					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\textbf{Cohort}}{\textbf{Homog.}}$	$\% \frac{\textbf{Non-homog.}}{\textbf{Homog.}}$
Mean horizon (1-yr) value	\$406,910,320	\$407,337,486	\$407,304,866	99.90%	99.99%
Std. dev. of horizon value	\$7,582,866	\$6,404,211	\$6,447,963	118.40%	100.68%
VaR (99%)	\$26,954,866	\$22,200,095	\$22,267,688	121.42%	100.30%
VaR (99.9%)	\$43,873,036	\$36,105,445	\$36,241,071	121.51%	100.38%
1999 (max m_{SVD} for Cohort minus Homog. Duration)					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\textbf{Cohort}}{\textbf{Homog.}}$	$\% \frac{\textbf{Non-homog.}}{\textbf{Homog.}}$
Mean horizon (1-yr) value	\$408,959,782	\$408,903,194	\$408,727,098	100.01%	99.96%
Std. dev. of horizon value	\$5,538,481	\$4,268,472	\$4,288,106	129.75%	100.46%
VaR (99%)	\$20,807,339	\$15,302,729	\$15,371,934	135.97%	100.45%
VaR (99.9%)	\$41,186,362	\$27,504,616	\$27,582,252	149.74%	100.28%
2001 (most recent)					
	Cohort	**Homogeneous**	**Non-Homogeneous**	$\% \frac{\textbf{Cohort}}{\textbf{Homog.}}$	$\% \frac{\textbf{Non-homog.}}{\textbf{Homog.}}$
Mean horizon (1-yr) value	\$405,423,695	\$405,638,699	\$405,741,009	99.95%	100.03%
Std. dev. of horizon value	\$7,756,065	\$6,560,998	\$6,469,952	118.21%	98.61%
VaR (99%)	\$29,064,517	\$23,414,913	\$23,236,163	124.13%	99.24%
VaR (99.9%)	\$52,052,920	\$39,123,757	\$39,013,106	133.05%	99.72%

Table 8: Credit Risk Capital: Comparing Estimation Methods. 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 5.2.1.