Credit Rating Dynamics and Markov Mixture Models[*]

Halina Frydman
Stern School of Business, New York University[†]

Til Schuermann[‡]
Federal Reserve Bank of New York and Wharton Financial Institutions Center

July, 2004

Abstract

Despite overwhelming evidence to the contrary, credit migration matrices, used in many credit risk and pricing applications, are typically assumed to be generated by a simple Markov process. In this paper we propose a parsimonious model that is a mixture of (two) Markov chains. We estimate this model using credit rating histories and show that the mixture model statistically dominates the simple Markov model and that the differences between two models can be economically meaningful. The non-Markov property of our model implies that the future distribution of a firm's ratings depends not only on its current rating but also on its past rating history. Indeed we find that two firms with identical current credit ratings can have substantially different transition probability vectors.

Keywords: Risk management, credit risk, credit derivatives
JEL Codes: C13, C41, G12, G20

Credit Rating Dynamics and Markov Mixture Models

Abstract: Despite overwhelming evidence to the contrary, credit migration matrices, used in many credit risk and pricing applications, are typically assumed to be generated by a simple Markov process. In this paper we propose a parsimonious model that is a mixture of (two) Markov chains. We estimate this model using credit rating histories and show that the mixture model statistically dominates the simple Markov model and that the differences between two models can be economically meaningful. The non-Markov property of our model implies that the future distribution of a firm's ratings depends not only on its current rating but also on its past rating history. Indeed we find that two firms with identical current credit ratings can have substantially different transition probability vectors.

1 Introduction

In the study of credit rating dynamics of firms, it is very convenient to assume that the ratings process is Markov. For one, the credit migration or transition matrix, which characterizes past changes in credit quality of these firms, is all that is needed to generate forecasts of the credit asset portfolio distribution in the future. For another, given an estimate of this matrix over any horizon, say one year, one may generate a forecast over any other horizon, say two years, by simply raising that matrix to the appropriate power (in this case to the power two).

However, the evidence for non-Markovian behavior of a rating process is mounting. Altman and Kao (1992), Carty and Fons (1993), Altman (1998), Nickell, Perraudin and Varotto (2000), Bangia et al. (2002), Lando and Skødeberg (2002), Duffie and Wang (2003), Hamilton and Cantor (2004) and others have shown the presence of non-Markovian behavior such as ratings drift, industry heterogeneity and time variation due in particular to the business cycle. Christensen, Hansen and Lando (2004) take a step towards addressing the non-Markovian behavior, albeit at the expense of parsimony, by considering the possibility of hidden "excited" states for certain downgrades.

In this paper we present a non-Markovian model for credit ratings histories. The proposed model is a mixture of (two) Markov chains, and we estimate the Markov mixture with credit rating histories using methodology developed by Frydman (2004). We show that the mixture model statistically dominates the simple Markov model and that the differences between two models can be economically meaningful. The non-Markov property of our model implies that the future distribution of a firm's ratings depends not only on its current rating but also on the past history of its ratings. Thus, unlike in a Markov model, all firms with a particular current rating are not assigned the same future distribution of ratings. Instead, our approach provides a way of computing firm or rating history specific future distributions of ratings. Indeed we find that two firms with identical credit ratings can have substantially different transition probability vectors.

A rating by a credit rating agency represents an overall assessment of an obligor's creditworthiness. There is some disagreement between the rating agencies about what exactly is assessed. 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 (Cantor and Packer (1995), BIS (2000)[1]).

Credit ratings and consequently credit migrations find wide applications in finance. These include bond pricing models like Jarrow and Turnbull (1995) and Jarrow, Lando and Turnbull (1997), credit derivative pricing models like Kijima and Komoribayashi (1998) and Acharya, Das and Sundaram (2002), as well as credit portfolio models such as CreditMetrics[R] by Gupton, Finger and Bhatia (1997). This topic also has significant policy relevance given the pending new banking regulation around the New Basel Capital Accord where capital requirements are driven in part by ratings migration (BIS (2001, 2003)).

The algorithm for the maximum likelihood estimation of the mixture of Markov chains is presented in Frydman (2004) with an illustration that uses a small data set of Moody's credit rating histories. In this paper we go further and derive explicit expressions for the conditional distribution of a firm's ratings process by considering past ratings histories. We show that this distribution will differ depending on the information available. One may know the entire history, only a subset (say the last five years) or just the current rating. We then take the model to corporate credit rating histories from Standard & Poor's spanning 1981 to 2002, reject the Markov alternative against the mixture model, document a strong aging effect by considering the time profile of mixture matrices, find that there is significant industry heterogeneity in this time profile, and finally show that firm, or rather rating history specific transition probabilities can vary a lot, a source of heterogeneity which is obscured by the Markov approach. This is further illustrated using a bond pricing example.

The plan for the remainder of the paper is as follows: Section 2 defines a Markov mixture model, discusses its estimation from continuous credit ratings histories and derives some of its probabilistic properties. In Section 3 we provide a synopsis of the dataset, discuss the estimation results, and compare a mixture model with a simple Markov model empirically and in terms of economic implications. Section 4 provides some concluding remarks.

2 Markov Mixture Modeling

In this section we briefly summarize the methodology developed in Frydman (2004) by considering a mixture of two continuous time Markov chains which captures population heterogeneity in the rate of movement among states. A mixture of two Markov chains seems a natural extension of a simple Markov chain as a model for bond ratings migration.[2] We consider a continuous time

---

[1] See especially Annex I.B in BIS (2000).

[2] See Norris (1997) for an overview on Markov chains.

version of the mixture model because the changes in ratings are recorded up to a calendar day thus providing us with essentially continuous ratings histories.

The mixture process is a continuous time stochastic process $X = \{X(t), t > 0\}$ with state space $R = \{1, 2, ..., w\}$, representing the different credit ratings, which, conditionally on the initial state, is a mixture of two continuous Markov chains, $X_Q = \{X_Q(t), t > 0\}$ and $X_A = \{X_A(t), t > 0\}$, with generators A and Q, respectively. It is assumed that the generators are related by

$$A = \Gamma Q, \Gamma = diag(\gamma_1, \gamma_{2,}, ...\gamma_w). \tag{1}$$

The discrete mixing distribution on these Markov chains is defined conditionally on the initial state. For the initial state i, there is a separate mixing distribution,

$$
\begin{aligned}
s_i &= P(X_A \mid X(0) = i), \\
1 - s_i &= P(X_Q \mid X(0) = i), \quad 0 \le s_i \le 1,
\end{aligned}
$$

that is, s_i is the proportion of firms with initial rating i that evolve according to X_A, and $1 - s_i$ is the proportion evolving according to X_Q.

The transition matrices $P_Q(t), P_A(t)$ of Markov chains X_Q and X_A are given by $P_Q(t) = \exp(tQ)$ and $P_A(t) = \exp(tA), t \ge 0$, respectively and the transition probability matrix of the mixture process over a time period $(0, t)$ is given by

$$P(0, t) = SP_A(t) + (I - S)P_Q(t), t \ge 0, \tag{2}$$

where $S = diag(s_1, s_2, ..., s_w)$.

We consider a probabilistic interpretation of the relationship in (1). Generator Q is a matrix with entries q_{ij} satisfying

$$q_{ii} \le 0, q_{ij} \ge 0, \sum_{j \ne i} q_{ij} = -q_{ii} \equiv q_i, i \in R,$$

and these entries have a probabilistic interpretation: $1/(-q_{ii})$ is the expected length of time that X_Q remains in state i, and q_{ij}/q_i is the probability that when a transition out of state i occurs, it is to state j, $i \ne j \in R$. Thus, it follows from (1) that X_Q and X_A in general differ in the rates at which they leave the states (i.e., $q_i \ne a_i$), but both chains, when leaving state i, have the same probability distribution for entering state j, given by $q_{ij}/q_i = a_{ij}/a_i, j \ne i$. Thus, depending on whether $\gamma_i = 0, 0 < \gamma_i < 1$, or $\gamma_i > 1$, the realizations of X_A never move out of state i, move out of state i at a lower, or at a higher rate, respectively, than those generated by X_Q. In this way the generators govern the migration speed of the firms. If $\gamma_i = 0, 1 \le i \le w$, the model reduces to a particular two-component mixture known in the literature as a mover-stayer model (see Frydman (1985) and Frydman and Kadam (2004)). If $\gamma_i = 1, 1 \le i \le w$, the mixture process collapses to a simple Markov chain. In fact, this restriction will guide hypothesis testing later on.

4

The EM algorithm for the maximum likelihood estimation of mixture parameters from a set of continuously observed rating histories is presented in Frydman (2004). The algorithm estimates the following parameters: $(s_i, \gamma_i, q_i, 1 \leq i \leq w)$. Denoting the MLEs of these parameters by $(\hat{s}_i, \hat{\gamma}_i, \hat{q}_i, 1 \leq i \leq w)$, the off-diagonal entries in the Q matrix are then estimated by

$$\hat{q}_{ij} = \frac{N_{ij}}{N_i} \hat{q}_i, i \neq j \in R,$$

where N_{ij} is the total number of $i \rightarrow j$ transitions for all rating histories in a sample (i.e. aggregated over all firms) and $N_i = \sum_{j \neq i} N_{ij}$. The ML estimate of A is obtained from the relationship in (1).

Next, in order to compare empirically the mixture model with the simple Markov chain we will carry out the test of the null hypothesis $H_0 : \gamma_i = 1, 1 \leq i \leq w$, that the credit rating process is a simple Markov chain, against the hypothesis that it is a mixture of two Markov chains. We will also carry out a number of other tests concerning mixing parameters in which the null hypothesis is of the general form $H_0 : \gamma_i = 1, i \in R'$, where $R' \subset R$ is a relevant subset of credit ratings, for example all investment grade ratings. All tests are based on the likelihood ratio statistic, which under the null is asymptotically χ^2 distributed with degrees of freedom equal to the number of constraints imposed by a null hypothesis.[3]

An important feature of our model specifically and of Markov mixture models generally is that the distribution of the future state of a mixture process X, conditional on its current state, *does* depend on its past history in contrast to the Markov process. The ability to capture serial or path dependence in the data is key not just for our application but for applications in economics and finance more broadly.

We now derive the explicit expression for the conditional distribution of a future state of X and illustrate its dependence on the past history of the process with examples. Suppose that at time t the process is in state i, and we want to compute the probability distribution of its state, say, one period from now:

$$P(X(t+1) = j \mid X(t) = i, \mathcal{I}_{t-}) \; i, j \in R, \tag{3}$$

where $\mathcal{I}_{t-}$ is available information about realization of X up to time $t-$. Clearly, under a Markov assumption the information contained in $\mathcal{I}_{t-}$ would be irrelevant for the computation of (3); however, under a mixture assumption, past history, contained in $\mathcal{I}_{t-}$, does matter for assessing future behavior of a process currently in state i. Set $\mathcal{I}_{i,t} = \mathcal{I}_{t-} \cup \{X(t) = i\}$ to be the observed realization of the rating path, i.e. the observed rating history, ending at time t in state i. We show in Appendix A that

$$P(X(t+1) = j \mid \mathcal{I}_{i,t}) = P(X_A \mid \mathcal{I}_{i,t}) p_{ij}^A + P(X_Q \mid \mathcal{I}_{i,t}) p_{ij}^Q, j \in R, \tag{4}$$

where p_{ij}^A and p_{ij}^Q are entries in $P_A(1), P_Q(1)$, respectively. We will refer to (4) as the basic forecasting equation. Thus, the distribution of a state one period later is a convex linear combinaton

---

[3] For details, see Frydman (2004).

5

of one period transition probabilities of Markov chains. The weight $P(X_A \mid \mathcal{I}_{i,t})$ is the probability of X evolving according to X_A based on the information in $\mathcal{I}_{i,t}$ and its value depends on the available information. The set $\mathcal{I}_{t-}$ may represent the complete information about the past, partial information or may be empty. In the last case we only know that at present time t the firm has rating i but have no information on its rating history.

In Appendix A we show that in the case of complete information, that is, when $\mathcal{I}_{t-} = \{X(s), 0 \leq s \leq t-\}$, or equivalently, $\mathcal{I}_{i,t} \equiv \{X(s), 0 \leq s \leq t\}$, the weight is given by

$$P(X_A \mid \mathcal{I}_{i,t}) = \frac{s_{i_0} L_A}{s_{i_0} L_A + (1 - s_{i_0}) L_Q}, \tag{5}$$

where i_0 is the initial state and

$$L_A = P(\mathcal{I}_{i,t} \mid X_A, i_0), \quad L_Q = P(\mathcal{I}_{i,t} \mid X_Q, i_0),$$

are the likelihoods of observing the realization $\mathcal{I}_{i,t}$ under X_A and X_Q respectively, conditional on initial state i_0. To compute these likelihoods we use the following information derived from the observed realization $\mathcal{I}_{i,t}$:

$$
\begin{aligned}
n_{kj} &= \# \; k \to j \text{ transitions for } \mathcal{I}_{i,t}, k \neq j \in R \\
\tau_k &= \text{ total time in state } k \text{ for } \mathcal{I}_{i,t}, k \in R,
\end{aligned}
$$

that is n_{kj} are the transition counts from state k to state j for a particular firm with history $\mathcal{I}_{i,t}$. With this information the likelihood functions are

$$
\begin{aligned}
L_Q &= \prod_{k \neq j} (q_{kj})^{n_{kj}} \prod_k \exp(-q_k \tau_k), \\
L_A &= \prod_{k \neq j} (a_{kj})^{n_{kj}} \prod_k \exp(-a_k \tau_k).
\end{aligned} \tag{6}
$$

We see from (6) that the probability in (4) is realization-specific: two realizations currently in state i have different probability distributions of a future state unless they have identical transition counts (n_{kj}), total times in states (τ_k), and initial states.

In the second example we assume no information is available other than the current state so that $\mathcal{I}_{t-} = \varnothing$. The appropriate weight for this example is

$$s_i(t) \equiv P(X_A \mid X(t) = i, \mathcal{I}_{t-}) = P(X_A \mid X(t) = i),$$

and is equal to (see Appendix A),

$$s_i(t) = \frac{\sum_x s_x p_{xi}^A(t) \pi_x}{\sum_x \pi_x \left[s_x p_{xi}^A(t) + (1 - s_x) p_{xi}^Q(t) \right]}, \quad k \geq 0, i \in R, \tag{7}$$

where $p_{xi}^A(t)$ $\left(p_{xi}^Q(t)\right)$ is the (x, i) entry in the $P_A(t), (P_Q(t))$, and $\pi_x = P(X(0) = x)$ is the initial distribution of X. Note that $s_i(t)$ is the probability that a realization in state i at time t evolves according to X_A, so that $s_i(0) = s_i$. The corresponding one period transition probability in this example is

$$P(X(t+1) = j \mid X(t) = i) = s_i(t)p_{ij}^A + (1 - s_i(t))p_{ij}^Q, i, j \in R. \tag{8}$$

Putting conditional probabilities in (8) into a matrix gives a period $(t, t+1)$, i.e. $(t+1)^{st}$ period transition matrix of the mixture process in the form

$$P(t, t+1) = S(t)P_A + (I - S(t))P_Q, t \geq 0, \tag{9}$$

where $P_A \equiv \exp(A)$ and $P_Q \equiv \exp(Q)$, and $S(t)$ is the diagonal matrix with entries $s_i(t)$. This example shows that a one-period transition matrix changes over time measured on the age scale of a realization.

In our final example we assume that in addition to the current state we may also know the initial state of the realization. In this case the weight can be obtained from (7) by setting $\pi_{i_0} = 1$ to get

$$P(X_A \mid X(t) = i, X(0) = i_0) = \frac{s_{i_0}p_{i_0,i}^A(t)}{s_{i_0}p_{i_0,i}^A(t) + (1 - s_x)p_{i_0,i}^Q(t)}.$$

The realization-specific transition probability distribution given by (4) and (5), and the age-specific transition matrices in (9) as well as other quantities of interest can be easily estimated by substituting maximum likelihood estimates for the true values of the parameters. In the next section we estimate a mixture of Markov chains as well as the related quantities for modeling credit ratings migration.

3   Data, Estimation and Results

Our data set of S&P ratings histories, CreditPro V. 6.2, is the same as used in Jafry and Schuermann (2004) and is described in some detail in Bangia et al. (2002). The dataset contains rating histories from January 1, 1981 to December 31, 2002 of mainly large corporate institutions around the world. Ratings for sovereigns and municipals are not included, leaving 9,929 unique obligors. 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. For our analysis we will restrict ourselves to U.S. obligors only to control for an important source of heterogeneity (see, for instance, Nickell, Perraudin and Varotto (2000)); there are 6,776 unique U.S. domiciled obligors in the sample yielding 75,278 firm years of data, including withdrawn ratings, and 1,013 defaults of which 842 were from a rating, for an average default rate of 1.35%. If we exclude the withdrawn ratings the average default rate is 1.66%.

3.1 Estimating and comparing the models

We estimate Markov and two-regime mixture models at the whole rating level meaning that rating modifiers (+/-) are ignored. Credit ratings $\mathcal{CCC}$ through $\mathcal{C}$ are grouped into $\mathcal{CCC}$. This is customary, largely because of limited sample size. Including rating modifiers and all $\mathcal{C}$-subgrades would increase the migration matrix from 9×9 to 21×21. We do, however, include the 'not rated' or 'rating withdrawn' category as a state, denoted by $\mathcal{NR}$. The last column in Table 1 contains the number of firm years by rating, and here it is clear that the total $\mathcal{CCC}$-states are visited rarely: only 945 firm years were spent in this state, just 1.3% of the total. By contrast the $\mathcal{NR}$ state is visited quite frequently: 24,677 firm years, about one-third of the total.

Transitions to $\mathcal{NR}$ may be due to any of several reasons, including expiration of the debt, calling of the debt, etc. Unfortunately, however, the details of individual transitions to $\mathcal{NR}$ are not known. In particular, it is not known whether any given transition to $\mathcal{NR}$ is "benign" or "bad." Bad transitions to $\mathcal{NR}$ 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.

Table 1 summarizes the parameter estimates for the mixture model and reports the initial distribution of firms by rating. The first column contains the initial mixing proportions for the model. These proportions will change over time according to (7). All firms that have initial rating $\mathcal{AAA}$ or $\mathcal{CCC}$ are driven entirely by regime A as $s_{\mathcal{AAA}}$ and $s_{\mathcal{CCC}}$ both equal one. The fourth column denoted $1/q_i$ gives the expected duration (in years) in rating i under regime Q while the fifth column denoted $1/(q_i \cdot \gamma_i)$ gives the expected duration in i under regime A. It is clear that for more than half of the rating categories the A regime is slower or stickier than the Q regime, exceptions being $\mathcal{BBB}$, $\mathcal{BB}$ and $\mathcal{B}$. This can also be seen by looking at the third column denoted γ_i. Recall that if $0 < \gamma_i < 1$ then regime A is slower than Q, but if $\gamma_i > 1$ then A is faster than Q.

We conduct a number of likelihood ratio tests. We state the null hypothesis in terms of the constraints on the gamma parameters and test it against the alternative hypothesis that the data follow an unconstrained mixture model. We first test Markov chain (the null hypothesis specifies all gammas to be one) against a mixture model. This test overwhelmingly rejected Markov chain in favor of the mixture model ($LR = 392$; $p < 0.001$). These results are consistent with Frydman (2004) who used a much smaller sample of Moody's ratings over a shorter sample window (1985 - 1995). Motivated by the fact that the estimates of gammas for $\mathcal{BBB}$, $\mathcal{BB}$, $\mathcal{B}$ and $\mathcal{AA}$, are relatively close to 1, we next test the null hypothesis that $\gamma_{\mathcal{AA}} = \gamma_{\mathcal{BBB}} = \gamma_{\mathcal{BB}} = \gamma_{\mathcal{B}} = 1$. This time we can not reject the null hypothesis at the 99% confidence level ($LR = 10.05$; $p = 0.018$). Hence it appears that expected durations in these ratings are the same for all firms. Nevertheless we proceed with the unconstrained model for the remainder of the analysis.

Constraining additional gammas to be equal to one results in the rejection of the null hypothesis. Thus, the strong rejection of the Markov chain is due to very different behavior of regime A and Q firms in the ratings $\mathcal{AAA}$, $\mathcal{A}$ and $\mathcal{CCC}$. The best and worst ratings, $\mathcal{AAA}$ and $\mathcal{CCC}$ respectively, and rating $\mathcal{A}$ indeed have values of γ most different from one: $\gamma_{\mathcal{AAA}} = 0.219$, $\gamma_{\mathcal{A}} = 0.638$ and $\gamma_{\mathcal{CCC}} = 0.178$. Natural subsets such as all investment grade gammas restricted to one (or all sub-investment grade gammas) result in similar rejections. Finally we see that there are two very different behaviors in state $\mathcal{NR}$ ($\gamma_{\mathcal{NR}} = 0.056$). This may be interpreted as reflecting different reasons for rating withdrawal. In a Markov chain generated by A, $\mathcal{NR}$ acts as an absorbing state corresponding to debt expiration wheras in a Markov chain generated by Q, $\mathcal{NR}$ represents temporary rating withdrawal caused, say, by non-payment of a required fee.

Just how different are the Markov and mixture transition matrices? In Tables 2a and 2b we show the one-year Markov and mixture (for the second year[4]) migration matrices, and in Tables 2c and 2d we show the one-year migration matrices for the two regimes, $P_A \equiv \exp(A)$ and $P_Q \equiv \exp(Q)$.[5] The diagonal entries, denoting no migration, are bolded for clarity. Recall that the mixture matrix $P = SP_A + (I - S)P_Q$, where $S = diag(s_{\mathcal{AAA}}, s_{\mathcal{AA}}, ..., s_{\mathcal{NR}})$, the initial weights given by the first column in Table 1. Comparing first the Markov with the mixture matrices, Tables 2a and 2b, the differences are small with the exception of the $\mathcal{CCC}$ and $\mathcal{NR}$ states where, for example, the staying probability $p_{\mathcal{CCC}}$ is 41.99% for the Markov model and 49.07% for the mixture. The A regime dominates the Q in the mixture with an estimated 78.68% of firms moving according to regime A.[6] These results are also consistent with the proportions reported in Frydman (2004), albeit on a shorter sample using histories from a different rating agency.

The differences are more marked when comparing the migration matrices of the two regimes, P_A and P_Q shown in Tables 2c and 2d, especially for those ratings where the value of γ_i is most different from unity, namely for $\mathcal{AAA}$ and $\mathcal{CCC}$. For example, the one-year probability of default for $\mathcal{CCC}$-rated firms is 31.67% under the A regime but 67% under the Q regime. The default likelihood for an $\mathcal{AAA}$-firm is more than 60 times higher in the Q (0.403%) than in the A regime (0.006%) and is, in fact, nearly twice as large as the default probability for a $\mathcal{BBB}$-rated firm, at 0.215%, in the Q regime. The likelihood of exiting the $\mathcal{NR}$ state for a firm governed by the A regime is much smaller ($1 - 0.9916 = 0.84\%$) than for the Q regime ($1 - 0.86 = 14\%$), so over time the fraction of $\mathcal{NR}$ governed by A should increase. Indeed it does 76.84% in year 1 to 86.74% by year 10, as can be computed using (7). However, many other cells are very similar or nearly identical.

---

[4]Since no firm starts in category $\mathcal{NR}$, we do not estimate the initial proportion s for $\mathcal{NR}$ (see Table 1). Thus, in the first year transition matrix the entries in the $\mathcal{NR}$ row are not defined. However, we do have an estimate of this proportion, see (7), for any later time. In order to compare the entries in the $\mathcal{NR}$ row of the Markov and mixture transition matrices, we display the second year transition matrix (i.e. the matrix in (9) when $t = 1$).

[5]The last row is trivially the unit vector, needed to make the matrix square, and has been omitted in the tables.

[6]The estimated overall proportion of histories (0.7868) moving according to regime A was computed by weighing the estimated mixing proportions by the initial distribution (see Table 1).

For example, $p_{\mathcal{BBB}\to\mathcal{AA}} = 0.25\%$ for both the A and Q regime. This makes clear the difficulties in casually comparing high dimensional objects such as credit migration matrices. We therefore make use of the scalar metric developed by Jafry and Schuermann (2004) in obtaining a measure of migration "size" which we briefly describe here.

Jafry and Schuermann (2004) propose a metric for comparing migration matrices based on singular values. Let P be the migration matrix of dimension w, and define the mobility matrix $\tilde{P} = P - I$, where I is an identity matrix of dimension w, i.e. the static (no migration) matrix. 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. The final metric M_{SVD} is simply the average of the singular values of $\tilde{P}$:

$$M_{SVD}(P) = \frac{1}{w} \sum_{i=1}^{w} \sqrt{\lambda_i}, \tag{10}$$

where λ_i are the eigenvalues of $\tilde{P}'\tilde{P}$. Jafry and Schuermann (2004) show that M_{SVD} approximates the average migration probability in P and satisfies several criteria for metrics M proposed in the literature including monotonicity (larger off-diagonal probabilities should yield larger values of M) and distribution discriminatory (the metric M should be more sensitive to far than to near migrations).

Denote P_M to be the Markov migration matrix and recall that $P(t)$ is the mixture model migration matrix for the tth year, with P_A and P_Q the two regime migration matrices. Then $M_{SVD}(P_M) = 0.199$, $M_{SVD}(P(2)) = 0.190$, $M_{SVD}(P_A) = 0.184$ and $M_{SVD}(P_Q) = 0.300$. This makes quite clear just how different the Q regime is from the A regime or the mixture which is, of course, dominated by the A regime.

3.2 Dynamic evolution of the mixture model

A feature of the mixture model is that the migration matrix is not static over time. The first period matrix will be different from the T-period matrix as the mixing proportions of the A and Q regimes will change over time. While $M_{SVD}(P(2)) = 0.190$, the average over twenty years $M_{SVD}\left(\overline{P(t)} = 0.189\right)$, meaning that the mixture matrix is less dynamic than the Markov matrix. In Figure 1 we show the evolution of $M_{SVD}(P(t))$ for $t = 2, ..., 21$, with $M_{SVD}(P_M)$ included as a visual reference.

The mixture model displays a clear vintage effect of ratings. Ratings dynamics are initially high and then decline rapidly over the first three to four years only to rise again after about a dozen years or so. By the time firms have been in the sample twenty years, their rating dynamics are similarly high as in the initial years. Altman (1998) argued for the importance of an aging or vintage effect in considering credit ratings migration, and here we are able to document for the first time more precisely this time profile. Altman also speculated what the impact of the NR category
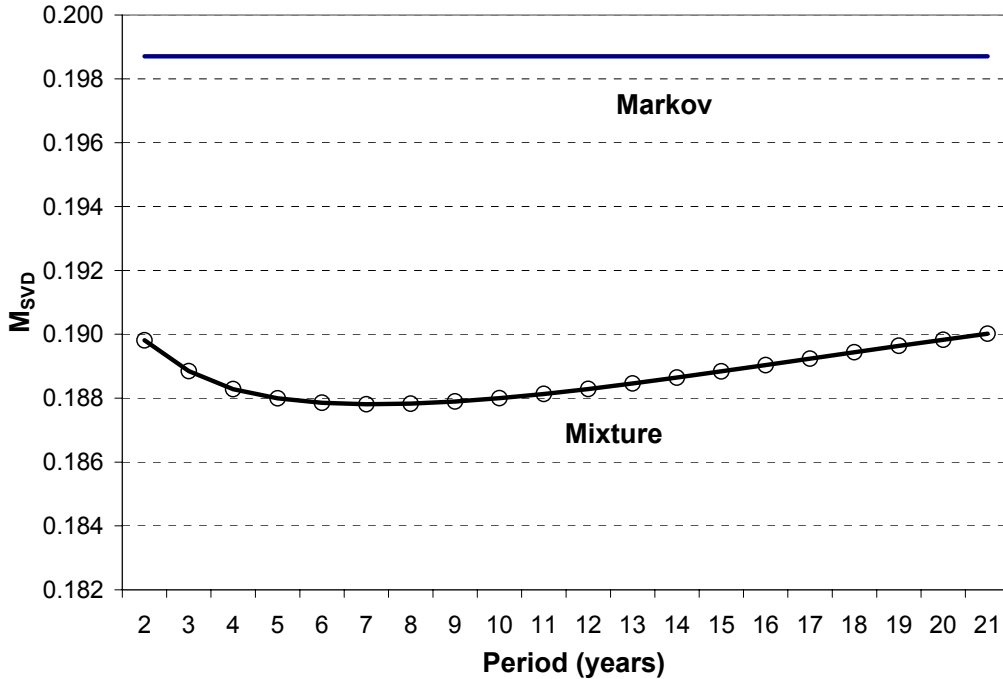
Figure 1: Time evolution of Markov and mixture matrices as measured by M_{SVD}

might be on the measured rating matrices and their dynamics. We re-compute the Markov and mixture matrices by following the usual approach in the literature: we proportionately re-allocate the probability of migration to $\mathcal{NR}$ to all other probabilities in that row. The results do change qualitatively. For instance, the mixture model curve in Figure 1 retains its shape but is shifted downwards reflecting the absence of a state with a high into migration rate, reducing the migration metric M_{SVD}.

3.3 Conditional mixture models: red or blue

Any given firm's rating evolution is governed by either the A or the Q process in the mixture model. However, since these regimes are not directly observed, there is no straight forward way to assign a particular firm to a regime. Put another way, a firm is either "red" or "blue", but we know neither its color nor the markers which might determine its color. A full treatment of this issue is well beyond the scope of this paper. However, it is well known that firm rating dynamics exhibit heterogeneity with regard to industry (Nickell, Perraudin and Varotto (2000)), and so we estimate Markov as well as mixture models by major industry groupings. Of the seven groups, two (agriculture, mining & construction, non-durable manufacturing) did not have sufficient number of observations for the EM algorithm to converge. Hence we report results only for the other five industry groups.

These results are summarized in Table 3 where we show the value of M_{SVD} by industry for the Markov and mixture models. As a benchmark we include the unconditional (all industries) model in the first column. The difference in dynamics between the A and Q regime is quite large: the M_{SVD} of P_A and P_Q for all industries differ by 63% while the Markov and the average (over 20 years) mixture model M_{SVD} differ by only 5.1%. Conditioning on industry seems to have a big impact for some but not all industry groups, suggesting that industry is a useful but not sufficient marker to help identify whether a firm is "blue" or "red." Focusing on the last two rows of the table across the different industries, we see that $M_{SVD}(P_A)$ and $M_{SVD}(P_Q)$ differ little for the finance, insurance and real estate (FIRE) sector ($\%\Delta = 37\%$) and service sector ($\%\Delta = 18\%$) compared with durable manufacturing ($\%\Delta = 75\%$) and wholesale, retail and trade ($\%\Delta = 73\%$). The FIRE sector is the least dynamic, whether using the Markov ($M_{SVD} = 0.189$) or the mixture model $\left(M_{SVD}\left(\overline{P(t)}\right) = 0.173 \right)$. The wholesale, retail and trade sector is the most dynamic for the Markov, and here the difference between the Markov and mixture models are most pronounced: $\%\Delta \left[M_{SVD}(P_M), M_{SVD}\left(\overline{P(t)}\right) \right] = 10.2\%$, twice as much as for the "all industries" model. Indeed the dynamics implied by the Markov model of this sector ($M_{SVD}(P_M) = 0.223$) are nearly the same as for the services sector where $M_{SVD}(P_M) = 0.221$. These two sectors are much more differentiated in the mixture model: $M_{SVD}\left(\overline{P(t)}\right)$ is 0.186 for wholesale, retail and trade and is 0.217 for the services sector.

Not only are there differences between the Markov and the average mixture model, but the time profile of the mixture models by industry also exhibit heterogeneity. This is captured in Figure 2 where the "all industries" time profile appears in the middle for easy comparison. All of the sectors share the convex M_{SVD} time profile with the "all industries" model except for communication, electric & gas which has a concave profile. This is likely due to the prevalence of long-maturity bonds in this industry, in turn driven by long lived assets (e.g. power stations). What is striking either way is how varied these profiles are, a source of heterogeneity that is clearly ignored by the Markov model.

3.4 Firm-specific migration vectors

At the extreme the mixture approach allows one to estimate firm or firm-history specific migration vectors. What is the range of one-year migration vectors across all firms of a given rating, say $\mathcal{A}$? We compute the one-year migration vector for all firms in our sample at the end of the sample period, namely December 31, 2002, conditional on the rating history available for each firm. These computations are done with the basic forecasting equation (4) using the weights in (7). The variation and range of those migration probabilities is rather substantial and is summarized in Table 4a for firms whose current rating is investment grade, $\mathcal{AAA}$ through $\mathcal{BBB}$, and in Table 4b for speculative (high yield) grades $\mathcal{BB}$ and below. The range from min to max in particular can be
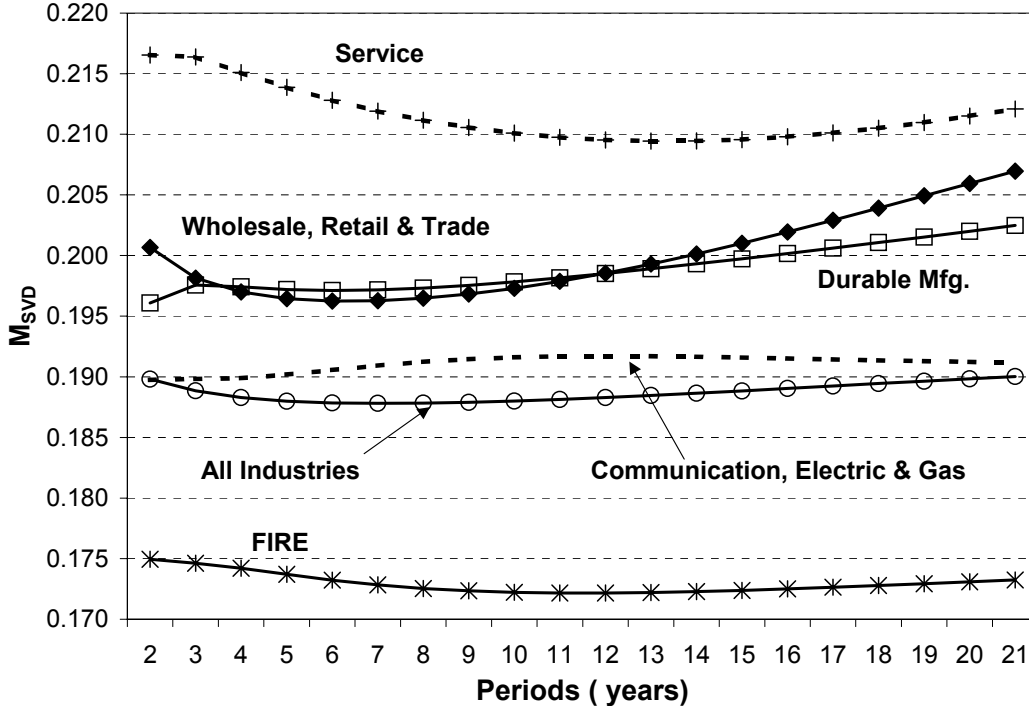
Figure 2: Time evolution of mixture matrices by major industry group (M_{SVD})

dramatic for the important default state. For example, the one-year ahead default probability for $\mathcal{A}$-rated firms ranges from 0.010% to 0.153%. While the relative difference is large, the absolute difference is not. This changes when we move down the rating spectrum towards the speculative grades. For instance, the default probability for $\mathcal{B}$-rated firms ranges from 4.110% to 6.467%. This is non-trivial since more firms were "born" with this rating than with any other rating (column 6 in Table 1). Moreover, about half of the corporate bond high yield market is rated $\mathcal{B}$ (including $\mathcal{B}+$ and $\mathcal{B}-$, of course).[7] For the $\mathcal{CCC}$-rated firms the range is 31.67% to 66%! In fact, the probability of no change has an even wider range: from 4.70% to 53.83%.

The minima and maxima should broadly correspond to one of the underlying Markov mixture matrices, i.e. P_A and P_Q, shown in Tables 2c and 2d respectively. For example, the maximum staying probability for $\mathcal{AAA}$ is 89.94% which corresponds to the A-regime staying probability; similarly for the $\mathcal{AA}$ rating. However, the maximum upgrade to $\mathcal{AAA}$ probability computed for $\mathcal{AA}$-rated firms, $p_{\mathcal{AA}\rightarrow\mathcal{AAA}}$ at 0.51%, corresponds to the Q-regime. Thus one regime does not necessarily correspond to either the mimimum or maximum migration probability. Perhaps the best

---

[7]As of April 2004, the $\mathcal{B}$ rating made up about 54% of the Merrill Lynch U.S. high yield index by number of issuers and about 45% by amount outstanding. For the $\mathcal{BB}$ rating it was 26% (33%) by issuers (amount) and for the $\mathcal{CCC}$ (and lower) rating 20% (21%). The total size of the high yield market as measured by the Merrill Lynch U.S. high yield index was around $580bn.

illustration of this regime mixing is the $\mathcal{B}$-rating (Table 4b). The maxima for the non-downgrade probabilities, namely $p_{\mathcal{B}\to\mathcal{AAA}}$ through $p_{\mathcal{B}\to\mathcal{B}}$ correspond to the Q-regime, while the maximum probabilities $p_{\mathcal{B}\to\mathcal{CCC}}$ and $p_{\mathcal{B}\to\mathcal{NR}}$ correspond to the A-regime. The reverse is true for the minima.

The $\mathcal{CCC}$ rating state seems more unusual than the others as it exhibits the most significant variation in migration vectors across firms. Firms "born" as $\mathcal{CCC}$ firms have a longer average duration (1.6 years) than firms downgraded into $\mathcal{CCC}$ (about $3\frac{1}{2}$ months). This is consistent with evidence presented by the rating agencies. For instance, Mah and Verde (2004) at Fitch find that an issuer rated $\mathcal{CCC}$ or lower that had been downgraded the previous year had a nearly 60% probability of defaulting the subsequent year.

3.5 Financial impact of mixture models

It is one thing to show that the Markov model is rejected by the mixture model; but does it matter economically? By way of illustration, we take fictitious three-year coupon paying corporate bonds (coupons are paid once a year), one per rating category, and compute their value at the end of one year using three different transition vectors. The baseline is, of course, the vector given by the Markov model. This is compared to the vector implied by the maximum and minimum staying probability from the mixture model, denoted by V1 and V2 in Table 5, which turns out to be the same as the max/min default probability for a given rating. Since the $\mathcal{NR}$ rating is not priced, we follow the conventional approach and re-allocate $p_{\mathcal{NR}}$ proportionately across the other ratings which assumes that transition to $\mathcal{NR}$ are non-informative (Carty (1997)).

We take the credit spreads and risk-free rate (proxied by the 10Y U.S. constant maturity Treasury rate) that prevailed on March 23, 2004, and assign coupons so that pricing does not depart too significantly from par $= 100$. The forward rate is derived in the standard way using the expectations hypothesis (see Saunders and Allen (2002)). In the event of default we assume a 40% recovery rate on par. The results are summarized in Table 5. The second column gives the price under the Markov model, the third and fifth under the mixture model, and the second and fourth columns shows the percentage differences to the Markov model. The last column contains the coupon payments. What stands out is the difference in pricing for the $\mathcal{CCC}$ rating: it can range from nearly 5% to 20% relative to the value implied by the Markov model. The difference is noticably less for the other rating categories, though it approaches one percentage point for the $\mathcal{B}$ category. Interestingly, γ alone is not a sufficient guide for anticipating pricing differences as $\gamma_{\mathcal{B}} = 1.079$ is not very different from unity. To be sure, this is merely illustrative, and different term structures of credit spreads would generate different valuation differences, some higher, some lower.

4   Concluding Remarks

We have presented a method for generating firm or rating history specific transition vectors using an alternative to the popular Markov approach to estimating credit transitions, namely Markov mixture models. We showed that the Markov model is rejected in favor of the mixture model. Indeed we find that two firms with identical credit ratings can have substantially different transition probability vectors. The model differences matter especially for the extreme ratings: $\mathcal{AAA}$ and $\mathcal{CCC}$.

The mixture model we estimate contains two regimes. More complicated regime structure could be explored, although it is not clear if the reduction in parsimony is merited. These regimes are not directly observed, but we can assess the likelihood of belonging to a regime for a given firm conditional on its rating history. The remaining challenges are the classification of firms, possibly with exogenously observable variables or markers. This is a natural topic of future research.

References

[1] Acharya, V.V., S.R. Das and R.K. Sundaram (2002), "Arbitrage-Free Pricing of Credit Derivatives with Rating Transitions," *Financial Analysts Journal* 58 (3), 28-44.

[2] Altman, Edward I. (1998), "The Importance and Subtlety of Credit Rating Migration," *Journal of Banking & Finance* 22, 1231-1247.

[3] Altman, Edward I. and D.L. Kao (1992), "Rating Drift of High Yield Bonds," *Journal of Fixed Income*, March, 15-20.

[4] Bangia, Anil, Francis X. Diebold, André Kronimus, Christian Schagen and Til Schuermann (2002), "Ratings Migration and the Business Cycle, With Applications to Credit Portfolio Stress Testing," *Journal of Banking & Finance,* 26 (2-3), 235-264.

[5] BIS (2003), "Third Consultative Paper," Basel Committee on Banking Supervision, April; available at http://www.bis.org/bcbs/bcbscp3.htm.

[6] BIS, (2001), "The New Basel Capital Accord," Basel Committee on Banking Supervision, January; available at <http://www.bis.org/publ/bcbsca03.pdf>.

[7] BIS (2000), "Credit Ratings and Complementary Sources of Credit Quality Information," Basel Committee on Banking Supervision, August; available at <http://www.bis.org/publ/bcbs_wp03.pdf>.

[8] Cantor, Richard and Frank Packer (1995), "The Credit Rating Industry," *Journal of Fixed Income*, December, 10-34.

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

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

[11] Christensen, Jens, Ernst Hansen, and David Lando (2004), "Confidence Sets for Continuous-Time Rating Transition Probabilities," forthcoming, *Journal of Banking & Finance.*

[12] Duffie, D. and Wang, K. (2003), "Multi-Period Corporate Failure Prediction with Stochastic Covariates," Working paper, Stanford University.

[13] Frydman, Halina (1984) "Maximum likelihood Estimation in the Mover-Stayer Model." *Journal of the American Statistical Association* 79, 632-638.

[14] Frydman, Halina and Ashay Kadam (2004), "Estimation in the Continuous Time Mover-Stayer Model with an Application to Bond Ratings Migration," *Applied Stochastic Models in Business and Industry* 20, 155-170.

[15] Frydman, Halina (2004), "Estimation in the Mixture of Markov Chains Moving with Different Speeds," working paper, NYU Stern School of Business.

[16] Gupton, G.M., C.C. Finger and M. Bhatia, 1997, *CreditMetrics^{TM} – Technical Document*, this version: April 2, 1997, J.P. Morgan, New York, NY.

[17] Hamilton, D. T. and Cantor, R. (2004), "Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History," Moody's Investor Service, *Special Comment,* February.

[18] Jafry, Yusuf and Til Schuermann (2004), "Measurement, Estimation and Comparison of Credit Migration Matrices", forthcoming, *Journal of Banking & Finance.*

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

[20] Jarrow, Robert A., David Lando, and Stuart M. Turnbull (1997), "A Markov Model for the Term Structure of Credit Risk Spreads," *The Review of Financial Studies* 10 (2), 481-523.

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

[22] Lando, David and Torben Skødeberg (2002), "Analyzing Ratings Transitions and Rating Drift with Continuous Observations," *Journal of Banking & Finance* 26 (2-3), 423-444.

[23] Mah, S.K. and M. Verde (2004), "Rating Path Dependency," Fitch Ratings *Report on Structured Finance*, March 4.

[24] Nickell, P., William Perraudin and S. Varotto (2000), "Stability of Rating Transitions," *Journal of Banking & Finance* 24 (1-2), 203-227.

[25] Norris, J. R. (1997), *Markov Chains,* Cambridge, UK: Cambridge University Press.

[26] Saunders, Anthony and Linda Allen (2002), *Credit Risk Measurement- New Approaches to Value at Risk and Other Paradigms*, 2nd Ed., New York: John Wiley & Sons.

Table 1

Parameter Estimates for Mixture Model

Credit Rating	s_i	q_i	γ_i	$1/q_i$	$1/(q_i \cdot \gamma_i)$	Initial Distribution	Total firm years
$\mathcal{AAA}$	1.000	0.485	0.219	2.062	9.415	311	2,495
$\mathcal{AA}$	0.941	0.153	0.821	6.520	7.937	652	6,940
$\mathcal{A}$	0.935	0.197	0.638	5.065	7.941	1,283	13,556
$\mathcal{BBB}$	0.820	0.150	1.146	6.675	5.823	1,171	10,606
$\mathcal{BB}$	0.675	0.221	1.285	4.533	3.527	1,333	7,445
$\mathcal{B}$	0.645	0.262	1.079	3.814	3.534	1,915	8,612
$\mathcal{CCC}$	1.000	3.498	0.178	0.286	1.604	111	945
$\mathcal{NR}$	0.000	0.155	0.056	6.450	115.108	–	24,677

Parameter estimates for the Markov mixture model and descriptive statistics

of the sample by rating category.

S&P U.S. corporate obligor histories, 1981-2002.

Table 2a

Markov Migration Matrix

(in percent)

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**89.73**	5.03	0.56	0.07	0.08	0.01	0.00	4.50	0.016
$\mathcal{AA}$	0.50	**88.11**	6.81	0.62	0.10	0.08	0.01	3.74	0.018
$\mathcal{A}$	0.08	1.82	**87.90**	5.19	0.47	0.18	0.01	4.31	0.024
$\mathcal{BBB}$	0.04	0.24	4.09	**84.88**	4.69	0.75	0.08	5.10	0.118
$\mathcal{BB}$	0.04	0.11	0.58	4.97	**77.30**	7.94	0.74	7.71	0.619
$\mathcal{B}$	0.01	0.08	0.27	0.49	3.99	**76.45**	5.16	9.10	4.464
$\mathcal{CCC}$	0.07	0.01	0.30	0.53	1.08	6.55	**41.99**	9.65	39.824
$\mathcal{NR}$	0.03	0.08	0.26	0.40	0.42	0.38	0.03	**97.69**	0.700

Annual Markov credit migration matrix

S&P U.S. corporate obligor histories, 1981-2002.

Table 2b

Markov Mixture Migration Matrix (2nd Year)*

(in percent)

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**89.88**	4.95	0.54	0.07	0.07	0.01	0.00	4.47	0.007
$\mathcal{AA}$	0.50	**88.11**	6.82	0.62	0.10	0.08	0.01	3.76	0.015
$\mathcal{A}$	0.08	1.82	**87.96**	5.16	0.47	0.18	0.01	4.30	0.023
$\mathcal{BBB}$	0.04	0.25	4.09	**84.86**	4.69	0.76	0.08	5.11	0.126
$\mathcal{BB}$	0.04	0.11	0.58	4.99	**77.31**	7.96	0.69	7.64	0.688
$\mathcal{B}$	0.01	0.08	0.29	0.52	4.04	**76.55**	4.52	9.04	4.944
$\mathcal{CCC}$	0.06	0.02	0.27	0.48	0.96	5.74	**49.07**	8.41	35.000
$\mathcal{NR}$	0.05	0.14	0.43	0.68	0.72	0.64	0.03	**96.11**	1.197

Annual Markov mixture credit migration matrix

* The 2nd year is used since no firm starts in category $\mathcal{NR}$, we do not estimate the initial proportions for $\mathcal{NR}$ (see Table 1). Thus, in the first year transition matrix the entries in the $\mathcal{NR}$ row are not defined. However, we do have an estimate of this proportion, see (7), for any later time.

S&P U.S. corporate obligor histories, 1981-2002.

Table 2c

$P_A \equiv \exp(A)$, **Migration Matrix for Regime 1**

(in percent)

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**89.94**	4.93	0.54	0.06	0.07	0.01	0.00	4.45	0.006
$\mathcal{AA}$	0.50	**88.24**	6.75	0.60	0.09	0.07	0.01	3.72	0.009
$\mathcal{A}$	0.08	1.76	**88.36**	4.98	0.45	0.17	0.01	4.18	0.013
$\mathcal{BBB}$	0.04	0.25	4.20	**84.50**	4.75	0.78	0.09	5.28	0.105
$\mathcal{BB}$	0.04	0.11	0.62	5.29	**75.70**	8.44	0.89	8.32	0.593
$\mathcal{B}$	0.00	0.08	0.26	0.49	4.02	**75.83**	5.92	9.29	4.108
$\mathcal{CCC}$	0.05	0.01	0.23	0.41	0.84	5.21	**53.83**	7.74	31.670
$\mathcal{NR}$	0.01	0.03	0.10	0.15	0.15	0.14	0.01	**99.16**	0.255

Annual Markov mixture credit migration matrix for regime 1 (A)

S&P U.S. corporate obligor histories, 1981-2002.

Table 2d

$P_Q \equiv \exp(Q)$, **Migration Matrix for Regime 2**

(in percent)

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**61.65**	18.50	2.32	0.51	0.52	0.24	0.01	15.83	0.403
$\mathcal{AA}$	0.50	**85.98**	7.86	0.91	0.18	0.14	0.01	4.30	0.113
$\mathcal{A}$	0.10	2.65	**82.42**	7.69	0.76	0.32	0.01	5.89	0.159
$\mathcal{BBB}$	0.03	0.25	3.61	**86.42**	4.39	0.70	0.03	4.35	0.215
$\mathcal{BB}$	0.03	0.11	0.52	4.38	**80.58**	6.97	0.29	6.24	0.882
$\mathcal{B}$	0.02	0.09	0.32	0.58	4.07	**77.87**	1.98	8.60	6.473
$\mathcal{CCC}$	0.11	0.10	0.65	1.16	2.13	10.79	**3.28**	14.80	66.996
$\mathcal{NR}$	0.18	0.50	1.53	2.46	2.60	2.31	0.09	**86.00**	4.322

Annual Markov mixture credit migration matrix for regime 2 (Q)

S&P U.S. corporate obligor histories, 1981-2002.

Table 3

$M_{\mathbf{SVD}}$ **Comparisons**

	All Industries*	Durable Mfg.	Communic., Electric & Gas	Wholesale, Retail & Trade	FIRE**	Service
# of firms	6,776	1,307	1,033	601	1,577	651
P_M	0.199	0.208	0.202	0.223	0.189	0.221
$P(2)$***	0.190	0.196	0.190	0.201	0.175	0.217
$\overline{P(t)}$	0.189	0.199	0.191	0.200	0.173	0.212
P_A	0.184	0.173	0.187	0.186	0.224	0.211
P_Q	0.300	0.303	0.312	0.321	0.307	0.250
$\%\Delta\left[P_A, P_Q\right]$	63%	75%	67%	73%	37%	18%
$\%\Delta\left[P_M, \overline{P(t)}\right]$	5.1%	4.2%	5.3%	10.2%	8.7%	4.4%

Comparison of matrix metric $M_{\mathbf{SVD}}$ across matrices and industries. $M_{\mathbf{SVD}}$ defined in equation (10).

 * "All industries" also includes two sectors, agriculture, mining&construction, non-durable manufacturing

   that did not have enough observations to allow for estimation of a separate mixture model.

** FIRE: finance, insurance and real estate.

*** Markov mixture matrix for year 2 (as in Table 2b above)

S&P U.S. corporate obligor histories, 1981-2002.

Table 4a

Firm-specific Migration Vectors Under Mixture Model

Investment Grade Firms

(in percent)

State at T		AAA	AA	A	BBB	BB	B	CCC	NR	D
					State at $T+1$					
AAA	$\widehat{\mu}$	89.78	5.01	0.55	0.07	0.07	0.01	0.0001	4.52	0.012
	$\widehat{\sigma}$	0.79	0.38	0.05	0.01	0.01	0.01	0.0003	0.32	0.011
	min	83.63	4.93	0.54	0.07	0.07	0.01	0.000	4.45	0.010
	max	89.94	7.95	0.94	0.17	0.17	0.06	0.002	6.99	0.097
AA	$\widehat{\mu}$	0.49	88.10	6.82	0.62	0.10	0.07	0.01	3.75	0.016
	$\widehat{\sigma}$	0.002	0.21	0.11	0.03	0.01	0.01	<0.001	0.06	0.010
	min	0.49	86.34	6.75	0.60	0.09	0.07	0.01	3.72	0.010
	max	0.51	88.24	7.69	0.86	0.17	0.14	0.01	4.21	0.094
A	$\widehat{\mu}$	0.08	1.83	87.87	5.20	0.48	0.18	0.02	4.32	0.022
	$\widehat{\sigma}$	0.003	0.12	0.79	0.36	0.04	0.02	0.001	0.23	0.020
	min	0.08	1.76	82.70	4.98	0.45	0.17	0.01	4.18	0.010
	max	0.10	2.61	88.36	7.57	0.75	0.31	0.02	5.81	0.153
BBB	$\widehat{\mu}$	0.04	0.25	4.07	84.93	4.67	0.76	0.08	5.07	0.134
	$\widehat{\sigma}$	0.002	<0.001	0.11	0.35	0.07	0.01	0.01	0.17	0.020
	min	0.03	0.25	3.61	84.51	4.39	0.71	0.03	4.35	0.110
	max	0.04	0.25	4.20	86.41	4.75	0.78	0.09	5.28	0.219

Descriptive statistics for firm-specific migration vectors using Markov mixture model

at the end of the sample period. Investment grades.

S&P U.S. corporate obligor histories, 1981-2002.

Table 4b

Firm-specific Migration Vectors Under Mixture Model

Speculative Grade Firms

(in percent)

State at T		State at $T+1$								
		$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{BB}$	$\widehat{\mu}$	0.04	0.12	0.58	4.97	77.44	7.92	0.67	7.58	0.639
	$\widehat{\sigma}$	0.002	0.002	0.02	0.16	0.86	0.26	0.11	0.37	0.51
	min	0.03	0.11	0.52	4.38	75.70	6.98	0.28	6.25	0.590
	max	0.04	0.12	0.62	5.29	80.56	8.44	0.89	8.32	0.879
$\mathcal{B}$	$\widehat{\mu}$	0.01	0.08	0.28	0.53	4.04	76.58	4.46	9.04	4.980
	$\widehat{\sigma}$	0.003	0.004	0.01	0.01	0.01	0.30	0.59	0.10	0.353
	min	<0.001	0.07	0.26	0.49	4.02	75.83	1.97	8.60	4.110
	max	0.02	0.10	0.32	0.59	4.07	77.87	5.92	9.29	6.467
$\mathcal{CCC}$	$\widehat{\mu}$	0.07	0.02	0.28	0.50	0.99	5.85	48.00	8.55	35.745
	$\widehat{\sigma}$	0.01	0.02	0.07	0.13	0.22	0.97	8.80	1.23	6.152
	min	0.06	0.01	0.23	0.41	0.84	5.21	4.70	7.74	31.670
	max	0.11	0.10	0.64	1.14	2.09	10.63	53.83	14.60	66.005

Descriptive statistics for firm-specific migration vectors using Markov mixture model

at the end of the sample period. Speculative grades.

S&P U.S. corporate obligor histories, 1981-2002.

Table 5

Bond Pricing Example

	Markov	V1 Mixture	% Δ V1 to Markov	V2 Mixture	% Δ V2 to Markov	Coupon (annual)
$\mathcal{AAA}$	101.973	101.909	0.063%	101.969	0.004%	2
$\mathcal{AA}$	102.635	102.583	0.051%	102.634	0.001%	2.5
$\mathcal{A}$	102.986	102.910	0.073%	102.985	0.0004%	3
$\mathcal{BBB}$	104.600	104.557	0.042%	104.594	0.006%	4
$\mathcal{BB}$	104.431	104.316	0.109%	104.424	0.006%	5
$\mathcal{B}$	103.998	103.028	0.933%	104.080	0.079%	6.5
$\mathcal{CCC}$	87.103	70.905	18.597%	90.827	4.275%	8.5

Example of pricing two bonds, V1 and V2, using Markov and mixture models.

V1 and V2 were chosen to correspond to the maximum and minimum migration to default probability of the firm-specific migration vectors respectively.

S&P U.S. corporate obligor histories, 1981-2002.

A Appendix

We first derive the basic forecasting equation in a slightly more general form than stated in (4). Let $p_{ij}^A(s)$ denote the i, j th entry in the transition probability matrix $P_A(s)$ and similarly for $p_{ij}^B(s)$.

Lemma: For any two times $\tau > t$,

$$P(X(\tau) = j \mid \mathcal{I}_{i,t}) = P(X_A \mid \mathcal{I}_{i,t}) p_{ij}^A(\tau - t) + P(X_Q \mid \mathcal{I}_{i,t}) p_{ij}^Q(\tau - t), i, j \in R. \quad (11)$$

Proof:

$$
\begin{aligned}
P(X(\tau) = j \mid \mathcal{I}_{i,t}) &= P(X(\tau) = j, X_A \mid \mathcal{I}_{i,t}) + P(X(\tau) = j, X_Q \mid \mathcal{I}_{i,t}) \\
&= P(X(\tau) = j \mid X_A, \mathcal{I}_{i,t}) P(X_A \mid \mathcal{I}_{i,t}) \\
&\quad + P(X(\tau) = j \mid X_Q, \mathcal{I}_{i,t}) P(X_Q \mid \mathcal{I}_{i,t}) \\
&= P(X_A \mid \mathcal{I}_{i,t}) p_{ij}^A(\tau - t) + P(X_Q \mid \mathcal{I}_{i,t}) p_{ij}^Q(\tau - t),
\end{aligned}
$$

where the first equality follows by the law of total probability, the second by the property of conditional probability and the last one by the Markov property of X_A and X_Q.

The equation in (4) is obtained by setting $\tau = t + 1$ and $p_{ij}^A = p_{ij}^A(1)$. In (11) the weight $P(X_A \mid \mathcal{I}_{i,t})$ depends on the information $\mathcal{I}_{i,t}$. Below we derive this weight under two different scenarios.

In the first scenario $\mathcal{I}_{i,t} = \{X(s), 0 \leq s \leq t\}$, and the weight is

$$
\begin{aligned}
P(X_A \mid \mathcal{I}_{i,t}) &= \frac{P(X_A, \mathcal{I}_{i,t})}{P(\mathcal{I}_{i,t})} \\
&= \frac{P(\mathcal{I}_{i,t} \mid X_A, i_0) P(X_A \mid i_0)}{P(\mathcal{I}_{i,t} \mid X_A, i_0) P(X_A \mid i_0) + P(\mathcal{I}_{i,t} \mid X_Q, i_0) P(X_Q \mid i_0)} \\
&= \frac{P(\mathcal{I}_{i,t} \mid X_A, i_0) s_{i_0}}{P(\mathcal{I}_{i,t} \mid X_A, i_0) s_{i_0} + P(\mathcal{I}_{i,t} \mid X_Q, i_0)(1 - s_{i_0})} \\
&= \frac{s_{i_0} L_A}{s_{i_0} L_A + (1 - s_{i_0}) L_Q},
\end{aligned}
$$

which is (5).

When the information set is $\mathcal{I}_{i,t} = \{X(t) = i\}$ the weight becomes

$$
\begin{aligned}
P(X_A \mid \mathcal{I}_{i,t}) &= \frac{\sum_x P(X_A, X(t) = i, X_0 = x)}{P(X(t) = i)} \\
&= \frac{\sum_x P(X(t) = i \mid X_A, X_0 = x) P(X_A \mid X_0 = x) P(X_0 = x)}{P(X(t) = i)} \\
&= \frac{\sum_x \pi_x s_x p_{xi}^A(t)}{\sum_x \pi_x \left[s_x p_{xi}^A(t) + (1 - s_x) p_{xi}^Q(t) \right]},
\end{aligned}
$$

which gives (7). In the above two derivations we used the property of conditional probability and the law of total probability.