

Banachewicz, Konrad; Lucas, André

**Working Paper**

## Quantile Forecasting for Credit Risk Management using possibly Mis-specified Hidden Markov Models

Tinbergen Institute Discussion Paper, No. 07-046/2

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Banachewicz, Konrad; Lucas, André (2007) : Quantile Forecasting for Credit Risk Management using possibly Mis-specified Hidden Markov Models, Tinbergen Institute Discussion Paper, No. 07-046/2, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/86275>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



TI 2007-046/2

Tinbergen Institute Discussion Paper

# Quantile Forecasting for Credit Risk Management using possibly Mis- specified Hidden Markov Models

*Konrad Banachewicz*

*André Lucas\**

*Faculty of Economics and Business Administration, VU University Amsterdam.*

*\* Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# Quantile Forecasting for Credit Risk Management using Possibly Mis-specified Hidden Markov Models\*

Konrad Banachewicz<sup>†</sup> André Lucas<sup>‡</sup>

June 8, 2007

## Abstract

Recent models for credit risk management make use of Hidden Markov Models (HMMs). The HMMs are used to forecast quantiles of corporate default rates. Little research has been done on the quality of such forecasts if the underlying HMM is potentially mis-specified. In this paper, we focus on mis-specification in the dynamics and the dimension of the HMM. We consider both discrete and continuous state HMMs. The differences are substantial. Underestimating the number of discrete states has an economically significant impact on forecast quality. Generally speaking, discrete models underestimate the high-quantile default rate forecasts. Continuous state HMMs, however, vastly overestimate high quantiles if the true HMM has a discrete state space. In the reverse setting, the biases are much smaller, though still substantial in economic terms. We illustrate the empirical differences using U.S. default data.

**Key words:** defaults; Markov switching; misspecification; quantile forecast; Expectation-Maximization; simulated maximum likelihood; importance sampling.

---

\*We are grateful to Standard and Poors for providing the CreditPro database for the empirical section.

<sup>†</sup>Vrije Universiteit Amsterdam, Department of Mathematics, phone +31-20-5987779, mail konradb@few.vu.nl.

<sup>‡</sup>Vrije Universiteit Amsterdam, Department of Finance and Tinbergen Institute.

# 1 Introduction

In this paper we employ Hidden Markov Models (HMMs) for forecasting quantiles of corporate default rates. Such quantiles play an important role in financial risk management. The class of HMMs has been popularized by Hamilton (1989) and has proven useful for modeling and forecasting a variety of economic time series, including as prime examples business cycles indicators and exchange rate. Over the last years, several authors have implemented HMMs to model and forecast credit risk dynamics. Crowder, Davis, and Giampieri (2005) uses the familiar model of Hamilton (1989) with two regimes: a high and low default rate regime. In the model of Crowder, the hidden Markov layer is observed via a binomial layer, where the binomial draw is the number of defaults. This is slightly different from Hamilton (1989), where the hidden layer is observed via a Gaussian series of observables. Banachewicz et al. (2006) extends the model of Crowder by making the transition probabilities in the MC dependent on macro-variables.

A different approach is taken by McNeil and Wendin (2007) and Koopman and Lucas (2005). They distinguish a continuum of possible states, where the state follows a simple time series process such as a low order autoregression. As in Crowder et al. (2005), the observed time series of default counts is modeled as a binomial process, where the success probability depends directly on the hidden Markov layer. Koopman, Kraeussl, Lucas, and Monteiro (2006) and Duffie, Eckner, Horel, and Saita (2006) further extend this approach to a continuous time setting, where the hidden layer drives the intensity of a point process.

The main drawback of the HMMs is that they are more cumbersome to estimate. Instead of straightforward maximum likelihood, one has to resort to simulated maximum likelihood, EM, or Bayesian methods. This is why some previous authors have abstained from modeling a hidden layer, and have used observable state variables instead. For example, Nickell, Perraudin, and Varotto (2000) and Bangia, Diebold,

Kronimus, Schagen, and Schuermann (2002) use GDP growth rates and NBER business cycle classifications to distinguish between high and low default rate regimes. McNeil and Wendin (2007), Koopman et al. (2006), and Duffie et al. (2006), however, show that even if one includes observable macro variables, a hidden layer is still needed to capture default rate dynamics correctly. Moreover, Lucas and Klaassen (2006) show that default regimes based on observables like GDP growth and NBER business cycle classifications result in a substantial underestimation of the annual default rate volatility. This is in line with Dacco and Satchell (1999), who show that a slight mis-classification of regimes can lead to a substantial deterioration in forecasting performance of regime switching models.

The main conclusion from the literature is that HMMs constitute a promising tool for dynamic credit risk modeling. So far, however, no systematic study has appeared that compares the adequacy of the proposed competing HMM specifications for forecasting. The models from the literature differ in the number of regimes they distinguish, in the hidden layer dynamics, and in whether they have a discrete or continuous state space.

The focus in our paper is on the effect of mis-specification in the hidden layer's state space dimension and dynamics on the quality of quantile forecasts. This appears particularly important given the limited number of time series observations typically available for default rate modeling: annual, quarterly, or monthly time series since the 1980s. For such data, it is far from trivial to reliably determine the appropriate dimension of the state space or the dynamics in the hidden Markov layer from the data.

We focus on forecasting quantiles rather than expected values. This is in line with the predominant use of quantiles rather than expected values for risk management purposes in the financial industry. Following the new Basel Capital Accord (2005), quantiles can be used directly by financial institutions to determine the size of required capital buffers. It will turn out that the quality of quantile forecasts

may be affected in a substantially different way by mis-specification than forecasts of means.

We contribute to the existing literature in three ways. First, we provide a systematic comparison of the forecast accuracy of different non-Gaussian state space models for credit risk. In particular, we are the first to compare both discrete-state and continuous-state HMM specifications as they have been put forward in the recent literature. Second, we concentrate on the effect of mis-specification in hidden layer dynamics on quantile forecasts. And finally, we apply the various methodologies to an empirical example.

We conduct a controlled simulation experiment, where we vary the number of regimes, i.e., the dimension of the state space. We systematically study the effect on quantile forecasts of over- or under-estimating the number of regimes. We find that underestimating the number of regimes has a significant impact on forecast quality. For the high quantiles typically used in practice, the differences appear economically significant. Overestimating the number of regimes appears to have less effect. Surprisingly, however, even correctly specified models have difficulty in adequately estimating the high quantiles of the true distribution of future default rates.

The continuous state models behave substantially differently from the discrete state models at high quantiles. Typically, continuous state models result in much higher quantiles. If the true data generating process (DGP) has a discrete state space, the continuous state model vastly overestimates the high quantiles. Conversely, if the true DGP has a continuous state space, the discrete state models substantially underestimate high quantile default rates. This holds even if we allow for four or five different discrete regimes.

We also apply the different methods to an empirical data set of U.S. corporate default rates, obtained from Standard and Poor's. We find that a discrete model with three states provides a good empirical description of the data. In terms of

forecasting performance, the continuous state model picks up the dynamics of the realized default rates better than the discrete state model. As usual, however, the forecasts lag the realizations if we forecast further out of sample. The main problem with the continuous-state specification is that its high-quantile forecasts are very large. Due to the lagging behavior of the forecast with respect to the realization in the more than one step ahead context, these large predicted quantiles might result in overly conservative default scenarios and, thus, in overly conservative capital requirements.

The rest of this paper is set up as follows. In Section 2 we describe the modeling framework and estimation methodology. In Section 3, we describe the simulation set-up. Section 4 discusses the simulation results, whereas Section 5 presents the empirical application. Finally, Section 6 concludes.

## 2 Model formulation, estimation, and selection

### 2.1 Model specification

Consider a portfolio of defaultable units or counterparties. At each time  $t = 1, \dots, T$  we observe a pair  $\{N_t, D_t\}$ . Here  $D_t$  is the number of counterparties that default during period  $(t - 1, t]$ , while  $N_t$  denotes the number of units at time  $t - 1$ . The number of units  $N_t$  is affected by the number of defaults over the previous period ( $D_{t-1}$ ). In addition,  $N_t$  may also increase because units enter the sample (births), or decrease because units leave the sample for reasons other than default. Along the lines of, e.g., Bangia et al. (2002), births and withdrawals are treated as exogenous.

We assume that the default rate ( $D_t/N_t$ ) dynamics are driven by a latent process  $W_t$ . The variable  $W_t$  captures current credit market conditions. As is commonly assumed in the literature, conditional on the realization of  $W_t$ , companies default independently.

We formulate our model in terms of a general non-Gaussian, non-linear state



space framework, along the lines of Durbin and Koopman (2001),

$$h(W_t|W_{t-1}), \tag{1}$$

$$g(D_t|N_t, W_t) = \text{dbin}(D_t, N_t, p(W_t)), \tag{2}$$

where  $W_t$  is a state vector,  $D_t$  is the number of defaults,  $N_t$  is the number of credit exposures,  $p(W_t)$  is between 0 and 1, and  $\text{dbin}(x, n, p)$  denotes the binomial density function, i.e., the probability of drawing  $x$  successes in  $n$  trials with success probability  $p$ . This flexible representation allows us to account directly for the discrete nature of the default process. The occurrence of zeros and very small default frequencies can now be dealt with in a natural way. Also, prediction, filtering and smoothing problems can be addressed naturally in this setting.

The State Space Formulation (SSF) in (1) and (2) has two sources of uncertainty. First, we have the familiar measurement errors in the measurement equation (2). In our current context, these measurement errors are non-Gaussian. The second source of uncertainty comes from the state equation (1). This uncertainty typically has to be integrated out for likelihood evaluation and, therefore, complicates the estimation process. We consider two specifications for the state dynamics (1). First, we consider a discrete state specification. Here, we assume the latent process follows a time-homogeneous Markov Chain taking values in the set  $\{1, \dots, s\}$ ,

$$\Pr(W_t = i|W_{t-1} = j) = P_{ij}, \quad i, j = 1, \dots, s, \tag{1a}$$

where  $s$  denotes the number of discrete states or regimes, and  $P_{ij}$  is the probability of moving from state  $j$  at  $t - 1$  to state  $i$  at  $t$ . The probabilities can be put into a transition matrix  $P = (P_{ij})_{i,j=1,\dots,s}$ . Models of type (1a) have been used in, e.g., Crowder et al. (2005) and Banachewicz et al. (2006).

As a convention in most parts of this paper, we interpret the state values 1 to  $s$  as corresponding to the lowest and highest default regimes, respectively. Other

interpretations are, however, also possible. For example, a four dimensional state vector can be interpreted as corresponding to two default regimes with second order dynamics. In this case, state 1 may be the low default regime if there was also a low default regime the period before. State 2 is then the low default regime if there was a high default regime the period before. State 3 and 4 are the high default regimes if there was a low or high default regime the period before, respectively. The transition matrix in this cases takes the restricted form

$$\begin{pmatrix} P_{11} & 0 & P_{13} & 0 \\ P_{21} & 0 & P_{23} & 0 \\ 0 & P_{32} & 0 & P_{34} \\ 0 & P_{42} & 0 & P_{44} \end{pmatrix}.$$

Using (1a), we can thus investigate both the effect of an incorrect number of regimes as was as the effect of a correct number of regimes, but an incorrect dynamic specification of the MC.

Our second specification for (1) has continuous states. We model this by a linear autoregressive time series process,

$$W_t = \Phi W_{t-1} + V\varepsilon_t, \tag{1b}$$

where  $\varepsilon_t$  is a Gaussian white noise with unit variance matrix. Using the standard way of mapping a higher order autoregression into a vector autoregression of order one, we are able to investigate the effect of dynamic mis-specification in the state equation. Models of type (1b) have been used in, e.g., McNeil and Wendin (2007), Koopman and Lucas (2005), Koopman et al. (2006), and Duffie et al. (2006).

To complete the model specification, we define the probability function  $p(W_t)$  in

the measurement equation (2). In the discrete state specification (1a), we set

$$p_t = p(W_t) = \alpha_{W_t}, \quad (3a)$$

where  $\alpha_i$  denotes the default probability in state  $W_t = i$ . For the continuous state model, we use a logistic mapping from the continuous  $W_t$  to the default probability. In this paper we restrict our attention to cases where  $p(W_t)$  only depends on the first element of  $W_t$ , or

$$p_t = p(W_t) = (1 + \exp(a + b \cdot e_1' W_t))^{-1}. \quad (3b)$$

where  $e_1$  is the first column of the unit matrix. The parameters  $b$  in (3b) and  $V$  in (1b) cannot be identified simultaneously. For identification, we typically normalize the unconditional variance of  $e_1' W_t$  in (1b) to unity. Alternatively, one can restrict the conditional variance  $V$  or the loading  $b$  to unity.

By combining (2) with (1a) or (1b), we obtain a discrete or continuous special case of the nonlinear and non-Gaussian state space model as defined in part II of Durbin and Koopman (2001). Parameter estimation in such models is not trivial and differs substantially between the discrete and continuous state specification. For completeness, we briefly review both estimation methodologies in the next two subsections.

## 2.2 Estimation: discrete case

For the discrete state case, (1a) and (2), we estimate the model parameters via EM using a modification of the Baum-Welch algorithm of Rabiner (1989). Details are given in Appendix A.1. The basic idea is to define so-called forward and backward variables  $\alpha_t(i)$  and  $\beta_t(j)$ . These variables are easy to compute recursively and represent probabilities of partial observation sequences. The recursions are used

to efficiently perform the Expectation step of the EM algorithm by computing the expectation of the log-likelihood, see Dempster et al. (1977). Here, we can exploit the special structure of the discrete state model. The expression for the expectation can be decomposed into separate parts, so that the problem of optimization over all model parameters simultaneously can be replaced by a set of univariate maximization problems. Moreover, the update formulas for maxima of particular parameters are expressed in terms of the forward and backward variables  $\alpha_t(i)$  and  $\beta_t(j)$ . New parameter values are then used to re-compute the expectation, and this iterative procedure is repeated until convergence. Further details on the implementation of the procedure are provided in Appendix A.1.

### 2.3 Estimation: continuous case

The flexibility gained by moving to the continuous state specification (1b) comes at the price of more complications in the likelihood evaluation. The principal difficulty is that the likelihood requires the evaluation of a high-dimensional integral,

$$\int \cdots \int \prod_{t=1}^T (\text{dbin}(D_t, N_t, p(W_t))h(W_t|W_{t-1})) \, dW_T \cdots dW_1. \quad (4)$$

This integral has to be computed for every evaluation of the likelihood. The first step to make the computations feasible is to replace the integrals by averages and use Monte Carlo maximum likelihood methods for parameter estimation. Let  $W^{(k)} = \{W_t^{(k)}\}_{t=1}^T$  for  $k = 1, \dots, K$  denote a set of simulated paths of (1b). Then we approximate (4) by

$$K^{-1} \sum_{k=1}^K \prod_{t=1}^T \text{dbin}(D_t, N_t, p(W_t^{(k)})). \quad (5)$$

Sampling directly from (1b), however, is highly inefficient as the paths  $W^{(k)}$  would have no relation to the realized sample and, would therefore contribute only little to the final likelihood. This problem can be avoided by using importance samplers

instead. Durbin and Koopman (2001) describe a way to create an efficient Gaussian importance sampler  $G(W^{(k)}|D_1, N_1, \dots, D_T, N_T)$  to approximate (4) by

$$K^{-1} \sum_{k=1}^K \frac{\prod_{t=1}^T \left( \text{dbin}(D_t, N_t, p(W_t^{(k)})) h(W_t^{(k)} | W_{t-1}^{(k)}) \right)}{G(W^{(k)}|D_1, N_1, \dots, D_T, N_T)}. \quad (6)$$

Koopman and Lucas (2005) extensively investigated the performance of this importance sampler for the model at hand and conclude that  $K = 30$  to  $K = 50$  importance samples already provide a very accurate approximation to (4). The basic idea of the Gaussian importance sampler is to linearize the measurement equation (2) around a linear Gaussian model with the same mean and mode and the same curvature of the log-density. This gives a new, approximate measurement equation with time varying mean and variance. The resulting *linear* approximate state space model can be used for efficient sampling of the states (1b) conditional on the observations using standard Kalman filter techniques. For a more detailed exposition, the reader is referred to Durbin and Koopman (2001) and Koopman and Lucas (2005).

## 2.4 Likelihood and model selection

In the general case of HMM modeling, we may want to determine the model on the basis of the sample. In particular, we want to choose between competing models with different dimensions  $s$  for the state space. We concentrate on likelihood based selection criteria. In the continuous case (1b), the (approximate) likelihood is obtained directly from the importance sampling scheme (6). In the discrete setting, the likelihood is not a by-product of the estimation procedure, but has to be computed separately. This can, however, be done in a computationally efficient way using the same mathematical apparatus that is employed for parameter estimation. Using the same, recursively defined variables  $\alpha_t(i)$  from Appendix A.1, we can write the

likelihood as

$$LL = \sum_{i=1}^s \alpha_T(i), \quad (7)$$

where  $s$  is the number of regimes in the HMM and  $T$  is the number of observations.

Once the likelihood has been computed, we can compare different models using for example the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). In the discrete case, these take the form

$$\begin{aligned} AIC &= -2 \ln(LL) + 2(s^2 + s - 1), \\ BIC &= -2 \ln(LL) + (s^2 + s - 1) \ln(T), \end{aligned}$$

where  $s$  is the dimension of the state space. For the continuous state specification (1b), the number of parameters is typically given by  $s + 2$ : the number of autoregressive parameters  $s$  and the logistic parameters  $a$  and  $b$  in (3b). A difference in these model selection criteria between the discrete and continuous-state specification, however, still has to be interpreted with care, as the behavior of these two models in out-of-sample forecasting is very different. This is further illustrated in Section 4.

### 3 Simulation set-up

In this section we describe our simulation experiment. The experiment is set up to investigate the quality of quantile forecasts under possible mis-specification of the hidden MC. The mis-specification may take the form of wrong sizes of the dimension of the state space or a wrong dynamic specification of the MC.

Let  $H_s(\theta_s)$  denote a discrete state HMM model, where  $s$  denotes the number of regimes in the HMM, and  $\theta_s$  gathers all the parameters of the HMM. The vector  $\theta$  in the discrete state case contains the levels of the default rates  $\alpha_i$ , the transition matrix  $P$ , and the vector of initial state probabilities  $\pi_0$ . The number of regimes  $s$  coincides

with the dimension of the discrete state space  $W_t$  in (1b). For the continuous state model, we use the notation  $H_\infty(\theta_\infty)$ .

Estimated values for the dimension and the parameters are denoted by  $\hat{s}$  and  $\hat{\theta}_s$ , respectively. For each model, we also define a quantile function for the  $h$ -step-ahead forecast. Let  $Q(x; h, H_s(\theta_s), \pi_T)$  be such that

$$Q(x; h, H_s(\theta_s), \pi_T) = \sup\{y \mid \Pr[Y \leq y] \leq x\},$$

where  $Y$  is the  $h$ -step-ahead forecast from the HMM  $H_s(\theta_s)$  with  $W_T$  following the distribution  $\pi_T$ , and where  $x \in [0, 1]$ . Each  $Q(\dots)$  is estimated as an average over 2500 simulated sample paths.

Our prime interest will be in the bias and variability of the the quantile functions  $Q$ . In our simulations, we use a pair  $(H_s(\theta_s), \pi_0)$  to generate a sample of  $D_t$ s. The  $N_t$ s are treated exogenously and are set to their empirical values from Section 5. Only out of sample we use a different scheme to set the value of  $N_{T+h}$ , namely  $N_{T+j} = N_{T+j-1} - D_{T+j-1}$  for  $j = 1, \dots, h$ . We thus abstract from units entering the sample after time  $T$ , and concentrate on the forecasts of defaults only.

We construct three benchmarks  $Q(x; h, H_s(\theta_s), \pi_T)$  that all use the correctly specified HMM, and only differ in their choice for the initial distribution  $\pi_T$ . We start off the MC with the (true) unconditional distribution  $\pi_T^{unc}$ , the conditional (smoothed) distribution  $\pi_T^{cond}$  of the state  $W_T$  using the simulated data and the true parameters, and the degenerate distribution  $\pi_T^{deg}$  that has a unit point mass on the realized value of  $W_T$ . Note that the latter two distributions are different in each simulation run. Let  $Q_{s,h}^{unc} = Q(x; h, H_s(\theta_s), \pi_T^{unc})$ , where  $Q_{s,h}^{cond}$  and  $Q_{s,h}^{deg}$  have similar definitions.

Using a simulated path of defaults, we can use the estimation methods discussed in the previous section to obtain  $\hat{\theta}_{s^*}$ , where  $s^*$  is unequal to  $s$  in the case of misspecification. If  $s^* = s$ , we only have the effect of parameter estimation error left. One can also investigate the case  $s^* = \hat{s}$ , where  $\hat{s}$  is the estimate of the state

dimension obtained by maximizing one of the selection criteria in Section 2.4. We consider values for  $s, s^*$  of 2 up to 5. In addition, we have the continuous state model, which we denote by  $s, s^* = \infty$ .

The specification for the continuous state model in our simulation experiment is an autoregressive model of order 1, AR(1),

$$W_t = \phi W_{t-1} + \sqrt{1 - \phi^2} \varepsilon_t, \quad (8)$$

where we normalized the unconditional variance to unity for identification, see Section 2. Models of this type have been used in previous empirical studies on default modeling and proven very useful, see McNeil and Wendin (2007), Koopman and Lucas (2005), Koopman et al. (2006), and Duffie et al. (2006).

For each comparison of quantile functions, we perform  $M = 1000$  simulations. The final curves presented in the next section present the averages over the simulations of  $\hat{Q}_{s^*,h}^{unc} - Q_{s,h}^{unc}$  and  $\hat{Q}_{s^*,h}^{cond} - Q_{s,h}^{deg}$  as a function of  $x$ , where

$$\hat{Q}_{s^*,h}^{unc} = Q(x; h, H_{s^*}(\hat{\theta}_{s^*}), \hat{\pi}_T^{unc}). \quad (9)$$

A similar definition holds for  $\hat{Q}_{s^*,h}^{cond}$ .

We consider forecast horizons of 4, 8, and 12 quarters. We set the sample size to  $T = 100$ . This corresponds to the typical size of for this type of data. Using  $T = 100$ , we should get a good impression of the effect of parameter uncertainty on the level and variability of quantile forecasts.

Finally, the parameters of the true DGPs in the simulations are chosen on the basis of empirical estimates using the data from Section 5. The parameter values are provided in Table 2 for 2, 3, and 5-state models. Interestingly, the 2-state model constitutes an asymmetric time series process. The high default rate regime is much less persistent than the low default rate regime. This may cause additional complications for the symmetric continuous state model. The model with 5 regimes



appears to be over-specified compared to the information in the data. Especially the two lowest default regimes lie very close together, and the regime with default probability 0.14% appears unstable. The model can nevertheless be used to investigate the effect of overspecification on quantile forecast stability. This is done in the next section.

<INSERT TABLE 2 AROUND HERE>

## 4 Simulation results

In this section, we present the outcome of the simulation experiments. In Subsections 4.1 and 4.2, we discuss the effect of overestimating and underestimating the state dimension, respectively. Subsection 4.3 is dedicated to the problem of approximating a continuous model with a discrete  $s$ -dimensional one, for different values of  $s$ .

### 4.1 Too many states

In this part of the simulation study we assume that the number of states is larger than necessary,  $s^* > s$ . The DGP is an HMM with a bivariate state space ( $s = 2$ ). We approximate this by HMMs of dimensions  $s^* = 2, 3, 4, \infty$ . The methodology works for higher dimensional discrete HMMs as well ( $s^* = 5, 6, 7, \dots$ ). Fitting such models, however, seems unrealistic given the limited sample size  $T = 100$ . As explained in section 3, we present the differences  $\hat{Q}_{s^*,h}^{unc} - Q_{s,h}^{unc}$  and  $\hat{Q}_{s^*,h}^{cond} - Q_{s,h}^{deg}$ . The first measure compares the unconditional quantile approximation to the one from the DGP. The second measure compares the quantile approximation based on the smoothed starting distribution at time  $T$  with the quantile from the true DGP starting from the actually realized  $W_T$ .

Figure 1 presents the result for a 4-quarters-ahead forecasting horizon. The first thing to note is that the quantile forecasts are very similar for  $s^* = 2, \dots, 4$ . The

bias is slightly negative for the fitted model with the correct  $s^* = s = 2$ . The same holds for the higher dimensional models. At the very high quantiles, the negative bias in the higher dimensional models becomes smaller and in some cases transforms to a slightly positive bias.

<INSERT FIGURE 1 AROUND HERE>

The second message from Figure 1 is the completely aberrant behavior of the quantiles from the continuous state model ( $s^* = \infty$ ). The bias for  $s^*$  is much higher than for the discrete state models. The effect is stronger the further we move out into the tails of the default rate distribution. The intuition for this is that the discrete state models account for the fact that the true DGP only has two different values for the default rate. The continuous state model, on the other hand, is calibrated partly by matching the unconditional volatility in the default rate for ( $s^* = \infty$ ) model to that of the true ( $s = 2$ ) DGP. Given that the volatilities are roughly matched, the continuous state model has the additional property that the predicted default rate can increase much further if we move far out into the tails of  $W_{T+h}$ . The interesting feature in Figure 1 is that the potential quantile bias due to this phenomenon already takes drastic forms for quantiles of 90% and higher.

The magnitude of the biases for the conditional quantile forecasts is illustrated further in panel A of Table 1. The numbers for the unconditional forecasts shows very similar results, and are therefore omitted. The bias in the predicted median default rate is slightly positive for all models. The relative biases for the median in the discrete state models hover around 0.2%. The continuous state model, however, reveals a bias of 3.6% for the median. Further out into the tails, the relative biases for the discrete state models remain modest and below 2% in absolute values. The relative bias for the continuous state model, on the other hand, rises from 77% for the 95th percentile to a staggering 149% for the 99th percentile. These percentage biases transform directly into percentage biases in economic capital if the loss given

default is constant across firms.

<INSERT TABLE 1 AROUND HERE>

We now turn to the stability of our results over increasing forecasting horizons  $h = 4, 8, 12$ . The results are reported in Figures 2. The figure not only presents the average as well as the confidence band of the complete forecasted quantile function. Since the outcomes are similar for  $s^* = 2, 3, 4$ , we only present the results for  $s^* = s = 2$ .

<INSERT FIGURE 2 AROUND HERE>

For all time horizons, the median default rate quantile is captured within the confidence bands. This follows from the fact that the bias at the 50th percentile is insignificantly different from zero. The biases for other quantiles, however, show a non-monotonic pattern. Moving from the median to the higher quantiles, the default rate quantiles are first significantly underestimated. For the very high quantiles, the bias reduces, but remains significant. This holds for both the unconditional and the conditional quantile forecasts. The different behavior of the forecasts over quantiles is important for risk management. It appears that the magnitude of the bias depends on the quantile one is interested in. Typical quantiles for risk management applications include 95%, 99%, and higher quantiles. At the 95th percentile, the bias appears close to its maximum and an underestimate of the true risk. For higher quantiles, this bias is somewhat smaller.

A general conclusion from this part of the study is that overestimating of the number of regimes has a limited effect on high quantile estimation. Relative biases are kept below 3% of the true quantile value. The main exception is the continuous state model, which vastly overestimates high quantiles by up to 130% or more. Moreover, the magnitude of biases varies over the quantile of interest. Biases are small near the median, increase further out in the tails, and decrease again in the extreme tails.

## 4.2 Too few states

We now turn to the opposite setting. We use a  $s = 5$  dimensional HMM model and approximate it by  $s^* = 2, 3, 5, \infty$ . For  $s^* = 2, 3$ , we thus underestimate the number of discrete regimes.

Figure 3 shows that all discrete state models underestimate the true quantile functions. This even holds for the true model  $s^* = 5$ . The models for  $s^* = 2, 3$  perform worse. The variation in high default rate values for the 2-state model ( $s^* = 2$ ) is apparently insufficient to capture the high annual default rates in the model with five regimes. This follows also from Table 2. The high default rate in the 2-state model of 0.69% roughly averages the high default probabilities of 0.53% and 0.86% in the 5-state model. Especially the high default rate of 0.86% can cause the 2-state model to significantly underpredict the high default rate quantiles.

The continuous state model again substantially overestimates the high quantiles. The biases are less than in Figure 1. This is due to the fact that the default rate in the true DGP also has a higher default rate volatility and, therefore, produces higher quantiles. The biases, however, appear still too large for practical risk management purposes. This is underlined further in panel B of Table 1. The percentage errors in the high quantile forecasts in panel B for  $s^* = \infty$  are of similar magnitude as in panel A. It is also clear from the table that the true model  $s^* = 5$  is the only one that keeps the biases in the forecast below 2%. Still, the biases for the 2-state model for the highest quantile computed ( $-12.45\%$ ) are only a fraction of those for the continuous state model. It is also interesting to see that the model with three regimes already performs much better than the 2-regime model. This illustrates that for the empirical data, adding a third regime forms an important improvement for the empirical validity of the model.

Figure 4 plots the unconditional and conditional quantile forecasts over 4, 8, and 12 quarter horizons. We again concentrate on the true model  $s^* = s = 5$ . The general picture is markedly different from that of the 2-regime model, see Fig-

ure 2. Whereas the bias for the model with two regimes was more or less symmetric around the median forecast, the 5-regime model shows a clear asymmetry. The unconditional quantile’s bias appears to reach its maximum positive value between the median and the 60th percentile. These biases can only be due to the uncertainty in the parameter estimation, as the model is correctly specified in the number of regimes. At the high quantiles, the biases are even larger in absolute terms. This is especially worrying given the importance of these quantiles for risk management and capital buffer determination.

If we now turn to the results for the conditional quantile forecasts in the right-hand column of graphs in Figure 4, we again notice some differences with the previous results. The resemblance between the unconditional and conditional forecast biases is less clear than for the model with two regimes. For a horizon of one year ( $h = 4$ ), the bias is again more or less (anti)-symmetric around the median. For more than one-year ahead ( $h = 8, 12$ ), however, the asymmetry in the bias again becomes apparent. For  $h = 8$ , the confidence band around the forecast of the median includes zero. For  $h = 12$ , this is no longer the case. As expected, all biases increase with the length of the forecasting horizon  $h$ . Also, the main conclusion that the absolute bias is largest at the high quantiles, remains robust.

The current experiment yields three conclusions. First, empirically congruent discrete HMM models on average under-estimate high credit risk quantiles. If the number of regimes is smaller than that of the true DGP, this is caused by an unwarranted averaging of high default rate regimes. If the number of regimes is specified correctly, the average bias is limited. Only the parameter uncertainty in these models for typically available sample sizes still causes some difficulties in capturing the location of high credit quantiles accurately. This problem worsens if we go further out of sample. Second, the increase in forecast accuracy of adding a regime can be substantial. In our empirically congruent simulation set-up, for example, the model with three regimes has much better properties than that with two regimes. Finally,

the continuous state model still vastly over-estimates high credit risk quantiles, even if the true DGP has five regimes. To investigate this issue further, we check whether discrete state models have similar difficulties in approximating a continuous state model.

### 4.3 Continuous versus discrete

We now examine how well a discrete HMM model can approximate a continuous state HMM model. As a benchmark, we estimate the continuous state model (1b) and (3b) on the empirical data. We obtain the parameter estimates<sup>1</sup>  $a = -5.89$ ,  $b = 0.72$ ,  $\Phi = 0.92$ , and  $V = 1 - \Phi^2$ . These parameters are used to simulate paths for  $s = \infty$ . Each path is used to estimate models of  $s^* = 2, 3, 4, \infty$ .

Figure 5 presents the results for the high quantiles  $[0.90, 1)$ . The recovered quantile function for the true model ( $s^* = s = \infty$ ) is quite close to its DGP counterpart, both for the conditional and unconditional forecasts. All the discrete models underestimate the high quantiles, though the bias in the unconditional forecasts for models with 3 or 4 regimes remain negligible up to the 95th percentile. The model with only two regimes, by contrast, shows a significant negative bias throughout the range of quantiles displayed.

<INSERT FIGURE 5 AROUND HERE>

The biases in the conditional forecasts are substantially larger than in the unconditional ones. The continuous state model can try to adapt to the realized state variable  $W_T$  arbitrarily closely if the signal is strong enough. The discrete state models, however, have to pick their estimate for  $W_T$  from a discrete set of different values. Though this discrete set with associated probabilities might provide a reasonable approximation to the unconditional distribution of the continuous state, it might be much less adequate for fitting a particular realization of  $W_T$ .

---

1

Panel C in Table 1 summarizes the absolute and relative biases. The bias at the median is smallest for the true model ( $s^* = \infty$ ). The models with 3 and 4 regimes have a comparable performance with biases below 3%, followed by the 2-regime model with a bias at the median of more than 5%. The interesting difference arises at the higher quantiles. Whereas the biases in the conditional forecasts for the correctly specified model ( $s^* = \infty$ ) remain below 9%, those for the discrete state models are substantially higher. The 2-state model has biases of 40% up to almost 60%. More important, however, is the performance of the models with 3 and 4 regimes. Their bias ranges from roughly 25% at the 95th percentile, to around 35% for the 99th percentile. Though these values are substantial, they appear far from the huge percentage (and absolute) biases of the continuous state model ( $s^* = \infty$ ) in case the true DGP has discrete states, see panels A and B in Table 1. This suggests that the discrete state models are more robust than the continuous state model for quantile forecasting. The biases, however, may still be large.

## 5 Empirical application: U.S. corporate defaults

We now illustrate the different forecasting methods for an empirical data set. The data for our study come from the CreditPro 7.0. database of Standard & Poor's. The time series of interest consist of U.S. corporate defaults between January 1981 and July 2005. We use quarterly observations. The sample period encompasses both expansions and contractions. This is important, as part of our interest concerns the difference of conditional default probabilities between economic regimes. In line with previous empirical work in this area, we define the number of exposures  $N_t$  for each sector as the number of active companies at the start of each quarter minus the number of withdrawals over the subsequent quarter. A withdrawal is defined as a company leaving the database for other reasons than default. If a company first withdraws and later defaults, this is recorded in the database. In such cases, we

skip the withdrawal event and only account for the default event. In this way, we mitigate any biases due to strategic default behavior.

We do a recursive out-of-sample forecasting exercise for the last 6 years of our sample. First, we drop the last 6 years of the sample and estimate a discrete and a continuous state HMM. The models are used to construct out-of-sample forecasts of the quantile functions. Next, we add a quarter to the sample and re-iterate the whole procedure. We use forecast horizons of 1 up to 4 quarters. We do not use the longer forecast horizons of 8 and 12 quarters as in the simulation section because of the limited data set.

In order to benchmark our forecasts, we also compute the semi-annual and annual default rates. Let  $p_t$  denote the quarterly default rate over quarter  $t$ . The realized default rate  $p_T^h$  over period  $[T, T + h]$  is computed as

$$p_T^h = 1 - \prod_{i=1}^h (1 - p_{T+i}).$$

The first issue we need to decide on is the dimension of the hidden state space. Based on the results from Table 3, we pick  $s = 3$  and compare its performance to that of the continuous state model. We provide no parameter estimates of the model, as all models are estimated recursively in our forecasting procedure.

<INSERT TABLE 3 AROUND HERE>

The results are reported in Figure 6. We plot the realized default rate, and the forecasts of the median default rate and the 5th and 95th percentiles. We do so for the 3-regime and the continuous state model and for horizons  $h = 1, 2, 4$ .

<INSERT FIGURE 6 AROUND HERE>

By dropping the last 6 years of the sample, we start the forecasting exercise at a historically interesting period, namely end of the 1990s. The early 2000s have shown a historically large and partially unexpected increase in corporate default rates. By



beginning the recursive forecasting procedure at this point, we are able to see how quickly the different models pick up with the increase of default rates over the years 2000 and 2001.

For the one-quarter-ahead forecasts, the median forecast appears to follow the realized default rate quite accurately. It only appears to miss the real surge in default rates around mid 2001. This appears from the realized default rate bumping into the 95% quantile around that time. The decrease in default rates around the end of the sample is also picked up by the 95% quantile, but not as much by the 5% quantile. As a result, the realized default rate (almost) crosses the 5% quantile here. Overall, the quantile bands remain pretty tight around the median forecast. They also capture long upward and downward trends in the default rate dynamics, but are incapable of reproducing short and sudden surges very accurately (spikes like these get “averaged” out).

Once we move beyond a one step ahead forecast, the accuracy of the discrete state model worsens. In particular, the forecast starts to lag the realization. This is a rather common phenomenon in time series forecasting, as the lag between the forecast and the realized value reflects the smaller amount of data that is at our disposal at the time the forecast is calculated. Especially during the peak years of the default crisis (2000, 2001), the model underpredicts the level of default rates. The 5% and 95 % quantiles vary over time along with the realized rate, although they are somewhat tight, which results in failure to capture the default rate surge in mid-2001. The median, on the other hand, picks up only a rough tendency of the realized process, but does not vary over time the way it should (while the realized default rate shows a clear dynamic pattern). These phenomena, showing deterioration of the forecast quality, are visible for  $h = 2$  and become quite prominent for  $h = 4$ .

The continuous model reveals a different behavior, both in qualitative and quantitative terms. For the  $h = 1$ , the dynamics of the forecast closely follow those of the realizations. The 95% quantile, however, is much larger than that for the

3-regime model. This is in line with the results from the simulation section. If we move further out, the mismatch between the forecasts and the realizations increases. Again, the forecast starts to lag the realization – as mentioned before, this is to be expected for  $h$ -step-ahead forecasts with  $h > 1$ . The 95% quantile forecast remains very large. Moreover, because it also lags the default rate realizations, it results in quantile forecasts (and therefore credit risk capital requirements) that still increase considerably, whereas default rates realizations are already decreasing. This clearly appears undesirable from an applied perspective.

## 6 Conclusions

In this paper we examine the performance of the Hidden Markov Model of Crowder, Davis, and Giampieri (2005) applied to predicting a quantile function of the future default rate. We address the issue of misspecification of the latent state dimension and show, that the impact on quantile forecasts is economically important. High quantiles are generally underestimated, and the problem becomes more pronounced if one uses too low a dimension for the forecasting model compared to the dimension of the true HMM. The quantile forecast biases also increase with the forecasting horizon.

We also compare the discrete state HMMs with a continuous state HMM model as proposed in the recent credit risk literature. The results show that if one uses the continuous state model to approximate a discrete state HMM, the resulting quantiles are vastly over-estimated. This obstructs the use of such models in practice, if in reality we would only have a few different default regimes. Vice versa, the discrete state models under-estimate high default rate quantiles if the true DGP has a continuous state. The biases in this latter case, however, are only 20% to 30% of the reverse bias, i.e., using a continuous state model for a discrete state DGP. The discrete state models, therefore, appear to be somewhat more robust. The bias

of 20% to 30%, however, still seems considerable for practical applications in risk management.

Our application to the empirical data corroborated our simulation results. The quantiles of the continuous state models were much larger than those of their discrete state counterparts. The discrete state models, on the other hand, reveal much more problems in picking up the dynamic behavior of default rates and produce fairly constant quantile forecasts, even if default rates vary substantially. The continuous state model does somewhat better in this respect, but only picks up the dynamics with a lag. Though this is quite usual in time series analysis, it casts doubts on the usefulness of these models for practical credit risk management. More research has to be put into ways to combine the best of both these approaches.

## References

- Banachewicz, K., A. Lucas, and A. van der Vaart (2006). Modelling portfolio defaults using Hidden Markov Models with covariates. Discussion Paper TI06-94/2, Tinbergen Institute.
- Bangia, A., F. X. Diebold, A. Kronimus, C. Schagen, and T. Schuermann (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* 26(2-3), 445-474.
- Basel Committee on Banking Supervision (2004, June). Basel ii: International convergence of capital measures and capital standards: A revised framework. Report 107, Bank Of International Settlements, Basel.
- Crowder, M., M. Davis, and G. Giampieri (2005). A Hidden Markov Model of default interaction. *Quantitative Finance* 5, 27-34.
- Dacco, R. and S. Satchell (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting* 18(1), 1-16.
- Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*

*B* 39, 1 – 38.

Duffie, D., A. Eckner, G. Horel, and L. Saita (2006). Frailty correlated default. Working paper, Stanford University.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Hamilton, J. D. (1989, March). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357 – 384.

Koopman, S. J., R. Kraeussl, A. Lucas, and A. Monteiro (2006). Credit cycles and macro fundamentals. Discussion Paper TI06-023/2, Tinbergen Institute.

Koopman, S. J. and A. Lucas (2005). A non-gaussian panel time series model for estimating and decomposing credit risk. Discussion Paper TI05-071/4., Tinbergen Institute.

Lucas, A. and P. Klaassen (2006). Discrete versus continuous state switching models for portfolio credit risk. *Journal of Banking and Finance* 30(1), 23 – 35.

McNeil, A. and J. Wendin (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* 14(2), 131–149.

Nickell, P., W. Perraudin, and S. Varotto (2000). Stability of rating transitions. *Journal of Banking and Finance* 24, 203–227.

Rabiner, L. R. (1989, February). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2), 257–286.

## A Appendix

### A.1 EM algorithm

As usual, we start with writing the down the likelihood of an observed sequence of defaults  $D = \{D_1, \dots, D_T\}$  along with an associated hidden state sequence  $W = \{W_1, \dots, W_T\}$  and macro process  $X = \{X_1, \dots, X_T\}$ . Define  $\tilde{W}_{it}$  as an indicator variable taking the value 1 if  $W_t = i$  and 0 otherwise. Sticking to the notation of Section 2, we have the

log-likelihood

$$\begin{aligned} \ln p_\theta(D, W|X) = & \sum_{i=1}^s \tilde{W}_{it} \ln \pi_i + \sum_{t=2}^T \sum_{i=1}^s \sum_{j=1}^s \tilde{W}_{i,t-1} \tilde{W}_{j,t} \ln q_{ij,t-1} \\ & + \sum_{t=1}^T \left[ \ln \binom{N_t}{D_t} + \sum_{i=1}^s \tilde{W}_{it} [D_j \ln \alpha_i + (N_j - D_j) \ln(1 - \alpha + i)] \right]. \end{aligned} \quad (10)$$

Assume that we have obtained some initial estimate  $\theta_0$  of the model's parameters. Given  $\theta_0$ , we can compute so called *forward* and *backward* variables (see Rabiner (1989)). In our case, define

- the probability of a particular number of defaults, given a state of the Markov Chain and the model parameters,

$$b_j(d_t) = P(D_t = d_t | W_t = j, X) = \binom{N_t}{d_t} \alpha_j^{d_t} (1 - \alpha_j)^{N_t - d_t} ;$$

- the probability of the partial observation sequence  $d_1, \dots, d_t$  and state  $i$  at time  $t$ , given the model parameters  $\theta$ ,

$$\begin{aligned} \bar{\alpha}_j(1) &= \pi_j b_j(d_1), \quad j = 1, \dots, s, \\ \bar{\alpha}_j(t) &= \left[ \sum_{i=1}^s \bar{\alpha}_i(t-1) q_{ij,t} \right] b_j(d_t), \quad j = 1, \dots, s \quad t = 2, \dots, T ; \end{aligned}$$

- the probability of a partial observation sequence from  $t+1$  to the end, given state  $i$  at time  $t$  and the model parameters,

$$\begin{aligned} \bar{\beta}_j(T) &= 1 \quad j = 1, \dots, s, \\ \bar{\beta}_i(t) &= \sum_{j=1}^s q_{ij,t} b_j(d_{t+1}) \bar{\beta}_{t+1}(j), \quad t = T-1, \dots, 1 \quad i = 1, \dots, s; \end{aligned}$$

- the probability of being in state  $i$  at time  $t$ , given the observed sequence and model parameters,

$$\gamma_j(t) = \frac{\bar{\alpha}_j(t) \bar{\beta}_j(t)}{\sum_{i=1}^s \bar{\alpha}_i(t) \bar{\beta}_i(t)}, \quad j = 1, \dots, s \quad t = 1, \dots, T ;$$

- the probability of the Markov chain being in state  $i$  at time  $t$  and state  $j$  at  $t + 1$ , given the model parameters and the observation sequence,

$$\xi_t(i, j) = \frac{\bar{\alpha}_t(i)q_{ij,t}b_j(d_{t+1})\bar{\beta}_{t+1}(j)}{\sum_{i=1}^s \sum_{j=1}^s \bar{\alpha}_t(i)q_{ij,t}b_j(d_{t+1})\bar{\beta}_t(t+1)} , \quad i, j = 1, \dots, s \quad t = 1, \dots, T-1 .$$

$$\gamma_j(t) = \frac{\bar{\alpha}_t(j)\bar{\beta}_t(j)}{\bar{\alpha}_t(1)\bar{\beta}_t(1) + \bar{\alpha}_t(2)\bar{\beta}_t(2)} , \quad j = 1, 2 \quad t = 1,$$

Let  $\mathcal{S}$  denote the space of all possible sample paths of the latent process. The E-step of the algorithm requires computing the expectation

$$\mathbb{E}_{\theta_0}[\ln p_{\theta}(D, W)|D, X] = \sum_{w \in \mathcal{S}} \ln p_{\theta}(D, w|X)p_{\theta_0}(D, w|D, X) .$$

For a fixed sequence of states (sample path)  $w$ , we have

$$p_{\theta}(D, w|X) = \pi_{w_0} \prod_{t=2}^T q_{w_{t-1}w_t,t} b_{w_t}(d_t) ,$$

such that

$$\begin{aligned} \mathbb{E}_{\theta_0}[\ln p_{\theta}(D, W)|D, X] &= \sum_{w \in \mathcal{S}} (\ln \pi_{w_0}) p_{\theta_0}(D, w|D, X) \\ &+ \sum_{w \in \mathcal{S}} \left( \sum_{t=2}^T \ln p(w_{t-1}|w_t) \right) p_{\theta_0}(D, w|D, X) + \sum_{w \in \mathcal{S}} \left( \sum_{t=2}^T \ln b_{w_t}(d_t) \right) p_{\theta_0}(D, w|D, X) . \end{aligned} \quad (11)$$

Our objective is to maximize (11) with respect to  $\theta$ . Since the parameters appear in groups, we can seek the maximum of each of the three sums in the above display separately. Update formulas for initial distribution vector are given by  $\pi_i^* = \gamma_i(1)$ . The second component of the likelihood contains the time-dependent probabilities  $\mathbf{Q}_{t,ij}$ . Those are described by the coefficients  $\Phi_{ij}$  and  $\eta_{ij}$ . As the maximum likelihood estimators do not have a closed form expression, we resort to numerical maximization for this part of the likelihood. If we can assume, that  $\mathbf{Q}_{t,ij} \equiv \mathbf{Q}_{ij}$  (i.e., transition probabilities are constant over time), than the

update formulas are given by

$$Q_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} .$$

Finally, the updates for  $\alpha_i$ 's are given by:

$$\alpha_i^* = \frac{\sum_{t=2}^T d_t \gamma_i(t)}{\sum_{t=2}^T N_t \gamma_i(t)} .$$

## Hamilton algorithm

Define iteratively for  $t = 1, \dots, T$  vectors of *prediction probabilities*

$$\xi_{t|t-1} = \begin{bmatrix} P_\theta(W_t = 1 | \mathcal{F}_{t-1}) \\ \dots \\ P_\theta(W_t = s | \mathcal{F}_{t-1}) \end{bmatrix} , \text{ where } \xi_{1|0} = \begin{bmatrix} \pi_1 \\ \dots \\ \pi_s \end{bmatrix} .$$

For  $t = 1, \dots, T$  it holds that

$$\xi_{t+1|t} = \frac{\mathbf{Q}'_t(\xi_{t|t-1} \circ \eta_t)}{\mathbf{1}'(\xi_{t|t-1} \circ \eta_t)} , \text{ with } \eta_t = \begin{bmatrix} P_\theta(D_t = d_t | W_t = 1, \mathcal{F}_{t-1}) \\ \dots \\ P_\theta(D_t = d_t | W_t = s, \mathcal{F}_{t-1}) \end{bmatrix} ,$$

with  $\mathbf{1}$  denoting a vector of ones, and  $\circ$  denoting an element-by-element multiplication.

As a by-product we obtain *filtered probabilities*, representing the distribution of the latent process at time  $t$  based on the information available at that time,

$$\xi_{t|t} = \frac{\xi_{t|t-1} \circ \eta_t}{\mathbf{1}'(\xi_{t|t-1} \circ \eta_t)} .$$

Furthermore, we can compute *smoothed probabilities*, useful for reproducing the evolution of the hidden state process,

$$\xi_{t|T} = \begin{bmatrix} P_\theta(W_t = 1 | \mathcal{F}_T) \\ \dots \\ P_\theta(W_t = s | \mathcal{F}_T) \end{bmatrix} ,$$

with  $\xi_{T|T}$  obviously equal to the most recent filtered probability. The probabilities  $\xi_{t|T}$  are obtained through a backward recursion

$$\xi_{t|T} = \xi_{t|t} \circ \{ \mathbf{Q}_t (\xi_{t+1|T} \doteq \xi_{t+1|t}) \}, \quad t = T - 1, \dots, 1, \quad (12)$$

where  $\doteq$  denotes element-by-element division.



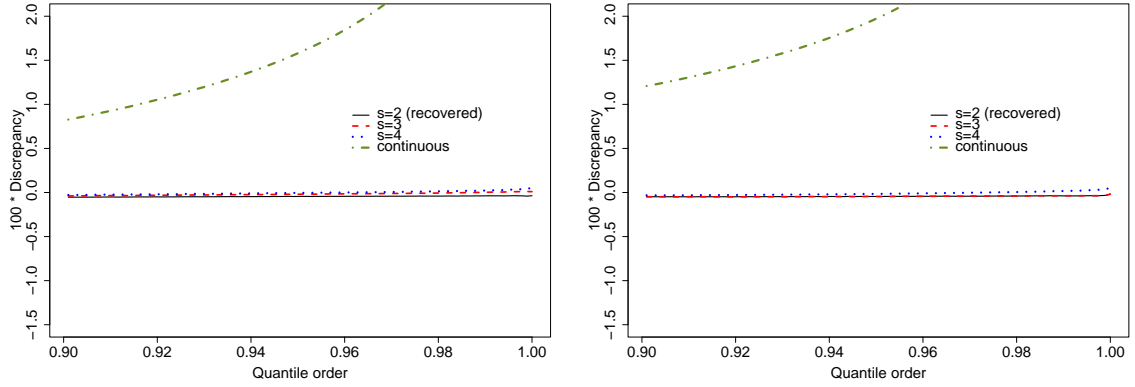


Figure 1: Bias in quantile functions for a 2-state HMM

This figure plots the average bias in estimated quantiles over 1,000 simulated samples. The DGP is a bivariate ( $s = 2$ ) HMM with parameters as given in Table 2. The focus is on the right-tail of the default rate distribution, i.e., quantiles 90%...100%. The left-hand graph plots the average (over 1,000 simulated samples) of  $100 \cdot (\hat{Q}_{s^*,h}^{unc} - Q_{s,h}^{unc})$  for  $h = 4$  quarters ahead as a function of the quantile. For each simulated sample,  $Q_{s^*,h}^{unc}$  as defined in (9) gives the quantile (based on 2,500 out-of-sample simulations) of the  $h$ -step-ahead forecast of the default rate from the HMM model of dimension  $s^*$ . The distribution of the hidden state  $W_T$  at the time of the forecast is the unconditional distribution given the true ( $Q$ ) or estimated ( $\hat{Q}$ ) parameters. The right-hand panel plots  $100 \cdot (\hat{Q}_{s^*,h}^{cond} - Q_{s,h}^{deg})$ . The initial distribution of  $W_T$  for  $\hat{Q}_{s^*,h}^{cond}$  is given by the smoothed distribution given the estimated parameters under an  $s^*$ -state HMM model. For  $Q_{s,h}^{deg}$  the true parameters are used and  $W_T$  is started from its simulated value.

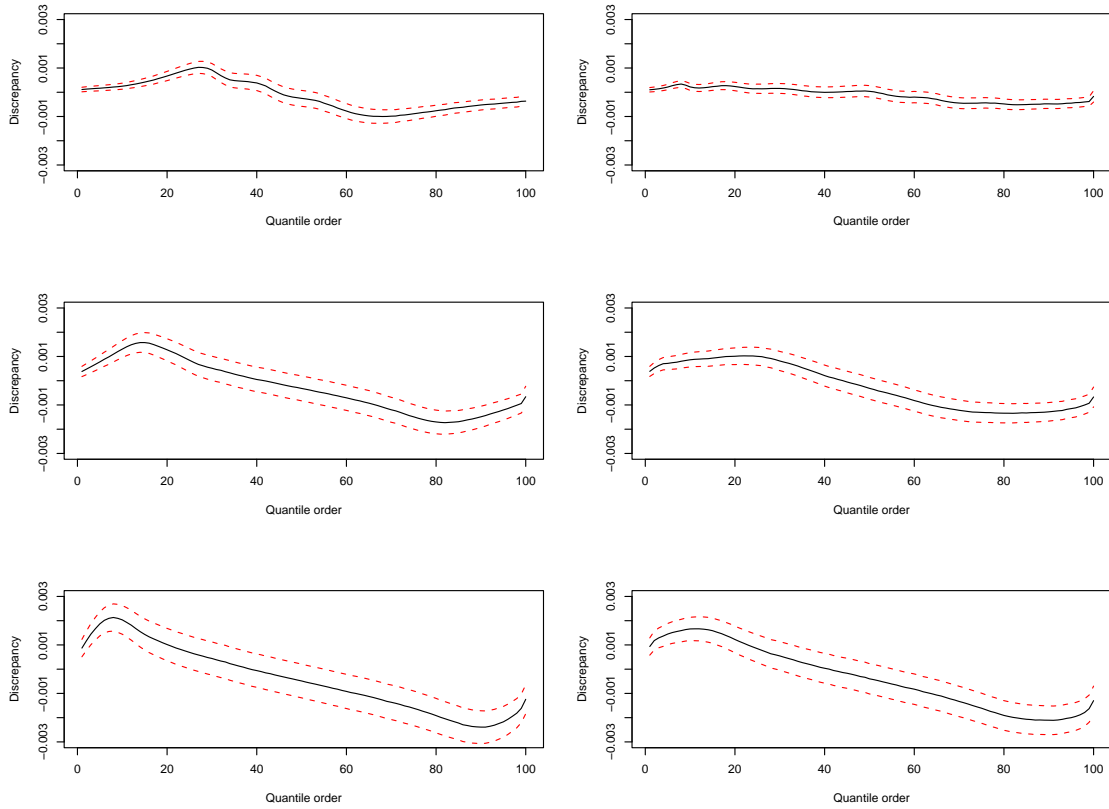


Figure 2: Bias in quantile functions for different forecast horizons  $h$

This figure plots the average bias in estimated quantiles over 1,000 simulated samples as a function of the quantile (horizontal axis). The DGP is a 2-regime HMM. Quantile biases are estimated as described in the note to Figure 1. The first, second, and third line of plots gives the results for 4, 8, and 12 quarters ahead, respectively. The left column of graphs plots the bias in the unconditional quantile functions, whereas the right-hand column gives the results for the conditional quantiles, where conditional is conditional on the state of the hidden Markov chain. The bands around the average are obtained by increasing or decreasing the average by 1.96 times the pointwise standard deviation over the simulations.

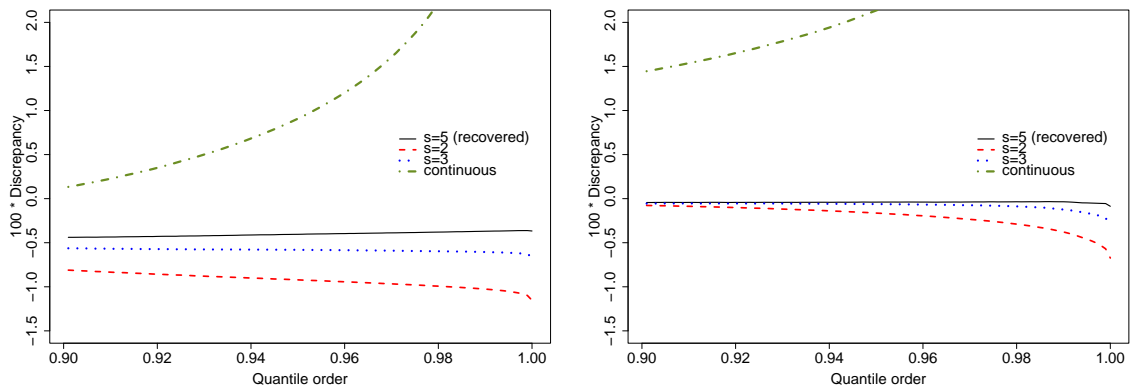


Figure 3: Bias in quantile functions for a 5-state HMM

This figure plots the average bias in estimated quantiles over 1,000 simulated samples. The DGP is a quinti-variate ( $s = 5$ ) HMM with parameters as given in Table 2. The focus is on the right-tail of the default rate distribution. Biases are computed as described in the note to Figure 1.

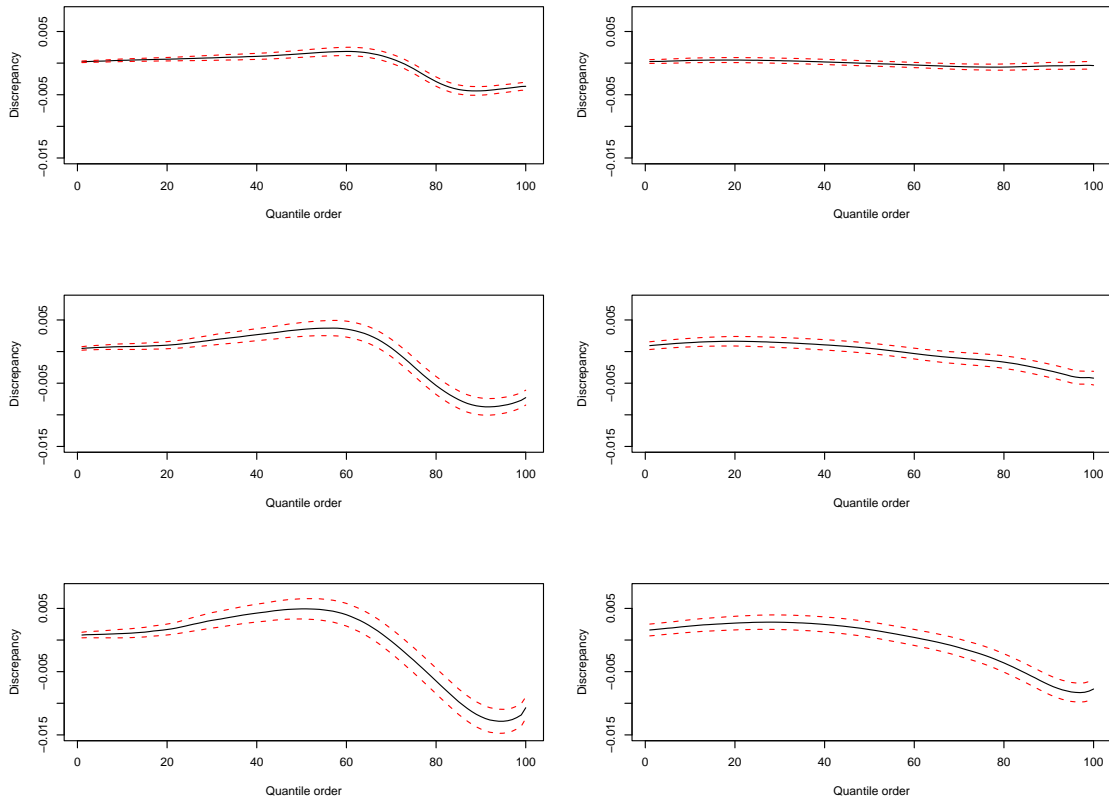


Figure 4: Bias in quantile functions for different forecast horizons  $h$

This figure plots the average bias in estimated quantiles over 1,000 simulated samples as a function of the quantile (horizontal axis). The DGP is a 5-regime HMM. Quantile biases are estimated as described in the note to Figure 1. The first, second, and third line of plots gives the results for 4, 8, and 12 quarters ahead, respectively. The left column of graphs plots the bias in the unconditional quantile functions, whereas the right-hand column gives the results for the conditional quantiles, where conditional is conditional on the state of the hidden Markov chain. The bands around the average are obtained by increasing or decreasing the average by 1.96 times the pointwise standard deviation over the simulations.

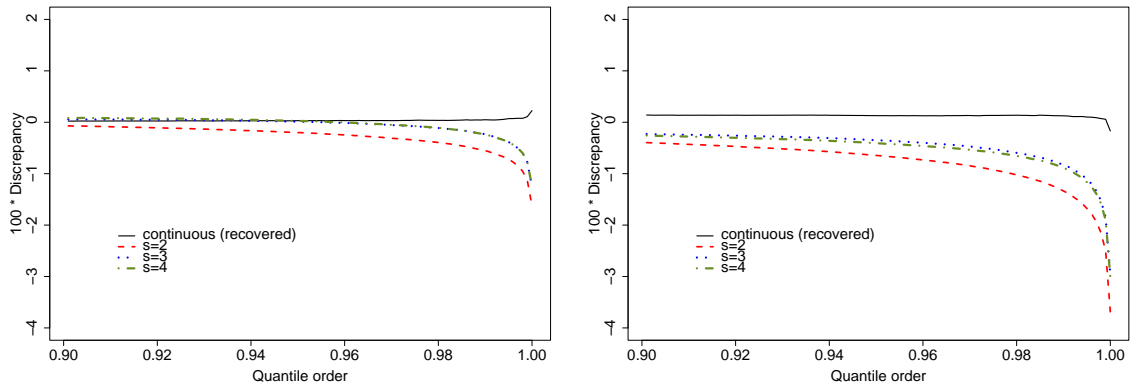


Figure 5: Bias in quantile functions for a continuous state HMM

This figure plots the average bias in estimated quantiles over 1,000 simulated samples. The DGP is a continuous state HMM with parameters as given in Table 2. The focus is on the right-tail of the default rate distribution, i.e., quantiles 90%...100%. Quantile biases are estimated as described in the note to Figure 1.

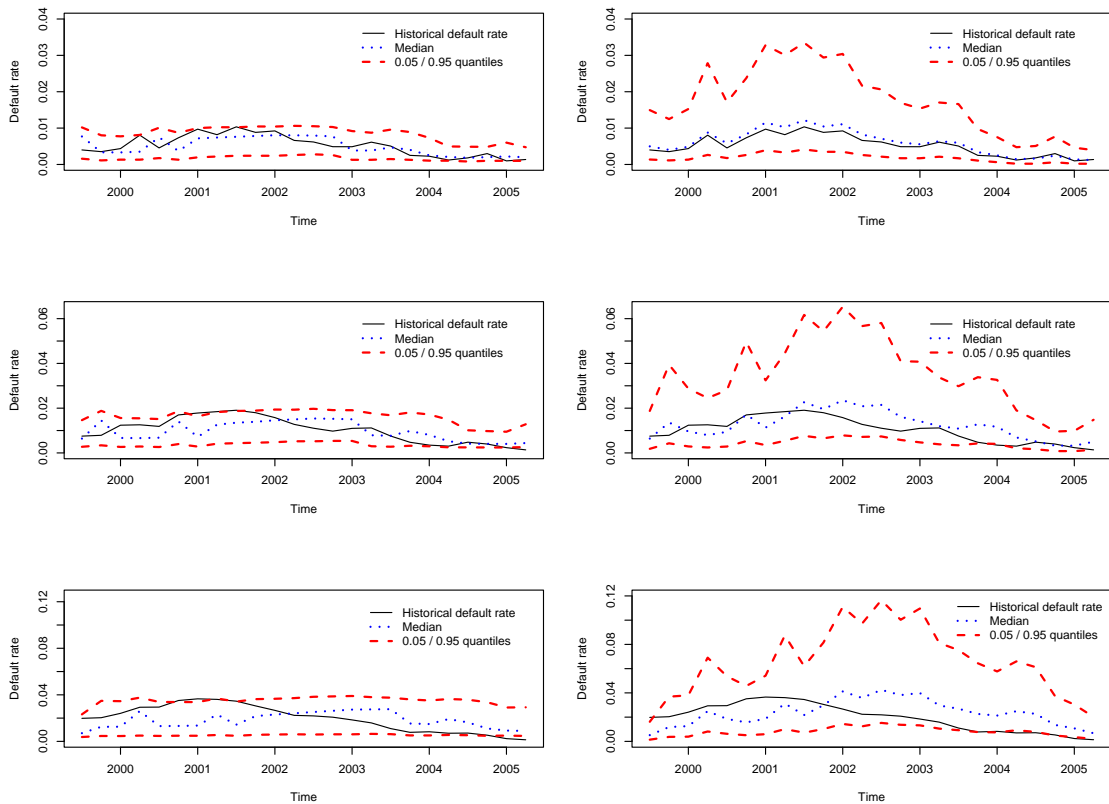


Figure 6: Empirical forecasts for S&P corporate default data  $h = 1, 2, 4$

This figure plots the  $h = 1$  (top line),  $h = 2$  (middle line), and  $h = 4$  (bottom line) quarter ahead forecasts of the corporate default rates based on a 3 regime (left column) and a continuous state (right column) HMM model. We plot the empirical, realized  $h$ -quarter default rate, the model median forecast, and the model forecast of the 5th and 95th percentile.

Table 1: Absolute and relative biases of conditional quantile forecasts

This table gives the absolute and relative bias of unconditional quantile forecasts for a 4-quarters-ahead forecasting horizon. The DGP is an  $s$ -regime HMM with parameters as given in Table 2, where  $s = \infty$  denotes the continuous state model. Quantile biases are estimated as described in the note to Figure 1. The numbers are expressed as percentages.

$s^*$	Absolute bias				Relative bias			
	$(100 \cdot (\hat{Q}_{s^*,4}^{cond} - Q_{s,4}^{deg}))$				$(100 \cdot (\hat{Q}_{s^*,4}^{cond}/Q_{s,4}^{deg} - 1))$			
	quantile				quantile			
	0.50	0.95	0.97	0.99	0.50	0.95	0.97	0.99
Panel A: $s = 2$								
2	0.00	-0.04	-0.04	-0.03	0.24	-1.74	-1.58	-1.36
3	0.00	-0.05	-0.04	-0.03	0.19	-1.75	-1.53	-1.36
4	0.00	-0.02	-0.02	0.01	0.23	-0.60	-0.07	0.58
$\infty$	0.55	1.97	2.62	4.18	3.63	76.54	98.46	149.44
Panel B: $s = 5$								
2	-0.04	-0.16	-0.23	-0.37	-2.42	-6.30	-8.53	-12.45
3	-0.01	-0.06	-0.07	-0.12	-0.52	-2.40	-2.71	-4.06
5	-0.01	-0.03	-0.03	-0.04	-0.34	-1.51	-1.37	-1.16
$\infty$	0.14	2.13	2.71	4.11	9.11	82.25	98.52	135.40
Panel C: $s = \infty$								
2	0.03	-0.64	-0.84	-1.34	5.36	-40.38	-46.16	-56.01
3	0.01	-0.34	-0.46	-0.82	2.47	-21.84	-25.61	-34.61
4	0.01	-0.41	-0.53	-0.89	2.31	-25.76	-29.32	-37.35
$\infty$	0.11	0.14	0.13	0.09	1.99	8.92	7.46	3.79

Table 2: Estimates for the Discrete state models, S&P data

This table gives the estimates of the transition matrix  $P$  of the hidden Markov chain  $W_t$  for the discrete state model of dimension  $s^* = 2, 3, 5$  using the S&P database for 1981–2005. The smoothed probabilities of  $W_T$  over the different states are in the vector  $\pi_T$ . The default probabilities in the different regimes  $W_t$  are denoted by  $\alpha_{W_t}$ . All numbers are denoted as percentages. States are denoted as *from* state  $i$  *to* state  $j$ .

	$i$	$s^* = 2$		$s^* = 3$			$s^* = 5$				
		$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$P_{ij}$	1	90.21	9.79	83.38	16.62	0.00	0.12	62.98	36.56	0.35	0.00
	2	15.99	84.01	25.47	58.80	15.74	6.32	73.37	20.29	0.02	0.00
	3			0.09	19.27	80.65	1.16	20.77	61.83	16.24	0.00
	4						0.06	2.49	25.02	53.68	18.75
	5						0.00	0.00	0.00	16.59	83.41
$\pi_T$		100.00	0.00	100.00	0.00	0.00	0.55	99.45	0.00	0.00	0.00
$\alpha_{W_t}$		0.22	0.69	0.18	0.43	0.82	0.14	0.19	0.31	0.53	0.86

Table 3: Information criteria for empirical data

$s^*$	$\ln \mathcal{L}$	AIC	BIC
2	-304.81	619.63	632.55
3	-284.97	591.95	620.38
4	-279.63	597.27	646.38
5	-279.30	616.60	691.56
$\infty$	-149.13	304.26	321.18