Confidence sets for continuous-time rating transition probabilities

Jens Christensen and David Lando
Department of Statistics and Operations Research
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø.
Denmark
fax: +45 3532 0678

This draft: October 16, 2002
First draft: May 2002

1We 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 have received valuable feedback from seminar participants when presenting 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.
1 Introduction

The Basel Committee on Banking Supervision, a regulatory body under the Bank of International Settlements, has in its 'New Capital Accord' proposed a regulatory setup in which banks are allowed to base the capital 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 formally 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 will have to assign default probabilities to each rating category and to probabilities of transition between non-default categories. There are many statistical issues, of course, in assigning such probabilities. A fundamental problem is the relatively small sample sizes which have to be used for the estimation. Not only are defaults rare in the top categories. Transitions to 'distant' rating categories are also rare. Most often, transitions involve the transition from a sub-category of a rating class (from, say Baa1 to Baa2 in Moody’s system or from BBB+ to BBB in Standard and Poor’s classification). Hence to observe more transition activity we may choose to include rating modifiers. However, that also greatly increases the number of rating transition probabilities to be estimated and in fact leaves us with even more rare events in our model. So whether or not we include 8 or (say) 18 rating categories leaves us with important estimates which have to be based on few events.

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. This is not possible using classical 'multinomial' techniques, such as those of Carty and Fons (1993) and Carty (1997). In this paper, we show that the continuous-time procedure also allows us to find significantly improved confidence sets for rare events. Our method is based on bootstrapping\(^1\) the generator and we contrast this method with a simple binomial approach and a multinomial approach. Both Nickell, Perraudin, and Varotto (2000) and Höse, Huschens, and Wania (2002) contain estimates of standard deviations and confidence sets, but since they are based on multi-

\(^1\)For an introduction to the bootstrap, see Efron (1982).
nominal type estimators, they cannot assign meaningful confidence sets to probabilities of rare events.

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 shortly, a proper use of the full information in the data set gives a better impression of the appropriateness of such a limit.

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. While the observed difference between the two quantities is not removed, a proper analysis of confidence sets may still alter the picture significantly.

The usefulness of the method becomes even more transparent when attacking some of the fundamental problems in the analysis of rating transitions, namely the inclusion of business cycle variables and the handling of non-homogeneities, such as industry and country effects, within rating classes. Whenever we try to analyze such problems we do so by introducing more covariates and hence limit the number of events even further for any given value of the covariates. In a related paper, Fledelius, Lando, and Nielsen (2002) consider the use of smoothing techniques to mitigate this problem. In this paper, our focus is on the estimation of transition probabilities without the inclusion of covariates. We consider instead the uncertainty of estimates by comparing a stable period (1995-1999) with a volatile period (1987-1991). There is clear evidence of variations in rating transition probabilities with time. Nickell, Perraudin, and Varotto (2000) demonstrate a dependence on the state of the business cycle, whereas Blume, Lim, and MacKinlay (1998) suggest a change in rating policy as an explanation for the variation.

The outline of the paper is as follows: Section 2 briefly gives a recapitulation of a discrete, multinomial estimation of the transition parameters and a continuous-time method based on the generator. We then describe two ways of obtaining confidence intervals for default events - one based on the discrete method and one based on a bootstrap method. Section 3 describes
our data and some of the data cleaning issues one faces when looking at the
data set. Section 4 gives our results and section 5 concludes.

2 Methodology

We first recapitulate how to obtain transition probability estimates in a dis-
crete setting where only yearly rating data are used and a continuous-time
setting where the entire history of data is used.

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 pre-
cisely, the maximum-likelihood estimator for the time-independent transition
probability becomes

\[
\hat{p}_{ij} = \frac{\sum_{t=0}^{T-1} n_{ij}(t)}{\sum_{t=0}^{T-1} n_i(t)}
\]

(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
depsends 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 prob-
ability 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.

Now we turn to estimation based on continuous observations. The generator matrix of a time-homogeneous Markov chain plays a central role in this procedure so we briefly recall its properties and interpretation.

Let \( P(t) \) denote the transition probability matrix of a continuous-time Markov chain with finite state space \( \{1, \ldots, K\} \) so that the \( ij \)'th element of this matrix is

\[
P_{ij}(t) = P(\eta_t = j | \eta_0 = i)
\]

The generator \( \Lambda \) is a \( K \times K \) matrix for which

\[
P(t) = \exp(\Lambda t) \quad \text{for all } t \geq 0
\]

where

\[
\exp(\Lambda t) \equiv \sum_{k=0}^{\infty} \frac{(\Lambda t)^k}{k!}
\]

The diagonal element of \( \Lambda \) we write \( -\lambda_i \) where

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

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

The waiting time for leaving state \( i \) has an exponential distribution with mean \( \frac{1}{\lambda_i} \). Once the chain leaves \( i \) it jumps to state \( j \neq i \) with probability \( \frac{\lambda_i}{\lambda_j} \).

The maximum-likelihood estimator of \( \lambda_{ij} \) based on observing realizations of the chain from time 0 to \( T \) is

\[
\hat{\lambda}_{ij} = \frac{N_{ij}(T)}{\int_0^T Y_i(s) ds}, \quad i \neq j^2
\]

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

\[
\hat{P}(1) = \exp(\hat{\Lambda}).
\]

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:

---

\(\footnote{Of course, \( \hat{\lambda}_i = \sum_{j \neq i} \hat{\lambda}_{ij} \).} \)
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 horizons can be obtained with no need to struggle with the embedding problem for Markov chains.

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.

In this paper we focus on the advantages of using the continuous-time estimator for confidence set construction. To understand how the generator based methodology improves our confidence set estimation as well, it is useful first to consider a simple binomial procedure for constructing confidence bands in the multinomial setting.

Consider a binomial random variable $X \sim b(\theta, N)$ where $\theta$ is the probability of failure. Given that we observe $X = 0$, 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 $X = 0$. 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 $\theta^{\text{max}}(N, \alpha)$ we find

$$\theta^{\text{max}}(N, \alpha) = 1 - \alpha^{\frac{1}{N}}$$

In Figure 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 $n_i(t) = \text{“number of firms in category } i \text{ at the beginning}$
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 $\bar{X}_i$ 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 $\theta_i^{\text{min}}$ 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 $\bar{X}_i$, that is $\theta_i^{\text{min}}$ must solve the following equation:

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

On the other hand let $\theta_i^{\text{max}}$ 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
Table 1: Upper 95% and 99% boundaries of $\theta$ for the Aaa, Aa, and A categories based on the sample size in the five year period 1995 through 1999. All six values correspond to values which could in principle be seen in Figure 1.

<table>
<thead>
<tr>
<th></th>
<th>$N_i$</th>
<th>$\theta_i^{\text{min}}(N_i, 0.05)$</th>
<th>$\theta_i^{\text{max}}(N_i, 0.05)$</th>
<th>$\theta_i^{\text{min}}(N_i, 0.01)$</th>
<th>$\theta_i^{\text{max}}(N_i, 0.01)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>189</td>
<td>0.015725478</td>
<td>0.024071526</td>
<td>0.007226003</td>
<td></td>
</tr>
<tr>
<td>Aa</td>
<td>635</td>
<td>0.004706578</td>
<td>0.007226003</td>
<td>0.001315361</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2276</td>
<td>0.001315361</td>
<td>0.002021316</td>
<td>0.002021316</td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>$N_i$</th>
<th>$X_i$</th>
<th>$\theta_i^{\text{min}}(N_i, 0.05)$</th>
<th>$\theta_i^{\text{max}}(N_i, 0.05)$</th>
<th>$\theta_i^{\text{min}}(N_i, 0.01)$</th>
<th>$\theta_i^{\text{max}}(N_i, 0.01)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baa</td>
<td>2092</td>
<td>1</td>
<td>0.000001210</td>
<td>0.00266040</td>
<td>0.00000240</td>
<td>0.00354623</td>
</tr>
<tr>
<td>Ba</td>
<td>883</td>
<td>1</td>
<td>0.00002867</td>
<td>0.00629359</td>
<td>0.00000568</td>
<td>0.00838407</td>
</tr>
<tr>
<td>B</td>
<td>1131</td>
<td>42</td>
<td>0.02689246</td>
<td>0.04986679</td>
<td>0.02418406</td>
<td>0.0541287</td>
</tr>
<tr>
<td>Caa</td>
<td>217</td>
<td>29</td>
<td>0.09136144</td>
<td>0.1862607</td>
<td>0.08045494</td>
<td>0.2035239</td>
</tr>
</tbody>
</table>

probability $\frac{\alpha}{2}$ to observe as few defaults as $\tilde{X}_i$:

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

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 will also be solved in the generator based method.
The only advantage in the binomial case is that we are able to derive genuine confidence sets, i.e. to analyze the set of parameters which an associated test would not reject based on the given observations. While it is in theory possible to develop asymptotic expressions for the distribution of test statistics in the continuous-time formulation and use those for building approximate confidence set, in practice the bootstrap method seems both easier to understand and to implement. To see why this is, note that the maximum-likelihood estimator does not have a simple closed form expression for its variance/covariance matrix and we would need this to say something about the confidence sets based on an asymptotic variance. In fact, we would need to use asymptotics twice. First to find the variance of $\hat{\Lambda}$ - using the Fisher information and then finding an expression for the variance of $\exp(\hat{\Lambda})$ - something which again only seems feasible using an asymptotic argument. But the asymptotic variance of $\hat{\Lambda}$ is hardly a good estimator, since we have many types of transitions which occur rarely in our data set.

We therefore choose to use bootstrap methods instead to build (one-dimensional) confidence sets\(^3\) for one-year transition probabilities. We proceed using what is sometimes referred to as the 'parametric bootstrap':

1. Estimate $\Lambda$ from our sample of issuers over a chosen time horizon (5 years, say). Call this estimate $\Lambda^*$.

2. Simulate a large number of 5 year histories for the same number of issuers in each rating category using $\Lambda^*$ as the data generating parameter.

3. For each history, compute the estimator $\hat{\Lambda}^*$ and $\exp(\hat{\Lambda}^*)$.

4. For the parameter of interest (such as a particular default probability), compute the relevant quantiles.

For a given parameter, this gives us a confidence set for that parameter. If we use 95% as our confidence set, we could expect to see 95% of our estimates of that parameter in the confidence interval. Of course, this only

\(^3\)Perhaps the terminology is loose here. In essence we are asking, assuming that our estimate is the true parameter, which values of our estimator are likely to occur. The estimator is a function of the data, and in a sense we are asking which data are compatible with the estimate - a different idea than the clean confidence procedure.
holds “marginally”. It would of course be extremely rare to see all parameters outside their confidence sets at once. But it is very difficult to report simultaneous confidence intervals for high-dimensional parameters as ours.

3 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. Moody’s rate every single debt issue of the issuers in the database. The overall ability of the issuer to honor future payments as well as for each issue seniority, call features, coupon structure, and maturity are taken into consideration. We consider for this study only issuers domiciled in the United States, and we exclude all but the senior unsecured issues. For a start this leaves us with 3,405 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 3,670 observations on 1,451 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. On the other hand if a ‘withdrawn rating’ is genuinely the last observation, and no other issues are alive, we leave out of consideration this issuer from that date on. Otherwise, the issuer is considered alive at the final date of the observation period which is the 9th of January 2002.

Having corrected for withdrawn ratings, we should in principle 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. It turns out that there are only 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

\footnote{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’.}
where we could not find a reason for the different ratings, but in these cases it was 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 do not use a default category as such, but do 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 304 recorded defaults manually. These defaults are recorded by Moody’s in their ‘Default Master’ database, and whenever reasonable, these default dates are given first priority as a definition of a default date.

All rating observations up till the date of default have in most cases been left unchanged but there are exceptions. If a transition from 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 default intensities for 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.5

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 ex-

---

5When 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.
change' 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 it is a case of just one issuer defaulting once or if it is a case, according to our rule above, of really two independent issuers each defaulting. After this procedure, we have now come down to 10,007 observations of rating changes distributed among 3,437 issuers.\footnote{The number has gone up from 3,405 due to our procedure of introducing new issuers, if the post-default information is judged to be real rating changes.}

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. Since withdrawals are treated as censored observations and the 'default' state is viewed as absorbing (or, at least, the recovery time is not analyzed), we have to estimate \(21 \times 22 = 462\) rating transition probabilities. This is hard with the sample we have at our disposal and 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, a much more reasonable number.

Figures 2 and 3 show the total number of issuers in respectively the 'Investment grade' and 'Speculative grade' categories since the 1st of January 1970.

During the last decade the number of Aaa-rated issuers have gone down some 50\%. Except for Aa, which has stayed fairly stable, we have seen a solid growth in all the other categories since the beginning of the 80s where corporate bonds got their revival, having been out of fashion since World War II, though the burst of the IT-bubble in year 2000 cut off the top and meant a general deterioration of the overall credit quality.\footnote{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.}

One study by Lucas and Lonski (1992) based on the Moody’s database of that time proves a general tendency towards more downgrades than upgrades already from the beginning of the 80s (approximately in a 2 to 1 relationship). Blume, Lim, and MacKinlay
Figure 2: The number of issuers in the four investment grade categories since the 1st of January 1970.

The implication of these figures is clear. The low number of Aaa- and Aa-rated issuers combined with their very stable rating behavior makes it very difficult by traditional multinomial analysis to estimate with reasonable accuracy anything but the transition probabilities to the neighboring categories and to compute confidence intervals. That is the task to which we now turn.

4 Results

We have chosen to estimate the generator over two 5 year periods. The period between January 1, 1995 and December 31, 1999 was a period with a relatively stable macroeconomic climate and the other period from January 1, 1987 to December 31, 1991 includes the savings and loan crisis and a macroeconomic recession. At the starting date of the estimation we had the

(1998), basing their study on Standard and Poor's data, take the analysis a step further and explain the observed pattern by more stringent standards employed by the rating agency. Moody's apparently have gone down the same path.
distribution of issuers described in Table 3.

The experiment we carry out for each period separately is the following. First, we estimate the generator matrix using the maximum-likelihood estimator given in (2). Think of this estimate as the 'true' generator. We then compute the exponential of the generator and record the corresponding 'true' one-year default probability for each rating category. We then simulate 50,000 5-year rating histories for each issuer using the 'true' generator. Based on these simulated rating paths we obtain 50,000 estimates of the generator. We exponentiate each of these generators thus obtaining an approximation of

<table>
<thead>
<tr>
<th></th>
<th>Aaa</th>
<th>Aa</th>
<th>A</th>
<th>Baa</th>
<th>Ba</th>
<th>B</th>
<th>Caa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>68</td>
<td>200</td>
<td>397</td>
<td>239</td>
<td>87</td>
<td>81</td>
<td>9</td>
</tr>
<tr>
<td>1995</td>
<td>46</td>
<td>127</td>
<td>434</td>
<td>332</td>
<td>131</td>
<td>203</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 3: The distribution of issuers at the beginning of each sample period. The numbers include issuers who have their rating withdrawn at a later date.
the distribution of the maximum-likelihood estimator of the one-year default probability for each category. We refer to this as the generator method. We also compute, for each simulation, estimates of one-year default probabilities using a standard multinomial method, thus obtaining a distribution of the multinomial estimator of the one-year default probabilities. The graphs show the value of the 'true' default probability and the simulated distributions of the one-year default probability estimator based on the generator and on the multinomial method. We compute 2.5% and 97.5% quantiles of these distributions. The distributions thus reflect the variations in estimates due to the limited sample size. The fact that our 'true' generator is of course only an estimate is only corrected for indirectly by considering two different periods.

The results of the generator estimation for the two periods are presented in Table 7 and Table 10. Note that there is a heavy concentration around the diagonal so that most observed transitions are to a neighboring category. Baa and Ba have more activity away from the diagonal. If we calculate the exponential of the generators we obtain the one-year transition probability estimates reported in Table 8 and Table 11. The default probabilities from these tables are referred to in the graphs as the 'true' probabilities.

Note that this method of estimating the transition probabilities gets rid of all the zeroes despite the fact that not all transitions are observed in the sample. For a discussion of this point, see Lando and Skødeberg (2002). For rare events, we thus obtain at least an estimate of how rare they are. We now show, that the generator gives much narrower bootstrapped confidence sets than what could be obtained using a multinomial method when events are rare either because of a low probability of occurrence, a low number of firms in a category or both.

Figure 4 shows the results for the investment grade categories in the period between 1987 and 1991. From the figure for the Aaa category, note that with a true default probability of 0.05 basis points, all multinomial estimators will produce an estimate of zero for the default probability. When using the generator method however, we obtain a nice distribution around the 'true' value and an upper 97.5% quantile of 0.13 bps, see Table 5. In the Aa category, Figure 4 shows that the multinomial based method produces an estimate of 0 for most paths. However, this time since there are 200 firms to begin with, more than 2.5% of the 50,000 simulations will contain at least one default. Whenever one default occurs, the estimate of the default probability is 1 divided by the total number of observed firms in Aa (i.e. the sum of the Aa row in Table 9). This number is roughly 800, and this explains...
the 97.5% confidence band of 12.2 bps. This is outside the graph, since the corresponding quantile for the generator method is 2.7 bps. The narrower confidence band of course reflects the fact that the estimator is much more precise. The bimodal distribution we see for the generator method in the Aa category can be explained as follows: The ‘true’ generator has positive entries in the transition intensities from Aa to Ba and B. However these transitions are rare. This creates a division in the simulations between those paths where these transitions occur and those where they do not. When the transitions do occur, the default probability increases significantly. The bimodal distribution which shows up for the multinomial method in the A category, cf. Figure 4, has a different explanation. Here, we see three peaks which are created from paths in which the total number of defaults equals 0, 1 or 2. Since the total number of firms (the denominator in (1)) that were in A over the period varies with the simulation, there is a slight variation in the default probability estimate even holding the number of defaults fixed. Again, the generator estimator creates a much nicer distribution around the true value. This is also true of the Baa category.

The picture changes somewhat in Figure 5. In the low rating categories, Ba, B and Caa, the number of observations and the number of direct transitions to default is so large, that the multinomial estimator becomes almost as precise as the generator-based method.

In Figures 6 and 7 we see very similar results for the period 1995 through 1999. However, the fact that events have become more rare, means that the equivalence of the two estimation methods does not show up before the B category. In a sense the pattern we see for one rating category in the volatile period, now shows up in the category one step below (whenever this makes sense). Whether the confidence set for the Aaa category is of any use here is questionable. The problem is of course, that the period was extremely stable and this emphasizes the importance of considering different stages of the business cycle.

As documented by Table 5 and Table 6 the generator estimates give tighter intervals than the multinomial method for all categories (neglecting the zero-intervals for the top categories in the multinomial-analysis). The two are comparable, though, for the low categories. The binomial method suffers from the dependence on sample size only. An additional problem is worth noting as well, since this explains the deviations of the binomial probabilities observed in the lower categories and it highlights the uncertainty of the estimators in the low categories. Consider for illustration the last years
Table 4: The last years of the rating history of Bethlehem Steel Corporation which filed for chapter 11 on October 15, 2001. This type of rating history illustrates the difference between a multinomial method and a generator method as explained in the text.

<table>
<thead>
<tr>
<th>Date</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 29, 1998</td>
<td>Ba3</td>
</tr>
<tr>
<td>December 27, 2000</td>
<td>B2</td>
</tr>
<tr>
<td>January 31, 2001</td>
<td>Caa1</td>
</tr>
<tr>
<td>October 1, 2001</td>
<td>Caa3</td>
</tr>
<tr>
<td>October 15, 2001</td>
<td>Default</td>
</tr>
<tr>
<td>October 16, 2001</td>
<td>Ca</td>
</tr>
</tbody>
</table>

of rating history for Bethlehem Steel Corporation which filed for Chapter 11 bankruptcy protection on October 15, 2001. The history is shown in Table 4. A binomial method, which only looks at the rating at the beginning of the year and registers default or no default at year end, would register a default from category B2, but would not see a movement from Caa3 to default. Since this is a typical pattern, the binomial method underestimates the default probability from Caa. In fact the same error would occur if we used the multinomial method to estimate one-year default probabilities without first estimating the generator. The generator (in the 8 category system) does register a movement from Caa to default and in fact the relatively short time of exposure in the category contributes to a high estimate of this transition intensity. When we exponentiate the generator we obtain a high default probability estimate from Caa, and this will show up in the simulations and since there are then enough firms in the lower categories to make the multinomial estimator work reasonably well, its distribution will look like that based on the generator.

9Or a movement from Caa1 to default if one feels that Caa3 is really a de facto default registration which is then confirmed two weeks later. The estimates of the generator for the junk categories is highly sensitive to the method chosen for handling junk ratings close to default dates.
Table 5: The 95% confidence intervals for the period 1987 to 1991.

<table>
<thead>
<tr>
<th></th>
<th>Generator method</th>
<th>Multinomial method</th>
<th>Binomial approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>4.54E-07, 1.32E-05</td>
<td>[0, 0]</td>
<td>[0, 0.008983]</td>
</tr>
<tr>
<td>Aa</td>
<td>1.97E-05, 0.000265</td>
<td>[0, 0.001224]</td>
<td>[0, 0.003633]</td>
</tr>
<tr>
<td>A</td>
<td>0.000263, 0.000650</td>
<td>[0, 0.001576]</td>
<td>[0, 0.001561]</td>
</tr>
<tr>
<td>Baa</td>
<td>0.001653, 0.003712</td>
<td>[0, 0.005766]</td>
<td>[0.000526, 0.007431]</td>
</tr>
<tr>
<td>Ba</td>
<td>0.016169, 0.038402</td>
<td>[0.013043, 0.041850]</td>
<td>[0.017593, 0.055483]</td>
</tr>
<tr>
<td>B</td>
<td>0.079289, 0.129622</td>
<td>[0.074879, 0.134663]</td>
<td>[0.051945, 0.104391]</td>
</tr>
<tr>
<td>Caa</td>
<td>0.508539, 0.718387</td>
<td>[0.471698, 0.769231]</td>
<td>[0.234224, 0.517122]</td>
</tr>
</tbody>
</table>

Table 6: The 95% confidence intervals for the period 1995 to 1999.

<table>
<thead>
<tr>
<th></th>
<th>Generator method</th>
<th>Multinomial method</th>
<th>Binomial approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>7.49E-10, 4.14E-08</td>
<td>[0, 0]</td>
<td>[0, 0.015726]</td>
</tr>
<tr>
<td>Aa</td>
<td>4.97E-08, 4.19E-07</td>
<td>[0, 0]</td>
<td>[0, 0.004707]</td>
</tr>
<tr>
<td>A</td>
<td>2.05E-06, 1.45E-05</td>
<td>[0, 0]</td>
<td>[0, 0.001315]</td>
</tr>
<tr>
<td>Baa</td>
<td>0.000055, 0.000479</td>
<td>[0, 0.001172]</td>
<td>[0.000012, 0.002660]</td>
</tr>
<tr>
<td>Ba</td>
<td>0.000901, 0.002584</td>
<td>[0, 0.004934]</td>
<td>[0.000029, 0.006294]</td>
</tr>
<tr>
<td>B</td>
<td>0.022340, 0.039788</td>
<td>[0.019157, 0.043250]</td>
<td>[0.026892, 0.049867]</td>
</tr>
<tr>
<td>Caa</td>
<td>0.267641, 0.391573</td>
<td>[0.259669, 0.402235]</td>
<td>[0.091361, 0.186261]</td>
</tr>
</tbody>
</table>
5 Conclusion

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 Basle Committee on Banking Supervision to impose a lower minimum probability of 0.0003 for rare events. The methods of this paper allow 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 use a continuous-time approach based on bootstrapping from the estimated generator and argue that this is superior to a method based on multinomial estimators.

We see that the value corresponds well to the estimate of the default probability of Baa issuers in the stable 5-year period beginning in 1995 but that it is somewhere between our estimate for the Aa and A category in the volatile 5-year period beginning in 1987. Note that if a 97.5% confidence limit was the base of this figure, then the Aa-default probability in the volatile period is close to 3 bps. The multinomial estimator produces a 97.5% quantile of 12 bps in this case. For the speculative grades the multinomial-analysis and the generator based methods produce roughly the same estimates.

The estimators which recognize the dynamics between classes inevitably are linked to a Markov formalism, but one should note that there are many ways in which dependencies can still be captured. If one believes that firms in a class which were downgraded into a class have a higher probability of further downgrades, an effect analyzed for example in Altman and Kao (1992b, Altman and Kao (1992c, Altman and Kao (1992a), Carty and Fons (1993) and Lando and Skodeberg (2002), then one could introduce a separate state for these firms, estimate the generator and do the exact same bootstrap procedure as proposed here. It would simply use a Markov chain with more states. This formulation would reduce the number of events observed for each category and hence very likely show the importance of using the generator-based method of this paper.
References

Altman, E. and D. L. Kao (1992a). *Corporate Bond Rating Drift: An Examination of Credit Quality Rating Changes over Time*.


Figure 4: Histograms of the default probability parameter estimates for the investment grade issuers for the period 1987-1991. The dotted vertical line shows the one-year default probability (PD) corresponding to the ‘true’ value derived from the estimated generator. The estimated generator is then used to simulate 50,000 new histories. From this we get simulated distributions of PD-estimates. The grey line shows the simulated distribution based on estimating the generator and the black line shows the simulated distribution of the PD obtained from the multinomial method. In the top categories, almost all of the mass of the multinomial is concentrated in 0, and the (tiny) rest of the distribution is far off-scale.
Figure 5: Histograms of the default probability parameter estimates for the speculative grade issuers for the period 1987-1991. The dotted vertical line shows the one-year default probability (PD) corresponding to the ‘true’ value derived from the estimated generator. The estimated generator is then used to simulate 50,000 new histories. From this we get simulated distributions of PD-estimates. The grey line shows the simulated distribution based on estimating the generator and the black line shows the simulated distribution of the PD obtained from the multinomial method. Note that the multinomial method performs almost as well in capturing variations of the PD estimate for the lowest categories due to the high number of events.
Figure 6: Histograms of the default probability parameter estimates for the investment grade issuers for the period 1995-1999. The dotted vertical line shows the one-year default probability (PD) corresponding to the 'true' value derived from the estimated generator. The estimated generator is then used to simulate 50,000 new histories. From this we get simulated distributions of PD-estimates. The grey line shows the simulated distribution based on estimating the generator and the black line shows the simulated distribution of the PD obtained from the multinomial method. In the top categories, almost all of the mass of the multinomial is concentrated in 0, and the (tiny) rest of the distribution is far off-scale. In this stable period the events in the top categories are estimated to be very rare.
Figure 7: Histograms of the default probability parameter estimates for the speculative grade issuers for the period 1995-1999. The dotted vertical line shows the one-year default probability (PD) corresponding to the 'true' value derived from the estimated generator. The estimated generator is then used to simulate 50,000 new histories. From this we get simulated distributions of PD-estimates. The grey line shows the simulated distribution based on estimating the generator and the black line shows the simulated distribution of the PD obtained from the multinomial method. Note that in this stable period the multinomial method does not become comparable to the generator method before the B category.
<table>
<thead>
<tr>
<th></th>
<th>Aaa</th>
<th>Aa</th>
<th>A</th>
<th>Baa</th>
<th>Ba</th>
<th>B</th>
<th>Caa</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>-0.072264</td>
<td>0.068650</td>
<td>0</td>
<td>0.003613</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aa</td>
<td>0.018045</td>
<td>-0.186465</td>
<td>0.162405</td>
<td>0.003008</td>
<td>0.001504</td>
<td>0.001504</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0.001410</td>
<td>0.022555</td>
<td>-0.135331</td>
<td>0.093040</td>
<td>0.014097</td>
<td>0.004229</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Baa</td>
<td>0.001182</td>
<td>0.009457</td>
<td>0.068564</td>
<td>-0.206875</td>
<td>0.105211</td>
<td>0.021279</td>
<td>0.001182</td>
<td>0</td>
</tr>
<tr>
<td>Ba</td>
<td>0</td>
<td>0.002925</td>
<td>0.008774</td>
<td>0.152087</td>
<td>-0.356820</td>
<td>0.166711</td>
<td>0.005850</td>
<td>0.020473</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0.002948</td>
<td>0.017690</td>
<td>0.082555</td>
<td>-0.300734</td>
<td>0.129729</td>
<td>0.067813</td>
</tr>
<tr>
<td>Caa</td>
<td>0</td>
<td>0</td>
<td>0.032373</td>
<td>0.097118</td>
<td>0.032373</td>
<td>0.064745</td>
<td>-1.327273</td>
<td>1.100665</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 7: The generator estimated for the US data for the period 1st of January 1987 to 31st of December 1991.
Table 8: The estimated one-year transition probability matrix $P = \exp(\Lambda^*)$ based on the generator estimated for the period 1987 to 1991 for the US data.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>61 68 66 71 66</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>191 177 158 148 149</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>381 381 395 384 377</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.008264</td>
<td>0 0.003876</td>
<td>214 221 242 242 258</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ba</td>
<td>0.038462</td>
<td>0.036145</td>
<td>0.026667</td>
<td>0.0125</td>
<td>0.05</td>
<td>78 83 75 80 80</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.052632</td>
<td>0.047059</td>
<td>0.032258</td>
<td>0.126316</td>
<td>0.116883</td>
<td>76 85 93 95 77</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caa</td>
<td>0.125</td>
<td>0.3</td>
<td>0.285714</td>
<td>0.444444</td>
<td>0.533333</td>
<td>8 10 7 9 15</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9: The one-year default probabilities for the period 1st of January 1987 to 31st of December 1991 estimated by the multinomial method.
Table 10: The generator estimated for the US data for the period 1st of January 1995 to 31st of December 1999.
<table>
<thead>
<tr>
<th></th>
<th>Aaa</th>
<th>Aa</th>
<th>A</th>
<th>Baa</th>
<th>Ba</th>
<th>B</th>
<th>Caa</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>0.931376</td>
<td>0.059901</td>
<td>0.008419</td>
<td>1.60E-05</td>
<td>8.05E-07</td>
<td>9.29E-08</td>
<td>1.13E-08</td>
<td></td>
</tr>
<tr>
<td>Aa</td>
<td>0.007756</td>
<td>0.885900</td>
<td>0.102042</td>
<td>0.004055</td>
<td>0.000232</td>
<td>1.25E-05</td>
<td>1.44E-06</td>
<td>1.85E-07</td>
</tr>
<tr>
<td>A</td>
<td>0.000727</td>
<td>0.029375</td>
<td>0.894370</td>
<td>0.070425</td>
<td>0.004718</td>
<td>0.000338</td>
<td>4.03E-05</td>
<td>6.72E-06</td>
</tr>
<tr>
<td>Baa</td>
<td>0.001341</td>
<td>0.001960</td>
<td>0.077222</td>
<td>0.857772</td>
<td>0.054361</td>
<td>0.006237</td>
<td>0.000897</td>
<td>0.000209</td>
</tr>
<tr>
<td>Ba</td>
<td>0.000119</td>
<td>0.000262</td>
<td>0.014365</td>
<td>0.148721</td>
<td>0.754308</td>
<td>0.074337</td>
<td>0.006282</td>
<td>0.001605</td>
</tr>
<tr>
<td>B</td>
<td>1.38E-05</td>
<td>5.05E-05</td>
<td>0.003018</td>
<td>0.019112</td>
<td>0.072230</td>
<td>0.790555</td>
<td>0.084592</td>
<td>0.030429</td>
</tr>
<tr>
<td>Caa</td>
<td>4.70E-07</td>
<td>1.72E-06</td>
<td>0.000141</td>
<td>0.000911</td>
<td>0.003955</td>
<td>0.081723</td>
<td>0.587025</td>
<td>0.326243</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 11: The estimated one-year transition probability matrix $P = \exp(\Lambda^*)$ based on the generator estimated for the period 1995 to 1999 for the US data.
<table>
<thead>
<tr>
<th></th>
<th>1995</th>
<th>1996</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>Number of issuers at year begin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>Aa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>125</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>424</td>
</tr>
<tr>
<td>Baa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0.041667</td>
<td>0</td>
<td>0.027523</td>
<td>0.049618</td>
<td>0.054745</td>
<td>192</td>
</tr>
<tr>
<td>Caa</td>
<td>0.210526</td>
<td>0.107143</td>
<td>0.142857</td>
<td>0.127660</td>
<td>0.126316</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 12: The one-year default probabilities for the period 1st of January 1995 to 31st of December 1999 estimated by the multinomial method.