ETH zürich

ETH Library

Generalized linear mixed models in portfolio credit risk modelling

Report

Author(s):
McNeil, Alexander J.; Wendin, Jonathan

Publication date:
2003

Permanent link:
https://doi.org/10.3929/ethz-a-004637463

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Generalized Linear Mixed Models in Portfolio Credit Risk Modelling

Alexander J. McNeil Jonathan Wendin
Departement Mathematik
Federal Institute of Technology
ETH Zentrum
CH-8092 Zürich
{mcneil,wendin}@math.ethz.ch

27 October 2003

Abstract

This paper introduces Generalized Linear Mixed Models (GLMMs), a well-known concept in statistics, to the world of portfolio credit risk modelling. A crucial point is that of dependence among default events and in the GLMM setting this is accomplished with so-called random effects. Default probabilities or default intensities are modelled as a result of fixed effects and random effects, where the latter can be used to create both dependence among defaults in the same year as well as between defaults in consecutive years. The paper gives a general introduction to GLMMs with problems relating to portfolio credit risk in mind and includes a survey of known fitting techniques and available software.

J.E.L. Subject Classification: G31, G11, C15

Keywords: Risk Management, Credit Risk, Dependence Modelling, Generalized Linear Mixed Models

1 Introduction

It is today well-accepted that default events (the inability of counter-parties to fulfill their financial obligations) 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 Figure 1, based on yearly default data for the years 1981–2000 (Standard and Poor's 2001), where years with many defaults are in general preceded and followed by years of high default intensities. A second reason to believe in dependence between defaults is that models including default dependence reproduce empirical default patterns better. Simulated output of models with independent defaults simply is not consistent with observed data. Capturing this dependence is of course 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. This makes the concept of counter-party default dependence one of the key elements of portfolio credit risk modelling.

The approach of the present paper is to partition the dependence between default events into two parts. Firstly, it seems reasonable to believe that when times are difficult, they are difficult for most companies, and this leads us to introduce dependence between defaults of different counter-parties in a given time period. For the purposes of this paper the time period will always be taken to be one year. We shall refer to this as *within-year* default dependence. Our second kind of dependence is the *between-year* default dependence. Roughly speaking, we expect recent adverse periods to exercise an influence on defaults in the near future. Dependence between consecutive years can be handled by introducing an unobservable, serially correlated "state of the economy"-process, which might even allow us to model forward-looking default frequencies.

Establishing and successfully fitting a model with the above traits is, however, easier said than done. Typically available credit data are often scarce and in general consist only of cohort sizes and default counts per credit rating group for at most 10–20 years. Nevertheless, both within-year as well as between-year default dependence can be captured with Generalized Linear Mixed Models (GLMMs), a well-known tool in statistics for modelling non-normal data such as binary and count data. The aim of the present paper is to show how these ideas can be used in credit risk with a number of concrete examples (including fitting techniques) as well as possible extensions.

In the GLMM framework default probabilities or default intensities of obligors are seen as a result of two different parts—*fixed effects* and *random effects*. Fixed effects are explanatory variables or covariates that are believed to have an impact on obligor default and are recorded on a regular basis, such as risk-free interest rate or key ratios on an obligor's balance sheet. Covariates are either quantitative (metrical or continuous) as in the case of an interest rate or qualitative, which is the case when observations are indexed by one or more classifying factors, e.g. credit rating according to a rating agency. Furthermore, covariates may be shared, regional, industry-related or obligor-specific, as in the case of ratios on a balance sheet.

Random effects, on the other hand, constitute a stochastic part of the default intensities or probabilities. A shared random effect—or *mixing variable* as we will also refer to it—can intuitively be thought of as a yearly state of the economy. If this state is identical across the portfolio all obligors face an increased default risk when the random effect is unusually high and this is the key to introducing within-year default dependence:

One usually makes the assumption that conditional on a realisation of the random effect, defaults are (conditionally) independent. But since the random effects are stochastic their influence on the joint default distribution must be integrated out, which creates within-year dependence among the responses.

In order to obtain between-year dependence, we add serial correlation to the realisations of the yearly state of the economy. This will be studied in more detail in Section 2. Models including between-year dependence can ultimately be used to forecast future default intensities. Models without this feature only allow us to make statements about the long-term stationary distribution.

The following section informally introduces a special instance of a GLMM, where default of a given company is seen as a Bernoulli-distributed event. Its general properties have been studied in more detail in Frey and McNeil (2003), although not in the language of GLMMs. For the time being we refrain from treating the question of between-year dependence.

1.1 One-factor Bernoulli mixture model

We consider the case with K credit rating classes $\{1,\dots,K\}$ with a higher number indicating a poorer credit worthiness. These could be internal groups or ratings according to some agency such as Moody's or Standard and Poor's. We denote by $\boldsymbol{m} := (m_1,\dots,m_K)'$ the size of each group and define $m = \sum_{k=1}^{K} m_k$ as the total number of obligors in our portfolio. For each $i = 1,\dots,m$ we let the function $\kappa : \{1,\dots,m\} \to \{1,\dots,K\}$ return the credit rating of obligor i and we proceed to model defaults as follows:

Introduce the random variable Y_i taking the value 1 if default of obligor i occurs during the next year and 0 otherwise. Formally, conditional on a random variable Φ (to be defined and interpreted shortly), we have

$$Y_i \,|\, \Phi \sim \mathrm{Be}(Q_i), \quad i = 1,\dots,m, \quad \text{where}$$
$$Q_i = Q_i(\eta_i) = \frac{\mathrm{e}^{\eta_i}}{1+\mathrm{e}^{\eta_i}} =: h(\eta_i) \quad \text{with} \quad \eta_i = \mu + \alpha_{\kappa(i)} + \boldsymbol{x}_i'\boldsymbol{\gamma} + \Phi.$$

The intercept μ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ are unobserved parameters, where the latter specifies the level of a factor describing the additional risk associated with the different rating groups compared to a baseline case.[1] $\boldsymbol{x}_i$ is a d-dimensional vector of covariates or explanatory variables and the unknown d-vector $\boldsymbol{\gamma}$ assigns appropriate weights to the covariates' impact on the default probability. The covariates chosen are global and/or obligor-specific and quantitative and/or qualitative as described in the introduction. Note that group affiliation is of course also a classifying fixed effect, although we do not include it in $\boldsymbol{x}_i$. Group affiliation as well as covariates $\boldsymbol{x}_i$ conclude the *fixed effects'* part of the Bernoulli model.

The *random effect* is given by the term Φ following a distribution of our choice. In this example we set $\Phi \sim \mathrm{N}(0, \sigma^2)$. The fixed and random effects together form η_i, which is modelled on the whole real line. The response function h maps $\mathbb{R}$ to an admissible value for the success probability of a Bernoulli trial, i.e. $(0, 1)$. The choice $h(x) = \mathrm{e}^x/(1+\mathrm{e}^x)$ together with a Gaussian random effect is known as the *logit model*. Another common choice for h is the cumulative standard normal distribution function $\Phi(x)$, yielding the *probit model*.

We assume that all default indicators $Y_1, \ldots, Y_m$ are conditionally independent (given Φ), thus for any $\boldsymbol{y} \in \{0, 1\}^m$

$$P(\boldsymbol{Y} = \boldsymbol{y} \,|\, \Phi) = \prod_{i=1}^{m} \left(Q_i(\mu, \boldsymbol{\alpha}, \boldsymbol{x}_i'\boldsymbol{\gamma}, \Phi)\right)^{y_i} \left(1 - Q_i(\mu, \boldsymbol{\alpha}, \boldsymbol{x}_i'\boldsymbol{\gamma}, \Phi)\right)^{1-y_i}.$$

If the value of Φ is very high, this means an increased default probability for all obligors in our portfolio, irrespective of group. The realisation of the random effect can therefore be regarded as the general state of the economy in the year under consideration. We obtain the unconditional distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_m)'$ by integrating out the distribution G of Φ

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \int_{\mathbb{R}} P(\boldsymbol{Y} = \boldsymbol{y} \,|\, \Phi = \phi) \, \mathrm{d}G(\phi) \tag{1}$$

and thus create within-year dependence among the components of $\boldsymbol{Y}$. In general, the above probability law of $\boldsymbol{Y}$ is not given analytically and this makes model fitting, i.e. estimation of the fixed effect parameters μ, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ as well as the random effect parameter σ, rather cumbersome. This in turn requires repeated, yearly observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ of $\boldsymbol{Y}$.[2]

If the we leave out the individual obligor covariates $\boldsymbol{x}_i$ and let the default probability be governed only by μ, $\boldsymbol{\alpha}$ and the random effect Φ, the number of defaults, M_k, in group k is

$$M_k \,|\, \Phi \sim \mathrm{Bin}(m_k, Q_k) \quad \text{with}$$
$$Q_k = h(\eta_k) = \frac{\mathrm{e}^{\eta_k}}{1 + \mathrm{e}^{\eta_k}}, \quad \text{where} \quad \eta_k = \mu + \alpha_k + \Phi.$$

Conditional on Φ the number of defaults in each rating class are (conditionally) independent binomial rvs:

$$P\left(\boldsymbol{M} = \boldsymbol{x} \,|\, \Phi\right) = \prod_{k=1}^{K} \binom{m_k}{x_k} \left(Q_k(\mu, \alpha_k, \Phi)\right)^{x_k} \left(1 - Q_k(\mu, \alpha_k, \Phi)\right)^{m_k - x_k},$$

where $\boldsymbol{M} := (M_1, \ldots, M_K)'$ denotes the number of defaults in each group respectively. The unconditional distribution of $\boldsymbol{M}$ is obtained by again integrating out the random effect Φ. In this case, all obligors in a given group are modelled *exchangeably*, i.e. given group affiliation the obligors cannot be distinguished, which may well be a realistic assumption.

Again, the model can only be fitted given a set of yearly data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ or $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_n$ along with cohort sizes. As soon as the parameters μ, $\boldsymbol{\alpha}$ and σ have been estimated,

[1] If group 1 is our benchmark case, we constrain $\alpha_1 \equiv 0$.

[2] The observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ may be of differing lengths depending on the number of obligors under study in each year.

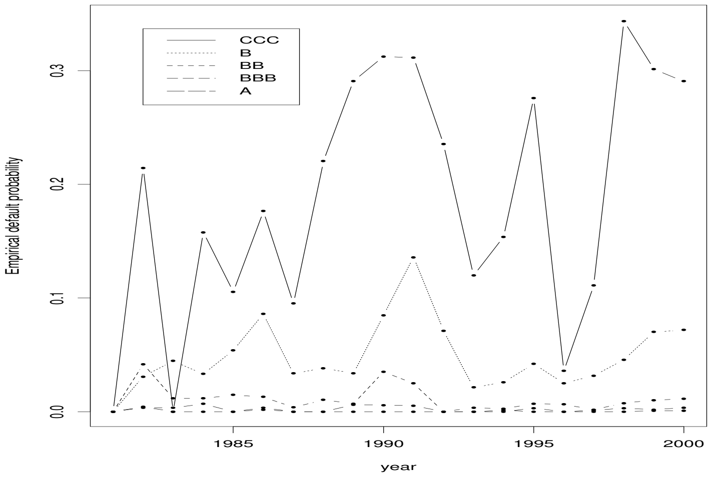

Figure 1: Empirical default probabilities according to Standard and Poor's (2001) for 20 years. Note the cyclic variation in default occurrences.

performing a Monte Carlo simulation of default data is easily accomplished. Simply generate $\Phi \sim \mathrm{N}(0, \sigma^2)$, calculate $Q_k = Q_k(\mu, \alpha_k, \Phi)$ as earlier on and finally simulate M_k as binomial with parameters m_k and Q_k for all groups $k = 1, \ldots, K$. Repeat this procedure the required number of times.

The above model with probit link ($h(x) = \Phi(x)$) is in effect a one-factor version of the well-known industry model CreditMetrics, where the factor is taken as the "yearly state of the economy" Φ. The different default probabilities for different groups are determined by the different values of $\mu + \alpha_k$. Also note that so far, we have refrained from introducing between-year dependence. This will be addressed in Example 2.3.

2 Generalized Linear Mixed Models

Generalized Linear Mixed Models (GLMMs) combine the idea of dependence through mixing variables with concepts from the world of linear modelling in order to include non-normal data such as binary and count data. Their strength is the potential of dealing with data involving multiple sources of random error, such as repeated measurements within subjects or *units*. Typical examples of units are students within a class room or hospitals within a region. When performing repeated measurements on the same unit, we expect the responses to be dependent. This is, as we shall see, accomplished with the mixing variable or random effect. In the present applications, we treat default or non-default of obligors in a given year as repeated measurements within that year, so that year is effectively the collecting unit; in this way we generate within-year dependence.

In what follows we study responses y_{ti} and covariates or explanatory variables $\boldsymbol{x}_{ti}$ for counterparty $i = 1, \ldots, m_t$ in year $t = 1, \ldots, n$. We define

$$(\boldsymbol{x}_t, \boldsymbol{y}_t) = (\boldsymbol{x}_{t1}, \ldots, \boldsymbol{x}_{tm_t}, y_{t1}, \ldots, y_{tm_t})$$

and interpret this as data on year t.

Given a random effect $\boldsymbol{b}_t$ of an arbitrary dimension p and covariates $\boldsymbol{x}_t$ we assume that the conditional density of the responses $\boldsymbol{y}_t$ belongs to the exponential family, such as Bernoulli or Poisson, with conditional mean

$$\mathbb{E}(Y_{ti} \,|\, \boldsymbol{b}_t) = h(\eta_{ti}), \quad \text{where} \quad \eta_{ti} = \boldsymbol{X}'_{ti}\boldsymbol{\beta} + \boldsymbol{Z}'_{ti}\boldsymbol{b}_t \quad \text{for individuals } \textbf{or} \tag{2a}$$

$$\mathbb{E}(\boldsymbol{Y}_t \,|\, \boldsymbol{b}_t) = h(\boldsymbol{\eta}_t) \quad \text{with} \quad \boldsymbol{\eta}_t = X_t\boldsymbol{\beta} + Z_t\boldsymbol{b}_t \quad \text{in vector form for whole units} \tag{2b}$$

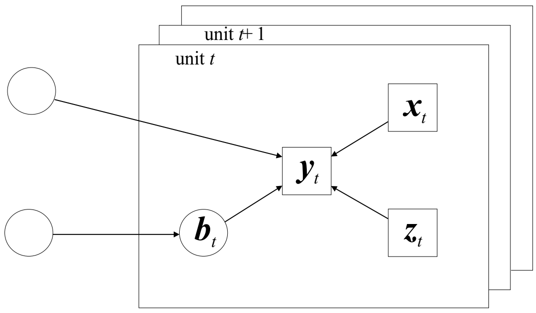

Figure 2: Hierarchical representation of the GLMM described in Equation 2. Observed quantities are in boxes and unobserved in circles. Between-year dependence is accomplished with dependence among the sequence of $\boldsymbol{b}_t$'s.

for $i = 1, \ldots, m_t$ and $t = 1, \ldots, n$. Eqn. 2 stresses the dependence of η_{ti} on both fixed effects and random effects. The random effect $\boldsymbol{b}_t$ is drawn from a distribution, parameters of which are called *hyperparameters* θ. Its contribution to the components of $\boldsymbol{\eta}_t$ is given by the design matrix Z_t. For the purposes of this paper it suffices to think of $\boldsymbol{b}_t$ as a scalar variate (and hence set $Z_{ti} = 1$ for all i and t). In more sophisticated models, a vector-valued random effect $\boldsymbol{b}_t$ could have each component interpreted as the general state of the economy according to industry sector and/or geographical location. The components of $\boldsymbol{b}_t$ would then typically be strongly correlated and the (observable) design vector $\boldsymbol{Z}_{ti}$ holds the corresponding (possibly weighted) exposures of obligor i.

The fixed effects are summarized in the design matrix X_t and may consist of covariates $\boldsymbol{x}_t$ and possibly past responses. As mentioned in the introduction, covariates may be shared or specific for each measurement. The picture is kept general by allowing both quantitative (metrical) and qualitative (categorical) covariates, where the latter need not be ordered. The impact of the fixed effects on η_{ti} is given by the unobserved parameter vector $\boldsymbol{\beta}$.

The function h is known as *response function* and serves as a link between $\mathbb{R}$, on which we model η_{ti}, and intervals of relevance for count data, e.g. $(0, \infty)$ for the intensity of Poisson data or $(0, 1)$ for the success probability of binary data. Usual requirements are that h is smooth and one-to-one. If no random effects are present a model like (2) is called a Generalized Linear Model (GLM), the theory of which is studied extensively in McCullagh and Nelder (1989) as are the concepts of exponential families and response functions.[3]

Conditional on the random effect $\boldsymbol{b}_t$, the responses $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tm_t})$ on unit t are treated as independent. The unconditional joint distribution of $\boldsymbol{y}_t$, however, is obtained by integrating out the effect of $\boldsymbol{b}_t$ and thus creates dependence among the responses $\boldsymbol{y}_t$ on unit t, cf. Section 1.1. In a credit risk context, where different years are treated as different units, this means that we have within-year dependence. The hierarchical structure of the GLMM becomes more evident if we interpret it graphically as in Figure 2.

A word on notation: one sometimes comes across GLMMs of the form

$$\boldsymbol{\eta}_t = X_t \boldsymbol{\beta} + Z_t \boldsymbol{b}_t + \boldsymbol{e}_t,$$

where $(\boldsymbol{e}_t)_{t=1}^n$ are termed (residual) random effects and usually are iid zero-mean Gaussian. These models can always be written in the form (2) by forming a random effects' vector $(\boldsymbol{b}_t, \boldsymbol{e}_t)$ and modifying the design matrix Z_t accordingly. Including $\boldsymbol{e}_t$ as a residual random effect may, however, have instructive advantages, especially if the random effect $\boldsymbol{b}_t$ is unit-specific or non-Gaussian.

[3]For GLMs it is customary to leave the index i out, since we in general consider only single and not repeated measurements on subject t. This follows since the dependence between repeated measurements vanishes if the random effect does.

So far, we have refrained from introducing the second kind of dependence; namely the (spatial) dependence among units. In the case of hospitals in a region we might expect spatial dependence according to geographical location. Likewise, as argued in the introduction of this paper, neighboring years seem to display dependent default patterns—between-year dependence. This is handled in an intuitively clear way in the GLMM framework. If realised random effects $\boldsymbol{b}_{t_1}, \boldsymbol{b}_{t_2}$ for different years t_1, t_2 are independent, then so are the responses $\boldsymbol{y}_{t_1}$ and $\boldsymbol{y}_{t_2}$. By incorporating suitable dependence among $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ we treat the problem of spatial or between-year dependence. For credit risk purposes, the most appealing dependence structure on $(\boldsymbol{b}_t)$ is that of a time series, which we illustrate in Example 2.3. For a brief overview of how spatial dependence is handled in other contexts, see Clayton (1996).

2.1 GLMMs in credit risk modelling

The ability of GLMMs to capture both within-year as well as between-year default dependence make them interesting for credit risk purposes. The following section aims at showing what kind of GLMMs are relevant for credit risk modelling and also how between-year dependence can be successfully incorporated. The topic of fitting these models, a task that naturally grows with model complexity, is saved until Section 3.

Example 2.1 (Section 1.1 as a GLMM). The Bernoulli model outlined in Section 1.1 fits into the GLMM framework as we now shall see. Assume we have n binary response vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ of size $m_1, \ldots, m_n$, respectively, describing the survival ($y_{ti} = 0$) or default ($y_{ti} = 1$) of obligor $i = 1, \ldots, m_t$ in year $t = 1, \ldots, n$. We interpret these as realisations of the rv $\boldsymbol{Y}$ specified in Eqn. 1. For each year t we also have a vector $(\boldsymbol{x}_{t1}, \ldots, \boldsymbol{x}_{tm_t})$ of d-dimensional covariates for obligor $i = 1, \ldots, m_t$. In our present analysis we make no distinction between new obligors and obligors that have survived through previous periods, unless follow-up time is included as a covariate.

Before identifying fixed and random effects we choose the response function h

$$E(Y_{ti} \,|\, \boldsymbol{b}) = \frac{\mathrm{e}^{\eta_{ti}}}{1 + \mathrm{e}^{\eta_{ti}}} =: h(\eta_{ti})$$

and collect all unknown parameters in the $(K + d)$-dimensional parameter vector $\boldsymbol{\beta}$:[4]

$$\boldsymbol{\beta} = (\mu + \alpha_1, \ldots, \mu + \alpha_K, \gamma_1, \ldots, \gamma_d)'.$$

We specify the design matrices X_t and Z_t in order to obtain a model of the form (2)

$$\eta_{ti} = \mu + \alpha_{\kappa(i)} + \boldsymbol{x}_{ti}'\boldsymbol{\gamma} + \Phi_t \equiv (X_t\boldsymbol{\beta} + Z_t\boldsymbol{b}_t)_i, \quad i = 1, \ldots, m_t, \quad t = 1, \ldots, n.$$

X_t has m_t rows (one for each obligor) and these are filled in turn as follows: the first K columns all have a zero-element except column $\kappa(i)$ which is 1 (recall that κ picks out the group affiliation and thus gives us the correct $\mu + \alpha_k$ for obligor i). The remaining d columns are filled with the covariates $\boldsymbol{x}_{ti}$. Note that X_t is indeed year-dependent, as the covariates are. This concludes the specification of the fixed effects.

The random effect is univariate and shared for all obligors in year t, which yields $Z_t = \mathbf{1}$ (vector of ones) and $\boldsymbol{b}_t = \Phi_t$, where $\Phi_t \sim \mathrm{N}(0, \sigma^2)$, $t = 1, \ldots, n$. Consequently, σ is our only hyperparameter. If the number of obligors in our portfolio varies from year to year, so will the size of our design matrices X_t ($m_t \times [K + d]$) and Z_t ($m_t \times 1$).

As mentioned in Section 1.1, $h(x) = \mathrm{e}^x/(1 + \mathrm{e}^x)$ (logit-model) is by no means the only possible choice of response function h for the Gaussian random effect Φ. Any smooth and increasing mapping $\mathbb{R} \to (0, 1)$ will do the job, such as a distribution function on $\mathbb{R}$. In

[4] Recall that the effective number of unknowns is $K + d$ and not $K + d + 1$ since the benchmark case k is constrained to have $\alpha_k \equiv 0$ as pointed out in Section 1.1.

particular, $h(x) = \Phi(x)$ (cumulative standard normal df) is a common choice called the probit-model. For more examples to this end, see Chapter 7 of Joe (1997).

A model very similar to that considered in Example 2.1 makes use of the Poisson distribution instead of the Bernoulli or binomial distributions. This well-known approximation works surprisingly well when the individual default probability of an obligor is very low, which is the case for obligors of high credit worthiness, and it may lead to considerable computational advantages when fitting the model to data with the Gibbs sampler. Although such models do not rule out the possibility of obligors defaulting more than once, the probability of this happening is usually very small if properly calibrated. Our next example illustrates these ideas for the case of K rating groups, whose respective members are indistinguishable.

Example 2.2 (One-factor Poisson GLMM). In the Poisson case we assume that given our mixing variable b_t, the number of defaults in group k in year t, M_{tk}, is

$$M_{tk}\,|\,b_t \sim \mathrm{Pois}(m_{tk}\Lambda_{tk}) \quad \text{for } k = 1,\dots,K \text{ and } t = 1,\dots,n \quad \text{where}$$
$$\Lambda_{tk} = h(\eta_{tk}) = \mathrm{e}^{\eta_{tk}} \quad \text{with} \quad \eta_{tk} = \mu + \alpha_k + b_t$$

and m_{tk} denotes the size of cohort k in year t. We implicitly assume that the number of observations in each year (i.e. number of rating classes) is the same for all n years. In this example the response function $h(x) = \mathrm{e}^x$ takes $\mathbb{R}$ to $(0,\infty)$—a typical range for the intensity of a Poisson distribution. Again, the mixing variable b_t is used for representing the overall economic state during year t. The setup is analogous to that of Section 1.1 and with conditional independence we obtain

$$P(\boldsymbol{M}_t = \boldsymbol{x}\,|\,b_t) = \prod_{k=1}^{K} \frac{1}{x_k!} \exp\left\{-m_k \mathrm{e}^{\mu+\alpha_k+b_t}\right\} \left(m_k \mathrm{e}^{\mu+\alpha_k+b_t}\right)^{x_k}$$

for any $\boldsymbol{x} \in \mathbb{N}_0^k$. The unconditional distribution of $\boldsymbol{M}_t$ is obtained by integrating out the effect of b_t as in Section 1.1. The design matrices X and Z are determined in analogy with Example 2.1. In this case we have $\log(m_{tk}\Lambda_{tk}) = \log(m_{tk}) + \log(\Lambda_{tk}) = \log(m_{tk}) + \eta_{tk}$ yielding an *off-set* $\log(m_{tk})$; a regressor whose coefficient is fixed to unity. The above model with $b_t \sim \mathrm{N}(0,\sigma^2)$ can be seen as an instance of CreditRisk$^+$ with a logistic factor.

Example 2.3 (Autoregressive random effect). In this example we abandon the assumption of the realisations $b_1,\dots,b_n$ of the random effects being independent. We denote by $\boldsymbol{m}_t = (m_{t1},\dots,m_{tK})'$ and $\boldsymbol{M}_t = (M_{t1},\dots,M_{tK})'$ the number of obligors and defaults, respectively, for each rating class $1,\dots,K$ in year t. In particular, we assume

$$M_{tk}\,|\,b_t \sim \mathrm{Bin}\left(m_{tk}, \mathrm{e}^{\mu_k+b_t}/(1+\mathrm{e}^{\mu_k+b_t})\right) \quad \text{for} \quad k=1,\dots,K;\; t=1,\dots,n.$$

We choose the random effects process univariate Gaussian satisfying

$$b_1 \sim \mathrm{N}\left(0, \tfrac{\sigma^2}{1-\alpha^2}\right), \quad b_t\,|\,b_{t-1} \sim \mathrm{N}\left(\alpha b_{t-1}, \sigma^2\right) \quad \text{for } t=2,\dots,n. \tag{3}$$

Consequently, our fixed effect parameters are $\boldsymbol{\mu} = (\mu_1,\dots,\mu_K)'$[5] and hyperparameters $\theta = (\sigma,\alpha)'$. Eqn. 3 is a first-order autoregressive time series, AR(1). It easily follows that $\mathrm{Cov}(b_s,b_t) = \sigma^2\alpha^{|s-t|}/(1-\alpha^2)$, i.e. decreases geometrically with the time distance $|s-t|$. This means that the sequence $\boldsymbol{M}_1,\dots,\boldsymbol{M}_n$ of defaults for years $1,\dots,n$ shows between-year dependence of an autoregressive kind. In Example 4.2 we will use the Gibbs sampler, introduced in the next section, to fit this model to data for the case $K = 1$. Note that random effects of this form can be used also for Poisson default counts of the type described in Example 2.2.

[5] We interpret μ_k as what in Section 1.1 is called $\mu + \alpha_k$.

2.2 Default probabilities and correlations

Specification of models for default counts such as those seen in the previous section introduces implied obligor default probabilities and default correlations, quantities which are of interest to evaluate given model parameters $\boldsymbol{\beta}$ and θ. In many cases, these are merely moments or other expectations of the random effect b_t, as seen e.g. in Frey and McNeil (2003).

2.2.1 Binomial models

The conditional default probability $\pi_{ti} \,|\, \boldsymbol{b}_t$ of obligor i in year t for binomial data is given by applying the response function h to $\eta_{ti} = \boldsymbol{X}'_{ti}\boldsymbol{\beta} + \boldsymbol{Z}'_{ti}\boldsymbol{b}_t$, cf. Equation 2. By taking expectations we easily obtain the unconditional default probability $\pi_{ti} = \mathbb{E}h(\eta_{ti})$.

Default correlations between obligors are obtained similarly, by calculating the covariance between the default indicators Y_{i_1}, Y_{i_2} for two obligors i_1, i_2. An important special case is that of K homogenous rating groups, where all members of a group are modelled exchangeably. Let π_{t,k_1k_2} be the covariance between two arbitrary default indicators in groups k_1 and k_2 in year t. Reasoning as above, we for binomial models have $\pi_{t,k_1k_2} = \mathbb{E}\left[h(\eta_{tk_1})h(\eta_{tk_2})\right]$, where $\eta_{tk_i} = \boldsymbol{X}'_{tk_i}\boldsymbol{\beta} + \boldsymbol{Z}'_{tk_i}\boldsymbol{b}_t$. Applying the usual norming returns the default correlation.

2.2.2 Poisson models

For Poisson models, the conditional default intensity $\pi_{ti}|\boldsymbol{b}_t$ of obligor i in year t is given by applying $h : \mathbb{R} \to (0, \infty)$ to $\eta_{ti} = \boldsymbol{X}'_{ti}\boldsymbol{\beta} + \boldsymbol{Z}'_{ti}\boldsymbol{b}_t$ and the unconditional version is obtained by taking expectations $\mathbb{E}h(\eta_{ti})$ in analogy with the previous subsection. The individual obligor default *probability* π_{ti}, on the other hand, is found as follows: Let S_{ti} be the number of defaults of a *single* obligor i in year t. Clearly, $S_{ti} \,|\, \boldsymbol{b}_t \sim \text{Pois}(h(\eta_{ti}))$ and consequently

$$
\begin{aligned}
\pi_{ti} &= 1 - P(S_{ti} = 0) = \mathbb{E}\left(1 - P(S_{ti} = 0 \,|\, \boldsymbol{b}_t)\right) \\
&= \mathbb{E}\left(1 - \mathrm{e}^{-h(\eta_{ti})}\right) \approx \mathbb{E}h(\eta_{ti}).
\end{aligned}
$$

This expectation can always be calculated, at least numerically, as soon as the parameters $\boldsymbol{\beta}$ and θ as well as covariates $\boldsymbol{x}$ have been determined. The covariance π_{t,k_1k_2} between two default indicators of rating k_1 and k_2 in year t (where obligors of a given group are indistinguishable) is found analagously:

$$
\pi_{t,k_1k_2} = \mathbb{E}\big((1 - \mathrm{e}^{-h(\eta_{tk_1})})(1 - \mathrm{e}^{-h(\eta_{tk_2})})\big)
$$

and can be used to find the default correlation between defaults of rating group k_1 and k_2. This expectation is a little more involved than that of the binomial model. One way to circumvent an analytical treatment of tricky expectations is given by the Gibbs sampler, which will be introduced in the next section on model fitting.

3 Fitting GLMMs

The examples of the previous section show that GLMMs are able to handle a wide range of binary and count data problems where intra-subject (within-year) and spatial (between-year) dependence are of importance. The only limitation when establishing a model is of course the possibility of later on fitting it to data. For GLMMs (or hierarchical models in general) the unconditional distribution of the responses is obtained by integrating out the effect of the random effects—an operation that is hard to carry out in practice and makes naive Maximum likelihood-fitting (ML) difficult. However, the Bayesian approach combined with Markov Chain Monte Carlo turns out to be particularly fruitful even for rather complex models. This is the most general tool for fitting GLMMs and will be the topic of Section 3.2.

3.1 Maximum-likelihood estimation

Recalling the notation of Section 2, the unconditional density or mass function f of response $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tm_t})$ takes the form

$$f(\boldsymbol{y}_t \,|\, \boldsymbol{\beta}, \theta, \boldsymbol{x}_t) = \int_{\mathbb{R}^p} \left(\prod_{i=1}^{m_t} P\left(Y_{ti} = y_{ti} \,|\, \boldsymbol{b}_t, \boldsymbol{x}_t, \boldsymbol{\beta}\right) f_{b_t}(\boldsymbol{b}_t \,|\, \theta) \right) \mathrm{d}\boldsymbol{b}_t, \tag{4}$$

where $p = \dim(\boldsymbol{b}_t)$ and f_{b_t} is the density of $\boldsymbol{b}_t$. This expression seldom has a closed form and must be evaluated numerically, which complicates ML-fitting. If the random effects $(\boldsymbol{b}_t)$ are mutually independent our sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ is a realization of independent f-distributed rvs and the likelihood function L is

$$L(\boldsymbol{\beta}, \theta \,|\, D) = \prod_{t=1}^{n} f(\boldsymbol{y}_t \,|\, \boldsymbol{\beta}, \theta, \boldsymbol{x}_t)$$

as expected. The independence among different subjects is the key to the factorization that nevertheless lets us perform the maximization, albeit numerically.

In order to capture the between-year dependence, the random effects $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ can no longer be independent. This makes the likelihood function even more involved. Conditional independence properties (given $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$) give the following likelihood:

$$L(\boldsymbol{\beta}, \theta \,|\, D) = \int \cdots \int \prod_{t=1}^{n} \prod_{i=1}^{m_t} P(Y_{ti} = y_{ti} \,|\, \boldsymbol{b}_t, \boldsymbol{\beta}, \boldsymbol{x}_t) f_b(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n \,|\, \theta)\, \mathrm{d}\boldsymbol{b}_1 \cdots \mathrm{d}\boldsymbol{b}_n,$$

If $(\boldsymbol{b}_t)$ is a sequence of independent random effects this expression clearly reduces to (4), but in the general case this is an integral over $\mathbb{R}^{n \times p}$. High-dimensional integration is difficult to master numerically and results are often inaccurate.

In this context the *Expectation-Maximization* (EM) algorithm is also of interest. In this way the difficult maximization is replaced by a sequence of easier maximization problems whose limit is the answer to our original problem. This is accomplished by augmenting the observed data $(\boldsymbol{x}_t, \boldsymbol{y}_t)_{t=1}^n$ by the unobserved random effects $(\boldsymbol{b}_t)$ and then performing successive expectation and maximization steps. See Robert and Casella (1999) for a general introduction and McCulloch (1997) for ideas on how these methods can be used in GLMMs.

3.1.1 Software for GLMM-fitting with ML

There are a number of existing software packages for GLMMs based on ML-methodology, for instance `glmmPQL` and `glmmML` which are a part of `MASS` for R. In Frey and McNeil (2003) a model very similar to that of the binomial model in Section 1.1 is fitted by maximum likelihood. Even for such simple models, the integrals in the likelihood function soon become very cumbersome.

3.2 Bayesian estimation

In classical statistics we regard parameters $\boldsymbol{\vartheta} = (\vartheta^{(1)}, \ldots, \vartheta^{(\nu)})'$ as fixed but unknown constants and our sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ is seen as a realization of a random variable indexed by $\boldsymbol{\vartheta}$. Based upon the observed values our knowledge about $\boldsymbol{\vartheta}$ is obtained. In the Bayesian setting the main distinction is instead made between known (i.e. observed) quantities D and unknown (i.e. unobserved) quantities $\boldsymbol{\vartheta}$. Unknowns $\boldsymbol{\vartheta}$ are quantities whose variation can be described by a probability distribution, comprising any information on $\boldsymbol{\vartheta}$ we may have before observing the data D. This distribution is called the *prior distribution*. The quantity of interest in the Bayesian framework is the *posterior distribution*—the law of unknowns $\boldsymbol{\vartheta}$

after having observed data. The posterior is calculated from the prior with *Bayes' rule* and inference on parameters is based on this distribution:

$$\begin{aligned} p(\boldsymbol{\vartheta} \,|\, D) &= \frac{p(\boldsymbol{\vartheta}, D)}{p(D)} = \frac{p(D \,|\, \boldsymbol{\vartheta})\, p(\boldsymbol{\vartheta})}{p(D)} \\ &= \frac{p(D \,|\, \boldsymbol{\vartheta})\, p(\boldsymbol{\vartheta})}{\int p(D \,|\, \boldsymbol{\vartheta})\, p(\boldsymbol{\vartheta})\, \mathrm{d}\boldsymbol{\vartheta}}. \end{aligned}$$

In our case $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \boldsymbol{b}, \theta)'$ contains the fixed effect regressors $\boldsymbol{\beta}$, the random effects $\boldsymbol{b}$ and the hyperparameters θ while D holds the responses and covariates $\{(\boldsymbol{y}_t, \boldsymbol{x}_t)\}_{t=1}^{n}$. This is depicted graphically in Figure 2, where unknowns are circled and observed quantities boxed, yielding the joint probability

$$p(\boldsymbol{\beta}, \boldsymbol{b}, \theta) = p(\boldsymbol{b} \,|\, \theta)\, p(\theta)\, p(\boldsymbol{\beta}). \tag{5}$$

The fact that $\boldsymbol{b}$, an inner random node, is treated the same as the parameters $\boldsymbol{\beta}$ and θ in the Bayesian setting may be a little confusing at first sight. For proper classical and Bayesian interpretations of GLMMs in this context, see the slightly technical text by Clayton (1996).

Simulating from posterior distributions is in general difficult. Firstly, the normalizing factor $p(D)$ is most likely out of reach by analytical means, meaning that the posterior distribution is known only up to a multiplicative constant. Secondly, only in very well-behaved cases will the posterior distribution, which is (at least in our case) multivariate, resemble a known distribution. This is for instance the case for low-dimensional problems where so called *conjugate priors* yield tractable posteriors. Over the years this has forced Bayesian statistics to deal exclusively with low-dimensional problems and more or less dubious priors.

Nevertheless, with recent techniques in computational statistics this is changing. Markov Chain Monte Carlo techniques such as the Gibbs sampler allow us to simulate from the posterior knowing only the functional form of the density and also to split annoying multivariate simulations into low-dimensional or even univariate simulation.

3.2.1 The Gibbs sampler

Assume we want to simulate data, possibly multivariate, from a distribution $p(x)$. If p is difficult to simulate from, we may construct an ergodic Markov chain with p as its stationary distribution and regard a sample from the Markov chain, possibly after a certain *burn-in*, as a sample from p. Constructing such a Markov chain is in fact surprisingly easy and methods of this kind are known as Markov Chain Monte Carlo (MCMC). The Gibbs sampler, which we are about to introduce, is a special instance of an MCMC algorithm.

Assume we have a vector $\boldsymbol{\vartheta} = (\vartheta^{(1)}, \ldots, \vartheta^{(\nu)})'$ of unknowns (parameters or unobserved latent random quantities like random effects). We denote by $p(\boldsymbol{\vartheta}|D)$ the posterior distribution of $\boldsymbol{\vartheta}$ given observed data D. The basic scheme of the Gibbs sampler for simulating from $p(\boldsymbol{\vartheta} \,|\, D)$ is as follows:

Step 0 Fix an arbitrary starting value $\boldsymbol{\vartheta}_0 = (\vartheta_0^{(1)}, \ldots, \vartheta_0^{(\nu)})'$ and set $i = 0$.

Step 1 Generate $\boldsymbol{\vartheta}_{i+1} = (\vartheta_{i+1}^{(1)}, \ldots, \vartheta_{i+1}^{(\nu)})'$ as follows:

- $\vartheta_{i+1}^{(1)} \sim p(\vartheta^{(1)} \,|\, \vartheta_i^{(2)}, \ldots, \vartheta_i^{(\nu)}, D)$
- $\vartheta_{i+1}^{(2)} \sim p(\vartheta^{(2)} \,|\, \vartheta_{i+1}^{(1)}, \vartheta_i^{(3)}, \ldots, \vartheta_i^{(\nu)}, D)$
  ...
- $\vartheta_{i+1}^{(\nu)} \sim p(\vartheta^{(\nu)} \,|\, \vartheta_{i+1}^{(1)}, \ldots, \vartheta_{i+1}^{(\nu-1)}, D)$

Step 2 Set $i = i + 1$ and go to Step 1.

It can be shown that under general regularity conditions the vector sequence $\{\boldsymbol{\vartheta}_i; i = 1, 2, \dots\}$ has stationary distribution $p(\boldsymbol{\vartheta}|D)$. The distributions in Step 1 are called *full conditional* distributions and by choosing the priors of $\vartheta^{(1)}, \dots, \vartheta^{(\nu)}$ appropriately we may obtain full conditionals following an easy-to-simulate-from distribution. Although this speeds up the simulation, well-known full conditionals are by no means a requirement and even in not so complex models they often do turn out to be non-standard, as we shall see in Example 4.2.

A very general remedy to this problem is including a so called Metropolis-Hastings' step. This way the non-standard full conditional is generated from a known proposal distribution of our choice and then accepted only with a certain probability. In fact, the standard Gibbs sampler is a special case of the Metropolis-Hastings algorithm where the acceptance probability is constantly 1. The Gibbs sampler and more general Metropolis-Hastings algorithms are discussed extensively in e.g. Robert and Casella (1999) and Ibrahim et al. (2001). Also, surprisingly often the full conditional has a log-concave density, which allows us to use the Adaptive Rejection Sampling (ARS) (Gilks 1992) to carry through our simulation.

A nice feature of the Gibbs sampler is that we in each iteration step may make inference on all unobserved nodes of our model; in the GLMM case $\boldsymbol{\vartheta} = (\boldsymbol{\beta}, \boldsymbol{b}, \theta)'$. Inference on the random effect $\boldsymbol{b}$ is of immediate interest in the context of obligor default probabilities and correlations, since these are merely moments or functions of $\boldsymbol{b}$ and parameters (which at times are tedious to handle analytically) as argued in Section 2.2. Consequently, inference on these quantities is generated at no extra cost when running the Gibbs sampler.

3.2.2 Software for Bayesian GLMM-fitting

The software package `GLMMGibbs` for R by Myles and Clayton can be downloaded on the Internet and allows fitting of binomial and Poisson GLMMs. It is based on ARS sampling and has some flexibility to handle correlated random effects, mainly for neighboring regions. Numerical instabilities may occur for the binomial family if the typical success probability is low, which may cause the ARS sampling to break down. In these cases the Poisson approximation might be worth considering as will be illustrated in Example 4.1, where `GLMMGibbs` fits a five-group model to the yearly default data of Figure 1. For a more detailed account of modelling possibilities as well as built-in distributional assumptions on priors in `GLMMGibbs`, see Myles and Clayton (2001). There is also `BUGS` for more general Gibbs sampling problems.

4 Practical implementations

The following section aims at showing how the GLMM ideas can be used in practice. We give one example of a five-group model with Poisson default counts using existing software and one example of a one-group model with autoregressive random effects based on customized code. The data used in the examples is yearly cohort sizes and default counts for the years 1981 to 2000 for rating groups A, BBB, BB, B and CCC according to Standard and Poor's (2001), see Figure 1.

Example 4.1 (Standard and Poor's Data). We use `GLMMGibbs` to fit the five-group Poisson GLMM of Example 2.2 using the above Standard and Poor's data. The random effect Φ is iid $\mathrm{N}(0, \sigma^2)$ for all years and groups are distinguished by their different $\alpha_{\mathrm{A}}, \dots, \alpha_{\mathrm{CCC}}$. The default probability π_k of an obligor in group k is by Section 2.2 approximately $\mathbb{E}\Lambda_k = \exp\{\mu + \alpha_k + \sigma^2\}$ and could of course be calculated exactly with the moment-generating function of a log-normal rv. Parameter estimates as well as in-group default probabilities can be found in Table 1. In Frey and McNeil (2003) the very same default data are fitted by maximum likelihood using a binomial model, yielding virtually the very same default probabilities, as seen in Table 1. Since the Poisson distribution is at its best for low-

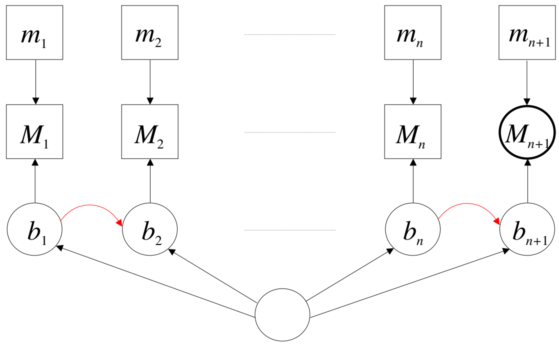

Figure 3: Hierarchical representation of the one-group model of Section 4.1 with a autoregressive time series structure on the random effects $(b_t)_{t=1}^n$. Rectangular nodes are observed and circular nodes unobserved. The nodes at time $n+1$ enter in Section 4.1.1 when predicting future defaults.

probability default groups the binomial ML-default probabilities differ slightly for groups B and CCC.

Parameter	A	BBB	BB	B	CCC
$\mu + \alpha_k$ (mean)	-7.84	-6.13	-4.64	-2.94	-1.53
$\mu + \alpha_k$ (median)	-7.79	-6.11	-4.62	-2.93	-1.52
s.e.$(\mu + \alpha_k)$	0.429	0.477	0.443	0.433	0.436
σ	(mean)	0.158		(median)	0.0633
s.e.(σ)	0.380				
π_k (mean)	0.00040	0.00223	0.0099	0.0542	0.2220
π_k (median)	0.00042	0.00222	0.0099	0.0536	0.1986
π_k (ML)	0.0004	0.00222	0.0098	0.0503	0.2066

Table 1: Parameter estimates and standard errors for a one-factor Bernoulli mixture model fitted to historical Standard and Poor's one-year default data, together with the implied estimates of default probabilities π_k for rating groups $k = \mathrm{A}, \ldots, \mathrm{CCC}$. The values π_k (ML) are from Frey and McNeil (2003).

4.1 Autoregressive random effects with Gibbs sampling

This section shows how the Gibbs sampler can be used to fit a one-group model with autoregressive serial dependence on the random effects process (b_t) as that of Example 2.3. Given b_t the number of defaults in year t, M_t, is

$$M_t \,|\, b_t \sim \mathrm{Bin}\left(m_t, \mathrm{e}^{b_t}/(1+\mathrm{e}^{b_t})\right), \quad t = 1, \ldots, n,$$

where m_t is the total number of obligors in year t. By $\boldsymbol{m}$ and $\boldsymbol{M}$ we shall mean the observed vectors $(m_1, \ldots, m_n)'$ and $(M_1, \ldots, M_n)'$ of cohort sizes and default counts, respectively, for all n years. Although in slight contrast to Example 2.3, we choose the hyperparameters of our model to be $\theta = (\mu, \sigma, \alpha)$ and the univariate random effects to satisfy

$$b_1 \sim \mathrm{N}\left(\mu, \tfrac{\sigma^2}{1-\alpha^2}\right), \quad b_t \,|\, b_{t-1} \sim \mathrm{N}\left(\alpha b_{t-1} + (1-\alpha)\mu,\, \sigma^2\right) \quad \text{for } t = 2, \ldots, n.$$

This is a first-order autoregressive process (AR(1)), which for $|\alpha| < 1$ has a stationary solution $b \sim \mathrm{N}(\mu, \frac{\sigma^2}{1-\alpha^2})$; an assumption that we shall impose on (b_t). For non-obvious computational reasons we let (i) μ be a part of the hyperparameters instead of the fixed effects $\boldsymbol{\beta}$ and (ii) σ^2 be the variance of the *increments* instead of b (which has variance $\frac{\sigma^2}{1-\alpha^2}$). These modifications will make the full conditionals easier to handle. It is easily seen that $\boldsymbol{b} = (b_1, \ldots, b_n)'$ is multivariate Gaussian with covariance matrix elements

$$\Sigma_{ij} = \mathrm{Cov}(b_i, b_j) = \tfrac{\sigma^2}{1-\alpha^2}\alpha^{|i-j|}, \quad i, j \in \{1, \ldots, n\}.$$

Its inverse Σ^{-1} is tridiagonal with diagonal elements $\frac{1}{\sigma^2}(1, 1+\alpha^2, \ldots, 1+\alpha^2, 1)$, off-diagonal elements $-\alpha/\sigma^2$ and determinant $\sigma^{-2n}(1-\alpha^2)$. 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 (or mass function) of the random quantity X, $[X \,|\, Y]$ is the conditional density of X given Y and by $[X \,|\, \cdot]$ we shall mean the density of X given data and all other parameters of the model, i.e. the full conditional of X. The key to all full conditional distributions is the joint distribution function of data and parameters

$$\begin{aligned} [\alpha, \mu, \sigma, \boldsymbol{M}, \boldsymbol{b}] &= [\boldsymbol{M} \,|\, \boldsymbol{b}][\boldsymbol{b} \,|\, \alpha, \mu, \sigma][\alpha][\mu][\sigma] \\ &= \left(\prod_{t=1}^{n} [M_i \,|\, b_i]\right) [\boldsymbol{b} \,|\, \alpha, \mu, \sigma][\alpha][\mu][\sigma], \end{aligned}$$

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)):

$$\begin{aligned} [\alpha|\cdot] &= \frac{[\alpha, \mu, \sigma, \boldsymbol{M}, \boldsymbol{b}]}{[\mu, \sigma, \boldsymbol{M}, \boldsymbol{b}]} \propto [\alpha, \mu, \sigma, \boldsymbol{M}, \boldsymbol{b}] \\ &\propto [\boldsymbol{b} \,|\, \alpha, \mu, \sigma][\alpha] \\ &\propto \sqrt{\det(\Sigma^{-1})} \exp\left\{-\tfrac{1}{2}(\boldsymbol{b} - \mu\mathbf{1})'\Sigma^{-1}(\boldsymbol{b} - \mu\mathbf{1})\right\}[\alpha] \\ &\propto \sqrt{1-\alpha^2} \exp\left\{-\tfrac{1}{2\sigma^2}(C_1(\boldsymbol{b}, \mu)\alpha^2 - C_2(\boldsymbol{b}, \mu)\alpha)\right\}[\alpha], \end{aligned} \tag{6}$$

where C_1 and C_2 are identified by performing the vector-matrix multiplications. By choosing a uniform prior $[\alpha]$ on $[0, 1]$ the density in (6) is log-concave and can be simulated from with the ARS algorithm (Gilks 1992).[6] The full conditional distribution of μ is found analogously: we assign a zero-mean Gaussian prior with a large variance (vague prior) and obtain

$$\begin{aligned} [\mu|\cdot] &\propto [\boldsymbol{b} \,|\, \alpha, \mu, \sigma][\mu] \\ &\propto \exp\left\{-\tfrac{1}{2}(\boldsymbol{b} - \mu\mathbf{1})'\Sigma^{-1}(\boldsymbol{b} - \mu\mathbf{1})\right\}[\mu] \\ &\propto \exp\Big\{-\tfrac{1}{2}\sum_i\sum_j \Sigma^{-1}_{ij}(b_i - \mu)(b_j - \mu)\Big\}[\mu], \end{aligned}$$

which again has the functional form of a Gaussian distribution and hence is easily simulated.

$$\begin{aligned} [\sigma|\cdot] &\propto [\boldsymbol{b} \,|\, \alpha, \mu, \sigma][\sigma] \\ &\propto \sqrt{\det(\Sigma^{-1})} \exp\left\{-\tfrac{1}{2}(\boldsymbol{b} - \mu\mathbf{1})'\Sigma^{-1}(\boldsymbol{b} - \mu\mathbf{1})\right\}[\sigma] \\ &\propto \sigma^{-n} \exp\left\{-C_3(\boldsymbol{b}, \mu, \alpha)\sigma^{-2}\right\}[\sigma]. \end{aligned}$$

If σ is assigned an inverse-gamma prior (i.e. $1/\sigma^2$ follows the gamma distribution), then $[\sigma|\cdot]$ is again inverse-gamma. The vector $\boldsymbol{b}$ can be treated as one entity and be updated *en bloc*, but for computational reasons it is wiser to deal with the components of $\boldsymbol{b}$ individually:

$$\begin{aligned} [b_t \,|\, \cdot] &\propto [M_t \,|\, b_t][b_t \,|\, \boldsymbol{b}_{-t}, \mu, \sigma, \alpha] \\ &\propto \left(\tfrac{e^{b_t}}{1+e^{b_t}}\right)^{M_t} \left(\tfrac{1}{1+e^{b_t}}\right)^{m_t - M_t} [b_t \,|\, \boldsymbol{b}_{-t}, \mu, \sigma, \alpha], \end{aligned} \tag{7}$$

[6] Again, the assumption $[\alpha] \sim U(0, 1)$ forces the time series (b_t) to be stationary.

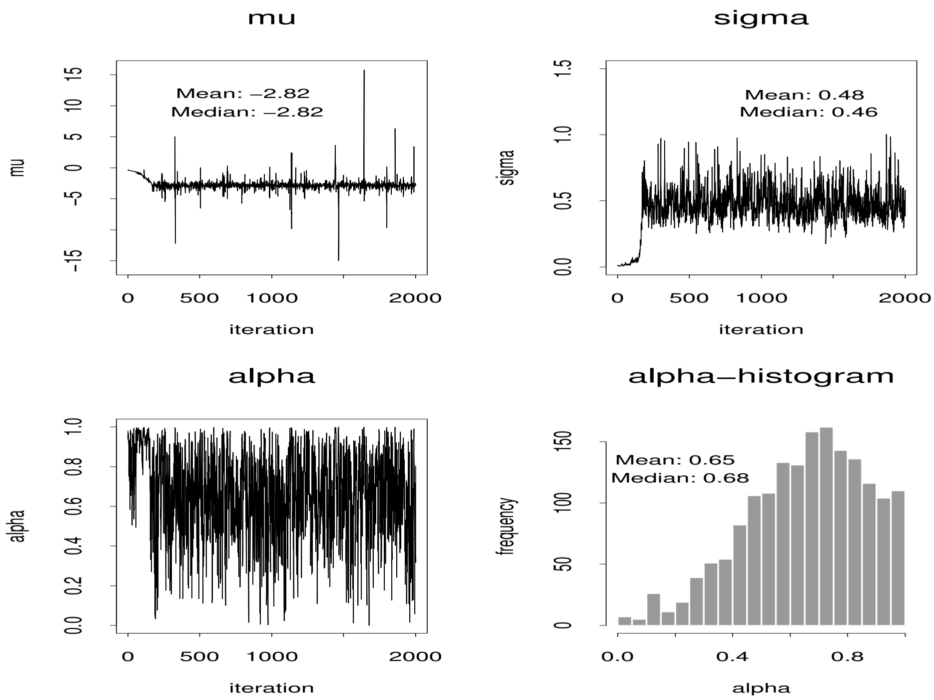

Figure 4: Gibbs sampler for 2000 iterations for the hyperparameters of Example 4.2. The first 300 iterations are considered a *burn-in*, i.e. all inference (including the histogram) is based on iterations 300–2000.

where $\boldsymbol{b}_{-t} = (b_1, \ldots, b_{t-1}, b_{t+1}, \ldots, b_n)$. Although $[b_t|\boldsymbol{b}_{-t}, \mu, \sigma, \alpha]$ is still Gaussian (depending only on the one or two neighboring b's as well as the hyperparameters) the density in (7) is not easily simulated from. With a Metropolis-Hastings step we may use the Gaussian factor $[b_t \mid \boldsymbol{b}_{-t}, \mu, \sigma, \alpha]$ as proposal density and accept the newly simulated value with a certain probability or else remain in the old state. Details on acceptance probabilities can be found in Chapter 6 of Robert and Casella (1999). The following example shows how Gibbs sampler fits this very model to actual credit data.

Example 4.2 (AR(1) random effect). In this example we add the B and CCC cohorts of Example 4.1 into a single group and fit it to an AR(1) random effects model. Figure 4 shows the trajectories of 2000-iteration Gibbs sampling for the hyperparameters μ, σ and α and hence (according to Section 3.2.1) their posterior distributions. For μ and σ, the posterior distributions soon concentrate about the values $\hat{\mu} = -2.82$ and $\hat{\sigma} = 0.46$, despite the vague priors. For α, which was given a uniform prior on $(0, 1)$, there is greater variation in the posterior distribution. This makes inference on α more difficult. However, the histogram rather clearly suggests that the hypothesis $\alpha = 0$ should be rejected.

Extending the above model to comprise also multi-group problems as in Example 2.3 is straightforward in theory, but may unfortunately turn out to pose some practical obstacles. The exact treatment of the multi-group case will depend on whether the driving random effect process $(\boldsymbol{b}_t)$ is chosen univariate or multivariate. Usual assumptions are that given $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ we not only have conditional independence among years, but also among groups in the same year. This turns some of the full conditional distributions into involved products, which in some cases are not as tractable as those seen in Example 4.2 (Gaussian and inverse-gamma). The more difficult cases can in principle be handled by the general Metropolis-Hastings algorithm as discussed in Section 3.2.1, but the exact choice of proposal density *may* affect the convergence of the output considerably and make model-fitting difficult.

4.1.1 Predicting defaults

A time series structure on the random effect lets us use historical default data to predict future defaults. If many defaults occurred in the last year (n) the concept of between-year dependence should make a large number of defaults in the next year ($n+1$) more probable than normal; an effect which is basically captured by the time series-parameter α. This is carried through by introducing M_{n+1} as an unobserved node of our model. Inference on M_{n+1} is made in the same way as with any other unobserved—namely with its posterior distribution, which we find by means of the Gibbs sampler:

Expand the graph of Figure 3 with the nodes b_{n+1}, M_{n+1} and m_{n+1}, where all but the latter are unobserved. For each iteration $i = 1, \ldots, N$ of the Gibbs sampler we now sample M_{n+1} according to

$$M_{n+1} \,|\, b_{n+1} \sim \text{Bin}\left(m_{n+1}, \mathrm{e}^{b_{n+1}}/(1+\mathrm{e}^{b_{n+1}})\right), \tag{8}$$

which is its full conditional distribution, and b_{n+1} according to Equation 7.[7] We shall refer to the posterior distribution of M_{n+1} as the *conditional* default distribution. An alternative way would have been to add only the node b_{n+1} in order to make inference on p_{n+1}, the default probability in year $n+1$, (rather than on M_{n+1} directly), by making the inverse logit-transformation

$$b_{n+1} \mapsto \mathrm{e}^{b_{n+1}}/(1+\mathrm{e}^{b_{n+1}}) =: p_{n+1}(b_{n+1}) \tag{9}$$

and then make inference on p_{n+1}. The default probability for year $n+1$ could furthermore be used to simulate default counts ($\sim$ Bin(m_{n+1}, p_{n+1})) and hence a M_{n+1}-distribution, without including M_{n+1} as a separate node. These two methods both deliver comparable results, although only one performs true inference on M_{n+1}.

This approach is fundamentally different from that lacking a dependence structure on the random effects, as in Example 4.1, where defaults in different years are treated as independent events and a stationary distribution for the yearly default patterns is fitted. The distribution of M_{n+1}, which we shall refer to as *unconditional*, the ordering of the observed defaults $M_1, \ldots, M_n$ will *not* affect the distribution of M_{n+1}, whereas in the conditional case this is essential.

Example 4.3 (Conditional vs. unconditional default distributions). In this example we show how the conditional (time series) model compares to the unconditional distribution seen in Example 4.1. Our goal will be to predict M_{2000}—the number of defaults in year 2000—using only historical default data for 1980–1999[8] and the current cohort size $m_{2000} = 1047$. In order to obtain the conditional distribution of M_{2000}, we include b_{2000} and M_{2000} as two unobserved nodes of our model and change the full conditional distributions accordingly. A histogram showing the posterior distribution of M_{2000} obtained through (8) and the Gibbs sampler is showed to the right in Figure 5.

The unconditional distribution of M_{2000} is obtained by fitting a stationary distribution as that of Example 4.1 to the data. Once the hyperparameters θ and possibly fixed effect regressors $\boldsymbol{\beta}$ have been estimated, we may simulate M_{2000}-distributed variates according to a two-stage scheme, which in our case takes the following form: (i) simulate $b \sim \mathrm{N}(\hat{\mu}, \hat{\sigma}^2)$ and form the default probability $p(b)$ according to (9) and (ii) simulate $M_{2000} \sim$ Bin(m_{2000}, p). We repeat this procedure 2000 times and obtain the unconditional M_{2000}-distribution seen to the left in Figure 5.

The actual number of defaults in year 2000 for the B and CCC rating classes is 85, which clearly is closer to the predictions of the conditional model than those of the unconditional.

[7] The only difference with M_{n+1} unknown as opposed to known is that a new value of M_{n+1} according to (8) must be simulated in each iteration. All other full conditionals are the same as if M_{n+1} were known.

[8] The data consists of cohort sizes (m_t) and default counts (M_t) for the B and CCC cohorts in Example 4.2 for 1980–1999.

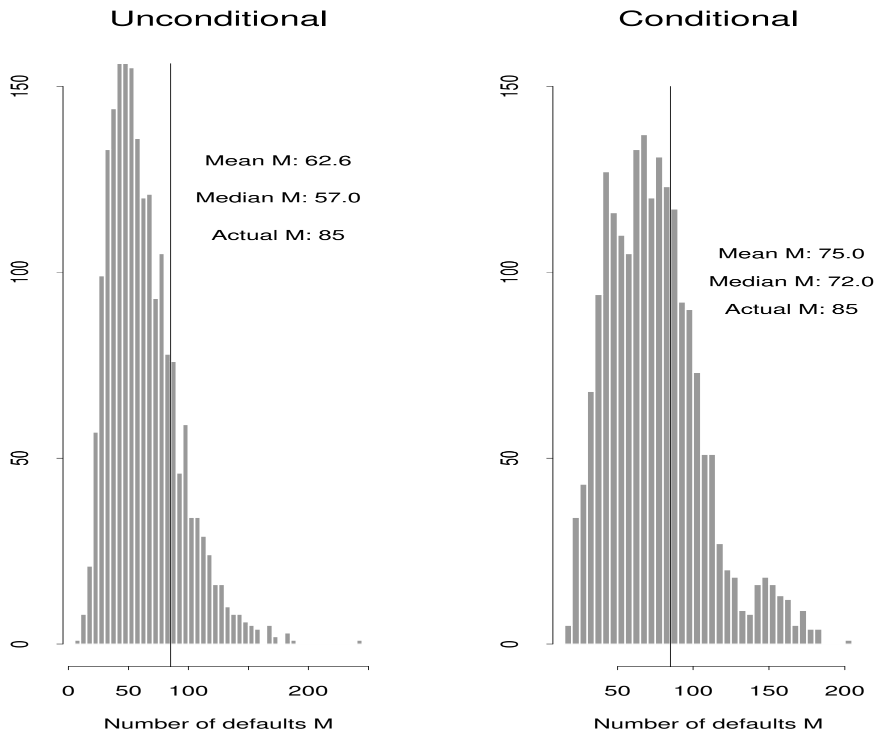

Figure 5: Conditional and unconditional distribution of the number of defaults in year 2000. 2000 simulations with the Gibbs sampler, burn-in 300.

A glance at Figure 1 shows that 1999 is a year of high default intensity, implying that the probability of having a high default intensity also in year 2000 should be higher than normal. The unconditional model, on the other hand, has no way of making use of this extra information and in this case performs worse.

Extending the above ideas to make predictions for years even further in the future is possible, by extending the graph of Figure 3 with nodes $M_{n+2}, \dots$ and $b_{n+2}, \dots$

5 Extensions

The Bernoulli and binomial GLMMs of this paper have focused only on the distinction between default and non-default of obligors. In practice also other events, such as *rating migrations*, may affect the profit-and-loss distribution of the credit portfolio considerably. This can be overcome by regarding the future state of an obligor as a *multinomial* trial with a range of possible rating states $\{1, \dots, K\}$ instead of just a binary default/non-default trial. In this setting the default count vectors $\boldsymbol{M} = (M_1, \dots, M_K)'$ have each component following the multinomial distribution—a special case of what we have studied so far, where the components are binomial.

Other possible extensions include fitting and testing models with not only categorical but also metrical covariates and fitting procedures for multivariate random effects.

References

Clayton, D.G. (1996): "Generalized Linear Mixed Models" in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter (Eds.), Chapman & Hall, London

Fahrmeir, L. and G. Tutz (1994): *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer-Verlag, New York.

Frey, R., and A. McNeil (2001): "Modelling dependent defaults," ETH E-Collection, URL `http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=85`, ETH Zürich.

Frey, R., and A. McNeil (2003): "Dependent Defaults in Models of Portfolio Credit Risk," *Journal of Risk*, (to appear).

Gilks, W.R. (1992): "Derivative-free adaptive rejection sampling for Gibbs sampling" in *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.). Oxford University Press, Oxford

Gilks, W.R. (1996): "Full conditional distributions" in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter (Eds.), Chapman & Hall, London

Ibrahim, J.G., M-H. Chen and D. Sinha (2001): *Bayesian Survival Analysis.* Springer-Verlag, New York.

Joe, H. (1997): *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

McCullagh, P. and J. Nelder (1989): *Generalized Linear Models.* Chapman & Hall, London.

McCulloch, C.E. (1997): "Maximum likelihood algorithms for generalized linear mixed models," *J. Amer. Statist. Assoc.*, **92**, pp. 162–170.

Myles, J., and D.G. Clayton (2001): "`GLMMGibbs`: An R Package for Estimating Bayesian Generalised Linear Mixed Models by Gibbs Sampling," URL `http://www.maths.lth.se/help/R/.R/doc/html/packages.html`

RiskMetrics-Group (1997): "CreditMetrics – Technical Document," available from `http://www.riskmetrics.com/research`.

Robert, C.P., and G. Casella (1999): *Monte Carlo Statistical Methods.* Springer-Verlag, New York.

Standard and Poor's (2001): "Ratings Performance 2000: Default, Transition, Recovery, and Spreads,".