Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 14 (2007) 131–149
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Bayesian inference for generalized linear mixed models of portfolio credit risk

Alexander J. McNeil, Jonathan P. Wendin *,[1]

Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland

Accepted 12 May 2006
Available online 1 September 2006

Abstract

The aims of this paper are threefold. First, we highlight the usefulness of generalized linear mixed models (GLMMs) in the modelling of portfolio credit default risk. The GLMM-setting allows for a flexible specification of the systematic portfolio risk in terms of observed *fixed effects* and unobserved *random effects*, in order to explain the phenomena of default dependence and time-inhomogeneity in historical default data. Second, we show that computational Bayesian techniques such as the *Gibbs sampler* can be successfully applied to fit models with serially correlated random effects, which are special instances of *state space models*. Third, we provide an empirical study using Standard and Poor's data on U.S. firms. A model incorporating rating category and sector effects, and a macroeconomic proxy variable for state-of-the-economy suggests the presence of a residual, cyclical, latent component in the systematic risk.

JEL classification: G31; G11; C15
Keywords: Credit risk; Generalized linear mixed model; State space model; Bayesian inference

1. Introduction

It is well accepted that credit default events show dependence. A first observation supporting this view is that default intensities seem to vary over time according to economic cycles. This can be seen in Fig. 1, which is based on six-monthly default data from Standard and Poor's (S&P) for

* Corresponding author. Tel.: +41 44 6322293; fax: +41 44 6321085.
E-mail addresses: mcneil@math.ethz.ch (A.J. McNeil), wendin@math.ethz.ch (J.P. Wendin).
[1] Financial support from NCCR Financial Valuation and Risk Management (a research program supported by the Swiss National Science Foundation) is gratefully acknowledged.

0927-5398/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2006.05.002

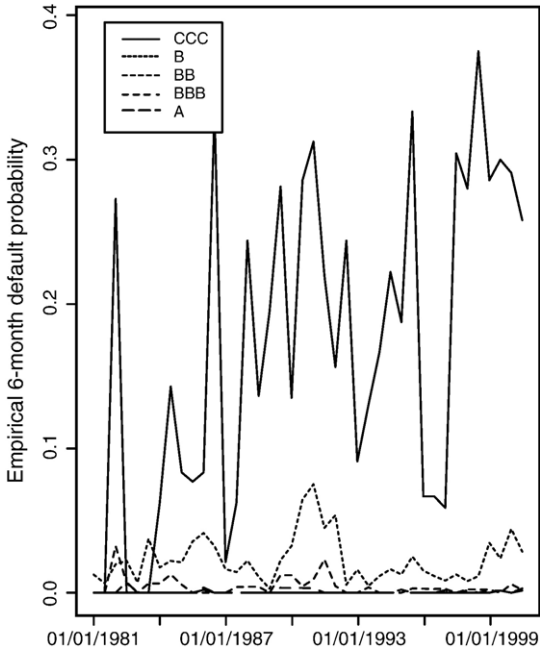

Fig. 1. Empirical bi-annual default rates according to Standard and Poor's for 20 years (U.S. obligors only). Note the cyclic behaviour of the B- and BB-default rates. (The CCC-class is very noisy due to its relatively small size.)

the years 1981–2000. Periods with many defaults are in general preceded and followed by other periods with many defaults, and this is particularly evident for B- and BB-rated companies. A second source of dependence is direct business liaisons between obligors, so-called contagion, meaning that a company may itself face increased risk if one of its major customers defaults; see for instance Egloff et al. (2004). Capturing the dependence is of immediate interest to financial institutions lending money or holding credit-risky investments, since a disproportionally large number of defaults over a fixed time horizon may have severe consequences.

On the one hand, we require statistical models of default that address the issue of *cross-sectional dependence* in default rates within a time period due to common economic conditions (the so-called *systematic risk*). On the other hand, we aim to capture *serial dependence* caused by the cyclical behaviour of economic factors. The starting point for most portfolio credit risk models is that, conditionally on the systematic risk, defaults occur independently. A key concern of ours will therefore be the systematic risk.

Several empirical studies, such as Nickell et al. (2000) and Bangia et al. (2002), have verified time variation in default rates, and confirmed that this time variation may to some extent be explained by observed macroeconomic variables. Unfortunately, observed variables as proxies for the systematic risk are seldom completely satisfactory. The first important issue is the identification of appropriate proxies. Moreover, there may also be a lag between the cycle of a proxy variable and that of the default activity, and this lag may vary stochastically over time. Mastering the lags is of critical importance for regulatory purposes, so that banks do not, for instance, lower their capital levels in an apparent upswing of the economy.

The above shortcomings can be remedied by allowing for unobserved (latent) systematic risk components that capture the residual systematic risk once any observed parts have been

accounted for. A more detailed account of the advantages of latent risk factors over observed ones is given in Koopman, Lucas and Klaassen (2005).

The requirement of being able to model both observed and unobserved elements of the systematic risk can be accommodated by the class of generalized linear mixed models (GLMMs). In what follows, we will often comply with standard GLMM-terminology and speak of *fixed effects* and *random effects* for observable and unobservable factors, respectively. As we shall see, wellchosen fixed effects and random effects offer great model versatility, and allow us to capture time-inhomogeneity in default rates and heterogeneity across individual obligors, industry sectors, or any other desired groupings.

From a practical point of view, the use of latent components in the systematic risk yields joint default distributions in the form of integrals, which in general lack closed forms (cf. Section 2.3). Analytical maximum likelihood techniques can be used for relatively simple models that do not incorporate serial dependence; examples include Gordy and Heitfield (2002), Frey and McNeil (2003), and Rösch (2005). However, for the models we consider, some form of Monte Carlo approach to inference seems unavoidable. While simulated maximum likelihood inference is one possible approach, we choose to demonstrate the attractions of computational Bayesian methodology (Gibbs sampling) in this paper.

The Gibbs sampler is capable of handling a variety of specifications of the systematic risk, including serially correlated, multivariate latent factors. It also allows us to calculate standard errors for both primary model parameters and derived model parameters (e.g., default correlations) in a straightforward manner. Moreover, the Bayesian approach opens up the possibility of using other sources of information about default, such as asset correlations, to set priors. Further discussion of the Bayesian approach, particularly with credit models in mind, is given in the Conclusions.

The models with serially dependent latent factors that we consider in this paper can be seen as *state space models* (Brockwell and Davis, 1991). There are some literatures on fitting such models to default data. Crowder et al. (2005) consider a model for default counts with a two-state latent systematic factor following a Markov chain. Gagliardini and Gouriéroux (2005), and Koopman, Lucas and Klaassen (2005) consider models where default risk is driven by continuous latent factors, although they model the ratio of defaulted obligors instead of the actual default counts. This simplification may show undesirable features, in particular when either the numerator or the denominator are small (Kurbat and Korablev, 2002).

The paper is organized as follows. In Section 2, we review the well known Bernoulli mixture model with fixed and random effects, and discuss its relationship to the class of GLMMs, highlighting the flexibility of this model class. Section 2.3 discusses both frequentist and Bayesian inference in such models. The empirical analysis is found in Section 3; the dataset is described and four models of increasing complexity are fitted. Section 4 presents our conclusions. Details of the implementation of the Gibbs sampler are given in Appendix A.

2. Generalized linear mixed models for credit portfolios

Consider a set $\mathcal{K} = \{1, \ldots, K\}$ of rating classes of increasing creditworthiness (the state default corresponds to state 0). Denoted by n_{tk} the number of firms in group $k \in \mathcal{K}$, so that $n_t = \sum_{k=1}^{K} n_{tk}$ is the total number of obligors in the portfolio at the beginning of period t.

For each obligor $i = 1, \ldots, n_t$ in period t, let Y_{ti} denote the default indicator variable, which assumes the value 1 if the obligor defaults during time period t and 0 otherwise, and let $\kappa(t, i)$ denote the rating. Note that the models will be of *repeated cross-sectional* type, so that Y_{ti} and Y_{si}, the ith indicator variables in two distinct time periods $s \neq t$, do not necessarily refer to the same obligors. Moreover, the vectors $\boldsymbol{Y}_t = (Y_{t1}, \ldots, Y_{tn_t})$ and $\boldsymbol{Y}_s = (Y_{s1}, \ldots, Y_{sn_s})$ may be of different lengths.

2.1. Bernoulli mixture models and GLMMs

2.1.1. Bernoulli mixture models

Given a (latent) random effect $\boldsymbol{b}_t$ (following a distribution F_b to be specified), we assume that the default indicators $Y_{t1},\ldots,Y_{tn_t}$ are conditionally independent with distribution given by

$$P(Y_{ti}=1|\boldsymbol{b}_t)=g(\mu_{\kappa(t,i)}+\boldsymbol{x}_{ti}'\boldsymbol{\beta}+\boldsymbol{z}_{ti}'\boldsymbol{b}_t). \tag{1}$$

In the above expression, $\boldsymbol{x}_{ti}$ and $\boldsymbol{z}_{ti}$ denote the known *design vectors* holding the corresponding covariates of obligor i in period t, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}:=(\mu_1,\ldots,\mu_K)$ are vectors of unknown regression coefficients, and $g(\cdot): R\rightarrow(0,1)$ is a smooth, strictly increasing mapping called the *response function*. Common choices are the *probit* response $g(x)=\Phi(x)$ (cumulative standard normal df) and the *logit* response $g(x)=1/(1+\exp\{-x\})$, but other choices are possible; see Chapter 7 of Joe (1997).

By assumption, the joint conditional distribution of $\boldsymbol{Y}_t=(Y_{t1},\ldots,Y_{tn_t})$ is

$$P(\boldsymbol{Y}_t=\boldsymbol{y}|\boldsymbol{b}_t)=\prod_{i=1}^{n_t}P(Y_{ti}=1|\boldsymbol{b}_t)^{y_i}(1-P(Y_{ti}=1|\boldsymbol{b}_t))^{1-y_i} \tag{2}$$

for all $\boldsymbol{y}\in\{0,1\}^{n_t}$. The unconditional distribution of $\boldsymbol{Y}_t$ is obtained by integrating out the effect of $\boldsymbol{b}_t$, and results in dependent Bernoulli indicators:

$$P(\boldsymbol{Y}_t=\boldsymbol{y})=\int P(\boldsymbol{Y}_t=\boldsymbol{y}|\boldsymbol{b}_t)\mathrm{d}F_b(\boldsymbol{b}_t).$$

The Bernoulli mixture model can be interpreted as a credit model of *threshold type* as follows. Let $\varepsilon_{t1},\ldots,\varepsilon_{tn_t}$ be iid rvs with distribution function g, which are also independent of $\boldsymbol{b}_t$. Set $V_{ti}:=-\boldsymbol{x}_{ti}'\boldsymbol{\beta}-\boldsymbol{z}_{ti}'\boldsymbol{b}_t+\varepsilon_{ti}$ for $i=1,\ldots,n_t$, and observe that the mixture model (1) corresponds to a model in which obligor i in period t defaults if and only if $V_{ti}\le\mu_{\kappa(t,i)}$. The random variable V_{ti} can be interpreted as the asset value of obligor i in period t and $\mu_{\kappa(t,i)}$ as the critical liability, as in the seminal work on structural models to credit risk in Merton (1974). This representation is particularly useful for the probit case, since each ε_{ti} is then standard Gaussian. The well known industry model CreditMetrics is of this kind with Gaussian random effects.

We refer to $\mathrm{corr}(V_{ti},V_{tj})$ as the *implied asset correlation* of obligors i and j. It readily follows that

$$\mathrm{corr}(V_{ti},V_{tj})=\mathrm{cov}(\boldsymbol{z}_{ti}'\boldsymbol{b}_t,\boldsymbol{z}_{tj}'\boldsymbol{b}_t)/\left(\sqrt{\mathrm{var}(\boldsymbol{z}_{ti}'\boldsymbol{b}_t)+w^2}\sqrt{\mathrm{var}(\boldsymbol{z}_{tj}'\boldsymbol{b}_t)+w^2}\right), \tag{3}$$

where $w^2:=\mathrm{var}(\varepsilon_{ti})$. In the probit case, we have $w^2=1$, while in the logit case $w^2=\pi^2/3$ (McCullagh and Nelder, 1989, p. 142).

The statistical challenge in a Bernoulli mixture model amounts to estimating $\boldsymbol{\beta}$, $\boldsymbol{\mu}$, and so-called *hyperparameters* θ of the distribution of $\boldsymbol{b}_t$ based on observations of $\boldsymbol{Y}_t$ as well as the covariates $\boldsymbol{x}_{ti}$ and $\boldsymbol{z}_{ti}$ for a number of time periods. For a more in-depth treatment of Bernoulli mixture models, see, e.g., Frey and McNeil (2003).

2.1.2. Fixed effects and random effects

The design vectors $\boldsymbol{x}_{ti}$ and $\boldsymbol{z}_{ti}$ can hold observed quantitative covariates as well as dummy variables indicating group membership. In the language of GLMMs the former vector specifies *fixed effects* through the fixed parameter vector $\boldsymbol{\beta}$, while the latter specify *random effects* through the random vector $\boldsymbol{b}_t$. Together, they determine the systematic risk $\boldsymbol{x}_{ti}'\boldsymbol{\beta}+\boldsymbol{z}_{ti}'\boldsymbol{b}_t$. Depending on the design of $\boldsymbol{z}_{ti}$, the model in (1) can feature random intercepts (if $\boldsymbol{z}_{ti}$ comprises dummy variables), random coefficients (if $\boldsymbol{z}_{ti}$ includes covariates), or both (Skrondal and Rabe-Hesketh, 2004, Chapter 3).

Fixed effects may be fully obligor-specific or shared for (parts of) the portfolio. Shared covariates that vary with time, such as macroeconomic variables or other observed risk factors, induce time-inhomogeneity in default rates. Wilson (1997) proposes a variety of variables that are relevant in portfolio credit risk. Obligor-specific covariates, such as balance sheet data, create heterogeneity among obligors. In effect, rating is an obligor-specific fixed effect, which for the sake of clarity has been kept separate from $\boldsymbol{\beta}$.

The design vectors $\boldsymbol{x}_{ti}$ and $\boldsymbol{z}_{ti}$ may contain time-dependent covariates that are known at the start of time period t or covariates that are realized during time period t, contemporaneously with the default indicators. This makes little difference when the model is fitted to historical data, but will be relevant if the model is used predictively. For variables realized during period t, an auxiliary time series model will be required to make predictions; see the discussion in Koopman, Lucas and Klaassen (2005).

The random effects $\boldsymbol{b}_t$ account for unobserved systematic risk, generating heterogeneity beyond that which can be captured with fixed effects. In other words, they are there to account for a phenomenon known as *overdispersion*, which is the tendency for data to show excess variance compared to what would be expected for independent responses. The random effects within a time period may be univariate or multivariate; the latter are useful in constructing hierarchical models with multiple tiers (cf. Section 2.2.1 and Model 4 in Section 3.2).

2.1.3. Generalized linear mixed models

The Bernoulli mixture model of Section 2.1.1 is a particular example of a generalized linear mixed model (GLMM). This family of models is well known in statistics, and has the potential to deal with data in continuous, discrete, or binary format involving multiple sources of random error. Another prominent industry model in credit that fits into the general framework is CreditRisk$^+$; where the number of defaults, conditionally on gamma-distributed latent factors, is Poisson (Credit Suisse First Boston, 1997).

The three basic constituents of a GLMM are (i) random effects $\boldsymbol{b}_t$ with distribution F_b and hyperparameters θ, (ii) a distribution from the exponential family for the conditional response variable Y_{ti} given $\boldsymbol{b}_t$, and (iii) a response function g (its inverse is known as *link function*) relating $\mu_{\kappa(t,i)}+\boldsymbol{x}_{ti}'\boldsymbol{\beta}+\boldsymbol{z}_{ti}'\boldsymbol{b}_t$ to the responses. The responses $Y_{t1},\ldots, Y_{tn_t}$ are assumed to be conditionally independent given $\boldsymbol{b}_t$. If no random effect $\boldsymbol{b}_t$ is present, the model is simply a generalized linear model (GLM), the theory of which is found in McCullagh and Nelder (1989), as are the concepts of exponential families and link functions. See also Chapter 7 of Fahrmeir and Tutz (1994).

2.2. Further modelling issues

2.2.1. The notion of units

GLMMs are well-suited to the modelling of responses on subjects, or units, that can be grouped into clusters (higher-level units) that are expected to be exposed to common latent factors. Classical examples of subjects (*level*-1 units) that form clusters (*level*-2 units) are students in a class or patients in a hospital. Repeated measurements on a subject are also encompassed in this setting by treating each measurement as a level-1 unit and the subject itself as level-2 unit. Clustering can be performed at more than two levels: for instance, with students as level-1 units, classes as level-2 units, and schools as level-3 units.

In this paper, we regard the particular time period under consideration as the highest-level unit, and all default indicators in the time period are repeated measurements. In its simplest form, this is a two-level model with companies within a time period as level-1 units and time periods as level-2 units. A three-level model; where period and industry sector affiliation define level-2 units, and

period alone defines level-3 units; will, however, also be considered. In this context, multivariate random effects are natural.

2.2.2. Homogeneous groups of obligors

When fully obligor-specific covariates are not considered, it makes sense to assume that all obligors in a given rating class and time period face the same risk of default, so that $\boldsymbol{x}_{ti} := \tilde{\boldsymbol{x}}_{tk}$ and $\boldsymbol{z}_{ti} := \tilde{\boldsymbol{z}}_{tk}$ for all i with $\kappa(t,i)=k$. This means that similarly-rated obligors in the same time period are modelled *exchangeably*, and is a simplifying assumption that is often made in practice. Defining $N_{tk} = \sum_{i=1}^{n_t} 1_{\{\kappa(t,i)=k\}} Y_{ti}$ to be the number of defaults for class k; we, under the assumptions of Section 2.1.1, have

$$N_{tk} | \boldsymbol{b}_t \sim \mathrm{Bin}\{n_{tk}, g(\mu_k + \tilde{\boldsymbol{x}}_{tk}' \boldsymbol{\beta} + \tilde{\boldsymbol{z}}_{tk}' \boldsymbol{b}_t)\}, \tag{4}$$

where $\tilde{\boldsymbol{x}}_{tk}$ and $\tilde{\boldsymbol{z}}_{tk}$ are the group-shared design vectors in period t. Note that the conditional independence property holds also for N_{tk_1} and N_{tk_2} for two different $k_1, k_2 \in \mathcal{K}$, and that the model (4) continues to fit into the GLMM-framework, albeit with conditionally independent binomial rather than Bernoulli responses.

2.2.3. Dynamic random effects

If the realizations of the random effects $\boldsymbol{b}_t$ are assumed to be independent for different units (i.e., time periods), then so are the responses on these units (i.e., defaults of companies in different time periods). This property is not always desired, and can be removed by imposing a dependence structure on the random effects. In the credit-risk context, it seems appropriate to assume that the current value of the latent systematic risk $\boldsymbol{b}_t$ depends on its value in the previous time period, and hence to consider a Markovian structure for the unobserved systematic component, in order to create the desired *serial* dependence. Traditional GLMM-applications feature spatial dependence rather than temporal, since units are often related by geographical location; see Clayton (1996) for an overview.

For the applications in Section 3, we investigate the particular case of a first-order auto-regressive, AR(1), time series

$$b_t = \alpha b_{t-1} + \phi \in_t, \qquad t \geq 1, \qquad \text{with } b_0 = \phi \in_0 / \sqrt{1-\alpha^2}, \tag{5}$$

where $\in_0, \in_1, \ldots$ are iid $N(0, 1)$. The AR(1)-time series in (5) is a Markov process that for $|\alpha|<1$ has a Gaussian stationary distribution with mean 0 and variance $\sigma^2 := \phi^2/(1-\alpha^2)$.

By interpreting (5) as the *state equation* and (4) as the *observation equation*, this defines a binomial *state space model* (or *hidden Markov model*) for the sequence of default counts (MacDonald and Zucchini, 1997, Chapter 2). In the credit-risk context, hidden Markov models have been proposed in Crowder et al. (2005), where a one-group model with a 2-state hidden Markov chain (b_t) is fitted to S&P data. In Section 3, we study a multi-group model driven by the Markov chain (5), evolving on R.

2.3. Estimation of GLMMs

2.3.1. Frequentist estimation[2]

In GLMMs, the unconditional distribution of the responses is obtained by integrating out the effect of the random effects, and this greatly complicates the use of standard maximum likelihood

[2] A number of software packages for GLMMs based on ML-methodology handle the case of independent units. In R, there is glmmPQL and glmmML, which are a part of the package MASS. For SAS, there is PROC NLMIXED, and for Splus 7, there is the Correlated Data library.

(ML) estimation in all but the simplest models. The general likelihood function for the responses on different units $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T$ follows from the conditional independence property, and is given by

$$L(\boldsymbol{\beta},\boldsymbol{\mu},\theta|D)=\int\cdots\int\prod_{t=1}^{T}\prod_{i=1}^{n_t}P(Y_{ti}=y_{ti}|\boldsymbol{b}_t,\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{x}_{ti},\boldsymbol{z}_{ti})f_b(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_T|\theta)\mathrm{d}\boldsymbol{b}_1\cdots\mathrm{d}\boldsymbol{b}_T, \quad (6)$$

where $D=\{\boldsymbol{y}_t,\boldsymbol{x}_t,\boldsymbol{z}_t\}_{t=1}^{T}$ denotes the observed data and $f_b(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_T|\theta)$ is the joint density of $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_T$. This is in general a high-dimensional integral (of dimension $\dim(\boldsymbol{b}_t)\times T$), but it can be considerably simplified if we assume iid random effects. In this case, (6) becomes a product of $\dim(\boldsymbol{b}_t)$-dimensional integrals, and for lower-dimensional random effects (particularly univariate) this can be evaluated numerically with high accuracy; see Frey and McNeil (2003) for an example. For higher-dimensional random effects, there are also a number of approximation methods that can be used, including penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL); see Breslow and Clayton (1993). However, models with correlated random effects are generally much too cumbersome to fit by numerical maximization of the likelihood.

In Chapter 3 of Gouriéroux and Monfort (1996), three generic approaches to likelihood inference in models of the form (6) with correlated random effects are identified: numerical approximations; application of the *expectation maximization* (EM) algorithm; or simulation of the full likelihood function using the importance sampling technique. The latter technique of *simulated maximum likelihood* is certainly very flexible and would be a possibility for our models, but we have chosen to use a computational Bayesian approach. For a recent application of simulated ML in the credit risk literature, see Koopman, Lucas and Daniels (2005). An application of the EM-algorithm in the context of GLMMs is given in McCulloch (1997).

2.3.2. Bayesian estimation[3]

As is well known, in Bayesian statistics we distinguish between known (i.e., observed) and unknown (i.e., unobserved) quantities. All unknowns, ϑ, are modelled as random quantities described by a *prior* probability distribution, comprising any information on ϑ we may have before observing the data, D. All model inference is based on the law of the unknowns after having observed the data, the posterior distribution $p(\vartheta|D)$, which is derived with Bayes' rule.

For the GLMMs of Section 2, ϑ contains the fixed effect coefficients $\boldsymbol{\beta}$ (including $\boldsymbol{\mu}$), the random effects $\boldsymbol{b}_t$, and the hyperparameters θ. This is depicted graphically in Fig. 2, where unknowns ϑ are circled and observed quantities $D=\{(\boldsymbol{y}_t,\boldsymbol{x}_t,\boldsymbol{z}_t)\}_{t=1}^{T}$ boxed. The joint prior distribution of ϑ can be written

$$p(\vartheta)=p(\boldsymbol{\beta},\boldsymbol{b}_t,\theta)=p(\boldsymbol{b}_t|\theta)p(\theta)p(\boldsymbol{\beta}), \quad (7)$$

where we have assumed that $\boldsymbol{\beta}$ and θ are independent a priori. For a detailed Bayesian interpretation of GLMMs in this context, we refer to Clayton (1996). The posterior distribution of ϑ will in particular contain posterior information about $\boldsymbol{b}_t$; a fact which we exploit in Section 3.

Unfortunately, the posterior distribution is generally unobtainable by analytical means; only its functional form is known. This is often enough to apply Markov chain Monte Carlo (MCMC) algorithms such as the *Gibbs sampler*. The objective of these procedures is to generate

[3] The software package GLMMGibbs for R fits binomial- and Poisson-type GLMMs (Myles and Clayton, 2001). It is based on ARS-sampling and allows some flexibility using correlated random effects, mainly for neighbouring regions. For general Gibbs sampling problems, there is also BUGS.

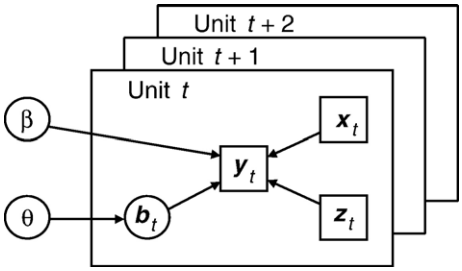

Fig. 2. Hierarchical representation (directed acyclic graph) of a GLMM. Observed quantities are in boxes and unobserved in circles. If a variable does not depend directly on another variable, then there is no arrow between them. Serial dependence is accomplished with dependence among the sequence of $\boldsymbol{b}_t$'s.

realizations from the joint posterior distribution of the unknowns through the use of a chain of simulation steps that produces realizations from the posterior in the limit. By construction, the Gibbs sampler reduces the multivariate simulation problem to a series of simulations of lower (or even univariate) dimension.

More explicitly, suppose we decompose ϑ into $(\vartheta_1,\ldots,\vartheta_J)$, where $J\geq 2$, and denote by $\vartheta_{-\ell}$ the vector obtained by removing ϑ_ℓ from ϑ. Note that ϑ_ℓ can be univariate, but need not be so. The basic building block of the Gibbs sampler is the distribution of ϑ_ℓ given $\vartheta_{-\ell}$ and D, which is known as the *full conditional distribution*. The Gibbs sampler proceeds by in turn simulating from all J full conditionals and updating the components of ϑ; the full algorithm can be found in Robert and Casella (1999). The algorithm is started, and after an initial number of iterations, known as the *burn-in* period, the simulated values are taken to be realizations from the posterior of ϑ.

By exploiting conjugacy, we obtain full conditionals that are easy to simulate from. Although this facilitates the implementation, this is by no means a requirement, and even in simple models some full conditionals often turn out to be non-standard (cf. Appendix A). In the univariate case, the ARS-(Adaptive Rejection Sampling) and ARMS-algorithms can often be employed (Gilks, 1992). While the former is intended only for log–concave densities, the latter is capable of handling arbitrary densities, and plays an important role in the fitting of our models.

3. An empirical study of S&P default data

3.1. Description of the data

The default data have been extracted from Standard and Poor's CreditPro^{TM} 6.6 database, and consist of 5676 U.S. obligors from all 13 industry sectors. The default counts have been collected for six-month periods, ranging from January 1981 to December 2000 ($T=40$ periods). Obligors whose rating is withdrawn have been excluded from consideration in the time period of the withdrawal. Obvious duplicates in the database (e.g., holding companies) have been removed as well. The rating classes included in the analysis are $\mathcal{K} = \{\mathrm{CCC}, \mathrm{B}, \mathrm{BB}, \mathrm{BBB}, \mathrm{A}\}$, where, for simplicity, qualifiers have been suppressed. This means that $k\in\mathcal{K}$ corresponds to one of the actual ratings k^+, k, or k^-. As is customary, we merge CCC, CC, and C into a single rating class: CCC. The rating classes AAA and AA have been excluded from the study, since these obligors hardly ever default directly.

Our most complex model considers sector effects as well as time period effects. When fitting this model, we analyse a subset of the above dataset, consisting of 3553 U.S. obligors from seven selected S&P industry sectors; see Table 1 for details.

Table 1
Industry sectors used in the analysis of Model 4

Sector	Name	# obligors
1	"Aerospace, automotive, capital goods, metal"	725
2	"Consumer, service sector"	888
3	"Leisure time, media"	545
4	"Utility"	477
5	"Health care, chemicals"	395
6	"High tech, computers, office equipment"+"Telecom"	523

Sector 6 is a merger of two regular S&P sectors.

3.2. Models

Since we do not consider individual obligor-level covariates, we group firms with a similar rating together in each time period as in Section 2.2.2. We use Gibbs sampling to fit four GLMMs of increasing complexity. In all models, we assume that the latent state process (b_t) is univariate, first-order autoregressive with variance $\sigma^2=\phi^2/(1-\alpha^2)$ as specified in (5). The response function employed is the *logit response* $g(x)=1/(1+\exp\{-x\})$, the canonical response in GLMMs for conditionally binomial data. The notation of Sections 2.2.2 and 2.2.3 applies.

Beginning with Model 2, we include an observed macroeconomic variable (x_t), in an attempt to partly explain the time-heterogeneity in the default rates. We use the Chicago Fed National Activity Index (CFNAI), which is published on a monthly basis. The Bayesian methodology also allows us to compare features of the posterior distribution of the latent systematic effect b_t (e.g., its mean) with the macroeconomic variable x_t in each time period. This provides a natural assessment of x_t as an explanatory variable for defaults, in particular with regard to detecting lags between the cycle of the index and the actual default cycle.

Model 1

We assume that $N_{t1},\ldots,N_{tK}$ are conditionally independent, given b_t, with

$$N_{tk}|b_t \sim \text{Bin}\{n_{tk}, g(\mu_k - b_t)\}, \qquad k\in\mathcal{K}. \tag{8}$$

The above sign convention implies that positive b_t's correspond to favourable economic conditions. The model set-up is described graphically in Fig. 3 as a directed acyclic graph. A

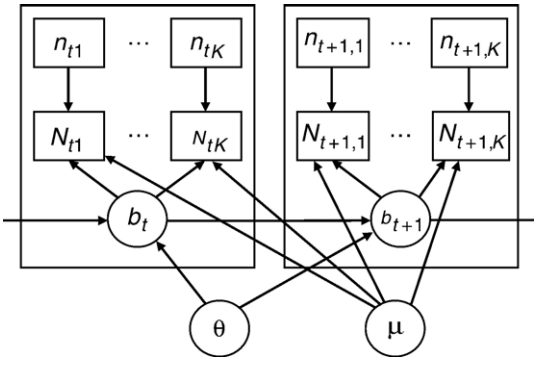

Fig. 3. Hierarchical representation of Model 1 with K homogeneous groups. Rectangular nodes are observed and circular nodes unobserved. The large rectangular nodes indicate the border area of the half-yearly top-level *units*.

derivation of the full conditional distributions is provided in Appendix A. In this model, the implied asset correlation (3) is simply given by $\sigma^2/(\sigma^2+\pi^2/3)$.

Model 2

Model 1 is extended by including the observed business cycle-covariate x_t, which in our case is the CFNAI for the first calendar month of period t. $N_{t1},\ldots,N_{tK}$ remain conditionally independent, given b_t, but with

$$N_{tk}|b_t \sim \text{Bin}\{n_{tk}, g(\mu_k - x_t\beta - b_t)\}, \tag{9}$$

where β is an additional parameter to be estimated. In this model, the implied asset correlation (3) in period t, given the value of x_t, is $\sigma^2/(\sigma^2+\pi^2/3)$.

Model 3

We now allow both the mean and variance of the systematic risk component to vary with rating category. $N_{t1},\ldots,N_{tK}$ are conditionally independent, given b_t, with

$$N_{tk}|b_t \sim \text{Bin}\{n_{tk}, g(\mu_k - x_t\beta_k - \phi_k b_t)\}, \tag{10}$$

where the variance of b_t is constrained to $1/(1-\alpha^2)$ for identifiability. Note that (10) falls into the class of GLMMs by writing $\boldsymbol{b}_t = b_t(\phi_1,\ldots,\phi_K)'$, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_K)'$, and setting $\tilde{\boldsymbol{x}}_{tk} = x_t\boldsymbol{e}_k$ and $\tilde{\boldsymbol{z}}_{tk} = \boldsymbol{e}_k$, where $\boldsymbol{e}_k$ denotes the kth unit vector in R^K. In this model, the implied asset correlation (3) for two obligors in rating class k, given the value of x_t, is $\sigma_k^2/(\sigma_k^2+\pi^2/3)$, where $\sigma_k^2 = \phi_k^2/(1-\alpha^2)$. See also Lopez (2004) for a discussion of rating-specific factor loadings.

Model 4

The clusters of companies in each time period are now divided into subclusters according to industry sector. These subclusters are allotted sector-specific random effects, as discussed in Section 2.2.1. Let $\mathcal{S} = \{1,\ldots,S\}$ be an index set for the industry sectors. We collect the number of obligors, n_{tsk}, and the numbers of defaults, N_{tsk}, for each rating category $k \in \mathcal{K}$, sector $s \in \mathcal{S}$, and time period t. Observe that $\sum_s N_{tsk}$ and $\sum_s n_{tsk}$ yield N_{tk} and n_{tk} of the previous models. (The calibration of Model 4 in this study, however, uses only seven selected S&P industry sectors and $S=6$, see Table 1.)

We assume that given b_t (as defined previously) and $b_{t1},\ldots,b_{tS}$, the observations $(N_{tsk})_{s\in\mathcal{S},k\in\mathcal{K}}$ are conditionally independent with

$$N_{tsk}|b_t, b_{ts} \sim \text{Bin}\{n_{tsk}, g(\mu_k - x_t\beta - b_t - b_{ts})\}. \tag{11}$$

The variates (b_{ts}) are assumed to be iid $N(0,\omega^2)$, and independent of (b_t). Intuitively, b_{ts} can be thought of as a random correction to b_t for sector s. It is therefore of interest to compare the variance of b_t with ω^2. If ω^2 is small compared to σ^2, the variation in default activity among industry sectors is negligible against the background of time period variability, and vice versa.

The notation can be simplified by introducing $b_{ts}^* := b_t + b_{ts}$. It is then clear that $(b_{t1}^*,\ldots, b_{tS}^*)$ is multivariate Gaussian with mean zero and

$$\text{cov}(b_{ti}^*, b_{tj}^*) = \begin{cases} \sigma^2+\omega^2 & \text{if } i=j, \\ \sigma^2 & \text{else.} \end{cases}$$

This fits into the GLMM-framework by setting $\boldsymbol{b}_t = (b_{t1}^*,\ldots,b_{tS}^*)'$ and $\boldsymbol{z}_{ti} = \boldsymbol{e}_s \in R^S$, where s is the sector affiliation of obligor i.

An important implication of sector-specific effects is that implied asset correlations depend on whether the obligors belong to the same sector; given x_t, the implied within-sector asset correlation is $(\sigma^2+\omega^2)/(\sigma^2+\omega^2+\pi^2/3)$, whereas the across-sector asset correlation is merely $\sigma^2/(\sigma^2+\pi^2/3)$.

3.3. Implementation details

We use customized code in C to fit the models of Section 3.2 by Gibbs sampling. The running time of a 10,000-iteration simulation amounts to a few minutes.

3.3.1. Choice of prior

As always in Bayesian analysis, the posterior results depend on the prior distribution. In the present study, we follow the most common approach of using *non-informative* priors for the intercepts, regression coefficients, and hyperparameters of our models.

In Model 1 and all subsequent models, the intercept parameters $\boldsymbol{\mu}$ are given a zero-mean, ordered Gaussian distribution with covariance matrix $\tau^2 I_{K\times K}$, where $\tau=100$. By ordered, we mean that we impose $\mu_1>\mu_2>\ldots>\mu_K$, so that default probabilities decrease with increasing creditworthiness. The autoregressive parameter α is given a uniform prior on the interval $(-1, 1)$. The variance of the innovations ϕ^2 in (5) is assigned an *improper* prior decaying as $1/x$; this corresponds to a limiting case of the inverse-gamma distribution (see Appendix A for precise details). This prior is standard practice for a scaling parameter, and has the advantage that we do not need to specify a modal value, which would be the case with a proper inverse-gamma prior. We apply similar priors to other scaling parameters in all models with the sole exception of the variance parameters $\phi_1^2,\ldots,\phi_K^2$ of Model 3. These are given independent, proper inverse-gamma priors for technical reasons addressed in Appendix A.

The coefficient β of Models 2 and 4 is assigned a zero-mean Gaussian prior with variance τ^2; likewise $\beta_1,\ldots,\beta_K$ of Model 3 are a priori independent with $N(0,\ \tau^2)$-priors. The variance parameter ω^2 of Model 4 is given the same improper inverse-gamma prior as ϕ^2.

3.3.2. Model comparison

We follow a modern approach to Bayesian model comparison based on *cross-validation predictive densities* and regression diagnostics derived from these. In particular, we consider the *conditional predictive ordinate* (CPO), defined as

$$\mathrm{CPO}_t := P(N_{t1,\mathrm{obs}},\ldots,N_{tK,\mathrm{obs}}|\{N_{t'1,\mathrm{obs}},\ldots,N_{t'K,\mathrm{obs}} : t'\in\mathcal{T}, t'\neq t\}),$$

where $\mathcal{T} := \{1,\ldots,T\}$. The cross-validation predictive density is attractive in that it suggests how likely the joint observation $N_{t1,\mathrm{obs}},\ldots,N_{tK,\mathrm{obs}}$ is, when the model is fitted to all observations except $N_{t1,\mathrm{obs}},\ldots,N_{tK,\mathrm{obs}}$. A further diagnostic of this kind is the univariate, standardized residual d_{tk}:

$$d_{tk} := \frac{N_{tk,\mathrm{obs}}-E(N_{tk}|\{N_{t'k',\mathrm{obs}} : t'\in\mathcal{T}, k'\in\mathcal{K}, (t',k')\neq(t,k)\})}{\sqrt{\mathrm{var}(N_{tk}|\{N_{t'k',\mathrm{obs}} : t'\in\mathcal{T}, k'\in\mathcal{K}, (t',k')\neq(t,k)\})}}. \tag{12}$$

The output of the Gibbs sampler can be utilized to implement the above quantities. By plotting $\{(t,\mathrm{CPO}_t) : t\in\mathcal{T}\}$ and/or $\{(t,d_{tk}) : t\in\mathcal{T}\}$ for all competing models, we obtain a graphical assessment of the different model fits. Obviously a good model should have CPO_t large and $|d_{tk}|$ small. An excellent survey of the above issues and more is given in Gelfand (1996).

3.4. Results

3.4.1. Model by model summary

Model 1

The trajectories of the Gibbs sampler output for the parameters μ_{BBB}, ϕ, and α for the first 2500 iterations are displayed in Fig. 4. Bayesian point estimates are obtained by taking the sample mean of the simulated output (after the burn-in), and standard errors are obtained by taking the corresponding sample standard deviation. The point estimates of all parameters are summarized in the upper part of Table 2.

The histogram in Fig. 4, and construction of a Bayesian credible set (from sample quantiles of the posterior sample), formally confirms that the hypothesis $\alpha=0$ is strongly rejected in favour of $\alpha>0$. This is evidence of profound serial dependence in the data. The point estimate of the implied asset correlation is 7.4%.

Model 2

Point estimates of the parameters are given in the lower part of Table 2. Although the coefficient β is highly significant (meaning that the CFNAI has explanatory power), the inclusion of the CFNAI-factor is not sufficient to fully capture the systematic risk present in the data. Firstly, there is substantial residual serial dependence, which can be read off by the virtually

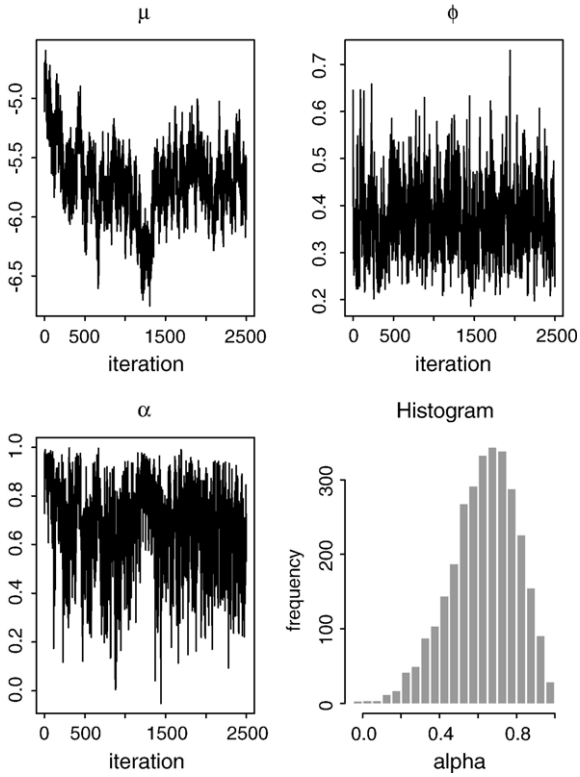

Fig. 4. The Gibbs sampler for the first 2500 iterations of the parameters μ_{BBB}, ϕ, and α of Model 1. The simulation stabilizes after approximately 1500 iterations. The histogram of α is based on iterations 10,000–15,000.

Table 2
Posterior mean (standard deviation) of the parameters of Model 1 and of Model 2

	μ_A	μ_{BBB}	μ_{BB}	μ_B	μ_{CCC}	ϕ	α	β
Model 1: Without CFNAI (x_t)								
Mean	−9.139	−7.185	−5.749	−3.907	−1.627	0.394	0.642	–
(S.D.)	(0.649)	(0.330)	(0.249)	(0.208)	(0.213)	(0.082)	(0.167)	–
Model 2: With CFNAI (x_t)								
Mean	−9.075	−7.121	−5.689	−3.841	−1.563	0.367	0.629	0.221
(S.D.)	(0.653)	(0.323)	(0.236)	(0.186)	(0.195)	(0.079)	(0.173)	(0.096)

The posterior probability of $\beta \leq 0$ is <1%, hence β is highly significant. The estimates are based on iterations 10,000–15,000 of the Gibbs sampler.

unchanged posterior estimate of α. Secondly the point estimate of ϕ^2 is reduced by only 16%, meaning that the implied asset correlation is 6.3%, given x_t.

The upper plot of Fig. 5 compares $\{(t, x_t\beta) : t \in \mathcal{T}\}$ of Model 2 with the posterior mean of $\{(t, b_t) : t \in \mathcal{T}\}$ of Model 1. There seems to be a fair amount of co-movement between the two series ($x_t\beta$) and (b_t), but it is obvious that x_t does not track b_t particularly accurately, and that x_t does not fully capture the default activity. This illustrates the problems associated with observed proxies for the systematic risk. The lower plot of Fig. 5 compares the posterior path $\{(t, b_t) : t \in \mathcal{T}\}$ of Models 1 and 2. Although the paths are very similar, the reduced variance of (b_t) when (x_t) is explicitly modelled can be detected.

Model 3

In Table 3, we see that for investment grade A, there is increased uncertainty in the parameter estimates. Due to the rare occurrence of A-rated obligors defaulting (3 time periods out of 40 with 1 default each), it is not surprising that the effect of the systematic risk is difficult to detect. This reasoning applies also to the BBB-class to a certain extent, but is less of an issue for rating classes that show more frequent defaults. The BB-category, for instance, displays a significant CFNAI-

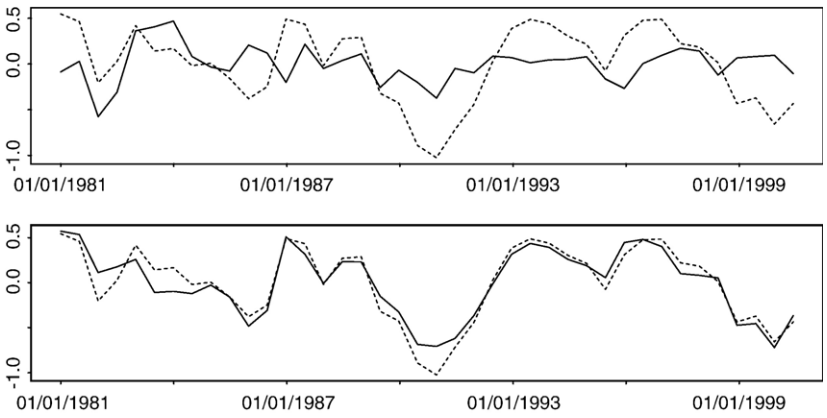

Fig. 5. Visual comparison of systematic risk factors. The upper plot displays $\{(t, x_t\beta): t=1,\ldots,T\}$ of Model 2 (solid line) and the unobserved factor $\{(t, b_t): t=1,\ldots,T\}$ of Model 1 (dashed line); in both cases their posterior means. The lower plot compares the posterior mean of the latent process (b_t) in Model 2 (solid) and Model 1 (dashed). All estimates are based on iterations 10,000–15,000 of the Gibbs sampler.

Table 3
Posterior mean (standard deviation) of the parameters of Model 3

Model 3: With CFNAI (x_t)						
k	μ_k		ϕ_k		β_k	
	Mean	(S.D.)	Mean	(S.D.)	Mean	(S.D.)
A	−10.079	(1.962)	1.191	(0.964)	−0.654	(0.892)
BBB	−7.170	(0.473)	0.375	(0.223)	0.457	(0.322)
BB	−5.759	(0.297)	0.257	(0.129)	0.683	(0.187)
B	−3.807	(0.381)	0.442	(0.097)	0.192	(0.121)
CCC	−1.464	(0.238)	0.261	(0.091)	0.141	(0.112)
α	0.680	(0.175)				

The estimates are based on iterations 10,000–15,000 of the Gibbs sampler.

coefficient β_{BB}. Moreover, its magnitude is twice that of σ_B, implying that the CFNAI-series appears to do a relatively good job of describing the default activity of the BB-category.

The CCC-class, however, appears largely unaffected by the CFNAI-loading, and shows a moderate dependence on the hidden risk factor b_t. One should bear in mind that one of the difficulties associated with inference about the CCC-class is its relatively small size (it ranges from approximately 10 obligors in 1981 to 70 obligors in 2000). Another is its possible heterogeneity, being a merger of C, CC, and CCC.

Model 4

In Table 4, we compare Model 4 with Model 2 when applied to our chosen subset of the data with sector information, cf. Table 1. The sector-specific variance ω^2 is only 24% smaller than σ^2, suggesting that there is heterogeneity among sectors that should be taken into consideration. In terms of implied asset correlations, we obtain a within-sector implied asset correlation of 10.9%, whereas the across-sector counterpart is only 6.5%. In contrast, the model without sector-specific random effects suggests an overall implied asset correlation of 7.5%. The different implied asset correlations have important regulatory consequences; see the Conclusions.

3.4.2. Model comparisons

Since the CFNAI clearly has explanatory power, Model 2 is our baseline case. The philosophy of our model comparison is to address the issues of rating specific factor loadings as well as industry-sector effects, where the latter comparison takes place in the reduced dataset. For reference, we include the simple GLM $N_{tk} \sim \mathrm{Bin}\{n_{tk},\ g(\mu_k - x_t\beta)\}$ without random effects as

Table 4
Posterior mean (standard deviation) of the parameters of Model 4 calibrated to the $S=6$ sectors of Table 1

	μ_A	μ_{BBB}	μ_{BB}	μ_B	μ_{CCC}	ϕ	ω	α	β
Model 4: Without sector-specific random effect ($\omega^2=0$)									
Mean	−30.103	−7.785	−5.859	−4.083	−1.547	0.415	–	0.596	0.224
(S.D.)	(11.607)	(0.525)	(0.306)	(0.244)	(0.253)	(0.106)	–	(0.224)	(0.122)
Model 4: With sector-specific random effect									
Mean	−30.118	−7.836	−5.904	−4.112	−1.550	0.349	0.417	0.683	0.237
(S.D.)	(11.603)	(0.538)	(0.325)	(0.271)	(0.279)	(0.111)	(0.117)	(0.223)	(0.120)

The estimates are based on iterations 10,000–15,000 of the Gibbs sampler.

Table 5
Model diagnostics of Section 3.4.2

$\frac{1}{T}\sum_{t=1}^{T}\log(\mathrm{CPO}_t)$			$\frac{1}{T}\sum_{t=1}^{T}\log(\mathrm{CPO}_t)$		
Model 0	Model 2	Model 3	Model 0	Model 2	Model 4
−8.02	−6.82	−6.69	−13.87	−12.89	−12.31
$\#\{t\in\{1,..,T\}:\mathrm{CPO}_t(j)>\mathrm{CPO}_t(i)\}$			$\#\{t\in\{1,..,T\}:\mathrm{CPO}_t(j)>\mathrm{CPO}_t(i)\}$		
$i\backslash j$	Model 2	Model 3	$i\backslash j$	Model 2	Model 4
Model 0	28	32	Model 0	30	33
Model 2	–	23	Model 2	–	33

Those for the full S&P dataset are to the left; those of the reduced dataset, with only seven sectors, to the right. The lower panel holds the number of time periods, out of $T=40$, in which the CPO of Model j is better than that of Model i. ($\mathrm{CPO}_t(i)$ is the CPO under Model i.) Model 0 is the GLM $N_{tk}\sim\mathrm{Bin}\{n_{tk}, g(\mu_k-x_t\beta)\}$. The other models are defined in Section 3.2.

Model 0. The relevant summary statistics are listed in Table 5: we consider the average log(CPO)-value under each model as well as the number of time periods in which the more elaborate models have a higher CPO-value. It is evident that any form of GLMM is superior to Model 0.

The comparison of Models 2 and 3 is given to the left in Table 5. Although the average log (CPO)-value is slightly better under Model 3, the number of time periods in which Model 3 has the better CPO-value is only 23 out of 40. This leads us to classify the fit of Model 3 as only marginally better than that of Model 2.

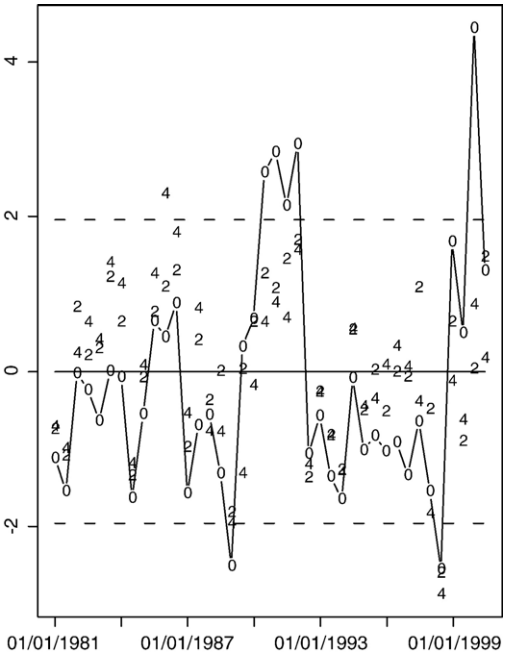

Fig. 6. Plot of $\{(t, d_{tk}):t=1,\ldots,T\}$ for $k=$B, see (12), under Model 0 (GLM with CFNAI-covariate), Model 2, and Model 4. Model indicated by number.

An analogous comparison of Models 2 and 4 clearly favours Model 4: in more than 30 out of 40 time periods, the fit of Model 4 is preferred (see the right part of Table 5). For sake of illustration, we include a plot of $\{(t, d_{tk}) : t \in \mathcal{T}\}$ for $k=B$ as well, see Fig. 6. The GLMMs perform noticeably better towards the second half of the time span, where the B-rating class is approximately four times its initial size.

4. Conclusions

We have highlighted the usefulness of the class of GLMMs as a flexible statistical framework for modelling correlated firm defaults. GLMMs allow for a convenient specification of systematic portfolio risk in terms of observed fixed effects and unobserved random effects, in order to capture inhomogeneities in default rates, either in time or across the portfolio. We have also pointed out that several well known industry models actually fall into this class.

Secondly, we have shown the feasibility of a Bayesian approach to inference for such models. MCMC-algorithms such as the Gibbs sampler are capable of dealing with models featuring complex latent structures, such as serially correlated and multivariate random effects. Without entering into the longrunning Bayesian versus frequentist debate, some further reasons for the Bayesian approach are: (i) standard errors of parameters are easily obtained from the simulation output; (ii) inference about derived parameters (default correlations, implied asset correlations) is as easy as inference about primary parameters; (iii) posterior information about random effects can be compared with macroeconomic variables; (iv) posterior information about future default events (i.e., forecasting) is easily achieved, as in Chapter 2 of Wendin (2006); (v) the simulation algorithms are fast. An extra benefit of the Bayesian approach could potentially be active use of prior distributions, which might be advantageous in situations where the default event is very rare: we could assess a prior estimate of σ_k (or ϕ_k) by analysing asset correlations inferred from equity data, as in the industry model KMV (Kealhofer and Bohn, 2001). This could, in a sense, allow us to draw stronger conclusions about default risk than is possible from an analysis of empirical defaults alone. This remains a subject for further research.

The third focus of this paper was an empirical analysis of half-yearly default data from S&P. Despite using an observed macroeconomic variable as proxy for the state of the business cycle, our analysis clearly suggests the presence of a residual, cyclical, latent component in the systematic risk. This in particular could be taken as evidence that S&P ratings are more *through-the-cycle* than *point-in-time*, cf. Rösch (2005). Although the macroeconomic variable was not able to capture the full variability in default rates, its presence reduced the variance of the random effects and thus the implied asset correlations.

The empirical analysis also addressed the two issues of rating-specific factor loadings as well as heterogeneity across industry sectors. Our analysis of the latter showed evidence of increased implied asset correlations within an industry sector. This raises some doubt about the applicability of single-factor models (i.e., Models 1–3) for the determination of regulatory capital as proposed by the BCBS (Basel Committee on Banking Supervision, 2002). Moreover, we do not necessarily concur with their proposed view that implied asset correlations fall monotonically with increasing probability of default; see also Lopez (2004). It should, however, be noted that inference about upper investment-grade factor loadings is highly uncertain due to the rare occurrence of defaults. For the subinvestment grade, the estimates are more precise, except possibly for the CCC-rating, which due to its small size is difficult to assess.

The estimates of implied asset correlations in the single-factor models agree with previous empirical studies such as Rösch (2005), and are generally lower than those recommended by the

regulator. It is, however, worth noting that the multivariate risk factor proposed in this paper, provides empirical evidence that implied asset correlations can in fact be substantially higher among obligors sharing industry sector. An important message, therefore, is that the issue of heterogeneity between industry sectors, or geographical regions, should be properly addressed.

Acknowledgment

We are grateful to Standard and Poor's for providing the dataset used in Section 3, and also to two anonymous referees whose detailed comments helped us considerably in our revision of the paper. The second author gratefully acknowledges financial support from NCCR Financial Valuation and Risk Management (a research program supported by the Swiss National Science Foundation).

Appendix A. Deriving the Full Conditionals of Model 1

In order to implement the Gibbs sampler, the full conditionals must be determined. In what follows, we use a notation that is common in literature on Gibbs sampling: $[X]$ denotes the (unconditional) density of X, $[X|Y]$ is the conditional density of X given Y, and by $[X|\cdot]$, we mean the full conditional distribution of X. It is convenient to use the notation $\underline{X}$ for the full sequence $X_1,\ldots,X_T$.

It is worth emphasizing that our implementation relies on the full conditional distribution of the random effect b_t, whereas other software (e.g., GLMMGibbs) simulate from the full conditional of the default probability instead. If the default probability is near zero, which is true for investment-grade ratings, the latter approach may lead to numerical instabilities. A detailed account of the derivations, including an implementation in C, is provided in Wendin (2006).

We first note that $\underline{b}:=(b_1,\ldots,b_T)$ is multivariate Gaussian with covariance matrix elements

$$\Sigma_{ij} = \mathrm{cov}(b_i, b_j) = \phi^2\alpha^{|i-j|}/(1-\alpha^2), \qquad i,j\in\{1,\ldots,T\}.$$

Its inverse Σ^{-1} is tridiagonal with diagonal elements $\phi^{-2}(1,\ 1+\alpha^2,\ldots,1+\alpha^2,1)$, off-diagonal elements $-\phi^{-2}\alpha$, and determinant $\phi^{-2T}\ (1-\alpha^2)$. This parameterization of the AR(1)-sequence makes the full conditional distribution of α log–concave.

Define $n_t:=(n_{t1},\ldots,n_{tK})$ and $N_t:=(N_{t1},\ldots,N_{tK})$. The key to all full conditional distributions is the joint distribution function of data and unknowns

$$\begin{aligned}[\underline{N},\underline{n},\underline{b},\boldsymbol{\mu},\alpha,\phi] &= [\underline{N}|\underline{n},\underline{b},\boldsymbol{\mu}][\underline{b}|\alpha,\phi][\boldsymbol{\mu}][\alpha][\phi]\\ &= \left(\prod_{t=1}^{T}\prod_{k\in\mathcal{K}}[N_{tk}|n_{tk},b_t,\mu_k]\right)[\underline{b}|\alpha,\phi][\boldsymbol{\mu}][\alpha][\phi],\end{aligned}$$

whose above factorization follows by conditional independence arguments, cf. Fig. 3. The full conditional of α is found by applying the conditional probability formula, and picking out only factors depending explicitly on α, for which we use the $\propto$ sign (see also Gilks (1996)):

$$[\alpha|\cdot] = \frac{[\underline{N},\underline{n},\underline{b},\boldsymbol{\mu},\alpha,\phi]}{[\underline{N},\underline{n},\underline{b},\boldsymbol{\mu},\phi]} \propto [\underline{N},\underline{n},\underline{b},\boldsymbol{\mu},\alpha,\phi] \propto [\underline{b}|\alpha,\phi][\alpha] \tag{A.1}$$

$$\propto \sqrt{\det\left(\Sigma^{-1}\right)}\ \exp\left\{-\frac{1}{2}\,\underline{b}\Sigma^{-1}\,\underline{b}'\right\}[\alpha]$$

$$\propto \sqrt{1-\alpha^2}\ \exp\left\{-\frac{1}{2}\phi^{-2}(C_1(\underline{b})\alpha^2 - C_2(\underline{b})\alpha)\right\}[\alpha]. \tag{A.2}$$

$C_1(\cdot)$ and $C_2(\cdot)$ are identified by performing the vector–matrix multiplications. By choosing a uniform prior $[\alpha]$ on $(-1, 1)$, Eq. (A.2) is log–concave in α and can be simulated from with the ARS-algorithm. For ϕ, we obtain

$$[\phi|\cdot] \propto [\underline{b}|\alpha,\phi][\phi] = \{\text{see Eqn.(A.1)}\} \propto \phi^{-T}\ \exp\{-C_3(\underline{b},\alpha)\phi^{-2}\}[\phi].$$

If $1/\phi^2$ is assigned a $\Gamma(\eta,\nu)$-prior (ϕ^2 is *inverse-gamma*), we for suitable choices of η and ν obtain a vague prior for ϕ. We choose the improper prior $(\eta, \nu)=(0, 0)$, that gives relatively large weight to the assertion that ϕ^2 is near 0; see Gelfand et al. (1990). The resulting full conditional is, however, proper, and furthermore remains inverse-gamma, which facilitates simulation. [4]

The full conditional distributions of all other model elements are found analogously and can all be simulated from with the ARMS-algorithm (Gilks, 1992). $\boldsymbol{\mu}$ is assigned a zero-mean, ordered Gaussian prior with variance $\tau^2 I_{K\times K}$, where τ is large, and $I_{K\times K}$ is the $K\times K$-identity matrix. Its density takes the form of an ordinary $N(0, \tau^2 I_{K\times K})$-vector times an indicator function, which imposes the ordering of the intercepts. For $k\in\mathcal{K}$, $[\mu_k]$ is $N(0,\tau^2)$ and

$$[\mu_k|\cdot] \propto \prod_{t=1}^{T}[N_{tk}|n_{tk},b_t,\mu_k][\boldsymbol{\mu}] \propto \prod_{t=1}^{T}[N_{tk}|n_{tk},b_t,\mu_k][\mu_k]I_{\{x\in R:\mu_{k+1}<x<\mu_{k-1}\}}(\mu_k).$$

(We use the convention $\mu_{K+1}=-\infty$ and $\mu_0=\infty$.) In order to apply the ARMS-algorithm to $\underline{b}$, we treat its components individually. The conditional distribution of b_t, given $\underline{b}_{-t}=(b_1,\ldots,b_{t-1},b_{t+1},\ldots,b_T)$, is Gaussian and depends on the one or two neighbouring b's. The full conditional reads

$$[b_t|\cdot] \propto \prod_{k\in\mathcal{K}}[N_{tk}|n_{tk},b_t,\mu_k][b_t|\underline{b}_{-t},\alpha,\phi] \propto \prod_{k\in\mathcal{K}}[N_{tk}|n_{tk},b_t,\mu_k][b_t|b_{t-1},b_{t+1},\alpha,\phi].$$

References

Bangia, A., Diebold, F., Kronimus, A., Schagen, C., Schuermann, T., 2002. Ratings migration and the business cycle, with application to credit portfolio stress testing. J. Bank. Finance 26, 445–474.

Basel Committee on Banking Supervision (2002, October). Quantitative impact, Study 3, Technical guidance. Bank of International Settlements.

Breslow, N., Clayton, D., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 88, 9–25.

Brockwell, P., Davis, R., 1991. Time Series: Theory and Methods, 2nd ed. Springer, New York.

Clayton, D., 1996. Generalized linear mixed models. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 275–301.

Credit Suisse First Boston, 1997. CreditRisk$^+$, a Credit Risk Management Framework. Technical report.

Crowder, M., Davis, M., Giampieri, G., 2005. Analysis of default data using hidden Markov models. Quantit. Finance 5 (1), 27–34.

Egloff, D., Leippold, M., Vanini, P., 2004. A simple model of credit contagion. Preprint, Swiss Banking Institute, University of Zurich.

[4] Notice that in Model 3, the scale parameter ϕ_k enters the joint density through the factor $[N_{tk} \mid n_{tk}, b_t, \mu_k, \phi_k, \beta_k]$. In particular, the inverse-gamma distribution is no longer a conjugate prior for ϕ_k^2. With a proper prior for ϕ_k^2 we may, however, use the ARMS-algorithm to simulate from $[\phi_k|\cdot]$. The calibration of Model 3 therefore employs independent inverse-gamma priors for $\phi_1^2,\ldots,\phi_K^2$ with $(\eta, \nu)=(\delta, \delta)$, where δ is set to 0.01. This is a proper prior that despite its mode at $\delta/(\delta+1)$ is very similar to the case $\eta=\nu=0$, if δ is small.

Fahrmeir, L., Tutz, G., 1994. Multivariate Statistical Modelling Based On Generalized Linear Models. Springer, New York.
Frey, R., McNeil, A., 2003. Dependent defaults in models of portfolio credit risk. J. Risk 6 (1), 59–92.
Gagliardini, P., Gouriéroux, C., 2005. Stochastic migration models with application to corporate risk. J. Financ. Econ. 3 (2), 188–226.
Gelfand, A., 1996. Model determination using sampling-based methods. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 145–161.
Gelfand, A., Hills, S., Racine-Poon, A., Smith, A., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Am. Stat. Assoc. 85, 972–985.
Gilks, W., 1992. Derivative-free adaptive rejection sampling for Gibbs sampling. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford, pp. 641–649.
Gilks, W., 1996. Full conditional distributions. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 75–88.
Gordy, M., Heitfield, E., 2002. Estimating default correlations from short panels of credit rating performance data. Technical Report. Federal Reserve Board.
Gouriéroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. CORE Lecture Series. Oxford University Press, New York.
Joe, H., 1997. Multivariate Models and Dependence Concepts. Chapman & Hall, London.
Kealhofer, S., Bohn, J.R., 2001. Portfolio management of default risk. Technical Report. KMV.
Koopman, S., Lucas, A., Daniels, R., 2005. A non-Gaussian panel time series model for estimating and decomposing default risk. Working Paper. Tinbergen Institute.
Koopman, S., Lucas, A., Klaassen, P., 2005. Empirical credit cycles and capital buffer formation. J. Bank. Financ. 29 (12), 3159–3179.
Kurbat, M., Korablev, I., 2002. Methodology for testing the level of the EDF™ credit measure. Technical Report, No. 020729. Moody's KMV.
Lopez, J., 2004. The empirical relationship between average asset correlation, firm probability of default, and asset size. J. Financ. Intermed. 13, 265–283.
MacDonald, I., Zucchini, W., 1997. Hidden Markov Models and Other Models for Discrete-valued Time Series. Chapman & Hall, London.
McCullagh, P., Nelder, J., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, London.
McCulloch, C., 1997. Maximum likelihood algorithms for generalized linear mixed models. J. Am. Stat. Assoc. 92, 162–170.
Merton, R., 1974. On the pricing of corporate debt: the risk structure of interest rates. J. Finance 29, 449–470.
Myles, J., Clayton, D., 2001. GLMMGibbs: An R Package for Estimating Bayesian Generalised Linear Mixed Models by Gibbs Sampling. http://www.maths.lth.se/help/R/.R/doc/html/packages.html.
Nickell, P., Perraudin, W., Varotto, S., 2000. Stability of rating transitions. J. Bank. Finance 24, 203–227.
Robert, C., Casella, G., 1999. Monte Carlo Statistical Methods. Springer, New York.
Rösch, D., 2005. An empirical comparison of default risk forecasts from alternative credit rating philosophies. Int. J. Forecast. 21, 37–51.
Skrondal, A., Rabe-Hesketh, S., 2004. Generalized Latent Variable Modeling. Chapman & Hall (CRC), Boca Raton.
Wendin, J., 2006. Bayesian Methods in Portfolio Credit Risk Management. Ph.D. thesis, Department of Mathematics, ETH Zürich.
Wilson, T., 1997. Portfolio credit risk I and II. Risk 10, 9–10.