Estimation in the continuous time mover-stayer model with an application to bond ratings migration *

Halina Frydman**
Stern School of Business, New York University

Ashay Kadam***
University of Michigan Business School

December 19th, 2002

Abstract

The usual tool for modeling bond ratings migration is a discrete, time-homogeneuous Markov chain. Such model assumes that all bonds are homogeneous with respect to their movement behavior among rating categories and that the movement behavior does not change over time. However, among recognized sources of heterogeneity in ratings migration is age of a bond (time elapsed since issuance). It has been observed that young bonds have a lower propensity to change ratings, and thus to default, than more seasoned bonds.

The aim of this paper is to introduce a continuous, time-nonhomogeneuous model for bond ratings migration, which also incorporates a simple form of population heterogeneity. The specific form of heterogeneity postulated by the proposed model appears to be suitable for modeling the effect of age of a bond on its propensity to change ratings. This model, called a mover-stayer model, is an extension of a time-nonhomogeneuous Markov chain.

This paper derives the maximum likelihood estimators for the parameters of a continuous time mover-stayer model based on a sample of independent continuously monitored histories of the process, and develops the likelihood ratio test for discriminating between the Markov chain and the mover-stayer model. The methods are illustrated using a sample of rating histories of young corporate issuers. For this sample, the likelihood ratio test rejects a Markov chain in favor of a mover-stayer model. For young bonds with lowest rating the default probabilities predicted by the mover-stayer model are substantially lower than those predicted by the Markov chain.

Keywords: Ratings migration, mover-stayer model, Markov chain, estimation

JEL Classification: C13, G33

1   Introduction

The usual tool for modeling bond ratings migration is a time-homogenous, discrete Markov chain. Such modeling assumes that all bond issues are homogeneous with respect to their movement behavior among rating categories and also that their behavior does not change over time. However, there are many sources of heterogeneity in the rating behavior, of which the age of a

bond has been recognized as an important one. Evidence presented by Altman(1998), Asquith et al, (1989), and Keenan, Soberhardt, and Hamilton (1999) and references therein suggests that propensity of bonds to change rating, and in particular to default, is lower during the early years after issuance than it is for seasoned bonds. However, this aspect of aging effect has not yet been incorporated in any systematic way in modeling of the evolution of ratings.

In this paper we propose a continuous time mover-stayer model to capture the aging effect described above.[1] This model is an extension of a time-nonhomogeneuous, continuous Markov chain. It postulates a simple form of heterogeneity: a population of bonds is assumed to consist of two subpopulations, "movers" and "stayers". "Movers" evolve according to a continuous time Markov chain, whereas "stayers" stay in their initial states. The proportion of stayers in each rating state is a parameter of the model and can be interpreted as a measure of immobility for the bonds in a given rating state. If this parameter is zero in each rating, the mover-stayer model reduces to a Markov chain. The mover-stayer model, which allows for a greater degree of immobility of bonds than a Markov chain, may be a better description of ratings migration for younger bonds. This paper develops a methodology for estimation and testing of this model against the continuous Markov chain.

We consider a time-nonhomogeneuous mover-stayer model with time measured since the issuance of the bond. More precisely the age-nonhomogeneity is modeled by assuming that the parameter of the mover-stayer model is a piece-wise constant function of age with a constant value for each year of life of the bond. In addition, according to the mover-stayer model, in each year of life a bond may exhibit a stayer or a mover type behavior. Thus, the proposed model incorporates both time (age) nonhomogeneity, as well as simple heterogeneity in the movement behavior of bonds that are of similar age.

Throughout we consider a continuous time framework, rather than a discrete one, because rating agencies (Moody's, Standard and Poor's) monitor

---

[1]The discrete time mover-stayer model which was introduced by Blumen, Kogan and McCarthy (1955) has been since employed in many areas (e.g. Colombo and Morrison (1989), Sampson (1990), Chaterjee and Ramaswamy (1996), and Chen et al (1997)). Frydman, Kallberg and Kao (1985) applied the estimation methods for the discrete time mover-stayer model developed in Frydman (1984) to the analysis of credit behavior. Subsequently, Altman and Kao (1991) used this methodology in their study of ratings migration.

changes in bond ratings on daily basis which gives rise to very detailed data. For a given bond issue the complete history of its rating changes is available, which includes the exact dates of rating changes, the types of the changes that occur, and the lengths of stay in different rating states. Modeling such data using only information obtained at discrete time points would necessarily entail a loss of information. In addition, a continuous time framework affords a possibility of making predictions of quantities of interest, such as of future rating state probabilities, at any relevant time horizon, whereas the predictions with a discrete version of the model can be made only at times that are multiples of the sampling interval. Until recently a discrete time framework has been employed almost exclusively to model rating migrations. The first statistical analysis of ratings migration with a continuous time Markov chain has been in Lando and Skodeberg (2000) to which the reader is referred for a more extensive discussion of the advantages of a continuous time framework and related recent references.

To implement the continuous time mover-stayer model, we derive the maximum likelihood (ml) estimators of its parameters based on a sample of independent continuously observed realizations from this process. The ml estimation in the continuous time mover-stayer model from continuously observed realizations has not been considered before.[2,3] Based on the derived ml estimators we formulate the likelihood ratio test for discriminating between the Markov chain and the mover-stayer models. The main tool for obtaining the ml estimators is the EM algorithm (Dempster, Laird, and Rubin (1977)). However, we show that when realizations are observed over the same fixed time horizon, the ml estimators can be easily obtained by direct maximization of the likelihood function.

We illustrate our methods using the rating histories of the sample of 856 corporate bond issuers that were observed in the period from January, 1985 to December, 1995. On the basis of this sample, we estimate the continuous age-nonhomogeneous mover-stayer model and Markov chain. The age-nonhomogeneity is modeled by assuming that the parameter of the process

---

[2]Maximum likelihood estimation in the discrete time mover-stayer model was discussed in Frydman (1984), Fuchs and Greenhouse (1988), and Swensen (1996). Bayesian estimation in the continuous time mover-stayer model from panel data was considered in Fougere and Kamionka (2002).

[3]Maximum likelihood estimation in the mixtures of continuous time Markov chains that generalize the mover-stayer model is considered in Frydman (2002). Another generalization of the discrete time mover-stayer model is in Cook et al. (2001).

is a step function which is constant within each one-year age interval, that is, the models are estimated separately for bonds in their first year of life, the second year of life, etc, yielding age specific one-year transition probability matrices.[4] Because of data limitation we estimate the two models only up to the fifth year of life of the bond.

We briefly summarize our empirical results reported in Section 4. The likelihood ratio test rejects the Markov chain in favor of the mover-stayer model in each of the one-year age intervals. The overall expected proportion of stayers estimated by the mover-stayer model is large for very young bonds and then decreases as bonds become more seasoned. This is consistent with the aging effect. The interesting aspect of our results is the large estimated proportion of stayers in the C (combined Caa, Ca and C) rating for young bonds. This has a substantial implication for the estimation of the probability of default from rating C: for young bonds one-year probabilities of default from rating C estimated by the mover-stayer model are substantially smaller than those estimated by a Markov chain. The difference in the default probabilities resulting from the two models is largest for very young bonds and then decreases with age. For both models the probability of default is largest in the fifth year and much smaller for younger bonds. We do note that our results are based on a small sample of bonds and thus we treat the empirical analysis solely as an illustration of the methodology. An application of the methodology developed here to a much larger sample is required to evaluate the usefulness of the continuous-time mover stayer model in modeling the aging effect.

This paper is organized as follows. In Section 2 we define the mover-stayer model in continuous time and its time nonhomogeneuous version which we use to model aging effect. Section 3 develops the ml estimation in the mover-stayer model from continuous observations. In Section 4 we report and discuss the results of the estimation of the mover-stayer model and the Markov chain for ratings migration of young bond issuers.

---

[4]We note that the one-year transition matrix for a portfolio containing bonds of different ages can then be estimated by the weighted average of the age specific one-year transition matrices with the weights representing the proportions of bond issuers of different ages.

2 The mover-stayer model

To define a mover-stayer model in continuous time we first consider a Markov chain in continuous time with state space $W = (1, 2, ..., w)$. The states correspond to different rating categories. Such chain is characterized by the generator matrix Q, which is the matrix with the following structure

$$q_{ii} \leq 0, q_{ij} \geq 0, \sum_{j \neq i} q_{ij} = -q_{ii} \equiv q_i, i \in W.$$

In the context of our application to ratings migration the entries in Q have the following probabilistic interpretation: each time a bond enters rating i it stays in it for the time that is exponentially distributed with parameter $(-q_{ii})$. When it exits from rating i, it makes a transition to rating $j, j \neq i$ with probability $q_{ij}/(-q_{ii})$. In particular,

$$\frac{1}{-q_{ii}} = \text{expected length of time for an issuer in}$$

$$\text{rating } i \text{ to remain in that rating.}$$

Matrix Q is called a generator, because it generates $M(t)$, the matrix of transition probabilities $m_{ij}(t)$ of a continuous time Markov chain. $M(t)$ is obtained, for every time t, by exponentiation of tQ, that is,

$$M(t) = \exp(tQ), t \geq 0.$$

For the definition of the matrix exponential and an exposition of Markov chains in continuous time see, for example, Norris (1997).

A continuous time mover-stayer model on state space $W = \{1, 2, .., w\}$ is a mixture of two independent Markov chains, one which evolves according to some infinitesimal generator Q, and the other whose transition probability matrix is an identity matrix I. The transition probability matrix, $P(t)$, of a continuous time mover-stayer model on state space W is then defined as

$$P(t) = SI + (I - S) \exp(tQ), t \geq 0, \tag{1}$$

where $S = \text{diag}(s_1, s_2, ..., s_w)$, with

$$s_i = \text{proportion of stayers in state } i, i \in W.$$

The Markov chain defined above is time homogeneous because its generator is constant in time. Similarly the mover-stayer model defined above is time homogeneous because it involves a time homogeneous Markov chain, and proportions of stayers that do not change over time.

A time-nonhomogeneous Markov chain has a generator $Q(t)$ which is a function of time. A simple, but for our purpose very useful, time-nonhomogeneous Markov chain can be defined by assuming that its generator, $Q(t)$, is a piecewise constant function of time on some time interval $(0, T)$ that corresponds to the time of the study. The particular specification of $Q(t)$ of interest to us is

$$
\begin{aligned}
Q(t) &= Q(1), 0 \le t \le 1, \\
&= Q(2), 1 < t \le 2, \\
&\vdots \\
&= Q(m), m - 1 < t \le m = T
\end{aligned}
\tag{2}
$$

where $1, 2, ..., m - 1$ are the times where regime changes occur and $Q(k)$ is the generator in the $k'th$ time subinterval, $1 \le k \le m$. With the view towards our application, we assume that time is measured in years since the issuance of the bond so that the generator in (2) depends on the age of the bond issuer and the one-year transition probability matrix for bonds age $(k - 1)$ is given by

$$
M(k - 1, k) = \exp(Q(k)), 1 \le k \le m - 1.
$$

We define the time-nonhomogeneous mover-stayer model by assuming that a generator of a Markov chain which describes the evolution of movers is as in (2). Furthermore, we assume that a bond in state i at the beginning of the k'th one-year age interval, $(k - 1, k)$, has probability $s_i(k)$ of being a stayer in that interval independently of its behavior in the preceding age intervals. Thus, the k'th age interval has its own vector of proportions of stayers, denoted by $s(k) = (s_i(k), 1 \le i \le w)$. The k'th one-year age interval transition probability matrix is then given by

$$
P(k - 1, k) = S(k) + (I - S(k)) \exp(Q(k)), 1 \le k \le m - 1,
$$

where $S(k) = \text{diag}(s_1(k), s_2(k), ..., s_w(k))$. The transition matrices $M(k - 1, k)$ and $P(k - 1, k), 1 \le k \le m - 1$, are age specific one-year transition matrices for the Markov chain and mover-stayer model respectively.

7

We note that for both models we can easily compute the transition probability matrix between arbitrary times s, t such that $0 \leq s \leq t \leq m$. For example, setting $s = 0.5$ and $t = 2.5$, the transition probability matrix $M(s, t)$ of a Markov chain with generator in (2) is

$$M(0.5, 2.5) = \exp(0.5Q(1)) \exp(Q(2)) \exp(0.5Q(3)), \qquad (3)$$

and the transition matrix $P(s, t)$ of a mover-stayer model is

$$
\begin{aligned}
P(0.5, 2.5) \;=\; & [S(1) + (I - S(1)) \exp(0.5Q(1))] \\
& \times [S(2) + (I - S(2)) \exp(Q(2))] \qquad (4) \\
& \times [S(3) + (I - S(3)) \exp(0.5Q(3))] .
\end{aligned}
$$

The expression in (3) follows by Markov property and the expression in (4) by the definition of the mover-stayer model which implies that the model has Markov property at discrete time (age) points $0, 1, ..., m$, but behaves as the mover-stayer model within each age interval.

For both a Markov chain and a mover-stayer model the estimation of their parameters can be done separately in each one-year age interval. By assumption in each such time interval both models are time homogeneous. Thus, to estimate the generator in (2) for a Markov chain, we use the ml estimate of $Q(k)$ using the data on bonds with age (time since issuance) in the interval $(k-1, k)$. This is a well known ml estimator of the generator of a time homogeneous Markov chain and is presented in Section 3.1. Similarly we estimate $Q(k)$ and $S(k)$ in the mover-stayer model based on bonds with age in the interval $(k-1, k)$. The ml estimators of these parameters, that is, of the parameters of a time homogeneous mover-stayer model are derived below.

3   Maximum Likelihood Estimation

Let $X = (X_t, t \geq 0)$, be a mover-stayer model with transition probability function defined in (1). Assume that we observe n independent realizations of X and that the k'th realization, X^k, is observed continuously on some time interval $[0, T^k]$ with $T^k \leq T$, where T is the time horizon of the study. Thus, $X^k = (X_t, 0 \leq t \leq T^k)$ and individual realizations may be observed over time intervals of different lengths. This may be the case when right censoring is present or when the mixture process has an absorbing state.

8

The right censoring is assumed to be independent. Let $\mathcal{A}$ be the set of all realizations that stayed continuously in an initial state, and $\mathcal{B}$ be the set of all realizations with at least one transition. We note that $\mathcal{A}$ may contain movers as well as stayers.

Let L_k^Q be the likelihood of observing X^k when it is generated by a Markov chain with an intensity matrix Q. Then conditional on knowing an initial state (see e.g., Albert (1962)),

$$L_k^Q = \prod_{i \neq j} (q_{ij})^{n_{ij}^k} \prod_i \exp(-q_i \tau_i^k),$$

where

$$
\begin{aligned}
n_{ij}^k &= \text{the number of times } X^k \text{ makes an } i \to j \text{ transition, } i \neq j, \\
\tau_i^k &= \text{the total time } X^k \text{ spends in state } i.
\end{aligned}
$$

Thus, the likelihood of $X^k \in \mathcal{B}$ under the mover-stayer model, conditional on knowing an initial state is

$$\prod_{i=1}^{w} (1 - s_i)^{I_i^k} L_k^Q,$$

and the similar likelihood for $X^k \in \mathcal{A}$ is

$$L_k = \prod_{i=1}^{w} (s_i)^{I_i^k} + \prod_{i=1}^{w} (1 - s_i)^{I_i^k} L_k^Q,$$

where

$$
\begin{aligned}
I_r^k &= 1 \text{ if } X_0^k = r \\
&= 0, \text{ otherwise.}
\end{aligned}
$$

It is seen that the likelihood function of n independent realizations, $L \equiv \prod_{k=1}^{n} L_k$, is difficult to maximize directly. Instead we develop the EM algorithm for obtaining the mles of Q and s. To implement this algorithm we require $\hat{Q}^c$ and $\hat{s}^c$, the mles of the parameters based on complete information. The derivation of $\hat{Q}^c$ and $\hat{s}^c$, is straightforward, but we include it below for completeness and also to introduce the notation needed for the formulation of the EM algorithm. Let

$$
\begin{aligned}
Y^k &= 1, \text{ if the k'th realization is generated by a stayer,} \\
&\quad\, 0, \text{ otherwise.}
\end{aligned}
$$

Obviously, for any realization with nonzero number of transitions, $Y^k = 0$, and for any realization with zero transitions, Y^k is not observed. Assuming that all $Y^k s$ are observed, L^c_k, the likelihood function of the kth realization of the mover-stayer model is simply

$$L^c_k = \prod_{r=1}^{w} (s_r)^{I^k_r Y_k} (1 - s_r)^{I^k_r (1-Y_k)} \left(\prod_{i \neq j} (q_{ij})^{n^k_{ij}} \prod_i \exp(-q_i \tau^k_i) \right)^{1-Y_k},$$

or

$$\log L^c_k = \sum_{r=1}^{w} I^k_r \log(1 - s_r) + Y_k \sum_{r=1}^{w} I^k_r \log \left[s_r / (1 - s_r) \right]$$
$$+ (1 - Y_k) \sum_{i \neq j} n^k_{ij} \log(q_{ij}) - (1 - Y_k) \sum_i \left(q_i \tau^k_i \right),$$

which for all realizations becomes

$$\log L^c = \sum_{r=1}^{w} m_r \log(1 - s_r) + \sum_{r=1}^{w} m^S_r \log \left[s_r / (1 - s_r) \right]$$
$$+ \sum_{i \neq j} n_{ij} \log(q_{ij}) - \sum_i q_i \tau_i + \sum_i q_i \tau^S_i,$$

where

$$m_r = \sum_{r=1}^{w} I^k_r = \text{total number of individuals that begin in state } r,$$

$$m^S_r = \sum_{k=1}^{n} I^k_r Y_k = \text{number of stayers in state } r$$

$$n_{ij} = \sum_{k=1}^{n} n^k_{ij} = \text{total number of } i \to j \text{ transitions in the sample,}$$

$$n_i = \sum_{j \neq i}^{w} n_{ij} = \text{total number of transitions out of state } i$$

$$\tau_i = \sum_{k=1}^{n} \tau^k_i = \text{total time in state } i \text{ for all individuals in the sample,}$$

$$\tau^S_i = \sum_{k=1}^{n} Y_k \tau^k_i = \text{the total time in state } i \text{ for stayers, } \tau^M_i = \tau_i - \tau^S_i$$

10

Solving the score equation $\partial \log L_c / \partial s_i = 0$, gives the natural estimator $\widehat{s}_i^c = m_i^S / m_i$. Now setting $\partial \log L_c / \partial q_{ij} = 0$, we obtain

$$\hat{q}_{ij}^c = \frac{n_{ij}}{\tau_i - \tau_i^S} = \frac{n_{ij}}{\tau_i^M}, \tag{5}$$

and from $\sum_{j \neq i}^w q_{ij} = q_i$,

$$\hat{q}_i^c = \sum_{j \neq i}^w \frac{n_{ij}}{\tau_i^M} = \frac{n_i}{\tau_i^M}, \tag{6}$$

From (5) and (6) we also get

$$\hat{q}_{ij}^c = \frac{n_{ij}}{n_i} \hat{q}_i. \tag{7}$$

3.1 The EM algorithm

Based on the mles assuming complete information we develop the EM algorithm for the estimation of the parameters $(s_i, q_i, i \in W)$. We note that because of (7) we don't have to update the value of $q_{ij}, j \neq i$, at each iteration of the algorithm. After the algorithm converges to $(\widehat{s}_i, \hat{q}_i, i \in W)$, we compute $\hat{q}_{ij}$ using (7).

1. **Initialize**
   At the $p + 1$st iteration, $p \geq 0$, set the values of $(s_i, q_i, i \in W)$ to:

   $$s_i^p, q_i^p, i \in W.$$

   Define Q^p to be the intensity matrix with the entries given by $q_{ij}^p = (n_{ij}/n_i)q_i^p, i \neq j$, and $q_{ii}^p = -q_i^p$.

2. **Expectation step**
   For the k'th history, which starts in state r, $1 \leq k \leq n, r \in W$, and does not make any transition compute the probability that it is generated by a stayer:

   $$E^p(Y_k) = \frac{s_r^p}{s_r^p + (1 - s_r^p) \exp(-q_r^p \tau_r^k)}.$$

   For the k'th history with at least one transition set

   $$E^p(Y_k) = 0.$$

11

Then compute the following expectations

$$E^p(\tau_i^S) = \sum_{k=1}^{n} \tau_i^k E^p(Y_k),$$

$$E^p(\tau_i^Q) = \tau_i - E^p(\tau_i^S),$$

$$E^p(m_i^S) = \sum_{k=1}^{n} I_i^k E^p(Y_k),$$

3. **Maximization step**

Compute the quantities

$$s_i^{p+1} = \frac{E^p(m_i^S)}{m_i},$$

and

$$q_i^{p+1} = \frac{n_i}{E^p(\tau_i^Q)}.$$

4. **Iterate**

Go back to Step 2 and iterate until convergence.

3.2   The special case of identical observation horizons

In a special case when all realizations are observed continuously during a fixed period of time $[0, T]$, the estimates of the parameters can be easily obtained by direct maximization of the likelihood function. In this case we define

a_r = number of realizations that stay continuously in state r

b_r = number of realizations with at least one transition that start in state r,

τ_i^A = total time in state i for histories with no transitions,

τ_i^B = total time in state i for histories with at least one transition,

The likelihood function, $L_A(Q, s)$, of the realizations in set $\mathcal{A}$ is

$$L_A(Q, s) = \prod_{k \in \mathcal{A}} \left\{ \prod_r [s_r + (1 - s_r) \exp(-q_r T)]^{I_r^k} \right\} \qquad (8)$$

$$= \prod_r [s_r + (1 - s_r) \exp(-q_r T)]^{a_r},$$

12

and the likelihood function, $L_B(Q, s)$, of the realizations in $\mathcal{B}$ is

$$
\begin{aligned}
L_B(Q, s) &= \prod_{k \in \mathcal{B}} \left\{ \prod_r (1 - s_r)^{I_r^k} \left[\prod_{j \neq i} q_{ij}^{n_{ij}^k} \prod_i \exp(-q_i \tau_i^k) \right] \right\} \qquad (9) \\
&= \prod_r (1 - s_r)^{b_r} \prod_{j \neq i} q_{ij}^{n_{ij}} \prod_i \exp(-q_i \tau_i^B),
\end{aligned}
$$

Thus, the overall loglikelihood function becomes

$$
\begin{aligned}
\log L(Q, s) &= \log L_A(Q, s) + \log L_B(Q, s) \\
&= \sum_r a_r \log s_r + (1 - s_r) \exp(-q_r T) \qquad (10) \\
&\quad + \sum_r b_r \log(1 - s_r) + \sum_{j \neq i} n_{ij} \log q_{ij} - \sum_i q_i \tau_i^B.
\end{aligned}
$$

The score equation with respect to s_r

$$
\frac{\partial \log L}{\partial s_r} = \frac{a_r [1 - \exp(-q_r T)]}{s_r + (1 - s_r) \exp(-q_r T)} - \frac{b_r}{1 - s_r} = 0,
$$

gives

$$
s_r = \frac{a_r - m_r \exp(-q_r T)}{m_r - m_r \exp(-q_r T)}. \qquad (11)
$$

Substituting (11) into (10), we obtain, up to the terms not depending on the parameters,

$$
\log L(Q, s) \sim - \sum_r b_r \log(1 - \exp(-q_r T)) + \sum_{j \neq i} n_{ij} \log q_{ij} - \sum_i q_i \tau_i^B. \quad (12)
$$

From the score equation

$$
\frac{\partial \log L}{\partial q_{ij}} = \frac{b_i T}{1 - \exp(q_i T)} + \frac{n_{ij}}{q_{ij}} - \tau_i^B = 0,
$$

we get

$$
q_{ij} = \frac{n_{ij}(\exp(q_i T) - 1)}{T b_i + \tau_i^B(\exp(q_i T) - 1)} = \frac{n_{ij}}{T b_i (\exp(q_i T) - 1)^{-1} + \tau_i^B}. \qquad (13)
$$

13

Now taking into account that $\sum_{j \neq i} q_{ij} = q_i$, gives the following equation for q_i

$$\frac{n_i}{Tb_i(\exp(q_iT) - 1)^{-1} + \tau_i^B} = q_i, \qquad (14)$$

which can be rewritten as

$$(n_i - q_i\tau_i^B)(\exp(q_iT) - 1) = q_iTb_i. \qquad (15)$$

It follows from (15), that $\hat{q}_i < n_i/\tau_i^B$. The equation (14) can be easily solved in an iterative fashion for $\hat{q}$. At the n'th iteration the left hand side of this equation is evaluated at $q_i^{(m)}$. This results in the value $q_i^{(m+1)}$. The iterations are repeated until convergence is achieved. The starting value for q_i could be taken as any value in the interval $(0, n_i/\tau_i^B)$. By (13), the estimates of the transition rates q_{ij} are given by $\hat{q}_{ij} = (n_{ij}/n_i)\hat{q}_i$, and the estimate of s_i is obtained from (11).

3.3 The likelihood ratio test

We note that the a Markov chain can be obtained from the mover-stayer model by setting all s_i equal to zero, that is, a Markov chain is nested in the mover-stayer model. This allows us to use the likelihood ratio statistic to test a Markov chain model against a mover-stayer model. The hypothesis test is of the form $H_0 : s = 0$ versus $H_1 : s \neq 0$, where the equality $s = 0$ and the inequality should be understood in the vector sense. The likelihood ratio statistic is

$$\Lambda = \sup_{Q,s=0} L(Q,s) / \sup_{Q,s} L(Q,s) = L(\hat{C}, 0)/L(\hat{Q}, \hat{s}).$$

Here $\hat{C}$ is the mle of the intensity matrix Q under H_0, that is, when the process is assumed to be a Markov chain, and $\hat{Q}, \hat{s}$ are the mles of the intensity matrix and fractions of stayers, respectively, in the mover-stayer model. By the standard result, under H_0, the asymptotic distribution of $-2 \log \Lambda$ is χ^2 with w degrees of freedom.

We now compute $-2 \log \Lambda$. We write the likelihood function for the mover-stayer model as $L(Q,s) = L_A(Q,s)L_B((Q,s)$, where, $L_B((Q,s)$ is given in (9), and

$$L_A(Q,s) = \prod_{k \in \mathcal{A}} \left\{ \prod_r \left[s_r + (1 - s_r) \exp(-q_r\tau_r^k) \right]^{I_r^k} \right\},$$

14

Note that this is more general than (8), because here we do not assume that realizations are observed over the same time horizons. The likelihood function of the observations under H_0 evaluated at $\hat{C}$ is

$$L(\hat{C}, 0) = \prod_k \left(\prod_{i \neq j} (\hat{c}_{ij})^{n_{ij}^k} \prod_i \exp(-\hat{c}_i \tau_i^k) \right),$$

where $\hat{c}_{ij} = (n_{ij}/n_i)\, \hat{c}_i$, and $\hat{c}_i = n_i/\tau_i$. In order to simplify the expression for $-2 \log \Lambda$, we write $L(\hat{C}, 0) = L_A(\hat{C}, 0) L_B(\hat{C}, 0)$, where

$$L_A(\hat{C}, 0) = \prod_{k \in \mathcal{A}} \left(\prod_i \exp(-\hat{c}_i \tau_i^k) \right) = \exp\left(-\sum_i \hat{c}_i \tau_i^A \right)$$

$$L_B(\hat{C}, 0) = \prod_{k \in \mathcal{B}} \left[\prod_{j \neq i} \prod_i \exp(-\hat{c}_i \tau_i^k)^{n_{ij}^k} \right] = \prod_{k \in \mathcal{B}} \left[\prod_{j \neq i} (\frac{n_{ij}}{n_i} \hat{c}_i)^{n_{ij}^k} \prod_i \exp(-\hat{c}_i \tau_i^k) \right]$$

Now, evaluating $L(Q, s)$ at $\hat{Q}, \hat{s}$, and noting (7), we obtain

$$\begin{aligned} \Lambda &= \frac{L_A(\hat{C}, 0)}{L_A(\hat{Q}, \hat{s})} \prod_i \left\{ \left(\frac{\hat{c}_i}{\hat{q}_i} \right)^{n_i} \exp\left[(\hat{q}_i - \hat{c}_i)\tau_i^B \right] \right\} \\ &= \prod_i \left\{ \left(\frac{\hat{c}_i}{\hat{q}_i} \right)^{n_i} \exp(\hat{q}_i \tau_i^B - \hat{c}_i \tau_i) \right\} / L_A(\hat{Q}, \hat{s}) \\ &= \prod_i \left\{ \left(\frac{\hat{c}_i}{\hat{q}_i} \right)^{n_i} \exp(\hat{q}_i \tau_i^B - n_i) \right\} / L_A(\hat{Q}, \hat{s}), \end{aligned}$$

or

$$-2 \log \Lambda = -2 \left\{ \sum_i n_i \log \left(\frac{\hat{c}_i}{\hat{q}_i} \right) + \sum_i (\hat{q}_i \tau_i^B - n_i) - \log L_A(\hat{Q}, \hat{s}) \right\}. \quad (16)$$

For the case of realizations observed over fixed time horizon $(0, T)$, using (8) and (11), we get

$$-2 \log \Lambda = -2 \left\{ \sum_i n_i \log \left(\frac{\hat{c}_i}{\hat{q}_i} \right) + \sum_i (\hat{q}_i \tau_i^B - n_i) - \sum_i a_i \log(a_i/m_i) \right\}.$$

4 Application to bond ratings migration

4.1 The Data and the Methods

The data consist of the rating histories of 856 corporate bond issuers in the industrial sector that were observed for some time in the period from January, 1985 to December, 1995. The data was obtained from Moody's and thus uses Moody's rating system. As is customary, and in our case necessary due to the small sample size, we grouped the original ratings into eight states: Aaa, Aa, A, Baa, Ba, B, C, D and WR where the ratings are ordered from the highest to the lowest with Aaa being the top ranking, D being the default state and WR denoting the state of rating withdrawal. This resulted in an initial distribution of 7, 50, 131, 119, 264, 256, 29, 0, 0 bonds in these states, respectively. Thus a majority of the bonds under consideration were issued with Ba or B rating.

About 30% of issuers entered the sample after 1990, thus providing us with only with relatively short times under study. Because of this sample limitation we decided to study aging effect only in the first five years of life of the bonds. To study this effect we estimate a time-nonhomogeneuous Markov chain and a mover-stayer model defined in Section 2 on the age interval $(0, 5)$. In both processes time t represents age of the bond issuer. Thus, we assume that a generator of a Markov chain, $Q(t)$, is a piecewise constant function of age:

$$
\begin{aligned}
Q(t) &= Q(1), 0 \leq t \leq 1, \qquad\qquad (17)\\
&= Q(2), 1 < t \leq 2,\\
&\vdots\\
&= Q(5), 4 < t \leq 5,
\end{aligned}
$$

and in case of a mover-stayer model, that the $k'th$ one-year age interval has its own vector of proportions of stayers, $s(k) = (s_i(k), 1 \leq i \leq 8), 1 \leq k \leq 5$, where i refers to a rating.

We now discuss estimation of the age specific transition matrices under the two models. To obtain these matrices under a Markov model we first estimate $Q(k)$ for each age interval. The mle $\hat{C}(k)$ of $Q(k)$, is given by (see for example, Andersen et al. (1993))

$$\hat{c}_{ij}(k) = \frac{n_{ij}(k-1,k)}{\int_{k-1}^{k} Y_i(s)ds}, i \neq j, \hat{c}_{ii}(k) = -\sum_{j \neq i} \hat{c}_{ij}(k) \equiv -\hat{c}_i(k) \qquad (18)$$

where $n_{ij}(k-1,k)$ is the total number of $i \rightarrow j$ transitions for all issuers in the $k'th$ year of their life and $Y_i(s)$ is the number of issuers in rating category i at time s. Thus, $\int_{k-1}^{k} Y_i(s)ds$ is the total exposure time in rating category i in the age interval $(k-1,k)$. We note that the estimators $\hat{c}_{ij}(k)$ use all of the available information and do not require that an issuer be present for the whole observation period.

To estimate the age specific transitions matrices under the mover-stayer model we estimate for each age interval the generator for the movers and the vector of proportions of stayers using the EM algorithm developed in Section 3.1. After testing that the algorithm converged to the same final values for different initial values we chose the initial values for the algorithm in the following way. For each one-year age interval we chose the observed proportion of bonds that stayed in a rating for that whole interval as an initial value for the proportion of stayers in this rating. For the $k'th$ age interval we chose $\hat{c}_i(k)$, as the initial value for $q_i, 1 \leq i \leq 8$, where $\hat{c}_i(k)$ is given in (18).

4.2   Estimation results

We summarize here the results of the estimation of the mover-stayer and the Markov chain models for the five age intervals. As default probabilities play an important role in the pricing of bonds and other related applications in finance, the main focus of the summary will be the comparison of default probabilities estimated by the two models.

First, for the purpose of the illustration we report and compare the results of the estimation of the two models for the age interval $(2,3)$. The results are in Tables 1-6. We note (Table 4) that for this age interval there is a very high proportion of stayers in rating C (81%). As a result the CC entry in the mover-stayer transition matrix (Table 6) is much larger (0.8594) than the corresponding one in the Markov chain transition matrix (Table 3) (0.7288). Since there is little movement out of the C rating to other nondefault ratings, this, in turn, results in the smaller default probability from the C rating in the mover-stayer model (0.0826) as compared to the Markov chain estimate (0.1569). Rating Baa also has a relatively high proportion of stayers (56%)

but the $BaaBaa$ entry is only somewhat larger (0.8715) in the mover-stayer matrix than in the Markov chain matrix (0.8558), and the default probabilities from this rating estimated by the two models are comparable. In fact the mover-stayer model predicts somewhat higher probability of default from this rating. Here the small increase in the $BaaBaa$ entry in the mover-stayer model as compared to the Markov chain results in the mover-stayer's model lower probability of the downgrade from Baa to Ba (0.0379) as compared to the Markov chain's estimate (0.0425). Thus, a large proportion of stayers in a rating may have a substantial (as in the case of the C rating) or relatively minor (as in the case of Baa rating.) implication for the default probability from that rating.

We now consider the estimation results for the five age intervals. For each one-year age interval including the one just discussed, we carried out the likelihood ratio for discrimination between the mover-stayer and Markov chain. The test rejected the Markov chain in favor of the mover-stayer model at less than 0.1 percent significance level in each of those intervals.

To obtain an overall heuristic measure of the proximity of the mover-stayer and Markov chain models we computed for each age interval the expected proportion of stayers as estimated by the mover-stayer model. This is defined for each one-year age interval as a weighted average of the estimated proportions of stayers in each state with weights given by the initial distribution. The expected stayer proportions estimated for the first through fifth year of age were 0.3417, 0.2483, 0.1748, 0.1549 and 0.1345 respectively, demonstrating that the estimated proportion of issuers that are not going to move out of their ratings during a one year horizon decreases as bond issuers age and therefore the two models are most different for the young issuer age and become similar with age.

We now consider the dependence of the default probabilities on age of a bond. In Table 7 we report the estimated by the two models one-year default probabilities from speculative ratings for bonds 0-4 years old. In the last column of this table we also report the proportion of stayers in each speculative rating for each age interval.

The one-year probabilities of default from rating C estimated by the mover-stayer model are substantially smaller than those estimated by a Markov chain. (For bonds just issued the one-year probability of default is estimated to be zero by both models.) The difference in the default probabilities estimated by the two models is largest for young bonds and then decreases monotonically with age. For both models the probability of default is largest

in the fifth year and much smaller for younger bonds.

The two models estimate similar probabilities of default from rating B for all considered ages with the mover-stayer proving somewhat larger estimates for 1-2 years old bonds. These probabilities increase with age in a monotonic fashion reflecting the aging of the bonds. The estimated proportion of stayers in rating B declines from 15% in the first year of bond issuer's life to 0 in the fifth year. This explains the virtual equality of probabilities of default from B estimated by the two models in the fifth year of life of the bonds. There is no clear age pattern in the estimated probabilities of default from rating Ba. However, the estimates of default probabilities from this rating in the two models are again closer for older bonds.

Our empirical results show that for 1-4 years old bonds the mover-stayer model estimates substantially lower default probabilities from rating C than a Markov chain. These probabilities are particularly different for 1 and 2 years old bonds. It also estimates somewhat different default probabilities than a Markov chain from other speculative ratings for 1 and 2 years old bonds. The empirical discrepancy suggests that a mover-stayer model, as a model subsuming a Markov chain, may provide, in particular for younger bonds, more accurate estimates of the default probabilities than a mover-stayer model. However, a much larger sample of bonds is needed to further assess this possibility.

References

Albert, A. (1962) Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process, *Annals of Mathematical Statistics.* 38, 727-753.

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

Altman, E. I. and Kao, D. L. (1991) Examining and Modeling Corporate Bond Rating Drift. New York University Salomon Center Working Paper Series s-91-39.

Andersen, P. K., Borgan, O., Gill, R.D., and Keiding, N. (1993) *Statistical Models Based on Counting Processes.* New York, Springer-Verlag.

Asquith, P., Mullins, D.W., and Wolff, E.D. (1989) Original Issue High Yield Bonds: Aging Analyses of Defaults, Exchanges, and Calls. *The Journal of Finance.* 44, 923-952.

Blumen, I., Kogan, M., and McCarthy, P.J. (1955) *The Industrial Mobility of Labor as a Probability Process*, Cornell Studies of Industrial and Labor Relations, vol. **6,** Ithaca, N.Y., Cornell University Press.

Chatterjee, R. and Ramaswamy, V. (1996), An Extended Mover-Stayer Model for Diagnosing the Dynamics of Trial and Repeat for a New Brand. *Applied Stochastic Models and Data Analysis,* Vol. 12, 165-178.

Chen, H. H., Duffy, S. W. and Tabar, L. (1997) A Mover-Stayer Mixture of Markov Chain Models for the Assessment of Dedifferentiation and Tumor Progression in Breast Cancer. *Journal of Applied Statistics* 24 (3), 265-278.

Colombo, R. A. and Morrison, D. G. (1989)) A Brand Switching Model with Implications for Marketing Strategies. *Marketing Science, 8 (*Winter*),* 89-99.

Cook, R. J., Kalbfleisch, J. D. and Yi, Y. (2001) A Generalized Mover-Stayer Model for Panel Data.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion).*Journal of Royal Statistical Society*, Series B, **39**, 1-38.

Fougère, D. and Kamionka, T., (2002) Bayesian Inference for the Mover-Stayer Model in Continuous Time with an Application to Labour Market Transition Data., forthcoming in *Journal of Applied Econometrics.*

Frydman, H., (1984) Maximum Likelihood Estimation in the Mover-Stayer Model., *Journal of the American Statistical Association*, **79**, 632-637.

Frydman, H., Kallberg, J.G., and Kao, D.L. (1985) Testing the adequacy of Markov chains and mover-stayer models as representations of credit behavior., *Operations Research*, **33**, 1203-1214.

Frydman, H. (2002) Estimation in the Mixture of Markov Chains Moving at Different Speeds. Manuscript available from the author.

Fuchs, C. and Greenhouse, J. B. (1988) The EM Algorithm for maximum likelihood estimation in the mover-stayer model.,*Biometrics*, **44**, 605-613.

Keenan, S. C., Sobehart, J. and Hamilton, D.T. (1999) Predicting Default Rates: A Forecasting Model for Moody's Issuer-Based Default Rates. *Moody's Special Comment,* August.

Lando, D. and Skodeberg, T. M. (2002) Analyzing Rating Transitions and Rating Drift with Continuous Observations. *The Journal of Banking and Finance,* 26, 423-444.

Norris, J. R. (1997) *Markov Chains.* Cambridge University Press.

Sampson, M. (1990) A Markov Chain Model for Unskilled Workers and the Highly Mobile. *Journal of the American Statistical Association,* 85, 177-180.

Swensen, A., (1996) On Maximum Likelihood Estimation in the Mover-Stayer Model., *Communications in Statististics-Theory and Methods,* **25**, 1717-1728.

Table 1: Initial distribution of two years old bonds and observed number of stayers for each rating in the age interval (2,3)

Rating	Aaa	Aa	A	Baa	Ba	B	C	WR
Initial Distribution	6	39	126	104	253	193	23	45
Observed number of stayers	5	31	106	91	208	157	20	43

Table 2: Maximum likelihood estimate of Markov chain generator for the (2,3) age interval

	Aaa	Aa	A	Baa	Ba	B	C	D	WR
Aaa	-0.1593	0.0000	0.1593	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Aa	0.0283	-0.2262	0.1697	0.0000	0.0000	0.0000	0.0000	0.0000	0.0283
A	0.0000	0.0166	-0.1745	0.1329	0.0000	0.0000	0.0000	0.0000	0.0249
Baa	0.0000	0.0000	0.0296	-0.1580	0.0494	0.0296	0.0000	0.0000	0.0494
Ba	0.0000	0.0000	0.0000	0.0128	-0.1967	0.0684	0.0086	0.0043	0.1026
B	0.0000	0.0000	0.0000	0.0000	0.0816	-0.2214	0.0350	0.0466	0.0583
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.3164	0.1808	0.1356
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WR	0.0000	0.0000	0.0159	0.0000	0.0000	0.0637	0.0000	0.0318	-0.1114

Table 3: Maximum likelihood estimate of Markov chain transition probability matrix for the (2,3) age interval

	Aaa	Aa	A	Baa	Ba	B	C	D	WR
Aaa	0.8527	0.0011	0.1350	0.0090	0.0001	0.0001	0.0000	0.0000	0.0019
Aa	0.0233	0.7987	0.1411	0.0095	0.0002	0.0009	0.0000	0.0004	0.0259
A	0.0002	0.0136	0.8429	0.1127	0.0028	0.0024	0.0000	0.0004	0.0248
Baa	0.0000	0.0002	0.0255	0.8558	0.0425	0.0274	0.0006	0.0015	0.0465
Ba	0.0000	0.0000	0.0009	0.0108	0.8241	0.0585	0.0076	0.0075	0.0906
B	0.0000	0.0000	0.0004	0.0005	0.0663	0.8053	0.0270	0.0456	0.0549
C	0.0000	0.0000	0.0009	0.0000	0.0001	0.0035	0.7288	0.1569	0.1098
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
WR	0.0000	0.0001	0.0138	0.0009	0.0022	0.0540	0.0009	0.0315	0.8965

Table 4: Maximum likelihood etimates of the proportion of "stayers" in the (2,3) age interval

Rating	Aaa	Aa	A	Baa	Ba	B	C		WR
s_i	0	0	0.11	0.56	0	0.07	0.81		0.75

Table 5: Maximum likelihood estimate of the generator for the movers in the (2,3) age interval

	Aaa	Aa	A	Baa	Ba	B	C	D	WR
Aaa	-0.1594	0.0000	0.1594	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Aa	0.0283	-0.2268	0.1701	0.0000	0.0000	0.0000	0.0000	0.0000	0.0283
A	0.0000	0.0187	-0.1967	0.1499	0.0000	0.0000	0.0000	0.0000	0.0281
Baa	0.0000	0.0000	0.0660	-0.3522	0.1101	0.0660	0.0000	0.0000	0.1101
Ba	0.0000	0.0000	0.0000	0.0129	-0.1972	0.0686	0.0086	0.0043	0.1029
B	0.0000	0.0000	0.0000	0.0000	0.0877	-0.2379	0.0376	0.0501	0.0626
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-1.3100	0.7486	0.5614
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WR	0.0000	0.0000	0.0314	0.0000	0.0000	0.1256	0.0000	0.0628	-0.2197

Table 6: Maximum likelihood estimate of transition probability matrix of the mover - stayer model in the (2,3) age interval

	Aaa	Aa	A	Baa	Ba	B	C	D	WR
Aaa	0.8527	0.0012	0.1336	0.0094	0.0004	0.0003	0.0000	0.0001	0.0022
Aa	0.0233	0.7988	0.1399	0.0099	0.0004	0.0017	0.0000	0.0008	0.0250
A	0.0002	0.0135	0.8458	0.1019	0.0059	0.0051	0.0001	0.0009	0.0266
Baa	0.0000	0.0002	0.0227	0.8715	0.0379	0.0254	0.0005	0.0022	0.0396
Ba	0.0000	0.0000	0.0017	0.0099	0.8246	0.0608	0.0050	0.0106	0.0874
B	0.0000	0.0000	0.0008	0.0004	0.0661	0.8078	0.0172	0.0519	0.0557
C	0.0000	0.0000	0.0010	0.0001	0.0001	0.0039	0.8594	0.0826	0.0530
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
WR	0.0000	0.0001	0.0063	0.0005	0.0011	0.0247	0.0003	0.0147	0.9524

Tables 7a-c: One-year default probabilities from ratings C, B and Ba using mover - stayer and Markov chain models.

Table 7a

C-->D			
Year	**Markov Chain**	**Mover - Stayer Model**	**Proportion of Stayers in C**
1	0.0000	0.0000	0.88
2	0.1920	0.1153	0.77
3	0.1569	0.0826	0.81
4	0.1000	0.0737	0.81
5	0.2118	0.1892	0.43

Table 7b

B-->D			
Year	**Markov Chain Model**	**Mover - Stayer Model**	**Proportion of Stayers in B**
1	0.0039	0.0039	0.15
2	0.0379	0.0435	0.3
3	0.0456	0.0519	0.07
4	0.1006	0.1038	0.13
5	0.1173	0.1189	0

Table 7c

Ba-->D			
Year	**Markov Chain Model**	**Mover - Stayer Model**	**Proportion of Stayers in Ba**
1	0.0000	0.0000	0.86
2	0.0379	0.0064	0
3	0.0075	0.0106	0
4	0.0202	0.0214	0
5	0.0077	0.0088	0