The real nature of credit rating transitions[#]

Axel Eisenkopf[†]

Goethe University Frankfurt, Finance Department

March 3rd, 2007, This Version: August 30th, 2007

Abstract

It is well known that credit rating transitions exhibit a serial correlation also known as a rating drift. This is clearly confirmed by this analysis, which also reveals that the credit rating migration process is additionally influenced by three non-observable hidden risk situations. This finding violates the common stationary assumption. The hidden risk situations in turn serially depend on each other in successive observation periods. Taken together, they represent the memory of a credit rating transition process and influence the future transient process. To take this into account, I introduce an extension of a higher order Markov model and a new Markov mixture model. These models allow me to capture these inherent serial correlation structures, to bypass the stationary assumption and to model the process as time non-homogeneous. An algorithm is introduced to derive a single transition matrix with the new additional information. Finally, by means of different CVaR simulations by CreditMetrics, I show that the standard Markov process overestimates the economic risk.

Key Words: Rating migration, rating drift, memory, higher order Markov process, Hidden Markov Model, Double Chain Markov Model, Markov Transition Distribution model, CVaR

JEL classification: C32 ; C41; G32

[†]Axel Eisenkopf, Goethe University Frankfurt, Finance Department, privat: Hundshager Weg 2, 65719 Hofheim, Germany, email: Axel_Eisenkopf@hotmail.com, phone: +49 6192 203686

[#] I thank André Berchthold for his great help on various programming tasks as well as André Güttler and Nobert Jobst for their helpful comments and very constructive discussion.

1 *Introduction*

Markov chains play a crucial role in credit risk theory and practise, especially in estimating credit rating transition matrices. A rating transition matrix is a crucial input for many credit risk models, such as CreditMetrics (see Gupton 1997) and CreditPortfolioView (see McKinsey&Co 1998). The most used basic Markov process is a time-homogeneous discrete time Markov chain, which assumes that future evolution is independent of the past and thus solely depends on the current rating state. The transition probability itself is independent of the time being. Ample empirical research has been done on the validity of these Markov properties and the behaviour of empirical credit rating migration frequencies.

The following non-Markovian properties have been found and confirmed. First, Altman and Kao (1992), Kavvathas, Carty and Fonds (1993), Lucas and Lonski (1992) and Moody's (1993) provided evidence for a so-called rating drift. They all found that the probability of a downgrade following a downgrade within one year significantly exceeds that of an upgrade following a downgrade and vice versa. The likelihood of an upgrade following an upgrade within a year significantly exceeds that of a downgrade following an upgrade. This means that the current rating somehow depends on the previous one, such that prior rating changes may carry predictive power for the direction of future ratings. This was also confirmed by more recent studies by Christensen et al. (2004), Lando and Skødeberg (2002) and Mah, Needham and Verde (2005). They partly used daily time rating data and more advanced statistical tests (such as bootstrapped confidence intervals) for default rates or rating intensities. Furthermore, the downward drift is much stronger than the upward drift, and obligors that have been downgraded are nearly 11 times more likely to default than those that have been upgraded; see Hamilton and Cantor (2004). On the other hand Krüger, Stötzel, and Trück (2005) found a Rating Equalization, i.e. a tendency that corporates receive a rating which they already received 2 or 3 years ago before they were up- or downgraded. In other words a downgrade following an upgrade is more likely than a transition in the same direction and vice versa. This might be driven by the fact that the rating system is based on logit-scores and financial ratios. More sophisticated approaches try to model the process, especially the non-Markovian behaviour, such as Markov mixture models. Frydman and Schuermann (2007) showed empirically that two companies with identical credit ratings can have substantially different future transition probability distributions, depending not only on their current rating but also on their past rating history. They proposed a mixture model based on two continuous-time Markov chains, which outperforms the simple Markov model.[1] The two chains differ in their rates of movement among ratings. Given a jump from one state, the probability of migrating to another state is the same for both chains, because they have the same embedded transition probability matrix. Hence, the model covers the observed heterogeneity in the rate of movement but does not explicitly capture the rating drift. The authors also conditioned their estimation on the state of the business cycle and industry group. However, this does not remove the heterogeneity with respect to the rate of movement. Second, Nickell et al. (2000) and Bangia et al. (2002) provided evidence that rating transitions differ according to the stage of the business cycle. They showed that downgrades seem to be more likely in recessions, whereas upgrades are more likely in expansions. In line with this finding, McNeil and Wendin (2005) used models from the family of hidden Markov models and found that residual, cyclical and latent components in the systematic risk remain even after the observed business cycle covariates are accounted for. Third, Altman and Kao (1992) found that the time since issuance of a bond seems to have an impact on its rating transitions. They analysed a pool of corporate bonds and their ratings from the initial issuance up to 10 years post-issuance and found that the older bonds are, the more likely they are to be downgraded or upgraded in comparison to newly issued bonds. They also came up with an additional ageing effect by showing that within these bond data, defaults peak at the third year and then decrease again. Kavvathas (2000) provided further evidence that upgrade and downgrade intensities increase with time since issuance (except for BBB and CCC

[1] It is an extension of the Continuous Time Mover-Stayer Model; see Frydman, Kadam (2004).

rated bonds regarding the downgrade intensity). Further in the analysis of Krüger, Stötzel, and Trück (2005), the time-homogeneity assumption itself is clearly rejected by an Eigenvalue and Eigenvector comparison. Fourth, Nickell et al. (2000) investigated the issuers' domicile and found for example that Japanese issuers are more likely to be downgraded in comparison to the international average. This was confirmed in later research by Nickell et al. (2002), providing fifth evidence that the issuers' domicile and business line in a multivariate set, along with the business cycle, also impact rating transitions. The credit cycle has the greatest impact thereupon. Finally, Nickell et al. (2000) found that the volatility of rating transitions is higher for banks and that large rating movements are just as likely or more likely for industrials.

In this study, I focus on the credit rating migration evolution, the serial correlation supposed by the rating drift and the time-homogeneity assumption. Hence, a comparison between different Markov models is conducted and the economic impact of all these assumptions is shown. The goal is to account for the non-Markovian findings with respect to the inherent serial correlation and the non-stationary. I introduce two new models, the Markov Transition Distribution model (MTD) for higher order dependencies and the Double Chain Markov Model (DCMM) for non-stationary higher order time series modelled by hidden states. I show that the rating transient behaviour is more complex than is commonly assumed and that serial correlation cannot be captured by simply taking the tuple of the current and the previous ratings into account, as the drift might suggest. In this analysis, I tackle the serial correlation in a dynamic way by taking into account the direction from where the previous rating migrated as well as the whole risk situation from the previous and current ratings. The results reject the stationary assumption in rating migrations and therefore confirm and endorse Lando and Skødeberg's study (2002). Furthermore, I show where this violation comes from and where it takes place. In the peer group, I find that the best model to capture all these issues is the double chain Markov model based on three hidden states. In a time-discrete world, each hidden state depends on its predecessor. This model also incorporates the idea proposed by Frydman and

Schuermann (2006), but supplements it with additional information about the risk intensities, the likelihood of occurrence of the hidden states and the “normal” most probable risk situation, represented by the first hidden state. The model does not assume that the probability of one particular state has to be the same for both chains; or, equivalently, in the manner of the DCMM for the hidden states. Given the underlying data, the process might be better described by introducing three risk situations compared to the two Markov chains. Additionally, the DCMM also covers the memory of a drift, which is not possible with those mixture models.

In the next section, the underlying data are described. In Section 3, the models necessary for the analysis are explained: the Independence Model, the standard time-homogeneous discrete time Markov chain, the higher order Markov models, the hidden Markov chain in brief and the Double Chain Markov Model in detail. In Section 4, the results are presented and validated with some test statistics. Furthermore, with the help of serially correlated random numbers, the difference to the assumed simple correlation structure by the rating drift is shown. Then an algorithm deriving a final matrix that preserves as much information as possible from the risk history is introduced. Finally, with the help of the final matrix, the economic impact is shown by a CreditMetrics simulation; Section 5 concludes.

2 Data description

This study is based on S&P rating transition observations and covers 11 years of rating history starting on 1 January 1994 and ending 31 December 2005. The data are taken from Bloomberg with no information on whether the rating was solicited by the issuer or not.[2] Given the broad range of different ratings for a given obligor, I use a rating history for the senior unsecured debt of each issuer. I treat withdrawn ratings as non-information, hence distributing these probabilities among all states in proportion to their values. In order to obtain an unbiased

[2] See Poon and Firth (2005) or Behr and Güttler (2006) for recent research in this area.

estimation of the rating transitions, I do not apply the full rating scale (including the + and - modifiers of S&P), because the sample size in each category would be too small. Instead, I use the mapped rating scale with 8 rating classes, from AAA to D, throughout.

I apply an international sample of 11,284 rated companies, distributed as 60% from the USA, 4.6% from Japan, 4.6% from Great Britain, 3.3% from Canada, 2.5% from Australia, 26% from France, and 2.4% from Germany. The rest of the sample is distributed over South America, Europe and Asia. The data set consists of 47,937 rating observations (31% upgrades, 69% downgrades). The rating categories D (default), SD (selected default) and R (regulated) are treated as defaults and I find 492 defaulted issuers for S&P. For 82 issuers, more than one default event is obtained, whereby the assumption is adopted that if a company is going into default, it will stay there. I therefore do not allow any cured companies, which means that I keep the current rating history until the first default occurs.

3 Model description

As a starting point, and to show that rating transitions do not follow complete random walks, I introduce the Independence Model. It assumes that each successive observation is independent of its predecessor. Next, the standard model in this area, the discrete time-homogeneous Markov chain in first order, is defined as: let X_t be a discrete random variable taking values in a finite set $N = \{1, \cdots, m\}$. The main property of a first order Markov chain is that the chain forgets about the past and only allows the future state to depend upon the current state. The time-homogeneous assumption states that the probability of changing from one state to another is independent from the time being. In other words, the future state at time $t+1$ and the past state at time $t-1$ are conditionally independent given the present state $X_t = i$. This Markov property indicates that the evolution of the direction is independent of the time being. The transitions between the different ratings states are captured in a time-independent transition probability matrix Q, where each row sums equal to one; see Brémaud (2001). The transition probabilities are then defined as:

$$q_{ij} = P\left(X_t = i_0 \middle| X_{t-1} = i_1\right) \quad \text{where } i_t, \dots, i_0 \in \{1, \dots, m\} \tag{1}$$

As the rating drift might suggest, the most straightforward way to incorporate serial correlation into the estimation process would be to take observations from an obligor's past rating history into account instead of merely conditioning the future rating on the current one. At first glance, the most intuitive way would be to model it as a homogeneous Markov chain in a higher order mode. In a higher order Markov chain of order l, the future state depends not only on the present state but also on $(l-1)$ previous states, which seems to cover the required path dependence assumed in this simple dependence structure. The transition probabilities of a higher order Markov chain are defined as:

$$q_{i_l,\dots,i_0} = P\left(X_t = i_0 \middle| X_{t-l} = i_l, \dots, X_{t-1} = i_1\right) \quad \text{where } i_t, \dots, i_0 \in \{1, \dots, m\} \tag{2}$$

For the purpose of illustration, we will assume a second order Markov chain where $l = 2$ with only three states $(m = 3)$. In this case, the future state $(t+1)$ depends on the current one (t_0) as well as the previous state $(t-1)$; see Pegram (1980). The transition matrix Q is then defined for the above example as:

$$Q = \begin{array}{ccc} & & X_{t+1} \\ X_{t-1} & X_t & \begin{matrix} 1 & \;\;2 & \;\;3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 3 \end{matrix} & \begin{bmatrix} q_{111} & q_{112} & q_{113} \\ q_{211} & q_{212} & q_{213} \\ q_{311} & q_{312} & q_{313} \\ q_{121} & q_{122} & q_{123} \\ q_{221} & q_{222} & q_{223} \\ q_{321} & q_{322} & q_{323} \\ q_{131} & q_{132} & q_{133} \\ q_{231} & q_{232} & q_{233} \\ q_{331} & q_{332} & q_{333} \end{bmatrix} \end{array} \tag{3}$$

For higher- order Markov chains, the number of different states rapidly grows quite large (in our example it would result in $m^l = 3^2 = 9$ states). Particularly if one applies it to credit rating data with at least 8 rating categories, it would expand in a second order mode to a matrix with a dimension of 64x8. Even if it seems to be a straightforward way to consider memory in the estimation process, the huge number of rating combinations necessary for a fully parameterised model is obviously a major drawback. Additionally, matrices with these kinds of dimensions are not feasible as input for other models (e.g. reduced form models), especially since the estimated matrices always result in a sparse matrix. Nevertheless, in order to see whether this estimation technique really best captures the migration behaviour and the serial correlation, I will take it into account.

In order to bypass this problem and to extend the idea of higher order Markov chains, I introduce the Mixture Transition Distribution (MTD) introduced by Raftery (1985) and further developed by Berchtold (1999 and 2002). The major advantage of this model is that it replaces the global contribution of each lagged period to the present by an individual contribution from each lag to the present. In this way, it bypasses the problem of the large number of estimated parameters from the MC_2 but is capable of representing the different order amounts in a very parsimonious way. In general, an *l*-th order Markov model needs to estimate $m^l(m-1)$ parameters, whereas the MTD model with the same order only needs to estimate $[m(m-1)]+l-1$ parameters, meaning that there is only one additional parameter for each lag. In general, the MTD model explains the value of a random variable X_t in the finite set $N=\{1,\ldots,m\}$ as a function of the *l* previous observations of the same variable. This Model allows the effect of each lag on the present to be considered separately Hence, the conditional probabilities are a mixture of linear combinations of contributions to the past and will be calculated as:

$$P\left(X_t=i_0 \middle| X_{t-1}=i_l,\ldots,X_{t-1}=i_1\right)=\sum_{g=1}^{l}\lambda_g P\left(X_t=i_0 \middle| X_{t-g}=i_g\right). \qquad (4)$$

Here λ_g denotes the weights expressing the effect of each lag g on the present value of X (i.e. i_0). This model is especially feasible if the current state does not depend on past *l* states, but the past states influence the future state (with each past state exerting a unique influence).

In order to model the estimation as accurately as possible and to account for (possibly non-Markovian) influencing factors without making explicit assumptions, the last two models are taken from the class of hidden Markov models (HMM). A migration to a certain state can thus be observed without having any assumptions about what really drives the process. However, one important assumption and a major drawback in a HMM is that the successive observations of the dependent variable are supposed to be independent of each other. But in contrast to Christansen (2004), I also specify it in a second order mode and let the hidden states depend on each other within two successive periods. Let us consider a discrete state discrete time hidden Markov model with a set of n possible hidden states in which each state is considered with a set of m possible observations. The parameter of the model includes an initial state distribution π describing the distribution over the initial state, a transition matrix Q for the transition probabilities q_{ij} from state i to state j conditional on state i and an observation matrix $b_i(m)$ for the probability of observing m conditional on state i. Note that also q_{ij} is time independent.[3]

In the last model, I focused on a combination of two models called a Double Chain Markov Model. It was first introduced by Berchtold (1999) and further developed by the Berchtold (2002). This model is a combination of a HMM and a non-homogeneous Markov chain and is thus especially feasible for modelling non-homogeneous time series. I will hereafter abbreviate it as DCMM. Here, the DCMM allows the observations to be dependent on each other, which overcomes the drawback of the standard HMM. The idea of such combinations is not new. First Poritzer (1982, 1988) and then Kenny et al. (1990) combined the HMM with an autoregressive model. Then a similar model was presented by Welkens (1987) in continuous time and by Paliwal (1983) in discrete time. If a time series is non-homogeneous and can be decomposed into a finite set of different risk situations during the time period, the DCMM can

[3] The parameters can be estimated using the Baum-Welch algorithm; see Rabiner (1989). For further details about HMM models, see Rabiner (1989), Cappé, Moulines and Rydén (2005) and MacDonald and Zucchini (1997).

be used to control the transition process with the help of individual transition matrices for each hidden state. Even in the higher order case, the model can handle very complex correlation structures. The discrete time DCMM combines the HMM governing the relation between states of a non-observable variable X_t and a visible non-homogeneous Markov chain governing the relation between successive outputs of an observed variable Y_t. In order to implement memory into the estimation, I allow the hidden states to depend in a higher order mode on each other. Let l denote the order of the dependence between the non-observable *X*'s and let f denote the order of the dependence between the observable *Y*'s. X_t then depends on $X_{t-l},\ldots,X_{t-1}$, whereas Y_t depends on X_t and $Y_{t-f},\ldots,Y_{t-1}$. Using these properties, the DCMM can account for memory in two different ways. First, it allows several hidden states with their respective transition matrices and therefore enables individual risk situations to interact for *l* successive periods with each other, hence incorporating a higher order dependency for the unobserved variable X_t. Second, as in an MC_2, the observable Y_t are allowed to depend on each other for *l* successive periods and therefore permit *l* successive rating observations to depend on each other. Obviously, due to the additional dependence of successive risk situations, the DCMM clearly adds explanatory power to the estimation compared to the MC_2.

A DCMM of order l for the hidden states and of order f for the observed states can be fully described by a set of hidden states $S(X)=\{1,\ldots,M\}$, a set of possible outputs $S(Y)=\{1,\ldots,K\}$, the probability distribution of the first l hidden states given the previous states $\pi=\{\pi_1,\pi_{2|1}\ldots,\pi_{l|1,\ldots l-1}\}$ and an l order transition matrix of the hidden states $A=\{a_{j_l,\ldots,j_0}\}$ where $a_{j_l,j_0}=P(X_t=j_0|X_{t-l}=j_l,\ldots,X_{t-1}=j_1)$. What makes this model very valuable and informative is that for every hidden state, the complete transient behaviour of every possible output state Y is calculated. Finally, for this output, a set of f order transition matrices between the successive observations Y given the particular state of X is calculated and defined as

$$C=\left(C^{(j_0)}\right). \qquad \text{with } C^{(j_0)}=\left\{c^{(j_0)}_{i_f,\ldots,i_0}\right\} \tag{5}$$

where $c^{(j_0)}_{i_f,\ldots,i_0}=P(Y_t=i_0|Y_{t-f}=i_f,\ldots,Y_{t-1}=i_1,X_t=j_0)$

determines the process. In the case of an order amount l > 1, the number of parameters for the transition matrix of the hidden states A and the transition matrix of the observations C can become quite large. In this case, A and each matrix of C can be replaced and approximated by an MTD Model; see Berchtold (2002).

In general, the probability of observing one particular value j_0 in the observed sequence Y_t at time t depends on the value of $X_{t-l},\ldots,X_{t-1}$. The problem is, that in order to initialise this process, l successive values of X_t are needed, but they are unobservable. The DCMM bypasses this problem by replacing these elements with probability distributions where the estimated probability of X_1 is denoted by π_1 and the conditional distribution of X_l given $X_1,\ldots,X_{l-1}$ is denoted as $\pi_{l|1,\ldots,l-1}$.

A DCMM is then fully defined by μ as $\mu=\{\pi,A,C\}$ with $\sum_{g=0}^{l-1} M^g (M-1)$ independent parameters for the set of distributions π, $M^l(M-1)$ independent parameters for the transition matrices between the hidden states A, and $MK^f(K-1)$ independent parameters for the transition matrices between the observations. As μ shows, three sets of probabilities have to be estimated, which is done using the EM algorithm.[4] Because of the iterative nature of the EM algorithm, it is a re-estimation rather than estimation. Instead of giving a single optimal estimation of the model parameters, the re-estimation formulas for π, A and C are applied repetitively, each time providing a better estimation of the parameters. After each iteration of the EM algorithm, the likelihood of the data also increases monotonically until it reaches a maximum. As in the standard EM algorithm, the joint probability of the hidden states (ε_t) and the joint distribution of the hidden states (γ_t) are used. For a higher order mode, π is then estimated as:

$$\hat{\pi}_{t|1,\ldots,t-1}(j_{t-1},\ldots,j_0)=\frac{\gamma_t(j_{t-1},\ldots,j_0)}{\gamma_{t-1}(j_{t-1},\ldots,j_1)} \; . \tag{6}$$

Finally the important higher order transition probabilities between the hidden states are estimated as

[4] This algorithm is also known in speech recognition literature as the Baum-Welch algorithm.

$$\hat{a}_{j_{l-1},\ldots,j_0,j} = \frac{\sum_{t=l}^{T-1} \epsilon_t\left(j_{l-1},\ldots,j_0,j\right)}{\sum_{t=l}^{T-1} \gamma_t\left(j_{l-1},\ldots,j_0\right)} \tag{7}$$

while the higher order transitions between the observations are estimated as

$$\hat{c}_{i_f,\ldots,i_0} = \frac{\sum_{\substack{t=1 \\ Y_{t-f}=i_f,\ldots,Y_t=i_0}}^{T} \sum_{j_{l-1}=1}^{M} \cdots \sum_{j_1=1}^{M} \gamma_t\left(j_{l-1},\ldots,j_0\right)}{\sum_{\substack{t=1 \\ Y_{t-f}=i_f,\ldots,Y_{t-1}=i_1}}^{T} \sum_{j_{l-1}=1}^{M} \cdots \sum_{j_1=1}^{M} \gamma_t\left(j_{l-1},\ldots,j_0\right)}. \tag{8}$$

After the model is estimated, one can search for the optimal sequence of the hidden states in order to maximise the conditional probability

$$P\left(X_1,\ldots,X_T \middle| Y_{-f+1},\cdots,Y_T\right) \tag{9}$$

and equivalently the joint probability

$$P\left(X_1,\ldots,X_T, Y_{-f+1},\cdots,Y_T\right). \tag{10}$$

This can be done with the Viterbi algorithm, which is an iterative dynamic programming algorithm for indicating the most likely sequence of hidden states – known as the Viterbi path. The goal of the algorithm is to find in an efficient way the best hidden path sequence with the help of hidden Markov models (see 1973). To achieve this, the Viterbi algorithm is run separately upon every single sequence, giving for each obligor the best non observable path of hidden states.

4 Results

4.1 In-sample assessment of various accuracy measures

As a starting point, the Independence Model is calculated, then the homogeneous Markov chain of orders 1 and 2, the MTD, a HMM with 3 hidden states in first and second orders and finally, different combinations of the DCMM model. In order to have a quantitative criterion for deciding which stochastic model fits the data best, the accuracy measures log likelihood, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) are computed. For the purpose of comparison, the initial *f* observations are dropped. Generally, this is based on the model order of the time series in order to have the same number of elements (59,969) in the log likelihood of each model. In other words, let $Y_{-f+1},\ldots,Y_0$ denote the first observations, then $Y_1,\ldots,Y_T$ are the observations used in the computation of the log likelihood. Here, the standard model (MC_1) is set as the benchmark model. The analysis shows that the most significant model is a Double Chain Markov Model (DCMM) with 3 hidden states in a second order dependency. The outcome results in the desirable dimension of a homogeneous first order Markov chain. Therefore, it will hereafter be labelled as DCMM_3_2_1.

The first model, the Independence Model, assumes that each successive observation is independent of its predecessor. As expected, this model strongly favours the rejection of the MC_1, which clearly confirms that rating transitions do not follow a random walk but are conditional on “something” previous (see Table 1 for the performance results). As described earlier, the most straightforward way to incorporate memory into the estimation would be to increase the order of a first order Markov chain (MC_1) to a second order Markov chain (MC_2). The results clearly show an improved accuracy measure for the MC_2, indicating that a dependency in successive rating observations does indeed exist. The Log Likelihood drops from -34,063 to -31,391 and the AIC as well as the BIC reduces from 68,211 to 63,038 and from 68,589 to 64,190 respectively. Based on a Likelihood Ratio test, Krüger, Stötzel, and Trück (2005) clearly confirmed this results for a second-order Markov chain. However the hypothesis whether a third order Markov property leads to even better results were rejected. Keeping this in mind and since a third order Markov chain would generate a very sparse matrix, it would not make any sense to compare it with the other models. The second order MC also needs to estimate a huge number of state combinations and would additionally result in a sparse matrix. However, as described earlier, the MTD_2 model has significantly fewer parameters to estimate (42) compared to the MC_2 (128). Furthermore, this model confirms that it is not only the current rating that determines the future rating but also its history. The log likelihood reduces to

-32,837 compared to -34,063 from the MC_1, the AIC to 65,758 compared to 68,211 of the MC_1 and the BIC drops from 68,589 to 66,136. It is obvious that the solely lagged rating one period before definitely influences the future rating, but with less informative power than in combination with the current rating, as with the MC_2.

At this point, it would be interesting to know whether the rating itself has the sole or most predictive power or whether other influencing factors (like the complete risk situation driven by several unobservable issues (e.g. the economy) in a non-stationary world) contribute significantly to explanatory power. In this sense, the class of higher order hidden Markov models provides another solution, as they do not make any assumptions as to what drives the output. In the case of the HMM (as expected from the independence assumption, which was already disproved through the results of the MC_2 and MTD), the HMM without any explanatory covariates is hardly a good model for the underlying data and application to credit rating migration data. The log likelihood as well as the AIC and BIC are closer to the Independence Model than to the MC_1. Interestingly, a HMM with three hidden states performs much better than a HMM with two states with an AIC of 141,216 and BIC of 141,639 compared to an AIC of 171,966 and BIC of 172,155. This can be seen as a further indication that a credit rating transition process is driven by three different unobservable drivers or situations. They may themselves be a combination of several risk dimensions, like the economic cycle, or even the previously described non-Markovian properties.

In comparison the DCMM, it seems obvious that the MC_2 can only partly model the correlation structure, since the DCMM is much more able to fit the data. The DCMM with three hidden states in a second order dependence structure beats every other model. Compared to the MC_1, the BIC was reduced by about 8,772 (12.8%); the AIC and the log likelihood were also reduced by significant amounts (9,762 (14.3%) and 4,991 (14.6%), respectively) (see Table 1).

In order to really know how many hidden states are driving the process, I also compute the DCMM with 1 up to 5 hidden states, but three hidden states clearly dominate every combination of hidden states. Next, in order to have a closer look at the correlation structure itself, I compute several DCMM models with different order amounts of the hidden states. In this way, the estimation of the transition probabilities are conditioned on different combinations of successive risk and rating situations in order to find the best suitable memory history driving the transient process of an obligor's rating history. I find that regarding the log likelihood, BIC and AIC, the DCMM with three hidden states in second order dependence structure clearly beats every other order combination. In order to facilitate comparison, I again drop the first l observations from each time series. If one increases the order amount to 3 and hence considers a risk situation of one additional period and one additional rating compared to the DCMM_3_2_1, the log likelihood increases from -29,066 to -29,132, whereas the AIC and BIC increase from 58,436 and 59,776 up to 58,673 and 60,472, respectively.[5] Even combinations of more than three hidden states with an order higher than two are beaten by the DCMM_3_2_1. Finally, as described above, the DCMM is capable of estimating the matrix of the hidden state as well as the matrices of the observations with the MTD. Even calculations with this approximation clearly support the finding that the DCMM_3_2_1 fits rating transition data best.

In summary, simply taking two successive rating observations into account and allowing this combination to determine the next future rating is not the best way to assess the oft-described rating drift. It is clearly one part of the memory and adds predictive power (as already indicated by the MC_2). Therefore, the best and most accurate way would be to consider two successive rating observations along with two successive complete risk situations, with individual risk intensities driving the process. By using this process, I also circumvent the resulting sparse matrix, which is clearly one of the MC_2's shortcomings. This approach confirms and particularly extends the results of Crowder, Daris and Giampierin (2004) with respect to their postulation that the process is driven by just two states, a risky state and non-risky state.

[5] Note that the figures of the DCMM in second order (Table 1) differ since one additional observation was dropped.

4.2 Estimation results: transient behaviour and transition matrices

To obtain an idea of how the transient behaviour and the correlation structure really behave and interact between the hidden states, it is necessary to focus more closely on the results of the DCMM_3_2_1 (see Table 2-3). As shown by the first hidden state distribution (π_1), the starting state in the process of credit rating migrations is, with a probability of 66.23%, the first hidden state and with a probability of nearly 30.27%, the third hidden state. With a probability of 3.51%, the second hidden state is unlikely to be the starting hidden state. Conditional on the previous hidden states, the distribution of the next hidden state distribution $(\pi_{2,1})$ clearly shows that if the first and second hidden states are the current states, it is very likely (95.33% and 100%, respectively) that the process will return to the first hidden state. The situations looks different if the process is currently in the third hidden state. Since this was not unlikely (30.27%), one can see that there is a reasonably good chance that the third hidden state (30.71%) will prevail. Again, the first hidden state is likely to dominate the process again (69.29%) (see Table 2).

The higher occurrence probability of the first hidden state indicates that the chance of being in a stationary world is very high, but that the probability of transitioning to the second or third hidden states, each with completely different risk intensities lies considerably in the future. In order to quantify how the hidden states depend on each other, a second order transition probability matrix of the three hidden states (Table 3) is computed. Again, as previously described, if in (t_0) the first hidden state is currently the actual one, it is very likely that it will also be the active one in the future state (t_{+1}) regardless from which hidden state in the previous period (t_{-1}) it migrates. However, if in the previous period the second hidden state was the active one, there is a 22% chance of migrating to the third hidden state in (t_{+1}) and an 8.4% chance of staying in the second one. The picture looks much different if the second or third hidden state is active in (t_0). In this case, if either one migrated from the first hidden state, it is almost certain that the process will revert back to the first hidden state. What is interesting to note is that the future transient behaviour of the second and third hidden states is almost identical conditional on the previous hidden state. In both cases, if the currently second or third hidden state migrated from the first hidden state, it is almost certain that the process will revert back to the first hidden state in (t_{+1}). On the other hand, if the process migrated from the third hidden state, there is no uncertainty that the process will occupy the second hidden state in (t_{+1}). Here one can clearly see that a rating history is not necessarily a stationary process, since the origin of the current hidden state -- and thus the corresponding previous risk situation -- definitely matters.

A change of hidden states in a process would not be remarkable if their associated risk intensities were also to stay the same. As previously described in the model, an associated individual transition matrix will be estimated for each hidden state (Tables 4-6). A comparison between the matrix estimated by the MC_1 (Table 7) and the three matrices shows tremendous differences in the distribution of the probability mass (see Table 8). This also confirms the finding of Krüger, Stötzel, and Trück (2005), hence they found that the entire transition probability matrix vary over time. The transition probability matrix for the first hidden state (Table 4) looks quite similar to the transition probability matrix normally derived from the MC_1. In other words, being in the first hidden state would result in a nearly normal risk situation. However, the risk situation in the first hidden state is more stable because more probability mass is located at the diagonal compared to the matrix estimated by the MC_1. On the other hand, the probability of defaulting increases slightly for every current rating. In contrast, the matrix of the second hidden state (Table 5) shows, with the exception of the default column, an absolute moving character (as proposed in Frydman and Schuermann’s Mover-Stayer Model (2006)). The DCMM provides additional information about the direction in which the rating is likely to move. For the investment grade area down to rating grade *A*, one can clearly see that the trend has a downward slope, meaning that the better a rating is, the more likely it will be downgraded. By contrast, in the speculative grade area from rating *BBB* down to the rating *CCC*, it is significantly more likely that the rating will be upgraded next. In other words, the second hidden state can be seen as a "mover state" with a "threshold" at rating *BBB*. This transient behaviour is absolutely comprehensible, since it demonstrates the common understanding of rating movements across the rating grades. However, with respect to the matrix of the third hidden state (Table 6), it can be seen as a very stable "stayer state" (as suggested by the second Markov model in the Mover-Stayer Model). Compared to this model, it also provides additional information about the risk intensities, the likelihood of occurrence of the hidden states and the "normal" most probable risk situation, represented by the first hidden state. A further important enhancement offered by the DCMM is that it does not assume that the probability of entering one state has to be the same for both chains; instead, these probabilities are determined by a separate transition probability matrix. The DCMM also covers the memory of a drift, which is not possible in this fashion with the mixture models.

4.3 Time dependent occurrence of the hidden states

As described earlier, the hidden states might be driven and influenced by several dimensions, such as the economic cycle and other exogenous effects. And for each sequence of observations, the most likely sequence of hidden states, known as the Viterbi path, can be estimated. In order to see the evolution of the hidden states in previous years, Picture 1 shows the hidden state distribution across the observation period. This distribution confirms that the most likely state will be the first hidden state. Interestingly, in 1997, credit rating transitions were as likely to be driven by the third hidden state as by the first hidden state in the underlying database. Starting in 1998, the second and third hidden states began to alternate in terms of their influence on the process every two years; every two successive years were dominated by one or the other hidden state. In other words, the migration volatility might have been higher and influenced by the second hidden state in 1998, 1999, 2002 and 2003. Additionally, the speculative grade issuer was more likely to upgrade, whereas the investment grade issuer faced a rating deterioration. In

1997, 2000, 2001, 2004 and 2005, however, the third hidden state dominated the second hidden state. Particularly in combination with the normal first hidden state, the transient behaviours were more stable and less volatile during these years.

4.4 Validation

In order to prove that the second order transient behaviour of the hidden states is not caused by spurious correlation, I calculated Cramer's V statistic (see Cramer 1999) for the hidden variables. It is a measure for the association between variables. The closer Cramer's V is to zero, the smaller the association between the hidden variables is. Here the three hidden states (with a value of 0.1256) do not depend very strongly on each other, which deflate any suspicion of a spurious correlation between the transition matrices of the second and third hidden states stemming from the correlated hidden states themselves.

Now that the inherent correlation structure and the transient property have been examined, it is important to investigate the estimation accuracy of the DCMM. To this end, Theil's U, which is the quotient of the root mean squared error (RMSE) of the forecasting model and the RMSE of the naive model, will be calculated (see Theil 1961). Hence, the results are compared against the "naive" model, which consists of a forecast repeating the most recent value of the variable. The naive forecast is a random walk specified as:

$$y_t = y_{t-1} + \varepsilon_t \quad \text{where} \quad \varepsilon_t \sim i.i.d.\ N\left(0, \sigma^2\right). \tag{11}$$

Behind this notion is the belief that if a forecasting model cannot outperform a naive forecast, then the model is not doing an adequate job. A naive model predicting no change will give a U value of 1, and the better the model, the closer Theil's U will be to 0. It is computed for the hidden states, resulting in a value of 0.0327, as well as for the observable variable, where I obtain a value of 0.0093. Both values indicate that the DCMM fits the data set nearly perfectly regarding the observable variables and, even more importantly, the hidden states as well. This should also be taken as evidence of the high explanatory power of the DCMM. In contrast, the single HMM with three hidden states performs much worse, with a value of 0.9021, which is nearly a completely naive guess. The value for the observed variables, 0.5551, is tremendously better but is still far less accurate than the one given by the DCMM. These differences clearly show that the DCMM's property of allowing dependence structures between the observations should be considered in estimating transition probabilities. This is not surprising, since this fact was already shown by the MC_2.

4.5 Out-of-sample performance

In order to ensure that these relationships are not the result of spurious correlations, the calculations should be repeated with both an out-of-sample and an in–the-sample data set. As can be seen in Table 1, the number of parameters of the MC_2 and DCMM_3_2_1 is too high to obtain unbiased estimates on the resulting small sub-samples.

A robustness check to prove the complex correlation structure itself is conducted with random numbers, once generated with serial correlation and once without. The serially correlated random numbers are calculated as

$$Y_t = ((\rho) \cdot Y_{t-1}) + (Y_{t-1} \cdot \varepsilon_t) \tag{12}$$

where ρ denotes the correlation coefficient and is assumed to be 40%. The random numbers themselves are assumed to be normally distributed and are scaled into the same 8 state rating scale $\{1,2,\ldots,8\}$ used in the original rating data. In order to make it comparable to the real rating data, the number of components in the log likelihood needs to be the same. Therefore, for each company, a random start rating is simulated. Afterwards, each company is assigned a sequence of random numbers equal in length to the number of rating observations in the original data set. Thus, the sample structure remains the same as in the original data. In the case of uncorrelated random numbers, the MC_1 performs best in terms of the AIC and BIC. In contrast to serially correlated random numbers, the MC_2 clearly beats the MC_1, which supports the idea that the MC_2 fits a simple serial correlated data set best, as supposed with the rating drift.[6] Even the DCMM_3_2_1 supports this idea, since the AIC and BIC beat the MC_1 but interestingly not the MC_2. On the other hand, the calculation based on the real rating data looks different, i.e. favours the DCMM_3_2_1 and hence confirms that the correlation structure in real credit rating data is much more complex than assumed and that the memory is not best captured by simply taking the combination of the current and previous ratings into account.

Deriving the final matrix

As previously shown, the memory information and the transition probabilities of the hidden states are spread over three transition probability matrices. At this point, the optimal way to handle the information would be a tractable matrix in the standard 8x8 dimension with the inherent transient and serial correlation structure. To derive such a matrix, a weighting approach is introduced. This approach is then also feasible for the DCMM model information in other areas (e.g. it is well known that the rating drift in Structured Finance is also evident and even stronger; see Cantor and Hu (2003)). The resulting matrix should approximate the non-stationary process and preserve its memory information. Since the rating migration process follows a non-homogeneous process, the new matrix will also be based on a non-homogeneous process. The new non-homogeneous transition probability matrix's first column would contain not only the current state (X_t) but also a functional relationship of the risk intensities in various possible risk situations. The following information is needed: the individual transition probability matrix $\{P_1, P_2, \ldots, P_n\}$ for each of the hidden states $h_1, \ldots, h_m$ (see Tables 4-6), the second order transition probability matrix of the hidden states (see Table 3) and information about the relative occurrence of the hidden states across the rating classes (see Table 9). Since

6 In support of the idea that the MC_2 captures simple serial correlation structures, BIC and AIC significantly increase if the calculations were based solely on random numbers without any serial correlation.

the second order transition matrix of the hidden state is used, memory is added to the process by allowing the future state to depend on the current and previous period risk situations. After the inputs are defined, the weighting approach is initiated by multiplying the elements for each hidden state of the second order transition probability matrix $ph_{i,\ldots,i_0} = P(H_i = i_0 | H_{t-l} = i_l, \ldots, H_{t-1} = i_1)$ by the corresponding relative occurrence frequency of the respective hidden state $prf_{ij} = P(X_t = i_0 | X_{t-1} = i_1)$. For m hidden states, it results in m column vectors (V) of size m^m. The resulting m vectors (V) are then summed together as $VW = \sum_i^m V_i$, where each element in the row vector is denoted as $\{v_1, v_2, v_3, \ldots, v_m\}$. Again, the new vector has the size m^m and is next divided sequentially into m buckets of size m starting from the first entry v_1. Now each bucket contains m entries, which are then summed together and denoted as ϖ_i. These will be the weighting factors for the transition probabilities of the respective hidden states, where ϖ_1 corresponds to the first hidden state, ϖ_2 corresponds to the second hidden state and so on $\{\varpi_1, \varpi_2, \ldots, \varpi_m\}$. Finally, the entries of the new matrix are calculated as the product of the weighting factors for the respective hidden state times the corresponding entries of the respective transition probability matrix $\{P_1, P_2, \ldots, P_n\}$ and are then summed together.

$$p_{n+1_{ij}} = \varpi_{1_{ij}} p_{1_{ij}} + \varpi_{2_{ij}} p_{2_{ij}} + \ldots + \varpi_{n_{ij}} p_{n_{ij}} \tag{13}$$

This is done for every entry in the new matrix. Finally, to ensure a row sum equal to one (as prescribed by the property of a stochastic matrix), each of the matrix's entries is divided by its respective row sum.

For purposes of illustration, let us consider our case with three hidden states and a situation in which it retains a rating of *AAA*. For the first hidden state, I start by multiplying each element of the first column $ph_{ij} = P(H_t = i_0 | H_{t-1}, H_{t-2})$ of the second order transition probability matrix for the hidden states by the relative frequency of the first hidden state for rating grade *AAA* (0.7318) and by the transition probability of the respective matrix *P1* (0.8677), resulting in the vector $V_1 = \{$0.5668492, 0.4402336, 0.610028, 0.6349829, 0, 0000635, 0.6349829, 0, 0}’. This is repeated for the remaining two hidden states in order to obtain two further weighted probability vectors, with $V_2 =$ {0, 0, 0, 0, 0, 0, 0, 0, 0}’ and $V_3 =$ {0.0212899, 0.0443077, 0.0071099, 0, 0.1986, 0, 0, 0.1986, 0}’. In the next step, the three vectors are summed together, resulting in vector *VW*={0.5881391, 0.4845413, 0.6171379, 0.6349829, 0.1986, 0.0000635, 0.1986, 0}’. Since we have 3 hidden states, the vector *VW* is split with its 9 entries into three buckets containing three entries each. The entries of each bucket are then summed together and divided by the total vector sum of *VW*. Now we have three weighting factors for the respective hidden states: $\varpi_1 = 0.436769$, $\varpi_2 = 0$ and $\varpi_3 = 0.248308$. In the last step, the weighting factors are each multiplied by the respective transition probability of the corresponding transition probability matrix $P_1 - P_3$ and then finally summed together. The derived transition probability expresses the weighted probability of the final matrix, which is in our example equal to (=0.436769*0.8677 + 0*0 + 0.248308*1 = 0.68508).

The final matrix (Table 10) exhibits the information of the transient behaviour of all three hidden states and the inherent serial correlation. Due to the second hidden state, the main diagonal shows lower probabilities than MC_1, the matrix for hidden state one (*P1*) and for hidden state three (*P3*). The probability mass is shifted by the second hidden state from rating state *AAA* to state *A*, towards a lower rating grade and from rating states *BBB* to *CCC* towards better rating states. This again is the idea of the mover characteristic.

4.6 Economic impact

After analyzing the transient behavior of credit rating migrations and their inherent correlation structure, it is important to obtain information about the economic impact. Since the class of reduced form models uses migration matrices as its main input, I conduct the analysis using the CreditMetrics model. Because the economic impact of transition probabilities with memory information from the successive risk situations is of major interest, a uniform correlation structure is assumed. Regarding Gupton (1997), the correlation is set equal to 0.20, which should be a reasonable value. Furthermore, I conduct simulations with asset correlations of 10%, 30% and 40%. The LGD is set equal to 45%. The value of the loan in one year for each rating is then computed as

$$V_t = EAD_t \bullet e^{(-(r_t + CS_t)t)} \tag{14}$$

where t denotes the time and is set equal to one year, r denotes the riskless rate, which is assumed to be 3%, and the EAD denotes the commitment. The credit spread with PD as the probability of default s is denoted by CS and calculated as:

$$CS_s = -\left(\ln\left(1 - PD_t\right)\right)/t \tag{15}$$

I set up a hypothetical portfolio consisting out of 500 obligors with a total value of €500 Mio. For simplicity's sake, the single exposures are assumed to be uniformly distributed with a net commitment of €1 Mio, and each obligor has only one loan. In order to be as realistic as possible, I apply a hypothetic portfolio composition taken from a large German bank portfolio. It consists of 1.2% exposure in rating class AAA, 9.6% in AA, and 16.4% in A, 41.8% in BBB, 27.2% in BB, 3.4% in B and 0.4% in CCC.

To obtain information regarding the economic impact, the simulation is conducted once with the matrix estimated by the MC_1 and once with the finally derived matrix. The simulations clearly show that the MC_1 overestimates the risk compared to a simulation based upon the information provided by the DCMM. Based on a confidence level of 99.0% (99.9%), the simulation conducted with the matrix from the MC_1 allocates a CVaR of €18,915,573 (€20,957,447), while the one generated by the finally derived matrix, including the inherent information of the DCMM, allocates a CVaR of €15,902,671 (€16,806,754). This result is in line with the observation that three different risk situations are obviously driving the transition. The first, most dominant hidden state shows a risk situation similar to the one proposed by the MC_1. The second hidden state is clearly moving, which results in a higher migration risk, but since the portfolio composition consists of 72.8% ratings below *A* and the second hidden state shows an upgrade trend, the result is very reasonable. In other words, within this portfolio composition, the second hidden state reduces the risk by moving to upgrade rating qualities. The third and even more likely state reduces the migration risk, since it is an absolute stayer state. Overall, it results in a lower risk situation as shown by the lower CVaR. Even if I assume that the exposures are equivalently distributed across the rating states, the MC_1 still overestimates the risk. In this case, for a portfolio with the same face value and the simulation based on the MC_1 matrix, I obtain a considerably higher CVaR (€38,796,557) compared to the one based on the information from the DCMM (€33,864,380).

In order to see what impact these transition probabilities might have under different correlation assumptions, I simulate the CVaR with the different correlations 0.1, 0.3 and 0.4 again. Even with these different correlation assumptions, the MC_1 clearly overestimates the risk based upon the rating observations within the time period between 1994 and 2005.

5 Conclusions

Credit rating transition probabilities are commonly estimated by a discrete time time-homogenous Markov chain. A large set of non-Markovian behaviors has already been discovered and unequivocally acknowledged in the literature. One very popular behavior is the so-called rating drift.

The goal of this paper is to overcome these non-Markovian behaviors, to account especially for the truth serial correlation and try to find out what really influences the transition probability. I introduce two new models, the Mixture Transition Distribution model (MTD) and the Double Chain Markov Model (DCMM), into the credit rating transition estimation area. I analyse these along with the most commonly used models in order to ascertain which model best fits the transient behaviour of a representative credit rating data set. The two new models perform best. In terms of AIC and BIC the MTD clearly outperforms the standard Markov chain (MC_1) but not the second-order Markov chain (MC_2). On the other hand, in light of the resulting sparse matrix from the MC_2 and the high number of parameters it requires, the Mixture Transition Distribution model is preferable. The DCMM beats every other model setting and furthermore discovers and emphasizes the true character of credit rating transitions. It is thereby obvious that the transition probability from one observation period to the next is not well captured by merely looking at a certain point in time and considering the frequencies of transitions one period later, as is done in the standard discrete time Markov chain. The underlying process is actually driven by three completely different risk situations determined by three hidden states instead of an average over the whole observation period. Each risk situation is determined by its individual risk intensity as given by a complete transition probability matrix. In this sense, the commonly assumed time-homogeneous assumption can also be rejected. The first and most probable hidden state can be summarised as a normal risk state with transition probabilities similar to the ones already known. However, the second hidden state can be seen as a "mover state" with a complete reversal trend depending on whether the obligor is rated in an investment grade area or in the speculative grade area. However, if an obligor is rated with a speculative grade rating, an upgrade trend is to be expected, whereas in the speculative area, the corporation would face a downgrade of its rating. The third hidden state is a very stable "stayer state" in which no migration risk seems likely. The serial correlation assumed by the well-known rating drift is clearly confirmed. I show that the correlation structure is not best captured by two successive rating observations, as is commonly assumed by the rating drift, but by the addition of two successive hidden risk situations. Therefore the memory of a credit rating transition process is determined by two successive risk situations with possible different risk intensities along with two successive rating observations. To combine the information of the process with three risk situations into one transition probability matrix, a weighting algorithm is introduced to incorporate the information from the DCMM output. The resulting matrix should be much more able to capture the true transient behaviour of credit rating transitions. Furthermore, several

CVaR simulations based on this weighted matrix are compared to simulations based on rating transition matrices calculated with the standard time-homogeneous discrete Markov model. These simulations show that in light of risk capital depending only on the current observation period, credit risk is, on average, clearly overestimated. As a result, not only the current rating and risk situation should be considered but also the previous one.

References

Altman, E.I., Kao, D.L., (1992). The *implications of corporate bond ratings drift.* Financial Analysts Journal 48. 64-67.

Asarnow, E., Edwards, D., (1995). *Measuring Loss on Defaulted Bank Loans: A 24 Year Study.* Journal of Commercial Lending: 11-23.

Aurora, D., Schneck, R., Vazza, D., (2005). *S&P Quarterly Default Update & Rating Transitions*. Global Fixed Income Research

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.

Behr, P., Güttler, A. (forthcoming). *The Informational Content of Unsolicited Ratings*. Journal of Banking and Finance.

Berchtold, A. (2002). *Higher-Order Extensions of the Double Chain Markov Model*. Stochastic Models. 18 (2). 193-227.

Berchtold, A., Raftery A.E., (2002). *The Mixture Transition Distribution Model for High-Order Markov Chains and Non-Gaussian Time Series. Statistical Science.* 17 (3), 328-356.

Berchtold, A. (1999). *The Double Chain Markov Model*. Communications in Statistics: Theory and Methods. 28 (11). 2569-2589.

Brémaud, (2001). *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and queues*. Springer

Cantor, R., Hu, J., (2003). *Structured Finance Rating Transitions: 1983-2002 Comparisons with Corporate Ratings and Across Sectors.* Moody's Investor Service, Global Comment.

Cappé, O., Moulines E., Rydén T (2005). *Inference in Hidden Markov Models.* Springer

Cramér, H., (1999). *Mathematical Methods of Statistics.* Princeton University Press.

Christensen, J.H.E., Hansen, E., Lando, D., (2004). *Confidence sets for continuous–time rating transition probabilities*. Journal of Banking and Finance 28, 2575–2602.

Frydman, H., Schuermann T., (2006). *Credit Rating Dynamics and Markov Mixture Models.*

Frydman, H., (2005). *Estimation in the Mixture of Markov Chains Moving with Different Speeds*. Journal of the American Statistical Association. 79. 632-638

Frydman, H., Kadam A., (2004). *Estimation in the Continuous time Mover-Stayer Model with an Application to Bond Ratings Migration.* Applied Stochastic Models in Business and Industry, 20. 155-170.

Giampieri, G., Davis M., Crowder M., (2005). *Analysis of Default Data Using Hidden Markov Models*. Quant. Finance, 5 27-34

Giampieri, G., Davis M., Crowder M., (2004). *A Hidden Markov Model of Default Interaction.* Working Paper, Department of Mathematics, Imperial College, London

Gupton, G.M., Finger C.C. and M. Bhatia, (1997), *CreditMetrics Technical Document*, J.P. Morgan.

Hamilton, D., Cantor R., (2004). *Rating Transitions and Defaults Conditional on Watchlist, Outlook and Rating History*. Moody's Investor Service, Special Comment, February.

Lucas, D., Lonski. J., (1992). *Changes in corporate credit quality 1970-1990*. Journal of Fixed Income 1. 7-14.

Kavvathas. D., (2000). *Estimating credit rating transition probabilities for corporate bonds.*

Kenny, P.M., Lenning P. Mermelstein., (1990). *A linear predictive HMM for vector valued observations to speech recognition.* IEEE Transitions on Acoustics, Speech and Signal Processing. Vol 38 (2). 220-225

Krüger, U., Stötzel, M., Trück S., (2005). *Time series properties of a rating system based on financial ratios.* Deutsche Bundesbank, Discussion Paper, Series2: Banking and Financial Studies, No 14/2005

Lando. D., Skødeberg. T.M., (2002). *Analyzing rating transitions and rating drift with continuous observations.* Journal of Banking and Finance 26. 423-444.

MacDonald LL., Zucchini W., (1997). *Hidden Markov and other Models for discrete-valued Time Series.* Chapman and Hall /CRC

Mah S., Needham C., Verde M., (2005). *Fitch Ratings Global Corporate Finance 2004 Transition and Default Study*. Fitch Ratings, Credit Market Research.

McKinsey&Co. (1998). *CreditPortfolioView - Approach Document.*

McNeil, A.J., Wendin, J., (2005). *Dependent Credit Migrations*. Department of Mathematik, ETH Zürich

Nickell. P., Perraudin. W., Varotto. S., (2000). *Stability of transition matrices*. Journal of Banking and Finance 24. 203-227.

Paliwal K.K., (1993). *Use of temporal correlation between successive frames in hidden Markov models based speech recognizer*. Proceedings. ICASSP. Vol. 2. 215.218

Pegram, G. G. S., (1980). *An autoregressive model for Multilag Markov chains*. J. Appl. Probab. 17 350-362.

Poritz, A.B., (1988). *Hidden Markov models: A guided tour.* Proceedings ICASSP. Vol.1. 7-13

Poritz. A.B., (1982). *Linear predictive hidden Markov models and the speech signal*. proceedings ICASSP. 1291-1294

Rabiner, L.R., (1989). *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.* Proc. IEEE. 77. 257-286.

Raftery, A.E. (1985). *A model for high-order Markov chains*. Journal of the Royal Statistical Society B, 47 (3), 528-539.

Raftery, A.E., (1985). *A new model for discrete-valued time series: autocorrelations and extensions.* Rassegna di Metodi Statistici ed Applicazioni, 3-4, 149-162.

Theil, H., (1961). *Economic Forecasts and Policy*. North-Holland Publishing Company, Amsterdam.

Trück, S.. Rachev S., (2006). *Changes in Migration Matrices and Credit VaR – a new Class of Differences Indices*

Trück, S., Rachev S., (2003). *Credit Portfolio Risk and PD Confidence Sets through the Business Cycle*

Wellekens, C.J., (1987). *Explicit time correlation in the hidden Markov models for speech recognition.* Proceedings ICASSP. 384-386.

Table 1: Qualitative performance of the models

The performance and the fit of the different models to the data is determined by the accuracy measures log likelihood, AIC and BIC. Here MC_# denotes the standard Markov Chain with order of # and HMM_#_# as the hidden Markov Model with # number of hidden states in a # order dependency. The Double Chain Markov model is denoted by DCMM with # hidden states in # order dependency with an output in a # dimension.

	Parameter	Log Likelihood	AIC	BIC
Independence model	7	-105,948	211,911	211,974
MC_1	42	-34,063	68,211	68,589
MC_2	128	-31,391	63,038	64,190
HMM_2_1	17	-79,643	159,322	159,475
HMM_3_1	29	-73,244	146,547	146,808
HMM_3_2	47	-70,560	141,216	141,639
MTD_2	42	-32,837	65,758	66,136
DCMM_2_2_1	91	-32,676	65,535	66,354
DCMM_3_2_1	**152**	**-29,072**	**58,449**	**59,817**

Table 2: First hidden state distribution π_1 and the conditional distribution $\pi_{1,2}$ of the second hidden state

This table shows the probability of which of the three hidden states might be the starting state π_1 in the rating sequence of each obligor and the conditional distribution $\pi_{1,2}$ of the further hidden states in the process given the first hidden state.

state distribution	States	1	2	3
π_1	1	0.6623	0.0351	0.3027
	1	0.9533	0.006	0.0407
$\pi_{1,2}$	2	1	0	0
	3	0.6929	0	0.3071

Table 3: Second order transition matrix of the hidden states

This table shows the transition probabilities of the hidden states in a second order dependency structure indicating how likely one of the three hidden states will be given the current one and the previous one.

		t+1	t+1	t+1
t-1	t0	1. hidden state	2. hidden state	3. hidden state
1	1	0.8927	0.0001	0.1072
2	1	0.6933	0.0837	0.2231
3	1	0.9607	0.0035	0.0358
1	2	1	0	0
2	2	0	0	1
3	2	0.0001	0.9999	0
1	3	1	0	0
2	3	0	0	1
3	3	0	1	0

Table 4: DCMM_3_2_1 Transition Probability Matrix for hidden state 1

This table shows transition probabilities calculated by the DCMM for the first hidden state based on a S&P issuer rating history for 1994 to 2005.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.8677	0.1249	0.0057	0.0016	0.0000	0.0000	0.0000	0.0000
AA	0.0040	0.8988	0.0897	0.0053	0.0005	0.0015	0.0000	0.0002
A	0.0009	0.0212	0.9076	0.0649	0.0028	0.0009	0.0007	0.0009
BBB	0.0002	0.0021	0.0378	0.9031	0.0437	0.0074	0.0029	0.0029
BB	0.0002	0.0015	0.0023	0.0468	0.8570	0.0697	0.0096	0.0128
B	0.0000	0.0007	0.0033	0.0037	0.0538	0.8435	0.0426	0.0524
CCC	0.0000	0.0000	0.0024	0.0000	0.0071	0.0737	0.6813	0.2355
Default	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 5: DCMM_3_2_1 Transition Probability Matrix for hidden state 2

This table shows transition probabilities calculated by the DCMM for the second ("mover") hidden state based on a S&P issuer rating history for 1994 to 2005.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AA	0.2121	0.0023	0.782	0.0036	0.0000	0.0000	0.0000	0.0000
A	0.0000	0.3664	0.0000	0.6336	0.0000	0.0000	0.0000	0.0000
BBB	0.0000	0.0000	0.6566	0.0004	0.3428	0.0000	0.0000	0.0002
BB	0.0000	0.0000	0.0023	0.7208	0.0000	0.2769	0.0000	0.0000
B	0.0000	0.0000	0.0000	0.0000	0.5348	0.0000	0.4539	0.0113
CCC	0.0000	0.0000	0.0000	0.0000	0.0106	0.8199	0.0952	0.0743
Default	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 6: DCMM_3_2_1 Transition Probability Matrix for hidden state 3

This table shows transition probabilities calculated by the DCMM for the third (“stayer”) hidden state based on a S&P issuer rating history for 1994 to 2005.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AA	0.0016	0.9984	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
A	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BBB	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
BB	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
B	0.0015	0.0000	0.0000	0.0000	0.0022	0.9961	0.0000	0.0002
CCC	0.0000	0.0000	0.0000	0.0245	0.0000	0.0470	0.9285	0.0000
Default	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 7: MC_1 Transition Probability Matrix

This table shows transition probabilities calculated as usually by a discrete homogeneous time Markov chain based on a S&P issuer rating history for 1994 to 2005.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.8402	0.1543	0.0043	0.0012	0.0000	0.0000	0.0000	0.0000
AA	0.0161	0.8617	0.1163	0.0043	0.0004	0.0011	0.0000	0.0001
A	0.0007	0.0399	0.864	0.0912	0.0022	0.0007	0.0005	0.0007
BBB	0.0002	0.0017	0.0705	0.8599	0.0568	0.0061	0.0024	0.0024
BB	0.0002	0.0013	0.0021	0.0736	0.8304	0.0730	0.0083	0.0111
B	0.0001	0.0006	0.0029	0.0033	0.0622	0.8339	0.0500	0.0469
CCC	0.0000	0.0000	0.0020	0.0020	0.0068	0.1191	0.6644	0.2057
Default	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Table 8: Deviation in percentage from the corresponding future rating grade calculated with MC_1

This table provides an overview regarding to the overall trend to migrate from a given rating to a certain rating class for the three matrices from the hidden states and the finally derived matrix. Hereby each column probability mass from each of the four matrices is compared to the respective one estimated by the MC_1.

	AAA	AA	A	BBB	BB	B	CCC	D
DCMM hidden state 1	1.81	-0.97	-1.25	-0.98	0.64	-3.60	1.58	2.98
DCMM hidden state 2	-75.27	29.18	35.67	31.18	-7.36	6.08	-24.32	-14.29
DCMM hidden state 3	16.98	-5.77	-5.85	-1.06	4.53	0.89	27.96	-21.05
final matrix	-9.85	10.54	13.98	14.45	2.66	1.85	5.99	-6.10

Table 9: Relative Frequency table rating distribution across the 3 hidden states

The table shows the relative occurrence frequencies of the hidden states for each rating during the observation period from 1994 to 2005.

	1. hidden state	2. hidden state	3. hidden state
AAA	0.7318	0.0697	0.1986
AA	0.7934	0.065	0.1416
A	0.8118	0.0679	0.1204
BBB	0.872	0.0645	0.0636
BB	0.9126	0.0473	0.0401
B	0.9338	0.0274	0.0387
CCC	0.874	0.0963	0.0297
Default	0.9941	0.0059	0.0000

Table 10: Final Matrix derived from the three hidden states

The transition probability matrix is derived through a weighting approach to keep as many information of the serial correlation and the transient characteristic of credit rating histories from the DCMM as possible. The transition probabilities are derived out of the second order transition probabilities of the hidden states, the respective relative frequencies of each hidden state for each rating grade, and the corresponding transition probabilities from the respective hidden state transition probability matrix.

	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	0.6690	0.3270	0.0031	0.0009	0.0000	0.0000	0.0000	0.0000
AA	0.0816	0.6675	0.2463	0.0035	0.0003	0.0008	0.0000	0.0001
A	0.0005	0.1192	0.6796	0.1978	0.0015	0.0005	0.0004	0.0005
BBB	0.0001	0.0011	0.2122	0.6760	0.1037	0.0039	0.0015	0.0016
BB	0.0001	0.0008	0.0017	0.2283	0.6606	0.0966	0.0051	0.0068
B	0.0006	0.0004	0.0017	0.0020	0.1613	0.6543	0.1495	0.0303
CCC	0.0000	0.0000	0.0013	0.0101	0.0061	0.2506	0.5878	0.1441
Default	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Figure 1: hidden state distribution across the years

This picture shows the hidden state distribution in percent over the observation period 1994-2005. The frequencies of the hidden states are derived for each obligors rating history through the Viterbi algorithms.