Modeling frailty-correlated defaults using many macroeconomic covariates*

Siem Jan Koopman[a,b] *André Lucas*[a,b,c] *Bernd Schwaab*[d]

[a] VU University Amsterdam

[b] Tinbergen Institute

[c] Duisenberg school of finance

[d] European Central Bank, Financial Markets Research

January 13, 2011

*Corresponding author: Siem Jan Koopman, VU University Amsterdam, Department of Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: s.j.koopman@feweb.vu.nl. We thank Darrell Duffie, Kay Giesecke, Peter Boswijk, Richard Cantor, Michel van der Wel, and participants at the 2010 Credit conference in Venice, the BIS workshop 'Stress Testing of Credit Portfolios' in Amsterdam, the Econometric Society European Meeting in Milan, the IBEFA meeting in San Francisco, and other workshops for helpful discussions. We thank four anonymous referees and the associate editor for constructive comments. We also thank Moody's to grant access to their default and ratings database for this research. The views expressed in this paper are those of the authors and they do not necessarily reflect the views or policies of the European Central Bank or the European System of Central Banks.

Modeling frailty-correlated defaults using many macroeconomic covariates

Siem Jan Koopman   André Lucas   Bernd Schwaab

Abstract

We propose a novel time series panel data framework for estimating and forecasting time-varying corporate default rates subject to observed and unobserved risk factors. In an empirical application to U.S. data, we find a large and significant role for a dynamic frailty component even after controlling for more than 80% of the variation in more than 100 macro-financial covariates and other standard risk factors. We emphasize the need for a latent component to prevent a downward bias in estimated default rate volatility and in estimated probabilities of extreme default losses on portfolios of U.S. debt. The latent factor does not substitute for a single omitted macroeconomic variable. We argue that it captures different omitted effects at different times. We also provide empirical evidence that default and business cycle conditions partly depend on different processes. In an out-of-sample forecasting study for point-in-time default probabilities, we obtain mean absolute error reductions of more than forty percent when compared to models with observed risk factors only. The forecasts are relatively more accurate when default conditions diverge from aggregate macroeconomic conditions.

Keywords: systematic default risk; frailty-correlated defaults; state space methods; credit risk management.

JEL classification: *G21, C33*

1 Introduction

Recent research indicates that observed macroeconomic variables and firm-level information are not sufficient to capture the large degree of default clustering in observed corporate default data. In an important study, Das, Duffie, Kapadia, and Saita (2007) reject the joint hypothesis of (i) well-specified default intensities in terms of observed macroeconomic variables and firm-specific information and (ii) the conditional independence (doubly stochastic default times) assumption. This is bad news for practitioners, since many textbook credit risk models build on conditional independence.

Excess default clustering is often attributed to frailty and contagion. The frailty effect captures default dependence above and beyond what is already implied by observed macroeconomic and financial data. In the econometric literature frailty effects are usually modeled by an unobserved risk factor, see McNeil and Wendin (2007), Koopman, Lucas, and Monteiro (2008), Koopman and Lucas (2008), Duffie, Eckner, Horel, and Saita (2009), and Azizpour, Giesecke, and Schwenkler (2010). When a model for non-Gaussian credit risk data contains dynamic latent components, the likelihood function is not available in closed form, and advanced econometric techniques based on simulation methods are often required. Contagion effects offer another explanation of excess default clustering. Contagion refers to the phenomenon that a defaulting firm can weaken other firms with which it has links. Examples of this literature are Giesecke (2004), Lando and Nielsen (2008), and Giesecke and Kim (2010). Such business links are particularly relevant at the industry level through supply chain relationships, see Lang and Stulz (1992), Jorion and Zhang (2007), and Boissay and Gropp (2010). Empirically distinguishing contagion from frailty effects, however, is notoriously hard given the limited datasets typically available. A statistical framework is provided in Azizpour et al. (2010).

In our current paper we develop a practical and feasible econometric framework for the measurement and forecasting of point-in-time default probabilities when excess default clustering is present. The underlying economic model allows for default correlations that originate from macroeconomic and financial conditions, frailty risk, and industry sector dynamics. The model is aimed to support credit risk management at financial institutions and stress tests at supervisory agencies. It may also have an impact on the assessment of systemic risk conditions at (macro-prudential) supervisory agencies such as the new European Systemic Risk Board for the European Union, and the Financial Stability Oversight Council for the United States. Time-varying default risk conditions contribute to overall financial systemic risk, and an assessment of the latter requires estimation of the former.

We present three contributions to the econometric credit risk literature. First, we show how to combine a nonlinear non-Gaussian panel data model for discrete default counts with an approximate dynamic factor model for continuous macroeconomic time series data. We argue that the resulting framework inherits the best features from both strands of literature. A linear Gaussian factor model permits the use of large arrays of relevant predictor variables. Conversely, a non-Gaussian panel data model in state space form allows for unobserved frailty effects, easily accommodates the cross-sectional heterogeneity of firms, and routinely handles missing values that arise in count data at a disaggregated level. Parameter and factor estimation are achieved by adopting a maximum likelihood framework for multivariate non-Gaussian models in state space form, see Durbin and Koopman (1997, 2001) and Koopman and Lucas (2008). Our model for default counts is set in discrete time rather than in in continuous time, such as for example Duffie, Eckner, Horel, and Saita (2009) and Azizpour et al. (2010). Though the model can be extended to such a context, this is not pursued in the present paper.

Our framework allows us to estimate a large dimensional model, accommodating more than 100 time series of disaggregated default counts and more than 100 macro-financial covariates, in only 20 to 90 minutes on a standard desktop PC. The computational speed and model tractability allow us to conduct repeated out-of-sample forecasting experiments, where parameters and factors are re-estimated based on expanding sets of data.

As our second contribution, we conduct an empirical study of U.S. default data from 1981Q1 to 2009Q4 and find a large and significant role for a dynamic frailty component after taking into account more than 80% of the variation from more than 100 macroeconomic and financial covariates, while controlling for standard measures of risk such as ratings, equity returns, and volatilities. The increase in likelihood from an unobserved component is large. Based on data including the recent financial crisis, and a different modeling framework and estimation methodology, we confirm and extend the findings of Duffie et al. (2009) who use a latent component to prevent a downward bias in the estimation of default rate volatility and extreme default losses on portfolios of U.S. corporate debt. Our results indicate that the presence of a latent factor is not due to a few omitted macroeconomic covariates, but rather appears to capture different omitted effects at different times. In general, the default cycle and business cycle appear to depend on different processes. As a result, inference on the default cycle using observed risk factors only is at best suboptimal, and at worst systematically misleading.

Third, we show that the three types of risk factors - common factors from observed data, a frailty factor, and simple observed industry-specific control factors - are all useful for out-of sample forecasting of default risk conditions. Reductions in the mean absolute forecasting errors are substantial, and far exceed the reductions achieved by standard models which use a limited set of observed covariates directly. For example, we find that mean absolute forecasting errors reduce about 43% through the cycle compared to a benchmark model with only observed macro-financial risk factors. Such reductions have clear practical implications for the computation of capital buffers, and for the stress testing of financial institutions' loan books. Reductions in MAE are most pronounced when frailty effects are highest. Examples are the year 2002, when default rates remain high while the economy is out of recession. Also, in the period 2005-07 leading up to the 2007-09 financial crisis, default conditions are substantially more benign than what is implied by observed macro data.

This paper proceeds as follows. In Section 2 we introduce the econometric framework which combines a nonlinear non-Gaussian panel time series model with an approximate dynamic factor model for many covariates. Section 3 shows how the proposed econometric model can be represented as a multi-factor firm value model for dependent defaults. In Section 4 we discuss the estimation of the unknown parameters. Section 5 introduces the data for our empirical study, presents the major empirical findings, and discusses the out-of-sample forecasting results. Section 6 concludes.

2 The econometric framework

In this section we present our reduced form econometric model for dependent defaults. We denote the default counts of cross section j at time t as y_{jt} for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. The model thus differs from for example Duffie, Eckner, Horel, and Saita (2009) and Azizpour et al. (2010) in that it is set in a discrete rather than continuous time framework and uses default counts rather than individual default times. The index j refers to a specific combination of firm characteristics, such as industry sector, current rating class, and company age. Defaults can be cross-sectionally dependent through shared exposure to the business cycle, financing conditions, monetary and fiscal policy, and waves of optimistic or pessimistic sentiment. The macroeconomic impact at time t is summarized by exogenous factors in the $R \times 1$ vector F_t. Other observed explanatory covariates, such as trailing equity returns and respective volatilities are collected in a vector C_t. A frailty factor f_t^{uc} (where 'uc' refers to unobserved component) captures default clustering above and beyond what is implied by observed macro data.[1] Observed and unobserved risk factors are collected in the factor path $\mathcal{F}_t = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_t\}$, where $\tilde{f}_t = (f_t^{uc}, F_t', C_t')'$. After conditioning on the factors, defaults y_{jt} in cross section j are assumed to be generated as sums over independent Bernoulli trials with a common time-varying default probability π_{jt}. We refer to Lando (2003, Chapter 9), McNeil, Frey, and Embrechts (2005, Chapter 8), McNeil and Wendin (2007), and CreditMetrics (2007) for a discussion of so-called binomial mixture models. The panel time series of defaults is given by

$$y_{jt}|F_t, C_t, f_t^{uc} \quad \sim \quad \text{Binomial}(k_{jt}, \pi_{jt}), \tag{1}$$

[1] As we do not explicitly include a contagion factor as in Azizpour et al. (2010), our estimated frailty component will also pick up excess default clustering due to (contagious) business links.

where y_{jt} is the total number of default 'successes' from k_{jt} exposures. In our model, k_{jt} represents the number of firms in cell j that are active at the beginning of period t. We re-count observed exposures k_{jt} at the beginning of each quarter.

The measurement and forecasting of the time-varying default probability π_{jt} is our central focus. We specify π_{jt} as the logistic transform of an index function θ_{jt}. Therefore θ_{jt} can be interpreted as the log-odds or logit transform of π_{jt}. The time-varying default probabilities are specified by

$$\pi_{jt} = (1+e^{-\theta_{jt}})^{-1}, \tag{2}$$

$$\theta_{jt} = \lambda_j + \beta_j f_t^{uc} + \gamma_j' F_t + \delta_j' C_t, \tag{3}$$

where λ_j is a fixed effect for the jth cross section and the coefficient vectors β_j, γ_j, and δ_j capture risk factor sensitivities that may depend on firm characteristics such as industry sector and rating class. The default signals θ_{jt} do not contain idiosyncratic error terms. Instead, idiosyncratic randomness is captured in (1). The log-odds of π_{jt} may vary over time due to variation in the macroeconomic factors, F_t, observed covariates, C_t, and the frailty component, f_t^{uc}.

The frailty factor f_t^{uc} is modeled as a stationary autoregressive process of order one,

$$f_t^{uc} = \phi f_{t-1}^{uc} + \sqrt{1-\phi^2}\eta_t, \qquad \eta_t \sim \mathrm{NID}(0,1), \qquad t = 1,\ldots,T, \tag{4}$$

where $0 < \phi < 1$ and where η_t is a serially uncorrelated, standardized Gaussian disturbance. We therefore have $\mathrm{E}(f_t^{uc}) = 0$, $\mathrm{Var}(f_t^{uc}) = 1$, and $\mathrm{Cov}(f_t^{uc}, f_{t-h}^{uc}) = \phi^h$. This specification identifies β_j in (3) as the frailty factor volatility, or standard deviation. Extensions to multiple unobserved factors and to other dynamic specifications for f_t^{uc} are possible.

Modeling the dependence of firm defaults on observed macro-financial variables is an active area of current research, see Duffie et al. (2009), Azizpour et al. (2010), and the references therein. The number of macroeconomic variables in the model differs across studies but is usually small. Instead of opting for a specific selection in our study, we collect a large number of macroeconomic and financial variables denoted by x_{nt} for $n = 1,\ldots,N$. The panel is assumed to adhere to a factor structure as given by

$$x_{nt} = \Lambda_n F_t + \zeta_{nt}, \qquad n = 1,\ldots,N, \tag{5}$$

where F_t is a vector of principal components, Λ_n is a row vector of loadings, and ζ_{nt} is an idiosyncratic disturbance term. The static factor representation of the approximate dynamic factor model (5) can be derived from a dynamic model specification, see Stock and Watson (2002a). The methodology of relating observed variables to a small set of factors has been employed in the forecasting of inflation and production data, see Massimiliano, Stock, and Watson (2003), asset returns and volatilities, see e.g. Ludvigson and Ng (2007), and the term structure of interest rates, see Exterkate, van Dijk, Heij, and Groenen (2010). These studies have reported favorable results when such macro factors are used for forecasting.

The factors F_t can be estimated consistently using the method of principal components. This method is expedient for at least three reasons. First, dimensionality problems do not occur even for large values of both N and T. This is particularly relevant for our empirical application, where $T, N > 100$ in both the macro and default datasets. Second, the method can be easily extended to account for missing observations which are present in many macroeconomic time series panels. Finally, the extracted factors can be used for the forecasting of particular time series in the panel, see Forni, Hallin, Lippi, and Reichlin (2005). Equations (1) to (5) combine the approximate dynamic factor model with a non-Gaussian panel data model by inserting the elements of F_t from (5) into the signal equation (3).

3 The financial framework

By relating the econometric model with the multi-factor model of CreditMetrics (2007) for dependent defaults, we can establish an economic interpretation of the parameters. In addition, we gain more intuition for the mechanisms of the model. Multi-factor models for firm default risk are widely used in risk management practice, see Lando (2003, Chapter 9).

In the special case of a standard static one-factor credit risk model for dependent defaults the values of the obligors' assets, V_i, are driven by a common random factor F, and an idiosyncratic disturbance ϵ_i. More specifically, the asset value of firm i, V_i, is modeled by

$$V_i = \sqrt{\rho_i}\, f + \sqrt{1 - \rho_i}\, \epsilon_i,$$

where scalar $0 < \rho_i < 1$ weights the dependence of firm i on the general economic condition factor f in relation to the idiosyncratic factor ϵ_i, for $i = 1, \ldots, K$, where K is the number of firms, and where $(f\,,\,\epsilon_i)'$ has mean zero and variance matrix I_2. The conditions in this framework imply that

$$\mathrm{E}(V_i) = 0, \qquad \mathrm{Var}(V_i) = 1, \qquad \mathrm{Cov}(V_i V_j) = \sqrt{\rho_i \rho_j},$$

for $i, j = 1, \ldots, K$. In our multivariate dynamic model, the framework is extended into a more elaborate version for the asset value V_{it} of firm i at time t and is given by

$$\begin{aligned} V_{it} &= \omega_{i0} f_t^{uc} + \omega_{i1}' F_t + \omega_{i2}' C_t + \sqrt{1 - (\omega_{i0})^2 - \omega_{i1}'\omega_{i1} - \omega_{i2}'\omega_{i2}}\ \epsilon_{it} \\ &= \omega_i' \tilde{f}_t + \sqrt{1 - \omega_i'\omega_i}\ \epsilon_{it}, \qquad\qquad t = 1, \ldots, T, \end{aligned} \tag{6}$$

where frailty factor f_t^{uc}, macro factors F_t and firm/industry-specific covariates C_t have been introduced in (1), the associating weight vectors ω_{i0}, ω_{i1}, and ω_{i2} have appropriate dimensions, the factors and covariates are collected in $\mathcal{F}_t = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_t\}$, where $\tilde{f}_t = (f_t^{uc}, F_t', C_t')'$, and all weight vectors are collected in $\omega_i = (\omega_{i0}, \omega_{i1}', \omega_{i2}')'$ with condition $\omega_i'\omega_i \leq 1$. The idiosyncratic standard normal disturbance ϵ_{it} is serially uncorrelated for $t = 1, \ldots, T$. The unobserved component or frailty factor f_t^{uc} represents the credit cycle condition after controlling for the first M macro factors $F_{1,t}, \ldots, F_{M,t}$ and the common variation in the covariates C_t. In other words, the frailty factor captures deviations of the default cycle from systematic macro-financial conditions. Without loss of generality we assume that all risk factors have zero mean and unit variance. Furthermore, we assume that the risk factors f_t^{uc} and F_t are uncorrelated with each other at all times.

In a firm value model, firm i defaults at time t when its asset value V_{it} drops below some threshold c_i, see Merton (1974) and Black and Cox (1976). In our framework, V_{it} is driven by systematic observed and unobserved factors as in (6). In our empirical specification, the threshold c_i depends on the current rating class, the industry sector, and the time elapsed since the initial rating assignment. For firms which have not defaulted yet, a default occurs when $V_{it} < c_i$ or, as implied by (6), when

$$\epsilon_{it} < \frac{c_i - \omega_i'\tilde{f}_t}{\sqrt{1 - \omega_i'\omega_i}}$$

for a given value of $\tilde{f}_t$. The (filtered) time varying default probability is then given by

$$\pi_{it} = \Pr\left(\epsilon_{it} < \frac{c_i - \omega_i'\tilde{f}_t}{\sqrt{1 - \omega_i'\omega_i}} \,\middle|\, \mathcal{F}_t \right). \tag{7}$$

Favorable credit cycle conditions are associated with a high value of $\omega_i'\tilde{f}_t$ and therefore with a low default probability π_{it} for firm i. Furthermore, equation (7) can be related directly to the econometric model specification in (2) and (3) where the firms $(i = 1, \ldots, I)$ are pooled into homogenous groups $(j = 1, \ldots, J)$ according to rating class, industry sector, and time from initial rating assignment. In particular, if ϵ_{it} is logistically distributed, we obtain

$$\begin{aligned} c_i &= \lambda_j \sqrt{1-a_j}, & \omega_{i0} &= -\beta_j \sqrt{1-a_j}, \\ \omega_{i1} &= -\gamma_j \sqrt{1-a_j}, & \omega_{i2} &= -\delta_j \sqrt{1-a_j}, \end{aligned}$$

where $a_j = \left(\beta_j^2 + \gamma_j'\gamma_j + \delta_j'\delta_j\right) / \left(1 + \beta_j^2 + \gamma_j'\gamma_j + \delta_j'\delta_j\right)$ for firm i that belongs to group j. The coefficient vectors λ_j, β_j, and γ_j are defined below (2) and (3). The parameters have therefore a direct interpretation in widely used portfolio credit risk models such as CreditMetrics (2007).

4 Estimation using state space methods

We next discuss parameter estimation and signal extraction of the factors for model (1) to (5). The estimation procedure for the macro factors is discussed in Section 4.1. The state space representation of the econometric model is provided in Section 4.2. We first estimate the parameters using a computationally efficient procedure for Monte Carlo maximum likelihood and then extract the frailty factor using a similar Monte Carlo method. A brief outline of these procedures is given in Section 4.3. All computations are implemented using the Ox programming language and the associated set of state space routines from SsfPack, see Doornik (2007) and Koopman, Shephard, and Doornik (2008).

4.1 Estimation of the macro factors

The common factors F_t from the macro data are estimated by minimizing the objective function given by

$$\min_{\{F,\Lambda\}} V(F,\Lambda) = (NT)^{-1} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t), \qquad (8)$$

where the $N \times 1$ vector $X_t = (x_{1t}, \ldots, x_{NT})'$ contains macroeconomic variables and F is the set $F = \{F_1, \ldots, F_T\}$ for the $R \times 1$ vector F_t. The observed stationary time series x_{nt} are demeaned and standardized to have unit unconditional variance for $n = 1, \ldots, N$. Concentrating out F and rearranging terms shows that (8) is equivalent to maximizing $\mathrm{tr}(\Lambda' S_{X'X} \Lambda)$ with respect to Λ and subject to $\Lambda'\Lambda = I_R$, where $S_{X'X} = T^{-1} \sum_t X_t X_t'$ is the sample covariance matrix of the data, see Lawley and Maxwell (1971) and Stock and Watson (2002a). The resulting principal components estimator of F_t is given by $\hat{F}_t = X_t'\hat{\Lambda}$, where $\hat{\Lambda}$ collects the normalized eigenvectors associated with the R largest eigenvalues of $S_{X'X}$.

When the variables in X_t are not completely observed for $t = 1, \ldots, T$, we employ the Expectation Maximization (EM) procedure as devised in the Appendix of Stock and Watson (2002b). This iterative procedure takes a simple form under the assumption that $x_{nt} \sim \text{NID}(\Lambda_n F_t, 1)$, where Λ_n denotes the nth row of Λ for $n = 1, \ldots, N$. Here, $V(F, \Lambda)$ in (8) is a linear function of the log-likelihood $L(F, \Lambda | X^m)$ where X^m denotes the missing parts of the dataset $X_1, \ldots, X_T$. Since $V(F, \Lambda)$ is proportional to $-L(F, \Lambda | X^m)$, the minimizers of $V(F, \Lambda)$ are also the maximizers of $L(F, \Lambda | X^m)$. This result is exploited in the EM algorithm of Stock and Watson (2002b) that we have adopted to compute $\hat{F}_t$ for $t = 1, \ldots, T$.

4.2 The factor model in state space form

We can formulate model (1) to (4) in state space form where F_t and C_t are treated as explanatory variables. In our implementation, F_t will be replaced by $\hat{F}_t$ as obtained from the previous section. The estimation framework can therefore be characterized as a two-step procedure. We first estimate the principal components to summarize the variation in macroeconomic data. We then concentrate on the estimation of the frailty factor f_t^{uc} by jointly considering $\hat{F}_t$ and C_t. In this way we establish a computationally feasible and relatively simple procedure. In Section 4.4 we present simulation evidence to illustrate the adequacy of our approach for parameter estimation and for uncovering the factors from the data.

The Binomial log-density function of model (1) is given by

$$\log p(y_{jt} | \pi_{jt}) = y_{jt} \log \left(\frac{\pi_{jt}}{1 - \pi_{jt}} \right) + k_{jt} \log(1 - \pi_{jt}) + \log \binom{k_{jt}}{y_{jt}}, \tag{9}$$

where y_{jt} is the number of defaults and k_{jt} is the number of firms in cross-section j, for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. By substituting (2) for the default probability π_{jt} into (9) we obtain the log-density in terms of the log-odds ratio $\theta_{jt} = \log(\pi_{jt}) - \log(1 - \pi_{jt})$ given by

$$\log p(y_{jt} | \theta_{jt}) = y_{jt} \theta_{jt} + k_{jt} \log(1 + e^{\theta_{jt}}) + \log \binom{k_{jt}}{y_{jt}}. \tag{10}$$

The log-odds ratio in (3) can be specified as

$$\theta_{jt} = Z_{jt} \alpha_t, \qquad Z_{jt} = (e_j', \ \hat{F}_t' \otimes e_j', \ C_t' \otimes e_j', \ \beta_j), \tag{11}$$

where e_j denotes the jth column of the identity matrix of dimension J, the state vector $\alpha_t = (\lambda_1, \ldots, \lambda_J, \gamma_{1,1}, \ldots, \gamma_{R,J}, \delta_1', \ldots, \delta_J', f_t^{uc})'$ consists of the fixed effects λ_j, the loadings

$\gamma_{r,j}$ and δ_j together with the unobserved component f_t^{uc}. The system vector Z_{jt} is time-varying due to the inclusion of $\hat{F}_t$ and C_t.

The state vector α_t contains all unknown coefficients that are linear in the signals θ_{jt}. The transition equation provides a model for the evolution of the state vector α_t over time and is given by

$$\alpha_{t+1} = T\alpha_t + Q\xi_t, \qquad \eta_t \sim \text{NID}(0, 1), \tag{12}$$

with transition matrix $T = \text{diag}(I, \phi)$ and with vector $Q = (0, \ldots, 0, \sqrt{1-\phi^2})'$. All initial elements of the state vector are subject to diffuse initial conditions except for f_t^{uc}, which has zero mean and unit variance.

The equations (10) and (12) belong to a class of non-Gaussian state space models as discussed in Durbin and Koopman (2001, Part II) and Koopman and Lucas (2008). In our formulation, most unknown coefficients are part of the state vector α_t and are estimated as part of the filtering and smoothing procedures described in Section 4.3. This formulation leads to a considerable increase in the computational efficiency of our estimation procedure. The remaining parameters are collected in a coefficient vector $\psi = (\phi, \beta_1, \ldots, \beta_J)'$ and are estimated by the Monte Carlo maximum likelihood methods that we will discuss next.

4.3 Parameter estimation and signal extraction

Parameter estimation for a non-Gaussian model in state space form can be carried out by the method of Monte Carlo maximum likelihood. Once we have obtained an estimate of ψ, we can compute the conditional mean and variance estimates of the state vector α_t. In both cases we make use of importance sampling methods. The details of our implementation are given next.

For notational convenience we suppress the dependence of the density $p(y; \psi)$ on ψ. The likelihood function of our model (1) to (4) can be expressed by

$$\begin{aligned} p(y) &= \int p(y, \theta) d\theta = \int p(y|\theta) p(\theta) d\theta \\ &= \int p(y|\theta) \frac{p(\theta)}{g(\theta|y)} g(\theta|y) d\theta = \text{E}_g \left[p(y|\theta) \frac{p(\theta)}{g(\theta|y)} \right], \end{aligned} \tag{13}$$

where $y = (y_{11}, y_{21}, \ldots, y_{JT})'$, $\theta = (\theta_{11}, \theta_{21}, \ldots, \theta_{JT})'$, $p(\cdot)$ is a density function, $p(\cdot, \cdot)$ is a joint density, $p(\cdot|\cdot)$ is a conditional density, $g(\theta|y)$ is a Gaussian importance density, and E_g denotes expectation with respect to $g(\theta|y)$. The importance density $g(\theta|y)$ is constructed as the Laplace approximation to the intractable density $p(\theta|y)$: both densities have the same mode and curvature at the mode, see Durbin and Koopman (2001) for details. Conditional on θ, we can evaluate $p(y|\theta)$ by

$$p(y|\theta) = \prod_{j,t} p(y_{jt}|\theta_{jt}).$$

It follows from (3) and (4) that the marginal density $p(\theta)$ is Gaussian and therefore $p(\theta) = g(\theta)$. Since $g(\theta|y)g(y) \equiv g(y|\theta)g(\theta)$ we obtain

$$p(y) = \mathrm{E}_g \left[p(y|\theta) \frac{p(\theta)}{g(y|\theta)} \frac{g(y)}{p(\theta)} \right] = \mathrm{E}_g \left[g(y) \frac{p(y|\theta)}{g(y|\theta)} \right] = g(y) \mathrm{E}_g \left[w(y, \theta) \right], \qquad (14)$$

where $w(y, \theta) = p(y|\theta)/g(y|\theta)$. A Monte Carlo estimator of $p(y)$ is therefore given by

$$\hat{p}(y) = g(y)\bar{w},$$

with

$$\bar{w} = M^{-1} \sum_{m=1}^{M} w^m = M^{-1} \sum_{m=1}^{M} \frac{p(y|\theta^m)}{g(y|\theta^m)},$$

where $w^m = w(\theta^m, y)$ is the value of the importance weight associated with the m-th draw θ^m from $g(\theta|y)$, and M is the number of Monte Carlo draws. The Gaussian importance density $g(\theta|y)$ is chosen for convenience and since it is possible to generate a large number of draws θ^m from it in a computationally efficient manner using the simulation smoothing algorithm of de Jong and Shephard (1995) or Durbin and Koopman (2002). We estimate the log-likelihood as $\log \hat{p}(y) = \log \hat{g}(y) + \log \bar{w}$, and include a bias correction term as discussed in Durbin and Koopman (1997).

The Gaussian importance density $g(\theta|y)$ is based on the approximating Gaussian model as given by

$$y_{jt} = c_{jt} + \theta_{jt} + u_{jt}, \qquad u_{jt} \sim \mathrm{NID}(0, d_{jt}), \qquad (15)$$

where the disturbances u_{jt} are mutually and serially uncorrelated, for $j = 1, \ldots, j$ and $t = 1, \ldots, T$. The unknown constant c_{jt} and variance d_{jt} are determined by the individual matching of the first and second derivative of $\log p(y_{jt}|\theta_{jt})$ in (10) and $\log g(y_{jt}|\theta_{jt}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log d_{jt} - \frac{1}{2} d_{jt}^{-1}(y_{jt} - c_{jt} - \theta_{jt})^2$ with respect to the signal θ_{jt}. The matching equations for c_{jt} and d_{jt} rely on θ_{jt} for each j, t. For an initial value of θ_{jt}, we compute c_{jt} and d_{jt} for all j, t. The Kalman filter and smoother compute the estimates for signal θ_{jt} based on the linear Gaussian state space model (15), (11) and (12). We compute new values for c_{jt} and d_{tj} based on the new signal estimates of θ_{jt}. We can repeat the computations for each new estimate of θ_{jt}. The iterations proceed until convergence is achieved, that is when the estimates of θ_{jt} do not change. The number of iterations for convergence are usually as low as 5 to 10 iterations. When convergence has taken place, the Kalman filter and smoother applied to the approximating model (15) compute the mode estimate of $\log p(\theta|y)$; see Durbin and Koopman (1997) for further details. A new approximating model needs to be constructed for each log-likelihood evaluation when the value for parameter vector ψ has changed. Finally, standard errors for the parameters in ψ are constructed from the numerical second derivatives of the log-likelihood function, that is

$$\hat{\Sigma} = \left[- \frac{\partial^2 \log p(y)}{\partial \psi \partial \psi'} \bigg|_{\psi = \hat{\psi}} \right]^{-1}.$$

For the estimation of the latent factor f_t^{uc} and fixed coefficients in the state vector, we estimate the conditional mean of α by

$$\begin{aligned} \bar{\alpha} &= \mathrm{E}\left[\alpha|y\right] = \int \alpha p(\alpha|y) d\alpha \\ &= \int \alpha \frac{p(\alpha|y)}{g(\alpha|y)} g(\alpha|y) d\alpha = \mathrm{E}_g \left[\alpha \frac{p(\alpha|y)}{g(\alpha|y)} \right]. \end{aligned}$$

In a similar way as the development in (14), we obtain

$$\bar{\alpha} = \frac{\mathrm{E}_g\left[\alpha w(\theta, y)\right]}{\mathrm{E}_g\left[w(\theta, y)\right]},$$

since $p(\alpha) = g(\alpha)$, $p(y|\alpha) = p(y|\theta)$ and $g(y|\alpha) = g(y|\theta)$. The Monte Carlo estimator for $\bar{\alpha}$ is then given by

$$\hat{\bar{\alpha}} = \widehat{\mathrm{E}}[\alpha|y] = \left[\sum_{m=1}^{M} w^m \right]^{-1} \sum_{m=1}^{M} \alpha^m w^m,$$

where $\alpha^m = (\alpha_{11}, \ldots, \alpha_{JT})'$ is the m-th draw from $g(\alpha|y)$ and where θ^m is computed using (11), that is $\theta_{jt}^m = Z_{jt} \alpha_{jt}^m$ for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. The associated conditional variances are given by

$$\widehat{\mathrm{Var}}[\alpha_{jt}|y] = \left(\left[\sum_{m=1}^{M} w^m \right]^{-1} \sum_{m=1}^{M} (\alpha_{jt}^m)^2 w^m \right) - \left(\hat{\bar{\alpha}}_{it} \right)^2,$$

and allow the construction of standard error bands.

In our empirical study we also present mode estimates for signal extraction and out-of-sample forecasting of default probabilities in (3). The mode estimates of α_{jt} are obtained by the Kalman filter smoother applied to the state space model (15), (11) and (12) where c_{jt} and

d_{jt} are computed by using the mode estimate of θ_{jt}. Finally, the mode estimate of $\pi = \pi(\theta)$ is given by $\bar{\pi} = \pi(\bar{\theta})$ for any nonlinear function $\pi(\cdot)$ that is known and has continuous support. We refer to Durbin and Koopman (2001, Chapter 11) for further details.

4.4 Simulation experiments

In this subsection we investigate whether the econometric methods of Sections 4.1 and 4.3 can distinguish default rate volatility due to changes in the macroeconomic environment from changes in unobserved frailty risk. The first source is captured by principal components F_t, while the second source is estimated via the unobserved factor f_t^{uc}. This exercise is important since estimation by Monte Carlo maximum likelihood should not be biased towards attributing variation to a latent component when it is in fact due to an exogenous covariate. For this purpose we carry out a simulation study that is close to our empirical application in Section 5. The variables are generated by the equations

$$
\begin{aligned}
F_t &= \Phi_F F_{t-1} + u_{F,t}, \qquad u_{F,t} \sim \mathrm{N}(0, I - \Phi_F \Phi_F'), \\
e_t &= \Phi_I e_{t-1} + u_{I,t}, \qquad u_{I,t} \sim \mathrm{N}(0, I - \Phi_I \Phi_I'), \\
X_t &= \Lambda F_t + e_t, \\
f_t^{uc} &= \phi_{uc} f_{t-1}^{uc} + u_{f,t}, \qquad u_{f,t} \sim \mathrm{N}(0, 1 - \phi_{uc}^2),
\end{aligned}
$$

where ϕ_{uc} and the elements of the matrices Φ_F, Φ_I, and Λ are generated for each simulated dataset from the uniform distribution $U[.,.]$, that is $\phi_{uc} \sim \mathrm{U}[0.6, 0.8]$, $\Phi_F(i,j) \sim \mathrm{U}[0.6, 0.8]$, $\Phi_I(i,j) \sim \mathrm{U}[0.2, 0.4]$, and $\Lambda(i,j) \sim \mathrm{U}[0, 2]$, where $A(i,j)$ is the (i,j)th element of matrix $A = \Phi_F, \Phi_I, \Lambda$. For computational convenience we consider F_t to be a scalar process ($R = 1$) and we have no firm-specific covariates ($C_t = 0$). The default counts y_{jt} in pooling group j are generated by the equations

$$
\begin{aligned}
\theta_{jt} &= \lambda_j + \beta f_t^{uc} + \gamma F_t, \\
y_{jt} &\sim \text{Binomial}\left(k_{jt}, (1 + \exp[-\theta_{jt}])^{-1}\right),
\end{aligned}
$$

where f_t^{uc} and F_t represent their simulated values, and exposure counts k_{jt} come from the dataset which is explored in the next section. The parameters λ_j, β, γ are chosen similar to their maximum likelihood values reported in Section 5. Simulation results are based on 1000 simulations. Each simulation uses $M = 50$ importance samples during simulated maximum likelihood estimation, and $M = 500$ importance samples for signal extraction, see Section 4.3.

[insert Figure 1 around here]

A selection of the graphical output from our Monte Carlo study is presented in Figure 1. We find that the principal components estimate $\hat{F}$ captures the factor space F well. The goodness-of-fit statistic R^2 is on average 0.94. The conditional mean estimate of f_t^{uc} is close to the simulated unobserved factor, with an average R^2 of 0.73. The sampling distributions of ϕ_{uc} and λ_0 appear roughly symmetric and Gaussian, while the distributions of factor sensitivities β_0 and γ_1 appear skewed to the right. This is consistent with their interpretation as factor standard deviations. The distributions of ϕ_{uc}, β_0, λ_0, and γ_1 are all centered around their true values. We conclude that our modeling framework enables us to discriminate between possible sources of default rate variation. The resulting parameter estimates are overall correct for both ψ and state vector α.

Finally, the standard errors for the estimated factor loadings γ do not take into account that the principal components are estimated with some error in a first step. We therefore need to investigate whether this impairs inference on these factor loadings. In each simulation we estimate parameters and associated standard errors using true factors F_t as well as their principal components estimates $\hat{F}_t$. The bottom panel in Figure 1 plots the empirical distribution functions of t-statistics associated with testing the null hypothesis $H_0 : \gamma_1 = 0$ when either F_t or $\hat{F}_t$ is used. The t-statistics are very similar in both cases. Other standard errors are similarly unaffected. We conclude that the substitution of F_t with $\hat{F}_t$ has negligible effects for parameter estimation.

5 Estimation results and forecasting accuracy

We first describe the macroeconomic, financial, and firm default data used in our empirical study. We then discuss our main findings from the study. We conclude with the discussion of out-of-sample forecasting results for time-varying default probabilities.

5.1 Data

We use data from two main sources. First, a panel of more than 100 macroeconomic and financial time series is constructed from the Federal Reserve Economic Database FRED (`http://research.stlouisfed.org/fred2`). The aim is to select series which contain information about systematic credit risk conditions. The variables are grouped into five broad categories: (a) bank lending conditions, (b) macroeconomic and business cycle indicators, including labor market conditions and monetary policy indicators, (c) open economy macroeconomic indicators, (d) micro-level business conditions such as wage rates, cost of capital, and cost of resources, and (e) stock market returns and volatilities. The macro variables are quarterly time series from 1970Q1 to 2009Q4. Table 1 presents a listing of the series for each category. The macroeconomic panel contains both current information indicators (real GDP, industrial production, unemployment rate) and forward looking variables (stock prices, interest rates, credit spreads, commodity prices).

[insert Table 1 around here]

A second dataset is constructed from the default data of Moody's. The database contains rating transition histories and default dates for all rated firms from 1981Q1 to 2009Q4. This data contains the information to determine quarterly values for y_{jt} and k_{jt} in (1). The database distinguishes 12 industries which we pool into $D = 7$ industry groups: banks and financials (fin); transport and aviation (tra); hotels, leisure, and media (lei); utilities and energy (egy); industrials (ind); technology and telecom (tec); retailing and consumer goods (rcg). We further consider four age cohorts: less than 3, 3 to 6, 6 to 12, and more than 12 years from the time of the initial rating assignment. Age cohorts are included since default probabilities may depend on the age of a company. A proxy for age is the time since the initial rating has been established. Finally, there are four rating groups, an investment grade group $Aaa - Baa$, and three speculative grade groups Ba, B, and $Caa - C$. Pooling over investment grade firms is necessary since defaults are rare in this segment. In total we distinguish $J = 7 \times 4 \times 4 = 112$ different groups.

In the process of counting exposures and defaults, a previous rating withdrawal is ignored if it is followed by a later default. If there are multiple defaults per firm, we consider only the first event. In addition, we exclude defaults that are due to a parent-subsidiary relationship. Such defaults typically share the same default date, resolution date, and legal bankruptcy date in the database. Inspection of the default history and parent number confirms the exclusion of these cases. Figure 2 presents the total number of default counts, exposure counts, and the corresponding fractions. We observe pronounced default clustering around the recession years of 1991, 2001, and the financial crisis of 2007-09.

[insert Figure 2 around here]

Aggregate default counts, exposure counts, and fractions are presented in the top panel of Figure 2. We observe pronounced default clustering around the recession years of 1991, 2001, and the financial crisis of 2007-09. Since defaults cluster due to high levels of latent systematic risk, it follows that systematic risk is serially correlated and may also account for the autocorrelation in aggregate defaults. Defaults may already rise before the onset of a recession, for example, in the years 1990 and 2000, and they may remain elevated as the economy recovers from recession, for example, in the year 2002. The bottom panel of Figure 2 presents disaggregated default fractions for four broad rating groups. Default clustering is visible for all rating groups.

Our proposed model considers groups of homogenous firms rather than individual firms. As a result it is not straightforward to include firm specific information beyond rating classes and industry sectors. Firm-specific covariates such as equity returns, volatilities and leverage are found to be important in Vassalou and Xing (2004), Duffie et al. (2007), and Duffie et al. (2009). Ratings alone may not be sufficient statistics for future default. To accommodate this concern, our set of explanatory covariates is extended with average measures of firm-specific variables across firms in the same industry groups. We use S&P industry-level equity index data from Datastream to construct trailing equity return and spot volatility measures at the industry level. The equity volatilities are constructed as realized variance estimates based on average squared returns over the past year. As a robustness check, we also follow Das, Duffie, Kapadia, and Saita (2007) and Duffie et al. (2009) by including the trailing 1-year return of the S&P 500 stock index, an S&P 500 spot volatility measure, and the 3-month T-bill rate from Datastream. These observed risk factors are then treated in the same way as the principal components from the macroeconomics dataset.

5.2 Observed macro factors and industry controls

Figure 3 presents the ten principal components obtained from the macro panel of Table 1 and computed by the EM procedure of Section 4.1. The NBER recession dates are depicted as shaded areas. The estimated first factor from the macroeconomic and financial panel is mainly associated with production and employment data; it accounts for a large share of 24% of total variation in the panel. The first factor exhibits clear peaks around the U.S. business cycle troughs. Overall, we select $M = 10$ factors which capture 82% of the variation in the panel.

[insert Figure 3 around here]

Intra-industry serial correlation may remain present in the data even after conditioning on macro and frailty factors that are common to all default data. Such leftover variation could be exploited for forecasting. We therefore regress trailing one year default rates at the industry-level on a constant and the trailing one year aggregate default rate (to concentrate out macro and frailty effects that are common to all data). Observed industry factors are then obtained as the resulting standardized residuals and included as controls in the log-odds equation (3) for all models.

5.3 Model specification

The model specification for the default counts of our $J = 112$ groups is as follows. The individual time series of counts is modeled as a Binomial sequence with log-odds ratio θ_{jt} as given by (3) or (11) where the scalar coefficient λ_j is a fixed effect, scalar β_j pertains to the frailty factor, vector γ_j to the principal components and vector δ_j to the observed industry control variables, for $j = 1, \ldots, J$. The model includes ten principal components that capture 82% of the variation from 107 macro-financial predictor variables, equity returns and volatilities at the industry level, industry-specific factors, and the firm-specific ratings, industry group, and age cohorts.

Since the cross-section is high-dimensional, we follow Koopman and Lucas (2008) in reducing the number of parameters by restricting the coefficients in the following additive structure

$$\bar{\chi}_j = \chi_0 + \chi_{1,d_j} + \chi_{2,a_j} + \chi_{3,s_j}, \qquad \bar{\chi} = \lambda, \beta, \gamma, \delta, \tag{16}$$

where χ_0 represents the baseline effect, $\chi_{1,d}$ is the industry-specific deviation, $\chi_{2,a}$ is the deviation related to age and $\chi_{3,s}$ is the deviation related to rating group. The deviations of all seven industry groups (fin, tra, lei, egy, tec, ind, and rcg) cannot be identified simultaneously given the presence of χ_0. To identify the model, we assume that $\chi_{1,d_j} = 0$ for the retail and consumer goods group, $\chi_{2,a_j} = 0$ for the age group of 12 years or more, and $\chi_{3,s_j} = 0$ for the rating rating group $Caa - C$. This normalization are innocuous and can be replaced by alternative baseline choices without affecting our conclusions. For the frailty factor coefficients, we do not account for age and therefore set $\beta_{2,a} = 0$ for all a. For the principal components coefficients, we only account for rating groups and therefore we have $\gamma_{1,d} = 0$ and $\gamma_{2,a} = 0$, for all d, s. For the industry factor coefficients, we only account for industry groups and therefore we have $\delta_{2,a} = 0$ and $\delta_{3,s} = 0$, for all d, s. Using this parameter specification, we combine model parsimony with the ability to test a rich set of hypotheses empirically given the data at hand.

5.4 Empirical findings

Table 2 presents the parameter estimates for three different specifications of the signal equation (3). Model 1 does not contain the macro factors, $\beta_j = 0$. Model 2 does not contain the latent risk factors, $\gamma_{rj} = 0$ for all r and j. Model 3 refers to specification (3) without restrictions.

[insert Table 2 around here]

When comparing the log-likelihood values of Models 1 and 3, we can conclude that adding a latent dynamic frailty factor increases the log-likelihood by approximately 65 points. This increase is statistically significant at the 1% level. Since in practice most default models rely on a small set of observed covariates, this finding indicates that a model without a frailty factor can systematically provide misleading indications of default conditions. Therefore, the industry practise is at best suboptimal, and at worst systematically misleading when used for inference on default conditions, e.g. during a stress testing exercise. Furthermore, our finding support Duffie et al. (2009) and Azizpour et al. (2010) who find that firms are exposed to a common dynamic latent component driving default in addition to observed risk factors.[2] Ignoring this component may lead to a significant downward bias when assessing default rate volatility and the associated probability of extreme default losses.

We further find that Model 2 produces a better in-sample fit to the data than Model 1 in terms of the maximized log-likelihood value. Hence, a single unobserved component captures default conditions better than the first ten principal components from the macroeconomic panel. We therefore conclude that business cycle dynamics and default risk conditions are different processes.

The principal components capture important covariation in defaults. The difference in the log-likelihood values of Models 2 and 3 is 44 points and is significant at a 5% level. We may therefore conclude that all risk factors in our model are significant. However, all principal components are not of equal importance to default rates. For example, factors 3 and 6 capture 10% and 4% of the variation in the macro panel, respectively, but they have no effect on default counts.

5.5 Interpretation of the frailty factor

We have given evidence in Section 5.4 that firms are exposed to a common dynamic latent factor driving default after controlling for observed risk factors. Given its statistical and economic significance, we may conclude that the business cycle and the default cycle are related but depend on different processes. The approximation of the default cycle by business cycle indicators may not be sufficiently accurate. Figure 4 presents the frailty factor estimates for Models 2 and 3. The recession periods of 1983, 1991, 2001, and 2008-09 are marked as

[2] One should note that our current frailty factor actually picks up the combined effect of pure frailty and contagion, see the distinction in Azizpour et al. (2010) and the discussion in Section 5.5.

shaded areas. Recession periods coincide with peaks in the default cycle in the top panel for Model 2. The bottom panel presents the estimated frailty effects for Model 3.

[insert Figure 4 around here]

Duffie et al. (2009) suggest that the frailty factor captures omitted relevant macro-financial covariates together with other omitted effects that are difficult to quantify. Our results suggest that the frailty factor captures predominantly other omitted effects as we have conditioned on macro factors that capture more than 84% of the variation in over 100 macro variables. The frailty effects in the period 2001-2002 could be attributed to the disappearance of trust in the accuracy of public accounting information following the Enron and WorldCom scandals. This development would cause lenders not to extend credit, causing illiquidity risks and eventually credit risks to go up. While such effects are important for risk conditions, they are difficult to quantify. Similarly, the downward movements of the frailty factor in 2005-2007 suggest that Model 3 is able to capture the positive effects of advances in credit risk transfer and securitization during that time. These advances have led to cheap credit access. The estimated frailty factor appears to capture different omitted effects at different times, rather than that it substitutes for a single missing covariate.

Figure 5 presents the estimated composite default signals θ_{jt} for investment grade firms (Aaa-Baa) against low speculative grade firms (Caa-C). The frailty effects are less important for investment grade firms, which typically have a highly robust access to credit. The default clustering implied by observed risk factors is sufficient to match the time-varying default probabilities in the recession periods 1983, 1991, 2001, and 2008. The speculative grade firms have a less robust access to credit that depends much more on market circumstances. For this group, frailty effects indicate additional default clustering in the 1980s, and also during the 1991 recession. The bottom panel of Figure 5 shows that the low default probabilities for bad risks in the years leading up to the financial crisis are attributed to the frailty component.

[insert Figure 5 around here]

Finally, we note that frailty in our current empirical framework may unintentionally pick up contagion effects across and within industry sectors and thus may overestimate 'pure' frailty effects. We refer to Azizpour et al. (2010) who undertake an important effort to disentangle contagion dependence from frailty. They argue that such contagion linkages remain an important source of variation even after frailty is taken into account. We do not make such a distinction explicitly here. We leave it to future research to investigate whether such a distinction is also useful for out-of-sample forecasting as dealt with in the next subsection.

5.6 Out-of-sample forecasting accuracy

We compare the out-of-sample forecasting performance between models by considering a number of competing model specifications. Accurate forecasts are valuable in credit risk management, for short-term loan pricing, and for credit portfolio stress testing. Also, out-of-sample forecasting is a stringent diagnostic check for modeling and analyzing time series. We present a truly out-of-sample forecasting study by estimating the parameters of the model using data up to a certain year and by computing the forecasts of the cross-sectional default probabilities for the next year. In this way we have computed our forecasts for the nine years of $2001, \ldots, 2009$.

The measurement of forecasting accuracy of time-varying probabilities is not straightforward. Observed default fractions are only a crude measure of default conditions. We can illustrate this inaccuracy by considering a group of, say, 5 firms. Even if the default probability for this group is forecasted perfectly, it is unlikely to coincide with the observed default fraction of either 0, 1/5, 2/5, etc. The forecast error may therefore be large but it does not necessarily indicate a bad forecast. The observed default fractions are only useful when a sufficiently large number of firms are pooled in a single group. For this reason we pool default and exposure counts over age cohorts, and focus on two broad rating groups, i.e., (i) all rated firms in a certain industry, and (ii) firms in that industry with ratings Ba and below (speculative grade). The mean absolute error (MAE) and the root mean squared error statistic (RMSE) are computed as

$$\mathrm{MAE}(t) = \frac{1}{D}\sum_{d=1}^{D} \left|\hat{\pi}^{an}_{d,t+4|t} - \bar{\pi}^{an}_{d,t+4}\right|, \quad \mathrm{RMSE}(t) = \left(\frac{1}{D}\sum_{d=1}^{D}\left[\hat{\pi}^{an}_{d,t+4|t} - \bar{\pi}^{an}_{d,t+4}\right]^2\right)^{\frac{1}{2}},$$

where index $d = 1, \ldots, D$ refers to industry groups.

The estimated and realized annual probabilities are given by

$$\hat{\pi}^{an}_{d,t+4|t} = 1 - \prod_{h=1}^{4}\left(1 - \hat{\pi}_{d,t+h|t}\right), \quad \bar{\pi}^{an}_{d,t+4} = 1 - \prod_{h=1}^{4}\left(1 - \frac{y_{d,t+h}}{k_{d,t+h}}\right),$$

respectively, where $\hat{\pi}_{d,t+h|t}$, for $h = 1, \ldots, 4$, are the forecasted quarterly probabilities for time $t + h$. To obtain the required default signals, we first forecast all factors $\hat{F}_t$, $\hat{f}^{uc}_t$ jointly using a low order vector autoregression and using the mode estimates of $\hat{F}_t$ and $\hat{f}^{uc}_t$, in-sample. Although mode estimates of f^{uc}_t are indicated by $\bar{f}^{uc}_t$, in our forecasting study we integrate them in a Gaussian vector autoregression for which mode and mean estimates are the same. This vector autoregressive model takes into account that the factors F_t and f^{uc}_t

are conditionally correlated with each other. Given the forecasts of $\hat{F}_t$ and $\hat{f}_t^{uc}$, we compute $\hat{\pi}_{d,t+h|t}$ using equations (2) and (3) and based on parameter estimates and mode estimates of the signal θ_{jt}.

Table 3 reports the forecast error statistics for five competing models. Model 0 does not contain common factors. It thus corresponds to the common practice of estimating default probabilities using long-term historical averages. We use a model with only baseline log-odds and three well-used macros (industrial production growth, changes in the unemployment rate, and the credit spread (Aaa – Baa)) as our benchmark. The benchmark model is denoted as $M0(X_t)$.

[insert Tables 3 around here]

Another version of Model 0 includes three observed variables instead of the common macro factors to forecast time-varying default probabilities; they are changes in industrial production, changes in unemployment rate, and the yield spread between Baa and Aaa rated bonds. We label the benchmark model M0(X_t). This approach is more common in the literature and here it serves as a more realistic benchmark. The results reported in Table 3 are based on out-of-sample forecasts from Models 1, 2, and 3, with their parameters replaced by their corresponding estimates as reported in Table 2.

As the main finding, 'observed' risk factors $\hat{F}_t$, the latent component f_t^{uc}, as well as the industry-specific risk factors in C_t, each contribute to out-of-sample forecasting performance for default rates, to different extents. Feasible reductions in forecasting error are substantial, and by far exceed the reductions achieved by using a few observed covariates directly.

The observed reduction in mean absolute forecasting error due to the inclusion of the three observed covariates from Model 0 is less than 2%. Using other observed risk factors provides similar results. Reductions in forecasting error increase when the observed covariates are replaced by principal components and are as high as 10% on average over the years 2001-2009. This finding shows that principal components from a large macro and finance panel can capture default dynamics more successfully.

Forecasts improve further when an unobserved component is added to the the principal components and observed industry factors. Mean absolute forecasting errors then reduce to 43%. Reductions in MAE are most pronounced when frailty effects are highest. This is the case in 2002, when default rates remain high while the economy is recovering from recession, and years 2005-2007, when default conditions are substantially better than expected from macro and financial data. Reductions of more than 40% on average are substantial and have clear practical implications for the computation of capital requirements. It is also clear that the simple AR(1) dynamics for the frailty factor are too simplistic to capture the abrupt changes in common credit conditions during the crisis of 2008. As the frailty factor is negative over 2007, the forecast of default risk over 2008 based on the AR(1) dynamics is too low. In 2009, we find that the full model including frailty again does better than its competitors. To further improve the forecasting performance of the full model in crisis situations, one could extend the dynamic behavior of the frailty factor further to include non-linearity. This is left for future research.

6 Conclusion

We have proposed a novel non-Gaussian panel data time series model with regression effects to estimate and forecast the dynamics of corporate default rates. The model combines a non-Gaussian panel data specification with the principal components of a large number of macroeconomic covariates. In an empirical application to U.S. data, the combined factors capture a significant share of the common dynamics in disaggregated default counts. We find a large and significant role for a dynamic frailty component, even after accounting for more than 80% of the variation in more than 100 macroeconomic and financial covariates. A latent component or frailty factor is thus needed to prevent a downward bias in the estimation of extreme default losses on portfolios of U.S. corporate debt. Our result also indicates that the presence of a latent factor may not be due to a few omitted macroeconomic covariates, but rather appears to capture different omitted effects at different times.

In an out-of-sample forecasting experiment, we obtain substantial reductions between 10% and 43% on average in mean absolute error when forecasting the point-in-time default probabilities using our factor structure. The forecasts from our model are particularly more accurate in times when frailty effects are important and when aggregate default conditions deviate from financial and business cycle conditions. A frailty component implies additional default rate volatility, and may contribute to default clustering during periods of stress. Practitioners who rely on observed macroeconomic and firm-specific data alone may underestimate their economic capital requirements and crisis default probabilities as a result.

References

Azizpour, S., K. Giesecke, and G. Schwenkler (2010). Exploring the sources of default clustering. *Stanford University working paper series*.

Black, F. and J. C. Cox (1976). Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance 31*(2), 351–367.

Boissay, F. and R. Gropp (2010). Trade Credit Defaults and Liquidity Provision by Firms. *Review of Financial Studies, forthcoming*.

CreditMetrics (2007). CreditMetrics (TM) - Technical Document, RiskMetrics Group. www.riskmetrics.com/pdf/dnldtechdoc/CMTD1.pdf.

Das, S., D. Duffie, N. Kapadia, and L. Saita (2007). Common Failings: How Corporate Defaults Are Correlated. *The Journal of Finance 62(1)*, 93–117(25).

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

Doornik, J. A. (2007). *Ox: An Object-Oriented Matrix Language*. Timberlake Consultants Press, London.

Duffie, D., A. Eckner, G. Horel, and L. Saita (2009). Frailty Correlated Default. *Journal of Finance 64(5)*, 2089–2123.

Duffie, D., L. Saita, and K. Wang (2007). Multi-Period Corporate Default Prediction with Stochastic Covariates. *Journal of Financial Economics 83(3)*, 635–665.

Durbin, J. and S. J. Koopman (1997). Monte Carlo Maximum Likelihood estimation for non-Gaussian State Space Models. *Biometrica 84(3)*, 669–684.

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 analysis. *Biometrika 89(3)*, 603–616.

Exterkate, P., D. van Dijk, C. Heij, and P. J. F. Groenen (2010). Forecasting the Yield Curve in a Data-Rich Environment using the Factor-Augmented Nelson-Siegel Model. *Erasmus University Rotterdam, Econometric Institute working paper*.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The Generalized Dynamic-Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association 100*, 830–840.

Giesecke, K. (2004). Correlated default with incomplete information. *Journal of Banking and Finance 28*, 1521–1545.

Giesecke, K. and B. Kim (2010). Systemic risk: What defaults are telling us. *Stanford University working paper*, 1–34.

Jorion, P. and G. Zhang (2007). Good and bad credit contagion: evidence from credit default swaps. *Journal of Financial Economics 84(3)*, 860–883.

Koopman, S. J. and A. Lucas (2008). A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk. *Journal of Business and Economic Statistics 26 (4)*, 510–525.

Koopman, S. J., A. Lucas, and A. Monteiro (2008). The Multi-Stage Latent Factor Intensity Model for Credit Rating Transitions. *Journal of Econometrics 142(1)*, 399–424.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *Statistical Algorithms for Models in State Space Form: Ssfpack 3.0.* Timberlake Consultants Press, London.

Lando, D. (2003). *Credit Risk Modelling - Theory and Applications.* Princeton University Press.

Lando, D. and M. S. Nielsen (2008). Correlation in corporate defaults: Contagion or conditional independence. *Copenhagen Business School, working paper*.

Lang, L. and R. Stulz (1992). Contagion and competitive intra-industry effects of bankrupcy announcements. *Journal of Financial Economics 32*, 45–60.

Lawley, D. N. and A. E. Maxwell (1971). *Factor Analysis as a Statistical Method.* American Elsevier Publishing, New York.

Ludvigson, S. C. and S. Ng (2007). The empirical risk-return relation: a factor analysis approach. *Journal of Financial Economics 83*, 171–222.

Massimiliano, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information . *European Economic Review 47(1)*, 1–18.

McNeil, A. and J. Wendin (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance 14(2)*, 131–149.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press.

Merton, R. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance 29(2)*, 449–470.

Stock, J. and M. Watson (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association 97(460)*, 1167–1179.

Stock, J. and M. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics 20(2)*, 147–162.

Vassalou, M. and Y. Xing (2004). Default risk in equity premiums. *Journal of Finance 109(2)*, 831–868.

Table 1: Macroeconomic and financial predictor variables

Main category	Summary listing		Total
(a) Bank lending conditions			
Size of overall lending	Total Commercial Loans Total Real Estate Loans Total Consumer Credit outst. Commercial&Industrial Loans Bank loans and investments Household obligations/income	Household debt/income-ratio Federal debt of Non-fin. sector Excess Reserves of Dep. Institutions Total Borrowings from Fed Reserve Household debt service payments Total Loans and Leases, all banks	12
Extend of problematic banking business	Non-performing Loans Ratio Net Loan Losses Return on Bank Equity Non-perf. Commercial Loans	Non-performing Total Loans Total Net Loan Charge-offs Loan Loss Reserves	7
(b) Macro and BC conditions			
General macro indicators	Real GDP Industr. Production Index Private Fixed Investments National Income Manuf. Sector Output Manuf. Sector Productivity Government Expenditure	ISM Manufacturing Index Uni Michigan Consumer Sentiment Real Disposable Personal Income Personal Income Consumption Expenditure Expenditure Durable Goods Gross Private Domestic Investment	14
Labour market conditions	Unemployment rate Weekly hours worked Employment/Population-Ratio	Total No Unemployed Civilian Employment Unemployed, more than 15 weeks	6
Business Cycle leading/ coinciding indicators	New Orders: Durable goods New orders: Capital goods Capacity Util. Manufacturing Capacity Util. Total Industry Light weight vehicle sales Housing Starts New Building Permits	Final Sales of Dom. Product Inventory/Sales-ratio Change in Private Inventories Inventories: Total Business Non-farm housing starts New houses sold Final Sales to Domestic Buyers	14
Monetary policy indicators	M2 Money Stock UMich Infl. Expectations Personal Savings Gross Saving	CPI: All Items Less Food CPI: Energy Index Personal Savings Rate GDP Deflator, implicit	8
Firm Profitability	Corp. Profits Net Corporate Dividends	After Tax Earnings Corporate Net Cash Flow	4
(c) Intern'l competitiveness			
Terms of Trade	Trade Weighted USD	FX index major trading partners	2
Balance of Payments	Current Account Balance Balance on Merchandise Trade	Real Exports Goods, Services Real Imports Goods & Services	4
(d) Micro-level conditions			
Labour cost/wages	Unit Labor Cost: Manufacturing Total Wages & Salaries Management Salaries Technical Services Wages Employee Compensation Index	Unit Labor Cost: Nonfarm Business Non-Durable Manufacturing Wages Durable Manufacturing Wages Employment Cost Index: Benefits Employment Cost Index: Wages & Salaries	10
Cost of capital	1Month Commerical Paper Rate 3Month Commerical Paper Rate Effective Federal Funds Rate AAA Corporate Bond Yield BAA Corporate Bond yield	Treasury Bond Yield, 10 years Term Structure Spread Corporate Yield Spread 30 year Mortgage Rate Bank Prime Loan Rate	10
Cost of resources	PPI All Commodities PPI Interm. Energy Goods PPI Finished Goods	PPI Industrial Commodities PPI Crude Energy Materials PPI Intermediate materials	6
(e) Equity market conditions			
Equity Indexes and respective volatilities	S&P 500 Nasdaq 100 S&P Small Cap Index	Dow Jones Industrial Average Russell 2000	10
			107

Table 2: Estimation results

We report the maximum likelihood estimates of selected coefficients in the specification for the signal or log-odds ratio (3) with parameterization $\bar{\chi}_j = \chi_0 + \chi_{1,d_j} + \chi_{2,a_j} + \chi_{3,s_j}$ for $\bar{\chi} = \lambda, \beta$. Coefficients λ refer to fixed effects or baseline log-odds, coefficients β refer to the frailty factor, and coefficients γ and δ refer to the macro and industry factors, respectively. Monte Carlo log-likelihood evaluation is based on $M = 5000$ importance samples. Data is from 1981Q1 to 2009Q4. Further details of the model specification are discussed in Section 5.3.

	Model 1: Only F_t		Model 2: Only f_t^{uc}		Model 3: All Factors	
par	val	t-val	val	t-val	val	t-val
λ_0	-2.62	12.72	-2.56	9.19	-2.88	10.24
$\lambda_{1,fin}$	0.01	0.10	0.06	0.45	0.03	0.19
$\lambda_{1,tra}$	0.19	1.13	0.18	1.16	0.19	1.24
$\lambda_{1,lei}$	0.00	0.00	-0.09	0.78	-0.04	0.31
$\lambda_{1,egy}$	-0.21	1.60	-0.05	0.25	-0.44	2.15
$\lambda_{1,ind}$	-0.11	1.28	-0.19	1.85	-0.12	1.19
$\lambda_{1,tec}$	-0.28	2.04	-0.25	2.10	-0.29	2.19
$\lambda_{2,0-3}$	-0.20	1.74	-0.23	2.06	-0.25	2.21
$\lambda_{2,4-5}$	0.26	2.39	0.17	1.46	0.14	1.38
$\lambda_{2,6-12}$	0.24	2.04	0.14	1.20	0.15	1.30
$\lambda_{3,IG}$	-7.55	14.46	-6.98	16.80	-7.56	13.69
$\lambda_{3,Ba}$	-3.26	12.07	-3.21	15.62	-3.88	13.68
$\lambda_{3,B}$	-1.21	6.51	-1.25	7.78	-1.79	7.86
β_0			0.60	4.90	0.53	4.30
$\beta_{1,fin}$			-0.14	0.81	-0.18	1.45
$\beta_{1,tra}$			0.03	0.15	0.04	0.30
$\beta_{1,lei}$			0.13	1.04	-0.00	0.03
$\beta_{1,egy}$			-0.37	1.92	0.52	2.80
$\beta_{1,ind}$			0.15	1.20	-0.17	1.93
$\beta_{1,tec}$			-0.02	0.15	-0.07	0.66
$\beta_{2,IG}$			0.36	1.18	0.07	0.18
$\beta_{2,Ba}$			0.23	1.38	0.44	1.92
$\beta_{2,B}$			0.20	2.22	0.35	2.54
γ_1^{IG}	1.37	4.02			1.44	4.47
γ_1^{Ba}	0.49	2.19			0.68	3.18
γ_1^{B}	0.44	5.01			0.63	3.86
γ_1^{Caa}	0.42	3.62			0.53	4.02
	...	...			...	...
δ_{fin}	0.18	2.33	0.17	2.31	0.17	2.00
δ_{tra}	-0.45	3.57	-0.37	2.72	-0.34	2.76
δ_{lei}	0.04	0.52	0.13	1.90	0.11	1.71
δ_{egy}	0.27	3.04	0.19	2.22	-0.03	0.34
δ_{ind}	0.02	0.29	-0.02	0.42	-0.05	1.13
δ_{tec}	0.30	5.21	0.25	3.98	0.25	3.53
δ_{rcg}	0.19	2.78	0.19	2.71	0.17	2.54
LogLik	-2660.98		-2639.43		-2595.87	

Table 3: Out-of-sample forecasting accuracy

The table reports forecast error statistics associated with one-year ahead out-of-sample forecasts of time-varying point-in-time default probabilities. Error statistics are relative to a benchmark model M0(X_t) with observed risk factors only, where X_t contains changes in industrial production, changes in unemployment rate, and the yield spread between Baa and Aaa rated bonds, see Section 5.6. We report mean absolute error (MAE) and root mean square error (RMSE) statistics for all firms (All) and speculative grade (SpG), respectively, based on all industry-group forecasts for the years $2001-2009$. The relative MAEs are also given for all industry-group forecasts, for each year. Model M0 contains constant only. Models M1, M2, and M3 contain in addition the factors F_t, f_t^{uc}, and both F_t, f_t^{uc}, respectively. The models may also contain covariates as indicated.

Model			TOTAL	Ch.MAE	2001	2002	2003	2004	2005	2006	2007	2008	2009
M0:	MAE	All	1.00	0.0%	1.05	0.66	1.62	1.08	1.01	1.04	1.02	1.06	0.80
no factors		SpG	0.99	-1.4%	1.01	0.65	1.58	1.08	1.01	1.04	1.03	1.05	0.76
	RMSE	All	1.01		1.06	0.70	1.49	1.09	1.01	1.05	1.04	1.07	0.85
		SpG	0.99		1.04	0.71	1.43	1.09	1.01	1.05	1.04	1.06	0.73
M0:	MAE	All	0.99	-0.8%	0.96	1.07	1.09	0.95	0.96	1.04	0.96	0.95	1.00
X_t, C_t		SpG	0.99	-1.3%	0.96	1.04	1.06	0.94	0.95	1.03	0.95	0.96	1.02
	RMSE	All	1.01		0.97	1.12	1.06	0.96	0.98	1.05	0.96	0.97	1.07
		SpG	1.00		0.97	1.07	1.03	0.95	0.97	1.04	0.95	0.97	1.07
M1:	MAE	All	0.91	-9.4%	0.93	0.84	1.10	0.74	0.66	0.83	0.96	0.89	1.27
F_t, C_t		SpG	0.90	-10.2%	0.99	0.82	1.04	0.75	0.64	0.81	0.96	0.91	1.25
	RMSE	All	0.92		0.89	0.77	1.04	0.77	0.71	0.85	0.97	0.88	1.34
		SpG	0.92		0.96	0.77	1.00	0.78	0.69	0.85	0.97	0.89	1.31
M2:	MAE	All	0.61	-38.7%	1.13	0.93	0.88	0.50	0.36	0.26	0.30	1.03	0.77
f_t^{uc}, C_t		SpG	0.62	-38.2%	1.07	0.79	0.89	0.52	0.35	0.30	0.34	1.05	0.76
	RMSE	All	0.65		1.15	0.88	0.80	0.55	0.41	0.28	0.35	1.17	0.76
		SpG	0.66		1.08	0.82	0.82	0.57	0.40	0.33	0.38	1.17	0.72
M3:	MAE	All	0.63	-36.9%	0.90	0.58	0.62	0.45	0.37	0.39	0.32	1.19	1.18
F_t, f_t^{uc}, no C_t		SpG	0.63	-37.4%	0.92	0.57	0.74	0.44	0.37	0.42	0.31	1.28	1.08
	RMSE	All	0.68		0.91	0.67	0.70	0.46	0.40	0.39	0.37	1.24	1.16
		SpG	0.68		0.95	0.70	0.77	0.46	0.40	0.43	0.37	1.32	1.09
M3:	MAE	All	0.57	-43.0%	0.95	0.58	0.61	0.37	0.35	0.31	0.29	1.10	0.98
F_t, f_t^{uc}, C_t		SpG	0.57	-43.2%	0.94	0.56	0.73	0.36	0.35	0.35	0.29	1.18	0.86
	RMSE	All	0.62		0.96	0.66	0.62	0.37	0.37	0.32	0.35	1.21	0.99
		SpG	0.63		0.97	0.68	0.70	0.38	0.39	0.37	0.35	1.26	0.90

Figure 1: Simulation analysis

The top panels contain the sampling distributions of R-squared goodness-of-fit statistics in regressions of $\hat{F}$ on simulated factors F, and conditional mean estimates $\hat{\mathrm{E}}[f^{uc}|y]$ on the true factor f^{uc}, respectively. The middle panels present the sampling distributions of key parameters ϕ_{uc}, β, λ_0, and γ_1. The bottom panel compares two empirical distribution functions of the t-statistics associated with the null hypothesis $H_0 : \gamma_1 = 0$. In each simulation either F or $\hat{F}$ are used to obtain simulated maximum likelihood parameter and standard error estimates. All distribution plots are based on 1000 simulations. The dimensions of the default panel are N=112, and T=100. The macro panel is of dimension N=120, and T=100.

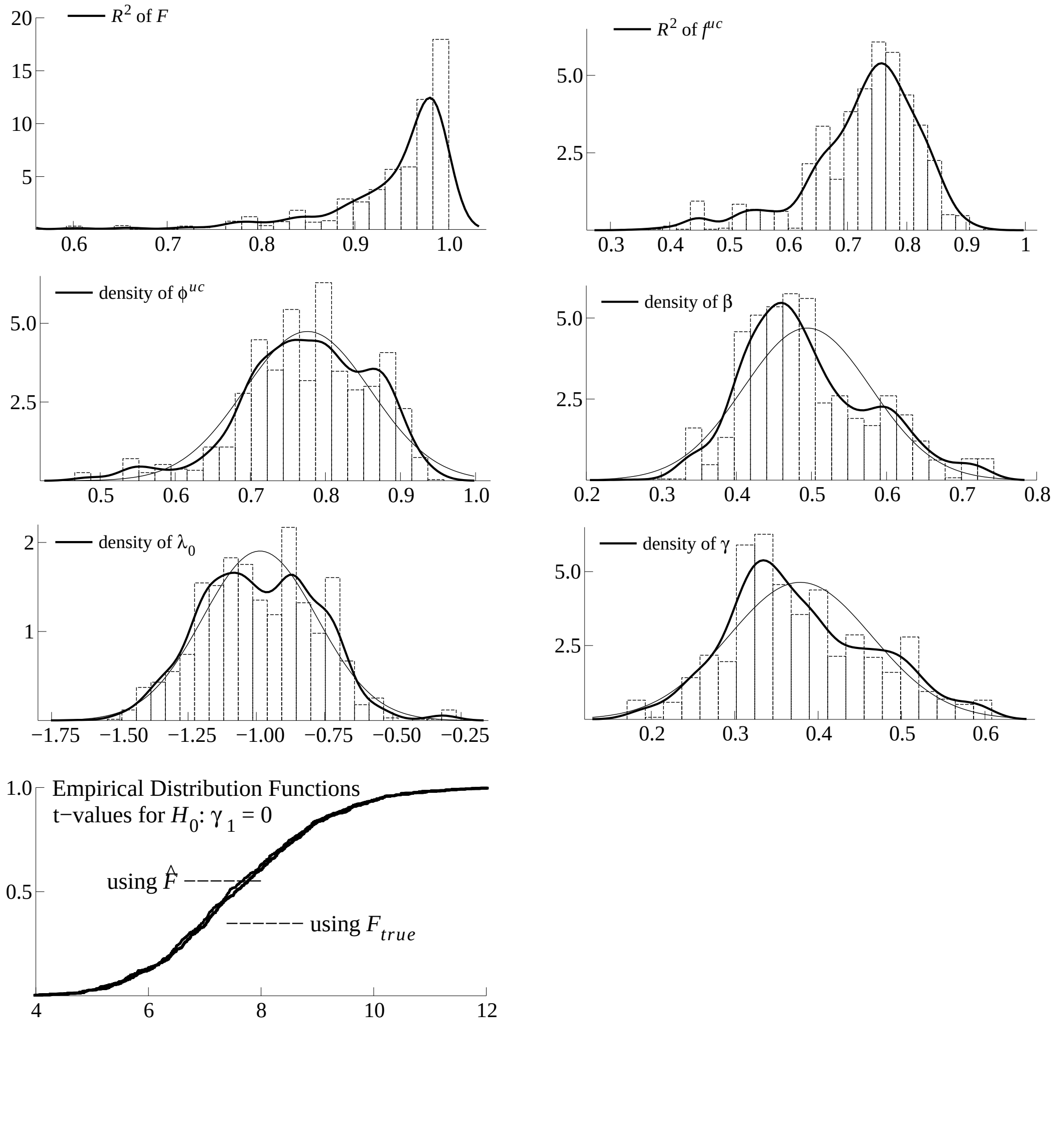

Figure 2: Aggregated default data and disaggregated default frequencies

The first three panels present time series plots of (a) the total default counts $\sum_j y_{jt}$ aggregated to a univariate series, (b) total number of firms $\sum_j k_{jt}$ in the database, and (c) aggregate default fractions $\sum_j y_{jt} \;/\; \sum_j k_{jt}$ over time. The bottom panel presents disaggregated default frequencies y_{jt}/k_{jt} over time for four the broad rating groups $Aaa - Baa$, Ba, B, and $Caa - C$. Each plot contains multiple default frequencies over time, disaggregated across industries and time from initial rating assignment.

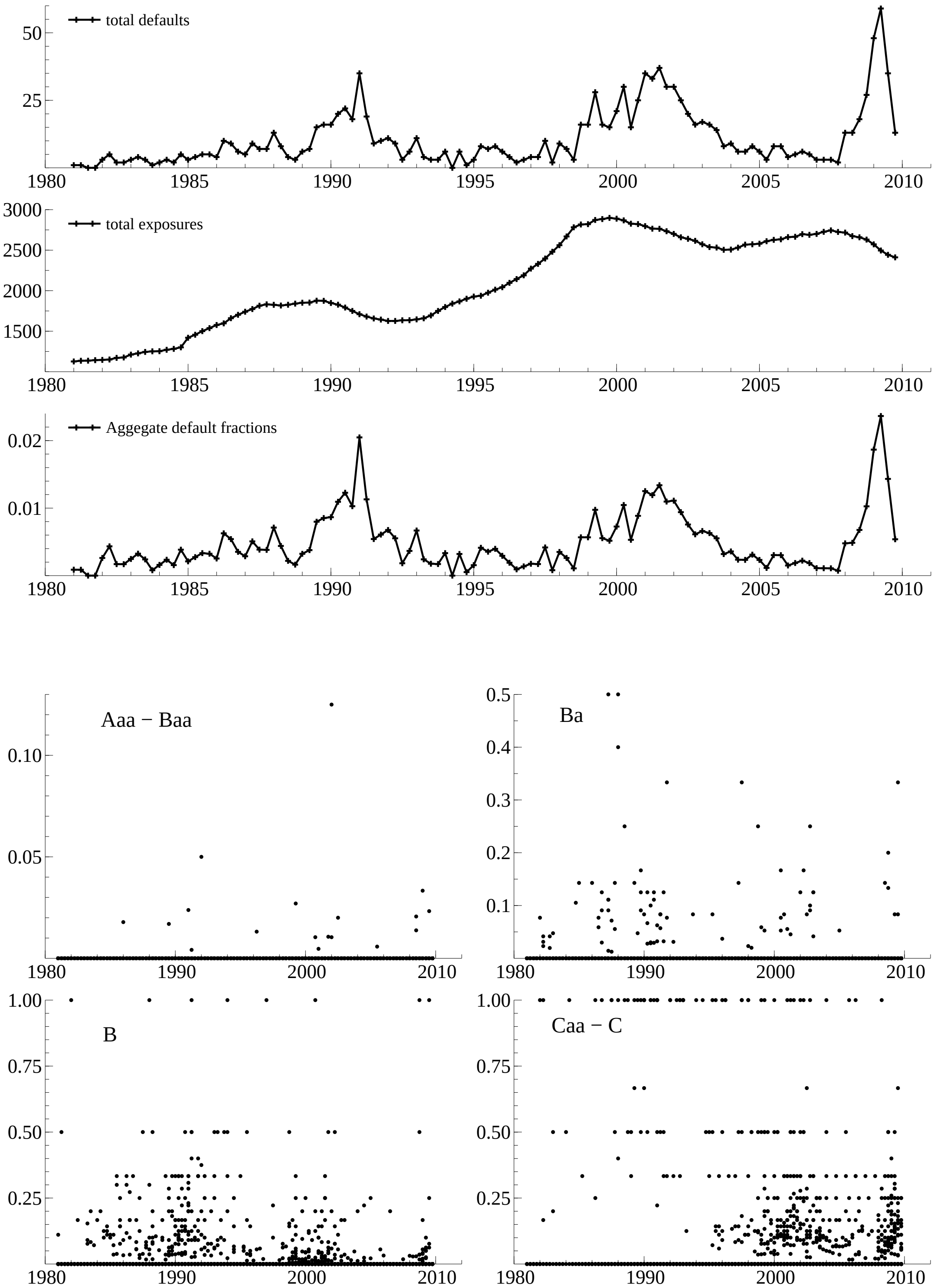

Figure 3: Macro factors from unbalanced data

We present the first ten principal components from our unbalanced panel of macro and financial time series as listed in Table 1. Shaded areas indicate NBER recession periods.

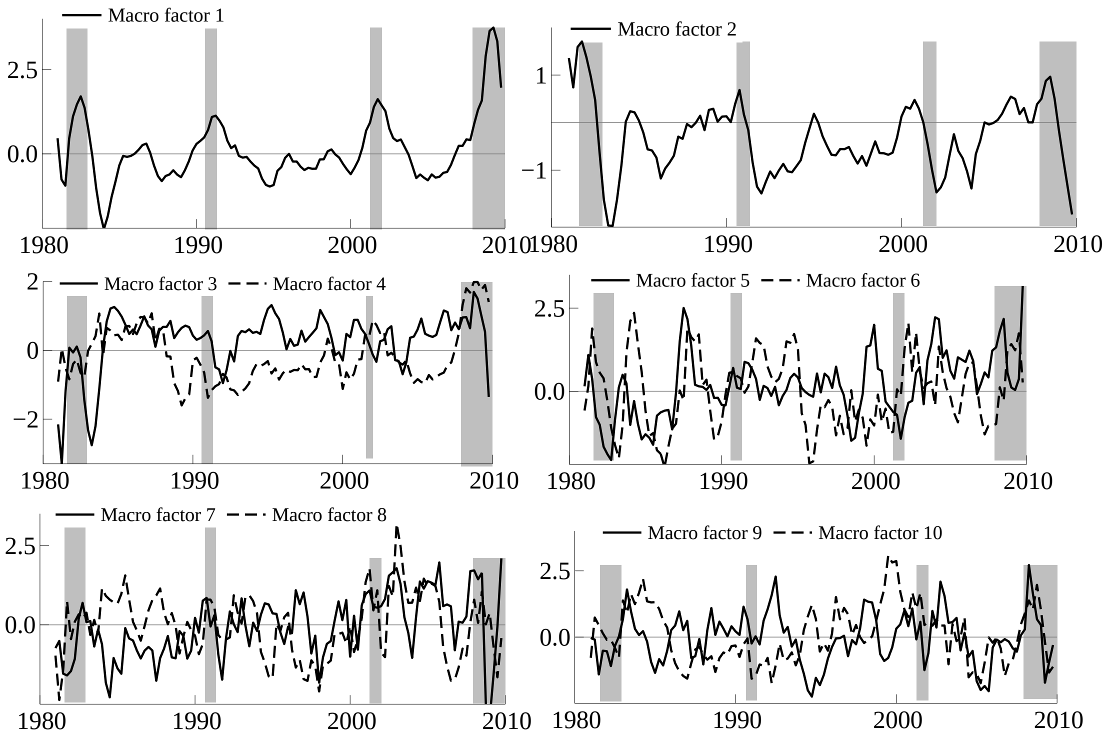

Figure 4: Frailty factor

We present estimates of f_t^{uc} from models M2 (top panel) and M3 (bottom panel). Standard error bands are obtained from conditional factor variance estimates, see Section 4.3, and correspond to a 95% confidence level.

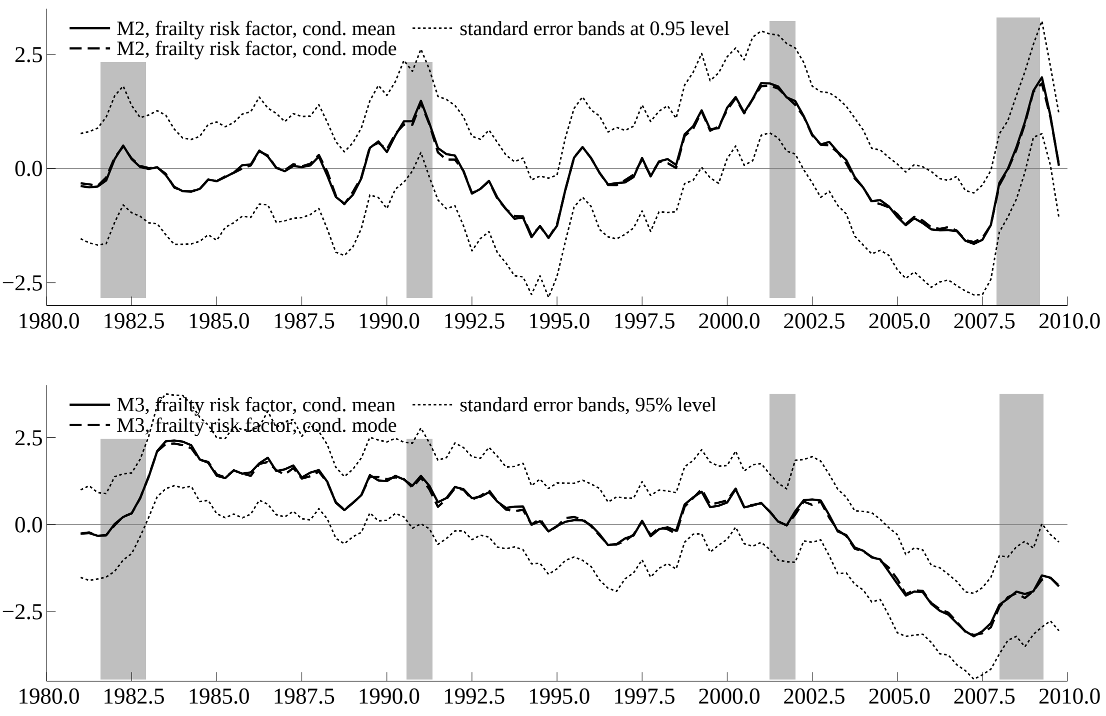

Figure 5: Smoothed default signals

The two panels present the smoothed default signals θ_{jt} for investment grade (Aaa-Baa) and low speculative grade (Caa-C) firms. The panels decompose the total default signal into estimated factors, scaled by their respective factor loadings (standard deviations). We plot variation due to the first principal component $\hat{F}_{1,t}$, all principal components $\hat{F}_{1,t}$ to $\hat{F}_{10,t}$, and all factors including the latent component $\hat{f}_t^{uc}$.