TI 2005-060/4
Tinbergen Institute Discussion Paper

A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk

Siem Jan Koopman[1,3,4]
André Lucas[2,3,4]
Robert Daniels[5]

[1] Department of Econometrics,
[2] Department of Finance and Financial Sector Management,
[3] Vrije Universiteit Amsterdam,
[4] Tinbergen Institute,
[5] De Nederlandsche Bank, Amsterdam.

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

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.
Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk*

This version: May 17, 2005

Siem Jan Koopman[‡§], **André Lucas**[¶§] **Robert J. Daniels**[†],

*Correspondence to: s.j.koopman@feweb.vu.nl, alucas@feweb.vu.nl, or r.j.o.daniels@dnb.nl.

[‡]Department of Econometrics, Vrije Universiteit Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, the Netherlands

[¶]Department of Finance and Financial Sector Management, ECO/FIN Vrije Universiteit Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, the Netherlands

[§]Tinbergen Institute Amsterdam, Roetersstraat 31, NL-1018 WB, Amsterdam, the Netherlands

[†]De Nederlandsche Bank, Postbus 98, NL-1000AB Amsterdam, The Netherlands

A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk

Abstract

We model 1981–2002 annual default frequencies for a panel of US firms in different rating and age classes from the Standard and Poor's database. The data is decomposed into a systematic and firm-specific risk component, where the systematic component reflects the general economic conditions and default climate. We have to cope with (i) the shared exposure of each age cohort and rating class to the same systematic risk factor; (ii) strongly non-Gaussian features of the individual time series; (iii) possible dynamics of an unobserved common risk factor; (iv) changing default probabilities over the age of the rating, and (v) missing observations. We propose a non-Gaussian multivariate state space model that deals with all of these issues simultaneously. The model is estimated using importance sampling techniques that have been modified to a multivariate setting. We show in a simulation study that such a multivariate approach improves the performance of the importance sampler.

Key words: credit risk; multivariate unobserved component models; importance sampling; non-Gaussian state space models.
JEL Codes: C32; G21.

1 Introduction

In this paper we propose a dynamic multivariate, non-Gaussian state space model to decompose panel time series of cohort and rating specific default hazard rates. In this way we disentangle differences in firm default behavior due to firm-specific causes from the effects due to macroeconomic conditions. The firm-specific effects may be correlated with observed heterogeneity, such as proxies for firm age or credit quality (ratings). We include such factors in our model. The common risk factor in the model is specified as a latent dynamic process and estimated explicitly. We test for the stability of the common risk factor's impact over time and across age classes.

The dynamic modeling of default rates is an important issue in the area of risk management, see, for example, the overview of Allen and Saunders (2003). First, there is significant empirical evidence that credit rating transition probabilities in general, and default probabilities in particular, vary over time, see Wilson (1997a,b), Nickell, Perraudin, and Varotto (2000), and Bangia, Diebold, Kronimus, Schagen, and Schuermann (2002). Specifically, default rates tend

to co-move with macro-economic conditions. This time variation can have a substantial impact on required capital buffers needed to cope with on-setting credit losses, see Bangia et al. (2002). Second, there is an increased liquidity on financial markets for credit risk related products, such as credit derivatives and asset backed securities (e.g., Collateralized Debt Obligations). This has led to a shift in the credit risk management process from an almost exclusive focus on counterparty screening and selection (using for example credit scoring models) at the loan contracting stage to a more active and dynamic management of credit risky portfolios. Dynamic credit risk management obviously requires adequate models for the dynamic behavior of default risk. Third, the proposed New Capital Accord (Basel Committee on Bank Supervision (2004)) allows banks to use their own internal estimates of counterparty default probabilities for computing capital requirements.[1] Dynamic default rate models can then help to predict these probabilities more accurately over different stages of the business cycle. As a result, capital buffers may be set more timely and efficiently by banks. In addition, predicted default probabilities may be closer to their empirical time-varying counterparts, a requirement that is explicitly tested under the New Capital Accord.

Several approaches have been suggested in the literature for modeling the dynamic behavior of default risk. Wilson (1997a,b) uses a logistic regression to relate default frequencies to a set of macro-economic variables. Nickell et al. (2000) and Bangia et al. (2002) compute rating transition probabilities based on Moody's or Standard and Poor's databases after first splitting their sample in expansion, neutral, and recession periods. For their sample splitting procedure, they use either observed GDP growth or NBER business cycle classifications, respectively. In this way, they end up with different default probabilities for different economic regimes. Formal testing reveals that the differences in probabilities between regimes are in some cases statistically significant. Kavvathas (2001) and Couderc, Renault, and Scaillet (2003) use a duration based modeling approach and also conclude that macro-economic variables such as the term structure of interest rates have a statistically significant impact on default intensities.

The basic framework underlying all of the previous papers is that credit risk is decomposed into a systematic, economy wide component, and a firm-specific component. The systematic component is captured by observed macro-economic variables, while the firm-specific component is modeled using distributional assumptions. In the present paper, we follow a similar line. The main difference with previous papers is that we model the systematic risk factor by an unobserved dynamic component, see Harvey (1989) and Durbin and Koopman (2001). Unob-

[1] There is a lively debate, also known as the pro-cyclicality debate, on whether risk sensitive weighting creates a feedback loop to the evolution of the business cycle itself, see Allen and Saunders (2003). For example, during downturns of the economy, default probabilities increase, thus increasing credit risk. This, in turn, results in higher capital requirements, making it more difficult for the bank to extend credit. The resulting credit rationing may increase default rates and credit risk even further, thus amplifying the speed and depth of the recession. We do not discuss this issue further in the present paper.

served components models may have an advantage over models based on observables. There is little theory as to which observables would be optimal or even correct as a proxy for systematic credit risk. Though the informal argument in favor of using business cycle indicators to model credit cycles is well known, there is both theoretical and empirical evidence that credit cycles may have their own separate dynamics, see for example Gorton and He (2003) and Koopman and Lucas (2005). The approach based on unobserved components gets around this issue by estimating the dynamics of any underlying systematic component directly from the data.

Obtaining data for credit risk analyses is generally a difficult issue, in particular if we are interested in the *dynamic* behavior of defaults. To determine the dynamics, the time series dimension of the data has to be sufficiently long. Internal bank data on defaults are usually proprietary and have a relatively short history. The closest relevant alternative is formed by default data obtained from the major rating agencies. Figure 1 shows the data we use in the present paper. Each sample year we form a new cohort of firms that received their initial rating in that year. We follow the default performance of each cohort for the next 15 years. This gives us a multivariate time series process as displayed in Figure 1.

<INSERT FIGURE 1 AROUND HERE>

The series in Figure 1 reveal several complicating factors. First, generally speaking, there is a similar pattern across all time series, peaking at around 1991 and 2002. This pattern is caused by common systematic or economy-wide risk factors. Second, there is significant variation across time series for different cohorts due to firm-specific risk factors. Third, the common risk factor appears to be correlated over time, as it should be if it is in some way related to business cycle conditions. Fourth, given the level of detail in the data, the default rate for specific cohorts in particular years may be zero. This problem is more pronounced for high ratings. The occurrence of zeros precludes the use of (inverse) logistic or probit type transformations for modeling the time series as in, for example, Wilson (1997a,b), and Koopman, Lucas, and Klaassen (2005). Finally, the cohort structure of the data intrinsically creates missing observations in the initial years of the samples. For example, in the first sample year we do not observe default hazard rates of the cohorts that started 2 years earlier or more.

To deal with all of these intricate econometric issues simultaneously is not straightforward. The main contribution of the present paper is to show that the problems can be overcome both conceptually and computationally by adopting a multivariate non-Gaussian time series perspective in state space representation. We apply this methodology empirically to the data shown in Figure 1. By the explicit multivariate nature of the model, we can introduce both firm-specific and common risk factors in a unified way. The general state space representation allows us to directly include a dynamic structure for the systematic risk factor. Moreover, by explicitly accounting for the discrete, non-Gaussian character of the default process, we can deal with the occurrence of zeros and small default frequencies in a natural way. Also the occurrence of missing observations is conveniently dealt with in the non-Gaussian state space

representation and the associated filtering and smoothing algorithms.

The generality of the model permits the development of tests for hypotheses related to aging effects in default probabilities as in Asquith, Mullins, and Wolff (1989), Altman (1983,1989), and Altman and Suggitt (2000), default rate dynamics as in Nickell et al. (2000), Bangia et al. (2002), Koopman and Lucas (2005) and Koopman et al. (2005), and aging effects in default correlations over longer holding periods, see in De Servigny and Renault (2003). Our empirical results in Section 4 show that allowing for age effects and sticky common factor risk dynamics is empirically relevant.

The generality of the model comes at the cost of complications for likelihood evaluation. In particular, due to the presence of both firm-specific stochastic factors and a dynamic unobserved systematic risk factor, we have to resort to Monte Carlo maximum likelihood methods for parameter estimation. Convenient and flexible algorithms and software to perform such computations, however, are nowadays readily available, see Shephard and Pitt (1997), Durbin and Koopman (1997, 2001) and other references given in Section 2. However, in this paper we develop the methodology further for a likelihood-based analysis of a multivariate non-Gaussian time series with possibly panel structure.

McNeil and Wendin (2004) in a recent paper build on a similar framework of generalized mixed linear models. Both their method of estimation and the nature of their empirical application, however, are substantially different from this paper. With regard to estimation, Wendin and McNeil use a Bayesian estimation perspective. By contrast, we implement a simulated maximum likelihood methodology based on efficient importance samplers. As mentioned before, this allows us to address multiple intricate econometric issues simultaneously, including missing observations. It is less clear how the latter issue is or could be dealt with naturally in the framework of McNeil and Wendin (2004). Note that missing values are an important ingredient of our empirical data shown in Figure 1. With regard to the empirical application, we test for a much richer set of model dynamics and parameter shifts than in the Wendin and McNeil paper. Our empirical results differ in that we find much more persistent dynamics of the common risk factor.

The paper is set up as follows. In section 2 we present the economic and the statistical model. The statistical properties of the model and the performance of the estimation procedures are studied in a thorough simulation experiment in Section 3. Section 4 presents our empirical estimation results. Section 5 concludes, while the Appendix gathers the details on the estimation and smoothing methodology used in this paper.

2 Modeling defaults

2.1 The economic model

The basic structure of our model is derived from the standard decomposition of default risk into systematic and firm-specific components, see the seminal framework of Gupton, Finger, and Bhatia (1997), which also underlies the new regulatory framework of the Basel Committee on Bank Supervision (2004). We consider a surplus variable S_{it} for each firm i at time t for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Default occurs if the surplus variable falls below a default threshold c_{it}. The default threshold may be firm-specific due to differences in credit quality between firms, as well as time varying because of fluctuating macroeconomic conditions. To allow for dependence among the different firms, we assume that the surplus variable S_{it} is driven by a systematic risk factor f_t and a firm-specific risk factor x_{it},

$$S_{it} = \rho_i f_t + \sqrt{1 - \rho_i^2} x_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{1}$$

compare Gupton et al. (1997). The firm-specific factors x_{it} are serially and cross-sectionally independent conditional on the systematic factor f_t. The parameter ρ_i in (1) determines the relative weights of the systematic and firm-specific risk component in the default mechanism. Its value is strictly between 0 and 1. For $\rho_i = 0$, default is entirely triggered by firm-specific factors. By contrast, for $\rho_i = 1$, default is completely driven by systematic risk. The weighting scheme in (1) is derived from Gupton et al. (1997) and is needed to ensure identification of the final model below. Identification is an issue, because S_{it} is not observed directly. For the same reasons it is assumed that the variances of f_t and x_{it} need to be restricted. For example, in the well-known CreditMetrics model for credit risk management, both f_t and x_{it} are assumed to be standard normally distributed. Note that if the variances of f_t and x_{it} are unity, the correlation between S_{it} and $S_{i't}$ for $i \neq i'$ equals $\rho_i \rho_{i'}$, or ρ_i^2 for homogeneous entities $\rho_i = \rho_{i'}$. As is common in the literature, we refer to the parameter ρ_i^2 as the asset correlation.

Instead of focusing on individual defaults, we concentrate on the aggregate default level within a cohort of firms that were initially rated in a particular year. A cohort is indicated by an index j, indicating the year in which the cohort started, where j_i is the cohort index for firm i. These cohorts are observed for different initial rating classes. Rating classes are indicated by the index s, where s represents the credit quality label assigned by the rating agency at the time when the firm was rated for the first time. The initial rating of firm i is denoted by s_i. Model (1) has too many parameters for the data at hand, see Figure 1. To reduce the number of parameters, we impose the following restrictions. First, we restrict the asset correlation parameter ρ_i^2 to be the same within a rating class, i.e., $\rho_i^2 = \bar{\rho}_{s_i}^2$. This restriction is usually imposed in empirical work on defaults due to the sparse nature of data, see for example Gordy (2000), Dietsch and Petey (2004) and Cowan and Cowan (2004). Second, we restrict the default thresholds c_i to depend only on the rating class s and the time since initial rating $t - j$,

i.e., $c_i = \bar{c}_{s_i,t-j_i}$. Dependence on the rating class is standard in credit risk research, as default probabilities vary significantly between rating classes, see for example Gupton, Finger, and Bhatia (1997). We also allow for age dependent default probabilities, where age is interpreted as the time since the company was rated for the first time. The inclusion of age effects in default probabilities is important, see for example Asquith, Mullins, and Wolff (1989), Altman (1983,1989), and Altman and Suggitt (2000). In particular, bankruptcy rates may be higher during the initial stages of a firm's life. By allowing the default threshold to vary with $t-j$, we can test explicitly for the existence of age effects in our data. It should be kept in mind, however, that $t-j$ is only a proxy for the actual age of the issuer. A firm can be founded a long time before the firm receives its initial rating.

Given the above restrictions, we can proceed to specify the economic model. Let k_{sjt} denote the number of firms of rating class s and cohort j still alive at the start of period t, and let y_{sjt} denote the number of defaulting firms in the tth period. Then y_{sjt} can be regarded as a realization of a binomial distribution conditional on the state of the systematic risk factor f_t, i.e.,

$$y_{sjt}|f_t \sim \text{Binomial}\left(k_{sjt}, \pi_{sjt}\right), \quad s = 1,\ldots,S, \quad j = 1,\ldots,J, \quad t = 1,\ldots,T, \tag{2}$$

where π_{sjt} is the conditional probability of default. From (1) and the restrictions imposed, it follows that

$$\pi_{sjt} = \Pr\left(x_{it} \leq \frac{\bar{c}_{s,t-j} - \bar{\rho}_s f_t}{\sqrt{1-\bar{\rho}_s^2}}\right) = F\left(\frac{\bar{c}_{s,t-j} - \bar{\rho}_s f_t}{\sqrt{1-\bar{\rho}_s^2}}\right) \tag{3}$$

where $F(\cdot)$ is a cumulative distribution function of the firm-specific risk factor. The default probability can now vary between initial rating categories s, age values $t-j$, and stages of the economy f_t. The economic process f_t can evolve slowly over time. Given that we have annual data for roughly a 20-year time span, we assume that f_t follows an autoregressive process of order one,

$$f_t = \phi f_{t-1} + \sqrt{1-\phi^2}\eta_t, \tag{4}$$

where η_t is standard normal and with $\phi \in [0,1]$. Extensions to more general processes for f_t is straightforward.[2] Note that the process f_t is the same across ratings and cohorts. We can thus use information from different rating and cohort default frequencies to improve our estimates of current systematic risk conditions. The parameterization in (4) imposes a unit (unconditional) variance restriction on f_t, because the variance of η_t and the asset correlation $\bar{\rho}_s^2$ cannot be identified simultaneously.

2.2 The statistical model

The unknown parameters of the model in Section 2.1 are ϕ, $\bar{c}_{s,t-j}$ and $\bar{\rho}_s$ for $s = 1,\ldots,S$ and $t-j = 0,\ldots,T-1$. To estimate the parameters and to measure the systematic risk factor f_t

[2]In the empirical application, for example, we also perform a robustness check using an AR(2).

(based on the other parameter estimates), we develop a statistical model that corresponds to the economic model and that allows for the numerical maximization of the likelihood function. The statistical model also provides the means for estimating the unobserved autoregressive process f_t (signal extraction). We should note that standard linear estimation methods cannot be applied since the model is non-Gaussian and depends on a dynamic latent process f_t. Monte Carlo simulation techniques are carefully designed and employed to carry out the task of estimation in a computationally efficient way. This is a challenging task since the dimensions of the integrals involved are high. In the empirical section we consider a set of models introduced in Section 2.1 with $S = 2$, $J = 15$ and $T = 21$ resulting in 21 observations of a 30×1 vector. Computations with dimensions up to 60 also proved computationally very feasible.

We start with (2) and consider the conditional log-density function

$$\log p(y_{sjt}|\pi_{sjt}) = y_{sjt}\{\log \pi_{sjt} - \log(1-\pi_{sjt})\} + k_{sjt}\log(1-\pi_{sjt}) + \log \left(\begin{array}{c} k_{sjt} \\ y_{sjt} \end{array} \right), \qquad (5)$$

with $y_{sjt} = 0, \ldots, k_{sjt}$. As the probability π_{sjt} is bounded between zero and one, we introduce the logit transformation

$$\theta_{sjt} = \log\{\pi_{sjt} \ / \ (1-\pi_{sjt})\} \qquad (6)$$

such that $F(\cdot)$ in (3) is effectively the logistic distribution. This leads to the following convenient expression for the log-density of y_{sjt},

$$\log p(y_{sjt}|\theta_{sjt}) = y_{sjt}\theta_{sjt} - k_{sjt}\log(1+\exp(\theta_{sjt})) + \log \left(\begin{array}{c} k_{sjt} \\ y_{sjt} \end{array} \right), \qquad (7)$$

from which it follows immediately that the binomial density belongs to the exponential family. The model for the signal θ_{sjt} is specified as

$$\theta_{sjt} = \lambda_{s,t-j} - \beta_s f_t, \qquad f_t = \phi f_{t-1} + \sqrt{1-\phi^2}\eta_t, \qquad (8)$$

where f_t is the systematic risk factor in the surplus variable S_{it} in (1). Equation (8) can be related to the model parameters in Section 2.1 by

$$\bar{\rho}_s^2 = \beta_s^2 \ / \ (1+\beta_s^2), \qquad \bar{c}_{s,t-j} = \lambda_{s,t-j} \Big/ \sqrt{1+\beta_s^2}. \qquad (9)$$

The unobserved, latent process f_t is modeled by the stationary autoregressive (AR) process (5). Other stationary and non-stationary processes for f_t can be considered in this context as well.

The multivariate defaults model can be represented as a non-Gaussian panel state space model with observation equation

$$y_{sjt} \sim p(y_{sjt}|\theta_{sjt}) = \text{Binomial}\left(k_{sjt}, \pi_{sjt}\right), \qquad (10)$$

where $\pi_{sjt} = (1 + \exp(-\theta_{sjt}))^{-1}$ and θ_{sjt} is given by (8). The dynamic latent variable f_t and the fixed effects $\lambda_{s,t-j}$ are placed in a so-called state vector α_t that can be modeled by the linear time series process

$$\alpha_{t+1} = T_t \alpha_t + R_t \xi_t, \qquad \xi_t \sim \text{NID}(0, Q_t), \tag{11}$$

for $t = 1, \ldots, T$ with $\alpha_1 \sim N(a, P)$. The vector a and the matrices T_t, R_t, Q_t and P are fixed and known matrices for $t = 1, \ldots, T$. The signal is then given by

$$\theta_{sjt} = Z_{sjt} \alpha_t, \tag{12}$$

where Z_{sjt} is a deterministic selection vector for $s = 1, \ldots, S$, $j = 1, \ldots, J$ and $t = 1, \ldots, T$. For example, for $S = 1$ and each cohort followed for 15 years, we set $\alpha_t = (\lambda_{1,1}, \ldots, \lambda_{1,15}, f_t)'$, and obtain $Q_t = I_{16}$,

$$T_t = \begin{pmatrix} I_{15} & 0 \\ 0 & \phi \end{pmatrix}, \qquad R_t = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{1-\phi^2} \end{pmatrix}, \qquad (Z'_{1,1,t}, \ldots, Z'_{1,15,t})' = (I_{15}, -\beta_1 \iota),$$

with ι a column vector of ones.

The joint model (10) and (11) is a special case of a nonlinear and non-Gaussian state space model as discussed in part II of Durbin and Koopman (2001). We should note that fixed effects can be part of the state vector α_t and they can be estimated simultaneously with the unobservable f_t. The estimation of the state vector with $\lambda_{s,t-j}$ and f_t is discussed in section 2.4. In the next section we focus on the estimation of the parameters ϕ and β_s for $s = 1, \ldots, S$ by Monte Carlo maximum likelihood.

2.3 Importance sampling estimation

Importance sampling techniques are used for various purposes in the statistical and econometric literature, see, for example, Kloek and Van Dijk (1978) and Geweke (1989). Recently, these techniques are also considered for the estimation of nonlinear and non-Gaussian time series models. Different variants of importance sampling for time series models have been developed, for example, see Danielsson and Richard (1993), Shephard and Pitt (1997) and Durbin and Koopman (1997). We follow the approach of Durbin and Koopman (2001, Part II). The Appendix explores this methodology for an univariate binomial time series model. In this section we generalize this methodology for the multivariate (panel) binomial time series model discussed in subsection 2.2.

Consider model (10) and (11) for which the likelihood function is given by

$$\begin{aligned} p(y) = \prod_{s,j} p(y_{sj1}, \ldots, y_{sjT}) &= \prod_{s,j} \int \left\{ \prod_{t=1}^{T} p(y_{sjt}|\alpha) \right\} p(\alpha)\, \mathrm{d}\alpha \\ &= \int \left\{ \prod_{t=1}^{T} \prod_{s,j} p(y_{sjt}|\alpha_t) \right\} p(\alpha)\, \mathrm{d}\alpha \end{aligned} \tag{13}$$

where T is the time series dimension and α denotes $(\alpha_1', \ldots, \alpha_T')'$. It follows from (12) that θ_{sjt} is a linear function of α_t and therefore $p(y_{sjt}|\alpha_t) = p(y_{sjt}|\theta_{sjt})$ is given by (7). An analytical expression for the high dimensional integral in (13) is not available in closed form. We therefore rely on numerical techniques for the evaluation of (13). Here we approximate the likelihood function by Monte Carlo integration based on efficient sampling methods.

In the Appendix the details of importance sampling for a univariate binomial time series model is presented. We adopt this approach below but extend the methodology for a multivariate panel of binomial time series. Importance sampling requires a linear Gaussian approximating model with corresponding density $p_G(y, \alpha)$ for which it holds that $p_G(y, \alpha) = p_G(y|\alpha)p(\alpha)$ since $p(\alpha) \equiv p_G(\alpha)$. On the basis of the linear Gaussian model we simulate α conditional on y using a simulation smoothing algorithm. This is a computationally efficient way to sample from $p_G(\alpha|y)$. In the case of the multivariate model of subsection 2.2, a linear Gaussian model should be considered for all observations, that is $y = (y_{111}, \ldots, y_{SJT})'$. The obvious multivariate extension of the linear Gaussian "approximating" model (25) of the Appendix is given by

$$\widetilde{y}_{sjt} = \theta_{sjt} + u_{sjt}, \qquad u_{sjt} \sim \text{NID}(0, H_{sjt}), \quad s = 1, \ldots, S, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T, \quad (14)$$

with $\theta_{sjt} = Z_{sjt}\alpha_t$ and where the general model for α_t is given by (11). The variables $\widetilde{y}_{sjt}$ and H_{sjt} are functions of y_{sjt} and the mode of state vector α_t, that is

$$\widetilde{y}_{sjt} = \theta_{sjt} + H_{sjt}y_{sjt} - (1 + \exp\theta_{sjt}), \qquad H_{sjt} = (1 + \exp(\theta_{sjt}))^2 \exp -\theta_{sjt}. \quad (15)$$

where $\theta_{sjt} = Z_{sjt}\alpha_t$ is replaced by its mode. The estimation of the mode is described in the Appendix and is based on linearization and maximization. The process consists of iteratively evaluating the conditional mean of α_t for the model (14) with $\widetilde{y}_{sjt}$ and H_{sjt} given by (15) but with α_t replaced by the current guess of its mode. The conditional mean of α_t is computed by the multivariate version of the Kalman filter and smoothing algorithm applied to the multivariate linear Gaussian time series model

$$\widetilde{y}_t^* = Z_t^*\alpha_t + u_t^*, \qquad u_t^* \sim N(0, H_t^*), \quad (16)$$

where $\widetilde{y}_t^*$ and u_t^* are stacked elements (column-wise) $\widetilde{y}_{sjt}$ and u_{sjt}, respectively, for the indices s and j, while Z_t^* and H_t^* are diagonal matrices with the diagonal elements Z_{sjt} and H_{sjt}, respectively. A computationally efficient implementation of the Kalman filter and smoothing algorithm is used to handle the high-dimensionality of $\widetilde{y}_t^*$'s in large panels, see Durbin and Koopman (2001, section 6.3). The resulting smoothed estimate of α_t from model (16) is taken as the new guess of the mode. The iterative process is initialized with an appropriate starting value for the mode. When the iterative process has converged, we adopt the model (14) as the basis for importance sampling, where (15) is evaluated at the mode. In this case we effectively sample α_t around its mode and with a curvature implied by model (14). The importance density is therefore $p_G(\alpha_t|y)$ for $t = 1, \ldots, n$.

The likelihood function (13) is equivalent to

$$p(y) = \int \left\{ \prod_{t=1}^{T} \prod_{s,j} p(y_{sjt}|\alpha_t) \right\} \frac{p(\alpha)}{p_G(\alpha|y)} p_G(\alpha|y)\, \mathrm{d}\alpha. \tag{17}$$

We construct the Monte Carlo estimator of the likelihood function by

$$\widetilde{p}(y) = \frac{1}{M} \sum_{m=1}^{M} \left\{ \prod_{t=1}^{T} \prod_{s,j} p(y_{sjt}|\alpha_t^m) \right\} \frac{p(\alpha^m)}{p_G(\alpha^m|y)}, \quad \text{with} \quad \alpha^m = \begin{pmatrix} \alpha_1^m \\ \vdots \\ \alpha_T^m \end{pmatrix}, \tag{18}$$

where M is the number of Monte Carlo simulations and α_τ^m is the mth simulated path based on the importance sampler $p_G(\alpha_\tau|y)$. These simulations are computed by the simulation smoother algorithm of Durbin and Koopman (2002). The Monte Carlo estimator can be simplified to

$$\widetilde{p}(y) = \frac{p_G(y)}{M} \sum_{m=1}^{M} \prod_{t=1}^{T} w(\alpha_t^m), \quad \text{where} \quad w(\alpha_t) = \prod_{s,j} \frac{p(y_{sjt}|\alpha_t)}{p_G(y_{sjt}|\alpha_t)}, \tag{19}$$

since $p_G(\alpha|y) = p_G(y)p_G(y|\alpha)p(\alpha)$ and $p_G(y|\alpha) = \prod_t \prod_{s,j} p_G(y_{sjt}|\alpha_t)$. Further note that $p_G(y)$ is effectively the Gaussian likelihood function of the approximating linear Gaussian model (16) that is computed by the Kalman filter via the prediction error decomposition, see Harvey (1989).

2.4 Estimation of state vector

The state vector α_t contains unobserved time series components and fixed effects. In the case of model (10) and (11). The state vector includes the latent autoregressive process f_t and the constants $\lambda_{s,t-j}$. The estimator of the state vector at time τ is the expected value of the state vector α_τ conditional on the data vector y and can be expressed as

$$\begin{aligned}
\widehat{\alpha}_\tau &= \int \alpha_\tau p(\alpha|y)\, \mathrm{d}\alpha \\
&= p(y)^{-1} \int \alpha_\tau p(y|\alpha) p(\alpha)\, \mathrm{d}\alpha \\
&= p(y)^{-1} \int \alpha_\tau \left\{ \prod_{t=1}^{T} \prod_{s,j} p(y_{sjt}|\alpha_t) \right\} p(\alpha)\, \mathrm{d}\alpha \\
&= p_G(y) p(y)^{-1} \int \alpha_\tau \prod_{t=1}^{T} w(\alpha_t) p_G(\alpha_t|y)\, \mathrm{d}\alpha
\end{aligned} \tag{20}$$

where density $p(y)$ is given by (13) and $w(\alpha_t)$ is given by (19). The scaling constant $p(y)$ in (20) can be estimated by (19) using the importance sampling method of the previous section. Further, by considering (17) and (19), an obvious importance sampling estimator for $\widehat{\alpha}_\tau$ is given by

$$\widetilde{\widehat{\alpha}}_\tau = \sum_{m=1}^{M} \alpha_\tau^m \prod_{t=1}^{T} w(\alpha_t^m) \Bigg/ \sum_{m=1}^{M} \prod_{t=1}^{T} w(\alpha_t^m), \qquad s = 1, \ldots, n \tag{21}$$

where α_τ^m is generated from the importance density $p_G(\alpha_\tau|y)$ for $\tau = 1, \ldots, n$. The estimation of the state vector is particularly interesting for the signal extraction of f_t and the estimation of $\lambda_{s,t-j}$ at the maximum likelihood estimate of the parameters.

3 Simulation experiment

3.1 Simulation design

To investigate the performance of the estimation methodology for datasets with a relatively small time series dimension, we have carried out a simulation study. The focus has been on the estimation of parameters that determine the dynamic characteristics of the data. We consider the multivariate Binomial state space model (2) for one rating class s, i.e., $S = 1$. The number of cohorts J varies from 1 to 16. For $J = 1$, the model reduces to a univariate Binomial time series model with signal $\theta_t = \lambda - \beta f_t$ and autoregressive process $f_t = \phi f_{t-1} + \sqrt{1-\phi^2}\eta_t$. For $J > 1$, a panel of J Binomial time series are treated simultaneously for the estimation of the common coefficient ϕ. We consider a model where each β_j, $j = 1, \ldots, J$, is estimated freely, and a 'pooled' model where $\beta_j \equiv \beta$. The constant λ is a scaling parameter and does not influence the dynamic properties of y_{sjt}, but only its level. Therefore, we do not consider it further in this simulation study. Finally, the sample sizes considered are $T = 20$, $T = 40$ and $T = 80$. The sample size $T = 20$ is of particular interest since this is close to the time series length considered in our empirical study in the next section. The results for $T = 20$ are then contrasted with the results for the larger sample sizes of $T = 40$ and $T = 80$.

The true data generation process has parameter values $\phi = 0.8$ and $\beta_j \equiv \beta = 0.6$. These values represent typical parameter values that are found in empirical work, cf. the empirical results in the next section. During the estimation, ϕ is restricted to be in the unit interval by employing a logistic transformation on the parameter. The simulation study is based on $1,000$ panel time series generated from the true model specification and for the three different sample sizes. The parameter vector is estimated for each generated sample using the importance sampling methods described in the previous section.

3.2 Simulation results

In Figure 2 the simulation results are presented graphically in a straightforward way. For a particular model with panel size 1 or 4 and for three different sample sizes $T = 20, 40, 80$, the $1,000$ estimates of ϕ and β_1 are presented as histograms. For the univariate model ($J = 1$) with sample size $T = 20$, the histogram of ϕ is bimodal, revealing a peak at the true value of $\phi = 0.8$, and a second peak at $\phi = 1.0$. The second peak is due to our parameterization $\phi \in [0, 1]$ and signals that for very short time series, the estimation procedure sometimes has

difficulty in distinguishing the correct stationary value of $\phi = 0.8$ from a unit root. The problem disappears if we extend the length of the sample to $T = 40$ or $T = 80$. The distributions for these sample sizes are unimodal again, peaking at the correct true value of $\phi = 0.8$.

The estimates of β_1 are presented in the bottom panel of Figure 2. For the models with $J > 1$ with the β_js estimated freely, the histograms for the estimates of β_j for $j = 2, \ldots, J$ are very similar to the one shown for β_1, and therefore omitted. We observe that for all considered choices of J and T the histograms are unimodal and that the bias is relatively small. Note that the Monte Carlo variation of the β estimates reduces mildly when J increases from 1 to 4, or significantly when the time series length increases. The latter effect is obvious. The former effect is due to the fact that we obtain multiple measurements of the same underlying signal f_t.

3.3 Diagnostic tests for the behavior of the importance sampler

The method of importance sampling provides effective means for the evaluation of the likelihood function and the estimation of the latent processes in the model. Geweke (1989) argues that importance sampling should only be used if the variance of the importance weights exists. In this case the central limit theorem applies and $\sqrt{M}(\bar{w} - w) \to N(0, \sigma^2)$ as $M \to \infty$, where w represents the ratio in (19), $\bar{w}$ is the sample mean of a series of M generated weights based on the importance draws α^m, and σ^2 is the (theoretical) variance of the stochastic function w.

In the case of non-Gaussian state space models, it is intractable to prove the existence of σ^2 analytically. Monahan (1993) and Koopman and Shephard (2004) have therefore developed diagnostic statistics for testing the existence of a variance for w empirically. To do this, we simulate a panel of multi-year defaults using the same set-up and parameter settings as before. Based on the simulated panel, maximum likelihood estimates of the parameters are obtained. These estimates are used to obtain a large number, say $100,000$, of importance sampling weights. We then examine the right-hand tail of the empirical distribution of these weights. In order to investigate the existence of second order moments for the importance sampler that is used in this paper, we have applied the tests of Monahan (1993) and Koopman and Shephard (2004) for different sample and panel sizes. In all cases considered, the statistics indicated that the weight function has a variance.

Further evidence of a regular weighting function is obtained by using some simple extreme value diagnostics, see Figure 3. These indicate how the performance of the importance sampler varies with the dimensions of the panel. We approximate the 50 largest tail observations of the 100,000 samples by a Pareto distribution with tail index γ,

$$1 - F(w) = w^{-\gamma}, \quad w \geq 1,$$

for $w = \tilde{w}_i/\tilde{w}_0$, $i = 1, \ldots, 50$, and $\tilde{w}_i$ the sorted largest sampling weights, $\tilde{w}_{50} \geq \tilde{w}_{49} \geq \ldots \geq \tilde{w}_0$. The Hill (or maximum likelihood) estimator for γ is also presented in Figure 3. The graphs

in Figure 3 plot the logarithm of one minus the empirical distribution function of w versus the logarithm of w itself. If the Pareto assumption holds, we would expect a line with slope coefficient $-\gamma$, see Embrechts et al. (1997). For the purpose of reference we also plot the line for $\gamma = 2$ (for which second moments do not exist). We see that the distributions' tails are generally steeper than the reference line. This is to be expected given the testing results mentioned earlier indicating that the variances exist.

<INSERT FIGURE 3 AROUND HERE>

It is interesting to observe the pattern in the weights over the different panel dimensions J and time series dimensions T. For $T = 20$ and $T = 40$, it is clear that the behavior of the sampling weights improves when the cross sectional dimension J increases. In particular, the average value of $\hat{\gamma}$ increases, indicating lighter tails and the existence of higher order moments as a result. We may conclude from these graphical displays that by increasing the panel size, importance sampling becomes even more reliable. This conclusion does not apply to the time series length. It seems that for shorter samples importance sampling is more stable. However, we should emphasize that in all cases the diagnostic tests indicate that the variance exists. Therefore we may conclude that the central limit theorem for the Monte Carlo likelihood estimator is likely to apply for this class of models.

Overall the Monte Carlo results are satisfactory. The performance of the estimation procedure becomes more reliable when more observations become available, in particular when this is due to an increase in the panel dimension. Finally we have shown that the estimation procedure is reliable even though it strongly depends on many computations and simulations.

4 Empirical results

In this section we apply our model to annual default data from the CreditPro6.2 database of Standard and Poor's. For each sample year from 1981 up to 2002, we construct cohorts of companies that were rated for the first time in that year and received a rating in a specific rating class. In our current application, we restrict ourselves to the broad classes of investment grade and subinvestment grade ratings. The methodology allows for extensions to more refined rating classes in a relatively straightforward manner. After the construction of each cohort, its default performance is followed over the subsequent 15 years. Note that the model needs two ingredients for each cohort, namely the number of defaults in a specific year (y_{sjt}) and the number of companies (in the cohort) at risk during the year (k_{sjt}). The number of companies at risk decreases over time due to two reasons. First, defaults obviously deplete the cohort size. Second, ratings may be withdrawn. To account for possible endogeneity biases and strategic behavior of firms withdrawing their ratings around default, the database records defaults even after ratings have been withdrawn. A graph of the data was provided in Figure 1.

We estimate a number of different models for each of the separate ratings. As can be seen

from Figure 1, the data are highly non-Gaussian. There is a significant number of zero values for the observed hazard rates. This precludes estimation of a dynamic model using an inverse probability transform of the hazard rates as in Koopman and Lucas (2005) or Koopman, Lucas, and Klaassen (2005). For example, the investment grade rating has many observed hazard rates equal to zero. Note that for the investment grade issuers, estimating the dynamics of the common risk factor is a challenging task given the limited number of defaults. By comparing the left-hand and right-hand graph in Figure 1, however, it is obvious that we can benefit from pooling. High default hazard rates for investment and subinvestment grade companies appear to coincide, giving a clearer signal on the value of the common underlying (macro) default climate f_t. Therefore, we also estimate all models using a system where all cohorts of both ratings are stacked into a simultaneous system that is subject to a single common risk factor f_t.

For a completely unrestricted estimation of the default thresholds λ_{t-j}, at least a few defaults are needed for each cohort in order to estimate the thresholds reliably. This difficulty may become problematic at longer horizons. We solve this problem by smoothing the thresholds over the different cohorts using a cubic spline structure on the aging effect. For this purpose, the interpolating spline methodology of Poirier (1976) is adopted. In particular, we put spline knots at the ages of 1, 3, 7, 11, and 15 years, thus allowing for much flexibility in the term structure of default probabilities, while at the same time reducing the number of parameters. The spline coefficients can be put into the state vector in the same way as the original thresholds λ_{t-j}. Estimates and standard errors are then obtained through the importance sampling methodology. The dimensionality of the estimation problem consisting of a non-linear search over the remaining parameters is then reduced significantly.

The estimation and testing results are presented in Tables 1 through 4. Tables 1 through 3 present the test results for the different model comparisons. Table 4 contains the parameter estimates and standard errors.

Table 1 shows that default probabilities vary with the age of the rating for different rating categories and different dynamic specifications for the common risk factor f_t. The effect is shown in Figure 4 for the AR(1) specification. The left-hand graph shows the age effect for investment grade ratings. Default probabilities show an increasing pattern for roughly the first 10 years. Then they level off and finally decrease. The subinvestment grade ratings displayed in the right-hand graph also reveal the familiar hump-shaped pattern. Default probabilities increase during the first 4 years after the firm's initial rating. Then they sharply decrease up to a rating age of 8 to 9 years, after which they level off and slightly decrease. The estimated patterns for the stacked rating categories is very similar and therefore omitted. Note that the p-values in Table 1 for the investment grade rating are highest. This can be explained by comparing Figure 1 and 4. The default events for investment grade companies are so scarce that identifying all spline parameters of the default curve from the available data is difficult.

Further parsimony could be imposed by reducing the number of spline knots to describe the aging effects for investment grade companies.

<INSERT TABLE 1 AND FIGURE 4 AROUND HERE>

Given the results from Table 1, all remaining models make use of aging default probabilities. Table 2 concentrates on the dynamic specification of the common risk component f_t. We take the modeling approach from 'specific to general', starting from a model without a common risk component, and then extending the model specification with a risk factor modeled as an independent sequence (i.i.d.), an AR(1) process, and an AR(2) process. The table shows that the estimation results are consistent for the different rating categories. There is clear statistical evidence for the presence of a common factor. The estimated factor shows persistence over time, as the i.i.d. specification is clearly rejected under the AR(1) alternative. The evidence for further dynamics is weaker. All ratings show that the AR(1) can be rejected under the AR(2) alternative at the 10% level, but not at the 5% level. Given the short time span of the data, it is not very surprising that it is hard to distinguish between these two dynamic specifications. This is confirmed by the results of the Monte Carlo study in the previous section. More data in the time series dimension is needed to answer this question in a satisfactory way. Note that the stickiness in the common factor process f_t characterizes a different data feature than the aging effect in default probabilities. The former is identified from the time series dimension of the data, whereas the latter is identified from the cross-sectional data properties across time points. The current model specification and the Kalman filter methods allow us to disentangle these different effects in one integrated estimation procedure. Interestingly, the AR(2) parameters reveal that the factor f_t may contain a cyclical component. This is discussed in more detail after the presentation of the estimation results in Table 4.

<INSERT TABLE 2 AROUND HERE>

We now focus on the final set of tests, checking for increases or decreases in the common factor loadings β_s in (3), either across time or across age. The tests are inspired by the controversy in the literature on whether default correlations or asset correlations change with the age of the company, and whether they should change if we predict defaults over longer horizons. De Servigny and Renault (2003) provide non-parametric estimates based on similar data as ours showing that default correlations increase with the holding period of the issuer's bonds. In contrast to previous papers like De Servigny and Renault (2003) and the references therein, our model based testing approach explicitly allows for the presence of common factor dependence with possibly sticky time series dynamics. The test is based on specifying the common factor loading as $\beta_s + \nu_{1s} \cdot (t-j) + \nu_{2s} \cdot t$. An age shift in correlations is then tested as $H_0 : \nu_{1s} = \nu_{2s} = 0$ versus the alternative $H_1 : \nu_{1s} \neq 0, \nu_{2s} = 0$. Similarly, the test for a time shift in correlations is performed using the same null hypothesis against the alternative $H_1 : \nu_{1s} = 0, \nu_{2s} \neq 0$. Table 3 clearly shows that there is no significant evidence in the data for either type of shift. Note that this test requires much of the available data. Both the time series

dynamics of f_t and the shifts in loadings can only be identified from the time series dimension of the data, which in our case is only 21 years. A longer sample period would allow us to obtain a more precise insight into the age change in correlations. Also note that the current results can still be reconciled with the changing default correlations found by De Servigny and Renault (2003). This is due to the presence of the AR dynamics of the common risk factor. Our estimate of β indicates the magnitude of the 'unconditional' asset correlations. The conditional, e.g. one-year, asset correlations, are much smaller, as they should be multiplied by $\sqrt{1-\phi_1^2}$.

<INSERT TABLE 3 AROUND HERE>

Table 4 contains the parameter estimates of the different models. The first thing to notice is that confidence intervals are very large for all parameters. This is due to the limited time span of the data, and the fact that the parameters displayed in Table 4 are mainly identified from the time series dimension. Except for the model with time shifts, the estimated factor loadings β_s for the different ratings are stable between the single rating models (I and S) and the stacked model (IS). As we have already seen, the models with loading shifts over age or time appear to be overparameterized, resulting in instable parameter estimates. The factor loadings for investment grade (β_I) and subinvestment grade (β_S) companies are generally very close for the current data set.

<INSERT TABLE 4 AROUND HERE>

The first order AR parameter is estimated at around 0.8 to 0.9, indicating strong persistence. This is in line with the persistence generally found in macro-economic time series, but substantially larger than the value of 0.5 found by McNeil and Wendin (2004). Interestingly, the AR(2) specification results in an AR polynomial with complex roots. The roots imply a (stochastic) cyclical behavior with a period of 9 to 10 years, and a persistence parameter of 0.8. The persistence parameter is in line with the size of ϕ_1 for the AR(1) specification. The length of the cycle is longer than that of typical business cycles, with an average length of 5 to 6 years for post-war data. Typical business cycle estimates are obtained from filters such as the one of Baxter and King (1999) and aim to extract business cycle frequencies ranging from 1.5 to 8 years. The period found empirically here is in line with earlier estimates based on a completely different default data set in Koopman and Lucas (2005) and Koopman et al. (2005). Note, however, that cycle periods have to be interpreted with care given the span of the data.

To conclude the empirical section, we present estimates of the common risk factor f_t using the Monte Carlo method for estimating the state vector, see Section 2.4. The smoothed estimates of f_t are presented in Figure 5. First we note that the estimates of f_t are very close across ratings. This is in line with the interpretation of f_t as a *common* risk component affecting all companies in an economy. The dynamic pattern of f_t based on the subinvestment grade rating is slightly more subtle than the one based on the investment grade companies only. This can be understood from the raw data in Figure 1. As the yearly number of defaults is much smaller for companies that are initially rated as investment grade, the investment grade companies can

only indicate roughly whether the common risk was generally high or low in a specific year. For subinvestment grade companies, more detailed dynamic patterns can be extracted due to the fact that default is just a more likely event. In this case, the measurement of the signal f_t is less subject to noise. When the two rating classes are stacked, the estimation procedure automatically extracts most information from the subinvestment grade rating. The more noisy signals from the investment grade data are then used to fine-tune the estimates of the signal f_t in the 'clear' peaks and troughs. From Figure 5 we also learn that the AR(2) specification finds some support in the data. The pattern of f_t exhibits some cyclical behavior. However, since the sample period only covers about 1.5 cycles, it is hard to find statistically significant evidence for cyclical behavior of the signal f_t.

<INSERT FIGURE 5 AROUND HERE>

5 Conclusions

In this paper we specified and implemented a non-Gaussian multivariate dynamic model for defaults. Despite the model's complicated structure, it can be estimated highly efficiently using importance sampling techniques and publicly available software (Ox and SsfPack, see Doornik (1999)). The estimation procedure works well in both a simulation and an empirical setting. We showed that increasing the cross-section size of a non-Gaussian panel time series may substantially improve the performance of the importance sampler.

Our empirical model allows for the simultaneous modeling of dynamic common risk factors and age effects in default probabilities or asset correlations. In this way, it allows for an explicit testing of hypotheses related to age changes in default correlations, see for example De Servigny and Renault (2003). For the Standard and Poor's dataset we found clear age effects in default probabilities. The evidence for shifts in asset correlations with rating age or with time, however, was not significant, although the short time span of the data should be taken into account.

The binomial panel time series model and its empirical results constitute a first step towards a more thorough application of multi-year default modeling in the presence of common dynamic risk factors. The model is able to pool across initial rating classes in a straightforward way. For example, we could further disaggregate the default frequencies along a finer rating scale (AAA – CCC). Preliminary experiments along these lines show that the results shown in this paper remain broadly unaffected. Other possible extensions of the current set-up include accounting for observed heterogeneity between firms other than the initial rating and age. Industry, for example, is a good candidate here. We leave this, however, for further research.

Appendix

Monte Carlo estimation of binomial time series models

Consider a univariate binomial time series model where the observation y_t is the number of successes in k_t independent trials with a probability of success π_t, that is

$$y_t \sim \text{Binomial}(k_t, \pi_t), \qquad t = 1, \ldots, T.$$

Further we assume that π_t is correlated over time and is modeled as a function of the dynamic process f_t. For example, the process f_t can represent a nonstationary random walk process or a stationary autoregressive moving average (ARMA) process. The link function between π_t and f_t is chosen in such a way that the density function can be formulated as a special case of the exponential family of density functions. We have

$$\pi_t^{-1} = 1 + \exp(-\theta_t), \quad \text{with} \quad \theta_t = \lambda - \beta f_t,$$

for $t = 1, \ldots, T$. Note that in the context of this paper, it is convenient to have $\theta_t = \lambda - \beta f_t$ rather than $\theta_t = \lambda + \beta f_t$, see subsection 2.2. The binomial log-density function conditional on the unobserved variable θ_t, a linear function of f_t, is given by

$$\log p(y_t|\theta_t) = y_t\theta_t - k_t \log\left(1 + \exp\theta_t\right) + \log \begin{pmatrix} k_t \\ y_t \end{pmatrix}. \tag{22}$$

The density $p(\theta)$, with $\theta = (\theta_1, \ldots, \theta_T)'$, depends on the dynamic properties of f_t but is usually taken as a Gaussian density. To generalize the specification for the unobservable θ_t further, we have

$$\theta_t = Z_t\alpha_t, \qquad \alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim \text{NID}(0, Q_t), \tag{23}$$

for $t = 1, \ldots, T$ with $\alpha_1 \sim N(a, P)$. The vector a and the matrices T_t, R_t, Q_t and P are fixed and known matrices for $t = 1, \ldots, T$. The joint model (22) and (23) is then a special case of a nonlinear and non-Gaussian state space model as discussed in part II of Durbin and Koopman (2001).

For this class of models, a so-called closed expression for the density $p(y)$, with $y = (y_1, \ldots, y_T)'$, is not available since

$$p(y) = \int p(y, \theta)\,\mathrm{d}\theta = \int p(y|\theta)p(\theta)\,\mathrm{d}\theta, \tag{24}$$

where $p(y|\theta)$ and $p(y)$ refer to different density functions and their product does not reduce to a convenient expression. We therefore rely on numerical techniques for the evaluation of $p(y)$. In particular, we follow the importance sampling methods of Shephard and Pitt (1997) and Durbin and Koopman (1997).

Importance sampling is based on the simulation of θ_t given the observations y. The simulations can be used to evaluate the likelihood function and to estimate θ_t (signal extraction). When the simulations are from a density conditional on y, the samples are informative with respect to y and therefore effective for this purpose. For example, a Monte Carlo estimator of the likelihood function $p(y) = \int p(y|\theta)p(\theta)\,\mathrm{d}\theta$ can also be based on the unconditional density, that is

$$\widehat{p}(y) = \sum_{i=1}^{M} p(y|\theta^i), \qquad \text{where} \quad \theta^i \sim p(\theta).$$

This simple estimator is poor since many simulations of θ from $p(\theta)$ will make no contribution to $p(y|\theta)$ and therefore a very large number of simulations M is needed to obtain only an inaccurate Monte Carlo estimate $\widehat{p}(y)$. Ideally, θ should be simulated from $p(\theta|y) = p(\theta, y)/p(y) = p(y|\theta)p(\theta)/p(y)$. In fact, in this case only one simulation for θ is needed. As it is the case for (24), a closed expression for density $p(\theta|y)$ is usually not available. We therefore need to find another device from which we can sample θ conditional on y. A convenient choice is to sample from a Gaussian observation model that depends linearly on θ since the density of a normal variable conditional on another normal variable is Gaussian. Further, sampling from a Gaussian density is straightforward. However, the Gaussian density $p_G(\theta|y)$ should be as close as possible to $p(\theta|y)$. In our search for an appropriate density $p_G(\theta|y)$, we start with a method for finding the mode of $p(\theta|y)$.

The mode $\widehat{\theta}$ of the log-density $\log p(\theta|y) = \log p(y|\theta) + \log p(\theta) - \log p(y)$ is the solution of

$$\frac{\partial \log p(\theta|y)}{\partial \theta} = \frac{\partial \log p(y|\theta)}{\partial \theta} + \frac{\partial \log p(\theta)}{\partial \theta} = 0,$$

which cannot be solved directly for θ. By adopting the arguments in Durbin and Koopman (2001, §§11.3, 11.4), the mode can be obtained by linearization through the formulation of a linear Gaussian model and by estimating the mode from this model. The linearization takes place around a current guess of the mode, say $\bar{\theta}^i$, and a new guess of the mode, say $\bar{\theta}^{i+1}$, is obtained from computing the conditional mean from the linear Gaussian model. Note that the mean and mode are the same in a Gaussian model. In more detail, we consider the Gaussian density $p_G(y|\theta)$ that corresponds to the linear Gaussian model

$$\widetilde{y}_t = \theta_t + u_t, \qquad u_t \sim \mathrm{NID}(0, H_t), \qquad t = 1, \ldots, T, \tag{25}$$

where $\widetilde{y}_t$ and H_t are obtained by the linearization conditions

$$\frac{\partial \log p_G(y|\theta)}{\partial \theta} = \frac{\partial \log p_(y|\theta)}{\partial \theta}, \qquad \frac{\partial^2 \log p_G(y|\theta)}{\partial \theta \partial \theta'} = \frac{\partial^2 \log p_(y|\theta)}{\partial \theta \partial \theta'}.$$

We note that $p(\theta) \equiv p_G(\theta)$ and therefore the linearization also applies to the first and second derivatives of the densities $p(\theta|y)$ and $p_G(\theta|y)$ with respect to θ. It follows that $\widetilde{y}_t$ and H_t depend on θ and for the linearization they are functions of y_t and the current guess $\bar{\theta}^i$. In the

case of the binomial density we obtain

$$\widetilde{y}_t = \theta_t + H_t y_t - 1 - \exp\theta_t, \qquad H_t = k_t^{-1}(1+\exp\theta_t)^2\exp(-\theta_t), \tag{26}$$

for $\theta_t = \bar{\theta}_t^i$. Given values for $\widetilde{y}_t$ and H_t, the conditional mean of θ_t for the Gaussian model (25) can be computed by the Kalman filter and smoothing algorithm. The resulting estimate of θ is taken as the new guess of the mode, $\bar{\theta}^{i+1}$. The sequence of new estimates of θ converges to the mode of $p(\theta|y)$. For many different non-Gaussian log-densities, convergence to $\bar{\theta}_t$ is very quick and typically takes five to ten iterations.

The model (25) with $\widetilde{y}_t$ and H_t evaluated at the mode $\bar{\theta}_t$ is referred to as the "approximating" linear Gaussian model and the corresponding density $p_G(\theta|y)$ is taken as the importance density. By generating samples from $p_G(\theta|y)$ we effectively simulate θ's at the location of the mode of $p(\theta|y)$ and with a curvature implied by the approximating model. Drawing samples from $p_G(\theta|y)$ is based on model (25) and is relatively straightforward given the simulation smoothing algorithms of de Jong and Shephard (1995) and Durbin and Koopman (2002).

Given a sequence of draws $\theta^i \sim p_G(\theta|y)$ for $i = 1, \ldots, M$, a Monte Carlo estimator of the likelihood is given by

$$\widehat{p}(y) = M^{-1}\sum_{i=1}^{M} p(y|\theta^i)p(\theta^i) \;/\; p_G(\theta^i|y) = p_G(y)M^{-1}\sum_{i=1}^{M} p(y|\theta^i) \;/\; p_G(y|\theta^i),$$

since $p_G(\theta) = p(\theta)$ when the model for θ_t is linear Gaussian. In a similar way, we can obtain estimates of θ by evaluating the conditional mean

$$\bar{\theta} = \int \theta p(\theta|y)\,\mathrm{d}\theta = p(y)^{-1}\int \theta p(\theta, y)\,\mathrm{d}\theta = p(y)^{-1}\int \theta p(y|\theta)p(\theta)\,\mathrm{d}\theta, \tag{27}$$

where density $p(y)$ is given by (24). The same argumentation as for the Monte Carlo estimator $\widehat{p}(y)$ applies such that the importance sampling estimator of $\bar{\theta}$ is given by

$$\widehat{\bar{\theta}} = \frac{\sum_{i=1}^{M} \theta^i \; p(y|\theta^i) \;/\; p_G(y|\theta^i)}{\sum_{i=1}^{M} p(y|\theta^i) \;/\; p_G(y|\theta^i)}, \tag{28}$$

where θ^i is sampled from the importance density $p_G(\theta|y)$ for $i = 1, \ldots, M$.

References

Allen, L. and A. Saunders (2003). A survey of cyclical effects in credit risk measurement models. Technical report, BIS Working Paper 126.

Altman, E. (1983). *Corporate financial distress. A complete guide to predicting, avoiding, and dealing with bankruptcy*. New York: Wiley.

Altman, E. (1989). Measuring corporate bond mortality and performance. *Journal of Finance 44*(4), 909–922.

Altman, E. and H. Suggitt (2000). Default rates in the syndicated bank loan market – a mortality analysis. *Journal of Banking and Finance 24*(1-2), 229–253.

Asquith, P., D. Mullins, and E. Wolff (1989). Original issue high yield bonds: Aging analyses of defaults, exchanges, and calls. *Journal of Finance 44*(4), 923–952.

Bangia, A., F. Diebold, A. Kronimus, C. Schagen, and T. Schuermann (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance 26*(2–3), 445–474.

Basel Committee on Bank Supervision (2004, June). Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework. Report 107, Bank of International Settlements, Basel.

Baxter, M. and R. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics 81*, 575–593.

Couderc, F., O. Renault, and O. Scaillet (2003). Credit risk dynamics: Forecasting rating transition matrices. FAMA Working paper, Geneva.

Cowan, A. M. and C. D. Cowan (2004). Default correlation: An empirical investigation of a subprime lender. *Journal of Banking and Finance 28*, 753–771.

Danielsson, J. and J. F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *J. Applied Econometrics 8*, S153–S174.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

De Servigny, A. and O. Renault (2003). Correlation evidence. *Risk* (July), 90–94.

Dietsch, M. and J. Petey (2004). Should SME exposures be treated as retail or corporate exposures? a comparative analysis of default probabilities and asset correlations in French and German SMEs. *Journal of Banking and Finance 28*, 773–788.

Doornik, J. (1999). *Object-oriented matrix programming using Ox* (3rd ed.). London: Timberlake Consultants Press. http://www.nuff.ox.ac.uk/Users/Doornik.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika 84*, 669–84.

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

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series models. *Biometrika 89*, 603–616.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modeling Extremal Events*, Volume 33 of *Applications of Mathematics; Stochastic Modelling and Applied Probability*. Heidelberg: Springer Verlag.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Gordy, M. B. (2000). A comparative anatomy of credit risk models. *Journal of Banking and Finance 24*(1–2), 119–149.

Gorton, G. B. and P. He (2003). Bank credit cycles. Working paper University of Pennsylvania.

Gupton, G., C. Finger, and M. Bhatia (1997). *CreditMetrics — Technical Document* (1st ed.). http://www.riskmetrics.com.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Kavvathas, D. (2001). Estimating credit rating transition probabilities for corporate bonds. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=252517.

Kloek, T. and H. K. Van Dijk (1978). Bayesian estimates of equation system parameters: an application of integration by monte carlo. *Econometrica 46*, 1–20.

Koopman, S. J. and A. Lucas (2005). Business and default cycles for credit risk. *Journal of Applied Econometrics 20*, 311–323.

Koopman, S. J., A. Lucas, and P. Klaassen (2005). Empirical credit cycles and capital buffer formation. *Journal of Banking and Finance*, forthcoming.

Koopman, S. J. and N. Shephard (2004). Estimating the likelihood of the stochastic volatility model: Testing the assumptions behind importance sampling. Working paper, Vrije Universiteit Amsterdam.

McNeil, A. and J. Wendin (2004). Generalised linear mixed models in portfolio credit risk modelling. Working paper, ETH Zurich.

Monahan, J. (1993). Testing the behavior of importance sampling weights. *Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface*, 112–117.

Nickell, P., W. Perraudin, and S. Varotto (2000). Stability of rating transitions. *Journal of Banking and Finance 24*(1–2), 203–227.

Poirier, D. J. (1976). *The Econometrics of Structural Change: with Special Emphasis on Spline Functions*. Amsterdam: North-Holland.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Wilson, T. (1997a, September). Portfolio credit risk, part I. *Risk*, 111–117.

Wilson, T. (1997b, October). Portfolio credit risk, part II. *Risk*, 56–61.

Table 1: Aging Effects in Default Probabilities

The table contains likelihood ratio tests for the aging effects in default thresholds. The restricted likelihoods LL^R give the likelihood for a model with an age-independent default threshold. The unrestricted likelihood LL^U corresponds to a model with a spline function for the default thresholds. The spline has 5 knots at ages of 1, 3, 7, 11, and 15 years, and thus has 4 additional parameters per rating class included. The column f_t gives the dynamics of the common risk factor for that model. The column LR is the likelihood ratio test with degrees of freedom parameter in the df column and significance in the p-value column.

Rating	f_t	LL^R	LL^U	LR	df	p-value
I.Grade	None	-458.44	-451.61	13.67	4	0.008
SubI.Grade	None	-815.95	-756.71	118.47	4	0.000
Stacked	None	-1274.38	-1208.38	132.00	8	0.000
I.Grade	IID	-454.22	-448.04	12.36	4	0.015
SubI.Grade	IID	-711.88	-681.79	60.16	4	0.000
Stacked	IID	-1157.32	-1122.92	68.80	8	0.000
I.Grade	AR1	-449.57	-443.55	12.04	4	0.017
SubI.Grade	AR1	-703.27	-674.61	57.32	4	0.000
Stacked	AR1	-1148.31	-1115.04	66.55	8	0.000

Table 2: Risk Factor Dynamics

The table contains likelihood ratio tests for the dynamic behavior of the common risk factor f_t. The columns Restr. and Unrestr. provide the dynamic behavior of the common risk factor, 'none' meaning that the model does not include a common risk factor. All models include aging effects for the default thresholds using the splines from the note in Table 1. The LL^R and LL^U columns give the likelihood of the restricted and unrestricted model, respectively. The column LR is the likelihood ratio test with degrees of freedom parameter in the df column and significance in the p-value column.

Rating	Restr.	Unrestr.	LL^R	LL^U	LR	df	p-value
I.Grade	None	IID	-451.61	-448.04	7.14	1	0.008
SubI.Grade	None	IID	-756.71	-681.79	149.84	1	0.000
Stacked	None	IID	-1208.38	-1122.92	170.91	2	0.000
I.Grade	IID	AR1	-448.04	-443.55	8.98	1	0.003
SubI.Grade	IID	AR1	-681.79	-674.61	14.37	1	0.000
Stacked	IID	AR1	-1122.92	-1115.04	15.77	1	0.000
I.Grade	AR1	AR2	-443.55	-441.70	3.69	1	0.055
SubI.Grade	AR1	AR2	-674.61	-673.02	3.17	1	0.075
Stacked	AR1	AR2	-1115.04	-1113.43	3.21	1	0.073

Table 3: Age and Time Shifts in Correlations

The table contains likelihood ratio tests for age and time shifts in the correlation parameters. The column type gives the type of the shift. The shifts are modeled by allowing the common factor loadings in the model to increase linearly in age and time, respectively. The common risk factor is modeled by an AR(1) process, and all models include aging effects for the default thresholds using the splines from the note in Table 1. The LL^R and LL^U columns give the likelihood of the restricted and unrestricted model, respectively. The column LR is the likelihood ratio test with degrees of freedom parameter in the df column and significance in the p-value column.

Rating	Type	LL^R	LL^U	LR	df	p-value
I.Grade	age	-443.55	-442.51	2.08	1	0.149
SubI.Grade	age	-674.61	-674.38	0.46	1	0.499
Stacked	age	-1115.04	-1113.45	3.17	2	0.205
I.Grade	time	-443.55	-442.17	2.76	1	0.096
SubI.Grade	time	-674.61	-674.41	0.41	1	0.522
Stacked	time	-1115.04	-1114.55	0.97	2	0.616

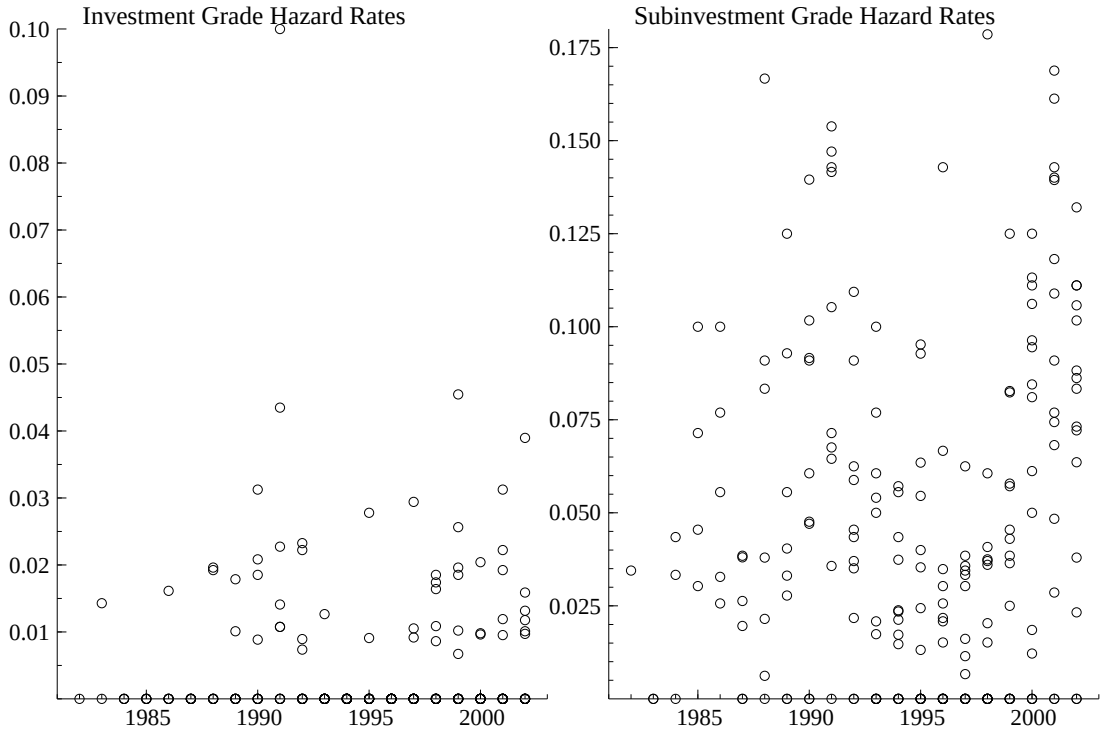

Figure 1: Cohort data for default hazard rates

This figure presents time series plots of age cohort specific default hazard rates. Each year, a cohort is formed of firms that received their initial rating in that specific year. Subsequently, the default performance of this cohort is followed over the next 15 years.

Table 4: Estimation results

The table contains the parameter estimates and standard errors of the cohort default model with two ratings. Each row contains the results for a different model. Rows starting with I, S, or IS, or for investment grade, subinvestment grade, and stacked investment and subinvestment grade data. The parameters ϕ_1 and ϕ_2 are the autoregressive parameters for the first and second lag of the latent common risk factor f_t, respectively. The parameters β_I and β_S are the investment and subinvestment grade loadings to the common risk factor, respectively. The loadings can increase with age or time. For example, the complete loading on the common factor for the investment grade firms equals $\beta_S + \nu_{1S} \cdot (t - j) + \nu_{2S} \cdot t$. Likelihood values corresponding to the models can be read from Tables 1 through 3. All models allow for aging effects in default probabilities as explained in the note of Table 1.

	ϕ_1	ϕ_2	β_I	β_S	ν_{1I}	ν_{1S}	ν_{2I}	ν_{2S}
				IID factor f_t				
I			0.53					
			[0.28,1.02]					
S				0.54				
				[0.37,0.79]				
IS			0.52	0.55				
			[0.30,0.91]	[0.38,0.80]				
				AR(1) factor f_t				
I	0.85		0.69					
	[-0.35,0.99]		[0.19,2.47]					
S	0.89			0.82				
	[-0.64,1.00]			[0.11,6.02]				
IS	0.89		0.76	0.82				
	[-0.53,1.00]		[0.12,5.04]	[0.13,5.26]				
				AR(2) factor f_t				
I	1.42	-0.75	0.67					
	[-1.17,4.01]	[-0.97,0.11]	[0.33,1.35]					
S	1.22	-0.55		0.56				
	[-1.55,3.99]	[-0.88,0.11]		[0.32,1.01]				
IS	1.25	-0.55	0.54	0.57				
	[-1.59,4.08]	[-0.87,0.11]	[0.27,1.10]	[0.32,1.03]				
				AR(1) factor f_t and age shift in loadings				
I	0.81		1.17		-0.06			
	[-0.18,0.98]		[0.36,3.78]		[-0.16,0.04]			
S	0.89			0.75		0.01		
	[-0.63,1.00]			[0.10,5.47]		[-0.04,0.06]		
IS	0.90		1.28	0.73	-0.32	-0.06		
	[-0.54,1.00]		[0.18,9.13]	[0.11,4.94]	[-0.46,-0.18]	[-0.11,-0.01]		
				AR(1) factor f_t and time shift in loadings				
I	0.77		0.00				0.06	
	[0.06,0.96]		[0.00,100]				[0.01,0.11]	
S	0.90			0.53				0.03
	[-0.70,1.00]			[0.04,7.81]				[-0.08,0.14]
IS	0.90		0.18	0.25			-1.41	0.05
	[-0.17,1.00]		[0.00,100]	[0.00,47.2]			[-1.53,-1.29]	[-0.07,0.17]

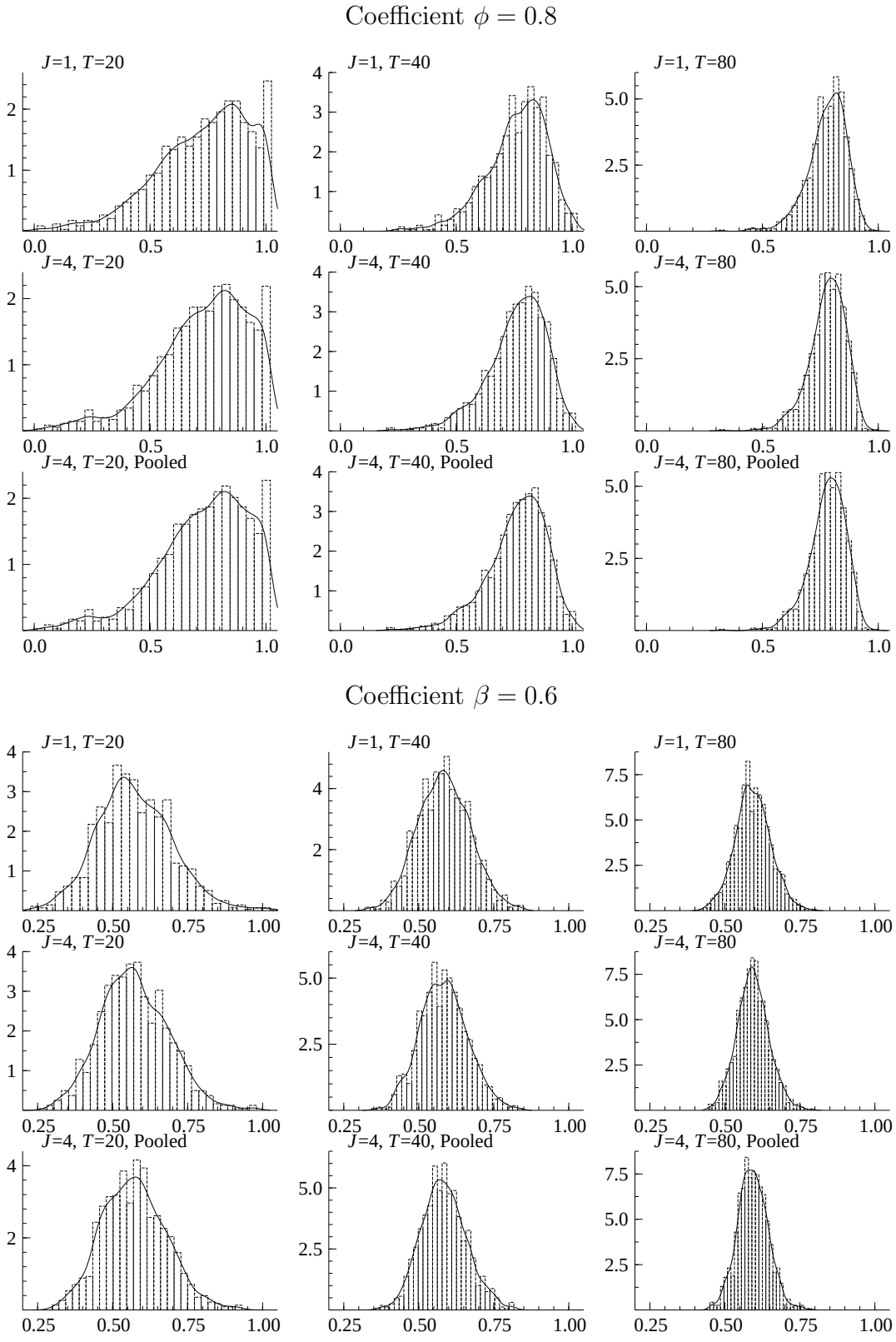

Figure 2: Results of a Monte Carlo simulation study

This figure presents the results of estimating the model parameters based on 1,000 simulated time series of size $T = 20$, $T = 40$, $T = 80$ and based on a panel of size $J = 1$, $J = 4$. In the middle row (for $J = 4$, different loadings β_i are estimated for each of the 4 series (the first of which is shown in the middle row of the lower panel). In the bottom row, estimates are based on pooling the different series by setting $\beta_j \equiv \beta$ for $j = 1, \ldots, 4$.

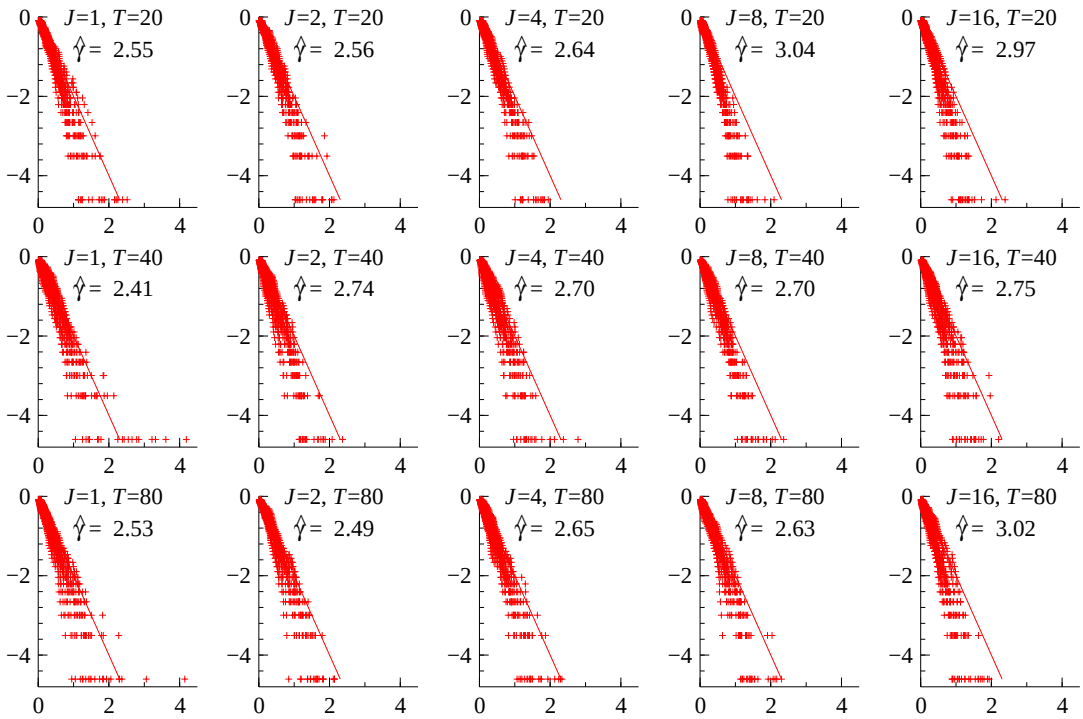

Figure 3: Tail behavior of importance sampling weights

For each time series length (T) and cross section size of the panel (J), we draw 100,000 importance samples. We take the 51 largest weights $\tilde{w}_i$, $i = 0, \ldots, 50$, sorted in ascending order, and transform them to $w_i = \tilde{w}_i / \tilde{w}_0$, $i = 1, \ldots, 50$. Each graph in the figure has on the horizontal axis $\ln(w_i)$, and plots the pairs $(\ln(w_i), \ln(1-(i-0.5)/n)$. If the weights w_i follow a Pareto distribution $F(w) = 1 - w^{-\gamma}$, we expect the pairs to lie around a straight line with slope γ. The Hill estimator ($\hat{\gamma}$) of γ is presented in each graph. The line for $\gamma = 2$ (2nd moments do not exist) is plotted as a benchmark.

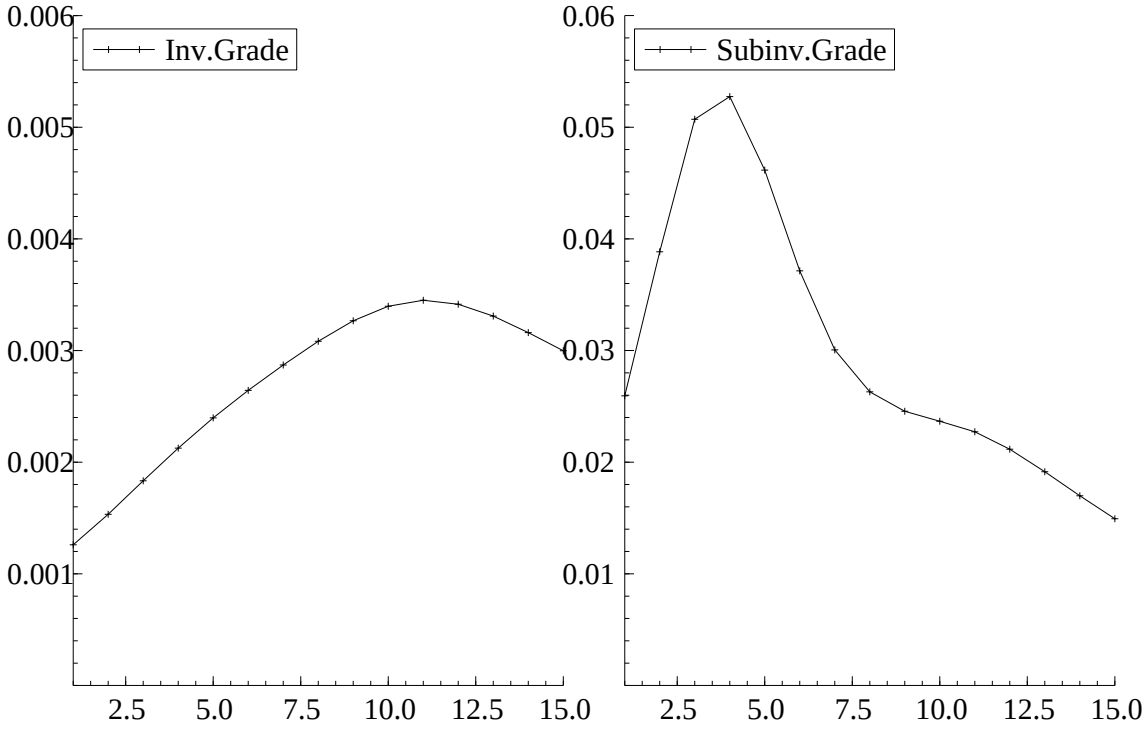

Figure 4: Default curves for investment and subinvestment grade

This figure presents the default probabilities of investment grade and subinvestment grade companies as a function of the age of the rating. The curves presented are $(1 + \exp(-c_{s,t-j}))^{-1}$, with $\bar{c}_{s,t-j}$ the default threshold for a company of age $t - j$ and initial rating class s, see (3). The horizontal axis gives the age of the issuer's rating.

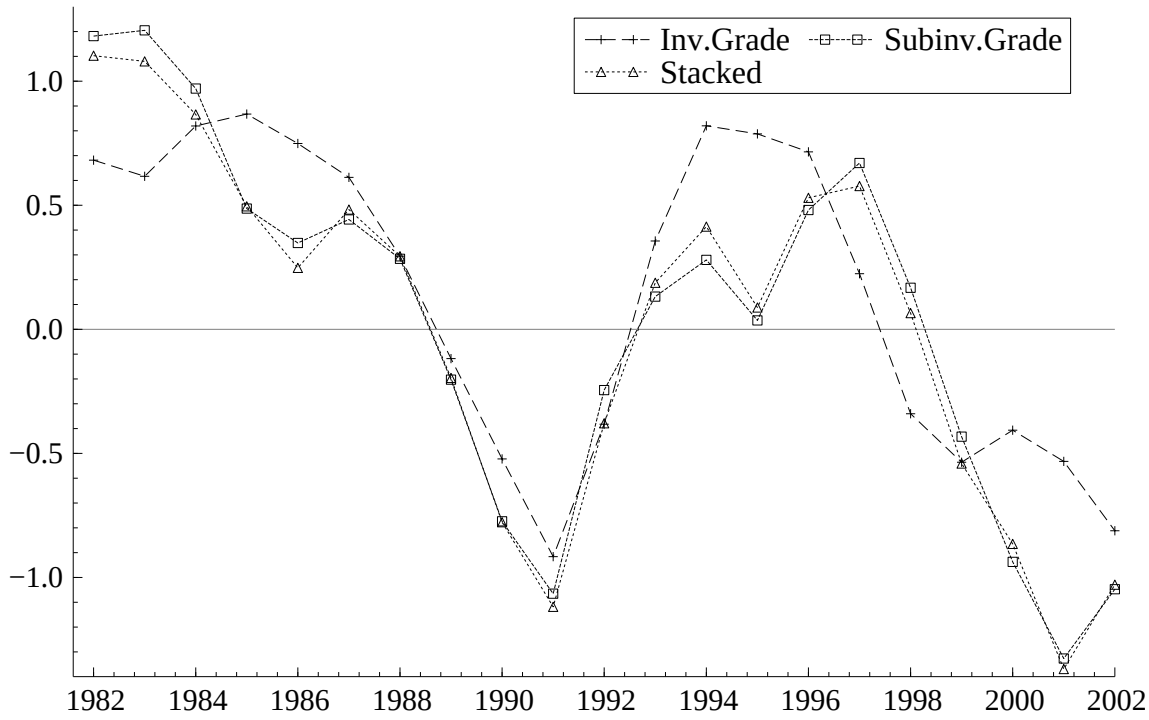

Figure 5: Smoothed estimates of the common risk factor f_t

The figure contains the smoothed common factor risk f_t estimates based on the investment grade data, the subinvestment grade data, and the investment and subinvestment grade data stacked together, respectively. The factor f_t is modeled as an AR(1) process with parameter estimates as displayed in Table 4.