



ELSEVIER

Journal of Banking & Finance 26 (2002) 423–444

Journal of
BANKING &
FINANCE

www.elsevier.com/locate/econbase

Analyzing rating transitions and rating drift with continuous observations

David Lando ^{a,*}, Torben M. Skødeberg ^b

^a Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

^b Nordea Investment Management, Strandgade 3, DK-1401 Copenhagen K, Denmark

Abstract

We consider the estimation of credit rating transitions based on continuous-time observations. Through simple examples and using a large data set from Standard and Poor's, we illustrate the difference between estimators based on discrete-time cohort methods and estimators based on continuous observations. We apply semi-parametric regression techniques to test for two types of non-Markov effects in rating transitions: Duration dependence and dependence on previous rating. We find significant non-Markov effects, especially for the downgrade movements. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C41; G33

Keywords: Rating transitions; Rating drift; Markov chains; Estimation

1. Introduction

Transition matrices are at the center of modern credit risk management. The reports on rating migrations published by Standard and Poor's and Moody's are studied by credit risk managers everywhere and several of the most

* Corresponding author. Tel.: +45-3532-0683; fax: +45-3532-0678.

E-mail addresses: dlando@math.ku.dk (D. Lando), torben.m.skodeberg@post.tele.dk (T.M. Skødeberg).

prominent risk management tools, such as J.P. Morgan's Credit Metrics and McKinsey's Credit Portfolio View are built around estimates of rating migration probabilities.

In essence, the estimates published by these agencies and in the published academic literature use a discrete-time setting and rely on a 'cohort' method which estimates the transition rates as follows: Given that there are N_i firms in a given rating category i at the beginning of the year and that out of this population N_{ij} have migrated to the category j , then the one year transition rate is estimated as

$$\hat{p}_{ij} = \frac{N_{ij}}{N_i}, \quad j \neq i. \quad (1)$$

An important consequence of this is that if a transition from i to j does not occur in a given period, the estimate of the corresponding rate is 0.

The rating agencies of course have access to continuous-time data on rating transitions and know the exact dates within a year that a company changes its rating or is downgraded. Similarly, a bank using an internal rating system will have access to a complete history of rating transitions. We argue in the following, that it is crucial to base the estimation of transition rates on these continuously observed histories to get efficient estimates of transition rates. This point is particularly important when estimating 'rare events' such as the transition from, say AAA in Standard and Poor's rating, to default. Very briefly stated, the maximum-likelihood estimator that one obtains for the one-year transition probability from AAA to default will be (and should be) non-zero even if there has been no direct *or indirect* defaults (i.e. default through a sequence of downgrades) in the period of observation. Briefly stated, if in a one year period there are no transitions from AAA to default, but there are transitions from AAA to AA and from AA to default (but by other firms), then the estimator for transitions from AAA to default should be non-zero, since evidently there is a chance of defaulting within a year after successive downgrades, even if it did not happen for one single firm in the sample. The continuous-time estimator captures this whereas the discrete-time method does not.

Apart from getting a better grip on the rare events, the continuous time methodology based on modern survival analytic techniques (see Skodeberg, 1998) and similar observations in the parallel work by Kavvathas (2000) has a number of additional advantages:

1. The framework permits a rigorous formulation and testing of assumptions 'rating drift' and other non-Markov type behavior (such as seasoning effects) investigated in for example Altman and Kao (1992a,b), Lucas and Lonski (1992), and Carty and Fons (1993).

2. The dependence on external covariates can be formulated and tested, and changes in 'regimes' either due to business cycles, as in for example Nickell

et al. (2000), or changes in rating policies, as indicated by Blume et al. (1998) can be quantified.

3. The continuous-time formulation hooks up nicely with rating-based term structure modeling in which one tries to estimate and calibrate yield curves for different rating classes, see for example Jarrow et al. (1997), Lando (1998) and Das and Tufano (1996).

4. Censoring is handled easily within the continuous-time framework. According to Carty (1997) only few (roughly 13%) of the migrations to the not-rated category are related to changes in credit quality and this observation is used there and in Nickell et al. (2000) as an argument in support of excluding the issuers who experience a transition to the not-rated category. Using the survival theoretic setting of this paper, the conclusion is the opposite: The very fact that transitions to not rated was caused by rating unrelated event justifies the inclusion of these events as censored variables, thus permitting a full use of the sample information. The time before a firm migrates to the not-rated category contains valuable information on the changes that did not occur in the time before this event. Note that the framework presented in Shumway (2001) would also allow for censoring to be treated rigorously, but his framework is still discrete-time.

5. When estimating homogeneous chains in continuous-time by estimating the generator of the continuous-time Markov chain, we avoid the ‘embedding problem’ for Markov chains (for more on this in a rating modeling context, see Israel et al. (1999)). This problem arises because not every discrete time Markov chain can be realized as a discretized continuous-time chain. Hence it may be impossible from a one-year transition matrix (for example if it contains zeros in some of the non-default rows) to construct a continuous-time chain which has the one-year transition matrix as its ‘marginal’. However, the continuous-time chain is very useful in that it allows cash flows occurring at all dates to be weighted by the exact survival probability corresponding to the chosen time horizon.

One of the most important goals behind the current effort to revise the Basel Capital Accord is to replace the existing risk weights with a system which more clearly recognizes the differences in risk of various instruments. It is likely that rating systems will play a larger role in quantifying these differences. The statistical framework presented in this paper is a natural framework for quantitatively assessing internal and external rating systems used by financial institutions.

The outline of the paper is as follows:

In Section 2, we briefly describe the data used in our study. In Section 3, we present the basic idea of continuous-time estimation in the framework of a homogeneous Markov chain. A simple example illustrates the importance of using continuous-time data. In Section 4 we describe how a time-inhomogeneous transition probability matrix may be estimated. Section 5 outlines the

statistical framework and formulates a rigorous notion of rating drift. In Section 6, the test results are presented, and Section 7 concludes. Appendix A summarizes the technical material needed for the paper.

2. The data

The data covers 17 years of rating history in the S&P system starting on 1 January, 1981 and ending 31 December, 1997. There are a total of 6659 firms which are rated at some point or another. The ratings are listed in the classification based on a total of 22 classifications. The top rating is AAA. Then follows AA+ and from then each of the categories AA, A, BBB, BB, B, CCC contain three ratings obtained by possibly adding '+' or '-' to the letter grade. Finally, there are some instances of CC and C ratings and a default category denoted D. For some results, we have chosen to look at groupings into eight categories which contain the seven letter categories (without plus or minus, and CC and C grouped into CCC) and the default category. There are a total of 7282 transitions recorded within the system of eight categories including transitions to NR. For the system consisting of 18 classes (in which all ratings including the letter C are grouped into one CCC category) there are a total of 11 606 transitions, again counting the number of transitions to the NR category. For each firm the exact transition dates between ratings (including default) are recorded and so are dates where the rating is discontinued. In these cases the firm receives the not rated (NR) assignment. There are 114 cases of transitions back from the NR category.

5405 out of the total population of 6659 firms are US companies. We do not have the names of these firms but we do have access to the distribution of the firms on industries. We are mainly looking at the aggregated data set to clearly illustrate our methods and to get enough observed transitions between categories to make inference possible. We do, however, briefly consider the results for the financial firms separately at the end.

3. The time-homogeneous case

Appendix A contains an overview of the necessary theory of Markov chain modeling that we need for the entire paper. For this section we only need to note the following facts: Throughout, we consider a K -state Markov chain where we think of state 1 as the highest rating category and state K as the default state. We collect the transition probabilities of the Markov chain for a given time horizon in a $K \times K$ matrix $P(t)$ whose ij 'th element is the probability of migrating from state i to state j in a time period of t . Just like a discrete-time Markov process on the rating classes can be obtained by matrix multiplication

from the one-period transition matrix, there exists a simple representation of the matrices $P(t)$ for arbitrary time horizons t for a continuous-time chain on the same state space. The generator matrix A is a $K \times K$ matrix for which

$$P(t) = \exp(At), \quad t \geq 0. \tag{2}$$

Here, At is the matrix A multiplied by t on every entry and the exponential function is a matrix exponential, as defined in Appendix A. The critical thing to note, is that the transition probabilities for every time horizon is a function of the generator. Hence, one can obtain maximum-likelihood estimators of the transition probability matrices by first obtaining the maximum-likelihood estimate of the generator and then applying the matrix exponential function on this estimate, scaled by the time horizon. The entries of the generator A satisfy

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

These entries describe the probabilistic behavior of the holding time in state i as exponentially distributed with parameter λ_i , where $\lambda_i = -\lambda_{ii}$, and the probability of jumping from state i to j given that a jump occurs is given by λ_{ij}/λ_i . To estimate the elements of the generator under an assumption of time-homogeneity we use the maximum likelihood estimator (see for example Küchler and Sørensen, 1997)

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

where $Y_i(s)$ is the number of firms in rating class i at time s and $N_{ij}(T)$ is the total number of transitions over the period from i to j , where $i \neq j$. The intuition is straightforward: the numerator counts the number of observed transitions from i to j over the entire period of observation. The denominator has the number of ‘firm-years’ spent in state i . Any period a firm spends in a state will be picked up through the denominator. In this sense all information is being used. We now illustrate the estimator both through a simple example and on our data set. The simple example will give the intuition of the procedure. The application on our data set will test the practical significance of using the continuous-time technique.

To illustrate the estimator, consider a rating system consisting of two non-default rating categories A and B and a default category D. Assume that we observe over one year the history of 20 firms, of which 10 start in category A and 10 in category B. Assume that over the year of observation, one A rated firm changes its rating to category B after one month and stays there the rest of the year. Assume that over the same period, one B rated firm is upgraded after two months and remains in A for the rest of the period and a firm which started

in B defaults after six months and stays there for the remaining part of the period. In this case we have for one of the entries

$$\hat{\lambda}_{AB} = \frac{N_{AB}(1)}{\int_0^1 Y_A(s) ds} = \frac{1}{9 + 1/12 + 10/12} = 0.10084.$$

Proceeding similarly with the other entries (and noting that the state D is assumed to be absorbing and the diagonal elements just make sure that rows sum to zero) we obtain the estimated generator

$$\hat{\Lambda} = \begin{pmatrix} -0.10084 & 0.10084 & 0 \\ 0.10909 & -0.21818 & 0.10909 \\ 0 & 0 & 0 \end{pmatrix}.$$

From this, we obtain the maximum likelihood estimator of the one-year transition matrix as

$$\widehat{P}(1) = \begin{pmatrix} 0.90887 & 0.08618 & 0.00495 \\ 0.09323 & 0.80858 & 0.09819 \\ 0 & 0 & 1 \end{pmatrix}.$$

Had we instead used a cohort method the result would have been

$$\widehat{P}(1) = \begin{pmatrix} 0.90 & 0.10 & 0 \\ 0.10 & 0.80 & 0.10 \\ 0 & 0 & 1 \end{pmatrix}.$$

As we see, the traditional method does not capture default risk in the A category simply because there is no firm defaulting directly from A. Note that the continuous-time method produces a strictly positive estimator for default from A over one year despite the fact that no firm in the sample defaults in one year from A. This is appropriate because the probability of migrating to B and the probability of default from B are clearly both positive. As a side remark, note that in a classical cohort analysis the firm upgraded from B does not provide more information than the upgrade. Here, it matters exactly when the upgrade took place, and the six months spent in A with no further change contributes information to the estimate of the transition intensity from rating class A.

We now consider how this difference materializes in our Standard and Poor’s data set. We consider a 10 year sub-period from 1988 to 1998 to have a reasonable large starting pool for the cohort method. If we use the cohort method and estimate the one-year transition rates for each of the ten years and then take an average of the estimated matrices we obtain the result presented in Table 1. We have chosen to include transitions to and from the NR category. As we will see later, we can easily modify the estimates to exclude that category by using estimation under censoring. In Table 2 we report the estimated gen-

Table 1

This shows the average of 10 one-year transition matrices, each estimated using a cohort method in the period 1988–1998^a

	NR	AAA	AA	A	BBB	BB	B	CCC	D
NR	0.9939	0.0001	0.0000	0.0001	0.0006	0.0006	0.0004	0.0000	0.0044
AAA	0.0266	0.9040	0.0607	0.0070	0.0000	0.0016	0.0000	0.0000	0.0000
AA	0.0302	0.0054	0.8786	0.0791	0.0039	0.0006	0.0008	0.0000	0.0000
A	0.0401	0.0004	0.0157	0.8903	0.0445	0.0068	0.0017	0.0001	0.0003
BBB	0.0583	0.0001	0.0028	0.0519	0.8375	0.0388	0.0068	0.0018	0.0018
BB	0.0906	0.0000	0.0003	0.0051	0.0795	0.7452	0.0587	0.0110	0.0095
B	0.1268	0.0000	0.0008	0.0015	0.0050	0.0730	0.7081	0.0326	0.0500
CCC	0.1658	0.0020	0.0000	0.0061	0.0089	0.0279	0.1003	0.4842	0.2048
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

^a We show also transitions to and from the NR category.

Table 2

The maximum-likelihood estimator of the generator based upon continuous-time observation over the 10-year period 1988–1998

	NR	AAA	AA	A	BBB	BB	B	CCC	D
NR	-0.0066	0.0000	0.0001	0.0003	0.0006	0.0010	0.0003	0.0000	0.0043
AAA	0.0248	-0.1062	0.0720	0.0071	0.0000	0.0024	0.0000	0.0000	0.0000
AA	0.0322	0.0068	-0.1301	0.0858	0.0044	0.0004	0.0004	0.0000	0.0000
A	0.0431	0.0004	0.0144	-0.1136	0.0499	0.0045	0.0011	0.0002	0.0000
BBB	0.0551	0.0003	0.0023	0.0548	-0.1691	0.0496	0.0061	0.0006	0.0003
BB	0.1017	0.0000	0.0012	0.0078	0.1101	-0.3213	0.0866	0.0108	0.0030
B	0.1713	0.0000	0.0027	0.0020	0.0061	0.0904	-0.4052	0.1038	0.0290
CCC	0.2099	0.0042	0.0000	0.0084	0.0000	0.0336	0.1301	-0.9697	0.5835
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

erator using the estimator given in Eq. (3) and by taking the matrix exponential of that estimator we obtain the estimated transition probabilities given in Table 3. As we can see from this table, the most important difference is the fact that with four decimal points there is a measurable default probability even for the highest rating categories. but note also the sizeable difference in the one-year default probability of a CCC-rated firm when using the continuous-time estimation method. One reason for this difference is the following: When using a cohort method based on yearly samples, we will only record a migration from CCC to default when a firm starts out in CCC in the beginning of the year in which the default occurs. Many firms in the sample are downgraded to CCC during the year and only stay there a short time before default. These will not be recorded as defaults from CCC in the cohort method, but they will be recorded in the method based on the continuous sample. This explains the increase in the CCC default frequency. It should also be noted, that in the sample, almost all ratings observed to be in the C and CC category ended up in

Table 3

The one-year transition matrix estimated from continuous-time data over the period 1988–1998 as the matrix exponential of the maximum likelihood estimator of the generator

	NR	AAA	AA	A	BBB	BB	B	CCC	D
NR	0.9935	0.0000	0.0001	0.0003	0.0006	0.0009	0.0003	0.0000	0.0043
AAA	0.0248	0.8995	0.0640	0.0091	0.0005	0.0020	0.0001	0.0000	0.0001
AA	0.0321	0.0061	0.8788	0.0761	0.0057	0.0006	0.0004	0.0000	0.0001
A	0.0424	0.0004	0.0129	0.8944	0.0436	0.0047	0.0011	0.0002	0.0002
BBB	0.0545	0.0003	0.0023	0.0479	0.8479	0.0393	0.0063	0.0008	0.0008
BB	0.0965	0.0000	0.0012	0.0090	0.0869	0.7303	0.0612	0.0084	0.0065
B	0.1518	0.0001	0.0022	0.0024	0.0084	0.0643	0.6734	0.0534	0.0440
CCC	0.1429	0.0025	0.0003	0.0053	0.0017	0.0215	0.0674	0.3824	0.3760
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

default or with a rating withdrawal. The fact that these are grouped into the CCC category may exaggerate the default frequency of this category.

The results above are mainly for illustration purposes. As we will see below there are many reasons to believe that the data are not from a homogeneous Markov chain and we need to modify our estimation methods to take into account both non-homogeneities and the influence from exogenous variables on the rating migration probabilities.

4. Estimating non-homogeneous chains

We have just seen how to use the maximum-likelihood estimator to estimate the generator and transition matrices using continuous data. Even if the method assumes time homogeneity, something which is hard to assume over the long run, it is a useful tool for estimating a one-year transition matrix. However, as we will see in this section another non-parametric method exists. This method is a useful tool for replacing the cohort methods over longer periods of time. Consider a non-homogeneous, continuous-time Markov process η with finite state space $S = \{1, 2, \dots, K\}$ whose transition probability matrix for the period from time s to time t is given by $P(s, t)$. Hence, the ij 'th element of this matrix describes the probability that the chain starting in state i at date s is in state j at date t .

In this section we will explain the so-called *product-limit* estimator, or Aalen–Johansen estimator, for the transition probabilities $P(s, t)$ and the relation to our example. Appendix A elaborates and provides further references. Given that our sample has m transitions over the period from s to t , we can estimate $P(s, t)$ consistently as

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

Here, T_i is a jump time in the interval $[s, t]$ and

$$\Delta \hat{A}(T_i) = \begin{pmatrix} -\frac{\Delta N_1(T_i)}{Y_1(T_i)} & \frac{\Delta N_{12}(T_i)}{Y_1(T_i)} & \frac{\Delta N_{13}(T_i)}{Y_1(T_i)} & \dots & \frac{\Delta N_{1p}(T_i)}{Y_1(T_i)} \\ \frac{\Delta N_{21}(T_i)}{Y_2(T_i)} & -\frac{\Delta N_2(T_i)}{Y_2(T_i)} & \frac{\Delta N_{23}(T_i)}{Y_2(T_i)} & \dots & \frac{\Delta N_{2p}(T_i)}{Y_2(T_i)} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \frac{\Delta N_{p-1,1}(T_i)}{Y_{p-1}(T_i)} & \frac{\Delta N_{p-1,2}(T_i)}{Y_{p-1}(T_i)} & \dots & -\frac{\Delta N_{p-1}(T_i)}{Y_{p-1}(T_i)} & \frac{\Delta N_{p-1,p}(T_i)}{Y_{p-1}(T_i)} \\ 0 & 0 & \dots & \dots & 0 \end{pmatrix}.$$

Here, $\Delta N_{hj}(T_i)$ denotes the number of transitions observed from state h to j at date T_i .¹

$\Delta N_k(T_i)$ counts the total number of transitions away from state k at date T_i and $Y_k(T_i)$ is the number of firms in state k right before date T_i and hence the diagonal element in row k counts, at a given date T_i , the fraction of the exposed firms $Y_k(T_i)$ which leaves the state at date T_i . The off-diagonal elements count the specific types of transitions away from the state divided by the number of exposed firms. Note that the variable Y automatically incorporates censoring in that nothing happens to the estimator on the day of a censoring event (if that is the only event). The number of exposed firms changes, however, and this will affect the estimate on the next date of an observed transition. Note that the bottom row is zero in ΔA because we do not model firms leaving the default state. Note also, that the rows of the matrix $I + \Delta A(T_i)$ automatically sum to 1. In summary, one may view this estimator as a cohort method applied to extremely short time intervals.

Let us briefly consider the method on the example of the previous section to see how the estimator produces yet another candidate for estimating a one-year transition probability matrix. To compute the one-year transition probability matrix non-parametrically, we first compute

$$\Delta A(T_{1/12}) = \begin{pmatrix} -0.1 & 0.1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

¹ This notation is used because $N_{hj}(t)$ counts the total number of transitions observed from h to j from the starting date until time t , and $\Delta N_{hj}(T_i)$ then is an increment of this process. Note that if observations were truly in continuous time, we would have no simultaneous jumps and for every time point t at most one off-diagonal element of $\Delta \hat{A}(t)$ would be non-zero. In practice there are ‘ties’ so that several off-diagonal elements can be non-zero at the same time point and the increment of a particular jump-type $\Delta N_{hj}(T_i)$ may even be larger than one. Given the relative richness of time points (days) and the many types of transitions possible from each class, ties and conventions for handling them do not seem to play an important role in our data set.

$$\Delta A(T_{2/12}) = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{11} & -\frac{1}{11} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\Delta A(T_{1/2}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -0.1 & 0.1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore we get the Aalen–Johansen estimator

$$P(\widehat{0}, 1) = \begin{pmatrix} 0.90909 & 0.08181 & 0.00909 \\ 0.09091 & 0.81818 & 0.09091 \\ 0 & 0 & 1 \end{pmatrix}.$$

As we can see, there is a difference between the estimator based on the generator and this estimator on the default probability. Hence it makes a difference whether we estimate the one-year transition probability based on continuous observations using the exponential of the generator or the non-parametric Aalen–Johansen estimator. One can view the matrix exponential as a smoothed version, and it is clearly this form which is most suited to risk management in that it allows computation of estimated default and transition intensities over arbitrarily short time intervals. The two methods are not dramatically different for large data sets, as illustrated in Tables 4 and 5. As we see, the difference is much less significant than the one between the cohort method and the methods based on continuous samples.

By comparing this Aalen–Johansen estimator over longer time horizons to estimators based on a time-homogeneity assumption, time inhomogeneities will become apparent. This non-parametric estimator does not give a way of detecting the sources of these inhomogeneities. To formulate statistical hypoth-

Table 4
One-year transition probability matrix estimated for the year 1997^a

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	0.95912	0.03982	0.00096	0.00010	0.00000	0.00000	0.00000	0.00000
AA	0.01249	0.93689	0.04519	0.00524	0.00015	0.00004	0.00000	0.00000
A	0.00011	0.01666	0.93097	0.04906	0.00274	0.00042	0.00001	0.00003
BBB	0.00002	0.00253	0.03635	0.90603	0.03955	0.01398	0.00030	0.00125
BB	0.00000	0.00012	0.00318	0.07866	0.85980	0.05411	0.00317	0.00096
B	0.00000	0.00005	0.00495	0.00385	0.07029	0.87618	0.02941	0.01527
CCC	0.00000	0.00004	0.00091	0.02523	0.02890	0.11823	0.52289	0.30380
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

^a The estimator is the maximum likelihood estimator based on continuous observations and derived from the maximum likelihood estimator of the intensity matrix assuming time-homogeneity. The estimator is only marginally different from the Aalen–Johansen estimator for the same period.

Table 5
Aalen–Johansen estimator for the year 1997^a

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	0.95866	0.03926	0.00184	0.00023	0.00001	0.00000	0.00000	0.00000
AA	0.01273	0.93714	0.04440	0.00544	0.00022	0.00007	0.00000	0.00000
A	0.00010	0.01682	0.93088	0.04880	0.00278	0.00061	0.00000	0.00001
BBB	0.00002	0.00252	0.03632	0.90736	0.03888	0.01353	0.00009	0.00128
BB	0.00000	0.00016	0.00347	0.07905	0.86016	0.05342	0.00294	0.00079
B	0.00000	0.00005	0.00507	0.00378	0.07066	0.87599	0.02797	0.01647
CCC	0.00000	0.00000	0.00045	0.02302	0.03104	0.12522	0.51784	0.30242
D	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

^aThis estimator is a non-parametric estimator of the one-year transition probability matrix. It is only marginally different from the parametric maximum likelihood estimator obtained in Table 4.

esis on this, we have to work with the transition intensities, and this is the topic of Section 5.

5. Introducing covariates: The statistical framework

We now set up the appropriate framework for testing whether the transition intensities of a Markov model depend on certain covariates. Our application is to test for non-Markovian behavior, i.e. so-called rating drift (i.e. dependence on previous rating) and waiting-time effects but the framework could also easily handle dependencies on macroeconomic variables (see Kavvathas (2000) for examples of this).

In Appendix A, we have recalled how a matrix of transition intensities characterize the evolution of a non-homogeneous Markov chain. The key assumption in the following is that the transition intensity for each type of rating migration is influenced by an external, time varying covariate. Let Y_{hi} denote an indicator process, which is 1 when the process is in state h and 0 otherwise. We assume that the intensity of transition from state h to state j for firm i is given as

$$\lambda_{hji}(t) = Y_{hi}(t)\alpha_{hji}(t, Z_i(t)),$$

where $\alpha_{hji}(t, Z_i(t))$ has the multiplicative form

$$\alpha_{hji}(t, Z_i(t)) = \alpha_{hj0}(t) \exp(\beta_{hj}Z_i(t)). \tag{5}$$

This statistical formulation is a semi-parametric multiplicative hazard model, which is a proportional intensities regression model. The theory behind this modeling is described in Andersen et al. (1991) and Section VII of Andersen et al. (1993). We have summarized in Appendix A.

Note that the time-varying ‘baseline’ intensity α_{hj0} is unspecified (but non-negative) and the parameter of interest is the regression coefficient β_{hj} . The

covariate Z_i is designed to keep track of an influence on i 's transition intensity. Non-zero values of the product of covariates and the parameter cause the intensities to deviate from the baseline hazard. Thus, if the process of rating changes exhibits non-Markov behavior the regression coefficients are significantly different from zero, and this is exactly what we investigate via statistical tests on β in the coming subsections. The theory behind the estimation of β is done by maximizing a so-called partial likelihood. The theory along with further references is outlined in Appendix A.

We have chosen to work with this model in this paper, since we are concerned with testing non-Markov effects of transitions, i.e. the covariates will be variables describing whether the previous move was an upgrade or a downgrade or the duration in the present state for each firm i . It is, of course, likely (see for example Bangia et al., 2000; Nickell et al., 2000; Kavvathas, 2000) that macroeconomic variables or other indicators of the business cycle influence rating intensities. Indeed, if a rating system attempts to keep the marginal default probabilities relatively constant for a given rating category, then one should see downgrades taking place more often in a recession. But since we do not want in this paper to explain the macroeconomic influences and do not want a fully parametric model to be misspecified due to business cycle variables, we absorb such fluctuations through the baseline intensity. Hence this test is specifically designed to allow us to focus precisely on the kind of deviation from a Markov hypothesis that we are interested in.

The first such deviation which we test for, is whether the last rating change influences the transition probabilities out of the present class.² The basic covariates are defined as

$$Z_i(t) = \begin{cases} 1, & \text{individual } i \text{ was upgraded to the present rating class,} \\ 0, & \text{otherwise.} \end{cases}$$

The statistical test for the hypothesis of no rating drift is seen to be the simple hypothesis

$$H : \beta = 0. \tag{6}$$

This hypothesis is equivalent to no serial correlation in any rating class of previous up- and downgrades. As recalled in the appendix, likelihood theory provides a (partial) likelihood ratio test for the hypothesis H .

Several studies³ address the issue of 'rating drift' which is essentially a 'non-Markov' property in that the history of the rating process – not just the current rating – carries information about the transition probabilities. But one needs to

² For this test, we only include data for firms that experience more than one change of rating. Similarly, for duration dependence, we need at least one rating change.

³ See e.g. Altman and Kao (1992a,b), Carty and Fons (1993, 1994).

be careful in defining what rating drift really means. To give an example of this, consider the notion of rating momentum as used for example in Carty and Fons (1993). That study found the following: for each rating category, the probability of a downgrade following a downgrade within a year significantly exceeds that of an upgrade following a downgrade. This way of tackling serial correlation effects is inappropriate since it does not recognize the dynamics of the Markov chain which may very well for several rating categories have a lower upgrade probability than downgrade probability. One should differentiate the direction in which one came into the current state, not the direction in which one leaves the current state. Indeed, one would expect for high rating categories to see a small probability of an upgrade compared to a downgrade and in a low grade like BB the picture could be reversed. In a continuous-time Markov model the binomial test hypothesis corresponds to asking whether $\sum_{j<h} \alpha_{hj}(t) = \sum_{j>h} \alpha_{hj}(t)$ for all states except for the best and the default rating class, where $\alpha_{hj}(t)$ is the transition intensity at time t between state h and j . This is not a reasonable hypothesis. If in addition (as in some studies) data are aggregated across rating categories such that only total number of up- and down-gradings are considered, then again the ‘drift’ could be a consequence of the composition of the firms in the sample: A high number of firms in low categories would show different results than a sample with a high number of firms in the high categories. Instead, a rigorous test of rating drift must check whether firms in a specific rating class exhibit the same rating behavior regardless of whether they obtained their current rating through an upgrade of a downgrade. Our specification takes care of this problem. But note that extensions could readily be made: one might be interested in differentiating which type of upgrade preceded the current rating, and not just note that it was an upgrade. While this of course gives a more precise statistical hypothesis, it also rapidly decreases the data underlying the test, and it becomes very hard to obtain statistical power.

Second, we study duration dependencies in this model. This requires the following definition of the basic duration covariates Z_i :

$$Z_i(t) = \text{“time since last entry into the present state”}.$$

Since previous empirical evidence has suggested a lower intensity as a function of time spent in a state, we have chosen the exponential form which not only keeps the intensity positive but also lets the effect ‘die out’ as the duration increases when the regression parameter is negative.

6. Test results

First, we consider the question of ‘momentum’ or ‘rating drift’ and ask if the intensity of being upgraded from a state depends on whether the current state

was reached through a downgrade or an upgrade. We ask the same question for downgrades: Is there a tendency of a downgrade to be more likely if the current state was reached through a downgrade. To get enough observed transitions to make meaningful inference, we consider only transitions from the current state to a neighboring state. We consider all possible ways of *reaching* the current state but group together all the downgrades into the current state by assigning the same value of the covariate for these firms. All the upgrades into the current state are then in the other group. The results are shown in Tables 6 and 7. In all cases, except for current ratings BB, CCC+ and CCC, we find a strong, downgrade momentum which in several cases increases the downgrade intensity by a factor of 3. For upgrades, the result is almost the opposite. There is virtually no detectable effect on the upgrade intensity of a previous upgrade except from ratings BBB+, BBB, BB+ and B+.

Next, we ask if the duration in a given rating influences the downgrade or upgrade intensity. Again, to make sure the data material is not too thin, we

Table 6
Results shown are for the test of an effect of a previous downgrade on the intensity of a downgrade to a neighboring state

Ratings		$\hat{\beta}$	std($\hat{\beta}$)	n_1	n_2	p
From	To					
AA+	AA	0.897	0.281	149	65	<0.01
AA	AA-	0.936	0.211	314	100	<0.01
AA-	A+	0.871	0.172	490	162	<0.01
A+	A	0.582	0.147	663	198	<0.01
A	A-	0.868	0.160	842	193	<0.01
A-	BBB+	1.180	0.196	780	161	<0.01
BBB+	BBB	0.714	0.168	721	180	<0.01
BBB	BBB-	1.180	0.222	712	140	<0.01
BBB-	BB+	1.090	0.241	641	95	<0.01
BB+	BB	0.970	0.303	513	59	<0.01
BB	BB-	0.144	0.227	571	82	0.53
BB-	B+	0.858	0.253	522	74	<0.01
B+	B	1.010	0.282	575	87	<0.01
B	B-	0.541	0.457	437	43	<0.01
B-	CCC+	2.030	1.040	271	28	<0.01
CCC+	CCC	6.170	23.5	194	15	0.20
CCC	CCC-	-0.929	0.873	150	18	0.32

The first column shows the precise type of transition studied. The second column reports the estimate of β . A positive (negative) β implies that the downgrade intensity is increased (decreased) by a factor of $\exp(\beta)$ compared to the case of a previous upgrade. The standard deviation of the estimate is provided. n_1 is the total number of times we have observed a firm exposed to the given type of transition, i.e. the total number of censored or uncensored observations in the 'From' rating category. n_2 reports the number of actual transitions observed. p is the test statistic reported as a probability. So a test statistic of <0.01 is significant at least at the level of one percent. We see highly significant effects in virtually all categories.

Table 7
 Results shown are for the test of an effect of a previous upgrade on the intensity of an upgrade to a neighboring state

Ratings		$\hat{\beta}$	std($\hat{\beta}$)	n_1	n_2	p
From	To					
AA+	AAA	-0.106	0.525	149	15	0.84
AA	AA+	-0.011	0.545	314	14	0.98
AA-	AA	-0.132	0.268	490	56	0.62
A+	AA-	0.337	0.233	663	85	0.14
A	A+	0.449	0.190	842	116	0.02
A-	A	0.261	0.151	780	177	0.08
BBB+	A-	0.720	0.168	721	153	<0.01
BBB	BBB+	0.508	0.173	712	137	<0.01
BBB-	BBB	0.143	0.173	641	144	0.405
BB+	BBB-	0.535	0.174	513	152	<0.01
BB	BB+	-0.100	0.187	571	122	0.60
BB-	BB	0.1947	0.190	522	114	0.315
B+	BB-	0.667	0.214	575	90	<0.01
B	B+	0.560	0.277	437	63	0.05
B-	B	0.490	0.477	271	22	0.31
CCC+	B-	-6.150	25.7	194	17	0.24
CCC	CCC+	-7.280	45.1	150	6	0.25

The first column shows the precise type of transition studied. The second column reports the estimate of β . A positive (negative) β implies that the upgrade intensity is increased (decreased) by a factor of $\exp(\beta)$ compared to the case of a previous downgrade. The standard deviation of the estimate is provided. n_1 is the total number of times we have observed a firm exposed to the given type of transition, i.e. the total number of censored and uncensored observations in the 'From' category. n_2 reports the number of actual transitions observed. p is the test statistic reported as a probability. A test statistic of <0.01 is significant at least at the level of 1%.

consider only transitions to neighboring states. In Table 8 it is shown that in almost every case of downgrades we reject the hypothesis that the duration has no influence. In fact, we see that $\hat{\beta}$ is negative, meaning that $\exp(\hat{\beta}) < 1$. Thus, the intensity $\alpha_{ji}(t) = \alpha_{j0}(t) \exp(\beta Z(t)_{ji})$ is negatively affected by a change in the duration. i.e. the longer the firm has been in the rating class—the smaller the probability to downgrade is. In Table 9 we again find that the longer the firm has been in the rating class – the smaller is the probability to upgrade. Combining the two duration analyses we may conclude, that a firm with a given rating has a lower probability of changing its rating the more time it spends in its current state. A possible explanation of these effects could be the reluctance of rating agencies to change a rating by more than one notch at a time. If this is the case, then firms on the way (say) down through the rating system, will spend relatively short time in the intermediate states. Hence those that stay there a short amount of time are often firms on the way down.

The results are striking. One should however note, that they build upon an aggregate treatment of the firms in which we do not separate the industries to

Table 8

Results shown are for the test of an effect of the waiting time in the initial category listed under 'From' on the intensity of a downgrade to a neighboring state

Ratings		$\hat{\beta}$	std($\hat{\beta}$)	n_1	n_2	p
From	To					
AAA	AA+	-0.348	0.114	61	13	<0.01
AA+	AA	-0.405	0.067	149	65	<0.01
AA	AA-	-0.282	0.037	314	100	<0.01
AA-	A+	-0.380	0.041	490	162	<0.01
A+	A	-0.351	0.035	663	198	<0.01
A	A-	-0.547	0.046	842	193	<0.01
A-	BBB+	-0.628	0.064	780	161	<0.01
BBB+	BBB	-0.360	0.047	721	180	<0.01
BBB	BBB-	-0.555	0.056	712	140	<0.01
BBB-	BB+	-0.679	0.095	641	95	<0.01
BB+	BB	-0.708	0.134	513	59	<0.01
BB	BB-	-0.453	0.099	571	82	<0.01
BB-	B+	-0.621	0.110	522	74	<0.01
B+	B	-0.529	0.085	575	87	<0.01
B	B-	-0.683	0.155	437	43	<0.01
B-	CCC+	-0.902	0.216	271	28	<0.01
CCC+	CCC	-2.241	0.690	194	15	<0.01
CCC	CCC-	-0.704	0.259	150	18	<0.01

The first column shows the precise type of transition studied. The second column reports the estimate of β . A negative (positive) β implies that the downgrade intensity is decreased (increased) after a duration of t by a factor of $\exp(\beta t)$ compared to a case where the duration has no effect. The standard deviation of the estimate is provided. n_1 is the total number of times we have observed a firm exposed to the given type of transition, i.e. the total number of censored or uncensored observations in the 'From' rating category. n_2 reports the number of actual transitions observed. p is the test statistic reported as a probability. A test statistic of <0.01 is significant at least at the level of one percent. We see highly significant effects in virtually all categories.

which the firms belong. Industry effects are shown to be significant in Nickell et al. (2000) and Kavvathas (2000). An analysis run separately on our data for the largest subsample of financial institutions does produce some deviations. For example, the downgrade momentum is no longer present for the categories from BB-up to BBB-. An explanation for this could be that financial institutions are typically unable to compete when the rating goes into the speculative categories. It will then often be overtaken or merge to get consolidation and more competitive funding rates. Hence a downgrade is not likely to be followed by another downgrade in these categories. But since we do not have the identities of the firms, we are unable to check that. Also, our subsamples would be relatively small in the various other industry groups and it would be hard to get enough statistical power to test the hypotheses we are interested in for each type of transition.

Table 9
 Results shown are for the test of an effect of the waiting time in the initial category listed under 'From' on the intensity of an upgrade to a neighboring state

Ratings		$\hat{\beta}$	std($\hat{\beta}$)	n_1	n_2	p
From	To					
AA+	AAA	-0.416	0.132	149	15	<0.01
AA	AA+	-0.226	0.096	314	14	<0.01
AA-	AA	-0.360	0.072	490	56	<0.01
A+	AA-	-0.331	0.057	663	85	<0.01
A	A+	-0.329	0.049	842	116	<0.01
A-	A	-0.376	0.045	780	177	<0.01
BBB+	A-	-0.449	0.057	721	153	<0.01
BBB	BBB+	-0.266	0.043	712	137	<0.01
BBB-	BBB	-0.346	0.051	641	144	<0.01
BB+	BBB-	-0.532	0.075	513	152	<0.01
BB	BB+	-0.540	0.085	571	122	<0.01
BB-	BB	-0.537	0.084	522	114	<0.01
B+	BB-	-0.383	0.071	575	90	<0.01
B	B+	-0.359	0.100	437	63	<0.01
B-	B	-0.430	0.189	271	22	0.012
CCC+	B-	-0.507	0.247	194	17	0.016
CCC	CCC+	-0.934	0.631	150	6	0.032

The first column shows the precise type of transition studied. The second column reports the estimate of β . A negative (positive) β implies that the upgrade intensity is decreased (increased) after a duration of t by a factor of $\exp(\beta t)$ compared to a case where the duration has no effect. The standard deviation of the estimate is provided. n_1 is the total number of times we have observed a firm exposed to the given type of transition, i.e. the total number of censored or uncensored observations in the 'From' rating category. n_2 reports the number of actual transitions observed. p is the test statistic reported as a probability. A test statistic of <0.01 is significant at least at the level of 1%. We see highly significant effects in virtually all categories.

7. Conclusion

There are two main conclusions from this paper: First, we show the importance of estimating transition data based on the full story of rating transitions. Using either the maximum likelihood estimator in the homogeneous case or the non-parametric Aalen–Johansen estimator in the non-homogeneous case, the default probabilities over (say) one year are non-zero even for the highest rating category. This will affect risk measures both of the VaR type (for small quantiles) and measures of risk taking into account expected loss given that a certain threshold has been passed.

Second, we have presented a rigorous formulation of the notion of 'rating drift' – a type of non-Markovian behavior – in the process of ratings. The conclusion from analyzing the data set provided by Standard and Poor's, is that there seem to be strong non-Markov effects for downgrades in the aggregate data set, i.e. working on the entire population of firms without

differentiating for example between industries. Both the duration in a given state and the direction from which the state was reached has significant effects on the downgrade intensity. These effects would be consistent with a policy of taking a downgrade through a series of mild downgrades. However, the effect becomes less pronounced (but still significant in several categories) when looking for example at financial firms only. For upgrades, a significant effect of the previous move is only present in a few states, whereas the duration again seems to be a significant factor.

Acknowledgements

We are grateful to Reza Bahar and Standard and Poor's for providing the data set used in this paper. Helpful comments from Tom Daula, Bharath Kumar, Stephen Smith and, especially, Peter Vlaar are gratefully acknowledged. We also thank seminar participants at the 2000 Global Derivatives Conference in Paris, University of Vienna, The Mathematical Research Institute at Oberwolfach, Morgan Stanley Dean Witter and at the Federal Reserve Bank of Chicago/Journal of Banking and Finance conference: Risk Management in the Global Economy: Measurement, Management, and Macroeconomic Implications. All errors and opinions are of course our own. Lando acknowledges partial support by the Danish Natural and Social Science Research Councils.

Appendix A. Markov chains, estimation and testing

A.1. Non-homogeneous Markov chains and transition intensities

This appendix provides a brief outline of the elements we need from the theory of finite state space, non-homogeneous Markov chains. The finite state space we consider consists of the rating categories including a default state and in some cases the not rated category as well.

The evolution of a continuous-time non-homogeneous Markov chain η is described through transition matrices of the form $P(s, t)$ where the ij 'th element contains the transition probability between states i and j from time s to t , i.e.

$$p_{ij}(s, t) = \text{Prob}(\eta_t = j | \eta_s = i), \quad s < t.$$

Recall that the Markov property says that

$$\text{Prob}(\eta_t = j | \eta_{s_0} = i_0, \eta_{s_1} = i_1, \dots, \eta_{s_{n-1}} = i_{n-1}, \eta_s = i) = \text{Prob}(\eta_t = j | \eta_s = i)$$

whenever $s_0, s_{1,1}, \dots, s_{n-1} < s$. This imposes the familiar restriction on the transition matrices:

$$P(s, u) = P(s, t)P(t, u) \quad \text{for } s < t < u.$$

To understand the time-inhomogeneity, note that for a time-homogeneous Markov chain the transition probability matrix is a function of the distance between dates and not the dates themselves, i.e. in the homogeneous case, there would exist a family of transition matrices indexed by one parameter $(P(t))_{(t \geq 0)}$ and we could then write

$$P(u - s) = P(t - s)P(u - t) \quad \text{for } s < t < u$$

keeping track only of the distance between the time points and not their location in calendar time. For the processes we consider, it is always assumed that there exist *transition intensities* for each type of transition, i.e. that for each t and each pair of states i, j the limit

$$\lambda_{ij}(t) := \lim_{h \rightarrow 0^+} p_{ij}(t, t + h)/h$$

exists. It is typically more natural to formulate statistical hypotheses in terms of transition intensities instead of through the probabilities. When these limits exist for all transitions, then we also have for each row a sum of all the intensities

$$\lambda_i(t) := \sum_{j \neq i} \lambda_{ij}(t)$$

which, if multiplied by (a small) Δt , approximates the probability of leaving the state i within Δt . This function also gives us the duration distribution in state i in that

$$P(\eta_u = i \text{ for all } u \in (s, t] | \eta_s = i) = \exp \left(- \int_s^t \lambda_i(u) \, du \right).$$

Note that this probability is not the same as $p_{ii}(s, t)$ which gives the probability of being in i both at times s and t , but does not restrict the chain to staying in i in the period between s and t . It is only in the case of a homogeneous Markov chain that one gets a simple formula from all the transition probabilities from the intensities. If the intensities are constant (time-independent) then we have

$$P(t) := P(0, t) = \exp(At) := \sum_{k=0}^{\infty} \frac{A^k t^k}{k!},$$

where we have given the definition of the matrix exponential as an infinite sum. This is Eq. (2) and this equation gives us the maximum likelihood estimator

used in Section 3 for the transition probabilities as a function of the estimated intensities. In the non-homogeneous case, the link between intensities and transition probabilities can be described as follows, cf. Gill and Johansen (1990): define the cumulative intensity function for a transition from state i to j , as

$$A_{ij}(t) = \int_0^t \lambda_{ij}(s) ds,$$

$$A_{ii}(t) = - \sum_{j \neq i} A_{ij}(t).$$

The transition matrix for a non-homogeneous chain is given from these cumulative intensities as a limit:

$$P(s, t) = \Pi_{[s,t]}(I + dA) \equiv \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \Pi_i(I + A(t_i) - A(t_{i-1})),$$

where $s \leq t_1 \leq t_n \leq t$, and where the ij 'th element of the matrix $A(t)$ is just $A_{ij}(t)$. The Aalen–Johansen estimator directly uses this link by estimating the increments of the individual intensity functions. These increments are computed from observed transitions divided by the number of exposed firms. All the cumulative intensities together produce the estimator for the transition probabilities.

When we test hypotheses on the influence of factors, such as previous state, on transitions we analyze each transition intensity separately. The outline of how this is done is provided in the next section.

A.2. Statistical theory

The data records transitions between states, and we let N_{hji} denote the number of observed transitions from state h to state j by firm i . We assume in the basic model that the intensity of transition from state h to j is given as

$$N_{hji}(t) = \int_0^t \alpha_{hji}(u) Y_{hi}(u) du + M_{hji}(t),$$

where

$$Y_{hi}(t) = \begin{cases} 1 & \text{if firm } i \text{ is in state } h \text{ at time } t, \\ 0 & \text{otherwise,} \end{cases}$$

and M_{hji} is a martingale. The term $\int_0^t \alpha_{hji}(u) Y_{hi}(u) du$ is the cumulative intensity for the transitions of firm i between state h and state j . Such a transition can occur several times if a firm i reenters the state h several times. The censoring variable Y_{hi} sets the intensity of jumping away from state h for firm i equal to zero when the firm leaves the state either due to a migration to another rating or to a NR category. When the firm leaves for state j this can be viewed as a

censored observation of the transition from i to any state different from j and the transition to NR is viewed as independent censoring as well.

The semi-parametric specification, explained in the paper,

$$\alpha_{hji}(t) = \alpha_{hj0}(t) \exp(\beta_{hj}Z_i(t))$$

is used to estimate and test for influence of the covariate process Z_i on the transition intensity from state h to j . The base-line intensity $\alpha_{hj0}(t)$ is left unspecified, and therefore a full likelihood function cannot be used. Instead, the regression parameter β_{hj} is found by maximizing the so-called partial likelihood

$$L(\beta_{hj}) = \prod_t \prod_i \frac{\exp(\beta_{hj}Z_i(t))}{S_{hj}^0(\beta_{hj}, t)},$$

where

$$S_{hj}^0(\beta_{hj}, t) = \sum_{i=1}^n Y_{hi}(t) \exp(\beta_{hj}Z_i(t)).$$

Note that the maximization is done for each type of transition separately: we use here the fact that the partial likelihood of all the observed rating transitions used for estimating the regression parameters of all transition types actually factors into a product of partial likelihood functions – one for each transition type – which therefore can be maximized separately. The maximization is done by setting the (partial) score function of $L(\beta)$ equal to zero, and this score function can be shown to equal

$$\frac{\partial \log L(\beta_{hj})}{\partial \beta_{hj}} = \frac{\partial}{\partial \beta_{hj}} \int_0^T \sum_{h,j,i} dN_{hji}(t) \left[\beta'_{hj}Z_i(t) - \log(S_{hj}^{(0)}(\beta_{hj}, t)) \right].$$

The asymptotic results is based on the martingale property of this expression (viewed, of course, as a process in t) and it can be shown (see for example Andersen et al., 1993) that the estimator is asymptotically normal. Furthermore, the (partial) likelihood ratio test for testing $\widehat{\beta}_{hj} = \overline{\beta}_{hj}$ given as

$$LR = -2 \log \left(\frac{L(\widehat{\beta}_{hj})}{L(\overline{\beta}_{hj})} \right) = 2(\log L(\widehat{\beta}_{hj}) - \log L(\overline{\beta}_{hj}))$$

has an asymptotic chi-square distribution with one degree of freedom. All our tests are based on this result. Once the estimate of β is obtained, one may go back and obtain a Nelson–Aalen type estimator of the baseline intensity, but we will not be concerned with that in this paper. For more on this, see Andersen et al. (1993).

References

- Altman, E., Kao, D., 1992a. The implications of corporate bond rating drift. *Financial Analysts Journal*, 64–75.
- Altman, E., Kao, D., 1992b. Rating drift in high yield bonds. *The Journal of Fixed Income*, 15–20.
- Andersen, P., Borgan, Ø., Gill, R., Keiding, N., 1993. *Statistical Models Based on Counting Processes*. Springer, New York.
- Andersen, P., Hansen, L.S., Keiding, N., 1991. Non- and semiparametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, 153–167.
- Bangia, A., Diebold, F., Schuermann, T., 2000. Ratings migration and the business cycle, with applications to credit portfolio stress testing. Working paper. Oliver Wyman & Co and New York University.
- Blume, M., Lim, F., MacKinlay, A., 1998. The declining credit quality of US corporate debt: Myth or reality. *Journal of Finance* 53 (4), 1389–1413.
- Carty, L., 1997. Moody's rating migration and credit quality correlation, 1920–1996. Special comment, Moody's Investors Service, New York.
- Carty, L., Fons, J., 1993. Measuring changes in corporate credit quality. Moody's Special Report, November.
- Das, S., Tufano, P., 1996. Pricing credit-sensitive debt when interest rates, credit ratings and credit spreads are stochastic. *The Journal of Financial Engineering* 5, 161–198.
- Gill, R., Johansen, S., 1990. A survey of product-integration with a view towards applications in survival analysis. *The Annals of Statistics* 184, 1501–1555.
- Israel, R., Rosenthal, J., Wei, J., 1999. Finding generators for Markov chains via empirical transition matrices. Working paper, University of British Columbia and University of Toronto.
- Jarrow, R., Lando, D., Turnbull, S., 1997. A Markov model for the term structure of credit risk spreads. *Review of Financial Studies*, 481–523.
- Kavvathas, D., 2000. Estimating credit rating transition probabilities for corporate bonds. Working paper. Department of Economics. University of Chicago.
- Küchler, U., Sørensen, M., 1997. *Exponential Families of Stochastic Processes*. Springer, New York.
- Lando, D., 1998. On Cox processes and credit risky securities. *Review of Derivatives Research* 2 (2), 99–120.
- Lucas, D., Lonski, J., 1992. Changes in corporate credit quality 1970–1990. *The Journal of Fixed Income*, 7–14.
- Moody's, 1994. Corporate bond defaults and default rates 1970–1993. Moody's Special Report, January.
- Nickell, P., Perraudin, W., Varotto, S., 2000. Stability of ratings transitions. *Journal of Banking and Finance* 24 (1–2), 203.
- Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*.
- Skodeberg, T., 1998. Statistical analysis of rating transitions – a survival analytic approach. Master's Thesis, University of Copenhagen.