APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind. (2008)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.707

Modeling dependencies between rating categories and their effects on prediction in a credit risk portfolio

Claudia Czado[1,*,†] and Carolin Pflüger[2]

[1] *SCA Zentrum Mathematik, Technische Universität München, Garching, Germany*
[2] *Harvard Business School, Cambridge, MA, U.S.A.*

SUMMARY

The internal-rating-based Basel II approach increases the need for the development of more realistic default probability models. In this paper, we follow the approach taken in McNeil A and Wendin J [7], (*J. Empirical Finance* 2007) by constructing generalized linear mixed models for estimating default probabilities from annual data on companies with different credit ratings. The models considered, in contrast to McNeil A and Wendin J [7], (*J. Empirical Finance* 2007), allow parsimonious parametric models to capture simultaneously dependencies of the default probabilities on time and credit ratings. Macro-economic variables can also be included. Estimation of all model parameters are facilitated with a Bayesian approach using Markov chain Monte Carlo methods. Special emphasis is given to the investigation of predictive capabilities of the models considered. In particular, predictable model specifications are used. The empirical study using default data from Standard and Poor's gives evidence that the correlation between credit ratings further apart decreases and is higher than the one induced by the autoregressive time dynamics. Copyright © 2008 John Wiley & Sons, Ltd.

Received 24 October 2007; Accepted 23 November 2007

KEY WORDS: credit risk; default probability; asset correlation; generalized linear mixed models; Markov chain Monte Carlo; prediction

1. INTRODUCTION

The Basel II (2004) agreement allows financial institutions to choose an internal-rating-based (IRB) approach to calculate the capital requirement for credit risk. McNeil *et al.* [1] provide in Chapter 8 a survey of default probability models used in credit risk management. In particular, threshold and

*Correspondence to: Claudia Czado, SCA Zentrum Mathematik, Technische Universität München, Garching, Germany.
†E-mail: cczado@ma.tum.de

Contract/grant sponsor: Deutsche Forschungsgemeinschaft; contract/grant number: CZ 86/1-1

Bernoulli mixture models are discussed. The risk weight formulas to be used in an IRB Basel II approach can be derived from a one-factor Gaussian threshold model with preassigned constant asset correlation (see, for example, Section 8.4.5 of [1]).

However, it has long been known that the value at risk and other risk indicators of a credit portfolio are sensitive to the accuracy of the estimation of default correlations. As the data are scarce, it is a challenge to estimate the default correlations among the creditors correctly. In particular, empirical studies have shown that credit default correlations vary over time [2], rating classes and industry sectors (see [3–5]) and macro-economic variables [6]. This empirical evidence of varying default correlations is not reflected so far in the Basel II approach. McNeil and Wendin [7] have utilized generalized linear mixed models (GLMM) to capture these dependencies. As [1] point out on page 403, a GLMM modeling approach allows for flexible models, which can incorporate macro-economic information as well as dynamic dependence structures. In contrast to standard industry credit risk models such as Credit Metrics or the KMV model, in a GLMM setup all model parameters are estimated jointly and no external data sources are needed for model parameters. In particular, McNeil and Wendin [7] studied Bernoulli mixture models with time-dependent random effects. The time dynamic is modeled through a latent autoregressive component. For their analysis, they used a Bayesian approach applying Markov chain Monte Carlo (MCMC) techniques to facilitate parameter estimation and inference in a dynamic setting. This is a very powerful estimation method, since estimation of all parameters is conducted in a single step and the dependence structure assumed allows one to borrow strength for the fit of an area with scarce data from areas with more information. MCMC methods are summarized in [8] and discussed in detail in [9], while many examples are provided in [10]. The empirical study presented in [7] using Standard & Poor's data on U.S. firms clearly demonstrated the usefulness and potential of their approach. In particular, they investigated an unstructured model for modeling the dependency among default events on rating categories using a large number of parameters.

We would like to extend their work in two directions: Firstly, it would be interesting to see whether one can use more parsimonious models to uncover the structure of this dependency on rating categories. For instance, we would like to analyze whether the dependence on the rating categories is constant over all categories or whether the dependence decays for categories with their risk rating further apart. Secondly, for the application of such models for credit portfolio management, it is vital to investigate the usefulness of these models for prediction. In light of the recent increase in U.S. morgage defaults, the aspect of predicting the default probabilities for each rating categories becomes very interesting for investors of mortgage portfolios.

For the first question, we propose to model the joint dynamic over time and rating categories using a vector autoregressive latent component with different correlation structures of the error model. This allows us to model different correlation structures among the rating categories. The model fit of the considered models was assessed using the well-established deviance information criterion (DIC) of [11] for models fitted by MCMC techniques in addition to graphical checks.

For the second question, we consider models for prediction of one time period, which allows information to be included up to the previous time period. For the macro-economic variables included in the models this means using a time-shifted version of the variable. To compare the predictions we used the Brier score [12], conditional predictive ordinate (CPO) and standardized predictive residuals proposed by Gelfand [13] and also utilized in [7]. Finally, we also investigated the information loss resulting from only using the macro-economic information available up to the previous year rather than the complete information.

DOI: 10.1002/asmb

To illustrate our approach, we analyze annual data from Standard & Poor's from 1981 to 2005. As in [7], we included the Chicago Fed National Activity Index (CFNAI) as a macro-economic variable to capture the cyclical component of the systematic risk due to common economic conditions. With regard to prediction, our analysis suggests that the dependence induced by the rating classes decays for rating classes further apart rather than being constant over all rating classes. When considering predictions using a time-shifted CFNAI, the model with decaying dependencies shows the best predictive capability among the models investigated. The information loss from using this time-shifted variable in contrast to the unshifted one was seen not to be severe in this data set.

2. BAYESIAN INFERENCE FOR BINOMIAL MIXED REGRESSION

We start with a similar setup as [7], in particular, we assume that there are K different rating categories and T periods under consideration. Let m_{tk} be the number of firms in category k in time period t and M_{tk} the number of defaults in category k in time period t, $t=1,2,\ldots,T$; $k=1,2,\ldots,K$. One can then consider indicator variables $Y_{s,t,k}$ such that in time period t the sth obligor of rating category k defaults $Y_{s,t,k}$ takes values 1 and 0 otherwise. We consider models of the form

$$M_{tk}|b_{tk} \sim \text{Bin}(m_{tk}, g(\mu_k - \mathbf{x}_t'\boldsymbol{\beta} - b_{tk})) \text{ independent} \tag{1}$$

where $\mathbf{b}_t = (b_{t1}, \ldots, b_{tK})'$ represents the unobserved risk vector in time period t and has a specified distribution; $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$ and $\boldsymbol{\beta}$ are fixed, unknown parameters; $\mathbf{x}_t$ is a p-dimensional covariate vector representing observed macro-economic risk factors in time period t; and $g(\cdot)$ is a known link function for binomial data such as the logit or probit link. In our empirical study, we will use the logit link. In the following, we will explore different distributions for $\mathbf{b}_t, t=1,\ldots,T$.

The above model can be motivated as follows: Given $\mathbf{b}_t$, the default indicators $Y_{s,t,k}$ are independent and take value 1 with probability $g(\mu_k - \mathbf{x}_t'\boldsymbol{\beta} - b_{tk})$ and value 0 otherwise. By defining

$$V_{s,t,k} = \mathbf{x}_t'\boldsymbol{\beta} + b_{tk} + \varepsilon_{s,t,k} \tag{2}$$

where $\varepsilon_{s,t,k} \sim g$ i.i.d. we can reformulate the model as follows: The sth obligor in rating category k defaults in time period t iff $V_{s,t,k} < \mu_k$. $V_{s,t,k}$ can be thought to represent the asset value of the sth obligor in rating category k in time period t and μ_k can be thought of as the critical liability as laid out in [14]. The component $\mathbf{x}_t'\boldsymbol{\beta}$ represents the asset value attributable to the observed macro-economic market conditions, while b_{tk} is the contribution of rating category k in time period t. The idiosyncratic term $\varepsilon_{s,t,k}$ captures the contributions, which cannot be explained by global or rating category factors.

The implied asset correlation $\text{cor}(V_{s,t,k}, V_{r,\tau,l})$ is given by

$$\text{cor}(V_{s,t,k}, V_{r,\tau,l}) = \frac{\text{cov}(b_{t,k}, b_{\tau,l})}{\sqrt{\text{var}(b_{t,k}) + \omega^2}\sqrt{\text{var}(b_{t,k}) + \omega^2}} \tag{3}$$

where $\text{var}(\varepsilon_{s,t,k}) = \omega^2$. For the logit link, we have $\omega^2 = \pi^2/3$. We see that the distribution of $\mathbf{b}_t, t=1,\ldots,T$, can be used to capture different aspects of the default correlations. We discuss several such choices.

DOI: 10.1002/asmb

Our baseline model is

$$\mathbf{b}_t = (b_t, \ldots, b_t)' \quad \text{where } b_t \sim N(0, \sigma^2) \text{ i.i.d. (Model0)}$$

Therefore, the layout asset values $V_{s,t,k}$ and $V_{r,\tau,l}$ are independent for $t \neq \tau$. This assumption is clearly not realistic, because one surely would expect asset values in subsequent years to be correlated. Furthermore, the correlation between asset values of obligors in the same time period is always $\text{var}(b_t)/(\text{var}(b_t)+\omega^2)$, whether or not they are in the same rating category.

For the next model, we assume that the asset value correlations are time dependent but independent of the rating category. In particular, we assume that

$$\mathbf{b}_t = (b_t, \ldots, b_t)'$$

where

$$b_t = \alpha b_{t-1} + \sigma \varepsilon_t, \quad t = 1, 2, \ldots, T \text{ (Model1)}$$
$$b_0 = \sigma \varepsilon_0 / \sqrt{1-\alpha^2} \quad \text{with } \varepsilon_0, \varepsilon_1, \ldots, T \quad \text{i.i.d } N(0, 1)$$

This AR(1) time series for b_t has a $N(0, \sigma^2/(1-\alpha^2))$ stationary distribution for $|\alpha|<1$. This model was considered in [7]. The asset values of obligors in subsequent years are now correlated with $\text{cov}(b_{t-1}, b_t) = \sigma^2/(1+\alpha^2)$, $t = 1, 2, \ldots, T$, but correlations between asset values of obligors in the same time period are still constant over rating categories.

The next two models allow for category-dependent asset correlations. In Model2, b_{tk} is assumed to be a first-order vector autoregressive AR(1) time series with

$$\mathbf{b_t} = \alpha \mathbf{b_{t-1}} + \boldsymbol{\varepsilon}_\mathbf{t}, \quad t = 1, 2, \ldots, T$$
$$\mathbf{b_0} = \boldsymbol{\varepsilon}_\mathbf{0} / \sqrt{1-\alpha^2}$$

where $\boldsymbol{\varepsilon}_\mathbf{0}, \boldsymbol{\varepsilon}_\mathbf{1}, \ldots$ are i.i.d. $N_K(\mathbf{0}, \Phi)$ with

$$\Phi = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \ldots & \rho^{K-1} \\ \rho & 1 & \ldots & \rho^{K-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{K-1} & \rho^{K-2} & \ldots & 1 \end{pmatrix} \quad \text{(Model2)} \tag{4}$$

Here $N_n(\boldsymbol{\mu}, \Sigma)$ denotes an n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Model2 introduces implied asset correlations

$$\text{cor}(V_{s,t,k}, V_{r,\tau,l}) = \frac{\Phi_{k,l}\alpha^{|t-\tau|}/(1-\alpha^2)}{\sigma^2/((1-\alpha^2)(1-\rho^2))+\omega^2} = \frac{\dfrac{\sigma^2}{(1-\rho^2)(1-\alpha^2)}\rho^{|k-l|}\alpha^{|t-\tau|}}{\sigma^2/((1-\alpha^2)(1-\rho^2))+\omega^2} \tag{5}$$

Here the asset values of obligors in the same rating category are most strongly correlated, and the asset values of obligors in similar rating categories are more closely correlated than those of obligors in disparate rating categories.

DOI: 10.1002/asmb

The final model considered is similar to Model2, only the covariance matrix Φ is replaced by

$$\Phi=\frac{\sigma^2}{1-\rho^2}\begin{pmatrix}1 & \rho & \dots & \rho\\ \rho & 1 & \dots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \dots & 1\end{pmatrix} \quad \text{(Model3)} \tag{6}$$

The implied asset correlations for Model3 are

$$\operatorname{cor}(V_{s,t,k}, V_{r,\tau,l})=\frac{\Phi_{k,l}\alpha^{|t-\tau|}/(1-\alpha^2)}{\sigma^2/(1-\alpha^2)+\omega^2}=\frac{\dfrac{\sigma^2}{(1-\rho^2)(1-\alpha^2)}\rho^{\mathbf{1}(k\neq l)}\alpha^{|t-\tau|}}{\sigma^2/(1-\alpha^2)+\omega^2} \tag{7}$$

where $\mathbf{1}(k\neq l)$ takes the value 1 if $k\neq l$ and 0 otherwise. This model incorporates the assumption that asset values of obligors in the same rating category are more closely correlated than those of obligors in different rating categories; however, for obligors in different rating categories it does not make a difference whether or not their rating categories are similar.

To complete the model formulation for a Bayesian setup, we have to specify the prior distributions. As in [7] we choose non-informative priors for the parameters and hyperparameters of our models. In all models, we chose an ordered normal $N_K(\mathbf{0},\tau_\mu^2 I)$ distribution as prior for $\boldsymbol{\mu}$ with $\tau_\mu=100\,000$, i.e. we require $\mu_1>\mu_2>\mu_3>\mu_4>\mu_5$ and the prior distribution then is $N_K(\mathbf{0},\tau_\mu^2 I)\mathbf{I}_{\mu_1>\mu_2>\mu_3>\mu_4>\mu_5}$. The variance σ^2 was given an improper prior decaying as $1/x$. This corresponds to the limiting case $\sigma^2\sim\operatorname{Inv}\Gamma(\eta,v)$ with $\eta=0$ and $v=0$, where $\operatorname{Inv}\Gamma(\eta,v)$ denotes the inverse Gamma distribution with parameters η and v. The coefficient $\boldsymbol{\beta}$ was given a $N(\mathbf{0},\tau_\beta^2 I)$ prior, where $\tau_\beta=10\,000$. In Model1, Model2 and Model3, α was given a normal prior with mean 0 and standard deviation $\frac{1}{4}$ truncated to the interval $(-1,1)$. This informative prior was chosen to improve convergence of the Markov chain and had little influence on the quality of the estimates. In Model2 and Model3, the parameter ρ was given a uniform prior on $(-1,1)$.

We used an MCMC algorithm to simulate from the posterior distributions. Our algorithms update parameters one at a time. To simulate from univariate full conditional distributions, which are only known up to a constant, we apply the ARS (adaptive rejection sampling) and ARMS algorithms [15, 16]. The former is intended for log-concave densities only, whereas the latter can be applied to a wide range of univariate densities. Only in one case this ARMS-algorithm was found not to work, and hence a Metropolis-sampling step had to be employed. For every model, 10 000 iterations were used as a burn-in to give the sampler the opportunity to settle down to equilibrium. The estimates were based on the following 200 000 iterations. Every 40th iteration was used so that there was a sample of 5000 simulations available for analysis. This sub-sampling frequency was chosen after having considered autocorrelation functions for the simulated values of the parameters.

For both prediction and estimation, we conducted our analyses using not only covariate values from the current year but also covariate values from the previous year. This was done to realistically simulate the situation of predicting default probabilities for the coming year, when only covariate values of the current year are available.

DOI: 10.1002/asmb

3. MODEL COMPARISON OF FIT AND PREDICTIVE POWER

To assess model fit, the complete data were used to estimate the posterior distributions. This gives estimates of quantiles, median, mean and standard deviation for all parameters for all models. Credible intervals, which are the Bayesian equivalent to confidence intervals, are used to assess the significance of parameters.

The DIC introduced by Spiegelhalter *et al.* [11] is used to compare the fit of different models. For a probability model $p(\mathbf{y}|\boldsymbol{\theta})$ with observed data $\mathbf{y}=(y_1, y_2, \ldots, y_n)$ it is defined as DIC$:= E[D(\boldsymbol{\theta}|\mathbf{y})]+p_D$. The posterior mean deviance is defined as $D(\boldsymbol{\theta})=-2\log(l(\mathbf{y}|\boldsymbol{\theta}))$ and corresponds to a Bayesian measure of fit or adequacy, whereas the effective number of parameters $p_D := E[D(\boldsymbol{\theta}|\mathbf{y})]-D(E[\theta|\mathbf{y}])$ is a measure of model complexity. In the case of a model with no random effects, p_D gives the number of parameters. Hence, this score considers both complexity and goodness of fit. When comparing models, the model with smallest DIC would be preferred.

Furthermore, seeing that in practical applications one is even more interested in the predictive quality of a model, we will consider this aspect carefully. To gain an idea of the predictive quality, we fitted the models using the data of all time periods except the last one and then computed the predictive distributions of the default probabilities for the last time period. To assess the goodness of those predictions, we will use the verification score introduced by Brier [12]. Let p_{tk}^{Obs} be the observed default probability in year t and rating category k and let p_{tkr} be the simulated value of the default probability in year t and rating category k from the rth iteration of the MCMC process. Assume that there are R iterations. As our predictions were made for 2005, the corresponding Brier score to measure the goodness of these predicted default probabilities is defined as

$$B=\frac{1}{R}\sum_{r=1}^{R}\sum_{k=1}^{K}(p_{2005kr}-p_{2005k}^{\text{Obs}})^2$$

Seeing that default probabilities vary strongly across rating categories, this Brier score assigns greater weight to riskier rating categories than to less risky rating categories. In order to adjust for this, we also considered a relative Brier score, which is defined by

$$B_{Re}=\frac{1}{R}\sum_{r=1}^{R}\sum_{k=1}^{K}((p_{2005kr}/p_{2005k}^{\text{Obs}})-1)^2$$

A model with small (relative) Brier score would be preferred.

Further, we used the category-specific CPO for 2005, which for rating category k is defined by

$$\text{CPO}_{2005,k}=p(N_{2005,k,\text{Obs}}|\{N_{t,1,\text{Obs}},\ldots,N_{t,K,\text{Obs}}\}, t\neq 2005)$$

$\text{CPO}_{2005,k}$ gives the conditional probability of observing $N_{2005,k,\text{Obs}}$ given all observations up to year 2004 and for a good model one would expect it to be large. Note that $\{\text{CPO}_{2005,k}, k=1,\ldots,K\}$ can be estimated using the MCMC iterates.

Finally, we considered the univariate, standardized predictive residual $d_{2005,k}$ defined by

$$d_{2005,k}:=\frac{N_{2005,k,\text{Obs}}-E(N_{2005,k}|\{N_{t,1,\text{Obs}},\ldots,N_{t,K,\text{Obs}}\}, t\neq 2005)}{\sqrt{\text{var}(N_{2005,k}|\{N_{t.1,\text{Obs}},\ldots,N_{t,K,\text{Obs}}\}, t\neq 2005)}}$$

Here, the model with small $|d_{2005,k}|$ would be preferred. The last two scores were also considered in [7], which facilitates comparison of results. For details of these scores, see [13]. Again $d_{2005,k}$ can be estimated using the MCMC iterates.

DOI: 10.1002/asmb

4. AN EMPIRICAL STUDY OF S&P DEFAULT DATA

4.1. Description of the data

The default data used are available from Standard and Poor's CreditPro^{TM} web site. It contains yearly default data from 1981 to 2005 in seven rating categories: 'CCC', 'B', 'BB', 'BBB', 'A', 'AA', 'AAA' ranked according to decreasing risk. Only categories CCC to A have been considered, because in categories AA and AAA, defaults are too rare to allow for statistical inference. Alternatively, one could also have combined the rating categories AAA, AA and A into one rating category 'A or above'. For simplicity, we will number the rating categories $1, \dots, 5$. The average number of firms per rating category per year is 450.

There are significant numbers of firms that were rated at the beginning of a year but not at the end of a year, so that there is no information available on whether or not they defaulted. These firms have been excluded from the analysis.

The CFNAI, which is published monthly, was used as a macro-economic indicator and its yearly average was used as the covariable. The CFNAI is based on data from the following broad categories: production and income; employment, unemployment and hours; personal consumption and housing; sales, orders and inventories and is thought to give a gauge on current and future economic activity and inflation.

4.2. Results

4.2.1. Estimation and model fit. Table I summarizes the posterior distributions of all parameters for all models considered for the unshifted and shifted CFNAI, respectively. One can see very clearly that β, when the unshifted CFNAI is used, is significantly $\neq 0$, i.e. the CFNAI is able to explain part of the inhomogeneity of default rates over time. Instead when the shifted CFNAI is used, the estimated regression coefficient is reduced and not quite significant, thus showing some information loss. The importance of the CFNAI is also illustrated in Figure 1, where one can see the fitted ($t < 2005$) and predicted ($t = 2005$) default probabilities from Model2, the observed default probabilities and (scaled) CFNAI in the same graph. One can observe that the CFNAI and the default probabilities behave very similarly over time. One can also see that for Model1, Model2 and Model3, the correlation parameter α for the time structure is significant, i.e. the time dependency of the unobserved risk helps to explain observed default probabilities. Moreover, in Model2 and Model3 correlation parameter ρ for the dependence between rating categories is distinct from 0 and higher than α, which indicates that the new correlation structures of Model2 and Model3 improve the fit. Further, the time correlation measured by α is lower than the correlation induced by the dependency between the categories measured by ρ.

Table II shows DIC scores using the unshifted CFNAI and the shifted CFNAI. For the unshifted and the shifted CFNAI, one can see that Model2 and Model3 have significantly lower DIC scores than Model0 and Model1, thus indicating a better fit when a dependency on the rating category is allowed. Since the DIC scores are the lowest for Model3, we see a slight preference in model fit for a equidistant correlation structure among the rating categories. The DIC values for Model0 and Model1 are quite similar for each of the two CFNAI specifications, which implies that the sole introduction of an unobserved autoregressive risk component does not improve the fit over the base model much. One can see that the DIC scores are consistently higher when using the shifted CFNAI than when using the unshifted CFNAI. This could be expected, because one would expect this year's CFNAI to give more relevant information than last year's CFNAI. However, the

DOI: 10.1002/asmb

Table I. Posterior mean estimates with estimated standard errors and estimated posterior quantiles based on complete data 1981–2005 with unshifted and shifted CFNAI, respectively.

	Unshifted CFNAI				Shifted CFNAI			
	10%	50%	Mean (Std dev.)	90%	10%	50%	Mean (Std dev.)	90%
μ_1								
Model0	−1.8	−1.0	−1.0 (0.13)	−0.9	−1.2	−1.0	−1.0 (0.14)	−0.8
Model1	−1.3	−1.0	−1.1 (0.18)	−0.8	−1.3	−1.0	−1.0 (0.19)	−0.8
Model2	−1.6	−1.2	−1.2 (0.46)	−0.9	−2.3	−1.1	−1.2 (0.96)	−0.7
Model3	−1.4	−1.1	−1.0 (0.42)	−0.8	−1.4	−1.1	−1.1 (0.23)	−0.9
μ_2								
Model0	−3.1	−3.0	−3.0 (0.12)	−2.8	−3.1	−3.0	−3.0 (0.13)	−2.8
Model1	−3.2	−3.1	−3.0 (0.17)	−2.8	−3.2	−3.0	−3.0 (0.19)	−2.8
Model2	−3.5	−3.1	−3.1 (0.35)	−2.8	−4.7	−3.1	−3.3 (0.83)	−2.7
Model3	−3.3	−3.0	−3.0 (0.40)	−2.7	−3.3	−3.0	−3.0 (0.23)	−2.8
μ_3								
Model0	−4.8	−4.6	−4.6 (0.14)	−4.4	−4.8	−4.6	−4.6 (0.15)	−4.4
Model1	−4.9	−4.6	−4.7 (0.19)	−4.4	−4.9	−4.6	−4.6 (0.20)	−4.4
Model2	−5.1	−4.7	−4.7 (0.39)	−4.4	−6.5	−4.7	−4.0 (0.86)	−4.4
Model3	−4.9	−4.7	−4.6 (0.47)	−4.3	−5.0	−4.7	−4.7 (0.24)	−4.4
μ_4								
Model0	−6.3	−6.1	−6.0 (0.19)	−5.8	−6.3	−6.0	−6.0 (0.19)	−5.8
Model1	−6.4	−6.1	−6.1 (0.22)	−5.8	−6.4	−6.1	−6.1 (0.24)	−5.8
Model2	−6.6	−6.2	−6.2 (0.37)	−5.8	−8.9	−6.2	−6.6 (1.20)	−5.8
Model3	−6.5	−6.1	−6.1 (0.49)	−5.8	−6.5	−6.1	−6.1 (0.28)	−5.8
μ_5								
Model0	−8.5	−8.0	−8.0 (0.38)	−7.5	−8.5	−8.0	−8.0 (0.38)	−7.5
Model1	−8.6	−8.0	−8.0 (0.39)	−7.6	−8.6	−8.1	−8.0 (0.41)	−7.5
Model2	−8.7	−8.1	−8.1 (0.53)	−7.5	−10.6	−8.2	−8.6 (1.30)	−7.6
Model3	−8.8	−8.1	−8.0 (0.50)	−7.5	−8.6	−8.1	−8.1 (0.43)	−7.6
β								
Model0	0.28	0.53	0.53 (0.19)	0.77	−0.06	0.21	0.22 (0.22)	0.49
Model1	0.25	0.49	0.49 (0.18)	0.72	−0.07	0.19	0.19 (0.20)	0.45
Model2	0.27	0.50	0.50 (0.18)	0.73	−0.05	0.21	0.20 (0.20)	0.44
Model3	0.28	0.51	0.51 (0.18)	0.72	−0.05	0.20	0.20 (0.20)	0.45
σ								
Model0	0.39	0.50	0.50 (0.10)	0.64	0.48	0.59	0.60 (0.11)	0.76
Model1	0.35	0.45	0.46 (0.09)	0.58	0.42	0.53	0.54 (0.10)	0.68
Model2	0.12	0.22	0.22 (0.08)	0.32	0.10	0.21	0.21 (0.08)	0.32
Model3	0.36	0.46	0.46 (0.09)	0.58	0.43	0.53	0.54 (0.09)	0.66
α								
Model0	—	—	—	—	—	—	—	—
Model1	0.10	0.35	0.35 (0.19)	0.59	0.10	0.34	0.33 (0.18)	0.57
Model2	0.22	0.49	0.51 (0.24)	0.87	0.22	0.52	0.56 (0.28)	1.00
Model3	0.16	0.44	0.46 (0.24)	0.81	0.16	0.40	0.40 (0.19)	0.64
ρ								
Model0	—	—	—	—	—	—	—	—
Model1	—	—	—	—	—	—	—	—
Model2	0.75	0.88	0.87 (0.090)	0.96	0.81	0.92	0.90 (0.07)	0.98
Model3	0.73	0.82	0.82 (0.068)	0.91	0.75	0.84	0.83 (0.06)	0.91

DOI: 10.1002/asmb

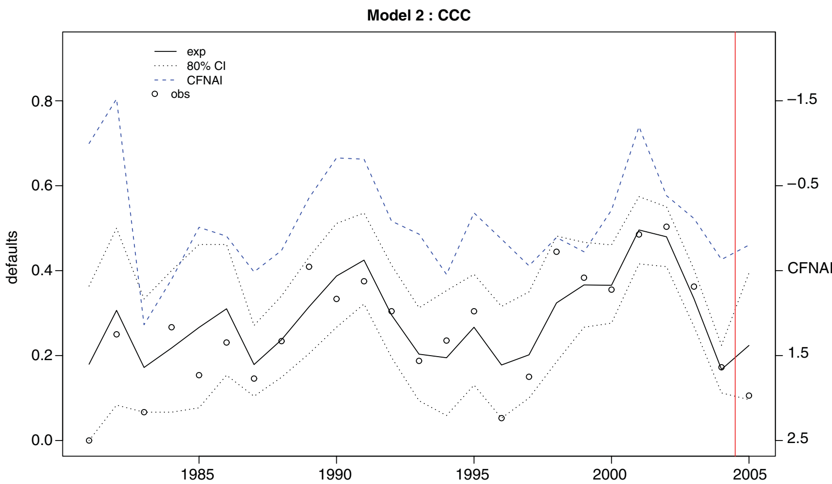

Figure 1. Fitted and predicted default probabilities (solid line) in Model2 with unshifted CFNAI (dashed line).

Table II. DIC score to assess model fit for the considered models.

Index not shifted	DIC	Effective number of parameters
Model0	7827.04	25.81
Model1	7826.28	24.93
Model2	7818.11	38.42
Model3	7816.40	38.64
Index shifted	DIC	Effective number of parameters
Model0	8102.44	26.23
Model1	8102.68	25.94
Model2	8096.70	38.86
Model3	8094.83	41.52

Models are fitted with data 1981–2005 using unshifted CFNAI (top) and shifted CFNAI (bottom).

DIC scores using the shifted CFNAI are in the range of those using the unshifted CFNAI, which indicates that using the shifted CFNAI one still obtains an acceptable model fit.

In Figure 2 one can see the observed default rates, the posterior fitted default probabilities for 1981–2004 and the predicted default probabilities for 2005 using the unshifted CFNAI. These probabilities are shown for all models and all rating categories except for category A (too few defaults). For instance, concentrating on rating category B and the year 1990, one can see that although for Model0 and Model1 the observed default probability is not in the fitted 80% credible interval, it is in this interval for Model2 and Model3. One can also see that the fitted expected default probability 1990 is clearly closer to observed default probability 1990 for Model2 and Model3. This again illustrates the improved fit of Model2 and Model3. For comparison, we also include a similar plot using the shifted CFNAI (see Figure 3). We see that there is no large difference in the fit and the prediction.

DOI: 10.1002/asmb

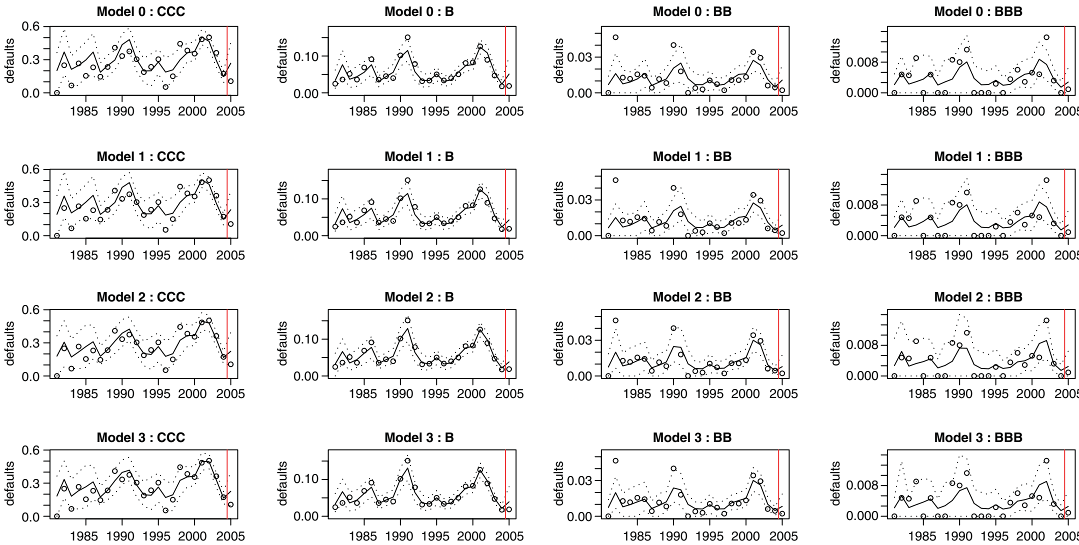

Figure 2. Estimated posterior mean default probabilities for $t = 1981, \ldots, 2004$ and predicted default probability for 2005 (solid line) with 80% credible intervals (dotted lines) and observed default probabilities (o) using the unshifted CFNAI.

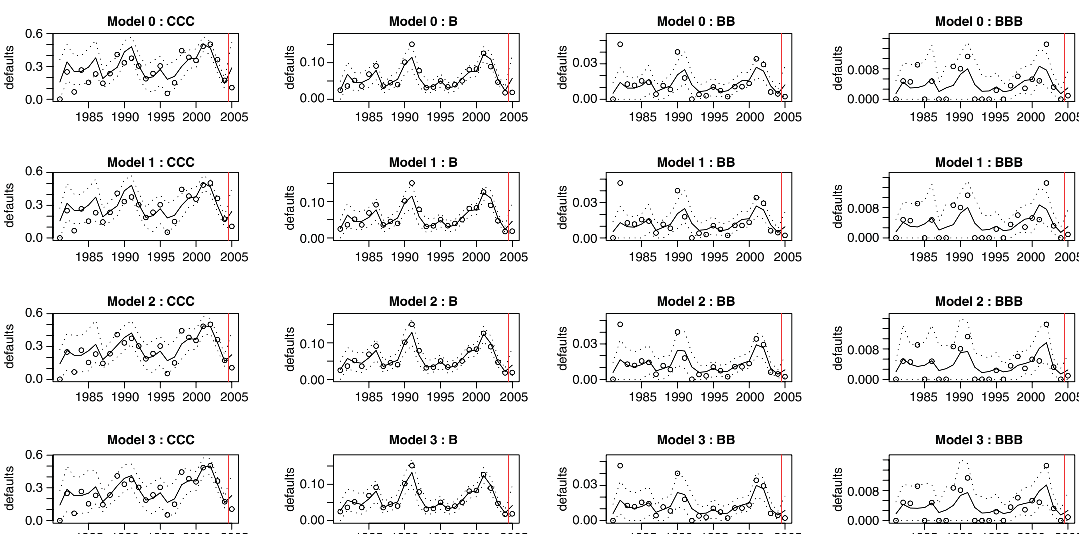

Figure 3. Estimated posterior mean default probabilities for $t = 1981, \ldots, 2004$ and predicted default probability for 2005 (solid line) with 80% credible intervals (dotted lines) and observed default probabilities (o) using the shifted CFNAI.

DOI: 10.1002/asmb

From a model fit perspective, one might be led to conclude that Model3 is slightly preferable to Model2. This is somewhat not expected, since we would expect Model2 to fit better than Model3. The reasoning behind this is that companies in adjacent rating categories share more characteristics and are exposed to more similar kind of risks than those in rating categories further apart. However, we like to note that the DIC is only approximate in exponential family models and problems with the DIC measure have been reported elsewhere. Therefore, DIC should be used only as a rough guideline. We place greater emphasis on checking the predictive capabilities of the models, since this is the primary interest of the data analyst.

4.2.2. Analysis of predictive distributions. Figure 4 and the top part of Table III summarize the predictive distributions for 2005 obtained for the different models using the unshifted CFNAI. In general, the point predictions such as the mode, mean and median of the predictive distribution are higher than the observed values. Further, the predictive distributions are skewed with a long right tail, so that mean and median are to the right of the mode of the distribution. Comparing the distributions obtained for rating category BBB, one can see in Figure 4 that the mode of the distribution is closest to the observed default probability for Model2. The same effect can be observed in the upper part of Table III. For comparsion, we also added the corresponding plot for the shifted CFNAI (see Figure 5). The predictive distributions are as expected less concentrated using the shifted CFNAI compared with those using the unshifted CFNAI.

The lower part of Table III shows the predicted default probabilities for 2005, but this time the shifted CFNAI was used. As one would expect, using this less informative covariate, one obtains

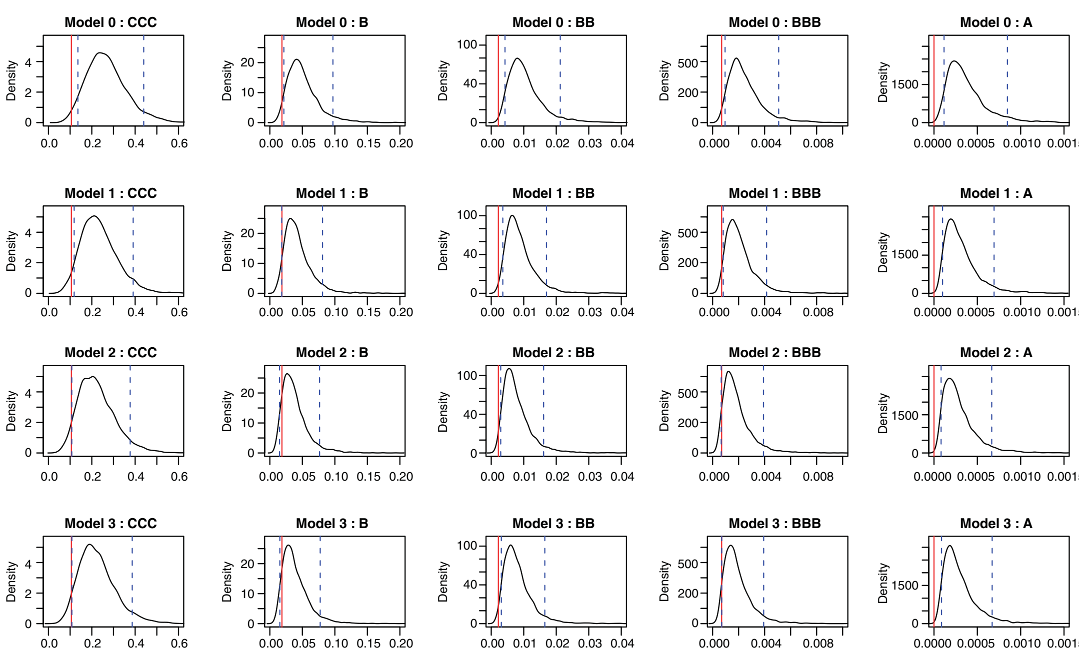

Figure 4. Predictive densities for 2005 in the different rating categories using unshifted CFNAI. The vertical solid line indicates the observed default probability in 2005 and the vertical dashed lines show the 90% credible interval.

DOI: 10.1002/asmb

Table III. Predicted default probabilities in 2005 with estimated standard errors and estimated quantiles of the predicted default distribution using the *unshifted* (top) and *shifted* CFNAI (bottom).

Rating	Observed	Model	5%	50%	Mean (Std dev.)	Mode	95%
Unshifted CFNAI							
CCC	0.11	Model0	0.14	0.26	0.27 (0.092)	0.24	0.43
		Model1	0.12	0.23	0.24 (0.085)	0.21	0.39
		Model2	0.11	0.21	0.22 (0.084)	0.20	0.38
		Model3	0.11	0.21	0.23 (0.085)	0.19	0.39
B	0.019	Model0	0.022	0.046	0.051 (0.025)	0.041	0.097
		Model1	0.018	0.039	0.043 (0.021)	0.032	0.081
		Model2	0.015	0.034	0.038 (0.020)	0.026	0.076
		Model3	0.015	0.035	0.039 (0.021)	0.028	0.077
BB	0.0022	Model0	0.0042	0.0094	0.0110 (0.0057)	0.0079	0.0212
		Model1	0.0036	0.0078	0.0089 (0.0048)	0.0064	0.0170
		Model2	0.0029	0.0068	0.0078 (0.0045)	0.0056	0.0161
		Model3	0.0031	0.0072	0.0082 (0.0047)	0.0060	0.0165
BBB	0.0007	Model0	0.0010	0.0022	0.0025 (0.0014)	0.0018	0.0050
		Model1	0.0008	0.0019	0.0021 (0.0012)	0.0015	0.0041
		Model2	0.0007	0.0016	0.0011 (0.0019)	0.0012	0.0039
		Model3	0.0007	0.0017	0.0019 (0.0011)	0.0014	0.0039
A	0	Model0	0.00012	0.00032	0.00038 (0.0003)	0.00024	0.00085
		Model1	0.00010	0.00026	0.00032 (0.0002)	0.00020	0.00069
		Model2	0.00008	0.00024	0.00029 (0.0002)	0.00018	0.00067
		Model3	0.00009	0.00025	0.00030 (0.0002)	0.00018	0.00067
Shifted CFNAI							
CCC	0.11	Model0	0.12	0.27	0.29 (0.12)	0.23	0.50
		Model1	0.10	0.23	0.24 (0.11)	0.19	0.44
		Model2	0.09	0.21	0.22 (0.10)	0.20	0.42
		Model3	0.09	0.21	0.23 (0.10)	0.20	0.42
B	0.019	Model0	0.019	0.049	0.057 (0.034)	0.037	0.122
		Model1	0.015	0.039	0.046 (0.028)	0.030	0.099
		Model2	0.012	0.033	0.039 (0.026)	0.023	0.086
		Model3	0.013	0.035	0.041 (0.027)	0.027	0.092
BB	0.0022	Model0	0.0037	0.0101	0.0120 (0.0080)	0.0077	0.0273
		Model1	0.0030	0.0080	0.0096 (0.0065)	0.0057	0.0213
		Model2	0.0024	0.0068	0.0081 (0.0059)	0.0044	0.0181
		Model3	0.0027	0.0072	0.0086 (0.0060)	0.0053	0.0191
BBB	0.0007	Model0	0.0009	0.0024	0.0029 (0.0020)	0.0018	0.0065
		Model1	0.0007	0.0019	0.0023 (0.0016)	0.0013	0.0051
		Model2	0.0005	0.0015	0.0019 (0.0014)	0.0011	0.0043
		Model3	0.0006	0.0017	0.0020 (0.0014)	0.0013	0.0046
A	0	Model0	0.00010	0.00034	0.00043 (0.0003)	0.00024	0.00108
		Model1	0.00008	0.00027	0.00034 (0.0003)	0.00018	0.00083
		Model2	0.00007	0.00023	0.00030 (0.0003)	0.00016	0.00074
		Model3	0.00008	0.00024	0.00031 (0.0003)	0.00016	0.00077

slightly larger standard deviations and larger 90% credible intervals. However, the modal values are closer to the observed values than in the upper part of Table III. This probably is due to the fact that the CFNAI mainly is an indicator for *future* economic activity and therefore is also highly relevant for the coming year.

DOI: 10.1002/asmb

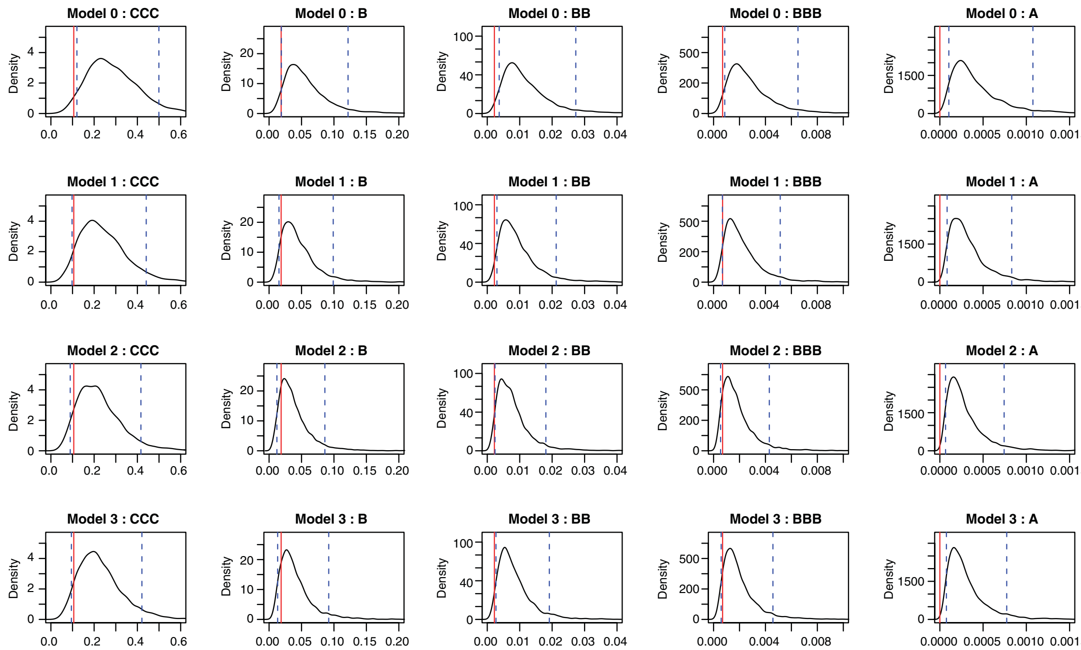

Figure 5. Predictive densities for 2005 in the different rating categories using shifted CFNAI. The vertical solid line indicates the observed default probability in 2005 and the vertical dashed lines show the 90% credible interval.

Figure 6 compares fit and prediction obtained for rating category B for all models using the shifted CFNAI and the unshifted CFNAI. The left panels of Figure 6 give the fitted ($t \neq 2005$) and the predicted ($t = 2005$) default probabilities, whereas the right panels of Figure 6 compare the predicted densities for 2005. Clearly, the base model Model0 gives the worst predictions, whereas the predictions in Model1 are better than those in Model0 for both unshifted and shifted CFNAI specifications. However, the predictive distribution of Model2 is the most concentrated predictive distribution with mode closest to the observed value.

The upper part of Table IV shows Brier scores and relative Brier scores for all models using the unshifted and the shifted CFNAI. Since in rating-category A the observed default probability is 0, we chose to divide by 10^{-4} instead, which is approximately the order of magnitude of the predictive default probabilities. Brier scores using the shifted CFNAI are higher for all models, but they are still in the range of Brier scores using the unshifted CFNAI, which means that while using the shifted CFNAI rather than the unshifted CFNAI impairs the predictive strength of our models; the predictions obtained using the shifted CNFAI are still reasonably good. These scores again support that Model2 has the best predictive qualities, closely followed by Model3. The same effects can also be observed in Table V. In the top and bottom parts, Model2 scores best for all rating categories and for shifted and unshifted CNFAI. Moreover, in the bottom part, one can see that scores using the non-shifted CFNAI are very close to the scores obtained using the shifted CFNAI. This again illustrates that using the shifted CFNAI does not impair predictions much.

DOI: 10.1002/asmb

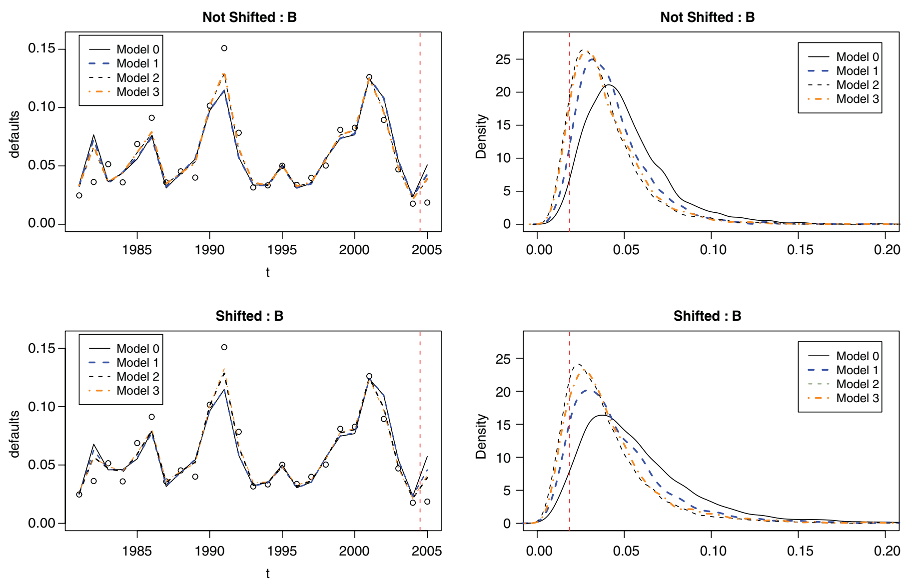

Figure 6. Left panels: fitted ($t \neq 2005$) and predicted ($t = 2005$) default probabilities for different covariate specifications; right panels: predictive default densities for rating category B in 2005.

Table IV. Brier scores for 2005 using 1981–2004 data with unshifted CFNAI and shifted CFNAI.

Index not shifted	Brier score	Relative Brier score
Model0	0.037	54
Model1	0.025	36
Model2	0.022	28
Model3	0.023	30
Index shifted	**Brier score**	**Relative Brier score**
Model0	0.049	86
Model1	0.032	51
Model2	0.026	37
Model3	0.027	41

We now consider the problem of predicting transition probabilities for each rating category. In Table VI we give the observed transition probabilities. Here a rating company could use their own rating rule. In lieu of this, we investigated the following *ad hoc* rating rule. To predict the transition probabilities, we use the predicted default probabilities p_{2005kr} for all rating categories and for 5000 MCMC recorded values after burn-in, which constitutes a sample from the posterior predictive distribution. We predict a default if this probability is greater than 0.5. We determine the empirical 20, 40, 60 and 80% quantiles of $p_{2005kr}, p_{2005kr} \leqslant 0.5$ over all k and r. The observation kr is classified as A if p_{2005kr} is less than or equal to 20% quantile, as BBB if it is between 20 and 40% quantiles, etc. These 5000 rating classifications are then used to construct the corresponding

DOI: 10.1002/asmb

Table V. Top part: absolute predictive residuals $|d_{2005,k}|$ for $k=1,\ldots,K$ using 1981–2004 data with unshifted CFNAI and shifted CFNAI; bottom part: conditional predictive ordinates $CPO_{2005,k}$ for $k=1,\ldots,K$ using 1981–2004 data with unshifted CFNAI and shifted CFNAI.

Index not shifted	$\lvert d_{2005,CCC}\rvert$	$\lvert d_{2005,B}\rvert$	$\lvert d_{2005,BB}\rvert$	$\lvert d_{2005,BBB}\rvert$	$\lvert d_{2005,A}\rvert$
Model0	3.66	3.96	2.33	1.27	0.63
Model1	3.00	3.15	2.00	1.05	0.58
Model2	2.75	2.57	1.76	0.89	0.55
Model3	2.79	2.69	1.85	0.94	0.56
Index shifted	$\lvert d_{2005,CCC}\rvert$	$\lvert d_{2005,B}\rvert$	$\lvert d_{2005,BB}\rvert$	$\lvert d_{2005,BBB}\rvert$	$\lvert d_{2005,A}\rvert$
Model0	4.00	4.36	2.51	1.38	0.67
Model1	3.08	3.25	2.06	1.08	0.59
Model2	2.71	2.47	1.74	0.85	0.55
Model3	2.83	2.72	1.87	0.94	0.57
Index not shifted	$CPO_{2005,CCC}$	$CPO_{2005,B}$	$CPO_{2005,BB}$	$CPO_{2005,BBB}$	$CPO_{2005,A}$
Model0	0.011	0.009	0.032	0.148	0.667
Model1	0.018	0.015	0.049	0.186	0.710
Model2	0.023	0.023	0.071	0.216	0.729
Model3	0.022	0.022	0.063	0.207	0.725
Index shifted	$CPO_{2005,CCC}$	$CPO_{2005,B}$	$CPO_{2005,BB}$	$CPO_{2005,BBB}$	$CPO_{2005,A}$
Model0	0.012	0.010	0.034	0.137	0.640
Model1	0.022	0.018	0.058	0.185	0.698
Model2	0.027	0.025	0.081	0.220	0.730
Model3	0.025	0.023	0.070	0.207	0.719

Table VI. Observed transition probabilities from 2004 to 2005.

	To A	To BBB	To BB	To B	To CCC	To default
From A	0.95	0.05	0.00	0.00	0.00	0.00
From BBB	0.07	0.90	0.03	0.00	0.00	0.00
From BB	0.00	0.06	0.86	0.08	0.00	0.00
From B	0.00	0.01	0.10	0.83	0.04	0.02
From CCC	0.00	0.01	0.01	0.31	0.57	0.11

predicted posterior transition probabilities given in Table VII. From these we see that transition probabilities are predicted reasonably accurately for the top four rating categories, whereas for CCC and the default probability estimates are less accurate. The sum of (weighted) squared differences gives 0.282 (0.133), 0.289 (0.136), 0.274 (0.129) and 0.276 (0.130) for Model0, Model1, Model2 and Model3, respectively. This again shows a slight preference to Model2.

Judging from the results on the predictive distributions, one would clearly prefer Model2 to Model3. This preference is not a surprise, since one would expect the risk structure of companies in adjacent rating categories to be more similar than that of companies with ratings further apart.

DOI: 10.1002/asmb

Table VII. Predicted transition probabilities from 2004 to 2005.

	To A	To BBB	To BB	To B	To CCC	To default
Model0						
From A	0.96	0.04	0.00	0.00	0.00	0.00
From BBB	0.04	0.89	0.07	0.00	0.00	0.00
From BB	0.00	0.07	0.88	0.05	0.00	0.00
From B	0.00	0.00	0.04	0.93	0.03	0.00
From CCC	0.00	0.00	0.00	0.02	0.98	0.02
Model1						
From A	0.96	0.04	0.00	0.00	0.00	0.00
From BBB	0.03	0.89	0.07	0.00	0.00	0.00
From BB	0.00	0.06	0.89	0.05	0.00	0.00
From B	0.00	0.00	0.04	0.93	0.03	0.00
From CCC	0.00	0.00	0.00	0.02	0.98	0.01
Model2						
From A	0.95	0.05	0.00	0.00	0.00	0.00
From BBB	0.05	0.87	0.08	0.00	0.00	0.00
From BB	0.00	0.08	0.86	0.06	0.00	0.00
From B	0.00	0.00	0.06	0.92	0.03	0.00
From CCC	0.00	0.00	0.00	0.02	0.98	0.01
Model3						
From A	0.96	0.04	0.00	0.00	0.00	0.00
From BBB	0.04	0.88	0.08	0.00	0.00	0.00
From BB	0.00	0.08	0.86	0.06	0.00	0.00
From B	0.00	0.00	0.05	0.92	0.03	0.00
From CCC	0.00	0.00	0.00	0.02	0.98	0.01

5. SUMMARY AND DISCUSSION

We have extended the Bernoulli mixture models considered in [7] by explicitly modeling the correlation structure between rating category and time period and studied their model fit and predictive capability. In contrast to [7], we place special emphasis on model prediction, which we believe is the primary focus of the data analyst. In particular, for the investigation of the predictive ability of a model we used predictable model specifications. Further, we utilized the Brier score and a standardized predictive residual. The results of our empirical study showed that the model extensions are useful for both model fit and prediction. In particular, the data provide evidence that the correlation effect between rating categories is decreasing when rating categories are further apart despite a rather limited database and it is larger than the correlation effect induced by the time dynamics. Model2 can also be extended to allow for rating-dependent α components or even more general vector autoregressive models for the risk vector. This will be the topic of future research.

For a larger data set with longer time history, one can extend the model to include dynamic model components for exogenous variables reflecting macro-economic information. In this context models as considered and fitted in [17] can be utilized. In addition, one could also consider different time dynamics such as general ARMA or stochastic volatility models. Further additional fixed

DOI: 10.1002/asmb

grouping variables such as industry sectors to allow for more homogeneous groups can be easily included.

With regard to the Bayesian approach, external data sources can be easily incorporated transforming this information into proper prior information. For example, if one expects the time correlation parameter ρ to be close to a value ρ_0, one can use, for example, a normal prior for ρ with mean value ρ_0 truncated to the interval $(-1, 1)$.

Because of the requirements for an IRB-based approach, it is to be expected that larger internal databases over longer time horizons will become available, where the model extensions discussed above will be feasible and expected to improve the default probability predictions. Finally, we like to note that models for default probabilities are only one component for credit risk management in addition to models for loss after default. Therefore, more realistic joint models that can be fitted and assessed by a Bayesian approach are to be envisioned.

APPENDIX A: CONDITIONAL DISTRIBUTIONS

In this section, we will denote unconditional densities by $[\cdot]$ and conditional densities by $[\cdot|\cdot]$. Further, we collect the observed data M_{tk} and m_{tk} for $t=1,\ldots,T$; $k=1,\ldots,K$ into the vector's $\mathbf{M}$ and $\mathbf{m}$, respectively. The complete risk vector is denoted by $\mathbf{b}=(\mathbf{b}_1,\ldots,\mathbf{b}_T)$. Finally, we also make repeated use of the following fact used in [7]:

If $\mathbf{Z}=(Z_1, Z_2, \ldots, Z_m)\sim N_m(\boldsymbol{\mu}, \Psi)$, where X denotes the inverse of Ψ, then

$$Z_r|\mathbf{Z}_{-r} \sim N(\tilde{\mu}, \tilde{\sigma^2}) \quad \text{with } \tilde{\mu}=\mu_r+\frac{1}{X_{rr}}\sum_{s=1,s\neq r}^{K} X_{sr}(\mu_s - Z_s) \quad \text{and} \quad \tilde{\sigma^2}=\frac{1}{X_{rr}} \tag{A1}$$

A.1. Model1

The joint density is given by

$$[\mathbf{M},\mathbf{m},\mathbf{b},\boldsymbol{\mu},\sigma,\boldsymbol{\beta},\alpha]=\prod_{t=1}^{T}\prod_{k=1}^{K}[M_{tk}|m_{tk},b_t,\mu_k,\boldsymbol{\beta}][\mathbf{b}|\sigma,\alpha][\boldsymbol{\mu}][\sigma][\boldsymbol{\beta}][\alpha]$$

where $[M_{tk}|m_{tk},b_t,\mu_k,\boldsymbol{\beta}]\propto g(\mu_k-\mathbf{x}_t\boldsymbol{\beta}-b_t)^{M_{tk}}(1-g(\mu_k-\mathbf{x}_t\boldsymbol{\beta}-b_t))^{m_{tk}-M_{tk}}$. Now, the complete risk vector $\mathbf{b}=(b_1,b_2,\ldots,b_T)^{\mathrm{T}}$ is multivariate normal with covariance matrix Σ given by $\Sigma_{st}=\operatorname{cov}(b_s,b_t)=\sigma^2\alpha^{|s-t|}/(1-\alpha^2)$, $s,t\varepsilon\{1,2,\ldots,T\}$. Its inverse is tridiagonal

$$\Sigma^{-1}=\frac{1}{\sigma^2}\begin{pmatrix} 1 & -\alpha & & & & \\ -\alpha & 1+\alpha^2 & -\alpha & & & \\ & -\alpha & 1+\alpha^2 & -\alpha & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\alpha & 1+\alpha^2 & -\alpha \\ & & & & -\alpha & 1 \end{pmatrix}$$

DOI: 10.1002/asmb

Moreover, $\det(\Sigma^{-1}) = \sigma^{-2\mathrm{T}}(1-\alpha^2)$. It then follows that the full conditional of α is

$$[\alpha|\sigma, \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{b}, \mathbf{m}, \mathbf{M}] \propto \sqrt{\det(\Sigma^{-1})}\exp\{-\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\Sigma^{-1}\mathbf{b}\}[\alpha]$$
$$\propto \sqrt{1-\alpha^2}\exp\{-\tfrac{1}{2}\sigma^{-2}(C_1(\mathbf{b})\alpha^2 - C_2(\mathbf{b})\alpha)\}[\alpha] \tag{A2}$$

where $C_1(\mathbf{b}) = \sum_{t=2}^{T-1} b_t^2$ and $C_2(\mathbf{b}) = 2\sum_{t=2}^{T} b_t b_{t-1}$. The posterior density of σ is given by

$$[\sigma^2|\alpha, \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{b}, \mathbf{m}, \mathbf{M}] \propto [\mathbf{b}|\sigma, \alpha][\sigma] \propto \sigma^{-\mathrm{T}}\exp\{-C_3(\mathbf{b})\sigma^{-2}\}[\sigma]$$

where $C_3(\mathbf{b}) = \frac{1}{2}(\sum_{t=1}^{T} b_t^2 + \alpha^2\sum_{t=2}^{T-1} b_1^2 - 2\alpha\sum_{t=2}^{T} b_t b_{t-1})$. But now, if σ^2 has an $\mathrm{Inv}\,\Gamma(\eta, v)$ prior then

$$[\sigma^2|\alpha, \boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{b}, \mathbf{m}, \mathbf{M}] \sim \mathrm{Inv}\,\Gamma(\eta + T/2, v + C_3(\mathbf{b}, \alpha))$$

The risk vector $\mathbf{b}|\alpha, \sigma$ is multivariate normal and for $\mathbf{b}_{-t} = (b_1, b_2, \ldots, b_{t-1}, b_{t+1}, \ldots, b_T)$, $t = 1, 2, \ldots, T$

$$[b_t|\underline{b}_{-t}, \alpha, \sigma] \sim \begin{cases} N(\alpha b_2, \sigma^2), & t = 1 \\ N(\alpha b_{T-1}, \sigma^2), & t = T \\ N\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1} + b_{t+1}), \dfrac{\sigma^2}{1+\alpha^2}\right) & \text{otherwise} \end{cases}$$

The full conditional density then is

$$[b_t|\mathbf{b}_{-t}, \alpha, \sigma, \mathbf{m}, \mathbf{M}] \propto \prod_{k=1}^{K} [M_{tk}|m_{tk}, b_t, \mu_k][b_t|\mathbf{b}_{-t}, \alpha, \sigma]$$

where $[M_{tk}|m_{tk}, b_t, \mu_k, \boldsymbol{\beta}] \propto g(\mu_k - \mathbf{x}_t\boldsymbol{\beta} - b_t)^{M_{tk}}(1 - g(\mu_k - \mathbf{x}_t\boldsymbol{\beta} - b_t))^{m_{tk} - M_{tk}}$. The full conditional density of $\boldsymbol{\beta}$ is given by

$$[\boldsymbol{\beta}|\alpha, \sigma, \boldsymbol{\mu}, \mathbf{b}, \mathbf{m}, \mathbf{M}] \propto \prod_{t=1}^{T}\prod_{k=1}^{K} [M_{tk}|m_{tk}, b_t, \mu_k, \boldsymbol{\beta}][\boldsymbol{\beta}]$$

The corresponding full conditionals for Model0 can easily be obtained by setting $\alpha = 0$.

A.2. Model2

Model2 has the joint distribution

$$[\mathbf{M}, \mathbf{m}, \mathbf{b}, \boldsymbol{\mu}, \sigma, \alpha, \boldsymbol{\beta}, \rho] = \prod_{t=1}^{T}\prod_{k=1}^{K} [M_{tk}|m_{tk}, b_{tk}, \mu_k, \beta][\mathbf{b}|\sigma, \alpha\rho][\boldsymbol{\mu}][\sigma][\alpha][\boldsymbol{\beta}][\rho] \tag{A3}$$

DOI: 10.1002/asmb

The risk vector $\mathbf{b}$ is again multivariate normal with zero mean vector and $\operatorname{cov}(b_{sk}, b_{tl}) = \Phi_{kl}\alpha^{|t-s|}/(1-\alpha^2)$, $s, t\varepsilon\{1,2,\ldots,T\}$, $k, l\varepsilon\{1,2,\ldots,K\}$ so that the covariance matrix has inverse

$$\Sigma^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} \Lambda & -\alpha\Lambda & & & & \\ -\alpha\Lambda & \Lambda(1+\alpha^2) & -\alpha\Lambda & & & \\ & -\alpha\Lambda & \Lambda(1+\alpha^2) & -\alpha\Lambda & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\alpha\Lambda & \Lambda(1+\alpha^2) & -\alpha\Lambda \\ & & & & -\alpha\Lambda & \Lambda \end{pmatrix} \tag{A4}$$

where

$$\Lambda = \begin{pmatrix} 1 & -\rho & & & & \\ -\rho & 1+\rho^2 & -\rho & & & \\ & -\rho & 1+\rho^2 & -\rho & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\rho & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix}$$

is the inverse of Φ/σ^2. This gives that $\det(\Sigma^{-1}) = \det(\Lambda)^{\mathrm{T}}(1-\alpha^2)\sigma^{-2TK} = \sigma^{-2TK}(1-\rho^2)^{\mathrm{T}}(1-\alpha^2)$. Now, $\mathbf{b}$ is again multivariate normal and it follows from (A1) that

$$[\mathbf{b}_t|\mathbf{b}_{-t}, \alpha, \boldsymbol{\beta}, \sigma, \rho] \sim \begin{cases} N(\alpha\mathbf{b}_2, \Phi), & t=1 \\ N(\alpha\mathbf{b}_{T-1}, \Phi), & t=T \\ N\left(\dfrac{\alpha}{1+\alpha^2}(\mathbf{b}_{t-1}+\mathbf{b}_{t+1}), \dfrac{1}{1+\alpha^2}\Phi\right) & \text{otherwise} \end{cases} \tag{A5}$$

Again using (A1) gives that if $t=1$

$$[b_{tk}|\mathbf{b}_{-tk}, \alpha, \boldsymbol{\beta}, \rho, \sigma]$$

$$\sim \begin{cases} N(\alpha b_{2,k} - \rho(\alpha b_{2,k+1} - b_{1,k+1}), \sigma^2), & k=1 \\ N(\alpha b_{2,k} - \rho(\alpha b_{2,k-1} - b_{1,k-1}), \sigma^2), & k=K \\ N\left(\alpha b_{2,k} - \dfrac{\rho}{1+\rho^2}(\alpha b_{2,k-1} - b_{1,k-1} + \alpha b_{2,k+1} - b_{1,k+1}), \dfrac{\sigma^2}{1+\rho^2}\right) & \text{otherwise} \end{cases}$$

DOI: 10.1002/asmb

If $t=T$

$$[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]$$

$$\sim\begin{cases} N(\alpha b_{T-1,k}-\rho(\alpha b_{T-1,k+1}-b_{T,k+1}),\sigma^2), & k=1\\ N(\alpha b_{T-1,k}-\rho(\alpha b_{T-1,k-1}-b_{T,k-1}),\sigma^2), & k=K\\ N\left(\alpha b_{T-1,k}-\dfrac{\rho}{1+\rho^2}(\alpha b_{T-1,k-1}-b_{T,k-1}+\alpha b_{T-1,k+1}-b_{T,k+1}),\dfrac{\sigma^2}{1+\rho^2}\right) & \text{otherwise}\end{cases}$$

For $1<t<T$

$$[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]\sim\begin{cases} N\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k}+b_{t+1,k})-\rho\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k+1}+b_{t+1,k+1})-b_{t,k+1}\right),\dfrac{\sigma^2}{1+\alpha^2}\right), & k=1\\ N\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k}+b_{t+1,k})-\rho\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k-1}+b_{t+1,k-1})-b_{t,k-1}\right),\dfrac{\sigma^2}{1+\alpha^2}\right), & k=K\\ N\left(\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k}+b_{t+1,k})-\dfrac{\rho}{1+\rho^2}\left[\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k-1}+b_{t+1,k-1})-b_{t,k-1}\right.\right. & \\ \quad\left.\left.+\dfrac{\alpha}{1+\alpha^2}(b_{t-1,k+1}+b_{t+1,k+1})-b_{t,k+1}\right],\dfrac{\sigma^2}{(1+\rho^2)(1+\alpha^2)}\right) & \text{otherwise}\end{cases}$$

Then $[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma,\mathbf{m},\mathbf{M}]\propto[b_{tk}|\mathbf{b}_{-tk},\alpha,\rho,\sigma,][N_{tk}|n_{tk},b_{tk},\mu_k,\boldsymbol{\beta}]$. The full conditional distribution of ρ is given by

$$\begin{aligned}[\rho|\alpha,\sigma,\boldsymbol{\beta},\mathbf{m},\mathbf{M},\mathbf{b}] &\propto [\mathbf{b}|\sigma,\rho,\alpha][\rho]\\ &\propto\sqrt{\det(\Sigma^{-1})}\exp\{-\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\Sigma^{-1}\mathbf{b}\}[\sigma]\\ &\propto\sqrt{(1-\rho^2)^{\mathrm{T}}}\exp\{-\tfrac{1}{2}\sigma^{-2}(S_1(\mathbf{b},\alpha)\rho+S_2(\mathbf{b},\alpha)\rho^2)\}\end{aligned}$$

Now, if $c_i(\mathbf{u},\mathbf{v})$ denotes the coefficient of ρ^i in $\mathbf{u}^{\mathrm{T}}\Lambda\mathbf{v}$, then here $c_1(\mathbf{u},\mathbf{v})=-(\sum_{k=2}^{K}u_k v_{k-1}+u_{k-1}v_k)$ and $c_2(\mathbf{u},\mathbf{v})=\sum_{k=2}^{K-1}u_k v_k$. Then for $i=1,2$

$$S_i(\mathbf{b},\alpha)=\sum_{t=2}^{T-1}(c_i(\mathbf{b}_t,\mathbf{b}_t)(1+\alpha^2))+c_i(\mathbf{b}_1,\mathbf{b}_1)+c_i(\mathbf{b}_T,\mathbf{b}_T)-2\alpha\sum_{t=2}^{T}c_i(\mathbf{b}_t,\mathbf{b}_t)\tag{A6}$$

The full conditional distribution of α can be determined as in (A2)

$$\begin{aligned}[\alpha|\sigma,\boldsymbol{\beta},\rho,\boldsymbol{\mu},\mathbf{b},\mathbf{m},\mathbf{M}] &\propto\sqrt{\det(\Sigma^{-1})}\exp\{-\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\Sigma^{-1}\mathbf{b}\}[\alpha]\\ &\propto\sqrt{1-\alpha^2}\exp\{-\tfrac{1}{2}\sigma^{-2}(C_1(\mathbf{b},\Lambda)\alpha^2-C_2(\mathbf{b},\Lambda)\alpha)\}[\alpha]\end{aligned}$$

DOI: 10.1002/asmb

where $C_1(\mathbf{b},\Lambda)=\sum_{t=2}^{T-1}\mathbf{b}_t{}^{\mathrm{T}}\Lambda\mathbf{b_t}$ and $C_2(\underline{\mathbf{b}})=\sum_{t=2}^{T}\mathbf{b_t^{\mathrm{T}}}\Lambda\mathbf{b_{t-1}}+\mathbf{b_{t-1}^{\mathrm{T}}}\Lambda\mathbf{b_t}$. The posterior density of σ^2 is given by

$$[\sigma^2|\alpha,\boldsymbol{\beta},\rho,\boldsymbol{\mu},\mathbf{b},\mathbf{m},\mathbf{M}]\propto[\mathbf{b}|\sigma,\rho,\alpha][\sigma]\propto\sqrt{\det(\Sigma)^{-1}}\exp\{-\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\Sigma^{-1}\mathbf{b}\}[\sigma]$$
$$\propto\sigma^{-TK}\exp\{-C_3(\mathbf{b},\Lambda,\alpha)\sigma^{-2}\}[\sigma]$$

where

$$C_3(\mathbf{b},\Lambda,\alpha)=\frac{1}{2}\left(\sum_{t=1}^{T}\mathbf{b_t^{\mathrm{T}}}\Lambda\mathbf{b_t}+\alpha^2\sum_{t=2}^{T-1}\mathbf{b_t^{\mathrm{T}}}\Lambda\mathbf{b_t}-\alpha\sum_{t=2}^{T}(\mathbf{b_t^{\mathrm{T}}}\Lambda\mathbf{b_{t-1}}+\mathbf{b_{t-1}^{\mathrm{T}}}\Lambda\mathbf{b_t})\right)$$

Then, if σ^2 has an $\mathrm{Inv}\,\Gamma(\eta,v)$ prior

$$[\sigma^2|\alpha,\boldsymbol{\beta},\rho,\boldsymbol{\mu},\mathbf{b},\mathbf{m},\mathbf{M}]\sim\mathrm{Inv}\,\Gamma(\eta+TK/2,v+C_3(\mathbf{b},\Lambda,\alpha))$$

The posterior densities for $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ can be found similar to those in Model1.

A.3. Model3

Model3 again has joint density (A3). Here, the covariance matrix has inverse as in (A4), where Λ is the inverse of $\Phi(1-\rho^2)/\sigma^2$ with Φ defined in (6), i.e.

$$\Lambda=\frac{1+3\rho}{1+3\rho-4\rho^2}\begin{pmatrix}1&\lambda&\lambda&\lambda&\lambda\\\lambda&1&\lambda&\lambda&\lambda\\\lambda&\lambda&1&\lambda&\lambda\\\lambda&\lambda&\lambda&1&\lambda\\\lambda&\lambda&\lambda&\lambda&1\end{pmatrix}\quad\text{with }\lambda=\frac{-\rho}{1+3\rho}\tag{A7}$$

This then gives

$$\det(\Sigma^{-1})=\det(\Lambda)^{\mathrm{T}}(1-\alpha^2)\sigma^{-2TK}=\left((1-\lambda)^4(1+4\lambda)\left(\frac{1+3\rho}{1+3\rho-4\rho^2}\right)^K\right)^{\mathrm{T}}\sigma^{-2TK}(1-\alpha^2)$$

(A5) and (A1) hold exactly as for Model2 and with Λ as above $[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]$ can be determined. For $t=1$

$$[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]\sim N\left(\alpha b_{2,k}+\lambda\sum_{s\neq k}(\alpha b_{2,s}-b_{1,s}),\frac{1+3\rho-4\rho^2}{1+3\rho}\sigma^2\right)$$

For $t=T$

$$[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]\sim N\left(\alpha b_{T-1,k}+\lambda\sum_{s\neq k}(\alpha b_{T-1,s}-b_{T,s}),\frac{1+3\rho-4\rho^2}{1+3\rho}\sigma^2\right)$$

DOI: 10.1002/asmb

For $1<t<T$

$$[b_{tk}|\mathbf{b}_{-tk},\alpha,\boldsymbol{\beta},\rho,\sigma]\sim N\left(\frac{\alpha}{1+\alpha^2}(b_{t-1,k}+b_{t+1,k})\right.$$
$$\left.+\lambda\sum_{s\neq k}\left(\frac{\alpha}{1+\alpha^2}(b_{t-1,s}+b_{t+1,s})-b_{t,s}\right),\sigma^2\frac{1}{1+\alpha^2}\frac{1+3\rho-4\rho^2}{1+3\rho}\right)$$

The full conditional density of ρ is determined by

$$[\rho|\alpha,\sigma,\boldsymbol{\beta},\mathbf{m},\mathbf{M},\mathbf{b}]$$
$$\propto[\mathbf{b}|\sigma,\rho,\alpha][\rho]\propto\sqrt{\det(\Sigma^{-1})}\exp\{-\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\Sigma^{-1}\mathbf{b}\}[\sigma]$$
$$\propto\left(\frac{1}{1+4\rho}\frac{1}{(1-\rho)^4}\right)^{T/2}\exp\left\{-\frac{1}{2\rho^2}\left(\frac{1+3\rho}{1+3\rho-4\rho^2}S_1(\mathbf{b},\alpha)-S_2(\mathbf{b},\alpha)\frac{\rho}{1+3\rho-4\rho^2}\right)\right\}$$

Now, if $c_i(\mathbf{u},\mathbf{v})$ this the coefficient of λ^i in $\mathbf{u}^{\mathrm{T}}(1+3\rho-4\rho^2)/(1+3\rho)\Lambda\mathbf{v}$ for Λ defined in (A7), then here $c_1(\mathbf{u},\mathbf{v})=\sum_k u_k v_k$ and $c_2(\mathbf{u},\mathbf{v})=\sum_{k\neq l}u_k v_l$. Then for $i=1,2$, S_i is again defined by (A6). Since the full conditional distributions of α, $\boldsymbol{\beta}$, σ and $\boldsymbol{\mu}$ do not depend on the actual form of Λ, these full conditional distributions are the same as those in Model2.

ACKNOWLEDGEMENTS

The first author was supported by the Deutsche Forschungsgemeinschaft grant CZ 86/1-1. We are also grateful to the editor and three anonymous referees for detailed suggestions that greatly improved this paper.

REFERENCES

1. McNeil A, Frey R, Embrechts P. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press: Princeton, NJ, 2005.
2. Nickell P, Perraudin W, Varotto S. Stability of rating transitions. *Journal of Banking and Finance* 2000; **24**: 203–227.
3. Demey P, Jouanin J, Roget C, Roncalli T. Maximum likelihood estimate of default correlations. *Risk* 2004; **November**:104–108.
4. Servigny A, Renault O. Correlation evidence. *Risk* 2003; **July**:90–94.
5. Gordy M, Heitfield E. *Estimating Default Correlations From Short Panels of Credit Performance Data*. Federal Reserve Board, 2002.
6. Bangia A, Diebold F, Kronimus A, Schlagen C, Schuermann T. Rating migration and the business cycle with application to credit portfolio stress testing. *Journal of Banking and Finance* 2002; **26**:445–474.
7. McNeil A, Wendin J. Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* 2007; **14**:131–149.
8. Chib S. Markov chain Monte Carlo methods: computation and inference. In *Handbook of Econometrics*, vol. 5, Heckman J, Leamer E (eds). North-Holland: Amsterdam, 2001; 3569–3649.
9. Gamerman D. *Markov Chain Monte Carlo Stochastic Simulation for Bayesian Inference* (1st edn). Chapman & Hall: London, 1997.
10. Gilks W, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice* (1st edn). Chapman & Hall: London, 1996.

11. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, *Series B* 2002; **64**:583–639.
12. Brier G. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 1950; **7**:1–3.
13. Gelfand A. Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, Gilks W, Richardson S, Spiegelhalter D (eds). Chapman & Hall: London, 1996; 145–161.
14. Merton R. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 1974; **29**:449–470.
15. Gilks W. Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian Statistics 4*, Bernardo J, Berger J, Dawid A, Smith A (eds). Chapman & Hall: London, 1992; 145–161.
16. Gilks W. Full conditional distributions. In *Markov Chain Monte Carlo in Practice*, Gilks W, Richardson S, Spiegelhalter D (eds). Chapman & Hall: London, 1996; 75–88.
17. Czado C, Song PX.-K. State space mixed models for longitudinal observations with binary and binomial responses. *Statistical Papers* 2007; in press. DOI: 10.1007/s00362-006-0039-y

DOI: 10.1002/asmb