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

First Draft: May 2005
Current Draft: March 2007

Abstract

Despite mounting 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. Based on empirical evidence we propose a parsimonious model that is a mixture of (two) Markov chains, where the mixing is on the speed of movement among credit ratings. 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. We also find that conditioning on the state of the business cycle or industry group does not remove the heterogeneity with respect to the rate of movement. We go on to compare the performance of mixture and Markov chain using out-of sample predictions.

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

---

*We would like to thank Sam Hanson, Benjamin Iverson, Petra Miskov and Mihaela Popescu for their excellent research assistance. Halina Frydman also thanks the Department of Biostatistics at the University of Copenhagen for its hospitality.

†hfrydman@stern.nyu.edu 212-998-0453

‡til.schuermann@ny.frb.org 212-720-5968

Any views expressed represent those of the authors only and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

In the study of credit rating dynamics of firms, it is very convenient to assume that the ratings process is time homogeneous Markov. The credit migration or transition matrix, which characterizes past changes in credit quality of these firms, is then all that is needed to generate forecasts of the credit asset portfolio distribution in the future. Moreover, in the continuous time homogeneous Markov framework the objective is to estimate a generator matrix which is used to compute the credit transition matrix, allowing for forecasts over any time horizon.

Against this convenience is mounting evidence of non-Markovian behavior of the rating process. Altman and Kao (1992), Carty and Fons (1993), Altman (1998), Nickell, Perraudin and Varotto (2000), Bangia et al. (2002), Lando and Skødeberg (2002), Hamilton and Cantor (2004) and others have shown the presence of non-Markovian behavior such as ratings drift and industry heterogeneity, and time variation due in particular to the business cycle. The literature is only recently beginning to propose modeling alternatives to address these departures from the Markov assumption. For example, Christensen, Hansen and Lando (2004) consider the possibility of latent "excited" states for certain downgrades in an effort to address serial correlation of ratings changes (or ratings drift). Giamperi, Davis, and Crowder (2005) use a hidden Markov model to back out the state of the economy from ratings dynamics, although their model focuses just on default prediction, and Stefanescu, Tunaru, and Turnbull (2006) consider a simulation-based Bayesian approach which allows for some ratings momentum.

This paper considers a different type of non-Markov behavior than discussed in the extant literature and proposes a non-Markov model which is a generalization of a continuous time homogeneous Markov chain. A continuous time homogeneous Markov chain has the property that durations in states (spells) follow an exponential distribution. An exponential distribution has a constant hazard function which implies that the plot of the integrated hazard function against time is a straight line. We show that this property is violated for credit rating dynamics, albeit differently for different ratings. Firms with the same rating migrate at different speeds, a feature which is not admitted in the Markov model.

Motivated by these empirical findings we propose a new model for the ratings migration process, namely a mixture of two independent continuous time homogeneous Markov chains. The two Markov chains differ only in their implied migration speed, specifically in the rates with which they exit the states; they have the same embedded transition probability matrix.[1] The model assumes that there are two subpopulations of firms, each moving according to its own Markov chain. It implies that a duration in a state is generated by one or the other of the two exponential

[1] Given that the jump from state x occurs, the probability that it is to state y is the same for both chains. The proposed model captures observed heterogeneity in the rate of movement, but is not intended to capture ratings drift per se.

distributions so that the observed durations in a given state come from a mixture of two exponential distributions.

This is a nontrivial generalization of the basic Markov migration model: instead of forcing all firms of a given rating to migrate at the same speed regardless of their rating experience to date, the mixture process allows for rich and nuanced migration behavior across firms which all have the same rating today (but arrived at that rating in different ways).

We estimate the mixture model with corporate credit rating histories from Standard & Poor's spanning 1981 to 2002 using an algorithm developed in Frydman (2005). We show that the mixture model not only statistically dominates the simple Markov model but that the differences between two models can be economically meaningful; for the $\mathcal{CCC}$ rating, pricing differences of the mixture model can range from 30% to 57% relative to the value implied by the Markov model. 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.

This paper's contribution is twofold, one methodological and the other empirical. Our methodological contribution to the approach developed in Frydman (2005) consists of deriving explicit expressions for a firm's future ratings distribution conditional on past information. We show how this distribution differs depending on the available information which has particular relevance for practitioners. One may know the entire history, only a subset (say the last five years), or just the current rating.

We find that, under our model, two firms with identical current credit ratings can have substantially different future distributions of ratings. We show that firm 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.

Despite this predicted variation, it remains an open question whether the simplicity of the Markov model results in markedly worse out-of-sample performance. With this in mind we conduct out-of-sample forecast evaluations of the Markov against the mixture model. Although the literature has recently proposed and developed some alternatives to the standard Markov model, to our knowledge we are the first to conduct systematic out-of-sample forecast evaluations. We find that the one year out-of-sample average error rate is about 4% lower for the mixture than for the Markov model.[2] Our findings of the Markov model's robust performance is consistent with recent work by Kiefer and Larson (2006) who find that for typical forecast horizons, say one or two years, credit rating dynamics can be adequately modeled as a Markov chain, albeit based only on in-sample analysis. Conditioning on industry or state of the business cycle does not alter the basic results.

[2] Stefanescu, Tunaru, and Turnbull (2006) conduct an out-of-sample exercise focusing only on transitions to default rather than all transitions.

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 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 2005).

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 data set, discuss the estimation results, and compare a mixture model with a simple Markov model empirically, in and out-of-sample, and in terms of economic implications. Section 4 provides some concluding remarks.

2 Markov Mixture Modeling

Firms may take on one of 17 credit ratings, including +/- modifiers, as well as the default ($\mathcal{D}$) rating, and withdrawn or not rated ($\mathcal{NR}$) state. A rating may be withdrawn for innocuous reasons such as debt retirement, or more interesting (but unobservable) reasons such as deteriorating firm quality. Sample size considerations lead us to group firms into the 7 ratings which exclude the +/- modifiers for a total of 9 states which may be visited. They are, from best credit quality to worst: $\mathcal{AAA}$, $\mathcal{AA}$, $\mathcal{A}$, $\mathcal{BBB}$, $\mathcal{BB}$, $\mathcal{B}$, and $\mathcal{CCC}$. The available migration data among 7 ratings and the $\mathcal{NR}$ state are essentially continuous; the changes in ratings are recorded up to a calendar day. Further details of the data set as well as discussion of the $\mathcal{NR}$ state are given below in Section 3

A continuous time-homogeneous Markov chain has the property that durations in states (spells) have an exponential distribution. A well known property of an exponential distribution is that it has a constant hazard function which implies that the plot of the integrated hazard function against time is a straight line. In Figure 1 we display Nelson-Aalen cumulative hazard plots for each rating, along with their 95% confidence interval. The straight line is the implied cumulative hazard for an exponential distribution, i.e. under the null of a Markov chain. It is clear that the $\mathcal{CCC}$ rating is most inconsistent with the constant hazard rate assumption. The Nelson-Aalen plots also show that the exponential assumption may be violated for the $\mathcal{A}$ rating.

Motivated by the empirical findings we propose a mixture of two Markov chains as a more appropriate model for bond ratings migration dynamics first considered in Frydman (2005).[3] Because of data availability and the by now well known advantages of using the continuous time Markov approach over the discrete one (Lando and Skødeberg 2002), we define and estimate a continuous time version of the mixture model. The proposed mixture captures population heterogeneity in the

[3] See Norris (1997) for an overview of Markov chains.

rate of movement among states and is briefly summarized in Section 2.1 below.

It is useful to bear in mind that the migration matrix, however derived or estimated, describes through its diagonal elements the migration speed of the firms. The stylized fact that these matrices tend to be diagonally dominant means that most of the time there is no migration at all. The key distinguishing feature of the mixture process considered here is the presence of two underlying Markov processes which differ in their implied migration speed, one slow and one fast. While the generator of each Markov chain is fixed, a given firm's predicted migration may change over its (rated) life as a result of the change in the probability that a firm is a slow (or equivalently a fast) regime firm. The change in probability is driven, in turn, by the rating history of a firm.

2.1 The Mixture Process

More formally, the mixture process is a continuous time stochastic process $X = \{X(t), t > 0\}$ with state space $\mathcal{R} = \{1, 2, ..., w\}$, representing the different credit ratings, which, conditional on the initial state, is a mixture of two continuous Markov chains, $X_Q = \{X_Q(t), t > 0\}$ and $X_G = \{X_G(t), t > 0\}$, with generators G and Q, respectively. These generators are related by

$$G = \Gamma Q, \quad \Gamma = diag(\gamma_1, \gamma_2, ..., \gamma_w). \tag{1}$$

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,

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

that is, s_i is the proportion of firms with initial rating i that evolve according to X_G, and $1 - s_i$ is the proportion evolving according to X_Q. As the firm ages, its likelihood of being driven by the G or Q regime changes depending on its rating experience or history. We describe this evolution in more detail in Section 2.2 below.

Generator Q is a matrix with entries q_{ij} satisfying

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

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, $i \ne j \in \mathcal{R}$. Thus, it follows from (1) that X_Q and X_G in general differ in the rates at which they leave the states (i.e., $q_i \ne g_i$), but both chains, when leaving state i, have the same probability distribution for entering state j, given by $q_{ij}/q_i = g_{ij}/g_i, j \ne i$. Thus, depending on whether $\gamma_i = 0, 0 < \gamma_i < 1$, $\gamma_i = 1$, or $\gamma_i > 1$, the realizations of X_G never move out of state i, move out of state i at a lower, the same rate or at a higher rate, respectively, than those generated by X_Q. If $\gamma_i = 0, 1 \leq i \leq w$, 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 $\gamma_i = 1, 1 \leq i \leq w$, the mixture process collapses to a simple Markov chain. In our application below, G is the slow regime while Q is the fast one.

The EM algorithm for the maximum likelihood estimation of mixture parameters from a set of continuously observed rating histories is presented in Frydman (2005). That algorithm estimates the set of 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

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

where N_{ij} is the total number of $i \to j$ 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 G is obtained from the relationship in (1).

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

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

where $S = diag(s_1, s_2, ..., s_w)$. The MLE of transition matrices can be evaluated by plugging in the MLEs of G, Q and S.

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

2.2 Prediction

We now discuss probabilistic implications of the Markov mixture which were not treated in Frydman (2005). We 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 in special cases. Namely we consider prediction under full information on the rating history and two cases with limited or partial information: only the current rating is known and both the current and the initial ratings are known. Whether the forecasts from the more nuanced mixture process beat the Markov model in an out-of-sample contest is discussed in Section 3.4.

2.2.1 Full Information

Suppose that at time t the process is in state i, and we want to compute the conditional distribution of X, say, one period from now:

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

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 available information, contained in $\mathcal{I}_{t-}$, does matter for assessing future behavior of a process currently in state i. Set $\mathcal{I}_{i,t} = \mathcal{I}_{t-} \cup \{X(t) = i\}$, namely all previous and current information about X. The probability of being in state j in period $t+1$, conditional on all information up to period t (this includes the rating in period t, denoted here by i, as well as the rating history or path) is given by (see Appendix A for details)

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

where p_{ij}^G and p_{ij}^Q are entries in $P_G(1), P_Q(1)$, respectively. We refer to (4) as the basic forecasting equation. Thus, the distribution of a given state one period later is a convex combinaton of one period transition probabilities of Markov chains. The weight $P(X_G \mid \mathcal{I}_{i,t})$ is the probability of X evolving according to X_G based on the information in $\mathcal{I}_{i,t}$. 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 other 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, $P(X_G \mid \mathcal{I}_{i,t})$, is given by

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

where i_0 is the initial state, and

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

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

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

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

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

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.

2.2.2 Limited Information: Current Rating

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

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

and is equal to (see Appendix A),

$$
s_i(t) = \frac{\sum_{j=1}^{w} s_j p_{ji}^G(t) \pi_j}{\sum_{j=1}^{w} \pi_j \left[s_j p_{ji}^G(t) + (1 - s_x) p_{ji}^Q(t) \right]}, \quad t \geq 0, i, j \in \mathcal{R}, \tag{7}
$$

where $p_{ji}^G(t)$ $\left(p_{ji}^Q(t)\right)$ is the (j, i) entry in the $P_G(t), (P_Q(t))$, and $\pi_j = P\left(X(0) = j\right)$ 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_G, so that $s_i(0) = s_i$. The corresponding one period transition probability in this example is

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

It is clear from (8) that the migration probability depends only on the current rating i (and, of course, on the age of the firm, t). 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

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

where, as before, $P_G \equiv \exp(G)$ 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 of a mixture process changes over time measured on the age scale of a realization.

2.2.3 Limited Information: Initial and Current Rating

In our final case 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

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

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

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[4]).

Our data set of S&P ratings histories, CreditPro V. 7.0, is similar to the one used in Jafry and Schuermann (2004) and is described in some detail in Bangia et al. (2002). The data set contains rating histories from January 1, 1981 to December 31, 2005 of mainly large corporate institutions around the world. Ratings for sovereigns and municipals are not included, leaving 11,338 unique obligors. The share of the most dominant region in the data set, North America, has steadily decreased from 98% to 63%, as a result of increased coverage of companies domiciled outside U.S. For our analysis we restrict ourselves to U.S. obligors only to control for an important source of heterogeneity, namely country (see, for instance, Nickell, Perraudin and Varotto 2000); there are 7,119 unique U.S. domiciled obligors in the sample yielding 79,051 firm years of data, including withdrawn ratings, and 1,024 defaults of which 820 were from a rating, for an average default rate of 1.30%. If we exclude the withdrawn ratings the average default rate is 1.50%.

Transitions are infrequent: 20% of the firms had no transitions at all in the sample from 1981 - 2002. Another 36% experienced only one transition, and only 10% of firms had more than three transitions. Since transition intensities, and therefore transition probabilities, are estimated using transition counts, off-diagonal entries will likely be estimated with low precision given so little variation in the data. This poses a challenge to more complex models, as well as to any attempts at conditioning on external factors such as industry or the state of the business cycle.

3.1 Estimating and comparing the models

We begin by estimating 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

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

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 $\mathcal{NR}$. For what follows we restrict our sample through year-end 2002, where the number of unique U.S. domiciled obligors is 6,455; the remaining three years will be used for out-of-sample forecast evaluation in Section 3.4. The last column in Table 1 (we return to the other columns later) contains the number of firm years by rating. Thus, for example, the $\mathcal{AA}$ state was visited for a total of about 6,306 firm-years and the $\mathcal{NR}$ state was visited most frequently: 24,430 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 to 1996, claims that only 1% of all rating withdrawals may have been due to deteriorating credit quality.

Table 1 summarizes the parameter estimates for the final mixture model and reports the initial distribution of firms by rating. The second column contains the initial mixing proportions for the model. These proportions will change over time according to (7) on the age (not calendar time) scale. All firms that have initial rating $\mathcal{AAA}$ or $\mathcal{CCC}$ are driven entirely by regime G 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 G. It turns out that the G regime is slower or stickier than the Q regime for most ratings. This can also be seen by looking at the third column denoted γ_i. Recall that if $0 < \gamma_i < 1$ then regime G is slower than Q, but if $\gamma_i > 1$ then G is faster than Q.

We conduct a number of likelihood ratio tests to test various versions of the mixture against the Markov model. 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. 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. We first test the Markov chain (the null hypothesis specifies all gammas to be one) against the mixture model. This test overwhelmingly rejects the Markov chain in favor of the mixture model ($-2 \log LR = 353.3$; $p < 0.001$). These results are consistent with Frydman (2005) who used a much smaller sample of Moody's ratings over a shorter sample window (1985 to 1995).

The strong rejection of the Markov chain is due to very different behavior of regime G and Q firms in the ratings $\mathcal{AAA}$, $\mathcal{AA}$, $\mathcal{A}$ and $\mathcal{CCC}$: $\gamma_{\mathcal{AAA}} = 0.147, \gamma_{\mathcal{AA}} = 0.688, \gamma_{\mathcal{A}} = 0.620$, and $\gamma_{\mathcal{CCC}} = 0.216$. 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 $\mathcal{NR}$ ($\gamma_{\mathcal{NR}} = 0.088$). This may be interpreted as reflecting different reasons for rating withdrawal. In a Markov chain generated by G, $\mathcal{NR}$ 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. Thus the $\mathcal{NR}$ state is more convincingly uninformative for the G than for the Q regime.

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[5]) migration matrices, and in Tables 2c and 2d we show the one-year migration matrices for the two regimes, $P_G \equiv \exp(G)$ and $P_Q \equiv \exp(Q)$.[6] The diagonal entries, denoting no migration, are bolded for clarity. 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 39.54% for the Markov model and 52.73% for the mixture. The G regime dominates the Q in the mixture with an estimated 74.6% of firms moving according to regime G.[7]

The differences are more marked when comparing the migration matrices of the two regimes, P_G and P_Q shown in Tables 2c and 2d, especially for those ratings where the value of γ_i is most different from one, namely for $\mathcal{AAA}$ and $\mathcal{CCC}$. For example, the one-year probability of default (PD) for $\mathcal{CCC}$-rated firms is 32.50% under the G regime but 65.56% under the Q regime. For this rating the G regime is the slowest of all ratings (same the $\mathcal{NR}$ state) at $\gamma_{\mathcal{CCC}} = 0.216$, suggesting that firms which do not start in this state but are downgraded to it, i.e. the so-called "fallen angels," do not stay long; two-thirds move quickly to the default state. This helps also to explain the large differences in one-year PDs between the cohort and duration (Markov) methods for the $\mathcal{CCC}$ rating, documented by Lando and Skødeberg (2002) and Jafry and Schuermann (2004). The cohort method ignores the within year migrations, resulting in PD estimates which are too small as many $\mathcal{CCC}$-rated firms spend little time in that state before defaulting.

At the other end of the credit spectrum, the default likelihood for an $\mathcal{AAA}$-firm is more than 60 times higher in the Q (0.387%) than in the G regime (0.006%) and is, in fact, one and a half times as large as the default probability for a $\mathcal{BBB}$-rated firm, at 0.258%, in the Q regime. The likelihood of exiting the $\mathcal{NR}$ state for a firm governed by the G regime is much smaller $(1 - 0.9906 = 0.94\%)$ than for the Q regime $(1-0.8992 = 10.08\%)$, so over time the fraction of $\mathcal{NR}$ governed by G should increase. Indeed it does: about 75% in year 1 to over 85% by year 10, as can be computed using (7). However, many other cells are very similar or nearly identical. For example, $p_{\mathcal{BBB}\to\mathcal{AAA}} = 0.03\%$ for

[5] Since no firm starts in category $\mathcal{NR}$, we do not estimate the initial proportion s for $\mathcal{NR}$ (see Table 2). 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 $t = 1$). However, it turns out that the mixture transition matrices change little with respect to age.

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

[7] The estimated overall proportion of histories (0.746) moving according to regime G was computed by weighing the estimated mixing proportions by the initial distribution (see Table 1).

both the G 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 $\tilde{P}$:

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

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 t^{th} year, with P_A and P_Q the two regime migration matrices. Then $M_{SVD}(P_M) = 0.210$, $M_{SVD}(P(2)) = 0.192$, $M_{SVD}(P_G) = 0.191$ and $M_{SVD}(P_Q) = 0.329$, so that $M_{SVD}(P_Q) - M_{SVD}(P_G) = 0.138$. As a basis for comparison, Jafry and Schuermann (2004) report that the average M_{SVD} difference between Markov and cohort estimates of migration matrices is 0.0149. This makes quite clear the dramatic difference of the Q regime from the G regime or the mixture which is, of course, dominated by the G regime.

3.2 Firm-specific migration vectors

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 full-information weights in (5). The variation and range of those migration probabilities is rather substantial and is summarized in Table 3. The range from min to max in particular can be dramatic for the important default state. For example, the one-year ahead PD for $\mathcal{A}$-rated firms ranges from 0.020% to 0.130%. 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 4.160% to 6.230%. This is non-trivial since more firms were "born" with this rating than with any other rating (column 7 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).[8] For the $\mathcal{CCC}$-rated firms the range is 32.50% to 60.90%!

The minima and maxima should broadly correspond to one of the underlying Markov mixture matrices, i.e. P_G and P_Q, shown in Tables 2c and 2d respectively. For example, the maximum staying probability for $\mathcal{AAA}$ is 89.40% which corresponds to the G-regime staying probability for that rating; similarly for the $\mathcal{AA}$ rating. However, the maximum upgrade to $\mathcal{A}$ probability computed for $\mathcal{BBB}$-rated firms, $p_{\mathcal{BBB}\to\mathcal{A}}$ at 4.14%, corresponds to the Q-regime. Thus one regime does not necessarily correspond to either the mimimum or maximum migration probability. Perhaps the best illustration of this regime mixing is the $\mathcal{B}$-rating (Table 3). The maxima for the upgrade probabilities $p_{\mathcal{B}\to\mathcal{AAA}}$ through $p_{\mathcal{B}\to\mathcal{BBB}}$ correspond to the Q-regime, while the maximum probabilities $p_{\mathcal{B}\to\mathcal{BB}}$, $p_{\mathcal{B}\to\mathcal{CCC}}$ and $p_{\mathcal{B}\to\mathcal{NR}}$ correspond to the G-regime. The reverse is broadly 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.

3.3 Industry and business cycle effects

There is ample evidence of industry and business cycle effects (Nickell, Perraudin and Varotto 2000, Bangia et al. 2002), and so we go on to condition on both dimensions and estimate industry and recession/expansion specific models. Specifically we estimate both mixture and Markov models for each of six industries and for recessions and expansions, where the latter uses the NBER business cycle dates.[9] We use the M_{SVD} metric to compare the conditional mixture and Markov matrices. The results are summarized in Table 4.

First, the top row repeats information previously presented: the difference in M_{SVD} between the unconditional mixture and Markov matrices is about 8.6%; it is negative since the mixture has a smaller M_{SVD}. The biggest difference is for the Durable Manufacturing sector where the

[8] 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.

[9] Originally we had classified firms into seven industries, but the EM algorithm for the mixture model failed to converge for the Services sector due to a total lack of firms with an initial rating of $\mathcal{AAA}$.

For business cycle dating, see http://www.nber.com/cycles/cyclesmain.html.

mixture M_{SVD} is 18.3% smaller, followed by Wholesale and Retail Trade at 15.6% smaller. Only for Agriculture, Mining and Construction is the M_{SVD} larger.

Similarly for comparisons of matrices estimated just for recessions and expansions, these difference are modest for expansions (5.8%) and more pronounced for recessions (18.2%). So while the conditional matrices using the Markov vs. the mixture approach are different from each other, this difference is not very large.

3.4 Out-of-sample forecasting

Although the literature on credit migration has started to propose alternatives to the simple Markov model, those alternatives have not been subjected to rigorous out-of-sample forecast evaluation. A full horse race of several alternatives is beyond the scope of this paper, but in this section we do compare the performance of the mixture with the Markov model. To that end we make use of the full sample, i.e. through the end of 2005, to conduct three one-year out-of-sample forecasts, starting with 2003 (using data through the end of 2002). We also look at a single three-year forecast, namely 2005 using data through the end of 2002.

We evaluate forecast performance by subtracting the forecast probability of the rating which the firm assumes in period $T+1$ from one. To take an example, suppose a firm is in state (rating) $i = \mathcal{BBB}$ at the end of the sample period T (2002). Under the Markov model the prediction vector can be taken directly from Table 2a. Suppose that in the following year, $T+1$ (2003), the firm is still rated $\mathcal{BBB}$, and the Markov model predicted this would occur with probability 0.8372, so that the forecast error is small, namely $1 - 0.8372 = 0.1628$. Mixture predictions are generated using all the information available through the end of the sample period with equation (4) and will differ depending on the firm's rating history.

Under a Markov mixture process, there is an alternative way of generating migration predictions. With each estimated mixture model is associated an estimated proportion of firms that move according to the fast regime (and slow regime). For instance, we know from Section 3.1 that in our intitial sample ending in 2002, 74.6% of firms move according to regime G, the slow regime, and therefore 25.4% move according to the fast regime, Q. If we knew that a given firm is a fast regime firm at year-end 2002, we would naturally use the fast regime transition matrix, $P_Q \equiv \exp(Q)$, to predict next year's ratings, and similarly for a slow regime firm. We clearly don't know precisely which firms are fast regime firms and which are slow regime firms, but we do know the the probability that a firm belongs to the G regime at time t, which is given by the full information weight $P(X_G \mid \mathcal{I}_{i,t})$ in (5). We thus sort on this weight and assign the bottom $N_Q(t)$ firms to the Q regime, where, for example, $N_Q(2002) = 0.254 * 6,455 = 1639$ firms.

The results are summarized in Table 5a. We report the average one-year out-of-sample forecast errors, in percentage points. For example, this average forecast error is 15.76% for the Markov model for 2003 (using data through the end of 2002). For the mixture model using the equation (4), i.e. a weighted average of the G and Q regime, that average forecast error is a bit lower, namely 15.69%. But when we use the classification rule with a cut-off value, that error drops to 15.12%. The pattern is similar for the other two years: the forecast gains are modest using the weighting but more substantial when using the classification rule. The three-year exercise, i.e. forecasting 2005 from 2002, also yields similar results (omitted from the table). The cut-off method works well because it uses explicitly the fact that there are two regimes and converts the problem to a classification one. The forecast based on the weighted transition matrix of a mixture model cannot be as powerful because it uses the aggregate transition matrix, which, as we learned in Section 3.1 via the M_{SVD} metric, is not so different from Markov matrix.

In Section 3.3 we presented estimates of industry and business cycle conditional models and found the differences to be modest. One would naturally expect the differences in out-of-sample forecast performance to be similarly modest. We test this hypothesis by producing one-year (2003) out-of-sample forecasts by industry and compare the conditional Markov and mixture models to each other as well as to the simplest model of all, the overall or unconditional Markov matrix. The results are summarized in Table 5b. For the mixture model we present average forecast errors for both the weighting and the cut-off method.

Two things stand out in the table. First, consistent with the previous forecast discussion, the mixture model using the cut-off approach has the best overall performance. It does better in five of six industries, and does nearly equally well for the sixth (Communication, Electric & Gas, average forecast error 0.6% worse). Second, it is tough to beat the simplest model of all, the overall or unconditional Markov matrix as can be seen in the last column of Table 5b where we compare the average error of mixture cut-off to the overall Markov matrix. While the industry-specific mixture model using cut-offs nearly always does better, at best the forecast improvement is 8.1% (for the Finance, Insurance & Real Estate sector).

To make use of the business cycle regime model specifications involves a conditional forecast: first a forecast of which regime the economy will be in, and conditional on this prediction the actual migration forecast. In Table 5c we show the one-year out-of-sample average error assuming that 2003 would have been either an expansion or a recession. Since the U.S. economy was in an expansion in 2003, we would expect the expansion-specific models to do better. That is indeed the case. For instance, the average error using the expansion Markov matrix was 15.76% while it was 16.06% using the recession matrix. Note that the expansion error is the same as using the unconditional Markov matrix; see Table 5a. When comparing Markov and mixture results we see the same pattern as with the unconditional specifications: mixture does better using the cut-off approach and only marginally better using the weighting approach. It turns out that the proportion of firms moving according to the faster Q regime was 0.231 for those alive during expansions and 0.352 for those alive during recessions.

3.5 Financial impact of mixture models

It is one thing to show that the mixture model dominates the Markov model statistically; 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 6, which turns out to be the same as the max/min default probability for a given rating. We also present the price under the G or Q regime which would be used under the cut-off decision rule. 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 6. The first column gives the price under the Markov model, columns two and four show the price under the mixture model, columns three and five show the percentage differences to the Markov model, columns six and eight show the pricing under the G and Q regime respectively (we would use those regimes under the cut-off rule), while columns seven and nine show their respective percentage difference to the Markov model. The last column contains the coupon payments.

Overall the most significant differences to the Markov model come about when pricing is done using the much faster Q regime. What stands out in particular is the difference in pricing for the $\mathcal{CCC}$ rating where the pricing under the Q regime is 28.22% lower than implied by the Markov model. Even under the mixture model using the weighting the differences are stark: they range from a 24.43% discount to a 5.95% premium. The difference is noticably less for the other rating categories, though it still exceeds one percentage point (discount) for the $\mathcal{B}$ category. To be sure, this is merely illustrative, and different term structures of credit spreads would generate different valuations, 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. The mixture model we estimate contains two regimes, one slow, one fast. These regimes are not directly observable but, for a given firm conditional on its rating history, we are able to compute the likelihood of the firm belonging to a given regime. With that likelihood we can then classify the firm to an appropriate regime.

The mixture model proposed here statistically dominates the Markov chain and is able to capture some of the empirical regularities exhibited by the data, such as a decreasing hazard rate of time in a rating. We showed that an out-of-sample one year prediction error rate is about 4% lower for the mixture compared to the Markov model. Indeed we find that two firms with identical current credit ratings can have substantially different future ratings distributions.

We also found that conditioning on the state of the business cycle or industry group does not remove the heterogeneity with respect to the rate of movement. Fitting the mixture model to these subgroups leads to qualitatively similar conclusions. Those factors do not seem to be at play here, but rather there is a more subtle ratings history dependence. In this paper we have proposed a particular modeling approach which captures this dependence.

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 (2005). "International Convergence of Capital Measurement and Capital Standards: A Revised Framework." Basel Committee on Banking Supervision, November; available at <http://www.bis.org/publ/bcbs118.htm>.

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

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

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

[9] Christensen, Jens, Ernst Hansen, and David Lando (2004), "Confidence Sets for Continuous-Time Rating Transition Probabilities," *Journal of Banking & Finance* 28, 2575-2602.

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

[11] Frydman, Halina (2005), "Estimation in the Mixture of Markov Chains Moving with Different Speeds," *Journal of the American Statistical Association* 100, 1046-1053.

[12] 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.

[13] Giampieri, Giacomo, Mark Davis, and Martin Crowder (2005), A Hidden Markov Model of default interaction, *Quantitative Finance* 5, 27-34

[14] 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.

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

[16] Jafry, Yusuf and Til Schuermann (2004), "Measurement, Estimation and Comparison of Credit Migration Matrices," *Journal of Banking & Finance* 28, 2603-2639.

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

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

[19] Kalbfleish, John D and Ross L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, New York, NY: John Wiley & Sons.

[20] Kiefer, N.M. and C.E. Larson (2006), "A Simulation Estimator for Testing the Time Homogeneity of Credit Rating Transitions," forthcoming, *Journal of Empirical Finance*.

[21] Kijima, M., and K. Komoribayashi (1998), "A Markov Chain Model for Valuing Credit Risk Derivatives," *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 Mariarosa Verde (2004), "Rating Path Dependency," Fitch Ratings *Report on Structured Finance*, March 4, New York.

[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] Ross, S. M. (1996), *Stochastic Processes,* New York, NY: John Wiley&Sons.

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

[28] Stefanescu, C., R. Tunaru and S.M. Turnbull (2006), "The Credit Ratings Process and Estimation of Transition Probabilities: A Bayesian Approach," mimeo, London Business School.

A Appendix

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

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

$$P(X(\tau) = j \mid \mathcal{I}_{i,t}) = P(X_G \mid \mathcal{I}_{i,t}) p^G_{ij}(\tau - t) + P(X_Q \mid \mathcal{I}_{i,t}) p^Q_{ij}(\tau - t), i, j \in R. \tag{11}$$

Proof:

$$\begin{aligned}
P\left(X(\tau) = j \mid \mathcal{I}_{i,t}\right) &= P(X(\tau) = j, X_G \mid \mathcal{I}_{i,t}) + P(X(\tau) = j, X_Q \mid \mathcal{I}_{i,t}) \\
&= P(X(\tau) = j \mid X_G, \mathcal{I}_{i,t}) P(X_G \mid \mathcal{I}_{i,t}) \\
&\quad + P\left(X(\tau) = j \mid X_Q, \mathcal{I}_{i,t}\right) P\left(X_Q \mid \mathcal{I}_{i,t}\right) \\
&= P\left(X_G \mid \mathcal{I}_{i,t}\right) p^G_{ij}(\tau - t) + P\left(X_Q \mid \mathcal{I}_{i,t}\right) p^Q_{ij}(\tau - t),
\end{aligned}$$

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_G and X_Q.

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

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

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

which is (5).

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

$$\begin{aligned}
P\left(X_G \mid \mathcal{I}_{i,t}\right) &= \frac{\sum_{j=1}^{w} P(X_G, X(t) = i, X_0 = j)}{P(X(t) = i)} \\
&= \frac{\sum_{j=1}^{w} P(X(t) = i \mid X_G, X_0 = j) P(X_G \mid X_0 = j) P(X_0 = j)}{P(X(t) = i)} \\
&= s_i(t) = \frac{\sum_{j=1}^{w} s_j p^G_{ji}(t) \pi_j}{\sum_{j=1}^{w} \pi_j \left[s_j p^G_{ji}(t) + (1 - s_x) p^Q_{ji}(t) \right]}, \quad t \geq 0, i, j \in \mathcal{R}, ,
\end{aligned}$$

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

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	0.766	0.147	1.306	8.901	247	2,059
$\mathcal{AA}$	0.946	0.191	0.688	5.249	7.625	563	6,306
$\mathcal{A}$	0.924	0.215	0.620	4.659	7.520	1,202	12,414
$\mathcal{BBB}$	0.822	0.203	0.868	4.935	5.683	1,154	10,202
$\mathcal{BB}$	0.598	0.249	1.184	4.021	3.396	1,311	7,238
$\mathcal{B}$	0.582	0.278	1.086	3.603	3.318	1,872	8,262
$\mathcal{CCC}$	1.000	2.977	0.216	0.336	1.552	106	870
$\mathcal{NR}$	0.000	0.110	0.088	9.126	103.567	--	24,430

Table 1: Parameter estimates for the final mixture model: S&P U.S. corporate obligor histories, 1981-2002. The second columns s_i denotes the initial mixing proportion of firms with initial rating i that evolve according to regime G; q_i denotes the exit rate from rating i so that $1/q_i$ gives the expected duration (in years) in rating i under regime Q; γ_i denotes the relative speed of leaving state i so that if $0 < \gamma_i < 1$ then regime G is slower than Q. The last two columns contain some descriptive statistics of the sample by rating category.

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**88.97**	6.22	0.60	0.08	0.09	0.02	0.00	3.99	0.017
$\mathcal{AA}$	0.59	**87.53**	7.13	0.66	0.11	0.09	0.01	3.85	0.022
$\mathcal{A}$	0.08	1.84	**87.16**	5.51	0.52	0.20	0.01	4.64	0.029
$\mathcal{BBB}$	0.03	0.21	3.86	**83.72**	4.81	0.76	0.07	6.41	0.120
$\mathcal{BB}$	0.04	0.09	0.49	5.15	**76.22**	7.93	0.68	8.79	0.616
$\mathcal{B}$	0.00	0.07	0.25	0.53	4.20	**75.26**	5.05	10.09	4.555
$\mathcal{CCC}$	0.00	0.01	0.32	0.63	1.13	6.30	**39.54**	10.46	41.605
$\mathcal{NR}$	0.03	0.09	0.26	0.41	0.44	0.39	0.02	**97.52**	0.838

Table 2a: Markov migration matrix, in percentage points. Annual Markov credit migration matrix using S&P U.S. corporate obligor histories, 1981-2002. Initial rating is in the first column so that, for example, the probability of an $\mathcal{AA}$ -rated firm being downgraded to $\mathcal{A}$ over one year is 6.81%. Diagonal entries, the staying probabilities, are bolded for convenience.

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**89.40**	5.99	0.57	0.08	0.08	0.01	0.00	3.87	0.006
$\mathcal{AA}$	0.58	**87.53**	7.14	0.65	0.10	0.08	0.02	3.87	0.017
$\mathcal{A}$	0.08	1.82	**87.23**	5.49	0.51	0.20	0.02	4.63	0.025
$\mathcal{BBB}$	0.03	0.22	3.83	**83.78**	4.78	0.77	0.08	6.39	0.127
$\mathcal{BB}$	0.03	0.09	0.50	5.14	**76.31**	7.92	0.62	8.70	0.690
$\mathcal{B}$	0.00	0.08	0.26	0.55	4.23	**75.37**	4.40	10.07	5.034
$\mathcal{CCC}$	0.00	0.01	0.24	0.48	0.87	4.93	**52.73**	8.25	32.503
$\mathcal{NR}$	0.11	0.37	1.03	1.66	1.81	1.61	0.05	**89.92**	3.444

Table 2b: Markov mixture migration matrix (2nd Year), in percentage points. Annual Markov mixture credit migration matrix using S&P U.S. corporate obligor histories, 1981-2002. The 2nd year is used because we do not have an estimate of the initial proportions for category $\mathcal{NR}$; see Footnote 5 for details. However, we do have an estimate of this proportion, see Eq. (7), for any later time. Diagonal entries, the staying probabilities, are bolded for convenience.

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**89.40**	5.99	0.57	0.08	0.08	0.01	0.00	3.87	0.006
$\mathcal{AA}$	0.58	**87.79**	7.00	0.63	0.10	0.08	0.02	3.79	0.011
$\mathcal{A}$	0.08	1.75	**87.74**	5.27	0.48	0.19	0.02	4.46	0.015
$\mathcal{BBB}$	0.03	0.21	3.76	**84.14**	4.64	0.74	0.08	6.29	0.099
$\mathcal{BB}$	0.03	0.09	0.51	5.44	**74.89**	8.30	0.82	9.34	0.574
$\mathcal{B}$	0.00	0.07	0.24	0.53	4.25	**74.46**	5.92	10.35	4.160
$\mathcal{CCC}$	0.00	0.01	0.24	0.48	0.87	4.93	**52.73**	8.25	32.503
$\mathcal{NR}$	0.01	0.03	0.10	0.15	0.16	0.15	0.01	**99.06**	0.312

Credit Rating	$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	**46.63**	29.07	3.53	0.66	0.64	0.23	0.01	18.85	0.387
$\mathcal{AA}$	0.61	**82.97**	9.52	1.10	0.21	0.16	0.01	5.31	0.110
$\mathcal{A}$	0.09	2.68	**81.08**	8.12	0.85	0.34	0.01	6.67	0.143
$\mathcal{BBB}$	0.03	0.27	4.14	**82.08**	5.46	0.87	0.05	6.85	0.258
$\mathcal{BB}$	0.03	0.09	0.49	4.71	**78.41**	7.36	0.32	7.74	0.861
$\mathcal{B}$	0.01	0.08	0.28	0.58	4.21	**76.62**	2.29	9.68	6.250
$\mathcal{CCC}$	0.01	0.06	0.58	1.13	1.99	9.74	**5.35**	15.57	65.564
$\mathcal{NR}$	0.11	0.37	1.03	1.66	1.81	1.61	0.05	**89.92**	3.444

Tables 2c (top) and 2d (bottom): Migration matrix from mixture model for regime G (top) and regime Q (bottom), in percentage points. $P_G \equiv \exp(G)$, $P_Q \equiv \exp(Q)$. S&P U.S. corporate obligor histories, 1981-2002.

State at T		State at T+1								
		$\mathcal{AAA}$	$\mathcal{AA}$	$\mathcal{A}$	$\mathcal{BBB}$	$\mathcal{BB}$	$\mathcal{B}$	$\mathcal{CCC}$	$\mathcal{NR}$	$\mathcal{D}$
$\mathcal{AAA}$	$\hat{\mu}$	89.23	6.08	0.58	0.08	0.08	0.01	0	3.93	0.011
	$\hat{\sigma}$	0.89	0.48	0.06	0.01	0.01	0	0	0.31	0.008
	min	82.77	5.99	0.57	0.08	0.08	0.01	0	3.87	0.010
	max	89.40	9.57	1.03	0.17	0.17	0.04	0	6.19	0.070
$\mathcal{AA}$	$\hat{\mu}$	0.58	87.50	7.16	0.66	0.10	0.08	0.02	3.88	0.016
	$\hat{\sigma}$	0	0.57	0.30	0.05	0.01	0.01	0	0.18	0.012
	min	0.58	83.79	7.00	0.63	0.10	0.08	0.01	3.79	0.010
	max	0.6	87.79	9.10	1.02	0.19	0.14	0.02	5.05	0.090
$\mathcal{A}$	$\hat{\mu}$	0.08	1.83	87.17	5.51	0.51	0.20	0.02	4.65	0.026
	$\hat{\sigma}$	0	0.12	0.89	0.38	0.05	0.02	0	0.30	0.017
	min	0.08	1.75	81.84	5.27	0.48	0.19	0.01	4.46	0.020
	max	0.9	2.57	87.74	7.80	0.81	0.32	0.02	6.42	0.130
$\mathcal{BBB}$	$\hat{\mu}$	0.03	0.22	3.84	83.75	4.79	0.77	0.08	6.40	0.130
	$\hat{\sigma}$	0	0.01	0.07	0.35	0.14	0.02	0.01	0.10	0.028
	min	0.03	0.21	3.76	82.11	4.64	0.74	0.05	6.29	0.100
	max	0.03	0.27	4.14	84.14	5.45	0.87	0.08	6.84	0.260
$\mathcal{BB}$	$\hat{\mu}$	0.03	0.09	0.50	5.15	76.30	7.92	0.62	8.70	0.689
	$\hat{\sigma}$	0	0	0.01	0.13	0.62	0.17	0.09	0.28	0.051
	min	0.03	0.09	0.49	4.71	74.89	7.37	0.33	7.75	0.570
	max	0.03	0.09	0.51	5.44	78.38	8.30	0.82	9.34	0.860
$\mathcal{B}$	$\hat{\mu}$	0	0.08	0.26	0.56	4.23	75.39	4.36	10.06	5.059
	$\hat{\sigma}$	0	0	0.01	0.01	0.01	0.33	0.55	0.10	0.317
	min	0	0.07	0.24	0.53	4.21	74.46	2.33	9.69	4.160
	max	0.01	0.08	0.28	0.58	4.25	76.60	5.92	10.35	6.230
$\mathcal{CCC}$	$\hat{\mu}$	0	0.02	0.29	0.57	1.03	5.62	45.90	9.30	37.27
	$\hat{\sigma}$	0	0.01	0.05	0.11	0.18	0.78	7.65	1.18	5.34
	min	0	0.01	0.24	0.48	0.87	4.93	12.04	8.25	32.50
	max	0.01	0.05	0.53	1.04	1.83	9.06	52.73	14.54	60.90

Table 3: Descriptive statistics for firm-specific migration vectors using the final Markov mixture model at the end of the sample period; investment grades. Min and max refer to migration to default. S&P U.S. corporate obligor histories, 1981-2002.

Industry	**Mixture**	**Markov**	**% Δ (to Markov)**
Unconditional	0.192	0.210	-8.6%
Agriculture, Mining, Construction	0.206	0.194	6.2%
Durable Manufacturing	0.178	0.218	-18.3%
Non-Durable Manufacturing	0.186	0.203	-8.5%
Communication, Electric & Gas	0.205	0.219	-6.3%
Wholesale & Retail Trade	0.187	0.222	-15.6%
Finance, Insurance & Real Estate	0.193	0.211	-8.3%
Expansion	0.196	0.208	-5.8%
Recession	0.183	0.224	-18.2%

Table 4: M_{SVD} metrics of Markov and mixture matrices, unconditional (overall) and conditioning on industry and state of the business cycle.

	2003	**2004**	**2005**	**2005 (3-year)**
Markov	15.76	13.86	14.25	28.78
Mixture (FI: weighting)	15.69	13.81	14.18	28.64
Mixture (FI: cut-off)	15.12	13.38	13.80	27.47
Cut-off proportion	0.254	0.216	0.221	0.254

Table 5a: Average one-year out-of-sample forecast errors in percent; last column is three-year out-of-sample prediction. For a given prediction year, say 2003, all information through year-end of the prior year, say 2002, is used. "Markov" uses the simple Markov model to generate rating predictions; "Mixture (weighting)" is the Markov mixture model using the full information weights given by equation (4), while "Mixture (cut-off)" assigns a proportion of firms as belonging to the fast (Q) regime, where the firms are sorted on the full information weight.

Industry	% Average Forecast Error (2003) Markov: Industry Specific	Markov: Overall	Mixture: Weighting	Mixture: Cut-off	% Δ to Industry Markov: Overall Markov	% Δ to Industry Markov: Weighting	% Δ to Industry Markov: Cut-off	**% Δ cut-off to overall Markov**
Agriculture, Mining, Construction	15.90	16.00	15.82	14.96	0.6	-0.5	-6.0	-6.5
Durable Manufacturing	14.71	14.46	14.65	14.59	-1.7	-0.4	-0.8	0.9
Non-Durable Manufacturing	15.67	16.15	15.57	15.07	3.1	-0.6	-3.8	-6.7
Communication, Electric & Gas	19.64	20.17	19.76	19.76	2.7	0.6	0.6	-2.0
Wholesale & Retail Trade	15.77	15.36	15.71	15.12	-2.6	-0.3	-4.1	-1.6
Finance, Insurance & Real Estate	15.40	15.76	14.89	14.48	2.3	-3.3	-6.0	-8.1

Table 5b: Average one-year out-of-sample forecast errors in percent by industry for 2003. The percent differences are as follows, using the first row as an example: industry to overall (0.6 = (16.00 / 15.90) – 1); industry to weighting (-0.5 = (15.82 / 15.90) – 1); industry to cut-off (-6.0 = (14.96 / 15.90) – 1); cut-off to overall Markov (-6.5 = (14.96 / 16.00) – 1).

	Expansion	**Recession**
Markov	15.76	16.09
Mixture (FI: weighting)	15.71	15.89
Mixture (FI: cut-off)	15.17	15.88
Cut-off proportion	0.231	0.352

Table 5c: Average one-year out-of-sample forecast errors in percent by business cycle regime for 2003. Expansion (recession) errors are obtained using the expansion (recession) matrices to forecast 2003, thus assuming that 2003 will be an expansion (recession) year; it turned out actually to be an expansion year. The proportion of firms moving according to the faster Q regime was 0.231 for those alive during expansions and 0.352 for those alive during recessions.

	Markov (1)	V1 Mixture (2)	% Δ V1 to Markov (3)	V2 Mixture (4)	% Δ V2 to Markov (5)	*G* Regime (6)	% Δ *G* to Markov (7)	*Q* Regime (8)	% Δ *Q* to Markov (9)	Coupon (annual)
$\mathcal{AAA}$	101.973	101.962	-0.01%	101.904	-0.07%	101.967	-0.01%	101.531	-0.43%	2
$\mathcal{AA}$	102.634	102.629	-0.005%	102.552	-0.08%	102.628	-0.01%	102.532	-0.10%	2.5
$\mathcal{A}$	102.984	102.974	-0.01%	102.890	-0.09%	102.978	-0.01%	102.881	-0.10%	3
$\mathcal{BBB}$	104.5794	104.448	-0.13%	104.5789	-0.0004%	104.580	0.0002%	104.450	-0.12%	4
$\mathcal{BB}$	104.306	104.132	-0.17%	104.310	0.004%	104.309	0.003%	104.131	-0.17%	5
$\mathcal{B}$	103.058	101.737	-1.28%	103.1195	0.06%	103.120	0.06%	101.725	-1.29%	6.5
$\mathcal{CCC}$	79.179	59.838	-24.43%	83.890	5.95%	83.887	5.95%	56.837	-28.22%	8.5

Table 6: Bond pricing example. Example of pricing two bonds, V1 and V2, using Markov and mixture models, as well as based on the *G* and *Q* regime alone using the cut-off. We assume a 40% recovery rate and use 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. 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.

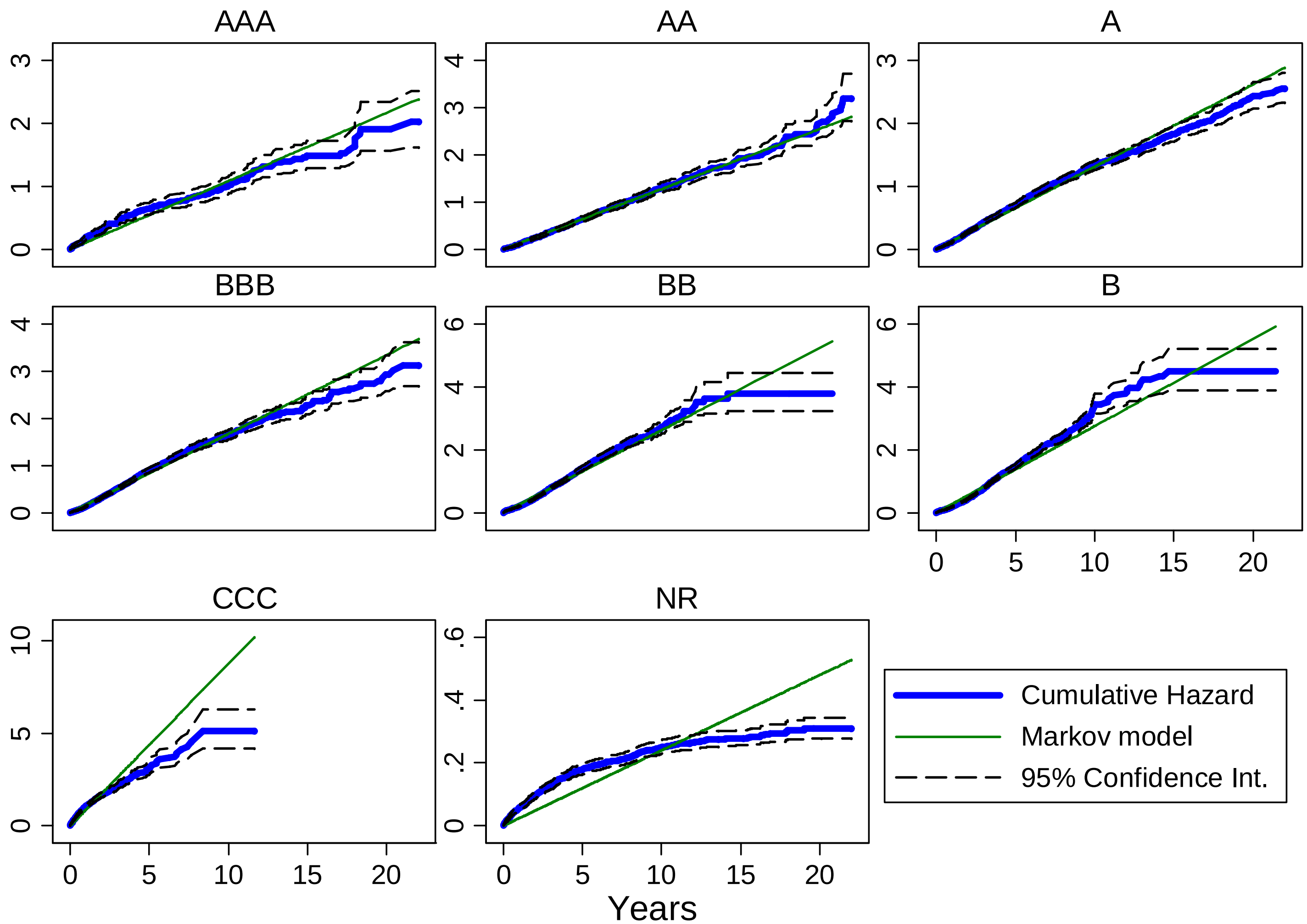

Figure 1: Nelson-Aalen plots for rating spells by rating. S&P rated U.S. firms, 1981-2002. Solid thick line denotes the empirical cumulative hazard, along with its 95% confidence interval (dashed lines). The thin solid line denotes the (constant) hazard as implied by the Markov model.