Mixed measurement dynamic factor models *

Bernd Schwaab [a,b] *Drew Creal* [d]

Siem Jan Koopman [a,c] *André Lucas* [a,b,e]

[a] Tinbergen Institute

[b] VU University Amsterdam, Department of Finance

[c] VU University Amsterdam, Department of Econometrics

[d] University of Chicago Booth School of Business

[e] Duisenberg School of Finance

*** Preliminary version ***

March 24, 2010

Mixed measurement dynamic factor models

Bernd Schwaab   *Drew Creal*

Siem Jan Koopman   *Andre Lucas*

Abstract

We propose a latent dynamic factor model framework for mixed-measurement mixed-frequency panel data. Time series observations may come from different families of parametric distributions, may be observed at different frequencies, and exhibit common dynamics and cross sectional dependence due to shared exposure to latent dynamic factors. As the main complication, the likelihood does not exist in closed form for this class of models. We therefore present three different approaches to parameter and factor estimation in this framework. First, assuming a factor structure for location parameters yields a parameter driven model that can be cast into state space form. Parameters and factors estimation is then accomplished by Monte-Carlo maximum likelihood based on importance sampling. Second, we propose a less complex observation driven alternative to the parameter driven original model, for which the likelihood exists in closed form. Finally, parameter and factor estimates can be obtained by Markov Chain Monte Carlo. We use the new mixed-measurement framework for the estimation and forecasting of intertwined credit and recovery risk conditions for US Moody's-rated firms from 1982 - 2008. The joint model allows us to construct predictive (conditional) loss densities for portfolios of bank loans and corporate bonds in the presence of non-standard sources of credit risk such as systematic frailty effects and systematic recovery risk.

Keywords: mixed-measurement mixed-frequency data; state space methods; generalized autoregressive score; latent dynamic factors; recovery rates; default risk.

JEL classification: *C32; G32*

1 Introduction

We develop a novel latent dynamic factor model for panels of mixed measurement time series data. In this framework, observations may come from different families of parametric distributions, may be observed at different frequencies, and are dependent in the cross-section due to shared exposure to latent dynamic factors. Consider available data

$$y_t = (y_{1t}, \ldots, y_{Nt})', \quad t = 1, \ldots, T, \tag{1}$$

where each row $y_i = (y_{i1}, \ldots, y_{iT})$, $i = 1, \ldots, N$, comes from a different density. We are not only thinking of simple differences in means or variances. Instead, some time series may be discrete, whereas others are continuous. Some time series may be Gaussian, while others are non-negative durations, or count data obtained from point processes. Time series data from the exponential family is often of particular interest. This family includes many well-known distributions, such as the binomial, Poisson, Gaussian, inverse Gaussian, Gamma, and Weibull distribution. The results of this paper allow us to analyze the joint variation in mixed data from the above densities, in a latent dynamic factor model setting.

In the absence of non-Gaussian or mixed data, latent factors underlying a panel of time series data can be analyzed using either (i) the method of principal components in an approximate dynamic factor model framework, see e.g. Connor and Korajczyk (1986, 1988, 1993), Stock and Watson (2002, 2005), Bai (2003), and Bai and Ng (2002, 2007), (ii) estimation procedures based on frequency domain methods, see e.g. Sargent and Sims (1977), Geweke (1977), Forni, Hallin, Lippi, and Reichlin (2000, 2005), or (iii) filtering and smoothing techniques in a state space framework, see e.g. Doz, Giannone, and Reichlin (2006), and Jungbacker and Koopman (2008). If data (1) come from different families of densities, however, none of the above methods can be used for parameter and factor estimation without modification. To the best of our knowledge, this paper is the first to present a likelihood-based analysis of a dynamic factor model for mixed measurement time series data. We refer to the model as the mixed-measurement dynamic factor model (MM-DFM).

In this paper, the main challenge is that the likelihood of the MM-DFM does not exist in closed form. Obviously, this hinders parameter and factor estimation and inference in a likelihood-based setting. We present three solutions to this problem. First, Shephard

and Pitt (1997), Durbin and Koopman (2000), and Jungbacker and Koopman (2007) show that maximum likelihood inference and latent factor estimation can be achieved by Monte Carlo maximum likelihood methods based on importance sampling techniques. We cast the MM-DFM in state space form, and demonstrate that this approach can be extended to the MM-DFM setting. Recent applications of importance sampling in a non-Gaussian framework include Koopman, Lucas, and Monteiro (2008), Koopman and Lucas (2008), and Koopman, Lucas and Schwaab (2008, 2010). .

Second, we consider a less complex observation driven model as an alternative to the parameter driven model in state space form. Here, the scaled score of the (local) log-likelihood function serves as the driving mechanism for the latent factors. This essentially eliminates the factor's second source of error. Creal, Koopman, and Lucas (2008) refer to such models as generalized autoregressive score (GAS) models. Effectively, this paper extends the class of GAS models to include a panel data model for observations from different families of parametric distributions (MM-GAS). Importantly, the likelihood exists in closed form, and can be maximized straightforwardly.

Third, we demonstrate that parameter and factor estimation in the MM-DFM framework can be performed by Bayesian techniques. Bayesian inference is particularly attractive when there is some prior information about the parameters. Also, Markov Chain Monte Carlo (MCMC) methods still work in settings with a very high dimensional factor space, where we may want to sample the factors in blocks.

As an example of a mixed-measurement and mixed-frequency panel data setting, we model the systematic variation in cross sections of corporate default counts, recovery rates on loans and bonds after default, and macroeconomic data. While defaults are discrete, the recovered percentages on the principal are continuous and bounded on the unit interval. Recovery values tend to be low precisely when defaults are high in an economic downturn, indicating important systematic covariation across different types of data from different families of parametric distributions. In addition, recovery rates, default counts, and macroeconomic indicators are available at different frequencies. It is not difficult to think of further applications from e.g. the actuarial sciences or financial market microstructure research. Mixed-measurement models are useful whenever different families of statistical distributions are appropriate for different types of data, while they may be driven by related

4

dynamics.

The remainder of this paper is as follows. We introduce the baseline mixed measurement dynamic factor model (MM-DFM) in Section 2, along with results regarding parameter estimation and signal extraction in this framework. We demonstrate how to speed up likelihood evaluations by collapsing observations, and address missing values. Section 3 introduces an observation driven MM-GAS alternative. Bayesian inference for the MM-DFM is treated in Section 4. Section 5 considers the estimation and forecasting of intertwined credit and recovery risk conditions. Section 6 concludes.

2 Mixed-measurement dynamic factor model

This section introduces a parameter driven latent dynamic factor model for variables from a broad range of densities, which we refer to as the mixed-measurement dynamic factor model (MM-DFM). Variables may be observed at different frequencies, such as monthly, quarterly, annually, etc.

2.1 Model specification

The mixed measurement dynamic factor model is based on a set of m dynamic latent factors that are assumed to be generated from a dynamic Gaussian process. For example, we can collect the factors into the $m \times 1$ vector f_t and assume a stationary vector autoregressive process for the factors,

$$f_{t+1} = \mu_f + \Phi f_t + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \Sigma_\eta), \qquad t = 1, 2, \dots, \tag{2}$$

with the initial condition $f_1 \sim \mathrm{N}(\mu, \Sigma_f)$. The $m \times 1$ mean vector μ_f, the $m \times m$ coefficient matrix Φ and the $m \times m$ variance matrix Σ_η are assumed fixed and unknown with the m roots of the equation $|I - \Phi z| = 0$ outside the unit circle and Σ_η positive definite. The $m \times 1$ disturbance vectors η_t are serially uncorrelated. The process for f_t is initialized by $f_1 \sim \mathrm{N}(0, \Sigma_f)$ where $m \times m$ variance matrix Σ_f is a function of Φ and Σ_η or, more specifically, Σ_f is the solution of $\Sigma_f = \Phi \Sigma_f \Phi' + \Sigma_\eta$.

Conditional on a factor path $\mathcal{F}_t = \{ f_1, f_2, \dots, f_t \}$, the observation $y_{i,t}$ of the ith variable

at time t is assumed to come from a certain density given by

$$y_{i,t}|\mathcal{F}_t \sim p_i(y_{i,t}; \mathcal{F}_t, \psi), \quad i = 1, \ldots, N. \tag{3}$$

For example, the observation $y_{i,t}$ could come from the exponential family of densities,

$$p_i(y_{i,t}; \mathcal{F}_t, \psi) = \exp\{a_i(\psi)^{-1} [y_{i,t}\,\theta_{i,t} - b_{i,t}(\theta_{i,t}; \psi)] + c_{i,t}(y_{i,t})\}, \tag{4}$$

with the signal defined by

$$\theta_{i,t} = \alpha_i + \sum_{j=0}^{p} \lambda'_{i,j} f_{t-j}, \tag{5}$$

where α_i is an unknown constant and $\lambda_{i,j}$ is the $m \times 1$ loading vector with unknown co-efficients for $j = 0, 1, \ldots, p$. The so-called link function in (4) $b_{i,t}(\theta_{i,t}; \psi)$ is assumed to be twice differentiable while $c_{i,t}(y_{i,t})$ is a function of the data only. The parameter vector ψ contains all unknown coefficients in the model specification including those in Φ, α_i and $\lambda_{i,j}$ for $i = 1, \ldots, N$ and $j = 0, 1, \ldots, p$. Scaling by a dispersion parameter $a_i(\psi)$ in (4) is not necessary for binary, binomial, Poisson, exponential, negative binomial, multinomial, and standard normal observations, as $a_i(\psi) = 1$ in these cases. Allowing for $a_i(\psi) \neq 1$ permits modeling observations from e.g. the Gamma, Gaussian, inverse Gaussian, and Weibull densities. In general, the results of Section 2.2 extend to densities $p_i(y_{i,t}; \mathcal{F}_t, \psi)$ which are twice differentiable with respect to their signal $\theta_{i,t}$, and $\partial^2 p_i(\cdot; \cdot)/\partial\theta_{i,t}^2 < 0$ to ensure positive implied variances. To enable the identification of all entries in ψ, we assume standard-ized factors in (2) which we enforce by the restrictions $\mu_f = 0$ and $\Sigma_f = I$ implying that $\Sigma_\eta = I - \Phi\Phi'$.

Conditional on $\mathcal{F}_t$, the observations at time t are independent of each other. It implies that the density of the $N \times 1$ observation vector $y_t = (y_{1,t}, \ldots, y_{N,t})'$ is given by

$$p(y_t|\mathcal{F}_t, \psi) = \prod_{i=1}^{N} p_i(y_{i,t}|\mathcal{F}_t, \psi).$$

The MM-DFM model is defined by the equations (2), (3) and (5).

2.2 Estimation via Monte Carlo maximum likelihood

An analytical expression for the maximum likelihood (ML) estimate of parameter vector ψ for the MM-DFM is not available. Let $y = (y_1', \ldots, y_T')'$ and $f = (f_1', \ldots, f_T')'$ denote the vector of all the observations and factors, respectively. Let $p(y|f; \psi)$ be the density of y conditional on f and let $p(f; \psi)$ be the density of f. The log-likelihood function is only available in the form of an integral

$$p(y; \psi) = \int p(y, f; \psi) \, df = \int p(y|f; \psi) p(f; \psi) \, df, \tag{6}$$

where f is integrated out. A feasible approach to computing this integral is provided by importance sampling; see, e.g. Kloek and van Dijk (1978), Geweke (1989) and Durbin and Koopman (2001). Upon computing the integral, the maximum likelihood estimator of ψ is obtained by direct maximization of the likelihood function using Newton-Raphson methods.

Importance sampling proceeds by finding a proposal distribution $g(f|y; \psi)$, called the importance density, which closely approximates $p(f|y; \psi)$ but has heavier tails. Assume that the conditions underlying the application of importance sampling hold, in particular that $g(f|y; \psi)$ is sufficiently close to $p(f|y; \psi)$ and simulation from $g(f|y; \psi)$ is feasible. Then a Monte Carlo estimate of the likelihood $p(y; \psi)$ can be obtained as

$$\tilde{p}(y; \psi) = g(y; \psi) \, M^{-1} \sum_{k=1}^{M} \frac{p(y|f^{(k)}; \psi)}{g(y|f^{(k)}; \psi)}, \qquad f^{(k)} \sim g(f|y; \psi), \tag{7}$$

where M is a large number of draws. Density $g(y; \psi)$ is the likelihood of an approximating model which is employed to obtain the samples $f^{(k)} \sim g(f|y; \psi)$, see below. A derivation of (7) is provided in the appendix A1.

For a practical implementation, the importance density $g(f|y; \psi)$ can be based on the linear Gaussian state space model

$$\tilde{y}_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \tilde{H}_t), \tag{8}$$

where the transition equation for θ_t is the same as in the original model of interest. The pseudo-observations $\tilde{y}_t$ and covariance matrices $\tilde{H}_t$ are chosen in such a way that the distri-

bution $g(f|y; \psi)$ implied by the approximating state space model is sufficiently close to the distribution $p(f|y; \psi)$ from the original non-Gaussian model. Shephard and Pitt (1997) and Durbin and Koopman (1997) argue that $\tilde{y}_t$ and $\tilde{H}_t$ can be uniquely chosen such that the mode and curvature at the mode of $g(f|y; \psi)$ match the mode and curvature of $p(f|y; \psi)$ for a given value of ψ. The following algorithm shows how an approximating model can be obtained in a MM-DFM setting.

Algorithm 1: A linear Gaussian approximating model for the mixed measurement DFM can be obtained by iterating the following steps until convergence. To this purpose define

$$\dot{p}_{i,t} = \left. \frac{\partial \log p_i(y_{i,t}|\theta_{i,t})}{\partial \theta_{i,t}} \right|_{\theta_{i,t}=\tilde{\theta}_{i,t}} \quad \text{and} \quad \ddot{p}_{i,t} = \left. \frac{\partial^2 \log p(y_{i,t}|\theta_{i,t})}{\partial \theta_{i,t} \partial \theta'_{i,t}} \right|_{\theta_{i,t}=\tilde{\theta}_{i,t}},$$

with $\theta_t = (\theta_{1,t}, \ldots, \theta_{Nt})$ and $\theta = (\theta_1, \ldots, \theta_T)$.

1. Initialize a guess $\tilde{\theta}$ of the mode.

2. Given a current guess $\tilde{\theta}$, compute $\tilde{y}_{i,t} = \tilde{\theta}_{i,t} - \ddot{p}_{i,t}^{-1}\dot{p}_{i,t}$, and $\tilde{H}_{i,t} = -\ddot{p}_{i,t}^{-1}$ for $i = 1, \ldots, N$. Let $\tilde{y}_t = (\tilde{y}_{1t}, \ldots, \tilde{y}_{Nt})'$ and $\tilde{H}_t = \text{diag}(\tilde{H}_{1,t}, \ldots, \tilde{H}_{N,t})$, for $t = 1, \ldots, T$.

3. With $\tilde{y}_t$ and $\tilde{H}_t$ from Step 2, apply the Kalman filter and smoother to the state space model (8) to obtain the smoothed estimates $\widehat{\theta}_t$ for $t = 1, \ldots, T$. Set $\tilde{\theta} = \widehat{\theta}$ as the next guess for the solution to the mode. Return to Step 2 until convergence.

A derivation of the updating equations is provided in the appendix A2. A possible metric of convergence is the sum of absolute percentage change between $\widehat{\theta}$ and $\tilde{\theta}$. We briefly describe how to implement this procedure for several examples; these will be useful in the application below.

Illustration 1: As an example for deriving the updating equations of Algorithm 1, we consider a univariate time series y_t, $t = 1, \ldots, T$, from a Binomial distribution with time-varying success probability π_t and time-varying number of trials k_t. The log-density $\log p(y_t|\pi_t) = y_t[\log \pi_t - \log(1 - \pi_t)] + k_t \log(1 - \pi_t) + \log \binom{k_t}{y_t}$ can be rewritten in terms of the canonical/natural parameter $\theta_t = \log[\pi_t/(1-\pi_t)]$ as $\log p(y_t|\theta_t) = y_t\theta_t - k_t \log[1+\exp(\theta_t)] + \log \binom{k_t}{y_t}$. The signal θ_t is assumed to exhibit factor structure (5), i.e., θ_t is an affine function of factors.

8

Differentiating the log-density with respect to its signal gives $\dot{p}_t = y_t - k_t e^{\theta_t}/(1 + e^{\theta_t})$, and $\ddot{p}_t = -k_t e^{\theta_t}/(1 + e^{\theta_t})^2$. Given a value of parameters ψ, Algorithm 1 now implies that we can match densities $p(f|y;\psi)$ and $g(f|y;\psi)$ by iterating on steps 2 and 3 with $\tilde{H}_t = k_t^{-1} e^{-\theta_t} (1 + e^{\theta_t})^2$ and $\tilde{y}_t = \tilde{\theta}_t + \tilde{H}_t \left[y_t - k_t \frac{e^{\theta_t}}{1+e^{\theta_t}} \right]$. After convergence, draws can be taken from the approximating density $g(f|y;\psi)$ to evaluate the likelihood as indicated in (7).

Illustration 2: Gaussian observations do not need to be updated because the approximating model and original model coincide. Consider Gaussian time series observations y_t with time-varying mean θ_t and fixed variance σ^2. The mean θ_t may vary due to exposure to latent dynamic factors f_t, see (5). Differentiating the Gaussian log-density with respect to its signal gives $\dot{p}_t = \sigma^{-2}[y_t - \tilde{\theta}_t]$, and $\ddot{p}_t = -\sigma^{-2}$. As a result, the updating takes the form $\tilde{H}_t = \sigma^2$, and $\tilde{y}_t = \tilde{\theta}_t + \left[y_t - \tilde{\theta}_t \right] = y_t$. The fact that no updating is necessary in this relevant case is fortunate, since it speeds up calculation of the approximating model. We also note here that Gaussian observations cancel in the calculation of the importance sampling weights, since the actual and approximating densities coincide.

Illustration 3: As a final example, we consider observations $0 < y_t < 1$ from a Beta(a,b) distribution, with $a, b > 0$. The first parameter a is often interpreted to capture mainly the 'location' of the observation according to $\mathrm{E}[y_t] = a/(a + b)$, while the second parameter b may determine the 'scale' according to $\mathrm{Var}[y_t] = ab/(a + b)^2(a + b + 1)$. Following this interpretation, we give a factor structure to the first parameter, $a = \theta_t$, to capture observed covariation with other time series of interest. The Beta log-density is given by $\log p(y_t|\theta_t; b) = -\log B(\theta_t; b) + (\theta_t - 1) \log y_t + (b - 1) \log(1 - y_t)$, with $\theta_t > 0$, $b > 0$, where $B(\theta_t; b) = \Gamma(\theta_t)\Gamma(b)/\Gamma(\theta_t + b)$ ensures that the density integrates to one. This implies $\dot{p}_t = \varphi(\theta_t + b) - \varphi(\theta_t) + \log y_t$ and $\ddot{p}_t = \varphi'(\theta_t + b) - \varphi'(\theta_t) < 0$, where $\varphi(x) = \Gamma'(x)/\Gamma(x)$ denotes the digamma function. The updating steps are formulated as above. Considering time variation in the second parameter is also possible.

To simulate values from the importance density $g(f|y;\psi)$, the simulation smoothing method of Durbin and Koopman (2002) can be applied to the approximating model (8). For a set of M draws of $g(f|y;\psi)$, the evaluation of (7) relies on the computation of $p(y|f;\psi)$, $g(y|f;\psi)$ and $g(y;\psi)$. Density $p(y|f;\psi)$ is based on (3), density $g(y|f;\psi)$ is based on the

Gaussian density for $y_{i,t} - \mu_{it} - \theta_{i,t} \sim \mathrm{N}(0, \sigma_{i,t}^2)$ (8) and $g(y; \psi)$ can be computed by the Kalman filter applied to (8), see Schweppe (1965) and Harvey (1989).

2.3 Estimation of the factors

Once an ML estimator is available for ψ, the estimation of the location of f can be based on importance sampling. It can be shown that

$$\mathrm{E}(f|y; \psi) = \int f \cdot p(f|y; \psi) \mathrm{d}f = \frac{\int f \, w(y, f; \psi) g(f|y; \psi) \mathrm{d}f}{\int w(y, f; \psi) g(f|y; \psi) \mathrm{d}f},$$

where $w(y, f; \psi) = p(y|f; \psi)/g(y|f; \psi)$. The estimation of $\mathrm{E}(f|y; \psi)$ via importance sampling can be achieved by

$$\tilde{f} = \sum_{k=1}^{M} w_k \cdot f^{(k)} \bigg/ \sum_{k=1}^{M} w_k, \tag{9}$$

with $w_k = p(y|f^{(k)}; \psi)/g(y|f^{(k)}; \psi)$, and $f^{(k)} \sim g(f|y; \psi)$. Similarly, the standard errors s_t of $\tilde{f}_t$ can be estimated by

$$s_t^2 = \left(\sum_{k=1}^{M} w_k \cdot (f_t^{(k)})^2 \bigg/ \sum_{k=1}^{M} w_k \right) - \tilde{f}_t^2, \tag{10}$$

with $\tilde{f}_t$ the tth elements of $\tilde{f}$. A derivation of (9) and (10) is provided in the appendix A1. The availability of conditional variance estimates allows us to construct estimated standard error bands around the conditional mean of the factors.

As an alternative estimator of the latent factors f_t, we may obtain the conditional mode as given by

$$\bar{f} = \mathrm{argmax}\, p(f|y; \psi). \tag{11}$$

The conditional mode indicates the most probable value of the factors given the observations. In practise, it is obtained automatically as a by-product when matching the modes of densities $p(f|y; \psi)$ and $g(f|y; \psi)$, see Algorithm 1. In practice, $\bar{f}$ and $\tilde{f}$ are usually very close, see also Section 5.

2.4 Collapsing observations

A recent result in Jungbacker and Koopman (2008) states that it is possible to collapse a $[N \times 1]$ vector of (Gaussian) observations y_t into a vector of transformed observations y_t^l of lower dimension $m < N$ without compromising the information required to estimate factors f_t via the Kalman Filter and Smoother. This subsection adapts their argument to a nonlinear mixed-measurement setting. We focus on collapsing the artificial Gaussian data $\tilde{y}_t$ with associated covariance matrices $\tilde{H}_t$, see (8) and (2).

Consider a linear approximating model for transformed data $\tilde{y}_t^* = A_t \tilde{y}_t$, for a sequence of invertible matrices A_t, for $t = 1, \ldots, T$. The transformed observations are given by

$$\tilde{y}_t^* = \begin{pmatrix} \tilde{y}_t^l \\ \tilde{y}_t^h \end{pmatrix}, \quad \text{with } \tilde{y}_t^l = A_t^l \tilde{y}_t \text{ and } \tilde{y}_t^h = A_t^h \tilde{y}_t,$$

where time-varying projection matrices are partitioned as $A_t = \begin{bmatrix} A_t^{l\prime} : A_t^{h\prime} \end{bmatrix}'$. We require (i) matrices A_t to be of full rank to prevent the loss of information in each rotation, (ii) $A_t^h \tilde{H}_t A_t^{l\prime} = 0$ to ensure that observations $\tilde{y}_t^l$ and $\tilde{y}_t^h$ are independent, and (iii) $A_t^h Z_t = 0$ to ensure that y_t^h does not depend on f. Several such matrices A_t^l that fulfill these conditions can be found. A convenient choice is presented below. Matrices A_t^h can be constructed from A_t^l, but are not necessary for computing smoothed signal and factor estimates.

Given matrices A_t, a convenient model for transformed observations $\tilde{y}_t^*$ is of the form

$$\begin{aligned} \tilde{y}_t^l &= A_t^l \theta_t + e_t^l, \\ \tilde{y}_t^h &= e_t^h \end{aligned} \quad , \quad \begin{pmatrix} e_t^l \\ e_t^h \end{pmatrix} \sim NIID \left(0, \begin{bmatrix} \tilde{H}_t^l & 0 \\ 0 & \tilde{H}_t^h \end{bmatrix} \right),$$

where $\tilde{H}_t^l = A_t^l \tilde{H}_t A_t^{l\prime}$, $\tilde{H}_t^h = A_t^h \tilde{H}_t A_t^{h\prime}$, $\theta_t = Z f_t$, and Z contains the factor loadings. Clearly, the $[N - m]$ dimensional vector $\tilde{y}_t^h$ contains no information about f_t. We can speed up computations involving the KFS recursions as follows.

Algorithm 2: Consider (approximating) Gaussian data $\tilde{y}_t$ with time-varying covariance matrices $\tilde{H}_t$, and $N > m$. To compute smoothed factors f_t and signals θ_t,

1. construct, at each time $t = 1, \ldots, T$, a matrix $A_t^l = C_t Z' \tilde{H}_t^{-1}$, with C_t such that $C_t' C_t = \left(Z' \tilde{H}_t^{-1} Z \right)^{-1}$ and C_t upper triangular. Collapse observations as $\tilde{y}_t^l = A_t^l \tilde{y}_t$.

2. apply the Kalman Filter and Smoother (KFS) to the $[m \times 1]$ low-dimensional vector $\tilde{y}_t^l$ with time-varying factor loadings $C_t^{-1\prime}$ and $\tilde{H}_t^l = I_m$.

This approach gives the same factor and signal estimates as when the KFS recursions are applied to the $[N \times 1]$ dimensional system for $\tilde{y}_t$ with factor loadings Z and covariances $\tilde{H}_t$.

A derivation is provided in Jungbacker and Koopman (2008, Illustration 4). Collapsing observations in the MM-DFM involves a tradeoff. One the one hand, less observations need to be passed through the KFS after collapsing observations. This leads to savings in computing time. On the other hand, collapsing observations requires the Choleski decomposition of a (small) $[m \times m]$ matrix at each time $t = 1, \ldots, T$, which is not required in a linear Gaussian dynamic factor model. As a result, the reductions in computing time depend on N, T, and m. Savings increase with N, and decrease with m and T.

2.5 Missing values due to mixed frequencies and forecasting

This section addresses the treatment of missing values. Missings arise easily when data is available at different sampling frequencies. Missing values also arise in out-of-sample forecasting at the end of the sample. For mixed frequency data, we suggest arranging the data on a grid at the higher frequency. For example, variables at a monthly and quarterly frequency can be arranged on a monthly grid. The quarterly series will then contains missing values. The precise arrangement may depend on whether data is a stock (point in time) or flow (a quantity over time, or average) measurement.

Missing values are accommodated easily in a state space approach. Most implementations of the Kalman filter (KF) and associated smoother (KFS) automatically assign a zero Kalman gain, zero prediction error, and large (infinite) prediction error variance to missing observations, see e.g. the implementation by Koopman, Shephard, and Doornik (2008). As a result, little extra effort is required. Some care must be taken when computing the importance sample weights $w_k = p(y|f^{(k)}; \psi)/g(y|f^{(k)}; \psi)$, $f^{(k)} \sim g(f|y; \psi)$. While $y = (y_1', \ldots, y_T')'$ may contain many missing values, the (mode) estimates of the corresponding signals $\theta = (\theta_1', \ldots, \theta_T')'$ and factors $f = (f_1', \ldots, f_T')'$ are available for all data. Some bookkeeping is therefore required to evaluate $p(y|f; \psi)$ and $g(\tilde{y}|f; \psi)$ at the corresponding values of f, or θ.

Forecasting in the MM-DFM framework has several advantages over the two-step approach of e.g. Stock and Watson (2002b). First, forecasting factors and observations in the MM-DFM framework does not require the formulation of an auxiliary model. Parameter estimation, signal extraction, and forecasting occurs in a single step. In a two-step approach, factors are extracted from a large panel of predictor variables first, and a second step relates the variable of interest to the estimated factors. A simultaneous modeling approach (i) is conceptually straightforward, (ii) retains valid inference which is usually lost in a two step approach, and (iii) ensures implicitly that the extracted common factors are related to the variable of interest.

Forecasting factors is straightforward. Forecasts f_{T+h}, for $h = 1, 2, \ldots, H$, can be obtained by treating future observations $y_{T+1}, \ldots, y_{t+H}$ as missing, and applying the estimation and signal extraction techniques of Section 2.2 to data $(y_0, \ldots, y_{T+H})$. The obtained conditional mean $\tilde{f}$ and mode forecasts $\bar{f}$ of the factors provide a location and maximum-probability forecast given observations, respectively. The mean (or median, mode) predictions of observations $(y_{T+1}, \ldots, y_{T+H})$ can be obtained as nonlinear functions of $(f_{T+1}, \ldots, f_{T+H})$.

3 Mixed measurement generalized autoregressive score models

This section introduces an observation driven alternative to the parameter driven MM-DFM by adjusting the factor (state) equation. We refer to Creal, Koopman, and Lucas (2008) who recently proposed a framework for observation-driven time-varying parameters models, which is referred to as generalized autoregressive score (GAS) models. This subsection extends the GAS family of models to include a dynamic factor model for mixed measurement panel data (MM-GAS).

3.1 Model specification MM-GAS

The observation and signal equation of the MM-DFM and MM-GAS model coincide, i.e.,

$$y_{i,t} \sim p_i(y_{i,t}|\theta_{i,t}; \psi) \qquad \theta_{i,t} = \alpha_i + \sum_{j=1}^{p} \lambda'_{i,j} f_{t-j}. \tag{12}$$

The observation densities are functions of a latent $m \times 1$ vector of factors that are assumed to come from a vector autoregressive specification. Instead of having their own source of error, the factors f_t in a GAS model are driven by the scaled score of the (local) log-density of y_t according to

$$f_{t+1} = \mu_f + \sum_{i=1}^{\bar{p}} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1}, \tag{13}$$

where μ_f is a vector of constants, and coefficient matrices A_i and B_j are of appropriate dimension $[m \times m]$ for $i = 0, \ldots, \bar{p} - 1$ and $j = 1, \ldots, q$. The scaled score s_t is a function of past observations, factors, and unknown parameters. Unknown coefficients from A_i, B_j, μ_f, etc., are collected in a vector ψ. The scaled score is given by

$$s_t = S_t \nabla_t, \tag{14}$$

where

$$\nabla_t' = \frac{\partial \log p(y_t|\theta_t; \psi)}{\partial \theta_t} \frac{\partial \theta_t}{\partial f_t}, \quad \text{and} \quad S_t = \mathrm{E}_{t-1}[\nabla_t \nabla_t']^{-1} = \mathcal{I}_t^{-1}, \tag{15}$$

such that the scaling matrix S_t is equal to the conditional Fisher information matrix. In most models of interest, the information matrix equality holds such that $S_t^{-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t'] = -\mathrm{E}_{t-1}\left[\frac{\partial^2 \log p(y_t|\theta_t; \psi)}{\partial f_t \partial f_t'}\right]$. The updating mechanism (14) for f_t is a Gauss-Newton iteration for each new observation y_t that becomes available. The updating equation is based on the (local) likelihood score and associated information matrix and therefore exploits the full density structure to update the factors. Given that factors are common across observations from different families of densities, scaling by (15) gives an automatic and model consistent way to weight the information provided by different observations.

3.2 Maximum likelihood estimation

Parameter and factor estimation by maximum likelihood for the MM-GAS model is simpler and less computationally demanding compared to the Monte Carlo methods required in the state space framework. The likelihood can be built recursively since current factors f_t, while stochastic, are perfectly predictable given past values of observations, factors, and coefficients ψ. Unknown parameters can be estimated by maximizing the log-likelihood

$$\max l(\psi) = \sum_{t=1}^{T} l(\psi; y^t, \mathcal{F}_t), \tag{16}$$

where $y^t = (y_1, \ldots, y_t)$, $\mathcal{F}_t = (f_1, \ldots, f_t)$, and $l(\psi; y^t, \mathcal{F}_t) = \log p(y_t | \mathcal{F}_t, \psi)$ for observed values y_t. Factors and likelihood increments are computed at each time t according to (13) and (16). Analytical derivatives for the score of the log-likelihood (16) can be obtained, but are usually complicated. In practise we therefore prefer to maximize the likelihood based on numerical derivatives. For a discussion whether standard asymptotic results apply, we refer to Creal, Koopman, and Lucas (2008, Section 3).

As in the MM-DFM setting of Section 2, we need to impose certain restrictions to ensure the identification of all parameters in ψ. As is common in factor models, a rotation of the factors by an invertible matrix, along with an inverse rotation of the factor loadings, yields an observationally equivalent model. As a result, we impose $\mu_f = 0$ in (13), and restrict certain factor loadings $\lambda_{i,j}$ in (12) to be rows of the corresponding identity matrix. We need to restrict as many rows of factor loadings as there are common factors in the model. Restricting the factor loadings identifies the unknown parameters in (13). This requirement is related to the scaling of $\Sigma_\eta = I - \Phi\Phi'$ in (2) to identify the factor loadings in the parameter driven framework.

We can still estimate (filtered) factors in the MM-GAS framework when portions of the panel are missing. For an unbalanced panel, we need to distinguish which part of the data is observed at each time time $t = 1, \ldots, T$. The increment in the log-likelihood for y_t, the score vector ∇_t, and scaling matrix S_{t-1} take contributions only from observed data. As in the state space model, forecasts f_{T+h} for $h = 1, 2, \ldots, H$ can be obtained by treating future observations $y_{T+1}, \ldots, y_{t+H}$ as missing. Alternatively, factors may be forecast based on the latest filtered values (imposing $A = I_m$).

4 Bayesian inference

Bayesian inference offers alternative approaches to overcome the complication that the likelihood of the MM-DFM is not available in closed form. Parameter and factor estimation by Markov Chain Monte Carlo (MCMC) is most useful when researchers have prior infor-

mation about parameters. In addition, MCMC may still work in the (rare) cases in which the importance sampler does not appear to possess a variance. In that case we would like to sample the factors in smaller chunks. An MCMC loop for parameters and factors can be constructed as follows.

4.1 Sampling the latent factors

We sample latent factors from its conditional density, i.e., $f^{(i)} \sim p(f|y, \psi^{(i-1)}, f^{(i-1)})$. For instance, this can be achieved by the simulation smoothing algorithm of Durbin and Koopman (2002) after constructing a Gaussian approximating model, see Section 2.2. The simulation smoother runs the Kalman Filter forward, the Kalman smoothing algorithm backward, and another run forward to simulate all factors in one step.

In case sampling all factors at once appears too ambitious, a single site (or block) random walk Metropolis sampler can be used. A new proposal value for factors $f_t^{(i)}$ can be constructed from previously sampled values $f_t^{(i-1)}$ by adding a vector of error terms. Adding Gaussian errors yields a (symmetric) Gaussian proposal density. The proposed new value is accepted with probability $\alpha_t = \min\left(\frac{P(f_t^{(i)})}{P(f_t^{(i-1)})}, 1\right)$, $t = 2, \dots, T - 1$, where $\frac{P(f_t^{(i)})}{P(f_t^{(i-1)})}$ is the likelihood ratio of the proposed sample and the previous sample. This likelihood ratio depends on data y and neighboring samples, $f_{t-1}^{(i-1)}$ and $f_{t+1}^{(i-1)}$. The cases $t = 1$ and $t = T$ can be handled similarly. If rejected, $f_t^{(i+1)} = f_t^{(i)}$. The scales of the error term is tuned to achieve roughly a 35% acceptance rate.

4.2 Sampling factor loadings and autoregressive parameters

We sample the parameters from their conditional density, i.e. $\psi^{(i)} \sim p(\psi|y, f^{(i)}, \psi^{(i-1)})$.

Under the assumption that the factor loadings $\lambda_{i,j}$ and signal intercepts $\lambda_{0,i}$ in (5) have a conjugate normal prior, they can be obtained by regression. They are sampled from a normal distribution in a Gibbs sampling step.

Drawing the factor autoregressive parameters in Φ requires a random-walk Metropolis step. This allows to assume a standard but non-conjugate beta prior for the autoregressive parameters. Again, the scales of the error term is tuned to achieve roughly a 35% acceptance rate.

# 5 Intertwined credit and recovery risk

Evidence from many countries in recent years suggests that collateral values and recovery rates on corporate defaults are volatile and, moreover, that they tend to go down just when the number of defaults goes up in economic recessions, see Altman, Brady, Resti, and Sironi (2003) for a survey. The inverse relationship between recovery rates and default rates has traditionally been neglected by credit risk models, treating the recovery rate as either constant or as a stochastic variable independent from the probability of default. It is now widely recognized that a failure to take these dependencies into account leads to incorrect forecasts of the loss distribution and the derived capital allocation, see Schuerman (2006).

According to the current Basel proposal, banks can opt to provide their own recovery rate forecasts for the calculation of regulatory capital, see Basel Committee on Banking Supervision (2004). As a result there is an immediate need for statistical modeling, in particular for the supervisory agencies who need to evaluate the banks' models. In credit risk practice, default counts are frequently modeled as conditionally binomial random variables, where default probabilities depend on unobserved systematic risk factors, see e.g. McNeil, Frey and Embrechts (2005, Chapter 9), and McNeil and Wendin (2007). Recovery rates take values on the unit interval, and may be given either a beta-distribution, see e.g. CreditMetrics (2007) and Gupton and Stein (2005), or a logit-normal distribution, see e.g. Düllmann and Trapp (2004) and Rösch and Scheule (2005).

Figure 1 illustrates the inverse relationship between observed default risk conditions and recovery rates. In bad times (high default rates), recoveries tend to be low, and vice versa. The effect of systematic recovery risk on the credit loss portfolio is illustrated in the second panel of 1. The unconditional loss distribution refers to a setting where loans are given to each firm in the Moody's database at the beginning of each quarter from 1982Q1 to 2008Q4. The Figure shows a histogram of the portfolio losses due to corporate defaults (i) when the recovery rate is held constant at its mean value, and (ii) when the recovery rates vary inversely with defaults, as observed in the data. Systematic recovery risk implies that credit losses become more extreme: good times become better (thicker left tail), and worse times become worse (thicker right tail). Clearly, neglecting recovery risk leads to an underestimation of risk.

Figure 1: Portfolio Loss Distributions with and without systematic recovery risk
The scatterplot in the top panel plots observed quarterly default rates for Moody's rated firms against average senior secured bond recovery rates over time. The regression line indicates an inverse relationship. The bottom panel presents a histogram of scaled historical default rates (the unconditional portfolio loss distribution) with and without systematic recovery rate risk. The panel compares the unconditional loss density (i) when recovery rates are held fixed at their mean value, and (ii) when historical recoveries vary inversely with the default rates.
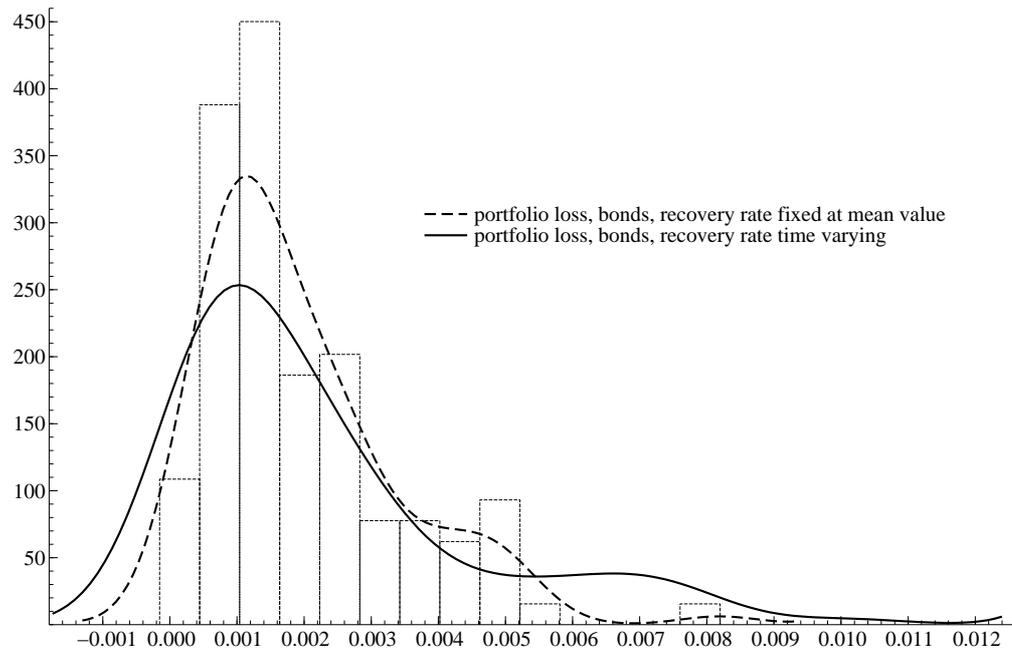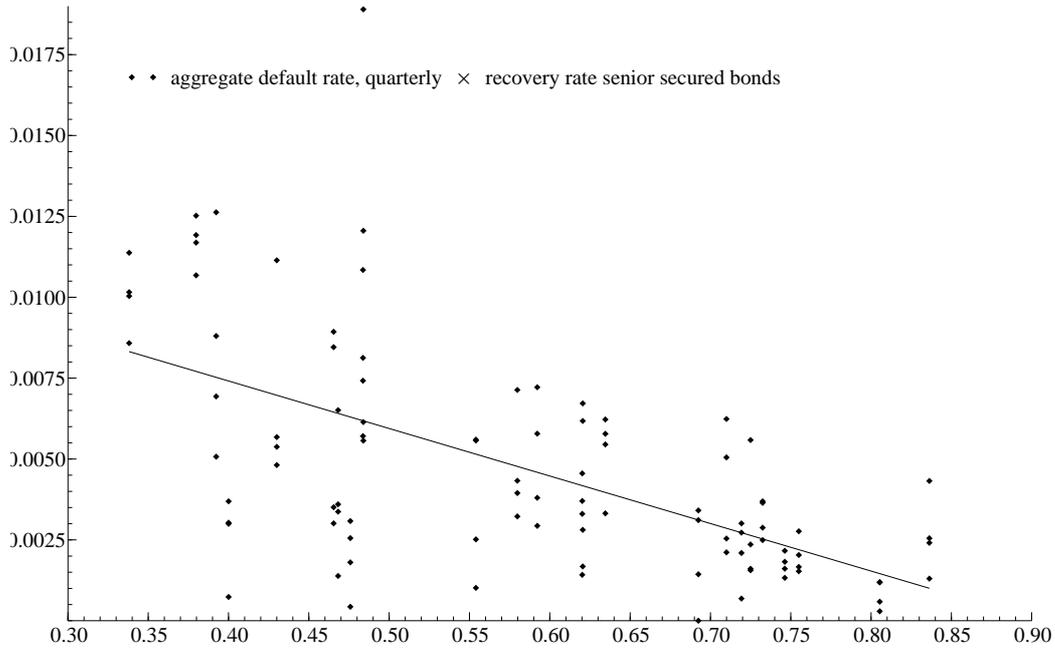
5.1 Data and mixed measurement model equations

Figure 2 contains time series plots of, from top to bottom, quarterly default counts of investment grade rated firms, quarterly default counts for firms with a speculative grade rating, annual recovery rates for collateralized bank loans, recovery rates for senior secured bonds[1], changes in the US unemployment rate, and the negative of the US industrial production growth rate. The macroeconomic indicators are standardized to zero mean and unit variance. The observations are denoted $d_{j,t}$, $r_{j,t}$, and $x_{j,t}$, respectively, where $j = 1, 2$. Macroeconomic data from January 1982 to December 2008 is obtained from the Fred St. Louis online database. Rating and default data is from Moody's. Figure 2 exhibits the clear inverse relationship between defaults and recovery rates.

No recovery rates are reported for senior secured bonds in 1984 and 1993, due to a lack of informative default events in that year. Loan recovery rates are available only from 1990 onwards, yielding a time series of 19 annual observations. Fortunately, missing values are easily accommodated using the results in Section 2.5.

A parsimonious model for mixed measurement data $y_t = (d'_t, r'_t, x'_t)'$, with common exposure to latent autocorrelated risk factors f_t, is given by

$$d_{j,t}|f_t \sim \text{Binomial}\left(k_{j,t}, [1 + e^{-\theta_{j,t}}]^{-1}\right), \quad r_{j,t}|f_t \sim \text{Beta}\left(a_{j,t}, b_j\right), \quad x_{j,t}|f_t \sim \text{N}\left(\mu_{j,t}, \sigma_j^2\right),$$
(17)

where $[1 + e^{-\theta_{j,t}}]^{-1} = \pi_{j,t}$ denotes a time-varying default probability within the unit interval. Location parameters for each observation are given by

$$\theta_{j,t} = c_{\theta,j} + \beta'_j f_t, \quad a_{j,t} = c_{a,j} + \gamma'_j f_t, \quad \mu_{j,t} = c_{\mu,j} + \delta'_j f_t,$$

where $c_{\theta,j}, c_{a,j}, c_{\mu,j}$ are intercept terms and $\beta_j, \gamma_j, \delta_j$ are factor loadings. Unknown coefficients and factors can be estimated as outlined in Section 2.

---

[1]Bond recovery rates are defined as the ratio of the market value of the bonds to the unpaid principal, one month after default, averaged across the bonds that default in a given year.

5.2 Major empirical findings

Figure 2 compares in-sample predictions for defaults, bond and loan recovery rates, and business cycle data to observed data. The single factor MM-DFM ($m = 1$) already gives an acceptable fit to the default counts and bond recovery rates. However, the fit to loan recovery rates and macroeconomic data is less satisfactory. This discrepancy may indicate that systematic default and recovery rate risk is related to, but different from, standard business cycle risk. This would confirm the related findings in Das, Duffie, Kapadia, and Saita (2007) and Bruche and Gonzalez-Aguado (2009). Extending the dimensionality of f_t yields a better fit, in particular for the macroeconomic indicators and bond recovery rates.

The corresponding plots for the MM-GAS model are reported in Figure 3. For the current data, the observation driven alternative is able to replicate the in-sample fit of the MM-DFM.

When estimating MM-DFM models, we assume that the assumptions underlying the application of importance sampling hold. In particular, $g(f|y; \psi)$ needs to approximate $p(f|y; \psi)$ sufficiently closely to ensure that the importance sampler possesses a variance. This guarantees a square root speed of convergence and asymptotic normality of the importance sampling estimators, see Geweke (1989). Some graphical diagnostics are presented in Figure 4. We present the largest 100 (log) importance sampling weights, a density plot, and a recursive variance estimate for 10000 importance sampling weights associated with models with $m = 1, 2, 3$ factors. There is no indication that a few extremely large weights dominate. The recursive variance estimates appear to converge. The largest weight accounts for less than 1% of the total sum of weights in all cases. However, the weights appear to become less well-behaved as more factors are added. Statistical tests for a finite variance are presented in Koopman, Shephard, and Creal (2009).

Figure 5 presents the conditional mean and conditional mode estimates for the three latent dynamic factors underlying the predictions in Figure 2. Both factor estimates are extremely close. The MM-GAS factors track the reported common factors from the MM-DFM.

Figure 2: MM-DFM: Actual vs predicted values

The figure plots the actual versus predicted values of (i) default counts of firms rated investment grade and speculative grade, respectively, (ii) bank loan recovery rates, and recovery rates for senior secured bonds, and (iii) changes in the unemployment rate, yoy, and negative changes in industrial production. Defaults are quarterly data, recovery rates are annual data, and macro data is monthly data. Predicted values are obtained from a model specification with $m = 2$ and $m = 3$ factors, respectively.
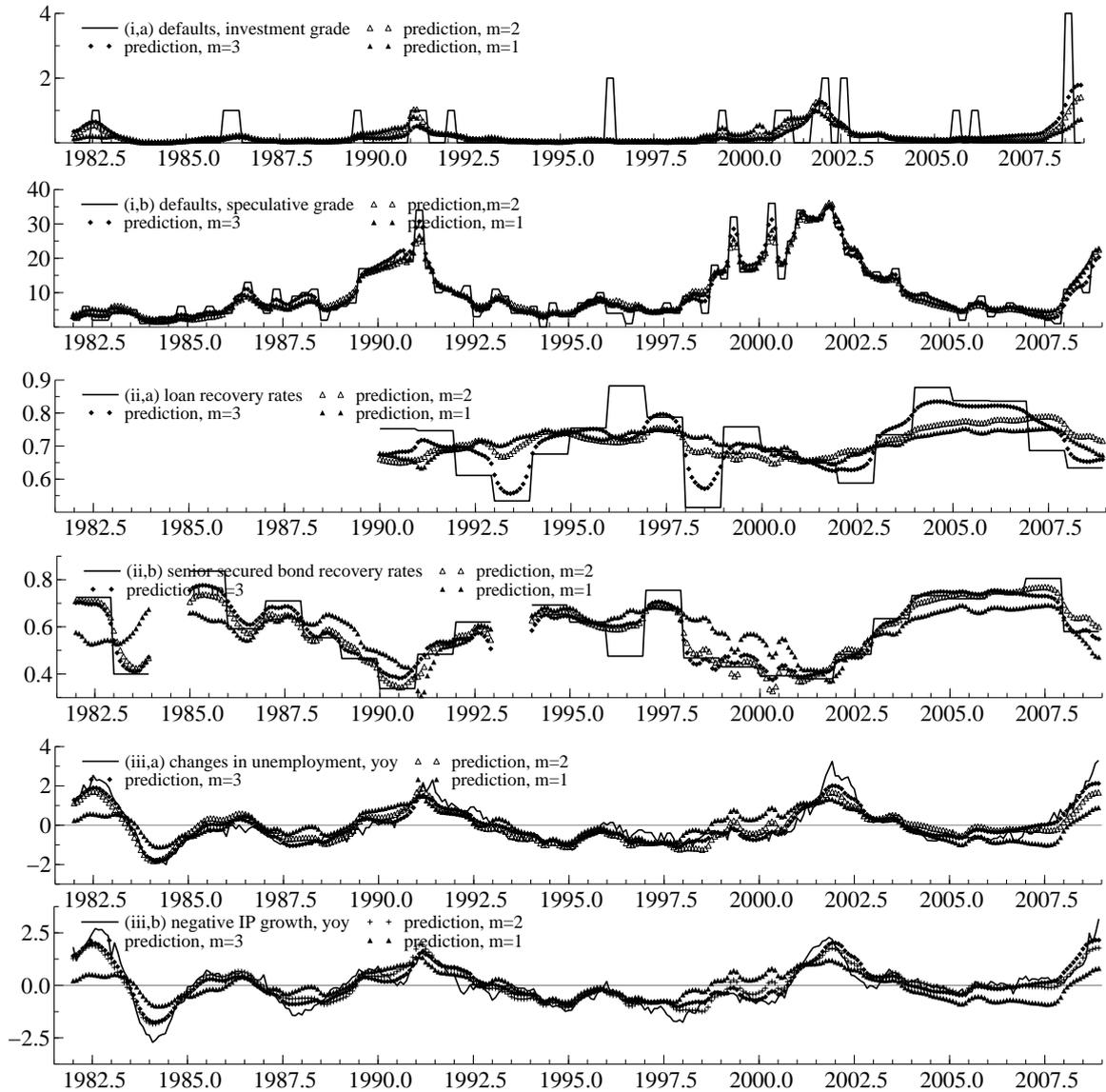
Figure 3: MM-GAS: Actual vs predicted values

The figure plots the actual versus predicted values of (i) default counts of firms rated investment grade and speculative grade, respectively, (ii) bank loan recovery rates, and recovery rates for senior secured bonds, and (iii) changes in the unemployment rate, yoy, and negative changes in industrial production. Defaults are quarterly data, recovery rates are annual data, and macro data is monthly data. Predicted values are from a multi-factor MM-GAS model specification with $m = 3$ factors.
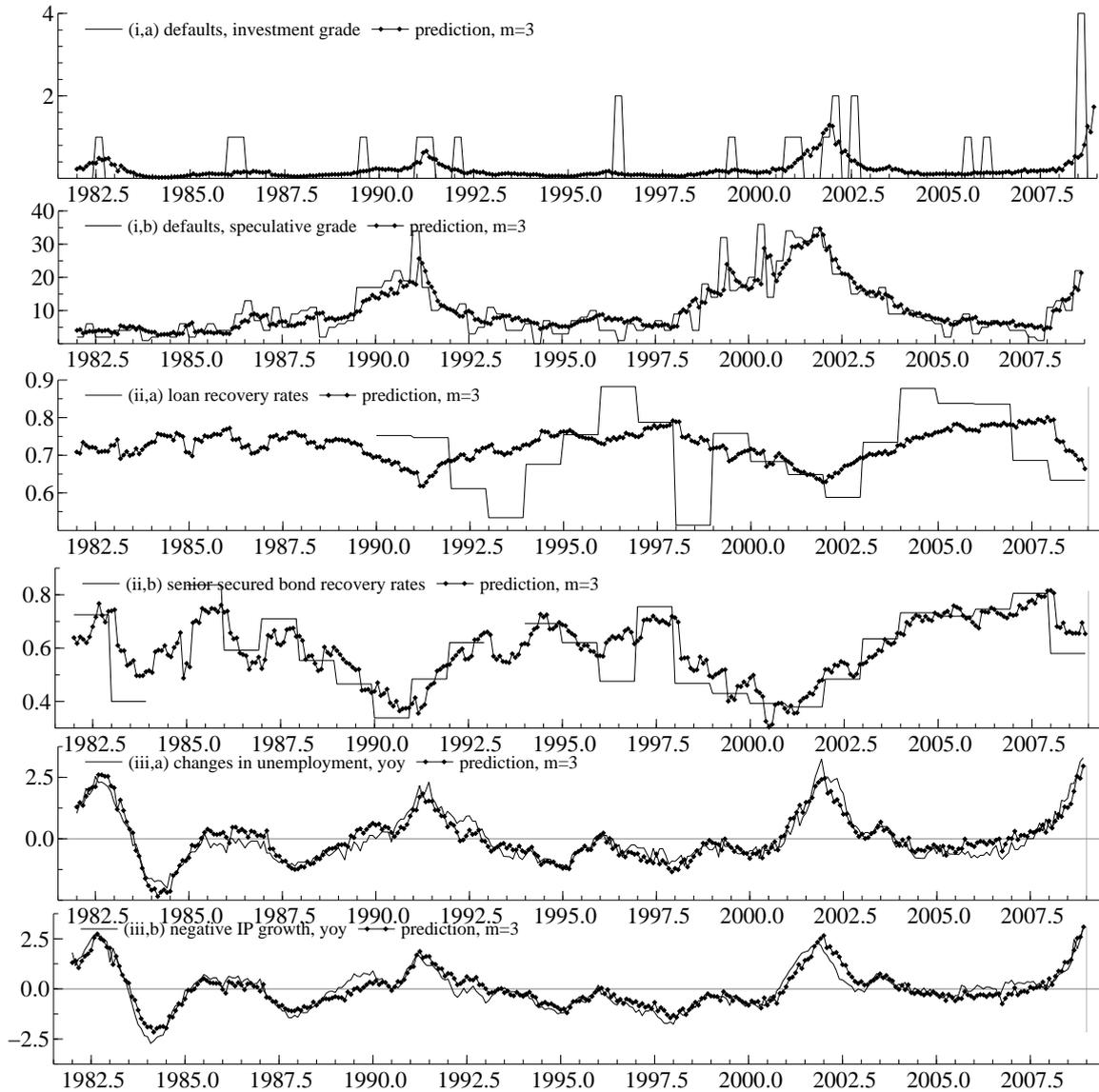
Figure 4: Importance sampling weights

The figure presents the largest 100 importance sampling weights, a density plot, and a recursive variance estimate for a total of 10000 simulated weights. The rows correspond to an empirical model specification with $m = 1$, $m = 2$, and $m = 3$ latent factors, respectively.
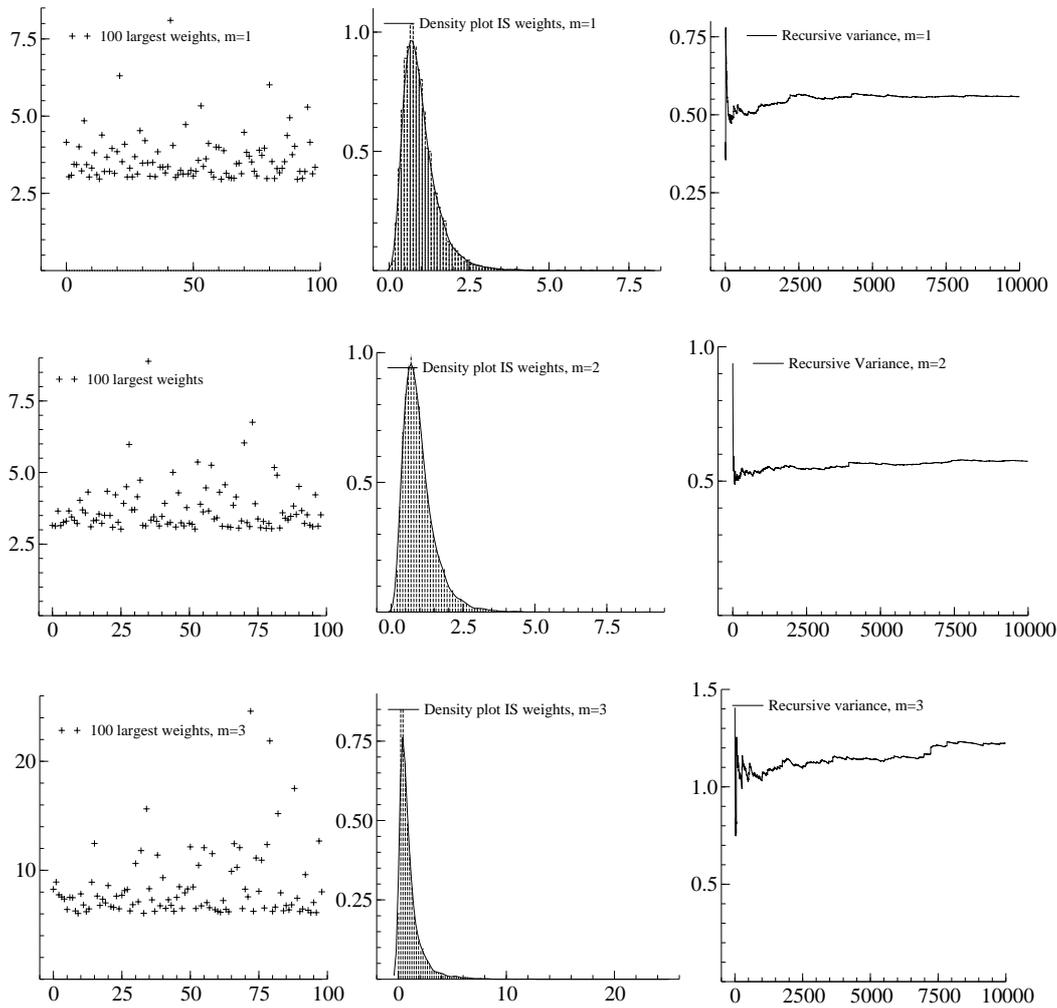
Figure 5: Latent factor estimates

The figure plots the estimates for three latent factors from a multi-factor ($m = 3$) MM-DFM and MM-GAS model specification, respectively. We report the conditional mean and mode estimates for the MM-DFM (left), and filtered factors for the MM-GAS (right). Standard error bands for the conditional mean of the factors are at a 0.95 confidence level.
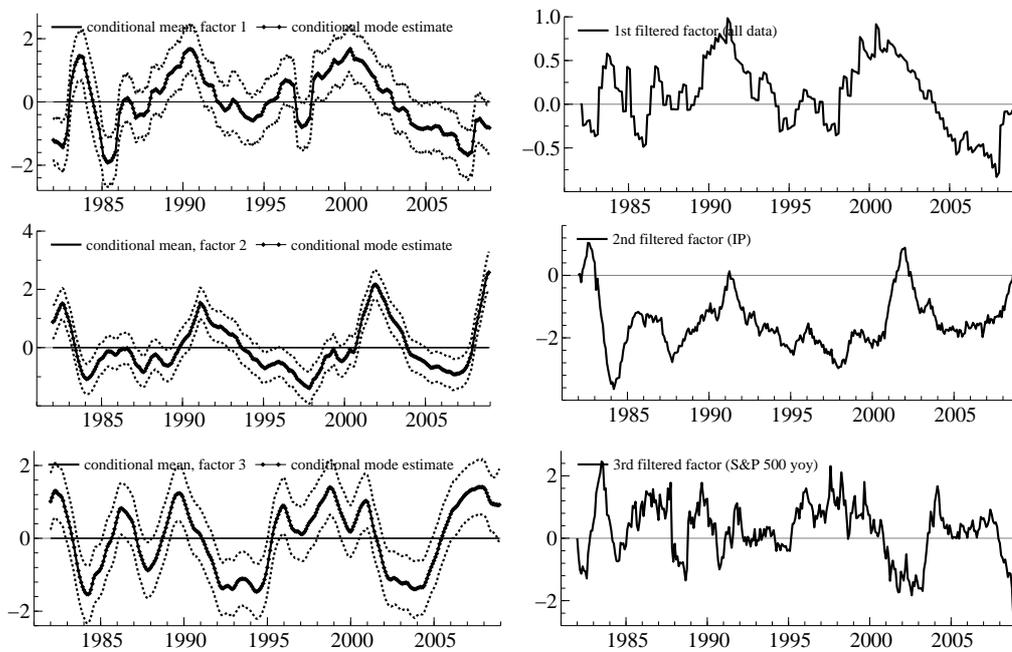
5.3   Out of sample evaluation

This section compares the out-of-sample predictions of several models for mixed measurement data. We consider four model which differ widely in their degree of sophistication. We consider

1. a random walk forecast, assuming the last year's rates will remain the same,

2. a low order unrestricted vector autoregression, VAR (2), fitted on quarterly data for default rates, recovery rates, and macroeconomic time series. Missing data is replaced straightforwardly by its last known values.

3. the parameter driven MM-DFM, estimated by state space methods for different values of m. Recovery rates are fitted using time-varying parameter versions of the beta and logit-normal distribution.

4. several observation driven MM-GAS models, for different values of m.

Each model is used to produce an out-of-sample forecast of the (i) default rate for both investment grade and speculative grade rated issuers over the next year, (ii) loan and senior secured bond recovery rates for defaulted debt over the next year, and (iii) the annual change in US industrial production.

Table 1 presents the mean absolute error (MAE) and root mean square error (RMSE) statistics associated with one year ahead forecasts. Simple models, such as the VAR(2) and the Random Walk, do relatively well in forecasting. This holds in particular for random walk forecasts for the recovery rates, and the VAR forecasts of the default rate.

The MM-GAS model does at least as well as the more complex MM-DFM when predicting default rates. It also beats the Random Walk forecasts for default rates. This means that the increase in model tractability and estimation speed of the MM-GAS model compared to the MM-DFM does not come at the cost of reduced forecasting power. Extending the dimensionality of f_t for the factor models (MM-DFM and MM-GAS) tends to help for the prediction of some variables (speculative grade default rates, loan and bond recovery rates), but not for others (investment grade default rates, annual IP growth). Table 1 further suggests that the conditionally logit-normal and beta specifications for recovery rates do

Table 1: Out of Sample Prediction Errors

The table presents the mean absolute error (MAE) and root mean square error (RMSE) statistics associated with out-of-sample point forecasts from different competing models.

MAE stats					
Model	IG def rate	SG def rate	loan rr	bond rr	ip growth
RW	1.09	0.49	0.11	0.14	2.03
VAR(2)	0.68	0.39	0.12	0.15	1.73
GAS(m=1)	0.88	0.43	0.13	0.16	1.77
GAS(m=2)	0.64	0.45	0.14	0.13	1.77
GAS(m=3)	0.75	0.41	0.15	0.14	1.94
GAS(m=4)	0.78	0.39	0.15	0.14	1.95
GAS(m=5)	0.76	0.32	0.13	0.17	1.94
SS(m=1, b)	0.69	0.49	0.12	0.18	1.72
SS(m=2, b)	0.77	0.46	0.13	0.16	2.07
SS(m=1, ln)	0.71	0.49	0.13	0.18	1.75
SS(m=2, ln)	0.74	0.44	0.14	0.14	1.98
SS(m=3, ln)	0.76	0.53	0.13	0.16	2.03
1/4 combi	0.78	0.42	0.12	0.15	1.80

RMSE stats					
Model	IG def rate	SG def rate	loan rr	bond rr	ip growth
RW	1.46	0.57	0.13	0.17	2.68
VAR(2)	0.85	0.48	0.14	0.18	2.31
GAS(m=1)	1.08	0.51	0.15	0.20	2.30
GAS(m=2)	0.81	0.53	0.15	0.18	2.30
GAS(m=3)	0.91	0.46	0.17	0.18	2.45
GAS(m=4)	0.92	0.43	0.17	0.18	2.44
GAS(m=5)	0.93	0.41	0.15	0.21	2.44
SS(m=1, b)	0.86	0.55	0.13	0.21	2.43
SS(m=2, b)	0.94	0.53	0.14	0.18	2.71
SS(m=1, ln)	0.86	0.55	0.14	0.20	2.45
SS(m=2, ln)	0.89	0.50	0.15	0.17	2.63
SS(m=3, ln)	0.93	0.61	0.16	0.18	2.63
1/4 combi	0.96	0.45	0.13	0.16	2.40

approximately equally well in prediction. The beta density seems slightly better for m=1, while the logit-normal specification is better for m=2. Both choices are reasonable.

A combined forecast from four models [the Random Walk, the VAR(2), a MM-GAS model with five factors, and a MM-DFM with one factor, equal weighting] is often among the best three forecasts, and never among the three worst forecasts. The combined forecast has low prediction RMSEs, in particular for both recovery rates and speculative grade default rates. We conclude that the combined forecasts from two relatively simple (random walk, VAR(2)) and two sophisticated (MM-DFM, MM-GAS) models appears to give good joint forecasts of default rates and recovery rates.

Figure 6 plots the out-of-sample point forecasts for default and recovery rates for the recession year 2008. The forecasted levels are similar for the MM-DFM and MM-GAS model. In this particular case, the MM-DFM delivers better joint forecasts of defaults and recoveries than the GAS model. This finding can be explained by the fact that we have restricted (identified) two GAS factors to load mostly on macro data.

Figure 6 also presents the predictive density for the 2008 credit portfolio loss, conditional on macro, default, and recovery data from 1982 to 2007. Both the MM-DFM and MM-GAS model imply capital buffers that would have ensured the solvency of a financial institution during that recession year.

6 Conclusion

We introduced a new latent dynamic factor model framework (MM-DFM) for time series observations from different families of distributions and mixed sampling frequencies. Parameter and latent factor estimates can be obtained by Monte-Carlo maximum likelihood methods based on importance sampling techniques. For the modeling of large data sets, we suggest two ways to increase computational speed. First, we collapse observations into a lower-dimensional space such that less observations are passed through the Kalman Filter and Smoother for each evaluation of the log-likelihood. Second, we consider a less complex observation-driven alternative model (MM-GAS) where the factors are driven by the scaled score of the (local) log-likelihood. Missing values arise due to mixed frequencies and forecasting, and can be accommodated straightforwardly in either the MM-DFM and MM-GAS

Figure 6: Out-of-sample forecasts for 2008

The two panels plots the 1982 to 2007 in-sample predictions for quarterly investment and speculative grade default rates, and loan and bond recovery rates. Out-of-sample point forecasts for 2008 are based on a multi-factor ($m = 3$) MM-DFM (top panel) and MM-GAS model (bottom panel), respectively. We also plot the simulated out-of-sample predictive density for the portfolio credit loss based on bond default and recovery rate data.
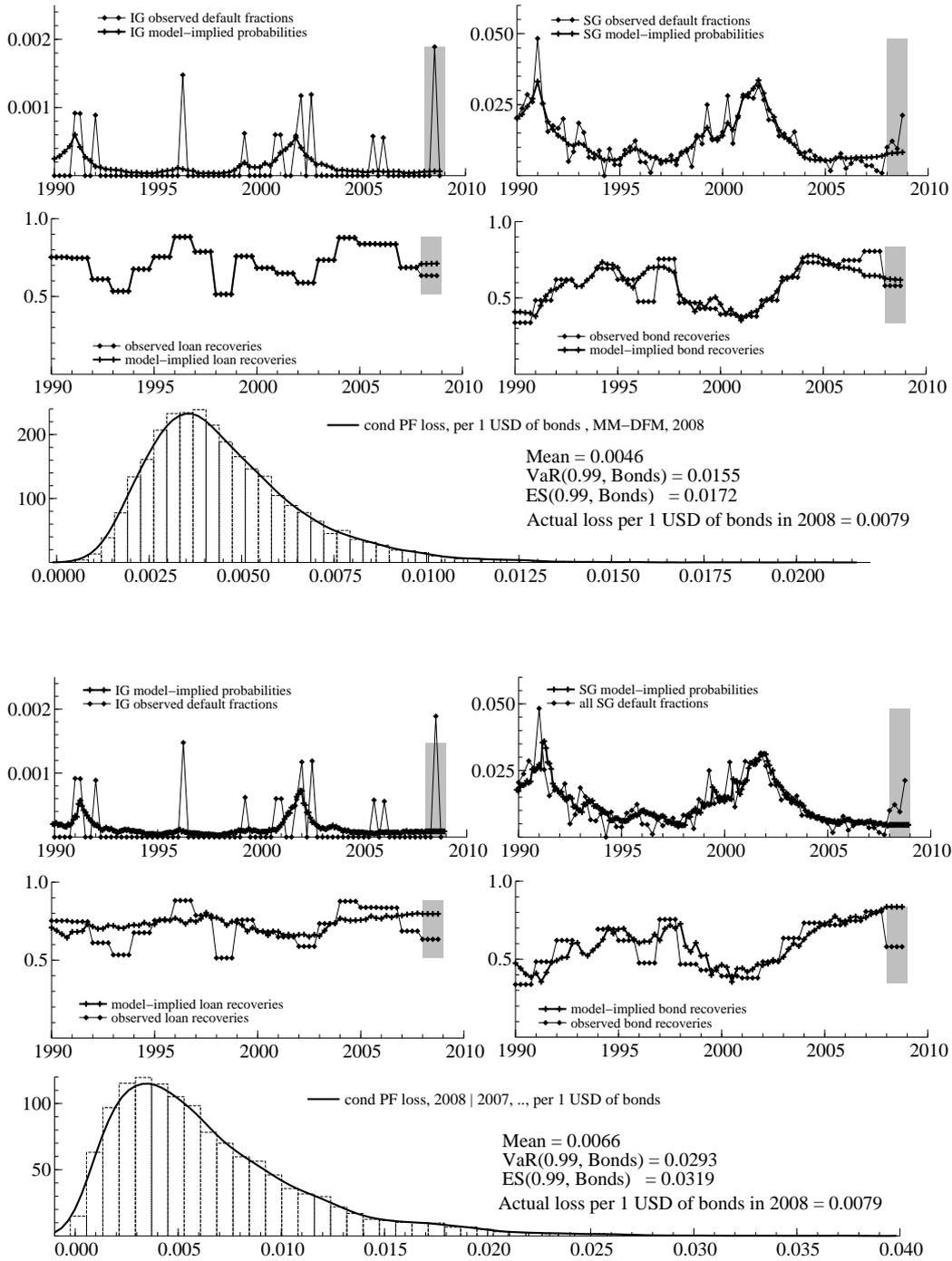
framework. In an empirical application of the mixed-measurement framework we model the systematic variation in US corporate default counts and recovery rates from 1982 - 2008. We estimate and forecast intertwined default and recovery risk conditions, and demonstrate how to obtain the predictive credit portfolio loss distribution. While the MM-GAS model is simpler and computationally more efficient than the MM-DFM, we do not nd that its reduced complexity comes at the cost of diminished out-of-sample (point) prediction accuracy.

A1. Derivation of importance sampling estimators

Equations (7), (9) and (10) are derived below. Using importance sampling to estimate parameters and factors in nonlinear non-Gaussian models is not new, we refer to Shephard and Pitt (1997), and Durbin and Koopman (1997, 2000). For given parameters ψ, consider the estimation of the mean of an arbitrary function of the factors, $x = x(f)$, where $f = (f_1', \ldots, f_T')'$, conditional on mixed measurement data $y = (y_1', \ldots, y_T')'$,

$$\tilde{x} = \mathrm{E}\left[x(f)|y\right] = \int x(f) p(f|y; \psi)\, \mathrm{d}f.$$

There is no analytical solution for this problem. Denoting a suitable Gaussian importance density by $g(f|y; \psi)$,

$$\tilde{x} = \int x(f) \frac{p(f|y; \psi)}{g(f|y; \psi)} g(f|y; \psi)\, \mathrm{d}f = \mathrm{E}_g\left[x(f) \frac{p(f|y; \psi)}{g(f|y; \psi)}\right] = \frac{g(y; \psi)}{p(y; psi)} \mathrm{E}_g\left[x(f) \frac{p(f, y; \psi)}{g(f, y; \psi)}\right], \qquad (A.18)$$

where E_g denotes expectation with respect to $g(f|y; \psi)$. Setting $x(f) \equiv 1$ gives

$$1 = \frac{g(y; \psi)}{p(y; \psi)} \mathrm{E}_g\left[\frac{p(f, y; \psi)}{g(f, y; \psi)}\right], \qquad (A.19)$$

and thus

$$p(y; \psi) = g(y; \psi) \mathrm{E}_g\left[\frac{p(f, y; \psi)}{g(f, y; \psi)}\right], \qquad (A.20)$$

The Monte Carlo estimator (7) is the empirical counterpart to (A.20). It is of the same form as the estimator presented in Durbin and Koopman (1997). A law of large numbers, such as Khinchin's WLLN, ensures convergence under relatively weak conditions, see Geweke (1989).

Dividing (A.20) by (A.19) yields

$$\tilde{x} = \frac{\mathrm{E}_g\left[x(f) w(f, y; \psi)\right]}{\mathrm{E}_g\left[w(f, y; \psi)\right]}, \qquad (A.21)$$

where

$$w(f, y; \psi) = \frac{p(f, y; \psi)}{g(f, y; \psi)} = \frac{p(y|f; \psi)\, p(f; \psi)}{g(y|f; \psi)\, g(f; \psi)} = \frac{p(y|f; \psi)}{g(y|f; \psi)}.$$

The last equality uses the fact that the marginal distribution of the state is Gaussian, $p(f; \psi) = g(f; \psi)$. The weights $w_k = p(y|f^{(k)}; \psi)/g(y|f^{(k)}; \psi)$, $f^{(k)} \sim g(f|y; \psi)$ are i.i.d. by construction. Choices $x(f) = f$ and $x(f) = f^2$ in (A.21) give an expression for the first two conditional moments of f. A law of large numbers implies convergence of the empirical counterparts in (9) and (10). ∎

A2. Derivation of Algorithm 1

We adapt a general argument for non-Gaussian models in state space form to the MM-DFM setting, compare Durbin and Koopman (2001), p. 192. For original work on importance sampling in a non-Gaussian framework we refer to Shephard and Pitt (1997), and Durbin and Koopman (1997, 2000). The dependence of observation

densities on unknown parameters ψ is suppressed. The linear Gaussian approximating model is of the form (8) and (2). Let $g(f|y)$ and $g(f,y)$ be generated by the Gaussian approximating model, and let $p(f|y)$ and $p(f,y)$ be the corresponding densities as generated by the mixed model (2), (3) and (5). We seek artificial data $\tilde{y}_t$ and variances $\tilde{H}_t$ such that the densities $g(f|y)$ and $p(f|y)$ have the same mode $\bar{f}$. The initialization condition for the unobserved factors is given by their stationary distribution, $g(f_1) = N(0, I_m)$. The (non-diffuse) initialization of factors and the time-invariance of MM-DFM system matrices Φ, Σ_η, .., simplify the exposition.

In the Gaussian model, the joint density $g(f,y)$ is given by

$$\log g(f,y) = \text{const} - \log g(f_1) - \frac{1}{2} \sum_t^T (f_{t+1} - \Phi f_t)' \Sigma_\eta^{-1} (f_{t+1} - \Phi f_t) - \frac{1}{2} \sum_t^T (y_t - Z f_t)' H_t^{-1} (y_t - Z f_t),$$

where $y_t \sim N(\theta_t, H_t)$ and signals are expressed as $\theta_t = Z f_t$. The conditional mode of $\log g(f|y) = \log g(f,y) - \log g(y)$ can be obtained as the solution to the first order condition

$$\frac{\partial \log g(f,y)}{\partial f_t} = (d_t - 1)f_1 - d_t \Sigma_\eta^{-1}(f_t - \Phi f_{t-1}) + \Phi' \Sigma_\eta^{-1}(f_{t+1} - \Phi f_t) + Z' H_t^{-1}(y_t - Z f_t) = 0, \qquad \text{(A.22)}$$

where $t = 1, \ldots, T$, $d_1 = 0$ and $d_t = 1$ for $t = 2, \ldots, T$, together with $\Sigma_\eta^{-1}(f_{T+1} - \Phi f_T) = 0$. Since $g(f|y)$ is Gaussian, the conditional mode $\bar{f}$ is equal to the conditional mean $\tilde{f} = E[f|y]$. The conditional mean is calculated efficiently by the Kalman filter and smoother (KFS), see e.g. Durbin and Koopman (2001), Chapter 4. It follows that the KFS recursions solve equation (A.22).

Assuming that the MM-DFM is sufficiently well-behaved, the mode of $\log p(f|y) = \log p(f,y) - \log p(y)$ is the solution to the vector equation

$$\partial \log p(f,y)/\partial f = 0, \qquad \text{(A.23)}$$

where $\log p(f,y) = \text{const} + \sum_{t=1}^T \log p(\eta_t) + \sum_{t=1}^T \log p(y_t|\theta_t)$, and $\eta_t = f_{t+1} - \Phi f_t$ as above. Thus, condition (A.23) becomes

$$\frac{\partial \log p(f,y)}{\partial f} = (d_t - 1)f_1 + d_t \frac{\partial \log p(\eta_{t-1})}{\partial \eta_{t-1}} - \Phi' \frac{\partial \log p(\eta_t)}{\partial \eta_t} + Z' \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} = 0, \qquad \text{(A.24)}$$

where $d_1 = 0$ and $d_t = 1$ for $t = 2, \ldots, T$. The first three terms of (A.24) and (A.22) are identical. The difference in the last terms is due to the observation component in the joint densities. It remains to linearize the last term of (A.24). Recall that

$$\dot{p}_t = \left. \frac{\partial \log p(Y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t = \tilde{\theta}_t} \quad \text{and} \quad \ddot{p}_t = \left. \frac{\partial^2 \log p(Y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t = \tilde{\theta}_t},$$

such that a first-order expansion about $\tilde{\theta}_t$ gives approximately

$$\partial \log p(y_t|\theta_t)/\partial \theta_t = \dot{p}_t + \ddot{p}_t (\theta_t - \tilde{\theta}_t). \qquad \text{(A.25)}$$

31

Substituting (A.25) in the last term of (A.24) gives the linearized form $Z'(\dot{p}_t + \ddot{p}_t \theta_t - \ddot{p}_t \tilde{\theta}_t)$. To obtain a form which coincides with the last term in (A.22), choose

$$\tilde{y}_t = \tilde{\theta}_t - \ddot{p}_t^{-1} \dot{p}_t \quad \text{and} \quad \tilde{H}_t = -\ddot{p}_t^{-1}. \tag{A.26}$$

These are the required updating equations. All elements in $y = (y_1', \ldots, y_T')'$ are independent after conditioning on the corresponding signal $\theta = (\theta_1', \ldots, \theta_T')'$. This implies that $\tilde{H}_t$ is diagonal for all $t = 1, \ldots, T$. As a result, each observation can be updated individually. ■

A3. GAS equations for mixed measurement credit risk model

We discuss the formulation of the MM-GAS model for the empirical application considered in Section 5. We consider the case of mixed measurements $y_t = (d_t', r_t', x_t')'$, where d_t is binomial, r_t is logit-normal, and x_t is Gaussian with time-varying parameters. The observations are dependent in the cross section since parameters depend on common factors.

$$d_{j,t}|f_t \sim \text{Binomial}\left(k_{j,t}, [1 + e^{-\theta_{j,t}}]^{-1}\right), \quad r_{j,t}|f_t \sim \text{Logit-normal}\left(\bar{\mu}_{j,t}, \bar{\sigma}_j^2\right), \quad x_{j,t}|f_t \sim \text{Normal}\left(\mu_{j,t}, \sigma_j^2\right),$$

where j indexes the cross section. The time-varying parameters depend on common factors as

$$\theta_{j,t} = c_{\theta,j} + Z_{d,j} f_t, \quad \bar{\mu}_{j,t} = c_{\bar{\mu},j} + Z_{r,j} f_t, \quad \mu_{j,t} = c_{\mu,j} + Z_{x,j} f_t.$$

The log-density for the observed variables y_t combines the multivariate normal, the binomial, and the logit-normal density. If all data is observed at time t, the local log-density is given by

$$\begin{aligned}
\log p(y_t|f_t, \psi) = \ & \text{const} + \sum_{j=1}^{n_1} d_{j,t}\theta_{j,t} - k_{j,t}\log[1 + \exp(\theta_{j,t})] \\
& -0.5 \sum_{j=1}^{n_2} 2\log[r_{j,t}/(1 - r_{j,t})] + \log \bar{\sigma}_j^2 + \bar{\sigma}_j^{-2}(\log[r_{j,t}/(1 - r_{j,t})] - \bar{\mu}_{j,t})^2 \\
& -0.5 \sum_{j=1}^{n_3} \log \sigma_j^2 + \sigma_j^{-2}(x_{j,t} - \mu_{j,t})^2,
\end{aligned}$$

where n_1, n_2, and n_3 are the dimensions of d_t, r_t, and x_t. As the log-density, the score and information matrix for the factors f_t also depend on which data is observed at time t.

$$\nabla_t = \sum_{j=1}^{n_1} \left(d_{j,t} - k_{j,t}[1 + \exp(-\theta_{j,t})]^{-1} \right) Z'_{d,j}$$

$$+ \sum_{j=1}^{n_2} \bar{\sigma}_j^{-2} \left(\log[r_{j,t}/(1 - r_{j,t})] - \bar{\mu}_{j,t} \right) Z'_{r,j}$$

$$+ \sum_{j=1}^{n_3} \sigma_j^{-2} \left(x_{j,t} - \mu_{j,t} \right) Z'_{x,j}.$$

The (inverse of) the scaling matrix is given as

$$S_{t-1}^{-1} = \mathrm{E}_{t-1}\left[\nabla_t \nabla'_t \right]$$

$$= Z'_d \Sigma_{d,t} Z_d + Z'_r \bar{\Sigma}^{-1} Z_r + Z'_x \Sigma^{-1} Z_x,$$

where $\Sigma_{d,t} = \mathrm{diag}\left(\pi_{1,t}(1 - \pi_{1,t})k_{1,t}, \ldots, \pi_{n_1,t}(1 - \pi_{n_1,t})k_{n_1,t} \right)$, $\bar{\Sigma} = \mathrm{diag}(\bar{\sigma}_1^2, \ldots, \bar{\sigma}_{n_2}^2)$, and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_{n_3}^2)$. In case data is missing at time t, the respective contributions to the sums are zero.

References

Altman, E. I., B. Brady, A. Resti, and A. Sironi (2003). The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business 78*, 2203–2227.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimension. *Econometrica 71(1)*, 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometica 70(1)*, 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics 25(1)*, 52–60(9).

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

Bruche, M. and C. Gonzalez-Aguado (2009). Recovery Rates, Default Probabilities and the Credit Cycle. *Journal of Banking and Finance, forthcoming*.

Connor, G. and R. A. Korajczyk (1986). Performance Measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15(3)*, 373–394.

Connor, G. and R. A. Korajczyk (1988). Risk and Return in equillibrium APT: Application of a new test methodology. *Journal of Financial Economics 21(2)*, 255–289.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *Journal of Finance 48(4)*, 1263–91.

Creal, D., S. J. Koopman, and A. Lucas (2008). A General Framework for Observation Driven Time-Varying Parameter Models. *Tinbergen Institute Discussion Paper Series 108/4*.

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).

Doz, C., D. Giannone, and L. Reichlin (2006). A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models. *ECB working paper 674*.

Düllmann, K. and M. Trapp (2004). Systematic Risk in Recovery Rates - An Empirical Analysis of U.S. Corporate Credit Exposures. *Deutsche Bundesbank working paper*.

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 (2000). Time Series Analysis of Non-Gaussian Observations based on State Space Models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society 62(1)*, 3–56.

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.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics 82(4)*, 540–554.

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.

Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series. *in Aigner, D and Goldberger, A, eds.: Latent variables in socio-economic models, North-Holland*.

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

Gupton, G. and R. Stein (2005). LossCalc V2: Dynamic Prediction of LGD. *Moodys KMV Investors Services*.

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

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika 94*, 827–839.

Jungbacker, B. and S. J. Koopman (2008). Likelihood-based Analysis of Dynamic Factor Models. *Tinbergen Institute Discussion Paper*.

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 (2008). A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk. *Journal of Business and Economic Statistics, forthcoming*.

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., A. Lucas, and B. Schwaab (2008). Forecasting Cross-sections of Frailty-correlated Default. *Tinbergen Institute Discussion Paper Series 029/04*.

Koopman, S. J., A. Lucas, and B. Schwaab (2009). Macro, Industry and Frailty-Effects in Defaults: The 2008 credit crisis in perspective. *working paper, available at www.tinbergen.nl/~bschwaab*.

Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics 149(1)*, 2–10.

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.

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.

Rösch, D. and H. Scheule (2005). A Multifactor Approach for Systematic Default and Recovery Risk. *The Journal of Fixed Income 15(2)*, 63–75.

Sargent, T. and C. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. *Federal Reserve Bank of Minneapolis, working paper No 55*.

Schuerman, T. (2006). What do we know about Loss-given-default? In D. Shimko (Ed.), *Credit Risk Models and Management*. Risk Books, London, UK.

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory 11*, 61–70.

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

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.

Stock, J. and M. Watson (2005). Implications of dynamic factor models for VAR analysis. *NBER working paper*.