

Dependent Credit Migrations

JONATHAN WENDIN* ALEXANDER J. McNEIL

Department of Mathematics, ETH Zürich

First version: July 2004; Current version: October 2005

Abstract

This paper examines latent risk factors in models for migration risk. We employ the standard statistical framework for ordered categorical variables and induce dependence between migrations by means of latent risk factors. By assuming a Markov process for the dynamics of the latent factors, the model can be interpreted as a *state space model*. The paper contains an empirical study on quarterly migration data from Standard & Poor's from the years 1981–2000, in which the ordered logit model with serially correlated latent factors is fitted by computational Bayesian techniques (Gibbs sampling). Apart from highlighting the usefulness of the Gibbs sampler for statistical inference in models of this kind, the survey in particular investigates the issues of rating-specific factor loadings and heterogeneity among industry sectors, with emphasis on their implications in terms of implied asset correlations.

J.E.L. Subject Classification: C23, C35, C15

Keywords: Credit Risk, State Space Models, Multivariate random effects, Gibbs sampling

1 Introduction

An ordered categorical variable expressing the ability of a company to fulfill its financial obligations is known as a *credit rating* (or simply *rating*). Such classifications are provided by agencies such as Moody's and Standard and Poor's, but could also be according to an internal system of a bank. With the recent Basel II accord the importance of ratings as a tool for risk management has increased (Basel Committee on Banking Supervision, 2005). This has led to interest in statistical models for the dynamics of ratings; the change in rating of an obligor is referred to as a *transition* or a *migration*.

Firstly, it is obvious that the present rating of an obligor is a strong predictor for its rating in the nearest future. A cardinal feature of any migration model is hence past and present ratings influencing the evolution. The *Markov chain* is a stochastic process of this kind, in which the migration probabilities given all past ratings depend only on the present state. This is the simplest form of a Markov chain, and it allows all migration probabilities for a specific time-horizon to be collected in a so-called *migration* (or *transition*) *matrix*. If no obligor-specific properties other than rating are considered, the Markov assumption is very convenient since all migrations taking place in a period can be summarized in cross-sectional migration counts without any loss of generality, cf. Section 2.2.1. The intuitive and comprehensible form of the migration matrix has made it a cornerstone of many current risk management applications, see Jafry and Schuermann (2004) for a review. In most applications migration matrices are estimated in discrete time with a monthly, quarterly or yearly time horizon. As obligors seldom change rating (and when they do it is often to

*e-mail: wendin@math.ethz.ch. Financial support from NCCR Financial Valuation and Risk Management (a research program supported by the Swiss National Science Foundation) is gratefully acknowledged.

a neighbouring state) migration matrices typically have their mass concentrated along the main diagonal; in fact, the low occurrence of certain transitions may be a problem when estimating migration probabilities naively by empirical proportions. One way of mitigating this is a continuous-time Markov approach as in Lando and Skødeberg (2002).

The assumption of firm rating paths following homogeneous Markov chains has been increasingly questioned and even statistically rejected in recent contributions. An empirically verified characteristic of agency rating data is *rating momentum*; for instance the tendency of recently downgraded obligors to be more at risk than other obligors. Such observations contradict the Markov property, see Lando and Skødeberg (2002) and the references therein. Statistical approaches to model non-Markovian rating paths can be found in Christensen et al. (2004) and Frydman and Schuermann (2005). The question of rating momentum will not be treated in this paper, since our analysis is based on repeated cross-sectional migration counts. Time inhomogeneity in migration probabilities is, however, an important topic which we treat next.

Several empirical surveys have found evidence of time variation in migration intensities and confirmed that this time variation may to some extent be explained by observed macro-economic variables, see Nickell et al. (2000), Bangia et al. (2002) and Hu et al. (2002). A second requirement of the migration model is therefore that of dependence among transitions taking place within a time period. The underlying economic conditions will be referred to as the *systematic risk* of the portfolio. 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 migration activity and this lag may vary stochastically over time. The above shortcomings have serious implications for regulation, see the discussion in Koopman, Lucas, and Klaassen (2005).

The approach of this paper will be to complement observed risk factors with unobserved latent ones, that capture the residual systematic risk once any observed parts have been accounted for. The unobserved systematic risk creates dependent migrations in each time period; we refer to this as *cross-sectional dependence*. We also expect the cyclical behaviour of economic factors to create *serial dependence* among migration events in different time periods; see McNeil and Wendin (2005) for a similar discussion in the context of default risk.

Relatively little work has been carried out on latent systematic risk in migration models. An early contribution is Kijima et al. (2002), where a (non-standard) statistical framework for correlated migrations is suggested. Gagliardini and Gouriéroux (2005b) present a general framework for rating dynamics with stochastic migration matrices. In particular, they consider serially correlated migration matrices, and perform an empirical analysis on French corporate data using a linearization of the likelihood function.

The modelling framework of this paper is conceptually close to that of Gagliardini and Gouriéroux (2005b). We consider a general statistical model for discrete-time migration counts based on standard techniques for ordered categorical response variables. The ratings are subject to both observed and unobserved systematic risk, where the unobserved risk is assumed to follow a Markov process. The latent risk factors and their serial dependence are able to capture the effects of cross-sectional dependence. The Markovian latent risk factors induce serial dependence; in fact, the models formally belong to the class of *state space models*.

The serially correlated latent risk factors yield joint migration distributions in terms of high-dimensional integrals, which are indeed awkward for standard maximum likelihood (ML) techniques. Koopman, Lucas, and Klaassen (2005) and Gagliardini and Gouriéroux (2005b) consider models with 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). An alternative approach is that of simulated ML; a recent application in the context of default risk is given in Koopman, Lucas, and Daniels (2005). Koopman, Lucas, and Monteiro (2005) use simulated ML to fit a duration model with stochastic intensity to credit transition data.

Instead of following a frequentist approach, we follow McNeil and Wendin (2005) and use computational Bayesian techniques (Gibbs sampling) for model inference. The algorithms are able to handle varying complex specifications of the latent systematic risk, such as serially correlated factors as well as general multivariate Gaussian risk factors that induce heterogeneity across industry sectors. A brief review of Bayesian statistics and Gibbs sampling is given in Section 3. The rest of the paper is organized as follows. Section 2 introduces the basic framework for models of cross-sectionally and serially dependent migration counts. The empirical analysis on quarterly migration data from Standard & Poor's follows in Section 4. Section 5 concludes.

2 State Space Model of Migration Counts

2.1 Notation

Consider a set $\mathcal{K} = \{1, \dots, K\}$ of rating classes of increasing creditworthiness. The state default, which we assume is absorbing, is denoted by 0 and is included in $\mathcal{K}_0 := \mathcal{K} \cup \{0\}$. Denote by m_{tk} the number of firms in group $k \in \mathcal{K}$ so that $m_t = \sum_{k \in \mathcal{K}} m_{tk}$ is the total number of obligors in the portfolio at the beginning of period t . For each obligor $i = 1, \dots, m_t$ in period t , let $\kappa(t, i)$ denote its initial rating and R_{ti} the rating at the end of the period t . Notice that the former is necessarily in \mathcal{K} , whereas the latter is in \mathcal{K}_0 .

Our aim is to derive a statistical model for $M_{t:k,\ell}$, the number of obligors with ratings $k \in \mathcal{K}$ and $\ell \in \mathcal{K}_0$ at the onset and at the end of period t , respectively. Although m_{t1}, \dots, m_{tK} are assumed to be known at the beginning of period t , the migration count variable $M_{t:k,\ell}$ is known only at the end of it. For notational convenience we introduce the *migration count vector* $\mathbf{M}_{tk} := (M_{t:k,\ell})_{\ell \in \mathcal{K}_0}$, which summarizes the migrations of the k -rated obligors during period t . Analogously, $\mathbf{M}_t := (\mathbf{M}_{tk})_{k \in \mathcal{K}}$.

The framework of this paper presupposes data of *repeated cross-sectional* type so that R_{ti} and R_{si} , the ratings of obligor i in two distinct periods $s \neq t$, do not necessarily refer to the same obligors. Cross-sectional migration counts are easily constructed given a dataset of rating paths (i.e. data of *panel* type, where the migration dates of each obligor are collected), but the reverse construction is in general not possible.

2.2 Cross-sectionally Dependent Migrations

Assumption 1. Conditional on a latent factor \mathbf{b}_t (following a non-degenerate distribution F to be specified) the ratings R_{t1}, \dots, R_{tm_t} at the end of period t are conditionally independent and satisfy

$$P(R_{ti} \leq \ell | \mathbf{b}_t) = g(\mu_{\kappa(t,i),\ell} - \mathbf{x}'_{ti}\boldsymbol{\beta} - \mathbf{z}'_{ti}\mathbf{b}_t), \quad \ell \in \mathcal{K}_0, \quad (1)$$

for some strictly increasing function $g : \mathbb{R} \rightarrow (0, 1)$.

\mathbf{x}_{ti} and \mathbf{z}_{ti} of Assumption 1 denote the known *design vectors* holding the corresponding covariates of obligor i , $\mu_{\kappa(t,i),\ell}$ and $\boldsymbol{\beta}$ are unknown intercepts and regression coefficients to be estimated, and g is known as the *response function*. Common choices of g are $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x \exp\{-u^2/2\} du$ and $1/(1 + \exp\{-x\})$, which lead to the so-called *probit* and *logit* models, respectively.

The formulation in (1) is a *generalized linear mixed model* (GLMM) for ordered polytomous responses, where the rating at the onset of period t has been included as a covariate. If no latent factor \mathbf{b}_t is present, the model is known as a generalized linear model (GLM), the theory of which is treated in the monograph by McCullagh and Nelder (1989). For more

information about GLMMs, see Fahrmeir and Tutz (1994) and Skrondal and Rabe-Hesketh (2004).

The design vectors \mathbf{x}_{ti} and \mathbf{z}_{ti} together with the latent factor \mathbf{b}_t constitute the *systematic risk* $\mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t$ of obligor i in period t . To simplify notation we will often refer to the systematic risk as $\gamma_{ti} := \mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t$. A fundamental role of \mathbf{x}_{ti} and \mathbf{z}_{ti} is to add observed elements to the systematic risk; they may hold quantitative variables as well as dummy variables indicating group membership. By including macro-economic variables or other observed risk factors we may capture time inhomogeneity in transition rates. Likewise, obligor-specific variables such as balance sheet data can be used to capture heterogeneity among obligors. In fact, current rating is an obligor-specific covariate, which for sake of clarity has been kept separate from $\boldsymbol{\beta}$. Observed explanatory variables are known as *fixed effects* in the GLMM-framework.

The latent factors \mathbf{b}_t (or *random effects* as they are now as in the literature on GLMMs) account for unobserved systematic risk, and hereby introduce heterogeneity beyond that which can be captured with observed covariates. In particular, they induce dependence among the responses. The latent factors within a time period may be univariate or multivariate; the latter are useful when treating migrations according to industry sector. Parameters of F are referred to as *hyperparameters* and will be denoted by θ .

It immediately follows from (1) that

$$P(R_{ti} = \ell | \mathbf{b}_t) = g(\mu_{\kappa(t,i),\ell} - \mathbf{x}'_{ti}\boldsymbol{\beta} - \mathbf{z}'_{ti}\mathbf{b}_t) - g(\mu_{\kappa(t,i),\ell-1} - \mathbf{x}'_{ti}\boldsymbol{\beta} - \mathbf{z}'_{ti}\mathbf{b}_t), \quad (2)$$

and that the intercepts $(\mu_{k,\ell})_{k \in \mathcal{K}, \ell \in \mathcal{K}_0}$, which are also known as *threshold values* or *cut-off levels*, for all $k \in \mathcal{K}$ must satisfy

$$-\infty = \mu_{k,-1} \leq \mu_{k,0} \leq \mu_{k,1} \leq \dots \leq \mu_{k,K-1} \leq \mu_{k,K} = \infty$$

in order for the probabilities in (2) to be non-negative. (For the absorbing state default we have $\mu_{0,\ell} = \infty$ for all $\ell \in \mathcal{K}_0$.) Note that a model for $K + 1$ ordered choices requires K thresholds. We shall refer to the vector of K thresholds associated with $k \in \mathcal{K}$ as $\boldsymbol{\mu}_k := (\mu_{k,\ell})_{\ell=0}^{K-1}$ and the full $K \times K$ -set of thresholds will be denoted by $\boldsymbol{\mu} := (\boldsymbol{\mu}_k)_{k \in \mathcal{K}}$.

Unconditionally, R_{t1}, \dots, R_{tm_t} of Assumption 1 are not independent, which is most easily seen by interpreting the threshold model (2) as follows: Let $\varepsilon_{t1}, \dots, \varepsilon_{tm_t}$ be iid rvs with df g , which are independent of \mathbf{b}_t . Set $V_{ti} := \varepsilon_{ti} + \mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t$ for $i = 1, \dots, m_t$, and notice that R_{ti} is generated according to

$$R_{ti} = \ell \iff V_{ti} \in (\mu_{\kappa(t,i),\ell-1}, \mu_{\kappa(t,i),\ell}].$$

ε_{ti} is referred to as the *idiosyncratic risk* of obligor i in period t . It is easy to see that V_{t1}, \dots, V_{tm_t} , despite being conditionally independent given \mathbf{b}_t , are dependent random variables. Moreover, their joint distribution fully determines that of $(R_{t1}, \dots, R_{tm_t})$. V_{ti} is often interpreted as the asset value of obligor i and $\boldsymbol{\mu}_i$ as critical liability levels as in the seminal work on structural models of credit risk in Merton (1974). This representation is particularly useful for the probit case with a Gaussian random effect \mathbf{b}_t (each ε_{ti} is standard Gaussian under the probit response); the joint distribution of V_{t1}, \dots, V_{tm_t} is then Gaussian, and fully determined by its correlation matrix as in the well-known industry model CreditMetrics.

The representation with V_{ti} allows us to quantify the cross-sectional migration dependence in period t in terms of the so-called *implied asset correlation*, $\text{corr}(V_{ti}, V_{tj})$, of two obligors i, j . Given relevant covariates, we trivially have

$$\text{corr}(V_{ti}, V_{tj}) = \text{cov}(\mathbf{z}'_{ti}\mathbf{b}_t, \mathbf{z}'_{tj}\mathbf{b}_t) / (\sqrt{\text{var}(\mathbf{z}'_{ti}\mathbf{b}_t) + w^2} \sqrt{\text{var}(\mathbf{z}'_{tj}\mathbf{b}_t) + w^2}), \quad (3)$$

where $w^2 := \text{var}(\varepsilon_{ti})$. In the probit case we have $w^2 = 1$ while in the logit case $w^2 = \pi^2/3$.

2.2.1 Homogeneous Groups of Obligators

When fully obligor-specific covariates are not considered it is often reasonable to divide the portfolio into homogeneous subgroups (or *buckets*), in which all ingoing obligors are assumed to share the same systematic risk. The index set of subgroups will be denoted by $\mathcal{H} = \{1, \dots, H\}$, where H is the number of homogeneous buckets.

A commonly seen assumption in practice is homogeneity within rating classes, which is accomplished by setting $\mathbf{x}_{ti} = \mathbf{x}_{tk}$ and $\mathbf{z}_{ti} = \mathbf{z}_{tk}$ for all i with $\kappa(t, i) = k$. This corresponds to the case $\mathcal{H} = \mathcal{K}$. Under Assumption 1 and given \mathbf{b}_t , we then have the vectors of migration counts $\mathbf{M}_{t1}, \dots, \mathbf{M}_{tK}$ (see Section 2.1 for their definition) conditionally independent with

$$\mathbf{M}_{tk} | \mathbf{b}_t \sim \text{Multinomial} \left\{ m_{tk}, \mathbf{p}_k(\mathbf{x}'_{tk}\boldsymbol{\beta} + \mathbf{z}'_{tk}\mathbf{b}_t) \right\} \quad \text{for all } k \in \mathcal{K}, \quad (4)$$

where

$$\mathbf{p}_k(\cdot) := \{p_{k,\ell}(\cdot)\}_{\ell \in \mathcal{K}_0} \quad \text{and} \quad p_{k,\ell}(x) := g(\mu_{k,\ell} - x) - g(\mu_{k,\ell-1} - x).$$

Thus, for each $\mathbf{x} = (x_0, x_1, \dots, x_K) \in \{0, 1, \dots, m_{tk}\}^{K+1}$ satisfying $x_0 + x_1 + \dots + x_K = m_{tk}$ we have

$$P(\mathbf{M}_{tk} = \mathbf{x} | \mathbf{b}_t) = \frac{m_{tk}!}{x_0! x_1! \dots x_K!} \prod_{\ell=0}^K p_{k,\ell}(\mathbf{x}'_{tk}\boldsymbol{\beta} + \mathbf{z}'_{tk}\mathbf{b}_t)^{x_\ell}.$$

The unconditional distribution of \mathbf{M}_{tk} is evidently *not* multinomial, since the effect of the latent factors \mathbf{b}_t must be integrated out. Furthermore $\mathbf{M}_{t1}, \dots, \mathbf{M}_{tK}$ are not independent. The systematic risk of group k in period t will be referred to as $\gamma_{tk} := \mathbf{x}'_{tk}\boldsymbol{\beta} + \mathbf{z}'_{tk}\mathbf{b}_t$.

For applications, it might be relevant to perform the grouping according to further attributes than merely rating class; natural candidates are industry sector and country membership. In the case of industry sector effects we then have $\mathcal{H} = \mathcal{K} \times \mathcal{S}$, where \mathcal{S} is an index set of industry sectors. A convenient way of introducing heterogeneity between subgroups is by means of multivariate latent factors. The transition properties of the portfolio are determined by the joint distribution of $(b_{t1}, \dots, b_{tH})'$, where b_{th} denotes the latent systematic risk of group $h \in \mathcal{H}$ at time t . This fits naturally into the GLMM-framework introduced in Assumption 1 by setting $\mathbf{b}_t = (b_{t1}, \dots, b_{tH})'$ and $\mathbf{z}_{th} = \mathbf{e}_h$, where \mathbf{e}_h is the h th unit vector in \mathbb{R}^H . Applications of this technique are given in Section 4.2.

2.2.2 Migration Correlations

A consequence of the latent factor \mathbf{b}_t is dependence between rating migrations occurring in a time period. It is often convenient to quantify this in terms of the migration correlation, an entity that relates to two particular obligors and a corresponding (bivariate) rating transition from $(k_1, k_2) \in \mathcal{K}^2$ to $(\ell_1, \ell_2) \in \mathcal{K}_0^2$. It is well-known that migration correlations alone do not determine the full distribution of the migrations. Nevertheless, they present useful summaries for comparisons between different statistical models for migrations. A review of issues regarding migration correlations is given in Gagliardini and Gouriéroux (2005a).

In the present context, migration correlations are derived with the conditional independence property and (2). Assuming that $\kappa(t, i) = k_i$ for two obligors $i = 1, 2$ we have:

$$\begin{aligned} & \text{cov} \left(I_{\{R_{t1}=\ell_1\}}, I_{\{R_{t2}=\ell_2\}} \right) \\ &= E \left[E \left[I_{\{R_{t1}=\ell_1\}} I_{\{R_{t2}=\ell_2\}} | \mathbf{b}_t \right] \right] - \prod_{i=1}^2 E \left[E \left[I_{\{R_{ti}=\ell_i\}} | \mathbf{b}_t \right] \right] \\ &= E \left[\prod_{i=1}^2 p_{k_i, \ell_i}(\mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t) \right] - \prod_{i=1}^2 E \left[p_{k_i, \ell_i}(\mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t) \right] \\ &= \int \prod_{i=1}^2 p_{k_i, \ell_i}(\mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t) dF(\mathbf{b}_t) - \prod_{i=1}^2 \int p_{k_i, \ell_i}(\mathbf{x}'_{ti}\boldsymbol{\beta} + \mathbf{z}'_{ti}\mathbf{b}_t) dF(\mathbf{b}_t). \end{aligned} \quad (5)$$

Since the number of combinations (k_1, k_2) and (ℓ_1, ℓ_2) above is large, we restrict our attention to the practice-relevant question of migrations in the same direction as well as default

correlations. In this spirit we define the joint *upgrade-correlation* and *downgrade-correlation* as

$$\text{corr}(I_{\{R_{t1} > k_1\}}, I_{\{R_{t2} > k_2\}}) \quad \text{and} \quad \text{corr}(I_{\{R_{t1} < k_1\}}, I_{\{R_{t2} < k_2\}}),$$

respectively. Naturally, the joint default correlation follows as $\text{corr}(I_{\{R_{t1}=0\}}, I_{\{R_{t2}=0\}})$. Given $\boldsymbol{\mu}$, $\boldsymbol{\beta}$ and θ these expressions are easily calculated numerically.

2.3 Serially Dependent Migrations

The cyclical behaviour of economic factors leads us to expect dependence between migration events taking place in different time periods. Given the interpretation of \mathbf{b}_t as a general state-of-the-economy in period t , it seems reasonable to let its value at time $t + 1$ depend on \mathbf{b}_t , and hence to impose a Markovian structure on the sequence (\mathbf{b}_t) .

Assumption 2. (i) The sequence (\mathbf{b}_t) is a Markov chain; and (ii) conditionally on (\mathbf{b}_t) , the \mathbf{M}_t 's are independent and \mathbf{M}_t depends on \mathbf{b}_t only.

Assumption 2 defines a *state space model* (or *hidden Markov model*, *HMM*) for the count sequence (\mathbf{M}_t) , see Künsch (2001). (Recall that m_{tk} is treated as a known variable, cf. Chapter 2 of MacDonald and Zucchini (1997) on binomial HMMs.) It is straight-forward to see that $\mathbf{M}_1, \dots, \mathbf{M}_T$ in general are not independent in the HMM framework. Assumption 2 thus provides a meaningful *serial dependence* for the migration counts (\mathbf{M}_t) , where the components of \mathbf{M}_t already exhibit cross-sectional dependence by Assumption 1.

The first-order autoregressive, AR(1), time series (b_t)

$$b_t = \alpha b_{t-1} + \phi \epsilon_t, \quad t \geq 1, \quad b_0 = \phi \epsilon_0 / \sqrt{1 - \alpha^2}, \quad (6)$$

where $\epsilon_0, \epsilon_1, \dots$ are iid $N(0, 1)$, is a real-valued Markov process in discrete time. For $|\alpha| < 1$ it has a Gaussian stationary distribution with mean 0 and variance $\sigma^2 := \phi^2 / (1 - \alpha^2)$. Its d -dimensional counterpart takes the form

$$\mathbf{b}_t = \alpha \mathbf{b}_{t-1} + A(\epsilon_{t,1}, \dots, \epsilon_{t,d})', \quad t \geq 1, \quad \mathbf{b}_0 = A(\epsilon_{0,1}, \dots, \epsilon_{0,d})' / \sqrt{1 - \alpha^2} \quad (7)$$

with $\alpha \in (-1, 1)$, $A \in \mathbb{R}^{d \times d}$ and $(\epsilon_{t,i})$ iid $N(0, 1)$. The stationary solution is Gaussian with mean zero and covariance matrix $\Sigma := \Phi / (1 - \alpha^2)$, where $\Phi := AA'$. The parameterizations in (6) and (7) turn out to be useful for the applications of Section 4.

Before turning to model calibration, we make a remark on the implications of Assumptions 1 and 2 on the rating path of a single obligor. The migration counts $\mathbf{M}_1, \dots, \mathbf{M}_T$ are easily seen to be independent *if* the underlying rating paths are Markovian. An important consequence of this observation is that serially dependent migration counts (as in the state space setting) necessarily imply non-Markovian rating paths. This feature should not be confounded with what is known as *rating momentum* on the obligor level, where a recent downgrade of an obligor increases its probability of further downgrades or default (Lando and Skødeberg, 2002). Models incorporating rating momentum treat each firm individually in order to model its rating path explicitly, and hence presuppose a dataset of actual rating paths. The non-Markovian rating paths under Assumptions 2, on the other hand, can be viewed as rating momentum due to general past adverse or favourable periods. Note that if $\mathbf{b}_1, \dots, \mathbf{b}_T$ instead are assumed to be independent, the rating paths are Markovian.

3 Calibration with Markov Chain Monte Carlo

The unconditional distribution of the migration count vectors (\mathbf{M}_t) as defined in Section 2 require the effect of the latent factors (\mathbf{b}_t) to be integrated out, which greatly complicates the use of standard maximum likelihood (ML) techniques. This difficulty applies even more

if the latent factors are serially correlated. Unless the likelihood function is approximated numerically, some kind of simulation-based approach to inference seems inevitable, see for instance the discussion of simulated ML in Chapter 3 of Gouriéroux and Monfort (1996). An overview of general approaches to frequentist and Bayesian inference in a very similar context is given in McNeil and Wendin (2005). The latter highlights the usefulness of computational Bayesian procedures for inference in models of portfolio credit risk, an approach we follow in this paper as well.

3.1 Bayesian Statistics and Markov chain Monte Carlo

In *Bayesian statistics*, all unobserved elements ϑ of the statistical model are treated as random variables. These are assigned a joint *prior distribution* $p(\vartheta)$, on a suitable space Θ , expressing any information about ϑ we might have before observing the data D . Inference is based on the *posterior distribution* of ϑ , $p(\vartheta | D)$, which is obtained by applying Bayes' rule. In the context of Section 2, ϑ holds the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$, the hyperparameters θ as well as the latent risk factors $\mathbf{b}_1, \dots, \mathbf{b}_T$, whereas the migration counts $\mathbf{M}_1, \dots, \mathbf{M}_T$ and design vectors constitute the observed data D .

The posterior distribution of ϑ is in general unobtainable by analytical means, but can be simulated from with *Markov chain Monte Carlo* (MCMC) techniques. The objective of these is to generate a realization $\vartheta^{(1)}, \vartheta^{(2)}, \dots$ of an ergodic Markov chain, whose stationary distribution is the desired posterior $p(\vartheta | D)$. The B iterations prior to convergence are known as *burn-in*. A prominent example is the *Gibbs sampler* which proceeds by simulating from the so-called *full conditional distributions*

$$p(\vartheta_i | \vartheta_1, \dots, \vartheta_{i-1}, \vartheta_{i+1}, \dots, \vartheta_r, D)$$

in a pre-specified order, where $\vartheta \equiv (\vartheta_1, \dots, \vartheta_r)$ is a decomposition of ϑ with $r \geq 2$ (Gilks, 1996).

By exploiting conjugacy we obtain full conditionals which are easy to simulate from. In many cases of practical interest, however, non-standard full conditionals necessarily emerge. Note that ϑ_i , $i = 1, \dots, r$, can be univariate, which greatly facilitates simulation since 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 Section 4. The Gibbs sampler along with other approaches to non-standard full conditionals are treated in detail in Chapters 6 and 7 of Robert and Casella (1999).

3.2 Inference with MCMC

Inference about a single parameter ϑ_i of the model is based on $\vartheta_i^{(B+1)}, \dots, \vartheta_i^{(B+N)}$, which is a sample of size N from $p(\vartheta_i | D)$, the marginal posterior distribution of ϑ_i . Point estimates and standard errors are given by forming the sample mean (or median) and standard deviation, respectively. Other features of $p(\vartheta_i | D)$ are estimated accordingly; the posterior probability of ϑ_i lying in $\Theta_0 \subset \mathbb{R}$ is estimated as $\#\{t : \vartheta_i^{(B+t)} \in \Theta_0\} / N$. This offers an intuitive way of conducting tests of hypotheses (Casella and Berger, 2002, Ch. 8), a technique that we will make use of when discussing the significance of regression coefficients $\boldsymbol{\beta}$. As ϑ also includes the latent risk factors, the output of the MCMC algorithm can be used to investigate the posterior path of (\mathbf{b}_t) . In particular, one can visually compare the path of (\mathbf{b}_t) with those of observed economic variables.

A further strength of the MCMC approach is that the above techniques can be applied to model quantities derived from several primary parameters as well, such as migration probabilities or correlations. Assuming that these can be written in the form $f(\vartheta)$ for some $f : \Theta \rightarrow \mathbb{R}$, a sample from their respective posterior distributions is instantly given by

$f(\vartheta^{(B+1)}), \dots, f(\vartheta^{(B+N)})$. This means that point estimates as well as standard errors of derived quantities are obtained at virtually no additional computational cost.

3.3 Bayesian Model Validation

Bayesian model validation involves two steps: firstly robustness (i.e. quantifying the impact of the prior distribution on the posterior point estimates) and secondly assessment of model fit, where the latter step is closely related to model selection. The first issue can be addressed by considering several prior specifications and comparing the resulting posterior estimates. By restricting ourselves to so-called *non-informative* (or *vague*) priors, the differences in posterior estimates are generally kept small.

For the second step we consider a modern approach to Bayesian model comparison based on *cross-validation predictive densities* and regression diagnostics derived from these. Due to the multivariate nature of \mathbf{M}_{tk} as basic observation, a marginal likelihood is a suitable diagnostic to work with for the purposes of Section 4. We consider the *conditional predictive ordinate* (CPO), defined as

$$\text{CPO}_t := p(\mathbf{M}_{t1,\text{obs}}, \dots, \mathbf{M}_{tK,\text{obs}} | \{\mathbf{M}_{s1,\text{obs}}, \dots, \mathbf{M}_{sK,\text{obs}} : s = 1, \dots, T, s \neq t\}),$$

where $\mathbf{M}_{tk,\text{obs}}$ denotes the observed realisation of \mathbf{M}_{tk} . The CPO is attractive in that it suggests how likely the joint observation $\mathbf{M}_{t1,\text{obs}}, \dots, \mathbf{M}_{tK,\text{obs}}$ is, when the model is fitted to all observations except $\mathbf{M}_{t1,\text{obs}}, \dots, \mathbf{M}_{tK,\text{obs}}$. It can be implemented promptly by re-using the output of the Gibbs sampler. By comparing the plots $\{(t, \text{CPO}_t) : t = 1, \dots, T\}$ as well as $T^{-1} \sum_t \log(\text{CPO}_t)$ for different models, we obtain an assessment of the model fits. See Gelfand (1996) and Carlin and Louis (2000) for an in-depth survey of the above issues and more.

4 Empirical Analysis

In this section we suggest a series of migration models of increasing complexity and fit these to S&P data by Gibbs sampling.

4.1 Dataset

The dataset analysed in this paper has been extracted from Standard & Poor's CreditPro™ 6.6 database and consists of 5,651 US and Canadian firms from 12 S&P industry sectors (see Table 1). The migration counts \mathbf{M}_{tk} have been collected for three-month periods, ranging from January 1981 to December 2000 ($T = 80$ quarters). Obligors whose rating is withdrawn during the period have been excluded; the same applies to obvious duplicates in the database such as holding companies. The rating classes under consideration are

$$\mathcal{K} = \{\text{CCC}, \text{B}, \text{BB}, \text{BBB}, \text{A}, \text{AA}, \text{AAA}\},$$

where for the case of simplicity qualifiers have been suppressed ($K = 7$). 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. All other notation follows Section 2.

4.2 Models

Throughout this section we employ the *logit response* $g(x) = 1/(1 + \exp\{-x\})$ in the framework of Sections 2.2 and 2.3. Apart from being the canonical response function for many GLMs (McCullagh and Nelder, 1989) the logit response is also straight-forward to evaluate (the *probit response* $g(x) = \Phi(x)$ requires numerical evaluation). We do not consider

obligor-specific covariates, so firms are arranged into homogeneous groups as discussed in Section 2.2.1.

The empirical study follows the common Bayesian approach of using *non-informative* priors; detailed information about prior distributions is given in Section 4.3.1.

Model Class (P) — Preliminary Analysis

Our preliminary analysis, also known as model class (P), explores the validity of the Chicago Fed National Activity Index (CFNAI) (x_t) as a proxy for the business cycle. The CFNAI is published on a monthly basis and can be downloaded from the internet, see McNeil and Wendin (2005).

We assume that all obligors in rating class k in period t are exposed to the same systematic risk γ_{tk} , $k \in \mathcal{K}$. Given $\gamma_{t1}, \dots, \gamma_{tK}$, the migration count vectors $\mathbf{M}_{t1}, \dots, \mathbf{M}_{tK}$ are assumed to be conditionally independent with

$$\mathbf{M}_{tk} | \gamma_{tk} \sim \text{Multinomial} \{m_{tk}, \mathbf{p}_k(\gamma_{tk})\},$$

as in (4). We give three different specifications of γ_{tk} . First we consider two simple GLMs which are free from latent factors:

- (P1) $\gamma_{tk} = x_t \beta$;
- (P2) $\gamma_{tk} = x_t \beta_k$,

where x_t is the CFNAI for the first calendar month of time period t . The unknown parameters of models (P1) and (P2) to be estimated are the threshold values $(\mu_{k,\ell})$ and the regression coefficients β and β_1, \dots, β_K , respectively.

We then investigate a GLMM which in addition to (x_t) features a sequence of serially dependent latent factors (b_t) , in order to capture any cross-sectional or serial dependence among the migrations:

- (P3) $\gamma_{tk} = x_t \beta + b_t$,

where (b_t) is the univariate AR(1) sequence with variance $\sigma^2 = \phi^2 / (1 - \alpha^2)$ introduced in (6). Given x_t , the implied asset correlation is zero under the first two models, and equals $\sigma^2 / (\sigma^2 + \pi^2 / 3)$ under model (P3). Model (P3) requires estimation of the hyperparameters ϕ and α in addition to the unknown parameters of model (P1).

Results

The posterior mean and standard deviation of all parameters of model class (P) are contained in Tables 2, 3 and 4, from which we draw the following conclusions. Four of the threshold parameters apparently exhibit large standard errors, as the corresponding transitions never take place in the dataset (see rows AAA and B). The large standard errors merely indicate the profound uncertainty in assigning these events a probability. It is also worth noting that the thresholds of rating class CCC exhibit higher standard errors than the cut-off levels of the other subinvestment-grade ratings. One reason for this is the small size of the CCC-cohort (rating classes B and BB are approximately 10 times larger than CCC).

Secondly, even though model (P1) suggests high explanatory power of the CFNAI, the results of model (P2) are slightly contradictory: the coefficients β_k remain significant (although the BB and CCC rating categories are borderline cases), but β_{AAA} carries a different sign than expected. We interpret this as empirical evidence for differences in exposure to systematic risk across rating classes.

Model (P3) is summarized in Table 4. The point estimates of ϕ and α clearly suggest the presence of latent systematic risk in the migrations; the variance of b_t , $\phi^2 / (1 - \alpha^2)$, suggests an implied asset correlation of 3.5%, and the time series parameter α points to profound serial dependence. Figure 1 presents the posterior mean of b_t in all time periods. For sake of

reference, we include the CFNAI in the graph for visual comparison. There is some degree of co-movement between the two time series $(x_t\beta)$ and (b_t) , but it is clear that the observed covariate does not capture the full variability in migration rates. Finally, the presence of (b_t) has reduced the explanatory power of CFNAI to the extent that it is no longer formally significant. We therefore leave out (x_t) in the remainder of the analysis.

The parameters of model (P3) can be used to calculate various model summaries: Table 11 holds the migration matrix, and Tables 12 and 13 the corresponding joint up- and downgrade correlation matrices, respectively, as defined in Section 2.2. Although migration correlations are small numbers (joint downgrade correlations usually higher than upgrade correlations), the values presented here are slightly higher than those of Gagliardini and Gouriéroux (2005b) based on yearly migration counts on French corporate data. (This holds also when we consider only the one-step up-up and down-down correlations.)

Model Class (K) — Rating-specific Factor Loadings

With the Basel II capital adequacy framework, in which risk loadings may depend on the credit quality of the obligor, the issue of rating-specific factor loadings has been subject to recent interest; see Basel Committee on Banking Supervision (2002) and Lopez (2004). We therefore propose the following. The findings of model (P2) motivate us to treat similarly-rated firms as exchangeable, but we allow the K rating classes to be subject to different systematic risk. The setting is hence that of Section 2.2.1 with $\mathcal{H} = \mathcal{K}$, which explains the label (K) of the current model class. The dependence structure of the model is specified by the joint distribution of $\mathbf{b}_t := (b_{t1}, \dots, b_{tK})'$, where b_{tk} denotes the latent risk of the k th rating class in period t . We will consider three Gaussian specifications of \mathbf{b}_t in turn: perfect dependence, exchangeable risk factors and finally an arbitrary covariance matrix. The results of model class (P) motivate us to refrain from introducing observable business-cycle covariates in our analysis, hence $\mathbf{x}_{tk} = \mathbf{0}$.

As in model class (P), we assume that $\mathbf{M}_{t1}, \dots, \mathbf{M}_{tK}$ are conditionally independent given $\gamma_{t1}, \dots, \gamma_{tK}$ with

$$\mathbf{M}_{tk} | \gamma_{tk} \sim \text{Multinomial} \{m_{tk}, \mathbf{p}_k(\gamma_{tk})\}.$$

Model (K1)

Let (b_t) be the univariate AR(1) sequence (6) with $\phi = 1$, and define

$$(K1) \quad \gamma_{tk} = b_{tk} = \phi_k b_t.$$

Note that we impose $\text{var}(b_t) = 1/(1 - \alpha^2)$ for identifiability. The components of $\mathbf{b}_t = b_t(\phi_1, \dots, \phi_K)'$ are perfectly dependent (i.e. the model is still in the single risk factor framework). Moreover, there is no reason to exclude negative ϕ_k 's a priori. The implied asset correlation for two obligors in the k th rating class is $\sigma_k^2/(\sigma_k^2 + \pi^2/3)$, where $\sigma_k^2 = \phi_k^2/(1 - \alpha^2)$. Model (K1) is a standard specification of rating-specific risk factor loadings and has been used in e.g. Gordy and Heitfield (2002) and Rösch (2005).

Model (K2)

Let (b_t) be the univariate AR(1) sequence in (6) with variance $\sigma^2 = \phi^2/(1 - \alpha^2)$, and define

$$(K2) \quad \gamma_{tk} = b_{tk} = b_t + \xi_{tk},$$

where (ξ_{tk}) are iid $N(0, \omega^2)$. The vector \mathbf{b}_t is multivariate Gaussian with mean zero and

$$\text{cov}(b_{tk}, b_{tl}) = \begin{cases} \sigma^2 + \omega^2 & \text{if } k = l, \\ \sigma^2 & \text{else,} \end{cases} \quad (8)$$

hence, its components are exchangeable (the covariance matrix exhibits *compound symmetry*). Furthermore, there is serial dependence between the \mathbf{b}_t 's due to the common factor b_t . Although model (K2) does not display rating-specific factor loadings, it allows us to obtain a feeling for the covariance structure of a multivariate Gaussian distribution on \mathbf{b}_t . In particular, the magnitude of ω^2 reveals the degree of variability in systematic risk across different rating classes. For two obligors of rating classes k and l , model (K2) suggests an implied asset correlation of $(\sigma^2 + \omega^2)/(\sigma^2 + \omega^2 + \pi^2/3)$ if $k = l$, and $\sigma^2/(\sigma^2 + \pi^2/3)$ otherwise.

Model (K3)

Assume that (\mathbf{b}_t) follows the K -dimensional AR(1) sequence (7) and set

$$(K3) \quad \gamma_{tk} = b_{tk}.$$

The covariance matrix of \mathbf{b}_t is $\Sigma = (\Sigma_{kl}) \in \mathbb{R}^{K \times K}$ and consists of $K(K+1)/2$ parameters, compared to only two in model (K2). For two firms in rating classes k and l the implied asset correlation equals $\Sigma_{kl}/(\Sigma_{kl} + \pi^2/3)$ under (K3).

Results

Point estimates of the parameters of models (K1), (K2) and (K3) are given in Tables 5, 6 and 7, respectively. The threshold parameters are similar to those of model class (P) and will not be commented.

The variance of the latent factor in model (P3) suggests an average implied asset correlation of 3.5%, whereas those of (K1) range between a fraction of a percent (class AAA) and 8.0% (class B). The Gaussian specification of \mathbf{b}_t in models (K2) and (K3) leads to higher implied asset correlations for firms sharing rating class. However, the point estimate of $\text{corr}(b_{tk}, b_{tl})$, $k \neq l$, of model (K2) is only 34%, which indicates low implied asset correlations for obligors in non-identical rating classes. The results of model (K3) are similar to those of the constrained model (K2), although there is large uncertainty about many of the elements of Σ .

Model Class (S) — Heterogeneity across Industry Sectors

This section addresses the issue of heterogeneity among industry sectors by means of multivariate latent risk factors. Let $\mathcal{S} = \{1, \dots, S\}$ be an index set of industry sectors for which migration counts \mathbf{M}_{tsk} are collected for each rating category $k \in \mathcal{K}$, sector $s \in \mathcal{S}$ and time period t . Notice that $\sum_s \mathbf{M}_{tsk}$ (componentwise summation) yield \mathbf{M}_{tk} as previously defined. Table 1 introduces the $S = 10$ sectors of our study. We treat all obligors sharing industry sector and rating class exchangeably, thus $\mathcal{H} = \mathcal{K} \times \mathcal{S}$. In this section we define $\mathbf{b}_t = (b_{t1}, \dots, b_{tS})'$ and denote by γ_{tsk} the systematic risk associated with \mathbf{M}_{tsk} .

It seems plausible to find non-perfect dependence between the systematic risk of different industry sectors. We therefore concentrate on Gaussian specifications of \mathbf{b}_t : both the compound symmetry model and the fully general covariance structure are treated. We also include the case of rating-specific factor loadings after industry effects have been accounted for in model (S3).

Given the systematic risks $\{\gamma_{tsk} : s \in \mathcal{S}, k \in \mathcal{K}\}$ we assume that the migration count vectors $\{\mathbf{M}_{tsk} : s \in \mathcal{S}, k \in \mathcal{K}\}$ are conditionally independent with

$$\mathbf{M}_{tsk} | \gamma_{tsk} \sim \text{Multinomial} \{m_{tsk}, \mathbf{p}_k(\gamma_{tsk})\}.$$

We now give three different specifications of the risks $\{\gamma_{tsk} : s \in \mathcal{S}, k \in \mathcal{K}\}$ in period t .

Model (S1)

Let (b_t) be the Gaussian AR(1) sequence (6) and assume that (ξ_{ts}) are iid $N(0, \omega^2)$ and independent from (b_t) . We define

$$(S1) \quad \gamma_{tsk} = b_{ts} = b_t + \xi_{ts}.$$

Under model (S1) we have \mathbf{b}_t Gaussian with mean zero and covariance matrix as in (8). In particular, the systematic risk is identical for all obligors sharing industry sector, irrespective of rating class. The implied asset correlation for two firms sharing sector equals $(\sigma^2 + \omega^2)/(\sigma^2 + \omega^2 + \pi^2/3)$, whereas that of two firms in different sectors is merely $\sigma^2/(\sigma^2 + \pi^2/3)$.

Model (S2)

A more general specification assumes that (\mathbf{b}_t) follows the S -dimensional AR(1) process (7) with

$$(S2) \quad \gamma_{tsk} = b_{ts}.$$

Consequently, the implied asset correlation between two firms in sectors k and l equals $\Sigma_{kl}/(\Sigma_{kl} + \pi^2/3)$, where $\Sigma = (\Sigma_{kl}) \in \mathbb{R}^{S \times S}$ is the covariance matrix of \mathbf{b}_t .

Model (S3)

Our final specification of γ_{tsk} merges models (K1) and (S1) by combining rating-specific factor loadings with sector effects:

$$(S3) \quad \gamma_{tsk} = \phi_k b_{ts} = \phi_k (b_t + \xi_{ts}),$$

where $\text{var}(b_t)$ has been constrained to $1/(1 - \alpha^2)$ for identifiability. (All other assumptions of models (K1) and (S1) apply.) This specification helps to avoid confounding heterogeneity across rating classes with events that can be traced back to an industry sector. The implied asset correlations under model (S3) will depend on both rating and sector membership.

Results

Point estimates of the parameters of models (S1), (S2) and (S3) are given in Tables 8, 9 and 10, respectively. The point estimate of ω in model (S1) is 51% larger than σ , and thus reveals material evidence of sector-specific variability (the correlation between b_{ts} of two different sectors is roughly 30%). Consequently, the overall implied asset correlation 3.5% of model (P3) is now 2.6 or 8.7%, depending on whether the two obligors belong to different sectors or not. Table 9, which allows for an arbitrary covariance matrix Σ for \mathbf{b}_t , yields a similar picture, albeit with large standard errors on many of the covariance parameters. Note that the correlation between certain sectors is high, whereas other appear to be rather uncorrelated.

Model (S3), which combines rating-specific factor loadings with industry sector effects, is summarized in Table 10. The factor loading of every rating class is significant; moreover, the magnitude of the ϕ_k 's shows less variability than in model (K1). A point estimate of the correlation between the b_{ts} 's of two different sectors is 37%.

4.3 Implementation Details

The models of the previous section are fitted by Gibbs sampling with self-customized code in C. The derivation of the full conditional distributions is sketched in Appendix A. The running time of a 10,000-iteration simulation of the algorithms ranges from a few minutes up to an hour, depending on the complexity of the model.

4.3.1 Choice of Priors

This study follows the common practice of using *non-informative* priors for the unknown parameters of the model.

The intercept vectors $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$ are assigned iid zero-mean, ordered Gaussian prior distributions with variance $\tau^2 I_{K \times K}$, where τ is large ($\tau = 100$) and $I_{K \times K}$ is the identity matrix. Likewise, β and β_1, \dots, β_K of models (P1) and (P2) are given iid $N(0, \tau^2)$ -priors. The autoregressive parameter α is assumed to be uniform on $(-1, 1)$ a priori.

The variance of the innovations ϕ^2 is assigned an *improper* prior decaying as $1/x$; this is a limiting case of the inverse-gamma distribution ($1/\phi^2$ is assumed to follow the $\Gamma(\eta, \nu)$ -distribution with $(\eta, \nu) = (0, 0)$). This is a standard vague prior for a scaling parameter, see McNeil and Wendin (2005). We employ the same prior for ω^2 of models (K2) and (S1) as well.

In models (K3) and (S2) Φ is assigned an *inverse-Wishart* distribution with parameters ν and Λ_0 , see Appendix A. This is a standard prior for a covariance matrix. For small values of ν the inverse-Wishart distribution is vague: we use $\nu = 0.001$ with Λ_0 set to the identity matrix. References on informative priors for a covariance matrix are given in Boscardin and Weiss (2004).

The scaling parameters ϕ_1, \dots, ϕ_K in models (K1) and (S3) occur in the form of regression coefficients rather than hyperparameters, which means that the inverse-gamma prior no longer leads to an inverse-gamma full conditional. As there is no prior reason for all these parameters to share sign, they are assigned independent $N(0, \tau^2)$ priors.

4.3.2 Model Comparisons

We apply the cross-validation density-approach described in Section 3.3 to compare the models of Section 4.2. Table 14 contains the relevant summaries of the conditional predictive ordinates (CPOs); we consider the average CPO value of each model (to the left) as well as the number of time periods (out of $T = 80$) in which the more advanced model(s) perform(s) better (see the right part of Table 14).

It is evident from Table 14 that the nine models under study can be divided into four groups of comparable model performance, which we now list in order of increasing model fit. The two simple GLMs (P1) and (P2) undoubtedly exhibit the poorest fit. The two univariate risk factor models (P3) and (K1) come second and are followed by models (K2) and (K3) with Gaussian distributions for $(b_{t1}, \dots, b_{tK})'$. Finally, model class (S) featuring sector-specific risk factors displays the superior model fit. The differences in model fit within these four groups are moderate and barely worth mentioning.

4.4 Discussion

One of the main conclusions of the preliminary analysis (P) and model class (K) is that the exposure to systematic risk may vary across rating classes. This is particularly evident for rating class AAA, whose coefficients β_{AAA} and ϕ_{AAA} in models (P2) and (K1) either have a different sign than expected or equal zero; moreover, its systematic risk in model (K3) shows very little correlation with the other rating classes. It is of course important to bear in mind that assessment of the systematic risk is difficult for rating classes where obligors are scarce. This applies especially to rating class CCC, whose parameter uncertainty is constantly large, but also to class AAA to some extent.

This being said, the results of model class (K) still point towards heterogeneity among rating classes, which is evident for instance from the large estimate of ω in model (K2). A consequence of ω in (K2) is that the implied asset correlations instantly fall by two thirds if the obligors belong to different rating classes. Clearly, there is no economic justification for massively reduced implied asset correlation just because two obligors belong to neighbouring rating classes. One therefore might suspect that the assumption of homogeneity within each

rating class is violated and that the correlation structure of $(b_{t1}, \dots, b_{tK})'$ is a substitute for effects not accounted for. Applying a finer classification scheme when forming the homogeneous groups or including other covariates could alleviate this problem. Before proceeding, we note that models (K2) and (K3) exhibit considerably larger estimates of $\text{var}(b_{tk})$ than models (P3) and (K1). Imposing perfect dependence on $(b_{t1}, \dots, b_{tK})'$ as in (K1) thus seems to force the variances to be modest. (Note that other empirical studies may treat the default or migration histories of each rating class as independent time series.)

In the next set of models homogeneous groups are composed with the two attributes rating class and industry sector. Each industry sector is allotted a latent risk factor, which in models (S1) and (S2) is identical for all rating classes. The correlation between the sector-specific risk factors is 30% in the exchangeable model (S1), whereas the correlation between industry sectors varies considerably in the general model (S2), despite the large parameter uncertainty. As industry sectors may be subject to different business conditions, a non-perfect dependence structure on $(b_{t1}, \dots, b_{tS})'$ makes good economic sense. Recall also that models featuring sector effects are clearly favoured in the model comparison, cf. Table 14. Nevertheless, a fully general covariance structure on \mathbf{b}_t as in model (S2) appears to be slightly over-ambitious given the historical data available.

When combining sector effects with rating-specific factor loadings as in (S3), we notice that the order of magnitude of the ϕ_k 's is much more comparable than in model (K1) without sector effects. This seems to suggest that the need for rating-specific factor loadings has decreased in presence of sector effects. As this specification is low in parameters and displays a good model fit, it will be our preferred model.

Finally, our findings do not necessarily support the view that factor loadings drop with increasing probability of default, at least not for the investment-grade obligors (rating BBB and above). We believe that properly addressing the issue of heterogeneity among industry sectors is more imperative. It should, however, be pointed out that inference about the investment grade is based mainly on migration events other than default. Thus, inferring factor loadings from default data only might lead to other conclusions (although estimates for the investment grade are usually highly uncertain due to the rare occurrence of defaults).

5 Conclusions

This paper presents a statistical framework for dependent rating migrations driven by latent, serially correlated systematic risk factors. Unobserved risk factors have several advantages over observed ones, which are briefly addressed in the Introduction. The paper also highlights the use of computational Bayesian statistics (MCMC) for inference in credit migration models with latent effects. MCMC is capable of handling a variety of specifications of the systematic risk, which are highly relevant for practice. In particular, the methodology can deal with multivariate Gaussian risk factors in order to capture heterogeneity among industry sectors.

The empirical study of quarterly cross-sectional migration data from S&P suggests the presence of serially dependent, latent, systematic risk in migration rates; the main empirical conclusions are very similar to those of McNeil and Wendin (2005). For practical purposes, the effect of the latent systematic risk is most easily interpreted in terms of implied asset correlations, which are defined in Section 2.2. The study provides empirical evidence of heterogeneity between rating classes, the origin of which lies in a violation of the assumed homogeneity among the obligors of a rating class. We therefore suggest taking e.g. industry affiliation into account before forming homogeneous groups of obligors.

We also find that a single risk factor often suggests smaller implied asset correlations than multivariate specifications of the latent risk. The implied asset correlations of firms sharing industry sector are not far from the values prescribed by regulators (Basel Committee on Banking Supervision, 2002), whereas the corresponding values for firms in different sectors

in general are substantially lower (although not much lower than those of the corresponding single risk factor model).

Some of the implied asset correlations of our analysis (especially for the investment-grade ratings) might be received as worryingly low if used as inputs in a risk management context, in particular when treating only defaults. As defaults of investment-grade obligors are very rare, the implied asset correlations of these rating classes will mainly be influenced by migration events other than default, but they nevertheless reflect the correlation structure of the implied asset values $(V_{t1}, \dots, V_{tm_t})'$ as defined in Section 2.2. However, care should be taken when applying the same estimates to events far out in the tails of $(V_{t1}, \dots, V_{tm_t})'$ such as joint defaults of investment-grade firms; it is well known, that the correlation structure alone does not fully determine the joint distribution of $(V_{t1}, \dots, V_{tm_t})'$, in particular not its tails. Frey et al. (2001) illustrate the substantial model risk in this context. A proper investigation of the tail behaviour of $(V_{t1}, \dots, V_{tm_t})'$ remains a subject for future research.

Acknowledgements

We thank Standard & Poor's for providing the dataset used in Section 4. The first author gratefully acknowledges financial support from NCCR Financial Valuation and Risk Management (a research program supported by the Swiss National Science Foundation).

A Deriving the Full Conditional Distributions

This section employs a notation that is common in literature on Gibbs sampling: $[X]$ denotes the (unconditional) density (or mass function) of the random quantity X , $[X|Y]$ the conditional counterpart given Y , and $[X|\cdot]$ is the full conditional distribution of X . \underline{X} denotes the full sequence (X_1, \dots, X_T) . The key to all full conditional distributions is the joint distribution function of data and parameters, which reads

$$\begin{aligned} [\underline{\mathbf{M}}, \underline{\mathbf{b}}, \underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \alpha, \phi, \omega] &\propto [\underline{\mathbf{M}}|\underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \underline{\mathbf{b}}][\underline{\mathbf{b}}|\alpha, \phi, \omega][\underline{\boldsymbol{\mu}}, \beta, \alpha, \phi, \omega] \\ &= \left(\prod_{t=1}^T \prod_{i \in \mathcal{K}} [M_{ti}|\boldsymbol{\mu}_i, x_t, \beta, b_t] \right) [\underline{\mathbf{b}}|\alpha, \phi, \omega][\underline{\boldsymbol{\mu}}][\beta][\alpha][\phi][\omega]. \end{aligned} \quad (9)$$

The last line follows by conditional independence arguments and a priori independence of the parameters. For more information about the Gibbs sampler and full conditional distributions, see Gilks (1996).

We exemplify the derivation of a full conditional distribution in the context of model (P3). Before proceeding, we observe that $\underline{\mathbf{b}} := (b_1, \dots, b_T)$ defined as in (6) is multivariate Gaussian with covariance matrix Σ_b :

$$\text{cov}(b_s, b_t) = \phi^2 \alpha^{|s-t|} / (1 - \alpha^2), \quad s, t \in \{1, \dots, T\}.$$

Its inverse Σ_b^{-1} is tridiagonal with diagonal elements $\phi^{-2}(1, 1 + \alpha^2, \dots, 1 + \alpha^2, 1)$, off-diagonal elements $-\phi^{-2}\alpha$ and determinant $\phi^{-2T}(1 - \alpha^2)$.

The full conditional of α is calculated by applying the formula for conditional probabilities and using (9). In each step of the derivation we drop factors without explicit dependence on α , which explains the use of the α -sign:

$$[\alpha|\cdot] = \frac{[\underline{\mathbf{M}}, \underline{\mathbf{b}}, \underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \phi, \alpha]}{[\underline{\mathbf{M}}, \underline{\mathbf{b}}, \underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \phi]} \propto [\underline{\mathbf{M}}, \underline{\mathbf{b}}, \underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \phi, \alpha] \propto [\underline{\mathbf{b}}|\phi, \alpha][\alpha].$$

The final step involves properties of $[\underline{\mathbf{b}}|\phi, \alpha]$

$$\begin{aligned} [\alpha|\cdot] &\propto [\underline{\mathbf{b}}|\phi, \alpha][\alpha] \propto \sqrt{\det(\Sigma_b^{-1})} \exp\left\{-\frac{1}{2}\underline{\mathbf{b}}\Sigma_b^{-1}\underline{\mathbf{b}}'\right\} [\alpha] \\ &\propto \sqrt{1 - \alpha^2} \exp\left\{-\frac{1}{2}\phi^{-2}(C_1(\underline{\mathbf{b}})\alpha^2 - C_2(\underline{\mathbf{b}})\alpha)\right\} [\alpha], \end{aligned} \quad (10)$$

where $C_1(\cdot)$ and $C_2(\cdot)$ are identified by performing the vector-matrix multiplications. If α is assigned a uniform prior, (10) is log-concave in α and can be simulated from with the ARS algorithm. The other full conditionals are derived analogously, see McNeil and Wendin (2005) for a similar application. Except for ϕ^2 , whose full conditional remains inverse-gamma, all other full conditionals of our study rely on the ARMS algorithm for their simulation (Gilks, 1992).

In case of multivariate $\mathbf{b}_t = (b_{t1}, \dots, b_{tH})'$, note that $\underline{\mathbf{b}} := (\mathbf{b}'_1, \dots, \mathbf{b}'_T)'$ is multivariate Gaussian with covariance matrix

$$\text{cov}(b_{si}, b_{tj}) = \Phi_{ij} \frac{\alpha^{|t-s|}}{1 - \alpha^2}, \quad \begin{array}{l} s, t \in \{1, \dots, T\}, \\ i, j \in \{1, \dots, H\}. \end{array}$$

As Φ has the properties of a covariance matrix, we assign it a *inverse-Wishart* prior distribution. This means that its inverse $\Phi^{-1} := \Lambda$ follows the Wishart distribution

$$[\Phi^{-1}] = [\Lambda] \sim \text{Wishart}(\nu, \Lambda_0),$$

where ν and Λ_0 are fixed real numbers and symmetric, non-singular $H \times H$ -matrices, respectively. Its prior density is thus proportional to

$$|\Lambda|^{(\nu-H-1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\Lambda_0\Lambda)\right\},$$

see Johnson and Kotz (1972) or Muirhead (1982). The inverse-Wishart distribution is a conjugate prior of a covariance matrix:

$$[\Lambda | \cdot] \sim \text{Wishart}(\nu + T, \Lambda_0 + C_3(\underline{\mathbf{b}})),$$

where $C_3(\underline{\mathbf{b}})$ is the following $H \times H$ -matrix:

$$C_3(\underline{\mathbf{b}}) = \sum_{t=1}^T \mathbf{b}_t \mathbf{b}'_t + \alpha^2 \sum_{t=2}^{T-1} \mathbf{b}_t \mathbf{b}'_t - \alpha \sum_{t=2}^T (\mathbf{b}_t \mathbf{b}'_{t-1} + \mathbf{b}_{t-1} \mathbf{b}'_t),$$

see Appendix A.2 of Bernardo and Smith (1994) or McCulloch and Rossi (1994). An algorithm for generating variates from the Wishart distribution is given in Johnson (1987). In each iteration of the Gibbs sampler the current value of Φ is obtained by inverting Λ .

Due to the form of $[\underline{\mathbf{M}} | \underline{\boldsymbol{\mu}}, \underline{\mathbf{x}}, \beta, \underline{\mathbf{b}}]$, the full conditional distribution of the latent risk factor \mathbf{b}_t does not follow a standard statistical distribution. We therefore update each b_{th} separately by means of the ARMS algorithm, an exercise which is greatly simplified by first deriving the conditional distribution of b_{th} given all other elements of $\underline{\mathbf{b}}$. Let $\underline{\mathbf{b}}_{-t} := (\mathbf{b}'_1, \dots, \mathbf{b}'_{t-1}, \mathbf{b}'_{t+1}, \dots, \mathbf{b}'_T)'$ and observe that under the multivariate AR(1) process in (7) we have $[\mathbf{b}_t | \underline{\mathbf{b}}_{-t}, \alpha, \Phi]$ multivariate Gaussian:

$$[\mathbf{b}_t | \underline{\mathbf{b}}_{-t}, \alpha, \Phi] \sim \begin{cases} N(\alpha \mathbf{b}_2, \Phi) & \text{if } t = 1, \\ N(\alpha \mathbf{b}_{T-1}, \Phi) & \text{if } t = T, \\ N\left(\frac{\alpha}{1+\alpha^2}(\mathbf{b}_{t-1} + \mathbf{b}_{t+1}), \frac{1}{1+\alpha^2}\Phi\right) & \text{else.} \end{cases} \quad (11)$$

Secondly, it is easily verified that if $\mathbf{Z} = (Z_1, \dots, Z_H)'$ is Gaussian with mean $(\mu_1, \dots, \mu_H)'$, then for each $h = 1, \dots, H$

$$Z_h | \{Z_i = z_i\}_{i=1, \dots, H; i \neq h} \sim N\left(\mu_h + \frac{1}{\tau_{hh}} \sum_{\substack{i=1 \\ i \neq h}}^H \tau_{ih}(\mu_i - z_i), \frac{1}{\tau_{hh}}\right), \quad (12)$$

where (τ_{ij}) denotes the *inverse* of the covariance matrix of \mathbf{Z} (the so-called *precision matrix*). Combining (11) and (12) the above observations yields the conditional distribution of b_{th} given the remainder of $\underline{\mathbf{b}}$ and the hyperparameters. All other full conditionals are simulated as in the case of univariate latent factors.

References

- Bangia, A., F. X. Diebold, A. Kronimus, C. Schlagen, and T. Schuermann (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *J. Banking Finance* 26, 445–474.
- Basel Committee on Banking Supervision (2002, October). Quantitative impact, Study 3, Technical guidance. Bank of International Settlements.
- Basel Committee on Banking Supervision (2005, May). Working paper No. 14, Studies on validation of internal rating systems. Bank of International Settlements.
- Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian Theory*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Chichester: John Wiley & Sons Ltd.
- Boscardin, W. J. and R. E. Weiss (2004). Fitting unstructured covariance matrices to longitudinal data. Working paper, UCLA School of Public Health.
- Carlin, B. P. and T. A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis* (2nd ed.). Chapman & Hall/CRC.
- Casella, G. and R. L. Berger (2002). *Statistical Inference* (2nd ed.). Duxbury, Pacific Grove CA.
- Christensen, J. H. E., E. Hansen, and D. Lando (2004). Confidence sets for continuous-time rating transition probabilities. *J. Banking Finance* 28, 2575–2602.
- Fahrmeir, L. and G. Tutz (1994). *Multivariate Statistical Modelling Based On Generalized Linear Models*. Springer, New York.
- Frey, R., A. J. McNeil, and M. Nyfeler (2001). Copulas and credit models. *Risk* 14(10), 111–114.
- Frydman, H. and T. Schuermann (2005, June). Credit rating dynamics and Markov mixture models. Preprint, Wharton Financial Institutions Center.
- Gagliardini, P. and C. Gouriéroux (2005a). Migration correlation: Definition and efficient estimation. *J. Banking Finance* 29, 865–894.
- Gagliardini, P. and C. Gouriéroux (2005b). Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics* 3(2), 188–226.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. In W. Gilks, S. Richardson, and D. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 145–161. Chapman & Hall, London.
- Gilks, W. R. (1992). Derivative-free adaptive rejection sampling for Gibbs sampling. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 641–649. Oxford University Press, Oxford.
- Gilks, W. R. (1996). Full conditional distributions. In W. Gilks, S. Richardson, and D. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 75–88. Chapman & Hall, London.
- Gordy, M. and E. Heitfield (2002). Estimating default correlations from short panels of credit rating performance data. Technical report, Federal Reserve Board.

- Gouriéroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods*. CORE Lecture series. New York: Oxford University Press.
- Hu, Y.-T., R. Kiesel, and W. Perraudin (2002). The estimation of transition matrices for sovereign credit ratings. *J. Banking Finance* 26, 1383–1406.
- Jafry, Y. and T. Schuermann (2004). Measurement, estimation and comparison of credit migration matrices. *J. Banking Finance* 28, 2603–2639.
- Johnson, M. E. (1987). *Multivariate Statistical Simulation*. Wiley.
- Johnson, N. L. and S. Kotz (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.
- Kijima, M. K., K. Komoribayashi, and E. Suzuki (2002). A multivariate Markov model for simulating correlated defaults. *J. Risk* 4(4), 1–32.
- Koopman, S. J., A. Lucas, and R. J. Daniels (2005). A non-Gaussian panel time series model for estimating and decomposing default risk. Working paper, Tinbergen Institute.
- Koopman, S. J., A. Lucas, and P. Klaassen (2005). Empirical credit cycles and capital buffer formation. *J. Banking Finance*, in press.
- Koopman, S. J., A. Lucas, and A. Monteiro (2005). The multi-state latent factor intensity model for credit rating transitions. Working paper, Tinbergen Institute.
- Künsch, H. R. (2001). State space and hidden Markov models. In O. Barndorff-Nielsen, D. Cox, and C. Klüppelberg (Eds.), *Complex Systems*, pp. 109–173. CRC Press.
- Kurbat, M. and I. Korablev (2002). Methodology for testing the level of the EDFtm credit measure. Technical Report #020729, Moody's KMV.
- Lando, D. and T. Skødeberg (2002). Analyzing rating transitions and rating drift with continuous observations. *J. Banking Finance* 26, 423–444.
- Lopez, J. A. (2004). The empirical relationship between average asset correlation, firm probability of default, and asset size. *Journal of Financial Intermediation* 13, 265–283.
- MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov Models and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.
- McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall, London.
- McCulloch, C. E. and P. E. Rossi (1994). An exact likelihood analysis of the multinomial probit model. *J. Econometrics* 64, 207–240.
- McNeil, A. J. and J. Wendin (2005). Bayesian inference for generalized linear mixed models of portfolio credit risk. Preprint, ETH Zürich.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *J. Finance* 29, 449–470.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.
- Nickell, P., W. Perraudin, and S. Varotto (2000). Stability of ratings transitions. *J. Banking Finance* 24, 203–227.

- Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. New York: Springer.
- Rösch, D. (2005). An empirical comparison of default risk forecasts from alternative credit rating philosophies. *International Journal of Forecasting* 21, 37–51.
- Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling*. Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, FL. Multilevel, longitudinal, and structural equation models.

Sector	Name	# Obligors	CAN
1	"Aerospace, automotive, capital goods, metal"	765	5.9
2	"Consumer, service sector"	920	3.2
3	"Leisure time, media"	571	5.3
4	"Utility"	486	5.4
5	"Health care, chemicals"	406	2.8
6	"High tech, computers, office equipment" + "Telecom"	531	6.1
7	"Financial institutions"	717	4.2
8	"Insurance"	444	4.0
9	"Energy and natural resources"	359	9.1
10	"Forest and building products, homebuilders" + "Real estate"	452	10.9

Table 1: The table displays the industry sectors used for the analysis of Section 4. Sectors 6 and 10 are mergers of two regular S&P sectors. The rightmost column shows the proportion (%) of Canadian firms in the sector.

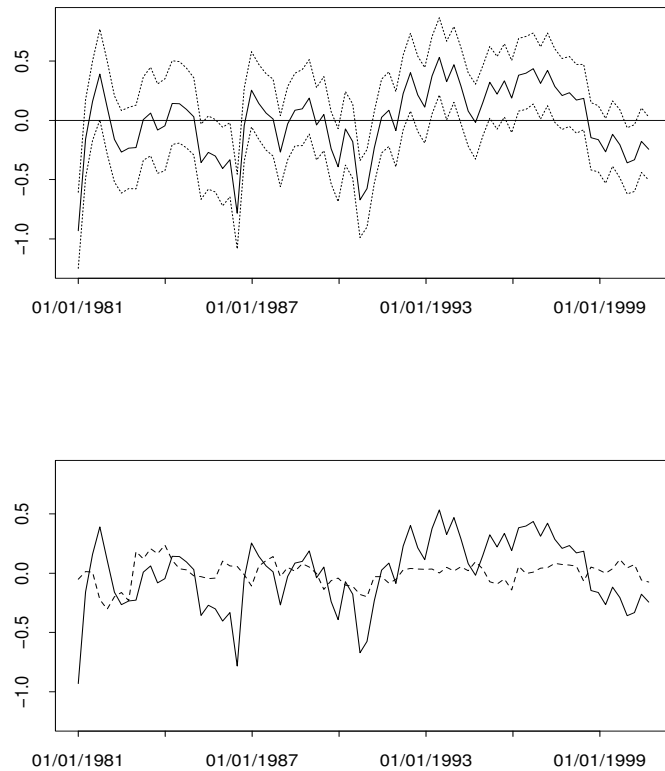


Figure 1: The upper plot shows the posterior mean of the latent process (b_t) under model (P3) including its 95% confidence bounds. Observe that the width of the confidence interval decreases with time, as the number of obligors in the sample grows. The lower plot shows the posterior mean (full line) together with the evolution of the CFNAI $\{(t, x_t\beta) : t = 1, \dots, T\}$ for reference. The point estimate of β in the lower plot is taken from model (P1).

k	Threshold parameters $\mu_{k,\ell}$						Remaining parameters					
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	$\mu_{k,A}$	$\mu_{k,AA}$	$\mu_{k,A}$	$\mu_{k,AA}$	β
AAA	-28.67 (4.93)	-22.48 (5.73)	-16.16 (4.95)	-8.65 (.82)	-7.62 (.54)	-6.16 (.27)	-3.81 (.08)	-3.81 (.08)	-3.81 (.08)	-3.81 (.08)	-3.81 (.08)	0.116 (.020)
AA	-11.34 (1.38)	-9.81 (.80)	-8.11 (.40)	-7.67 (.32)	-6.77 (.20)	-3.97 (.05)	6.37 (.17)	6.37 (.17)	6.37 (.17)	6.37 (.17)	6.37 (.17)	
A	-10.11 (.70)	-9.77 (.62)	-7.60 (.22)	-6.47 (.13)	-4.19 (.04)	5.25 (.07)	8.37 (.33)	8.37 (.33)	8.37 (.33)	8.37 (.33)	8.37 (.33)	
BBB	-8.15 (.33)	-7.73 (.26)	-6.18 (.13)	-4.22 (.05)	4.27 (.05)	7.25 (.22)	9.93 (.80)	9.93 (.80)	9.93 (.80)	9.93 (.80)	9.93 (.80)	
BB	-6.58 (.18)	-5.86 (.13)	-3.66 (.04)	4.04 (.05)	6.54 (.18)	8.35 (.43)	9.79 (.80)	9.79 (.80)	9.79 (.80)	9.79 (.80)	9.79 (.80)	
B	-4.65 (.06)	-3.68 (.04)	4.20 (.05)	6.29 (.14)	6.95 (.20)	8.67 (.47)	20.37 (5.57)	20.37 (5.57)	20.37 (5.57)	20.37 (5.57)	20.37 (5.57)	
CCC	-2.01 (.07)	3.25 (.12)	4.92 (.26)	5.72 (.38)	6.95 (.65)	8.84 (1.34)	9.84 (1.67)	9.84 (1.67)	9.84 (1.67)	9.84 (1.67)	9.84 (1.67)	

Table 2: Posterior mean and standard deviation (in brackets) of **model (P1)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000). The posterior probability of $\{\beta < 0\}$ is 0.000.

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters		
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	β_k	$p(\beta_k < 0 D)$				
AAA	-28.62 (4.93)	-22.47 (5.65)	-16.32 (4.95)	-8.69 (.83)	-7.65 (.54)	-6.19 (.26)	-3.85 (.09)	-0.315 (.107)	0.999				
AA	-11.30 (1.35)	-9.81 (.79)	-8.11 (.37)	-7.68 (.30)	-6.79 (.20)	-3.98 (.05)	6.39 (.16)	0.216 (.058)	0.000				
A	-10.10 (.73)	-9.77 (.64)	-7.60 (.22)	-6.47 (.13)	-4.19 (.04)	5.25 (.07)	8.37 (.33)	0.164 (.043)	0.000				
BBB	-8.14 (.33)	-7.72 (.27)	-6.18 (.13)	-4.22 (.05)	4.27 (.05)	7.26 (.22)	9.94 (.80)	0.113 (.045)	0.008				
BB	-6.58 (.18)	-5.86 (.12)	-3.66 (.04)	4.04 (.05)	6.54 (.18)	8.34 (.43)	9.78 (.81)	0.067 (.045)	0.069				
B	-4.65 (.07)	-3.68 (.04)	4.20 (.05)	6.30 (.15)	6.96 (.21)	8.69 (.47)	20.37 (5.53)	0.136 (.044)	0.001				
CCC	-2.01 (.07)	3.25 (.11)	4.90 (.26)	5.70 (.37)	6.93 (.64)	8.78 (1.32)	9.79 (1.66)	0.124 (.077)	0.054				

Table 3: Posterior mean and standard deviation (in brackets) of **model (P2)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

k	Threshold parameters $\mu_{k,\ell}$						Remaining parameters					
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	β	ϕ	α		
AAA	-28.94 (4.78)	-22.61 (5.70)	-16.32 (5.04)	-8.61 (.84)	-7.59 (.55)	-6.14 (.28)	-3.82 (.13)	0.060				
AA	-11.28 (1.38)	-9.79 (.80)	-8.08 (.38)	-7.66 (.32)	-6.78 (.22)	-3.98 (.11)	6.46 (.19)	0.256				
A	-10.08 (.68)	-9.75 (.60)	-7.62 (.24)	-6.48 (.15)	-4.19 (.10)	5.33 (.12)	8.44 (.34)	0.672				
BBB	-8.15 (.34)	-7.73 (.28)	-6.18 (.16)	-4.22 (.11)	4.36 (.11)	7.34 (.24)	10.03 (.80)					
BB	-6.60 (.20)	-5.88 (.15)	-3.67 (.10)	4.13 (.11)	6.63 (.20)	8.42 (.42)	9.84 (.81)					
B	-4.69 (.12)	-3.72 (.11)	4.25 (.11)	6.35 (.17)	7.02 (.22)	8.74 (.48)	20.53 (5.48)					
CCC	-2.04 (.12)	3.29 (.15)	4.93 (.27)	5.71 (.39)	6.91 (.65)	8.72 (1.33)	9.72 (1.68)					

Table 4: Posterior mean and standard deviation (in brackets) of **model (P3)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000). The posterior probability of $\{\beta < 0\}$ is 0.119.

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters		
	$\mu_{k,D}$	$\mu_{k,B}$	$\mu_{k,CCC}$	$\mu_{k,BB}$	$\mu_{k,AAA}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	ϕ_k	$p(\phi_k < 0 D)$			
AAA	-28.69 (4.91)	-16.18 (4.86)	-22.46 (5.66)	-8.67 (.83)	-7.62 (.53)	-6.16 (.27)	-3.83 (.10)	-0.020 (.110)	0.554				
AA	-11.25 (1.43)	-8.05 (.40)	-9.75 (.82)	-7.62 (.34)	-6.74 (.23)	-3.94 (.11)	6.47 (.19)	0.235 (.050)	0.000				
A	-10.03 (.73)	-7.58 (.22)	-9.70 (.64)	-6.44 (.14)	-4.16 (.08)	5.32 (.10)	8.43 (.33)	0.191 (.038)	0.000				
BBB	-8.13 (.34)	-6.16 (.15)	-7.71 (.28)	-4.19 (.10)	4.35 (.10)	7.34 (.24)	10.03 (.80)	0.222 (.041)	0.000				
BB	-6.58 (.23)	-3.67 (.15)	-5.87 (.19)	4.21 (.16)	6.71 (.23)	8.53 (.45)	9.99 (.82)	0.379 (.051)	0.000				
B	-4.68 (.16)	4.33 (.16)	-3.71 (.16)	6.43 (.21)	7.09 (.25)	8.81 (.50)	20.20 (5.60)	0.387 (.051)	0.000				
CCC	-2.01 (.12)	4.96 (.29)	3.31 (.16)	5.74 (.42)	6.95 (.69)	8.76 (1.34)	9.73 (1.66)	0.250 (.066)	0.000				
$\alpha :$										0.689 (.115)			

Table 5: Posterior mean and standard deviation (in brackets) of **model (K1)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters		
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	$\mu_{k,A}$	$\mu_{k,AA}$	$\mu_{k,AA}$	ϕ	ω	α
AAA	-28.82 (4.89)	-22.48 (5.66)	-16.26 (5.04)	-8.72 (.81)	-7.67 (.54)	-6.20 (.28)	-3.85 (.13)				0.167		(.034)
AA	-11.31 (1.38)	-9.84 (.83)	-8.15 (.39)	-7.72 (.33)	-6.82 (.22)	-4.01 (.11)	6.54 (.19)				0.387		(.026)
A	-10.08 (.71)	-9.74 (.63)	-7.61 (.24)	-6.48 (.15)	-4.20 (.10)	5.40 (.12)	8.50 (.33)				0.801		(.090)
BBB	-8.16 (.35)	-7.74 (.29)	-6.18 (.15)	-4.22 (.10)	4.42 (.11)	7.40 (.23)	10.11 (.83)						
BB	-6.64 (.20)	-5.93 (.16)	-3.71 (.11)	4.20 (.11)	6.70 (.20)	8.47 (.43)	9.90 (.83)						
B	-4.76 (.12)	-3.79 (.10)	4.31 (.11)	6.41 (.17)	7.08 (.22)	8.82 (.49)	20.55 (5.64)						
CCC	-2.12 (.12)	3.34 (.15)	5.01 (.27)	5.81 (.38)	7.03 (.64)	8.86 (1.30)	9.85 (1.61)						

Table 6: Posterior mean and standard deviation (in brackets) of **model (K2)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

Threshold parameters $\mu_{k,\ell}$											Remaining parameters
k	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	$\mu_{k,AAA}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	α
AAA	-28.88 (4.75)	-22.73 (5.49)	-16.65 (4.86)	-9.08 (.85)	-8.03 (.57)	-6.55 (.35)	-4.19 (.25)				0.470 (.076)
AA	-11.42 (1.37)	-9.91 (.83)	-8.19 (.39)	-7.77 (.33)	-6.88 (.23)	-4.08 (.12)	6.51 (.20)				
A	-10.19 (.71)	-9.85 (.63)	-7.67 (.24)	-6.53 (.15)	-4.25 (.09)	5.33 (.11)	8.44 (.34)				
BBB	-8.21 (.34)	-7.79 (.28)	-6.24 (.15)	-4.28 (.09)	4.34 (.10)	7.33 (.24)	10.03 (.82)				
BB	-6.69 (.21)	-5.98 (.17)	-3.76 (.12)	4.17 (.12)	6.67 (.21)	8.46 (.45)	9.92 (.84)				
B	-4.86 (.12)	-3.88 (.11)	4.24 (.11)	6.34 (.18)	7.01 (.23)	8.74 (.48)	20.49 (5.55)				
CCC	-2.32 (.16)	3.29 (.18)	4.95 (.28)	5.74 (.38)	6.95 (.63)	8.73 (1.30)	9.72 (1.65)				

Covariance matrix Σ														
Posterior mean														
	AAA	AA	A	BBB	BB	B	CCC	Posterior standard deviation						
	AAA	AA	A	BBB	BB	B	CCC	AAA	AA	A	BBB	BB	B	CCC
AAA	1.174	-0.065	0.008	0.025	-0.105	0.109	-0.011	(.467)	(.203)	(.187)	(.193)	(.202)	(.195)	(.221)
AA	-0.065	0.322	0.072	0.061	0.091	0.065	0.018	(.203)	(.099)	(.060)	(.060)	(.066)	(.064)	(.076)
A	0.008	0.072	0.207	0.060	0.032	0.046	0.031	(.187)	(.060)	(.060)	(.044)	(.046)	(.047)	(.055)
BBB	0.025	0.061	0.060	0.210	0.083	0.071	0.020	(.193)	(.060)	(.044)	(.062)	(.048)	(.048)	(.058)
BB	-0.105	0.091	0.032	0.083	0.321	0.031	-0.041	(.202)	(.066)	(.046)	(.048)	(.085)	(.065)	(.075)
B	0.109	0.065	0.046	0.071	0.031	0.293	0.134	(.195)	(.064)	(.047)	(.048)	(.065)	(.080)	(.071)
CCC	-0.011	0.018	0.031	0.020	-0.041	0.134	0.466	(.221)	(.076)	(.055)	(.058)	(.075)	(.071)	(.152)

Table 7: Posterior mean and standard deviation (in brackets) of **model (K3)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters		
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$				ϕ :	ω :	α :
AAA	-28.95 (4.86)	-22.77 (5.62)	-16.50 (4.97)	-8.63 (.83)	-7.58 (.54)	-6.13 (.28)	-3.80 (.13)				0.191		(.037)
AA	-11.35 (1.43)	-9.84 (.87)	-8.09 (.41)	-7.67 (.34)	-6.78 (.24)	-3.97 (.12)	6.67 (.20)				0.471		(.024)
A	-10.24 (.75)	-9.89 (.67)	-7.65 (.25)	-6.50 (.17)	-4.21 (.11)	5.53 (.12)	8.64 (.35)				0.792		(.096)
BBB	-8.19 (.34)	-7.77 (.28)	-6.22 (.16)	-4.26 (.12)	4.51 (.12)	7.49 (.25)	10.21 (.81)						
BB	-6.68 (.21)	-5.96 (.16)	-3.76 (.11)	4.21 (.12)	6.71 (.20)	8.53 (.44)	9.97 (.82)						
B	-4.78 (.12)	-3.81 (.11)	4.31 (.12)	6.41 (.18)	7.07 (.22)	8.78 (.47)	20.53 (5.50)						
CCC	-2.10 (.13)	3.40 (.16)	5.07 (.27)	5.87 (.38)	7.07 (.65)	8.86 (1.32)	9.84 (1.65)						

Table 8: Posterior mean and standard deviation (in brackets) of **model (S1)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters					
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	α								
AAA	-28.94	(4.72)	-22.83	(5.60)	-16.50	(5.02)	-8.72	(.82)	-7.68	(.53)	-6.22	(.27)	-3.87	(.11)	0.573	(.055)
AA	-11.38	(1.37)	-9.90	(.83)	-8.18	(.39)	-7.76	(.32)	-6.87	(.22)	-4.06	(.09)	6.64	(.18)		
A	-10.21	(.72)	-9.88	(.63)	-7.73	(.23)	-6.60	(.15)	-4.31	(.08)	5.48	(.10)	8.60	(.34)		
BBB	-8.32	(.35)	-7.90	(.29)	-6.35	(.15)	-4.38	(.09)	4.44	(.09)	7.43	(.23)	10.09	(.81)		
BB	-6.83	(.19)	-6.11	(.15)	-3.91	(.09)	4.10	(.09)	6.60	(.19)	8.41	(.42)	9.86	(.83)		
B	-4.94	(.10)	-3.96	(.08)	4.18	(.09)	6.29	(.17)	6.96	(.22)	8.67	(.47)	20.44	(5.58)		
CCC	-2.25	(.10)	3.29	(.14)	4.94	(.26)	5.73	(.37)	6.95	(.64)	8.79	(1.30)	9.65	(1.70)		

Covariance matrix Σ

	Posterior mean										Posterior standard deviation									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
1	.363	.063	.095	.018	.096	.083	.050	.060	.247	.091	(.11)	(.07)	(.09)	(.09)	(.09)	(.08)	(.11)	(.10)	(.12)	(.11)
2	.063	.269	.089	.001	.123	.057	.022	-.022	.065	.099	(.07)	(.08)	(.08)	(.08)	(.08)	(.07)	(.10)	(.09)	(.11)	(.10)
3	.095	.089	.365	-.020	.062	.059	.090	.035	.072	.169	(.09)	(.08)	(.12)	(.09)	(.09)	(.08)	(.11)	(.10)	(.12)	(.11)
4	.018	.001	-.020	.444	.029	.071	.024	.013	-.052	.056	(.09)	(.08)	(.09)	(.14)	(.09)	(.09)	(.11)	(.10)	(.12)	(.11)
5	.096	.123	.062	.029	.406	.076	.100	.020	.101	.051	(.09)	(.08)	(.09)	(.09)	(.13)	(.09)	(.11)	(.10)	(.12)	(.11)
6	.083	.057	.059	.071	.076	.339	.097	.048	.139	.092	(.08)	(.07)	(.08)	(.09)	(.09)	(.11)	(.11)	(.10)	(.11)	(.11)
7	.050	.022	.090	.024	.100	.097	.582	.149	.111	.203	(.11)	(.10)	(.11)	(.11)	(.11)	(.11)	(.15)	(.10)	(.12)	(.11)
8	.060	-.022	.035	.013	.020	.048	.149	.442	.076	.063	(.10)	(.09)	(.10)	(.10)	(.10)	(.10)	(.10)	(.14)	(.12)	(.11)
9	.247	.065	.072	-.052	.101	.139	.111	.076	.611	.053	(.12)	(.11)	(.12)	(.12)	(.12)	(.11)	(.12)	(.12)	(.16)	(.11)
10	.091	.099	.169	.056	.051	.092	.203	.063	.053	.536	(.11)	(.10)	(.11)	(.11)	(.11)	(.11)	(.11)	(.11)	(.11)	(.15)

Table 9: Posterior mean and standard deviation (in brackets) of **model (S2)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

k	Threshold parameters $\mu_{k,\ell}$										Remaining parameters		
	$\mu_{k,D}$	$\mu_{k,CCC}$	$\mu_{k,B}$	$\mu_{k,BB}$	$\mu_{k,BBB}$	$\mu_{k,A}$	$\mu_{k,AA}$	ϕ_k	$p(\phi_k < 0 D)$				
AAA	-28.86 (.78)	-22.46 (5.68)	-16.08 (5.06)	-8.67 (.78)	-7.65 (.53)	-6.18 (.29)	-3.83 (.15)	0.202 (.068)	0.003				
AA	-11.30 (1.37)	-9.80 (.81)	-8.11 (.41)	-7.68 (.34)	-6.79 (.24)	-3.99 (.13)	6.59 (.20)	0.215 (.038)	0.000				
A	-10.16 (.72)	-9.82 (.63)	-7.64 (.24)	-6.51 (.16)	-4.22 (.12)	5.45 (.14)	8.56 (.35)	0.214 (.032)	0.000				
BBB	-8.20 (.37)	-7.78 (.31)	-6.23 (.17)	-4.26 (.12)	4.44 (.12)	7.42 (.24)	10.08 (.80)	0.210 (.033)	0.000				
BB	-6.83 (.25)	-6.12 (.21)	-3.90 (.18)	4.29 (.19)	6.80 (.26)	8.63 (.47)	10.09 (.84)	0.334 (.046)	0.000				
B	-4.84 (.16)	-3.87 (.15)	4.31 (.16)	6.42 (.21)	7.08 (.26)	8.80 (.50)	20.62 (5.50)	0.277 (.040)	0.000				
CCC	-2.061 (.11)	3.28 (.14)	4.94 (.27)	5.73 (.39)	6.95 (.67)	8.82 (1.33)	9.82 (1.65)	0.148 (.041)	0.000				
								ω :	1.936 (.269)				
								α :	0.741 (.111)				

Table 10: Posterior mean and standard deviation (in brackets) of **model (S3)** with logit response function. Gibbs sampling with 2,000 iterations (burn-in 5,000).

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.73	2.04	0.17	0.03	0.02	0.00	0.00	0.00
AA	0.17	97.89	1.82	0.07	0.02	0.03	0.00	0.00
A	0.02	0.49	97.91	1.42	0.11	0.05	0.00	0.00
BBB	0.00	0.06	1.27	97.13	1.31	0.17	0.02	0.03
BB	0.01	0.02	0.12	1.54	95.70	2.33	0.15	0.14
B	0.00	0.02	0.08	0.09	1.30	96.01	1.54	0.96
CCC	0.01	0.01	0.09	0.24	0.41	3.03	84.24	11.97
D	0	0	0	0	0	0	0	100.00

Table 11: Unconditional quarterly migration matrix (%) of model (P3) with $x_t = 0$, cf. Table 4.

	AA	A	BBB	BB	B	CCC
AA	0.02	0.04	0.06	0.07	0.06	0.10
A		0.06	0.10	0.12	0.11	0.17
BBB			0.17	0.19	0.18	0.27
BB				0.21	0.20	0.31
B					0.18	0.29
CCC						0.45

Table 12: Joint quarterly upgrade-correlation matrix (%) of model (P3), cf. Table 4.

	AAA	AA	A	BBB	BB	B	CCC
AAA	0.28	0.26	0.23	0.23	0.30	0.29	0.59
AA		0.24	0.22	0.21	0.28	0.27	0.55
A			0.20	0.19	0.25	0.24	0.50
BBB				0.19	0.25	0.24	0.49
BB					0.32	0.31	0.64
B						0.31	0.62
CCC							1.27

Table 13: Joint quarterly downgrade-correlation matrix (%) of model (P3), cf. Table 4.

i	$\frac{1}{T} \sum_{t=1}^T \log(\text{CPO}_t^{(i)})$	$i \setminus j$	(P2)	(P3)	(K1)	(K2)	(K3)	(S1)	(S2)	(S3)
(P1)	-124.56	(P1)	34	60	60	76	74	79	78	78
(P2)	-124.46	(P2)		60	59	76	74	78	75	76
(P3)	-121.98	(P3)			41	72	62	78	75	78
(K1)	-121.73	(K1)				73	64	74	73	75
(K2)	-118.92	(K2)					20	57	54	55
(K3)	-119.26	(K3)						60	57	59
(S1)	-116.62	(S1)							40	36
(S2)	-116.74	(S2)								39
(S3)	-116.60									

Table 14: The leftmost part of the table shows the mean CPO-value, see Section 3.3; the rightmost part holds $\#\{t \in \{1, \dots, T\} : \text{CPO}_t^{(j)} > \text{CPO}_t^{(i)}\}$, where $\text{CPO}_t^{(i)}$ is the CPO under model i .