



## Confidence sets for continuous-time rating transition probabilities

Jens H.E. Christensen<sup>a</sup>, Ernst Hansen<sup>b</sup>, David Lando<sup>a,\*</sup>

<sup>a</sup> Department of Finance, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark

<sup>b</sup> Department of Applied Mathematics and Statistics, University of Copenhagen, Denmark

Available online 14 August 2004

---

### Abstract

This paper addresses the estimation of default probabilities and associated confidence sets with special focus on rare events. Research on rating transition data has documented a tendency for recently downgraded issuers to be at an increased risk of experiencing further downgrades compared to issuers that have held the same rating for a longer period of time. To capture this non-Markov effect we introduce a continuous-time hidden Markov chain model in which downgrades firms enter into a hidden, ‘excited’ state. Using data from Moody’s we estimate the parameters of the model, and conclude that both default probabilities and confidence sets are strongly influenced by the introduction of hidden excited states.

© 2004 Elsevier B.V. All rights reserved.

*JEL classification:* C41; G21; G28

*Keywords:* Rating transitions; Rating drift; Continuous-time Markov chains; Confidence sets; Default probabilities

---

### 1. Introduction

The Basel Committee on Banking Supervision has in its ‘New Capital Accord’ proposed a regulatory setup in which banks are allowed to base the capital

---

\* Corresponding author. Tel.: +45 3815 3613; fax: +45 3815 3600.

E-mail address: [dl.fi@cbs.dk](mailto:dl.fi@cbs.dk) (D. Lando).

requirement on their own internal rating systems and to use external rating systems as well. The increased reliance on rating systems as risk measurement devices has further increased the need for focusing on statistical analysis and validation methodology for rating systems. While the formal definitions of ratings by the major agencies do not employ a probability, or an interval of probabilities, in their definition of the various categories, any use of the ratings for risk management and capital allocation must rely on default probabilities from each rating category and on probabilities of transition between non-default categories. There are many statistical issues, of course, in estimating such probabilities and in assessing the precision of the estimates.

In [Lando and Skødeberg \(2002\)](#), it is shown that using a continuous-time analysis of the rating transition data enables us to meaningfully estimate probabilities of rare transitions, even if the rare transitions are not actually observed in our data set. When using classical ‘multinomial’ techniques, as those of [Carty and Fons \(1993\)](#) and [Carty \(1997\)](#), this is only possible in the limit where the data are observed at all dates and the transition matrices are estimated over very short periods (daily). The information gain in using the full information of exact transition dates is important. The continuous-time approach using generator matrices is mainly a question of ease of formulation and application.

However, the computation of the one-year transition probability estimates from a generator (or by taking powers of – say – daily transition matrices) implicitly assumes that the underlying process is Markov. There is overwhelming evidence that the rating evolutions are non-Markov, and one of the most well-documented facts is that there is a ‘downward momentum’ effect in ratings. This means that firms that are downgraded into a class have a higher probability of experiencing a further downgrade from this class than the companies that were not downgraded into that same class. Similar effects are documented with respect to upward movements.

In this paper we expand the state space of the rating process to take into account such non-Markov effects. We estimate a Markovian model in which firms which are downgraded into certain categories enter into an ‘excited state’ in which they stay for a stochastic amount of time. For example, a firm downgraded into Baa, will be assumed to enter into a latent state Baa\* and (if no further rating action is observed) to migrate into the ‘normal’ state Baa after a random amount of time. We estimate the time it takes for the latent process to make the unobserved migration from the excited state into the normal state, but more importantly, we obtain estimates for transition probabilities from excited as well as normal states. We compare the estimates obtained in this larger model with those obtained by using a ‘basic model’ based on the observed rating classes only. Our data set does not include outlooks and watchlists which Moody’s uses to convey information in addition to the rating on the likely direction of future rating movements. With access to these data, it would be natural to let the excited states correspond to a ‘negative outlook’. At the completion of this study, the study [Hamilton and Cantor \(2004\)](#) appeared which documents that outlooks have highly significant effects on migration probabilities.

A second focus of our paper is the use of a bootstrap<sup>1</sup> procedure to obtain confidence sets for the default probability estimates, and again compare the confidence sets obtained in the enlarged state space model with those obtained by using only the observed ratings. Our main focus is again on the rare events.

Our main conclusions are as follows. First of all, we confirm that the description of states as ‘excited’ is well chosen. Firms in the excited states have higher default probabilities over a one-year horizon than firms in the normal state. Second, using the extended model increases default probabilities from normal states compared to what is obtained in the basic model. The intuition is, that the typical path to default is through a sequence of downgrades. In the extended model, successive downgrades are more likely because the entrance into an excited state increases the probability of a further downgrade to yet another excited state, etc. As in Lando and Skødeberg (2002) the fact that we use ‘continuously’ observed (i.e. daily) rating transitions allows us to estimate probabilities of rare events (such as default from Aaa, Aa or A) which are not observed in the data set. However, we show here that the modification of the estimates of default for such rare transitions needs to be larger than that proposed in Lando and Skødeberg (2002). Finally, the confidence sets for rating transition probabilities also become wider in the extended model, partly as a consequence of the increased number of parameters.

The improved understanding of transition probability variability has consequences for a number of issues in credit risk management and in the analysis of credit risk models in general. First, removing the zeros from all events in the matrix of estimated transition probabilities and supplying confidence bands gives us a concrete method of assessing, for example, the proposal put forward by the Basel Committee, of imposing a minimum probability of 0.03% in these cases. As we will see, a proper use of the full information in the data set gives a better impression of the appropriateness of such a limit. Both Nickell et al. (2000) and Höse et al. (2002) contain estimates of standard deviations and confidence sets, but since they are based on multinomial type estimators, they cannot assign meaningful confidence sets to probabilities of rare events where an estimate of standard deviation is either not well-defined or the asymptotic methods are not applicable.

Second, the comparison of actual default probabilities for high rating categories and the default probabilities implied by spreads in corporate bond prices relies on a point estimate of the default probability for a given rating category. If this probability is significantly changed when one takes into account the non-Markovian behavior of the rating transitions, especially for short maturities, then the actual default probability is capable of explaining a larger part of the observed credit spreads.

Our focus in this paper is on non-Markov effects that are in a sense ‘internal’ to the rating process, i.e. which can be characterized by the history of the rating process

---

<sup>1</sup> For an introduction to the bootstrap, see Efron (1982). Our method is an example of a *parametric* bootstrap.

itself. The study of such effects is performed for example in Altman and Kao (1992b), Altman and Kao (1992c), Altman and Kao (1992a), Hamilton and Cantor (2004), Carty (1997), Kavvathas (2000), Lucas and Lonski (1992) and Lando and Skødeberg (2002). A study of particular relevance to our study is the non-parametric estimation of Fledelius et al. (2004). By means of non-parametric kernel smoothing techniques they show that conditional on a previous downgrade there is a temporary increased intensity of being further downgraded. And the other way round, a previous upgrade leads to a temporary increased intensity for being further upgraded. This effect is observed in all four rating categories they consider, and whether they condition on previous upgrades or downgrades the effect seems to disappear after some 20–30 months if no further rating transitions are observed. This is consistent with the stated objective of Moody's of avoiding rating reversals, see Cantor (2001). These findings lend support to the extended state model of this paper. However, for practical reasons even our model does not allow for all heterogeneities. First of all, when we consider downgrade intensities, we only distinguish between firms that were downgraded into the current state. We do not distinguish between those for which we have no prior movements available and those where the current state was reached through an upgrade. Based on the study of Fledelius et al. (2004), the grouping of these two types is justified for the purpose of estimating downgrade intensities. We also keep this grouping when we study upgrade intensities which is not consistent with the analysis of Fledelius et al. (2004). However, since our focus is on the probability of default, we are mainly concerned with the effect of incorporating a temporary increased intensity of further downgrades upon a previous downgrade, since the other effect will have only second order effects. The dominant contribution to the default probability is the probability of taking a series of downgrades into default which is not interrupted by temporary upgrades. As a consequence we leave out of consideration the non-homogeneity arising from upgrades. Due to data limitations, we also do not include excited states for the Aaa, Aa and A categories. The number of firms downgraded from Aaa into Aa, for example, is so small that we cannot obtain a meaningful estimate of the transition away from an excited Aa state.

Non-Markov effects or non-time homogeneity may of course also arise from business cycle effects documented for example in Nickell et al. (2000), Bangia et al. (2002) and Kavvathas (2000), or possible changes in rating practices, as argued in Blume et al. (1998). We perform here an analysis which is conditional on a particular phase in the business cycle. We consider estimates from a stable period (1995–1999) and from a volatile period (1987–1991). While there may still be cyclical effects within these small periods, the homogeneity assumption is much more plausible over these time periods. This conditional analysis also limits the problems with correlation between rating migrations arising from common dependence of exogenous variables. It is well known from intensity-based models of default, that to induce significant correlation of default events through common variation in intensities, one must have extremely volatile intensities, see for example Duffie and Gârleanu (2001) for evidence of this. We see no evidence even close to such fluctuations of transition intensities over the five year periods we consider. Correlation of defaults might also occur due to 'domino effects' or 'parent–subsidiary' relations but such events are very rare.

The outline of the paper is as follows: Section 2 describes our data and some of the data cleaning issues one faces when looking at the data set. Section 3 recalls the discrete-time, multinomial estimation of the transition parameters based on discretely observed ratings and contrasts this with the continuous-time case. Furthermore, it discusses the problems with confidence sets for rare events in the multinomial model. In Section 4, we describe in detail the continuous-time Markov chain model with extended state space, we present an estimator for this model, and we discuss the problems of determining the initial state of the issuers at the beginning of the estimation period. In Section 5 we give a description of the bootstrap experiment we apply to obtain confidence sets. Section 6 contains our results and Section 7 concludes the paper. Appendix A contains a description of the technique applied when estimating the extended model.

## 2. Data description

The rating transition histories used for this study are taken from the complete ‘Moody’s Corporate Bond Default Database’, that is the edition containing complete issuer histories since 1970. We consider for this study only issuers domiciled in the United States. Moody’s rate every single debt issue of the issuers in the database but we exclude all but the senior unsecured issues. This leaves us with 3405 issuers with 28,177 registered debt issues. Including all the rating changes for every issue we reach a total of 90,203 rating change observations, ranging from just one observation for some small issuers up to General Motors Acceptance Corporation, the finance subsidiary of GM, with 3670 observations on 1451 debt issues since 1973.

Our first task is to produce a single rating history for the senior unsecured debt of each company – a task which is not completely simple. A first step is to eliminate all irrelevant ‘withdrawn rating’ observations. A ‘withdrawn rating’ is defined as irrelevant if it is not the last observation of the issuer implying that the issuer is still present in the database.<sup>2</sup> On the other hand if a ‘withdrawn rating’ is genuinely the last observation, and no other issues are observed after the withdrawal, we leave out of consideration this issuer from that date on and treat the issuer as ‘right censored’. The issuer is also right censored if alive at the final date of the observation period which is the 9th of January 2002.

Having corrected for withdrawn ratings, we should be able to obtain the senior unsecured rating for the company by looking at any senior unsecured issue which has not matured or has not been repaid. There are, however, 62 exceptions to this principle which we handle individually. The typical case is that a single debt issue has some special covenants, is traded in the Euro-market, or possesses similar peculiarities, which makes it fair to neglect that specific issue. There were a few cases where we could not find a reason for the different ratings, but in these cases it was

---

<sup>2</sup> We can draw this conclusion because we work with the complete database, where all loans matured or called in the past have as final observation a ‘withdrawn rating’.

easy to assign a rating by looking at the rating of the majority of issues. Having done the above cleaning of our data set, we are left with 13,390 rating change observations.

The next critical step is to get a proper definition of default. Recall, that Moody's does not use a default category as such, but does record a default date in the data base. The lower categories (from B and downward) may include firms in default, and the rating then measures the severity of the default. The problem is that to measure transition rates from non-default categories to default, we must know whether the firm migrated from (say) B to default or whether the assignment of the rating B was already part of the default scenario. As our primary focus is the estimate of the probability of default for the various categories, it has been essential to make sure that these cases were treated in the right way. Hence we decided to look at each of the 305 recorded defaults manually. These defaults are recorded by Moody's in their 'Default Master Database', and whenever reasonable, these default dates are used to define a default date. In this case, all rating observations until the date of default have typically been left unchanged. However, if a transition from say B1 to Caa occurs a few days (up to a week) before the default date, we interpret this event as a B1-issuer jumping directly to default. It is clear in cases like this that the rating Caa has reflected the imminent default and that only legal issues have made the default date different from the date at which that lower rating was assigned. There is some arbitrariness in this choice and it means that one should be very careful interpreting the estimated default probabilities for the very low rated firms.

Rating changes observed after the date of default are eliminated, *unless* the new ratings reach the B1/B2-level or higher *and* the ratings are related to debt issued after the time of default. In these cases we have treated the last rating observations after the recovery to the higher rating as related to a new (or 'refreshed') independent issuer.<sup>3</sup>

Finally, there are 17 incidences of issuers with two default dates attached to their name in the database, where the first might refer to a 'distressed exchange' and the second to the date of a 'chapter 11 action'. By cross-checking all the information in the database and looking at the rating observations before and after both default dates, it has in most cases been possible clearly to determine whether we have just one issuer defaulting once or it is a case which according to our rule set up above should be treated as two defaults of independent issuers. After this procedure, we have now come down to 9991 observations of rating changes distributed among 3446 issuers.<sup>4,5</sup>

---

<sup>3</sup> When rating a company, Moody's give weight to estimating not just the probability of default but also the expected recovery of principal given default, which are combined into one single rating. Most post-default rating changes are therefore not real changes but mere reflections of the improved or deteriorated expectation of recovery, why they are of no interest for our purposes.

<sup>4</sup> The number has gone up from 3405 due to our procedure of introducing new issuers, if the post-default information is judged to be real rating changes.

<sup>5</sup> Amongst the remaining issuers some might be affiliates of others. However, as remarked by Lucas and Lonski (1992), affiliate companies need not follow the same rating path as the parent company, so we will not pursue this issue any further.

Since the introduction of additional notches (Aa1, Aa2, etc.) in the beginning of the 80s there has been a total number of 21 categories in the Moody's system, and adding a default and a 'withdrawn rating' category we would have to work with 23 categories in all. Because we treat withdrawals as censored observations and the 'default' state is viewed as absorbing (or, at least, the recovery time is not analyzed), there are  $21 \times 22 = 462$  rating transition probabilities to estimate in the full system. This model is hard to estimate with the sample we have at our disposal and it becomes impossible with the addition of latent states. Hence we have chosen to reduce the number of categories to a total of 8 in the usual way: Aaa is kept as an independent category. Aa1, Aa2, and Aa3 are merged into one single category Aa. The same procedure is applied to A, Baa, Ba, and B. For the Caa-category we merge Caa, Caa1, Caa2, Caa3, Ca, and C into one category. Having done this simplification we only have to estimate 56 rating transition probabilities in the standard model, a much more reasonable number.

### 3. Discretely observed vs. continuously observed ratings

The advantage of using continuously observed data when estimating probabilities of unobserved or rare transitions was observed in Lando and Skødeberg (2002). In this paper we focus on two extensions of that approach. First, to correct for observed non-Markov effects, we expand the state space to include latent rating categories for down-graded firms. We investigate, whether this adjustment has serious consequences for the estimates. Second, we use a bootstrap method to obtain confidence sets for the probabilities of default, something which is not possible using traditional methods. Before we turn to these models, we briefly summarize the difference between the discrete time, 'multinomial' approach and the continuous-time approach. To motivate the bootstrap procedure, we also show why the traditional multinomial method is less suitable for obtaining confidence sets.

Estimation in a discrete-time Markov chain is based on the fact that the transitions away from a given state  $i$  can be viewed as a multinomial experiment. Let  $n_i(t)$  denote the number of firms recorded to be in state  $i$  at the beginning of year  $t$ . Disregarding withdrawn ratings for the moment, each of the firms may be in one of  $K$  states at the beginning of year  $t+1$ . Let  $n_{ij}(t)$  denote the number of firms with rating  $i$  at date  $t$  which are in state  $j$  at time  $t+1$ . The estimate of the one-year transition probability at date  $t$  is then

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t)}.$$

If the rating process is viewed as a time-homogeneous Markov chain which we observe over time, then the transitions away from a state can be viewed as independent multinomial experiments. This allows us to in essence collect all the observations over different years into one large data set. More precisely, the maximum-likelihood estimator for the time-independent transition probability becomes

$$\hat{P}_{ij} = \frac{\sum_{t=0}^{T-1} n_{ij}(t)}{\sum_{t=0}^{T-1} n_i(t)}, \quad (1)$$

where  $T$  is the number of years for which we have observations. In practice there are rating withdrawals, and typically this is handled by elimination of the observation for the year in which the withdrawal occurs. This procedure depends on the withdrawal being ‘non-informative’, an assumption which we make throughout, both in the discrete- and the continuous-time setting. In the special (but unlikely) case where the number of firms in a rating category stays the same (i.e. the inflow is equal to the outflow), the estimator for the transition probabilities is the average of the one-year transition probability matrices. But this average only serves as an approximation when the number of firms in a given rating category changes from year to year. The estimator above correctly weighs the information according to the number of firms observed each year.

Time-homogeneity is used in the estimator (1) to aggregate transitions and exposures over different time periods. Without the homogeneity assumption, we may estimate the probability of a transition between two categories over a particular time-period  $t$  to  $T$  as

$$\hat{p}_{ij}(t, T) = \frac{n_{ij}(t, T)}{n_i(t)}, \quad (2)$$

where  $n_{ij}(t, T)$  is the observed number of transitions from  $i$  to  $j$  over that particular time period. This is a so-called cohort estimator, and this is also a multinomial type estimator and it can be interpreted as such even without a Markov assumption. However, both estimators are 0 when no transitions from  $i$  to  $j$  occur, and both have problems with obtaining meaningful confidence sets in this case, as we shall see below.

Estimation based on continuous observations relies on estimating the generator matrix of a time-homogeneous Markov chain. Let  $P(t)$  denote the transition probability matrix of a continuous-time Markov chain with finite state space  $\{1, \dots, K\}$  so that the  $ij$ th element of this matrix is

$$P_{ij}(t) = P(\eta_t = j | \eta_0 = i).$$

The generator  $A$  is a  $K \times K$  matrix for which

$$P(t) = \exp(At) \quad \text{for all } t \geq 0,$$

where

$$\exp(At) \equiv \sum_{k=0}^{\infty} \frac{(At)^k}{k!}.$$

The diagonal element of  $A$  we write  $-\lambda_i$  where



$$\lambda_i = \sum_{j \neq i} \lambda_{ij}, \quad \lambda_{ij} \geq 0 \quad \text{for all } i \neq j,$$

and from this we note that the rows of a generator sum to zero.

The maximum-likelihood estimator of  $\lambda_{ij}$  based on observing realizations of the chain from time 0 to  $T$  is <sup>6</sup>

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

where  $N_{ij}(T)$  counts the total number of transitions from  $i$  to  $j$  in the time interval and  $Y_i(s)$  is the number of firms in state  $i$  at time  $s$ . Hence the maximum-likelihood estimator for the one-period transition matrix  $P(1)$  is

$$\hat{P}(1) = \exp(\hat{A}).$$

Lando and Skødeberg (2002) analyze the importance of using this estimator compared to the discrete-time estimator based on annual observations. Briefly summarized, the advantages of the continuous-time estimator are:

1. We obtain non-zero estimates for probabilities of events which the multinomial method estimates to zero.
2. We obtain estimates for the generator from which transition probabilities for arbitrary time horizons can be obtained without having to worry about finding roots of discrete-time transition matrices.
3. The estimator uses all available information in the data set by using information from firms up until the date of a withdrawn rating and by including information of a firm even when it enters a new state. In the multinomial estimator, we cannot distinguish the exact date within the year that a firm changed its rating.

However, the interpretation relies on a Markov-assumption, and this is not satisfied for the process of observed ratings. Hence it is natural to look for representations for which the Markov assumption is more palatable. This we will do in the next section. Now, we will consider the problem of obtaining confidence sets using the multinomial type estimators.

To construct confidence bands for default probabilities in the multinomial model we will use the following simple binomial procedure. Consider a binomial random variable  $X \sim b(\theta, N)$  where  $\theta$  is the probability of failure. Given that we observe  $\tilde{X} = 0$ , we may ask which is the largest  $\theta$  “consistent” with this observation, i.e. for a given level  $\alpha$ , what is the smallest  $\theta$  we can reject based on the observation of  $\tilde{X} = 0$ . This  $\theta$  of course depends on  $N$  since more observations give us more power to reject. The smallest  $\theta$  is the solution to the equation

$$(1 - \theta)^N = \alpha,$$

i.e. denoting the solution  $\theta^{\max}(N, \alpha)$  we find

---

<sup>6</sup> Of course,  $\hat{\lambda}_i = \sum_{j \neq i} \hat{\lambda}_{ij}$ .

$$\theta^{\max}(N, \alpha) = 1 - \alpha^{1/N}.$$

In Fig. 1 we have illustrated this function as a function of  $N$  for  $\alpha=0.01$  and  $\alpha=0.05$ .

In a multinomial analysis we could use this procedure as follows: For a given rating category  $i$  and a given number of firms  $N_i$  in this category, consider the binomial distribution obtained by considering default/no default as the only possible outcomes. We could then obtain a confidence band for the default probability  $\theta_i$  in the cases where we observe no defaults by following the above procedure. To be precise, if we have  $T$  years of observations, we consider the quantities  $n_i(t)$  = “number of firms in category  $i$  at the beginning of year  $t$  which have a rating or a default recorded at the beginning of year  $t+1$  as well.” Let  $N_i = \sum_{t=0}^{T-1} n_i(t)$ . This is the size of the binomial vector to be used. For our data set this produces the confidence bands shown in Table 1 for the top categories where in fact no defaults are observed. We may of course also assign confidence sets to the remaining default probabilities where transitions to default are observed. For this, assume that we have observed  $\tilde{X}_i$  defaults over the period of  $T$  years and that the total number of issuers having started in category  $i$  is defined as  $N_i$  above. This means that  $X_i$  is  $b(\theta_i, N_i)$ . Now a two-sided confidence set for the true parameter  $\theta_i$  for a  $(1 - \alpha)$  level of significance is calculated in the following way. Let  $\theta_i^{\min}$  denote the lower end of the interval. This must be a value so low, that with probability  $1 - \frac{\alpha}{2}$  we will not be able to experience as many defaults as  $\tilde{X}_i$ , that is  $\theta_i^{\min}$  must solve the following equation:

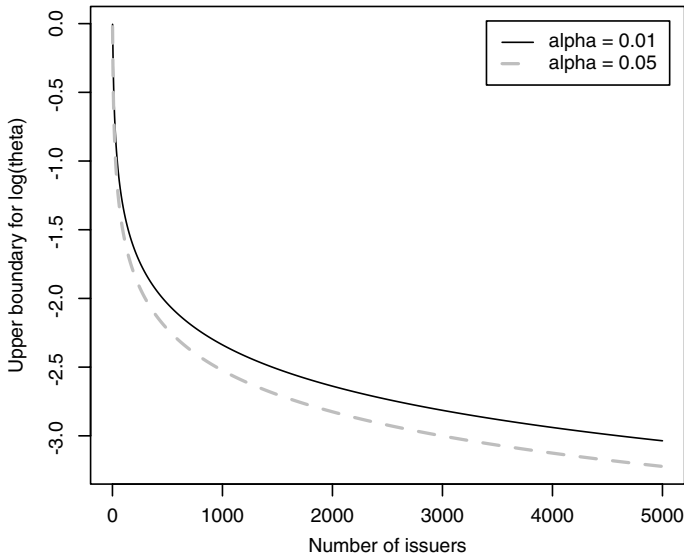


Fig. 1. 95% and 99% upper confidence boundaries of a default probability estimate given an observation of  $\tilde{X} = 0$  and viewed as a function of the number of issuers.

Table 1

Upper 95% and 99% boundaries of  $\theta_i$  for the Aaa, Aa, and A categories based on the exposure over the period 1995 through 1999 measured in company years as calculated in the denominator of Eq. (1)

	$N_i$	$\theta_i^{\max}(N_i, 0.05)$	$\theta_i^{\max}(N_i, 0.01)$
Aaa	189	0.015725	0.024072
Aa	635	0.004707	0.007226
A	2277	0.001315	0.002020

All six values correspond to values which could in principle be seen in Fig. 1.

$$P(X_i \leq \tilde{X}_i - 1 | \theta_i = \theta_i^{\min}) = 1 - \frac{\alpha}{2}.$$

On the other hand let  $\theta_i^{\max}$  denote the upper end of the interval. To find this we must let  $\theta_i$  take on a value so high, that it will only just be possible with probability  $\frac{\alpha}{2}$  to observe as few defaults as  $\tilde{X}_i$ :

$$P(X_i \leq \tilde{X}_i | \theta_i = \theta_i^{\max}) = \frac{\alpha}{2}.$$

Solving this for the lower categories gives us the results shown in Table 2.

There are at least two important problems with this procedure. The first problem is that the confidence sets are very wide. This will become more transparent later as we improve our estimation methodology. Loosely speaking, since the estimator based on discrete-time observations is inefficient when we have access to continuous-time data, the confidence sets based on this estimation methodology become wide. More importantly, the confidence sets for the zero-event categories depend on  $N_i$  and  $\alpha$  only. Hence if there are fewer firms in category Aaa than in Aa, the confidence set will be wider – something which is counterintuitive when we consider the dynamics of the rating transitions. This problem can be solved in a continuous-time setting where the entire information content of the data set is used, but the cost is loss of an analytical expression for confidence bounds. Even asymptotic methods are of no use here since our main focus is precisely the cases with few observed transitions.

A key concern is to understand the role of the Markov assumption in our estimation procedure. In the next section, we therefore turn to the specification of our

Table 2

Two-sided confidence intervals for  $\theta$  for the Baa, Ba, B, and Caa categories based on a binomial method and the sample data from 1995 through 1999

	$N_i$	$\tilde{X}_i$	$\theta_i^{\min}(N_i, 0.05)$	$\theta_i^{\max}(N_i, 0.05)$	$\theta_i^{\min}(N_i, 0.01)$	$\theta_i^{\max}(N_i, 0.01)$
Baa	2091	1	0.000012	0.002662	2.39e–06	0.003548
Ba	880	1	0.000029	0.006315	5.70e–06	0.008413
B	1132	42	0.026869	0.049823	0.024163	0.054081
Caa	217	29	0.091361	0.186261	0.080455	0.203524

model with an extended state space which allows us to capture downward momentum effects and hence to compare estimates in the extended model with those obtained using just the observed ratings.

#### 4. The extended state space model

As noted in the introduction, there is ample evidence that the rating process has ‘momentum’. However, as indicated in Fledelius et al. (2004), there is evidence that the above-average probability for downgraded firms of being further downgraded is temporary. This is also consistent with results on the duration of outlooks shown in Hamilton and Cantor (2004). Accordingly, we introduce latent states that we will refer to as *excited states*. They are meant to capture a heterogeneity in the population of the rating categories, possibly caused by a preference of the rating agencies for applying a sequence of single-notch downgrades instead of one multi-notch downgrade.

The extended model we have chosen to work with adds the following four excited states to the state space: Baa\*, Ba\*, B\*, and Caa\*. Due to data limitations we do not add any excited states to the Aaa, Aa, and A categories.<sup>7</sup> In total this gives us latent background processes with an extended state space given by

$$E = \{Aaa, Aa, A, Baa^*, Baa, Ba^*, Ba, B^*, B, Caa^*, Caa, D\}$$

while the actually observed rating processes still have a set of ratings equal to

$$A = \{Aaa, Aa, A, Baa, Ba, B, Caa, D\}.$$

The structure of the allowed transitions of the background process is as follows. The excited states receive firms which in the data set are observed as having a downgrade into the corresponding rating category. For example, when a company is downgraded into Baa, we assign it in our extended model to the state Baa\*. Once a company reaches an excited state, it can subsequently make four types of transitions. First, it can migrate from the excited state to the corresponding normal or non-excited state, i.e. from Baa\* to Baa. The interpretation is that after a random amount of time it is no longer at an increased risk of a further downgrade. Second, it can be upgraded to a non-excited state, i.e. for example from Baa\* to A. Third, it can be downgraded into an excited state, i.e. for example from Baa to Ba\*, and finally it can default.

The addition of the excited states has the consequence that many transitions become unobserved. We do observe the entrance of a firm into the excited states, since by definition a company which is downgraded enters into this state. However, we do not observe a transition from the excited state to the corresponding non-excited state, and as a consequence, if a company reached its current rating through a downgrade, we do not know whether it leaves that category from the excited or the non-

<sup>7</sup> For the Aaa, Aa, and A categories there are too few rating transitions, and the issuers stay too long in the new categories even upon a downgrade for this type of analysis to make sense.

excited state. If it reached its current state through an upgrade or if no prior move is observed, then it is by definition in a non-excited state, and the transition will be known to occur from this state.

Hence all transitions away from a rating class by firms which were not downgraded into that class, will be known to have occurred and recorded as a move away from the non-excited rating category. However, all transitions away from a rating class by companies which were downgraded into that class are latent in our model since we do not know if they occur from the excited or non-excited version of the class. In particular, since all firms in excited categories were downgraded into their current rating, we do not have any direct observations of moves away from excited categories.

The decision on which categories we split in excited and non-excited states is driven in part by whether we have enough data to estimate transition probabilities of the excited states. The choice has consequences for our numerical results. Since almost all defaults occur after migrating down through the system to the lowest categories, the probability of a default for a better rated company within a fixed time period, will essentially be a reflection of the probability of moving down to these categories. And this probability will in turn be increased if there are many composite ratings along the way. If we had enough data to model an excited state in A, default probabilities from AA would be likely to increase.

Summing up, the purpose is to model a possible tendency for downgraded issuers to have a temporary above-average probability of further downgrades as documented for example by [Fledelius et al. \(2004\)](#). This effect would increase the estimated downgrade probabilities of the excited categories as compared with the non-excited categories. Ultimately, this should turn up as significantly larger estimated probabilities of default for the excited categories, i.e. we would be able to confirm that recently downgraded issuers have a higher temporary PD than issuers well-established in the same rating category.

## **5. The data sets and the bootstrap experiment**

We have fixed two time windows, the first being from January 1, 1987 to December 31, 1991, the second being January 1, 1995 to December 31, 1999. We consider it reasonable to assume that the processes involved behave in a time-homogeneous way under such relatively short time horizons. Of course, considerable changes seem to occur between the two time windows and the separation between the windows is of the same magnitude as the duration of each of them. Therefore the assumption of time-homogeneity may seem questionable. We maintain it because the variations within the two periods is much smaller than that between the two periods.

The study population corresponding to a time window, consists of all issuers with an annotation at the onset of the time window, and all issuers with a first annotation within the time window. It is assumed that the first annotation of the newcomers corresponds to a non-excited state. The issue is somewhat more complicated for the issuers alive at the onset: they might or they might not be in an excited state.

As we want to condition on the starting state, this is a nuisance. We have resolved the problem by following these issuers backward in time until their last change of rating. At that moment in time (known as the issuer's onset), the issuer was in a uniquely defined state. And so we follow the study population in a somewhat larger period of time, than the time window indicates.

It is well known in survival analysis, that defining a study population as the persons alive at a specific day, and then following these persons backward in time, is problematic. The phenomenon is known as left truncation, and it is intimately related to the so-called waiting time paradox. If care is not taken, the waiting time paradox introduces a bias in the estimation of the survival. In short: the study population is missing persons with short survival times, because these persons tend to have died before the onset of the time window. We consider the problems related to left truncation to be quite mild in this context. The point is, that we have multiple events per issuer (at least for most of them), and left truncation only applies to the first of these events.

Not all issuers in the study population are observed throughout the time window. As we have mentioned, some are late starters. Others are censored in the window meaning that they have a withdrawn rating, usually because their bonds have expired or have been called. As is common practice, we treat such instances of censoring as non-informative. And finally some experience a default – and we use no more information after the default date for those issuers unless they reappear as a refreshed firm in which case they are treated as a new issuer.

The issuers in the study population are followed from their individual onset until explicit withdrawal or the end of the time window (implicit right censoring), whichever comes first. Finally, the observed rating history of the various issuers are considered to be independent realizations of the hidden Markov model.

In order to obtain confidence sets for the estimated transition probabilities, in particular the default probabilities, we make the following (parametric) bootstrap experiment. We simulate  $N=500$  fake datasets for each time window. Each dataset is generated as follows: For each of the issuers in the real dataset, we have a starting state and an observation period (from the individual onset to the implicit or explicit censoring, whichever comes first). A fake dataset has the same number of issuers as the real dataset, and each fake issuer is paired with the corresponding real issuers, meaning that it has the same starting state and the same observation period. The issuer's history background Markov process is simulated using the estimated transition structure for the first and the last period, respectively, and translated into a history of observed rating transitions.

The hidden Markov chain model is then reestimated, using each fake dataset. From the estimated transition structure we calculate the one-year default probability for each true state. And this vector of one-year default probabilities is considered the outcome of the analysis of a single fake dataset.

We end up with  $N$  of these vectors. For each coordinate (that is, each true state of the background process), we have  $N$  one-year default probabilities. In the graphs to follow, a kernel estimator of the densities of these default probabilities are depicted together with boundaries, indicating the middle 95% of the observations. We actu-

ally know the true value of the default probabilities behind the simulation experiment, this is also added to the pictures. The bootstrap experiment of the standard model without the extended state space is undertaken using a similar method. To validate that 500 is a reasonable number of simulations, we tried decreasing the number to 100 and repeated this five times. The estimators obtained using the smaller samples showed almost no variability and they were extremely close to the results from the larger sample. Since the simulations are very time-consuming we did not want to increase beyond 500 simulations and the experiment above suggested we could perhaps have done with less.

## 6. Results

The technical issue of estimating this hidden Markov chain model is outlined in [Appendix A](#). We have chosen to estimate the generators over two 5-year periods. The period between January 1, 1995 and December 31, 1999 had a relatively stable macroeconomic climate whereas the other period from January 1, 1987 to December 31, 1991 included the savings and loan crisis and a macroeconomic recession. The main results of our analysis are presented in [Tables 4 and 5](#). All other tables and the figures elaborate on these results ([Tables 6 and 7](#)). Instead of estimating the generator directly we have used an approach based on a discretization with period length equal to one day. This increases the mathematical tractability of the estimation of the latent variables in the extended state space model and it is a harmless approximation in the standard case since the approximation of the transition probability  $P(\Delta t)$  from the generator  $A$

$$P(\Delta t) \approx I + A\Delta t$$

is very accurate when  $\Delta t$  is just one day. This means that the estimated one-year transition probability matrices obtained by raising the one-day matrix to the power 365, is very close to the matrix exponential of the generator.

Table 3

The 95% confidence sets for the one-year probabilities of default for the binomial approach and the bootstrap simulation of the standard model based on the estimated transition probability matrices for the period 1987–1991

Rating	Binomial approach		Bootstrap standard model	
	Estimate	Conf. set	Estimate	Conf. set
Aaa	0	[0, 0.00896]	3.57e-07	[6.7e-09, 9.8e-07]
Aa	0	[0, 0.00363]	1.59e-05	[9.1e-07, 4.9e-05]
A	0	[0, 0.00156]	5.95e-05	[3.1e-05, 0.0001]
Baa	0.00257	[0.000530, 0.00748]	0.000513	[0.00029, 0.00082]
Ba	0.0331	[0.0178, 0.0559]	0.0122	[0.00062, 0.018]
B	0.0706	[0.0481, 0.0992]	0.0482	[0.035, 0.063]
Caa	0.367	[0.234, 0.517]	0.435	[0.34, 0.55]

Table 4

The estimated one-year default probabilities and their bootstrapped 95% confidence sets for both the extended and the standard model based on the estimated transition probability matrices over the period 1987–1991

Rating	Bootstrap extended model		Bootstrap standard model	
	Estimate	Conf. set	Estimate	Conf. set
Aaa	7.40e-07	[1.3e-08, 2.8e-06]	3.57e-07	[6.7e-09, 9.8e-07]
Aa	2.96e-05	[1.8e-06, 9.2e-05]	1.59e-05	[9.1e-07, 4.9e-05]
A	0.000106	[5.1e-05, 0.0002]	5.95e-05	[3.1e-05, 0.0001]
Baa*	0.00115	[0.00054, 0.0024]	N.A.	N.A.
Baa	0.000676	[0.00035, 0.0013]	0.000513	[0.00029, 0.00082]
Ba*	0.0193	[0.0086, 0.037]	N.A.	N.A.
Ba	0.00808	[0.004, 0.02]	0.0122	[0.00062, 0.018]
B*	0.0910	[0.055, 0.13]	N.A.	N.A.
B	0.0334	[0.025, 0.056]	0.0482	[0.035, 0.063]
Caa*	0.479	[0.37, 0.6]	N.A.	N.A.
Caa	0.106	[0.15, 0.55]	0.435	[0.34, 0.55]

Table 5

The estimated one-year default probabilities and their bootstrapped 95% confidence sets for both the extended and the standard model based on the estimated transition probability matrices over the period 1995–1999

Rating	Bootstrap extended model		Bootstrap standard model	
	Estimate	Conf. set	Estimate	Conf. set
Aaa	8.38e-11	[4.2e-12, 4e-10]	1.09e-11	[1.5e-12, 3.2e-11]
Aa	2.15e-08	[1.5e-09, 9.4e-08]	2.89e-09	[5.2e-10, 6.8e-09]
A	2.36e-06	[1.9e-07, 9.8e-06]	3.34e-07	[6.6e-08, 7.7e-07]
Baa*	2.51e-04	[9.6e-06, 0.001]	N.A.	N.A.
Baa	1.16e-05	[3.1e-06, 0.00014]	3.04e-05	[4.5e-06, 8.2e-05]
Ba*	0.000665	[0.00027, 0.0013]	N.A.	N.A.
Ba	0.000849	[0.00027, 0.0017]	0.000325	[0.00016, 0.00054]
B*	0.0238	[0.012, 0.038]	N.A.	N.A.
B	0.0217	[0.015, 0.028]	0.0110	[0.0074, 0.015]
Caa*	0.444	[0.33, 0.53]	N.A.	N.A.
Caa	0.0376	[0.032, 0.083]	0.171	[0.14, 0.21]

Before discussing the main results, we motivate the choice of bootstrapped confidence sets by comparing the computation for one of the periods based on the binomial method (from discrete data) with the bootstrap method (based on continuously observed data), see Table 3. For the period 1987–1991 the binomial approach gives us zero-estimates for the top categories Aaa, Aa, and A, whereas the continuous-time approach leads us to a non-zero estimate of the actual default probability over the period. For the categories Baa down to B, the binomial approach gives much higher default probability estimates, and in both cases the confidence sets from the binomial approach are much wider compared with those obtained by bootstrapping the continuous-time model. But more seriously, in the binomial approach the width of the confidence sets is larger in Aaa than in Aa, for example, simply due



Table 6

The estimated one-year transition probability matrix for the period 1987–1991

	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	0.973195	0.024699	0.000746	0.001313	0.000033	1.25e−05	5.02e−07	3.57e−07
Aa	0.006218	0.934759	0.055947	0.001853	0.000639	0.000543	1.53e−05	1.59e−05
A	0.000457	0.006776	0.958695	0.02792	0.004549	0.001495	4.90e−05	5.95e−05
Baa	0.000453	0.003542	0.025301	0.924547	0.036552	0.008514	0.000578	0.000513
Ba	2.24e−05	0.001487	0.005157	0.069724	0.835357	0.072105	0.003962	0.012185
B	2.91e−06	5.40e−05	0.002058	0.010984	0.036749	0.858297	0.043610	0.048245
Caa	1.32e−05	0.000137	0.013102	0.037577	0.012966	0.024217	0.476543	0.435445
D	0	0	0	0	0	0	0	1

Table 7

The estimated one-year transition probability matrix for the period 1995–1999

	Aaa	Aa	A	Baa	Ba	B	Caa	D
Aaa	0.980954	0.018711	0.000332	2.69e−06	1.13e−07	2.18e−09	2.21e−10	1.09e−11
Aa	0.002565	0.962642	0.034355	0.000419	1.82e−05	4.58e−07	4.72e−08	2.89e−09
A	1.87e−05	0.009717	0.965604	0.023521	1.09546	3.92e−05	4.16e−06	3.34e−07
Baa	0.000473	0.000382	0.028419	0.948274	0.020343	0.001804	0.000275	3.04e−05
Ba	1.96e−05	3.40e−05	0.005145	0.078280	0.874130	0.039655	0.002411	0.000325
B	1.93e−06	7.20e−06	0.0012967	0.008322	0.041973	0.884010	0.053386	0.011003
Caa	3.24e−08	1.22e−07	3.19e−05	0.000201	0.001109	0.045791	0.781791	0.171075
D	0	0	0	0	0	0	0	1

to the lower number of firms in the Aaa category. The confidence sets should reflect the dynamics of transitions and therefore we use bootstrapped confidence sets in all that follows. Note also, that for the Caa-category the binomial approach leads to relatively low estimates because the Caa-category is in most cases just a transitional category that issuers pass through on their way to default. The continuous-time model catches this effect resulting in a higher estimate for the probability of default.

We now turn to the extended model and the comparison of the results of that model to those of the standard model. For the period 1995 through 1999, the estimated one-year default probabilities and the corresponding bootstrapped confidence sets are presented in Table 5. For the investment grade categories the point estimate of the default probability in this period is scaled up approximately by a factor of 10 in the extended model, and the same holds for the 97.5% quantile of the simulated distribution. For the Ba-category of the extended model there is no big difference between the excited and non-excited state, but they are both twice as big as the point estimate of the standard model. For the B-category of the extended model we see a similar picture. The cause for this difference between the two models is to be found in the Caa-category, where the extended model has a significant difference between the excited category Caa\* and the non-excited category Caa. In the standard model the two types of issuers are merged into one category with an estimated one-year probability of default of 0.17.





The results for the period 1987 through 1991 shown in [Table 4](#), show the same picture for the investment grade categories. However, now the difference between the point estimates of the standard and the extended model is roughly a factor of 2. In this period the default probabilities were already at an extremely high level in historical terms.

For the Ba- and B-categories we see a two-type system where the members of the excited states are being downgraded through the categories much faster than the more stationary subset in each category. In the standard model these two types of issuers are simply merged into one category for each rating, and as a consequence the estimated one-year default probability for these two categories lies between the estimates for the corresponding excited and non-excited categories in the extended model.

A more detailed description and explanation can be obtained by looking at the estimated one-year transition probability matrices for the extended model in [Tables 8 and 9](#). For the period 1987 through 1991 we see a marked difference between excited and non-excited states in the size of the diagonal elements. An issuer in a non-excited state is more likely than an issuer in an excited state to stay in the same state for one year. The issuer in the excited state has a measurable probability of falling into the normal state and an increased probability of suffering downgrades. In the period 1995–1999, the rating activity is much smaller and the difference between the two types of states is harder to detect.

Note that only in the extended state space model, and only when considering the worst five-year period of the 32 year long dataset, i.e. the period January 1, 1987 to December 31, 1991, do we see the 97.5% quantile of our bootstrapped confidence sets reach a level of 0.02% for the A-category.

To complete the discussion, [Figs. 2 and 3](#) give representative graphical illustrations whose construction are described in detail at the end of [Section 5](#). These show the value of the ‘true’ one-year default probability and the kernel smoothed density following from the simulated distribution of the one-year default probability. The 2.5% and 97.5% quantiles of these distributions are marked with lines in the graphs. The figures thus reflect the variations in estimates due to the limited sample size. The fact that our ‘true’ transition structure is of course only an estimate is only corrected for indirectly by considering two different periods. [Fig. 2](#) shows the important point that there is an effect of using an extended model even for the category A in which we do not have an excited state. The fact that there are excited states in the system below A, significantly increases the default probability. [Fig. 3](#) shows the heterogeneity between firms in the excited state Baa\* and those in the ‘normal’ state Baa.

As a final remark, note that this effect also holds for 3- and 5-year probabilities of default estimated in the two models. As shown in [Table 10](#) the default probability estimates for rare events based on the volatile period show the same relative differences as the one-year estimates simply because of the almost linearity of the small transition probabilities in the time period lengths considered.

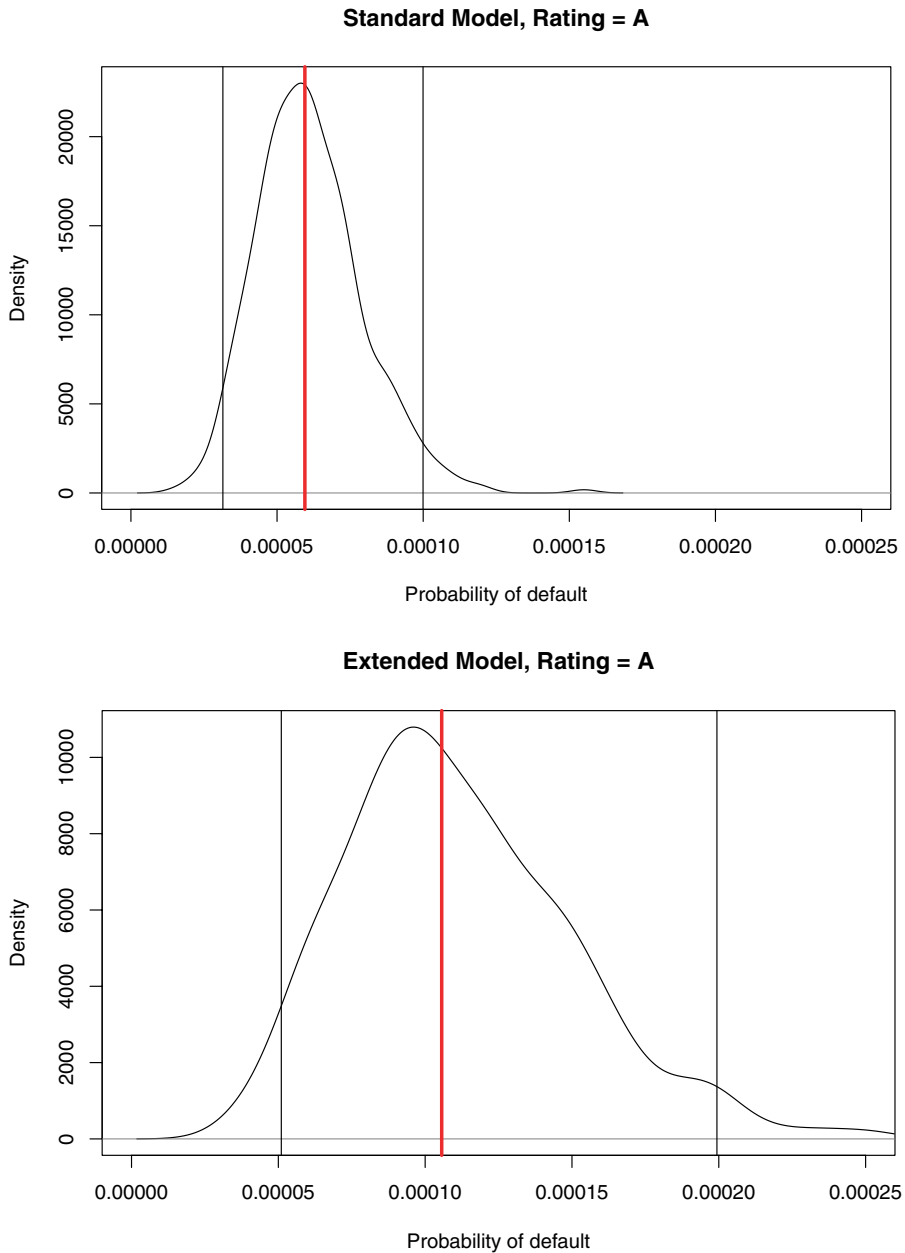


Fig. 2. Simulated density distribution of the default probability estimates for A-rated issuers over the period 1987–1991. In the top graph is depicted the distribution obtained from the standard model, whereas the bottom graph shows the distribution obtained from the extended model. The bold vertical line in each graph shows the one-year default probability (PD) corresponding to the ‘true’ value derived from the estimated transition probability matrix. The thin vertical lines mark the 2.5% and the 97.5% quantile.

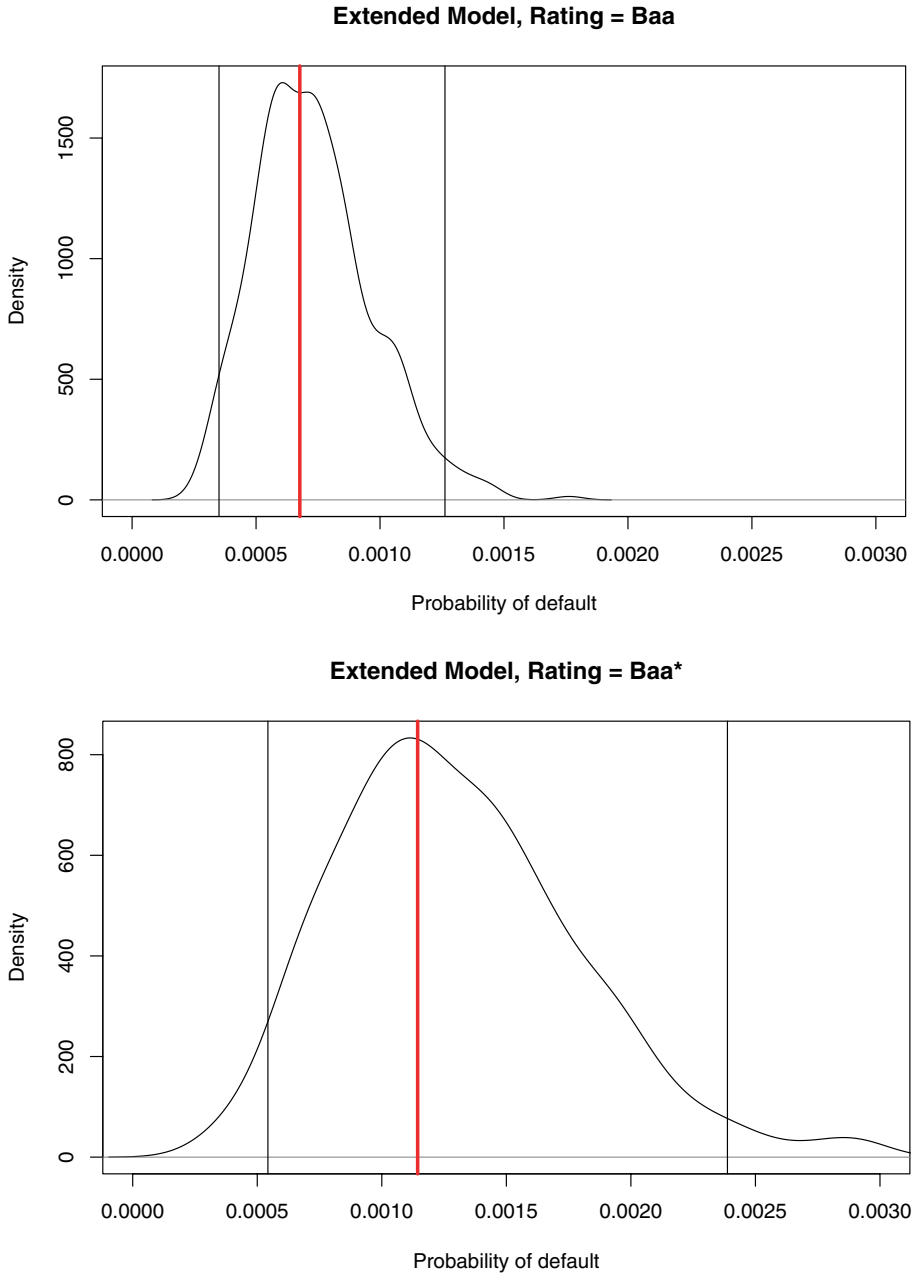


Fig. 3. Simulated density distribution of the default probability estimates for Baa (the top graph) and Baa\* in the extended model over the period 1987–1991. The bold vertical lines show the one-year default probability (PD) corresponding to the 'true' value derived from the estimated transition probability matrix. The thin vertical lines mark the 2.5% and the 97.5% quantile.

Table 10

The estimated 3-year and 5-year default probabilities for both the standard and the extended model based on the estimated one-day transition probability matrix over the period 1987–1991

Rating	3-year default probability		5-year default probability	
	Standard	Extended	Standard	Extended
Aaa	0.0000108	0.0000225	0.0000531	0.000107
Aa	0.000193	0.000339	0.000641	0.00107
A	0.000707	0.00128	0.00230	0.00405
Baa*	N.A.	0.0110	N.A.	0.0287
Baa	0.00477	0.00614	0.0132	0.0161
Ba*	N.A.	0.0768	N.A.	0.139
Ba	0.0458	0.0364	0.0877	0.0737
B*	N.A.	0.270	N.A.	0.380
B	0.171	0.120	0.282	0.203
Caa*	N.A.	0.780	N.A.	0.839
Caa	0.746	0.269	0.821	0.387

## 7. Conclusion

We have investigated the importance of the Markov assumption in the method of estimating one-year transition probabilities using the continuous-time Markov technique as advocated in Lando and Skødeberg (2002). By taking into account one of the most pronounced non-Markov effects in ratings, namely the higher downgrade intensity of recently downgraded firms, we find that the non-zero default probability estimates obtained from the generator method for investment grade classes are increased further. This is true for both the ‘quiet’ period 1995–1999 and the more volatile period 1987–1991. If we compare the estimated default probabilities for the top categories in the extended state space model with the standard state space model, we find larger estimates in the model with extended state space simply because the risk of migrating down through a sequence of excited states is larger. We also find larger probabilities of migrating down from an excited state than from the corresponding non-excited state in the extended model, thereby confirming that the terminology is appropriate.

We also use a bootstrap procedure to estimate the confidence bands for the default probabilities. This is used to address a particular problem discussed in the Basel Accord. Estimates of rating transition probabilities often suffer from small samples, either in the number of firms which are actually rated or in the number of events which take place. This often results in estimates which are 0 even if there may be reason to believe that the events can and will actually occur given a large enough sample. This insecurity has led the Basel Committee on Banking Supervision to impose a lower minimum probability of 0.0003 for rare events. This paper allows us to assess whether this is a reasonable limit for sample sizes corresponding to the number of US corporate issuers in the Moody’s Default Database.

We find that the minimum of the Basel Committee corresponds well to the estimate of the default probability for Ba issuers in the standard model or Baa\* issuers in the extended model in the stable 5-year period beginning in 1995, and that this minimum

is somewhere between our estimate for the A and Baa category in the volatile 5-year period beginning in 1987 for both types of models. If a 97.5% confidence limit was the base of this figure, we can conclude that the A-default probability in the volatile period is below the 3 bps (at only 2 bps). The minimum level of 0.03% put forward by the Basel Committee in its latest proposal for a new capital accord therefore is conservative based on the evidence in our data for the top three investment grade categories.

### Acknowledgments

We are grateful to Moody's Investors Services for providing the data set and for supporting the research. In particular, we would like to thank Richard Cantor, David Hamilton and Roger Stein for useful conversations. We are grateful for receiving valuable feedback from an anonymous referee and seminar participants at a meeting of Moody's Investors Service Academic Advisory Board, at the Credit Risk workshop at Carnegie-Mellon University and at the C.R.E.D.I.T 2002 conference in Venice. In particular, we thank our two discussants Paul Embrechts and Michael Gallmeyer for their valuable comments. All errors and opinions expressed in this paper are of course our own.

### Appendix A

In this appendix we will present the technique used for estimating the extended state space model from Section 4.

We start by recalling the classical terminology for hidden Markov Models (abbreviated HMM in the following). A HMM consists of a background process  $(X_t)_{t \in T}$ , in discrete or continuous time as the case may be, and an observed process  $(Y_t)_{t \in T}$ . The background process has values in a set  $E$ , denoted the set of states. The observed process has values in another set  $A$ , denoted the set of symbols. Usually we assume that there are only finitely many states and finitely many symbols.

In the extended model of this paper the background process has a state space given by

$$E = \{Aaa, Aa, A, Baa^*, Baa, Ba^*, Ba, B^*, B, Caa^*, Caa, D\}$$

and the observed process has a set of symbols equal to

$$A = \{Aaa, Aa, A, Baa, Ba, B, Caa, D\}.$$

There are four ingredients in the description of the distribution of the processes. Firstly, the background process is assumed to be a time-homogeneous Markov process. Secondly, conditionally on the  $X_t$ -variables in any given time window  $[a, b]$ , the corresponding observed variables in that time window,  $(Y_t)_{t \in [a, b]}$ , are independent. Thirdly, conditionally on  $X_{t_0}$ , the corresponding observed variable  $Y_{t_0}$  is independent of all other background variables,  $(X_t)_{t \neq t_0}$ . Finally, the conditional distribution of  $Y_t$  given  $X_t$  does not depend on  $t$ .



The most obvious example of a HMM, and in fact the only type we consider in this paper, is given by the relation  $Y_t = f(X_t)$ , where  $(X_t)_{t \in T}$  is a time-homogeneous Markov process with state space  $E$ , and where  $f: E \rightarrow A$  is some function. It is well known that the observed process  $(Y_t)_{t \in T}$  in this case may be non-Markovian. But of course, if  $f$  is simply the identity  $E \rightarrow E$ , the observed process is identical to the background process, and consequently Markovian. So the class of processes that are functions of Markov processes, form a broader larger class than the Markov processes themselves. And the class of HMM's is even wider, at least if we insist that the background Markov process has a finite state space.

A parsimonious description of a HMM is given in terms of the transition structure of the background process  $(X_t)_{t \in T}$ ,<sup>8</sup> the initial distribution of the background process, and the emission probability, that is, the conditional distribution of  $Y_t$  given  $X_t$ . If the observed process is simply a function of the background process, the emission probability is trivial, and the parsimonious description is given in terms of the parameters for the background Markov process.

We now turn to a discussion of the HMM we study in this paper. We consider a symbol set  $A$  which is ordered (from 'good' to 'bad'), a state space  $E$  and a function  $f: E \rightarrow A$  with the property that each symbol  $a \in A$  has one or two pre-images in  $E$ . If  $a$  has just one pre-image (a simple symbol), this is also denoted  $a$ , if it has two pre-images (a composite symbol), they are denoted  $a$  and  $a^*$ . We can extend the order structure from  $A$  to  $E$  by considering  $a$  and  $a^*$  to be next to each other, and by considering  $a^*$  'better' than  $a$ .

We consider Markov processes on  $E$ , which respect this ordering. This means that we *disallow* the following three types of transitions:

- 
1.  $a \rightarrow a^*$  if  $a$  is a composite symbol
  2.  $x \rightarrow a$  if  $a^*$  exists and if  $x$  is a state that is better than  $a^*$
  3.  $x \rightarrow a^*$  if  $a^*$  exists and if  $x$  is a state that is worse than  $a$
- 

There is a well-developed methodology for estimating the parameters of an HMM in discrete time, with observations from a single realization of the model. This methodology is usually referred to as the *Baum–Welch algorithm* (Baum et al., 1970; Baum, 1972) – see Koski (2001) for a careful recent discussion. In modern language, the Baum–Welch algorithm is best viewed as a special case of the EM-algorithm, see Dempster et al. (1977), even though the algorithm itself and the clarification of its convergence properties predates the work of Dempster et al. considerably.

In the E-step of the EM-algorithm, a new objective function is constructed on the basis of a preliminary estimate of the parameters. This objective function is subsequently maximized in the M-step, and the argument maximizing it is the updated parameter estimate. In the HMM-setting, the new objective function is

<sup>8</sup> "Transition structure" is a common term for the transition intensity if time is continuous, and the one-step transition probability if time is discrete.

$$\theta \mapsto E_{\theta_0} \log P_{\theta}(X | Y = y),$$

where  $\theta_0$  is the present parameter estimate,  $X$  is shorthand for the entire background process, and  $Y$  is shorthand for the entire observed process. This objective function turns out to have a product-multinomial form, just like the ordinary log-likelihood for a standard Markov model, and it is easily maximized: The transition probability from  $a$  to  $b$  is estimated as

$$\hat{P}_{ab} = \frac{\hat{N}_{ab}}{\sum_c \hat{N}_{ac}}.$$

The sum in the denominator is over all possible states  $c$ . In a usual Markov model the numbers  $\hat{N}_{ab}$  would simply mean the number of observed transitions from  $a$  to  $b$ , here it means the expected number of transitions, given the observed  $Y$ -sequence (note that this expectation is with respect to the present parameter estimate),

$$\hat{N}_{ab} = E \left( \sum_{n=1}^{N-1} \mathbf{1}_{(X_n=a, X_{n+1}=b)} \mid Y_1 = y_1, \dots, Y_N = y_N \right).$$

It is easy to write up  $\hat{N}_{ab}$  as a sum over all possible paths of the background process. But the sum is not directly calculable if the process is observed over a long time – there are simply too many terms. The main contribution of Baum et al. was an organization of the calculation of  $\hat{N}_{ab}$  in terms of two linear programming algorithms, known as the *forward* and *backward* algorithm, which both run in time directly proportional to the number of observations. The forward algorithm calculates

$$F(n, a) = P(Y_1 = y_1, \dots, Y_n = y_n, X_n = a)$$

for each  $a$ , inductively in  $n$ , utilizing the induction formula

$$F(n + 1, b) = \sum_a F(n, a) P_{ab} P(Y_{n+1} = y_{n+1} \mid X_{n+1} = b), \tag{A.1}$$

where  $P_{ab}$  is the present transition probability from  $a$  to  $b$ . The backward algorithm calculates

$$B(n, a) = P(Y_n = y_n, Y_{n+1} = y_{n+1}, \dots, Y_N = y_N \mid X_n = a)$$

for each  $a$ , inductively in  $n$ , but this time going from  $n=N$  down to  $n=1$ . The backward algorithm utilizes the induction formula

$$B(n, a) = \sum_b P_{ab} B(n + 1, b) P(Y_n = y_n \mid X_n = a). \tag{A.2}$$

From the output of these algorithms we may compute the expected number of transitions from  $a$  to  $b$  as

$$\hat{N}_{ab} = \frac{\sum_{n=1}^{N-1} F(n, a) P_{ab} B(n + 1, b)}{\sum_c F(N, c)}. \tag{A.3}$$

We have implemented the Baum–Welch procedure described above, with a few adjustments. The first has to do with the nature of time. We have data on a daily basis, so it is rather natural to think of time as discrete and to consider day-to-day transitions. But in that perspective our observations are extremely non-volatile: the issuers typically have the same rating for years. And this makes the direct calculations of the forward and backward algorithms unnecessarily time consuming, considering how little is happening in each step. It turns out to be dramatically more efficient to calculate  $F(n,a)$  and  $B(n,a)$  only at the times of observed jumps in the  $Y$ -process. Our rather simple HMM-structure makes this possible, at the expense of slightly more complicated versions of (A.1) and (A.2).

The change in viewpoint on how the induction is performed, has a conceptual advantage, as well as a numerical: The standard Baum–Welch algorithm does not make sense in continuous time, as there is nothing to do induction on. But induction from jump to jump makes perfect sense. In the application it turns out that there is no difference between results obtained if we think of ratings as being given on a day-to-day basis, or if we think of the ratings as being given in continuous time, though.

The second modification of the Baum–Welch procedure has to do with the fact that the standard procedure is designed for one long sequence of observations, where we have many independent, but shorter sequences. We have adapted a standard trick of adding a fictitious state, corresponding to censoring, at the end of each sequence, and then gluing the sequences together to form one long sequence. Analyzing this artificial sequence, we get ‘estimates’ for the transitions to and from the censor state. The transitions from the censor state are more or less nonsensical, but can be disregarded without any problems. Transitions to the censor state is a slight problem, as these transitions give a bias downwards to all other transitions, but we can simply normalize them out.

## References

- Altman, E., Kao, D.L., 1992a. Corporate Bond Rating Drift: An Examination of Credit Quality Rating Changes over Time. The Research Foundation of the Institute of Chartered Financial Analysts, Charlottesville, VA.
- Altman, E., Kao, D.L., 1992b. The implications of corporate bond rating drift. *Financial Anal. J.* 48 (3), 64–75.
- Altman, E., Kao, D.L., 1992c. Rating drift in high yield bonds. *Journal of Fixed Income*, 15–20.
- Bangia, A., Diebold, F., Kronimus, A., Schagen, C., Schuermann, T., 2002. Ratings migration and the business cycle, with applications to credit portfolio stress testing. *Journal of Banking and Finance* 26, 445–474.
- Baum, L.E., 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In: *Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, CA, 1969; dedicated to the memory of Theodore S. Motzkin)*. Academic Press, New York, pp. 1–8.
- Baum, L.E., Petrie, T., Soules, G., Weiss, N., 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematics and Statistics* 41, 164–171.
- 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.

- Cantor, R., 2001. Moody's investors service response to the consultative paper issued by the Basel Committee on bank supervision "a new capital adequacy framework". *Journal of Banking and Finance* 25, 171–185.
- Carty, L., 1997. Moody's Rating Migration and Credit Quality Correlation, 1920–1996. Special Comment. Moody's Investors Service, New York.
- Carty, L.V., Fons, J.S., 1993. Measuring Changes in Corporate Credit Quality. Moody's Special Report. Moody's Investors Service, New York.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39 (1), 1–38 (with discussion).
- Duffie, D., Gârleanu, N., 2001. Risk and valuation of collateralized debt obligations. *Financial Analysts Journal* 57 (1), 41–59.
- Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. CBMC-NSF Regional Conference Series in Applied Mathematics.
- Fledelius, P., Lando, D., Nielsen, J.P., 2004. Non-parametric analysis of rating transition and default data. *Journal of Investment Management* 2, 71–85.
- Hamilton, D., Cantor, R., 2004. Rating Transitions and Defaults Conditional on Watchlist. Outlook and Rating History. Special Comment. Moody's Investors Service, New York.
- Höse, S., Huschens, S., Wania, R., 2002. Rating migrations. In: Härdle, W., Kleinow, T., Stahl, G. (Eds.), *Applied Quantitative Finance*. Springer, Berlin, pp. 53–76.
- Kavvathas, D., 2000. Estimating credit rating transition probabilities for corporate bonds. Working Paper. Department of Economics, University of Chicago.
- Koski, T., 2001. Hidden Markov Models for Bioinformatics. *Computational Biology Series*, vol. 2. Kluwer Academic Publishers, Dordrecht.
- Lando, D., Skødeberg, T.M., 2002. Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking and Finance* 26, 423–444.
- Lucas, D., Lonski, J., 1992. Changes in corporate credit quality 1970–1990. *Journal of Fixed Income* 1, 7–14.
- Nickell, P., Perraudin, W., Varotto, S., 2000. Stability of ratings transitions. *Journal of Banking and Finance* 24, 203–227.