DYNAMIC FACTOR MODELS WITH STOCHASTIC VOLATILITY

WILLIAM J. MCCAUSLAND

Abstract. I introduce posterior simulation methods for a dynamic latent factor model featuring both mean and variance factors. The cross-sectional dimension may be large, so the methods are applicable to data-rich environments. I apply the new methods in two empirical applications. The first involves a panel of 134 real activity and financial indicators observed monthly from 1959 to 2015; the second, a panel of 10 currencies, with daily log returns observed over a decade.

1. Introduction

Since their introduction by Geweke (1977), dynamic factor models have become important tools for the analysis of macroeconomic and financial data. A great deal of the data variation in large macroeconomic panels can be accounted for by a small number of factors. The data reduction that factor structure allows is a huge computational advantage for forecasting. Dynamic factor models are commonly used in data-rich environments, where the number of observed series (which we denote N) is in the order of one hundred or more. The number of observation periods (T) is usually not large, as data are collected at monthly or quarterly frequency, and the availability of data prior to 1960 is very limited. See surveys on dynamic factor models by Bai and Ng (2008) and Stock and Watson (2011).

Dynamic factor models have also been applied in multivariate stochastic volatility models for asset returns. Here, N is usually much smaller and T much larger. See for example, Aguilar and West (2000), Chib, Omori, and Asai (2009) and Nakajima and West (2013).

Many dynamic factor models used in macroeconomics feature constant variances. However, measures of real activity, such as production and employment, are known to vary more in some periods than in others: the Great Moderation, for example, was a period of relative tranquility for many such measures. Macroeconomic panels often feature asset

Date: July 22, 2016.
Département de sciences économiques, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada.
e-mail: william.j.mccausland@umontreal.ca, file: DFM-SV.tex. Preliminary and incomplete.

prices, exchange rates and commodity prices; these financial indicators are also relevant for macroeconomic forecasting and exhibit even greater fluctuations in volatility.

Bayesian methods for inference in dynamic factor models have many advantages. Computing posterior moments of parameters and unobserved factors is an exercise in numerical integration, and as such it is often more stable than maximizing an objective function such as a likelihood. Posterior simulation only requires a joint density function for all observable and unobservable quantities, including data, latent variables and parameter values; computation of a likelihood function is unnecessary. This facilitates analysis in non-linear and non-Gaussian models. Bayesian methods allow for a formal probabilistic specification of a priori information, such as theoretical restrictions, or regularization priors. Bayesian methods naturally accommodate a flexible approach to parsimony: discipline can be achieved either with a small number of parameters, or by imposing tight prior distributions on a larger number of parameters.

The purpose of this paper is to develop posterior simulation methods for a dynamic factor model with stochastic volatility, suitable for data-rich macroeconomic applications. In this model, both the mean and the variance of the observed series have a factor structure. Since the focus of this paper is the variance factor model, we use a relatively simple mean factor model. Mean factors follow a Gaussian first order vector autoregressive process and the factor loadings are all contemporaneous. Factor variances and factor loadings do not vary in time, which implies that conditional correlations between the common parts of the observed series do not vary in time either. With the exception of constant factor loadings, all of these restrictive assumptions could be relaxed without too much trouble. However, it would be more difficult to introduce time-varying factor loadings in data-rich environments due to the explosion in the number of latent variables.

The variance factor model captures fluctuating variance of the idiosyncratic term of the various observed series. In previous studies, such as Del Negro and Otrok (2008), the idiosyncratic variance of each series evolves independently according to a univariate stochastic volatility model. Posterior simulation for such a model would be a considerable computational burden in a data-rich environment. At the same time, we would expect some common variation in the various idiosyncratic variances. I use a factor model to capture this common variation using a small number of factors.

In order to work well with large panels, simulation methods need to be computationally efficient. The cost of an MCMC draw should not rise very quickly in N. Gibbs sampling effectively decouples the computation associated with the mean and variance factor models. The computational cost of updating the mean factors and their parameters in linear in N. The same is true of the variance factors and their parameters.

Once we condition on variance factors and their parameters, we are left with a linear Gaussian state space model, and we can decompose the joint Gaussian distribution of mean factors and observed series into the conditional distribution of factors given series and the marginal distribution of the observed series, both Gaussian. The first gives us the conditional posterior distribution of factors, We can use the second to evaluate the conditional likelihood with factors integrated out; this allows us to update the conditional posterior of the factors' autoregressive coefficients marginally of factors.

Usually this decomposition is done by applying the Kalman filter, together with one of the simulation smoothing procedures proposed by Carter and Kohn (1994), Frühwirth-Schnatter (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). See, for example, Kim and Nelson (2000). However, the Kalman filter is computationally costly for data rich environments, due to the need to solve a system of N equations for each of T observations. An alternative is to compute the block band precision (inverse of variance) matrix of the conditional distribution of the factors. From there, it is easy to compute its Cholesky factor, also a block band matrix, draw factors from their conditional posterior distribution using forward- and back-substitution, and evaluate the conditional posterior density of the autoregressive coefficients with factors integrated out. All of this is accomplished using standard operations for matrices with block band structure. Methods involving block band matrix (or more general sparse matrix) operations have been discussed and applied in Chan and Jeliazkov (2009), McCausland, Miller, and Pelletier (2011) and Rue (2001). McCausland, Miller, and Pelletier (2011) gives a detailed comparison of computational costs for the Kalman filter and for block band matrix operations.

The block band matrix operations allow for computationally efficient joint draws of mean factors and their autoregressive coefficient matrices from their conditional posterior distribution. This computational efficiency does not rely on, or even benefit from, the conditional conjugacy of Gaussian priors for the autoregressive coefficient matrices. Thus free to

choose non-conjugate priors, I normalize the distribution of means factors by setting the unconditional—rather than the conditional—variance of the factors to a constant. The advantages here are reducing the posterior dependence between the autoregressive coefficient matrix and the mean factor loading matrix.

Given the mean factors and their parameters, the variance factor model is multivariate state space model with Gaussian latent states and non-linear non-Gaussian observations. We apply the HESSIAN method described in McCausland (2012) to draw the variance factors, one entire univariate factor series at a time. This method is based on a very close approximation of the conditional posterior distribution of the latent factor in univariate state space models with Gaussian latent states and non-linear non-Gaussian observations. I use such an approximation as a proposal distribution for each variance factors series. The method can be applied one factor series at a time: once one conditions on all other factor series, the k'th factor series is the unknown state of a univariate state-space model with Gaussian latent states and non-linear non-Gaussian observations. This is true whether or not the variance factor series are *a priori* mutually independent. I do not impose such independence; while the rotation I use to identify factors gives a time-t marginal distribution with cross-sectionally independent factors, there remains dynamic dependence across factors.

Section 2 describes the model for observables, discusses my identification strategy and specifies a prior distribution, up to the selection of its various hyper-parameters. Sections 3 and 4 describe Bayesian computation for the mean and variance factor models, respectively. Section 5 reports the results of an artificial data exercise and two empirical applications. The "Getting it Right" exercise, based on Geweke (2004), verifies the correctness of the algorithms used and their implementation in code. The first empirical application involves a panel of 134 real activity and financial indicators observed monthly from 1959 to 2015; the second, a panel of 10 currencies, with daily log returns observed over a decade. Section 6 concludes.

2. A factor model with mean and variance factors

2.1. **The model.** We observe y_t, an $N \times 1$ vector of dependent variables and x_t, a $J \times 1$ vector of independent variables, at times $t = 1, \ldots, T$. The model for the dynamics of y_t features two latent dynamic factors, a mean factor $F_{\mu t}$ of length K_μ and a variance factor

$F_{\sigma t}$ of length K_σ. Given the latent factors, y_t evolves according to

$$y_t = \mu_t + [\mathrm{diag}(\sigma_t^2)]^{1/2}\epsilon_t, \tag{1}$$

where the conditional mean vector μ_t has the factor structure

$$\mu_t = \Lambda_\mu F_{\mu t} + B_\mu x_t, \tag{2}$$

and the conditional variance vector σ_t^2 has the exponential factor structure

$$\sigma_t^2 = \exp(\Lambda_\sigma F_{\sigma t} + B_\sigma x_t). \tag{3}$$

Here, Λ_μ and Λ_σ are factor loading matrices, $\epsilon_t \sim \mathrm{iid}\,\mathrm{N}(0, I_N)$, and the exponential function is applied elementwise. The latent factor series $F_{\mu t}$ and $F_{\sigma t}$ are first order Gaussian vector autoregressions:

$$F_{\mu t} = \Phi_\mu F_{\mu,t-1} + u_{\mu t}, \tag{4}$$

$$F_{\sigma t} = \Phi_\sigma F_{\sigma,t-1} + u_{\sigma t}, \tag{5}$$

where $u_{\mu t} \sim \mathrm{iid}\,\mathrm{N}(0, \Sigma_\mu)$ and $u_{\sigma t} \sim \mathrm{iid}\,\mathrm{N}(0, \Sigma_\sigma)$. The ϵ_t, $u_{\mu t}$ and $u_{\sigma t}$ are mutually independent. The same vector x_t of exogenous observed variables appears in both the mean factor (2) and variance factor (3) equations; one can always impose exclusion restrictions. In data-rich environments, N is much larger than K_μ and K_σ. To simplify notation, let $F_\mu \equiv (F_{\mu,1}^\top, \ldots, F_{\mu,T}^\top)^\top$, $F_\sigma \equiv (F_{\sigma,1}^\top, \ldots, F_{\sigma,T}^\top)^\top$, $y \equiv (y_1^\top, \ldots, y_T^\top)^\top$, $\theta_\mu \equiv (B_\mu, \Lambda_\mu, \Phi_\mu)$, $\theta_\sigma = (B_\sigma, \Lambda_\sigma, \Phi_\sigma)$ and $\theta = (\theta_\mu, \theta_\sigma)$.

2.2. **Identification.** The lack of identification in factor models is well known; there are rotation, permutation, scale and sign transformations of factors and their parameters that leave the distribution of y unchanged. Here I describe my identification strategy.

Two popular normalizations of Σ_μ are $\Sigma_\mu = I$ and $\Sigma_\mu = I - \Phi_\mu\Phi_\mu^\top$, which set the conditional and unconditional variances of $F_{\mu t}$, respectively, to I. The corresponding restrictions on Φ_μ require that the eigenvalues of Φ_μ and $\Phi_\mu\Phi_\mu^\top$, respectively, lie in the unit circle.

The $\Sigma_\mu = I - \Phi_\mu\Phi_\mu^\top$ normalization is commonly used in principal components analyses. The $\Sigma_\mu = I$ normalization is more often used in Bayesian analyses, largely because a

Gaussian prior distribution for Φ_μ is nearly conjugate for the conditional distribution of F_μ given Φ_μ, and exactly conjugate if one conditions on a pre-sample $F_{\mu,0}$.

Although my analysis is Bayesian, I adopt the $\Sigma_\mu = I - \Phi_\mu \Phi_\mu^\top$ normalization. It has several advantages over the $\Sigma_\mu = I$ normalization, and the advantages of the latter may be overestimated.

Among the advantages are: (1) It is easy to compare results of a Bayesian analysis with the static factors and loadings given by a principal components analysis. (2) It is easy to decompose the variance of each y_{ti} into common and idiosyncratic terms: $\mathrm{Var}[y_{ti}|\theta, x_t] = \sum_{k=1}^{K_\mu} \Lambda_{ik}^2 + E[\sigma_{ti}^2|\theta, x_t]$; the first term is a particularly transparent expression that does not involve Φ_μ since the unconditional variance of $F_{\mu t}$ does not depend on Φ_μ. (3) The autoregression coefficient Φ_μ is more easily interpreted. It is an autocorrelation matrix, and thus invariant to units of measurement. Symmetry of Φ_μ is necessary and sufficient for the time reversibility of $F_{\mu t}$; under the normalization $\Sigma_\mu = I$, the time reversibility condition is much less transparent. (4) Under the $\Sigma_\mu = I$ normalization, the unconditional variance of $F_{\mu t}$ is highly sensitive to Φ_μ near the boundary of Φ_μ, while under the $\Sigma_\mu = I - \Phi_\mu \Phi_\mu^\top$ normalization, it is constant. Thus, there is less posterior dependence between Φ_μ and the factor loadings Λ_μ, through the varying scale of factors and their loadings, and this makes posterior simulation using Gibbs sampling more numerically efficient. (5) A principal components analysis gives good initial values of factors and loadings, which shortens the transient burn-in period the MCMC chain spends in regions of low posterior probability.

The principal advantage of the $\Sigma_\mu = I$ normalization is the approximate conditional conjugacy of a Gaussian prior for Φ_μ: such a prior makes the conditional posterior distribution of Φ_μ nearly Gaussian, so that a single draw from a Gaussian proposal is almost as good as an exact conditional draw. Under the $\Sigma_\mu = I - \Phi_\mu \Phi_\mu^\top$ normalization, there is no convenient conditionally conjugate prior distribution for Φ_μ, Gaussian or otherwise. However, the conditional density of F_μ given Φ_μ depends on F_μ through a low dimensional "sufficient statistic", and once this statistic is computed, the conditional density can be evaluated very cheaply for each new value of Φ_μ. In this way, a large number of repetitions of a simple generic MCMC update, such as random-walk Metropolis or slice sampling, can achieve almost the same numerical efficiency as an exact conditional draw, at a cost not much larger than a single such update. See Appendix C for details.

Under the $\Sigma_\mu = I$ normalization, the advantages of conjugacy are most completely captured when the prior for Φ_μ is Gaussian, while under the $\Sigma_\mu = I - \Phi_\mu \Phi_\mu^\top$ normalization, there is no particular advantage in using a Gaussian prior. Thus, any advantages of a non-Gaussian prior come at no additional cost. Section 2.3 below describes a prior that is more flexible than a Gaussian one, especially for highly persistent factors, and incorporates many, though not all, of the necessary conditions on Φ_μ for stationarity, before truncation.

I use exactly the same normalization for Σ_σ and Φ_σ and realize most of the same advantages.

I restrict Λ_μ and Λ_σ to be row-permutations of lower triangular matrices with positive diagonal elements. This normalization is slightly more general than the usual one, in which factor loading matrices are lower triangular with positive diagonal elements. It affords more flexibility on the choice of the factor founder[1] series, those observed series with exclusion restrictions in the form of loadings set to zero. Since I can (and do) choose different permutations of the observed series to identify Λ_μ and Λ_σ, this is substantive. Section A.2 describes how I determine the factor founders. Briefly, I choose them so that the factors resemble the static factors obtained using principal components.

I will call the j'th mean factor founder the observed series i for which the exclusion restriction $\Lambda_{\mu,ik} = 0$, $k < j$, and the sign restriction $\Lambda_{\mu,ij} > 0$ hold; I define the j'th variance factor founder analogously.

2.3. **Prior distributions.** Here I describe prior distributions for the parameters of the model, up to the values of various hyper-parameters. I assign numerical values of the hyper-parameters in Section 5. I will assume that the first exogenous variable is a constant: $x_{t1} = 1$, $t = 1, \ldots, T$. Thus $B_{\mu,i1}$ and $B_{\sigma,i1}$ are the constant coefficients of the i'th mean and variance equations, respectively.

The parameters have the conditional independence properties implied by the following density decomposition:

$$f(\theta) = f(\Phi_\mu) f(\Phi_\sigma) f(B_\mu) f(B_\sigma) f(\Lambda_\sigma) f(\Lambda_\mu | B_\sigma).$$

I choose a prior for Φ_μ (Φ_σ) that makes the K_μ (K_σ) latent factor series exchangeable. At the same time, I want to treat diagonal elements, which are autocorrelations, and

[1] I believe the term "factor founder" is due to Carvalho, Chang, Lucas, Wang, and West (2008).

off-diagonal elements, which are cross-correlations, differently, as well as ensure that the eigenvalues of $\Phi_\mu \Phi_\mu^\top$ ($\Phi_\sigma \Phi_\sigma^\top$) are in the unit circle.

I choose a truncated prior distribution for Φ_μ (Φ_σ). Truncation is to the stationary region, where $\Phi_\mu \Phi_\mu^\top$ ($\Phi_\sigma \Phi_\sigma^\top$) has eigenvalues in the unit circle. In the pre-truncation distribution, the elements of Φ_μ (Φ_σ) are independent, diagonal elements are identically distributed and off-diagonal elements are identically distributed. The pre-truncation distributions of the elements of Φ_μ and Φ_σ are

$$(\Phi_{\mu,kl}+1)/2 \sim \begin{cases} \mathrm{Be}(\phi_1,\phi_2) & k=l, \\ \mathrm{Be}(\phi_3,\phi_4) & k \neq l. \end{cases} \qquad (\Phi_{\sigma,kl}+1)/2 \sim \begin{cases} \mathrm{Be}(\phi_5,\phi_6) & k=l, \\ \mathrm{Be}(\phi_7,\phi_8) & k \neq l. \end{cases}$$

Here, Be denotes the Beta distribution. This distribution allows much more flexibility than the Gaussian distribution for the behaviour of the prior of diagonal elements near one. According to the value of ϕ_2, for example, the density of $\Phi_{\mu,kk}$ at one can be zero ($\phi_2 > 1$), finite ($\phi_2 \geq 1$), or infinite $\phi_2 < 1$; setting $\phi_2 > k$ makes the $k-1$'th derivative of the density at one equal to zero. Another advantage is that the restricted support of the Beta distribution incorporates necessary conditions for stationarity, reducing the impact of truncation.

The parameters B_μ and B_σ have rows that each relate to a specific observed series. These series may be very different types of variables, with different scales and units of measurement, and their priors should reflect this. Suppose, for example, that series i is an asset return series, which might be measured either as a difference of log prices or as a percentage. The value of $B_{\mu,i}$ associated with the latter will be 100 times larger than the value of $B_{\mu,i}$ associated with the former; the constant coefficient $B_{\sigma,i1}$ should be larger by $2\ln 100$. I choose priors for B_μ and B_σ where their rows are independent and have Gaussian distributions that are not necessarily identical. For the purpose of the computational sections of this paper, we will specify the prior distributions of $B_{\mu,i}$ and B_σ in terms of the precision $\bar{H}_{\mu,i}$ and covector $\bar{c}_{\mu,i}$, rather than the mean and variance. Thus,

$$B_{\mu,i} \sim N(\bar{H}_{B_{\mu,i}}^{-1}\bar{c}_{B_{\mu,i}}, \bar{H}_{B_{\mu,i}}^{-1}), \qquad B_{\sigma,i} \sim N(\bar{H}_{B_{\sigma,i}}^{-1}\bar{c}_{B_{\sigma,i}}, \bar{H}_{B_{\sigma,i}}^{-1}).$$

In the Getting it Right and Bank of Canada currency panel applications, each $B_{\mu,i}$ has mean ϕ_9 and standard deviation ϕ_{10}, so that $\bar{H}_{\mu,i} = (1/\phi_{10}^2)I$ and $\bar{c}_{B_{\mu,i}} = (\phi_9/\phi_{10}^2)\iota$, where ι is a J-vector of ones. In the FRED-MD panel application,

The elements of Λ_σ are dimensionless, and so despite the fact that each pertains to a specific series, I choose a prior that does not distinguish among series, except to accommodate the exclusion and sign restrictions serving to identify Λ_σ. Λ_σ has the distribution of a $N \times K_\sigma$ matrix whose elements are iid $\mathrm{N}(0, \phi_{14}^2)$, right-multiplied by the orthogonal matrix that makes Λ_σ satisfy the identification restrictions.[2] Here, ϕ_{14} is a prior scale hyperparameter. Thus, elements of Λ_σ are independent, with

$$\frac{\Lambda_{\sigma,ij}}{\phi_{14}} \sim \begin{cases} 0 & i \text{ is a variance factor founder before the } j\text{'th}, \\ \chi(K_\sigma - j + 1) & i \text{ is the } j\text{'th variance factor founder}, \\ \mathrm{N}(0,1) & \text{otherwise}, \end{cases}$$

where $\chi(\nu)$ denotes the chi (sic; not chi-squared) distribution with ν degrees of freedom. Note that for each row i, with or without exclusion restrictions, $\phi_{14}^{-2} \sum_{k=1}^{K_\sigma} \Lambda_{\sigma,ik}^2 \sim \chi^2(K_\sigma)$, where $\chi^2(\nu)$ denotes the chi-*squared* distribution with ν degrees of freedom.

The elements of Λ_μ, unlike those of Λ_σ, have the same units of measurement as the observed series they relate to. I therefore scale the prior for each row $\Lambda_{\mu,i}$ by $\phi_{13}\exp(B_{\sigma,i1}/2)$, where $B_{\sigma,i1}$ is the coefficient of the constant term of the variance equation for indicator i and ϕ_{13} is a prior scale hyper-parameter. Otherwise, the prior for Λ_μ is similar to that of Λ_σ. Elements of Λ_μ are conditionally independent given B_σ, with

$$\frac{\exp(-B_{\sigma,i1}/2)}{\phi_{13}}\Lambda_{\mu,ij}|B_\sigma \sim \begin{cases} 0 & i \text{ is a mean factor founder before the } j\text{'th}, \\ \chi(K_\mu - j + 1) & i \text{ is the } j\text{'th mean factor founder}, \\ \mathrm{N}(0,1) & \text{otherwise}. \end{cases}$$

For each row i, with or without exclusion restrictions, the conditional distribution of $\phi_{13}^{-2}\exp(-B_{\sigma,i1})\sum_{k=1}^{K_\mu}\Lambda_{\mu,ik}^2$ given B_σ is $\chi^2(K_\mu)$.

[2] A draw of Λ_σ from its prior is obtained by permuting the rows of $R^\top$, where QR is the QR decompostion of a matrix of iid Gaussian variates.

FIGURE 1. Overview of posterior simulation

(1) Set $(\Lambda_\mu^{(-B-1)}, B_\mu^{(-B-1)}, \Phi_\mu^{(-B-1)}, F_\mu^{(-B-1)}, \Lambda_\sigma^{(-B-1)}, B_\sigma^{(-B-1)}, \Phi_\sigma^{(-B-1)}, F_\sigma^{(-B-1)})$ (Section A.3)
(2) For $m = 1, \ldots, M$, draw updates preserving the conditional posterior distributions
  (a) $\Lambda_\mu | B_\mu, \Phi_\mu, F_\mu, \theta_\sigma, F_\sigma, x, y$ (Section 3.1)
  (b) $B_\mu | \Lambda_\mu, \Phi_\mu, F_\mu, \theta_\sigma, F_\sigma, x, y$ (3.2)
  (c) $\Phi_\mu, F_\mu | \Lambda_\mu, B_\mu, \theta_\sigma, F_\sigma, x, y$ (3.3)
    (i) $\Phi_\mu | \Lambda_\mu, B_\mu, \theta_\sigma, F_\sigma, x, y$
    (ii) $F_\mu | \Lambda_\mu, B_\mu, \Phi_\mu, \theta_\sigma, F_\sigma, x, y$
  (d) $\Phi_\mu | \Lambda_\mu, B_\mu, F_\mu, \theta_\sigma, F_\sigma, x, y$ (3.4)
  (e) $\Lambda_\sigma | \theta_\mu, F_\mu, B_\sigma, \Phi_\sigma, F_\sigma, x, y$ (4.1)
  (f) $B_\sigma | \theta_\mu, F_\mu, \Lambda_\sigma, \Phi_\sigma, F_\sigma, x, y$ (4.2)
  (g) $\Phi_\sigma, F_\sigma | \theta_\mu, F_\mu, \Lambda_\sigma, B_\sigma, x, y$ (4.3)
    (i) $\Phi_\sigma | \theta_\mu, F_\mu, \Lambda_\sigma, B_\sigma, x, y$
    (ii) $F_\sigma | \theta_\mu, F_\mu, \Lambda_\sigma, B_\sigma, \Phi_\sigma, x, y$
  (h) $\Phi_\sigma | \theta_\mu, F_\mu, \Lambda_\sigma, B_\sigma, F_\sigma, x, y$ (3.4)

3. Bayesian computation, mean factor model

In this section and the next, I describe Markov chain Monte Carlo methods for posterior simulation of parameters and factors. A single step of the Markov chain consists of eight draws, updating the Gibbs blocks B_μ, Λ_μ, (Φ_μ, F_μ), Φ_μ, B_σ, Λ_σ, (Φ_σ, F_σ) and Φ_σ. Each update preserves the posterior distribution $B_\mu, \Lambda_\mu, \Phi_\mu, F_\mu, B_\sigma, \Lambda_\sigma, \Phi_\sigma, F_\sigma | y$, although the blocks (Φ_μ, F_μ) and (Φ_σ, F_σ) are joint blocks consisting of marginal and conditional parts. Figure 1 gives an overview and directs the reader to the appropriate sections describing the updates. For notational simplicity, I assume that y_{ti} is observed for all t and i; if any data are missing, this can be easily accommodated, as described in Appendix A.1. Appendix A.3 describes how the initial value of the MCMC chain is determined.

Gibbs sampling, a divide-and-conquer approach which breaks a simulation problem into simpler problems, neatly decouples simulation for the mean and variance factor models. Given variance factors and their parameters, the model reduces to a mean factor model with known but time varying variances; given mean factors and their parameters, it reduces to a variance factor model for the residual term of the mean factor model. An important implication of this modularity is that the variance factor model, the central contribution of this paper, can be paired with alternate mean factor models with no modification of the code for the variance factor model.

I discuss Bayesian computation in two separate sections. Here in Section 3, I describe Gibbs blocks for the unknown quantities B_μ, Λ_μ, Φ_μ and F_μ of the mean factor model of equations (2) and (4). In Section 4, I do the same for the unknown quantities B_σ, Λ_σ, Φ_σ and F_σ of the variance factor model, equations (3) and (5).

To focus attention on the mean factor model, I condition on σ_t, $t = 1, \ldots, T$, and suppress notation for this conditioning, for the rest of the section. The remainder of this section is divided into four parts, where I discuss the Gibbs blocks B_μ, Λ_μ, (Φ_μ, F_μ) and Φ_μ.

3.1. **Drawing Λ_μ.** In the conditional posterior distribution of Λ_μ, the rows $\Lambda_{\mu,i}$ are conditionally independent. If i is not a mean factor founder, the prior for $\Lambda_{\mu,i}$ is Gaussian and therefore conditionally conjugate. In this case, the conditional posterior distribution of $\Lambda_{\mu,i}$ is $N(\bar{\bar{H}}_{\Lambda_{\mu,i}}^{-1}\bar{\bar{c}}_{\Lambda_{\mu,i}}, \bar{\bar{H}}_{\Lambda_{\mu,i}}^{-1})$, where the posterior precision $\bar{\bar{H}}_{\Lambda_{\mu,i}}$ and covector $\bar{\bar{c}}_{\Lambda_{\mu,i}}$ are given by

$$\bar{\bar{H}}_{\Lambda_{\mu,i}} = \bar{H}_{\Lambda_{\mu,i}} + \sum_{t=1}^{T} e^{-\sigma_{ti}^2} F_{\mu,t} F_{\mu,t}^\top, \tag{6}$$

$$\bar{\bar{c}}_{\Lambda_{\mu,i}} = \bar{c}_{\Lambda_{\mu,i}} + \sum_{t=1}^{T} e^{-\sigma_{ti}^2} (y_{ti} - B_{\mu,i} x_t) F_{\mu,t}. \tag{7}$$

Here $\bar{H}_{\Lambda_{\mu,i}} = \phi_1^{-2} \exp(-B_{\sigma,i1}) I$ and $\bar{c}_{\Lambda_{\mu,i}} = 0$ are the prior precision and covector, with I being the $K_\mu \times K_\mu$ identity matrix and 0, a K_μ-vector of zeros.

If i is a mean factor founder, say the j'th, the situation is a little more complicated. The exclusion restriction $\Lambda_{\mu,ik} = 0$, $k > j$, applies and the conditional posterior distribution of the non-zero elements of $\Lambda_{\mu,i}$ is not Gaussian, due to the non-conjugate χ prior distribution of $\Lambda_{\mu,ij}$. I draw a Metropolis-Hastings proposal $\Lambda_{\mu,i}^*$ with $\Lambda_{\mu,ik}^* = 0$ for $k > j$ and $(\Lambda_{\mu,i1}^*, \ldots, \Lambda_{\mu,ij}^*) \sim N(\bar{\bar{H}}_{\Lambda_{\mu,i}}^{-1}\bar{\bar{c}}_{\Lambda_{\mu,i}}, \bar{\bar{H}}_{\Lambda_{\mu,i}}^{-1})$, where the expressions for $\bar{\bar{H}}_{\Lambda_{\mu,i}}$ and $\bar{\bar{c}}_{\Lambda_{\mu,i}}$ are the same as above, except that the right hand side of (6) is replaced with its $j \times j$ leading submatrix and the right hand side of (7) is replaced with its $j \times 1$ leading subvector. The proposal is accepted with probability $\min[1, R(\Lambda_{\mu,i}, \Lambda_{\mu,i}^*)]$, where the Hastings ratio $R(\Lambda_{\mu,i}, \Lambda_{\mu,i}^*)$ is given by

$$R(\Lambda_{\mu,i}, \Lambda_{\mu,i}^*) = \left(\frac{|\Lambda_{\mu,ij}^*|}{\Lambda_{\mu,ij}} \right)^{K_\mu - j}.$$

Then if $\Lambda_{\mu,i}^*$ is accepted and $\Lambda_{ij}^* < 0$, multiply the following quantities by minus one: the j'th column of Λ; F_{tj}, $t = 1, \ldots, T$; Φ_{jk} and Φ_{kj}, $k \neq j$. This maps a value of (Λ, F, Φ)

outside the sign-restricted region to an observationally equivalent value inside. With enough data and well chosen factor founders, this mapping will be rarely be needed.

3.2. **Drawing B_μ.** Drawing B_μ is very straightforward, as its Gaussian prior distribution is conditionally conjugate. Its conditional posterior distribution is Gaussian, with independent rows $B_{\mu,i} \sim N(\bar{\bar{H}}_{B_{\mu,i}}^{-1}\bar{\bar{c}}_{B_{\mu,i}}, \bar{\bar{H}}_{B_{\mu,i}}^{-1})$, where the posterior precision $\bar{\bar{H}}_{B_{\mu,i}}$ and covector $\bar{\bar{c}}_{B_{\mu,i}}$ are given by

$$\bar{\bar{H}}_{B_{\mu,i}} = \bar{H}_{B_{\mu,i}} + \sum_{t=1}^{T} e^{-\sigma_{ti}^2} x_t x_t^\top, \tag{8}$$

$$\bar{\bar{c}}_{B_{\mu,i}} = \bar{c}_{B_{\mu,i}} + \sum_{t=1}^{T} e^{-\sigma_{ti}^2}(y_{ti} - \Lambda_{\mu,i}F_{\mu,t})x_t. \tag{9}$$

Here $\bar{H}_{B_{\mu,i}} = \phi_2^{-2} I$ and $\bar{c}_{B_{\mu,i}} = (\phi_2/\phi_2^2)\iota$ are the prior precision and covector, and ι is a J-vector of ones.

3.3. **Drawing Φ_μ and F_μ.** I draw Φ_μ and F_μ jointly from their conditional posterior distribution, taking advantage of the fact that the conditional posterior distribution of F_μ is multivariate Gaussian, with a block band precision matrix. I first perform a sequence of random walk Metropolis updates of columns of Φ_μ, preserving the conditional posterior distribution $\Phi_\mu|\Lambda_\mu, B_\mu, y$—with F_μ marginalized out. I then draw F_μ from its conditional posterior distribution $F_\mu|\Phi_\mu, \Lambda_\mu, B_\mu, y$.

Computing the Hastings ratio for the random walk Metropolis updates requires evaluating the density $f(\Phi_\mu|\Lambda_\mu, B_\mu, y)$, with F_μ integrated out. Usually, evaluating $f(\Phi_\mu|\Lambda_\mu, B_\mu, y)$ and drawing F_μ from its conditional posterior distribution are done using the Kalman filter. Instead, I adopt the approach outlined in McCausland, Miller, and Pelletier (2011), which involves computing the block band Cholesky factor of the negative Hessian matrix of $\log f(F_\mu|\Phi_\mu, \Lambda_\mu, B_\mu, y)$. The computational advantages of this approach are outlined in McCausland, Miller, and Pelletier (2011), and these are especially important in the case where N is much greater than K_μ. The Kalman filter requires solving a symmetric system of N equations at each observation t. The sparse Hessian matrix approach—see also Chan and Jeliazkov (2009) and Rue (2001)—requires solving a symmetric system of K_μ equations at each observation.

I use a sequence of random walk Metropolis steps to update Φ_μ. I update each column $\Phi_{\mu,k}$ multiple times, using a proposal distribution $\Phi^*_{\mu,k} \sim \mathrm{N}(\Phi_{\mu,k}, \Omega_k)$. WJM: what to say about Ω_k? The proposal Φ^*_μ constructed by replacing row k of Φ_μ with $\Phi^*_{\mu,k}$ is accepted with probability

$$\min\left[1, \frac{f(\Phi^*_\mu) f(y|\Phi^*_\mu, \Lambda_\mu, B_\mu, x)}{f(\Phi_\mu) f(y|\Phi_\mu, \Lambda_\mu, B_\mu, x)}\right]. \tag{10}$$

For two reasons, it is quite cheap to perform several repetitions of the Φ^*_μ each time. First, while two evaluations of $f(y|\Phi_\mu, \Lambda_\mu, B_\mu, F_\mu, x)$—numerator and denominator—are required for the first iteration, each subsequent iteration involves only one evaluation. Second, for each evaluation after the first, one can reuse computations depending only on Λ_μ. These additional evaluations have a computational cost that does not depend on N. This means that a large number of repetitions, enough to achieve the same numerical efficiency as an exact conditional draw, can be performed at modest total cost.

Appendix B gives details on how to compute $f(y|\Phi_\mu, \Lambda_\mu, B_\mu, x)$ and draw $F_\mu|\Phi_\mu, \Lambda_\mu, B_\mu, y$.

3.4. **Drawing Φ_μ.** In this block, I draw Φ_μ from its conditional distribution given data and all unknown quantities, including F_μ. The identification $\mathrm{Var}[F_{\mu,t}|\Phi_\mu] = I$ means that the factor innovation variance is $I - \Phi_\mu \Phi_\mu^\top$, a function of Φ_μ. This means that the posterior distribution of Φ_μ does not closely approximate a Gaussian distribution, as it does when the factor innovation variance is constant. Fortunately, there is a low-dimensional sufficient statistic—a function of F_μ—for Φ_μ. This allows us to make a large number of extremely cheap draws using random walk Metropolis steps. The overall result is a update of Φ_μ that is nearly independent of the previous value, obtained at low total computational cost.

Appendix C gives details.

4. Bayesian computation, variance factor model

This section discusses Bayesian computation for the variance factor model. To focus attention on this part of the model, I condition on $\mu_t = \Lambda_\mu F_{\mu t} + B_\mu x_t$—and suppress notation for this conditioning—for the rest of the section. I define $\tilde{y}_t = y_t - \mu_t$ and $\tilde{y} = (\tilde{y}_1^\top, \ldots, \tilde{y}_T^\top)^\top$. Thus, $\tilde{y}_t$ follows a pure variance factor model.

We can write the conditional density of $\tilde{y}$ as

$$\log f(\tilde{y}|\Lambda_\sigma, B_\sigma, F_\sigma, x) = -\frac{1}{2}\left[NT\log 2\pi + \iota^\top \Lambda_\sigma \sum_{t=1}^{T} F_{\sigma,t} + \iota^\top B_\sigma \sum_{t=1}^{T} x_t \right.$$
$$\left. + \sum_{t=1}^{T}\sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)\right], \tag{11}$$

where $\Lambda_{\sigma,i}$ and $B_{\sigma,i}$ are the i'th rows of Λ and B.

4.1. **Drawing Λ_σ.** In the conditional posterior distribution of Λ_σ given F_σ and B_σ, rows $\Lambda_{\sigma,i}$ are conditionally independent. I update the $\Lambda_{\sigma,i}$ one at a time, using a Metropolis-Hastings transition. I first evaluate the gradient and an approximation to the Hessian of $\log f(\tilde{y}|\Lambda_\sigma, B_\sigma, F_\sigma, x)$, both with respect to $\Lambda_{\sigma,i}$, at the current value of $\Lambda_{\sigma,i}$. I then use these to construct a multivariate Gaussian proposal $\Lambda^*_{\sigma,i}$. I then evaluate the same gradient and approximate Hessian at $\Lambda^*_{\sigma,i}$, construct the appropriate Hastings ratio, and perform a Metropolis-Hastings accept/reject.

I now provide some details on the computations. As with the mean factor loadings, the situation is a little more complicated when i is a factor founder, and so I begin with the simpler case when i is not a factor founder.

I compute the gradient and an approximation to the Hessian as follows. First, define v_i as what remains of $\log f(\Lambda_{\sigma,i}) + \log f(\tilde{y}|\Lambda_\sigma, B_\sigma, F_\sigma, x)$ after eliminating additive terms that do not depend on $\Lambda_{\sigma,i}$:

$$v_i = -\frac{1}{2}\left[\phi_5^{-2}\Lambda_{\sigma,i}\Lambda_{\sigma,i}^\top + \Lambda_{\sigma,i}\sum_{t=1}^{T} F_{\sigma,t} + \sum_{t=1}^{T}\tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)\right]. \tag{12}$$

The gradient and Hessian of v_i with respect to $\Lambda_{\sigma,i}$ are given by g_i and Ψ_i:

$$g_i \equiv \frac{\partial v_i}{\partial \Lambda_{\sigma,i}^\top} = -\phi_5^{-2}\Lambda_{\sigma,i}^\top - \frac{1}{2}\sum_{t=1}^{T} F_{\sigma,t}[1 - \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)] \tag{13}$$

$$\Psi_i \equiv \frac{\partial^2 v_i}{\partial \Lambda_{\sigma,i}\partial \Lambda_{\sigma,i}^\top} = -\phi_5^{-2}I - \frac{1}{2}\sum_{t=1}^{T} F_{\sigma,t}F_{\sigma,t}^\top \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t) \tag{14}$$

I compute g_i and an approximation $\hat{\Psi}_i = -(\phi_5^{-2} + T/2)I$ to Ψ_i which is much faster to compute than Ψ_i. The approximation is based on the fact that the conditional expectation

of each term in the sum (14), given Φ, Λ and B—and *not* $\tilde{y}$—is one. Using the approximation may reduce numerical efficiency, but since the approximation is incorporated into the Hastings ratio, the update still preserves the correct conditional posterior distribution.

I draw $\Lambda^*_{\sigma,i}|\Lambda_{\sigma,i}$ from the proposal distribution $N(\Lambda_{\sigma,i} - \hat{\Psi}_i^{-1} g_i, -\hat{\Psi}_i^{-1})$, whose log density has gradient g_i and Hessian $\hat{\Psi}_i$. Let $q(\cdot|\Lambda_{\sigma,i})$ denote the density of this distribution.

Next, I evaluate v_i and g_i at $\Lambda^*_{\sigma,i}$ and denote these by v_i^* and g_i^*. Let $q(\cdot|\Lambda^*_{\sigma,i})$ be the density of the distribution $N(\Lambda^*_{\sigma,i} - \hat{\Psi}_i^{-1} g_i^*, -\hat{\Psi}_i^{-1})$.

I accept $\Lambda^*_{\sigma,i}$ as the new state with probability $\min[1, R(\Lambda_{\sigma,i}, \Lambda^*_{\sigma,i})]$, where the Hastings ratio R is given by:

$$R(\Lambda_{\sigma,i}, \Lambda^*_{\sigma,i}) = \exp(v_i^* - v_i) \frac{q(\Lambda_{\sigma,i}|\Lambda^*_{\sigma,i})}{q(\Lambda^*_{\sigma,i}|\Lambda_{\sigma,i})}.$$

Now let's turn to the case where i is a factor founder, say the j'th. I need to add the term $(K_\sigma - j)\log \Lambda_{\sigma,ij}$ to v_i to account for the scaled χ distribution of $\Lambda_{\sigma,ij}$ and the term $(K_\sigma - j)/\Lambda_{\sigma,ij}$ to the j'th element of g_i. Then, with $\Lambda^*_{\sigma,i}$ and g_i^* interpreted as their leading $j \times 1$ subvectors and $\hat{\Psi}_i$ as its leading $j \times j$ submatrix, we proceed as before to obtain a draw $\Lambda^*_{\sigma,i}$ and accept or reject.

Then if $\Lambda_{\sigma,i}$ is accepted and $\Lambda^*_{\sigma,ij} < 0$, I multiply the following quantities by minus one: the j'th column of Λ; $F_{\sigma tj}$, $t = 1, \ldots, T$, $\Phi_{\sigma,jk}$ and $\Phi_{\sigma,kj}$, $k \neq j$. As before, this maps a value of $(\Lambda_\sigma, F_\sigma, \Phi_\sigma)$ outside the truncation region to an observationally equivalent value inside.

In practice, it is helpful to limit the size of the proposal step when the term $-\hat{\Psi}_i^{-1} g_i$ is very large. The acceptance probability in these cases may be so small that the MCMC chain gets stuck. The term is often very large during the initial burn-in period, when the distribution of the MCMC chain is far from its invariant distribution. Let $\hat{\Psi}_i = LL^\top$ be the Cholesky decomposition of $\hat{\Psi}_i$. When the Euclidean length of $L^{-1} g_i$ is greater than $2K$, I replace $\hat{\Psi}_i^{-1} g_i$ with $(2K/\|L^{-1} g_i\|)\hat{\Psi}_i^{-1} g_i$ (and similarly with $L^{-1} g_i^*$ to preserve detailed balance).

4.2. **Drawing B_σ.** I update B_σ in a similar way as I do Λ_σ. In the conditional posterior distribution of B_σ, rows $B_{\sigma,i}$ are independent and I draw them one at a time. Mechanically, the procedure is the same as for Λ_σ; only the expressions for the value, gradient and Hessian of the log posterior density differ.

Up to an additive term not depending on $B_{\sigma,i}$, the log conditional posterior density of $B_{\sigma,i}$ is

$$v_i = -\frac{1}{2}[B_{\sigma,i}\bar{H}_{B_{\sigma,i}}B_{\sigma,i}^\top - 2B_{\sigma,i}\bar{c}_{B_{\sigma,i}}] - \frac{1}{2}[K_\mu B_{\sigma,i1} + \phi_2^{-2}\exp(-B_{\sigma,i1})\Lambda_{\mu,i}\Lambda_{\mu,i}^\top]$$
$$-\frac{1}{2}\left[B_{\sigma,i}\sum_{t=1}^T x_t + \sum_{t=1}^T \tilde{y}_{ti}^2\exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)\right].$$

The first term comes from $f(B_\sigma)$, the second from $f(\Lambda_\mu|B_\sigma)$ and the third from $f(\tilde{y}|\Lambda_\sigma, B_\sigma, F_\sigma)$.

The gradient and Hessian of v_i with respect to $B_{\sigma,i}$ are given by g_i and Ψ_i:

$$g_i \equiv \frac{\partial v_i}{\partial B_{\sigma,i}^\top} = -\bar{H}_{B_{\sigma,i}}B_{\sigma,i}^\top + \bar{c}_{B_{\sigma,i}} - \frac{1}{2}\left[\sum_{t=1}^T x_t[1 - \tilde{y}_{ti}^2\exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)]\right]$$
$$-\frac{1}{2}[K_\mu - \phi_2^{-2}\exp(-B_{\sigma,i1})\Lambda_{\mu,i}\Lambda_{\mu,i}^\top]e_1^{(1)}$$

$$\Psi_i \equiv \frac{\partial^2 v_i}{\partial B_{\sigma,i}\partial B_{\sigma,i}^\top} = -\bar{H}_{B_{\sigma,i}} - \frac{1}{2}\left[\sum_{t=1}^T x_t x_t^\top \tilde{y}_{ti}^2\exp(-\Lambda_{\sigma,i}F_{\sigma,t} - B_{\sigma,i}x_t)\right]$$
$$+\frac{1}{2}\phi_2^{-2}\exp(-B_{\sigma,i1})\Lambda_{\mu,i}\Lambda_{\mu,i}^\top E_{11}^{(1)},$$

where $e^{(1)}$ is a K_σ-vector whose only non-zero element is $e_1^{(1)} = 1$ and $E^{(1)}$ is a $K_\sigma \times K_\sigma$ matrix whose only non-zero element is $E_{11}^{(11)} = 1$.

Here I use $\hat{\Psi}_i = -\bar{H}_{B_{\sigma,i}} - \frac{1}{2}\sum_{t=1}^T x_t x_t^\top$ as an approximation to the Hessian matrix.

4.3. **Drawing Φ_σ and F_σ.** I update the k'th row $\Phi_{\sigma,k}$ of Φ_σ and the k'th variance factor series $F_{\sigma,k} \equiv (F_{\sigma,1k}, \ldots, F_{\sigma,Tk})$, jointly, one k at a time. The proposal $(\Phi_{\sigma,k}^*, F_{\sigma,k}^*)$ consists of a random walk proposal $\Phi_{\sigma,k}^* \sim N(\Phi_{\sigma,k}, \Omega_k)$ followed by a conditional proposal of $F_{\sigma,k}^*$ given $\Phi_{\sigma,k}^*$. The joint proposal $(\Phi_{\sigma,k}^*, F_{\sigma,k}^*)$ is accepted or rejected as a unit. The proposal density $q(F_{\sigma,k}^*|\Phi_{\sigma,k}^*)$ is a very close approximation of the conditional posterior density of $F_{\sigma,k}$ given $\Phi_{\sigma,k}^*$ and the current values of all other quantities, including the other factor series. I construct it according to the HESSIAN method described in McCausland (2012).

The HESSIAN method is a generic method for building an approximation of the conditional posterior distribution of states in state space models where states are univariate and Gaussian, but observations can be non-Gaussian, non-linear and multivariate. I can apply it here because each factor series, together with $\tilde{y}$, becomes such a state space model when we condition on all the other factor series. What this means, more precisely, is that

the Hessian matrix of $\ln f(F_{\sigma,k}|F_{-k}, \tilde{y}, \Lambda, B, \Phi)$—where F_{-k} denotes all factors except the k'th—is tri-diagonal and the diagonal elements do not depend on $\tilde{y}$.

To apply the HESSIAN method, I need to supply inputs specifying the conditional Gaussian factor distribution $F_{\sigma,k}|F_{-k}, \Phi$ and the measurement distribution $\tilde{y}|F, \Lambda, B$. To specify the factor distribution, I need to provide its tridiagonal precision matrix H_k and its covector c_k. To specify the measurement distribution $\tilde{y}|F, \Lambda, B$, I need to supply (routines to perform) multiple derivatives of $\log f(\tilde{y}_t|F, \Lambda, B)$ with respect to F_{tk}, for $t = 1, \ldots, T$. We spend the rest of this Section 4.3 doing both.

First, we find the precision and covector of the conditional factor distribution $F_{\sigma,k}|F_{-k}, \Phi$. The conditional distribution of F given Φ is Gaussian, with mean zero and block band precision

$$\mathbf{H} \equiv \begin{bmatrix} I & -\Phi^\top & & \\ & I & \ddots & \\ & & \ddots & -\Phi^\top \\ & & & I \end{bmatrix} \begin{bmatrix} I & & & \\ & I - \Phi\Phi^\top & & \\ & & \ddots & \\ & & & I - \Phi\Phi^\top \end{bmatrix}^{-1} \begin{bmatrix} I & & & \\ -\Phi & I & & \\ & \ddots & \ddots & \\ & & -\Phi & I \end{bmatrix}.$$

The derivation of this is very similar to the derivation in Appendix B. The matrix $\mathbf{H}$ has $T \times T$ blocks, each of dimension $K \times K$. The non-zero blocks are

$$\mathbf{H}_{11} = I + \Phi^\top (I - \Phi\Phi^\top)^{-1}\Phi, \quad \mathbf{H}_{nn} = (I - \Phi\Phi^\top)^{-1},$$

$$\mathbf{H}_{tt} = (I - \Phi\Phi^\top)^{-1} + \Phi^\top (I - \Phi\Phi^\top)^{-1}\Phi, \quad t = 2, \ldots, n-1,$$

$$\mathbf{H}_{t,t-1} = -(I - \Phi\Phi^\top)^{-1}\Phi, \quad t = 2, \ldots, n.$$

Now I specify the conditional distribution $F_{\sigma,k}|F_{-k}, \Phi \sim N(H_k^{-1}c_k, H_k^{-1})$. The non-zero elements of the $T \times T$ conditional precision matrix H_k and the $T \times 1$ covector c_k are

$$H_{k,11} = (\mathbf{H}_{11})_{kk}, \quad H_{k,TT} = (\mathbf{H}_{TT})_{kk},$$

$$H_{k,tt} = (\mathbf{H}_{tt})_{kk}, \quad t = 2, \ldots, n-1,$$

$$H_{k,t,t-1} = (\mathbf{H}_{t,t-1})_{kk}, \quad t = 2, \ldots, n,$$

$$c_{k,1} = \sum_{l \neq k} (\mathbf{H}_{tt})_{kl} F_{tl} + (\mathbf{H}_{t+1,t})_{lk} F_{t+1,l}, \quad c_{k,n} = \sum_{l \neq k} (\mathbf{H}_{tt})_{kl} F_{tl} + (\mathbf{H}_{t,t-1})_{kl} F_{t-1,l}$$

$$c_{k,t} = \sum_{l \neq k} (\mathbf{H}_{tt})_{kl} F_{tl} + (\mathbf{H}_{t,t-1})_{kl} F_{t-1,l} + (\mathbf{H}_{t+1,t})_{lk} F_{t+1,l}, \quad t = 2, \ldots, n-1.$$

We now turn to the specification of the measurement distribution $\tilde{y}_t | F_t, x_t$. We need to supply code to evaluate multiple derivatives of $\psi(F_t) \equiv \log f(\tilde{y}_t | F_t, x_t)$ with respect to F_{tk}, at any point F_{tk}. For $t = 1, \ldots, T$,

$$\tilde{y}_t | F, x \sim N \left(0, \begin{bmatrix} \exp(\Lambda_1 F_t + B_1 x_t) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \exp(\Lambda_N F_t + B_N x_t) \end{bmatrix} \right).$$

Therefore

$$\psi(F_t) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^{N} (\Lambda_{\sigma,i} F_t + B_{\sigma,i} x_t) - \frac{1}{2} \sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i} F_t - B_{\sigma,i} x_t), \tag{15}$$

and we compute

$$\frac{\partial \psi(F_t)}{\partial F_{tk}} = -\frac{1}{2} \sum_{i=1}^{N} \Lambda_{ik} + \frac{1}{2} \sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i} F_t - B_{\sigma,i} x_t) \Lambda_{ik}, \tag{16}$$

$$\frac{\partial^n \psi(F_t)}{\partial F_{tk}^n} = \frac{(-1)^{n+1}}{2} \sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i} F_t - B_{\sigma,i} x_t) \Lambda_{ik}^n, \quad n = 1, 2, \ldots \tag{17}$$

Using the HESSIAN method, we need to evaluate derivatives of $\psi(F_t)$ with respect to F_{tk} many times, so it is helpful to pre-compute quantities not depending on the F_{tk}. For any k, write

$$\sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp(-\Lambda_{\sigma,i} F_t - B_{\sigma,i} x_t) = \sum_{i=1}^{N} \tilde{y}_{ti}^2 \exp\left[-\sum_{l=1, l \neq k}^{K} \Lambda_{il} F_{tl} - B_{\sigma,i} x_t \right] \exp(-\Lambda_{ik} F_{tk})$$

and then define

$$c_{ti}(k) \equiv \tilde{y}_{ti}^2 \exp\left[-\sum_{l=1, l \neq k}^{K} \Lambda_{il} F_{tl} - B_{\sigma,i} x_t \right].$$

Pre-computing the $c_{ti}(k)$, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ before applying the HESSIAN method to draw $F_{\sigma,k}$ allows for computationally efficient code.

Hyper-parameters	Description	GIR	FRED-FM	BofC
(ϕ_1, ϕ_2)	Φ_μ diagonal, Beta shape	$(200, 200)$	$(4, 2)$	$(200, 200)$
(ϕ_3, ϕ_4)	Φ_μ off-diagonal, Beta shape	$(200, 200)$	$(20, 20)$	$(200, 200)$
(ϕ_5, ϕ_6)	Φ_σ diagonal, Beta shape	$(18, 2)$	$(10.5, 1.5)$	$(18, 2)$
(ϕ_7, ϕ_8)	Φ_σ off-diagonal, Beta shape	$(50, 50)$	$(50, 50)$	$(50, 50)$
(ϕ_9, ϕ_{10})	B_μ, Gaussian mean, std	$(0, 0.01)$	(see below)	$(0, 0.001)$
(ϕ_{11}, ϕ_{12})	B_σ, Gaussian mean, std	$(-11, 0.1)$	(see below)	$(-11, 1)$
ϕ_{13}	Λ_μ, Gaussian std	0.5	1.0	2.0
ϕ_{14}	Λ_σ, Gaussian std	0.1	0.4	0.4

TABLE 1. Prior hyper-parameter values used in Getting It Right (GIR), FRED-MD panel and Bank of Canada panel applications.

Description	$E[B_{\mu,i}]$	$\mathrm{sd}[B_{\mu,i}]$	$E[B_{\mu,i}]$	$\mathrm{sd}[B_{\mu,i}]$
Real production, income or consumption, $\Delta \log$	$\ln 10^{-5}$	$\ln 10$	0.002	0.002
Institute for Supply Management index	$\frac{1}{2} \ln 10$	$\ln 10$	50	10
Numbers employed, $\Delta \log$	$\ln 10^{-5}$	$\ln 10$	0.001	0.001
Numbers unemployed, $\Delta \log$	$\ln 10^{-2}$	$\ln 10$	0.001	0.001
Housing starts and permits, $\Delta \log$	$\ln 10^{-2}$	$\ln 10$	0.001	0.001
Price index, $\Delta^2 \log$	$\ln 10^{-5}$	$\ln 10$	0	10^{-6}
Money and Credit aggregates, $\Delta^2 \log$	$\ln 10^{-5}$	$\ln 10$	0	10^{-6}
Nominal Interest rates and yields, Δ	$\ln 10^{-1}$	$\ln 10$	0	10^{-6}
Nominal Interest rate spreads	$\ln 10^{-2}$	$\ln 10$	1	1
Nominal and Real Exchange rates, $\Delta \log$	$\ln 10^{-4}$	$\ln 10$	0	10^{-4}
Hourly Earnings, $\Delta^2 \log$	$\ln 10^{-5}$	$\ln 10$	0	10^{-6}
Asset price, $\Delta \log$	$\ln 10^{-4}$	$\ln 10$	0.002	0.002
Price index, materials and commodities, $\Delta^2 \log$	$\ln 10^{-4}$	$\ln 10$	0	10^{-6}
Hours, Δ	$\ln 10^{-1}$	$\ln 10$	0	10^{-4}
Miscellaneous, low error	$\ln 10^{-4}$	$\ln 100$	0	0.001
Miscellaneous, high error	$\ln 1$	$\ln 100$	0	0.1

TABLE 2. Prior hyper-parameter values used in FRED-MD panel analysis

5. RESULTS

Here I report results from an artificial data exercise designed to test the algorithms used, and two empirical applications. The first empirical application uses data from a panel of macroeconomic indicators. The second analyses data from a panel of currency returns.

Tables 1 and 2 give numerical values for the various hyper-parameters used in the three simulations.

5.1. **Getting it right.** Here I describe a simulation whose sole purpose is to test the correctness of the posterior simulation methods. This is a purely pre-data exercise, involving

only artificial data. The tests described here are similar to those described in Geweke (2004). We draw a sample from the *joint* distribution of parameters, latent factors *and data*, using the simulation methods described in Sections 3 and 4. If the posterior simulation methods are conceptually sound and correctly implemented, then the marginal distribution of the parameters in the sample is the same as their prior distribution. This is a testable implication of program correctness.

I set the dimension parameters $N = 3$, $J = 1$, $K_\mu = K_\sigma = 2$ and $T = 10$. These values are very small, to avoid excessive serial dependence in the sample and to obtain a very large sample in reasonable time, but sufficiently large to test the correctness of our code. The exogenous series x_t is a constant: $x_t = 1$ for $t = 1, \ldots, T$. Hyper-parameter values are shown in Table 1. I generate a posterior sample of size of $M = 4 \times 10^6$ and for analysis, I use a subsample of size 100000 consisting of every 40'th draw.

The sample is the output of a Markov chain whose invariant distribution is the joint distribution of parameters, latent factors and data. This Markov chain is a Gibbs sampler, consisting of exactly the same blocks used for posterior simulation, described in Sections 3 and 4, plus an additional block, updating data y from its conditional distribution given all parameters and both factor series. The additional block simply involves simulating data from the model.

Table 3 shows the results for various parameters. The second, third and fourth columns give the population and sample mean of the parameter and a numerical standard error for the sample mean. The fifth and sixth columns give the population and sample mean of the squared difference between the parameter and its true population mean. The seventh column gives a numerical standard error of the latter sample mean. Numerical standard errors are computed using the R package `coda`, which uses a time series method.

Sample means and variances are close to true prior means and variances, measured in terms of numerical standard error, except for the parameters $\Phi_{\sigma kl}$, where the given "true" prior means and variances are pre-truncation and the simulated values come from the truncated distribution.

Tables 4 and 5 show results for the mean and variance factor series, respectively. Recall that the marginal distribution of all factors is standard Gaussian. Each row corresponds to a time period. The first and second columns of Table 4 give the sample mean of $F_{\mu,t1}$

	$\mu \equiv E[x]$	$\bar{x}$	$\hat{\sigma}_{\text{nse},\mu}$	$\text{Var}[x]$	$\overline{(x-\mu)^2}$	$\hat{\sigma}_{\text{nse},\sigma^2}$
$\Lambda_{\mu,11}$	0.62666	6.278e-01	6.9e-04	1.073e-01	1.082e-01	3.4e-04
$\Lambda_{\mu,21}$	0.00000	-1.158e-03	1.1e-03	2.500e-01	2.502e-01	7.3e-04
$\Lambda_{\mu,22}$	0.39894	3.988e-01	6.1e-04	9.085e-02	9.117e-02	3.2e-04
$\Lambda_{\mu,31}$	0.00000	-2.363e-04	1.0e-03	2.500e-01	2.493e-01	7.2e-04
$\Lambda_{\mu,32}$	0.00000	-9.767e-04	1.0e-03	2.500e-01	2.507e-01	7.2e-04
$B_{\mu,11}$	0.00000	1.907e-04	1.1e-04	1.000e-04	9.895e-05	2.8e-07
$B_{\mu,21}$	0.00000	-5.488e-05	1.1e-04	1.000e-04	9.986e-05	2.8e-07
$B_{\mu,31}$	0.00000	-6.607e-05	1.1e-04	1.000e-04	9.832e-05	2.8e-07
$\Phi_{\mu,11}$	0.00000	-3.525e-04	1.0e-04	2.494e-03	2.503e-03	7.2e-06
$\Phi_{\mu,12}$	0.00000	-6.996e-05	1.0e-04	2.494e-03	2.494e-03	7.1e-06
$\Phi_{\mu,21}$	0.00000	-1.279e-04	1.0e-04	2.494e-03	2.490e-03	7.1e-06
$\Phi_{\mu,22}$	0.00000	-1.070e-04	1.0e-04	2.494e-03	2.500e-03	7.1e-06
$\Lambda_{\sigma,11}$	0.12533	1.253e-01	1.4e-04	4.292e-03	4.299e-03	1.5e-05
$\Lambda_{\sigma,21}$	0.00000	-1.073e-04	2.0e-04	1.000e-02	9.984e-03	2.8e-05
$\Lambda_{\sigma,22}$	0.07979	7.986e-02	1.2e-04	3.634e-03	3.643e-03	1.2e-05
$\Lambda_{\sigma,31}$	0.00000	-8.351e-05	2.0e-04	1.000e-02	9.999e-03	2.8e-05
$\Lambda_{\sigma,32}$	0.00000	1.363e-04	2.0e-04	1.000e-02	9.969e-03	2.8e-05
$B_{\sigma,11}$	-11.00000	-1.100e+01	2.0e-04	1.000e-02	1.002e-02	2.8e-05
$B_{\sigma,21}$	-11.00000	-1.100e+01	2.0e-04	1.000e-02	9.956e-03	2.9e-05
$B_{\sigma,31}$	-11.00000	-1.100e+01	2.0e-04	1.000e-02	1.001e-02	2.9e-05
$\Phi_{\sigma,11}$	0.80000	7.904e-01	5.1e-04	1.714e-02	1.712e-02	8.0e-05
$\Phi_{\sigma,12}$	0.00000	-2.125e-05	1.9e-04	9.901e-03	8.829e-03	3.3e-05
$\Phi_{\sigma,21}$	0.00000	-2.852e-05	1.9e-04	9.901e-03	8.828e-03	3.2e-05
$\Phi_{\sigma,22}$	0.80000	7.892e-01	5.0e-04	1.714e-02	1.717e-02	7.9e-05

TABLE 3. Sample and population means and variances in "Getting it right" experiment

and a numerical standard error for this sample mean. The third and fourth columns give the sample mean of $F^2_{\mu,t1}$ and a numerical standard error. The fifth through eighth columns show the same information as the first four columns, for the second factor series, F_{t2}. Table 5 shows the same information as Table 4, for the variance factor series. Again, sample means and variances are close to true prior means and variances, measured in terms of numerical standard error.

5.2. **An application to a panel of macroeconomic and financial data.** Here I analyse a panel of 134 U.S. macroeconomic and financial indicators, observed monthly from January 1959 to January 2015. The data are from the FRED-MD database described by McCracken and Ng (2014). I transform the data as described in that paper, obtaining an unbalanced panel with $N = 134$ indicators and $T = 670$ observation periods. I also transform the data so that the sample mean and variance of each series are zero and one.

	$F_{\mu,t1}$	nse, $F_{\mu,t1}$	$F^2_{\mu,t1}$	nse, $F^2_{\mu,t1}$	$F_{\mu,t2}$	nse, $F_{\mu,t2}$	$F^2_{\mu,t2}$	nse, $F^2_{\mu,t2}$
1	0.0020	0.0012	1.000	0.0017	-0.0028	0.0017	0.999	0.0014
2	0.0025	0.0013	0.999	0.0014	0.0006	0.0013	1.000	0.0013
3	0.0025	0.0013	1.001	0.0015	-0.0008	0.0011	0.998	0.0016
4	-0.0001	0.0013	1.001	0.0019	0.0006	0.0012	1.002	0.0016
5	-0.0012	0.0017	1.000	0.0017	0.0000	0.0012	0.999	0.0013
6	-0.0006	0.0014	0.999	0.0016	0.0006	0.0010	0.997	0.0012
7	-0.0010	0.0013	0.997	0.0019	0.0007	0.0010	0.998	0.0016
8	0.0011	0.0019	0.999	0.0014	-0.0012	0.0013	1.000	0.0016
9	0.0002	0.0013	1.001	0.0019	-0.0002	0.0010	1.002	0.0020
10	0.0010	0.0011	1.001	0.0017	-0.0003	0.0009	1.000	0.0014

TABLE 4. Getting it right: moments of mean factors

	$F_{\sigma,t1}$	nse, $F_{\sigma,t1}$	$F^2_{\sigma,t1}$	nse, $F^2_{\sigma,t1}$	$F_{\sigma,t2}$	nse, $F_{\sigma,t2}$	$F^2_{\sigma,t2}$	nse, $F^2_{\sigma,t2}$
1	0.0024	0.0012	0.999	0.0013	0.0008	0.0010	1.000	0.0014
2	0.0028	0.0015	1.000	0.0016	-0.0003	0.0011	0.999	0.0019
3	0.0027	0.0011	1.000	0.0015	-0.0000	0.0010	1.000	0.0020
4	0.0022	0.0012	0.999	0.0011	-0.0009	0.0010	1.002	0.0014
5	0.0011	0.0012	0.998	0.0009	-0.0002	0.0010	1.001	0.0013
6	0.0014	0.0012	0.997	0.0012	-0.0005	0.0009	1.001	0.0009
7	0.0009	0.0009	1.000	0.0013	-0.0005	0.0009	1.000	0.0008
8	0.0003	0.0008	1.000	0.0015	-0.0000	0.0010	0.998	0.0014
9	-0.0009	0.0010	1.000	0.0015	0.0014	0.0009	0.999	0.0012
10	0.0007	0.0011	1.000	0.0014	0.0005	0.0009	1.000	0.0012

TABLE 5. Getting it right: moments of variance factors

Table 6 shows the posterior mean of mean factor loadings for the 40 indicators i with the largest sums $\sum_{k=1}^{K_\mu} \Lambda^2_{\mu,ik}$. The sum is given in the first column of the table. Table 7 shows the posterior mean of variance factor loadings for the 40 indicators i with the largest sums $\sum_{k=1}^{K_\sigma} \Lambda^2_{\sigma,ik}$.

Table 8 shows the posterior mean, the posterior standard deviation, the numerical standard error and the relative numerical efficiency of the elements of Φ_μ. Each of the four panels shows a 7×7 array of values, corresponding elementwise to the matrix Φ_μ. Table 9 shows the same information for the Φ_σ matrix.

Figures 2, 3, 4, 5, 6, 7 and 8 show time series plots pertaining to the mean factors 1 through 7. The upper panel of these figures shows the posterior mean as a function of t. The lower panel shows the posterior standard deviation and the numerical standard error for the mean. Figures 9 and 10 show time series plots for the two variance factors. Again,

the upper panel shows the posterior mean and the lower panel shows the posterior standard deviation and the numerical standard error for the mean.

	var	$L_{\mu,i1}$	$L_{\mu,i2}$	$L_{\mu,i3}$	$L_{\mu,i4}$	$L_{\mu,i5}$	$L_{\mu,i6}$	$L_{\mu,i7}$
S and P 500	0.011	0.002	0.004	0.002	0.014	0.010	0.000	0.956
S and P: indust	0.011	0.003	0.003	0.005	0.008	0.010	0.000	0.960
NAPM	0.016	0.478	0.001	0.000	0.009	0.493	0.001	0.002
IPFPNSS	0.019	0.355	0.002	0.000	0.003	0.010	0.612	0.000
AAAFFM	0.021	0.002	0.976	0.000	0.000	0.001	0.000	0.000
TB6MS	0.025	0.037	0.008	0.000	0.923	0.006	0.000	0.000
PAYEMS	0.027	0.923	0.008	0.001	0.000	0.031	0.009	0.000
GS1	0.028	0.032	0.006	0.001	0.933	0.001	0.000	0.000
T10YFFM	0.031	0.005	0.961	0.000	0.002	0.002	0.000	0.000
BAAFFM	0.033	0.017	0.950	0.000	0.000	0.000	0.000	0.000
IPFINAL	0.033	0.276	0.001	0.001	0.002	0.008	0.679	0.000
CUSR0000SAC	0.082	0.000	0.000	0.918	0.000	0.000	0.000	0.000
T5YFFM	0.088	0.008	0.897	0.000	0.006	0.001	0.000	0.000
INDPRO	0.095	0.446	0.008	0.000	0.006	0.002	0.443	0.000
DNDGRG3M086SBEA	0.097	0.000	0.000	0.902	0.000	0.000	0.000	0.000
TB3MS	0.098	0.037	0.012	0.000	0.838	0.013	0.000	0.002
CUSR0000SA0L5	0.100	0.000	0.000	0.899	0.000	0.000	0.000	0.000
CPIAUCSL	0.100	0.000	0.001	0.898	0.000	0.001	0.000	0.000
NAPMPI	0.110	0.432	0.042	0.000	0.012	0.394	0.003	0.007
S and P div yield	0.111	0.000	0.009	0.001	0.026	0.015	0.000	0.839
SRVPRD	0.114	0.763	0.022	0.002	0.001	0.076	0.021	0.000
IPMANSICS	0.126	0.458	0.010	0.000	0.005	0.005	0.395	0.000
NAPMNOI	0.137	0.405	0.061	0.001	0.019	0.357	0.009	0.012
IPCONGD	0.138	0.145	0.006	0.000	0.003	0.024	0.685	0.000
CPIULFSL	0.142	0.000	0.000	0.857	0.000	0.000	0.000	0.001
CP3M	0.142	0.057	0.016	0.001	0.769	0.013	0.000	0.002
USGOOD	0.167	0.833	0.000	0.000	0.000	0.000	0.000	0.000
CAPUTLB00004S	0.168	0.361	0.053	0.000	0.008	0.000	0.411	0.000
NAPMEI	0.172	0.423	0.004	0.000	0.007	0.395	0.000	0.000
S and P PE ratio	0.193	0.000	0.016	0.000	0.014	0.020	0.000	0.757
CPITRNSL	0.198	0.000	0.000	0.801	0.000	0.000	0.000	0.000
MANEMP	0.206	0.767	0.001	0.000	0.000	0.020	0.005	0.000
PCEPI	0.215	0.000	0.000	0.783	0.000	0.001	0.000	0.000
GS5	0.219	0.021	0.007	0.005	0.748	0.000	0.000	0.000
DMANEMP	0.236	0.725	0.002	0.001	0.001	0.030	0.006	0.000
USTPU	0.249	0.695	0.011	0.003	0.000	0.013	0.028	0.000
CUUR0000SA0L2	0.294	0.000	0.000	0.706	0.000	0.000	0.000	0.000
NAPMSDI	0.318	0.186	0.012	0.000	0.007	0.474	0.000	0.003
GS10	0.342	0.017	0.008	0.008	0.625	0.000	0.000	0.000
IPBUSEQ	0.349	0.345	0.001	0.003	0.000	0.002	0.299	0.001

TABLE 6. FRED panel, posterior mean of mean factor loadings for indicators with large mean factor loadings

	var	$L_{\sigma,i1}$	$L_{\sigma,i2}$
EXUSUK	2.490	-0.421	1.521
EXSZUS	2.397	-0.330	1.513
EXJPUS	2.186	-0.471	1.401
TB6MS	1.831	1.325	0.274
PAYEMS	1.824	1.337	-0.188
S and P PE ratio	1.498	-0.149	1.215
COMPAPFF	1.301	1.141	0.000
CP3M	1.264	1.042	0.422
FEDFUNDS	1.251	1.103	0.187
T5YFFM	1.226	-0.171	1.094
S and P 500	1.165	-0.075	1.077
CUSR0000SA0L5	1.126	0.913	-0.541
TB3MS	1.099	0.964	0.413
SRVPRD	1.085	1.035	-0.116
T10YFFM	1.076	-0.048	1.036
CPIULFSL	1.070	0.872	-0.556
TB6SMFFM	0.996	0.991	0.121
TB3SMFFM	0.935	0.962	0.093
S and P div yield	0.884	0.181	0.923
CPIAUCSL	0.863	0.871	-0.322
AMBSL	0.839	-0.200	0.894
INDPRO	0.786	0.882	0.089
RPI	0.766	0.446	0.753
CUSR0000SAS	0.757	0.815	-0.304
NAPMSDI	0.750	0.856	-0.131
IPMAT	0.732	0.855	-0.014
DMANEMP	0.728	0.853	-0.009
W875RX1	0.715	0.217	0.817
AAAFFM	0.675	0.791	-0.222
MZMSL	0.653	0.310	0.747
IPDMAT	0.647	0.797	-0.105
USCONS	0.641	0.758	-0.260
GS10	0.636	-0.018	0.797
DTCOLNVHFNM	0.629	-0.362	0.706
T1YFFM	0.616	0.716	0.320
CES2000000008	0.599	0.731	-0.255
BAA	0.583	-0.024	0.763
S and P: indust	0.578	-0.297	0.700
USGOOD	0.576	0.751	0.113
BAAFFM	0.575	0.723	-0.230

TABLE 7. FRED panel, posterior mean of variance factor loadings for indicators with large variance factor loadings

i/j	1	2	3	4	5	6	7
1	0.765	0.054	0.040	0.064	0.127	-0.016	0.067
2	-0.059	0.967	0.013	-0.092	-0.022	-0.004	-0.022
3	-0.014	0.006	-0.214	0.011	0.001	0.001	0.146
4	0.025	0.085	0.115	0.333	-0.067	-0.043	0.142
5	0.262	0.036	-0.016	0.066	0.714	0.104	0.008
6	-0.093	0.020	0.002	0.070	0.027	-0.121	-0.007
7	-0.007	0.029	-0.080	-0.108	-0.055	-0.021	0.219
1	0.026	0.018	0.033	0.028	0.028	0.032	0.026
2	0.017	0.004	0.015	0.014	0.014	0.014	0.012
3	0.033	0.015	0.043	0.041	0.045	0.043	0.036
4	0.028	0.014	0.039	0.039	0.040	0.041	0.034
5	0.027	0.019	0.046	0.040	0.042	0.038	0.031
6	0.030	0.015	0.042	0.040	0.040	0.045	0.038
7	0.025	0.011	0.038	0.035	0.031	0.038	0.034
1	0.001	0.001	0.001	0.001	0.001	0.001	0.001
2	0.002	0.000	0.000	0.001	0.001	0.000	0.001
3	0.001	0.000	0.001	0.001	0.001	0.001	0.001
4	0.001	0.001	0.001	0.001	0.002	0.001	0.001
5	0.001	0.002	0.001	0.002	0.002	0.001	0.001
6	0.001	0.001	0.001	0.001	0.001	0.001	0.001
7	0.001	0.000	0.001	0.001	0.001	0.001	0.001
1	0.179	0.070	0.461	0.766	0.377	0.534	0.816
2	0.064	0.128	0.631	0.259	0.120	0.993	0.262
3	0.554	0.572	0.972	0.504	0.524	0.888	0.671
4	0.412	0.309	0.859	0.699	0.358	0.812	0.657
5	0.292	0.070	0.715	0.272	0.190	0.411	0.455
6	0.709	0.250	0.879	0.703	0.438	0.728	0.503
7	0.750	0.686	0.655	0.591	0.737	0.491	0.778

TABLE 8. FRED panel, posterior mean and standard deviation, numerical standard error and relative numerical efficiency of Φ_μ, elementwise

5.3. **An application to daily exchange rates.** This section provides an analysis of a panel of daily log returns of 10 currencies relative to the US dollar: the Australian Dollar (AUD), Brazilian Real (BRL), Euro (EUR), Japanese Yen (JPY), Mexican Peso (MXN), New Zealand Dollar (NZD), Singapore Dollar (SGD), Swiss Franc (CHF), British Pound (GBP), and Canadian Dollar (CAD). The sample covers the period from July 8, 2005 to July 8, 2015 We use noon spot rates of the 10 currencies and the US dollar, denominated in Canadian dollars, obtained from the Bank of Canada. We compute the log returns of the exchange rates and remove returns for those days when one or more of the markets was closed, giving 2505 observations for each return series.

i/j	1	2
1	0.785	-0.063
2	-0.067	0.832
1	0.019	0.019
2	0.018	0.018
1	0.001	0.001
2	0.001	0.001
1	0.306	0.266
2	0.371	0.162

TABLE 9. FRED panel, posterior mean and standard deviation, numerical standard error and relative numerical efficiency of Φ_σ, elementwise

FIGURE 2. FRED panel, mean factor 1. Upper panel: posterior mean of $F_{\mu,t1}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t1}$ and numerical standard error for the posterior mean

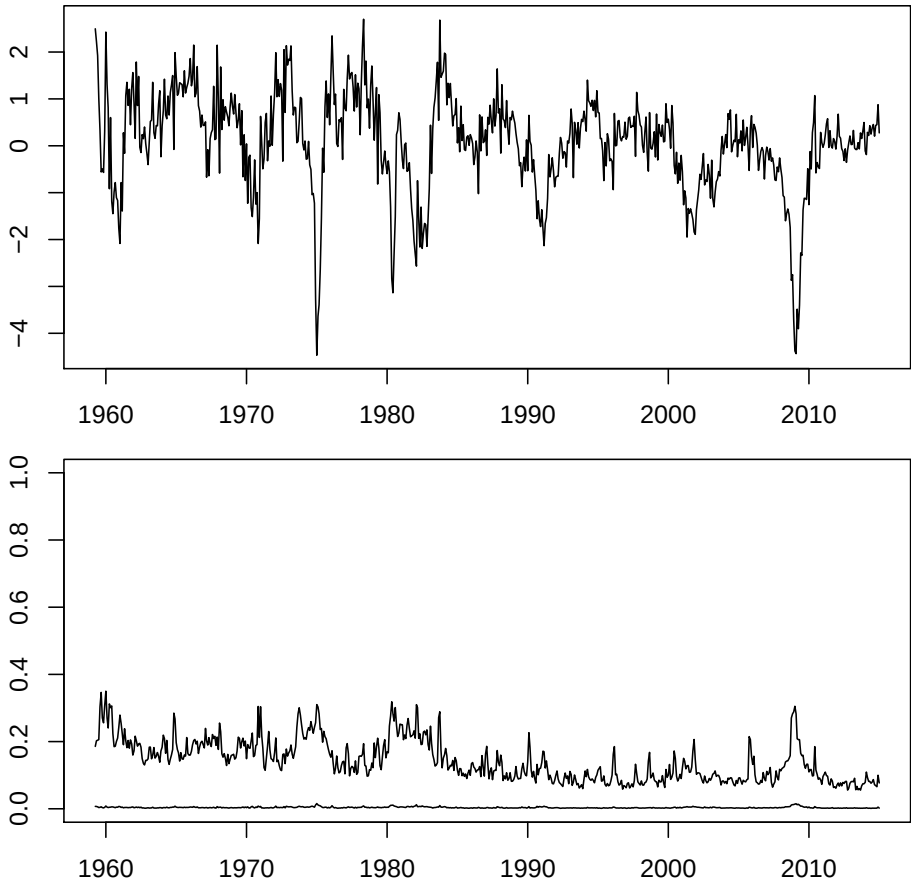
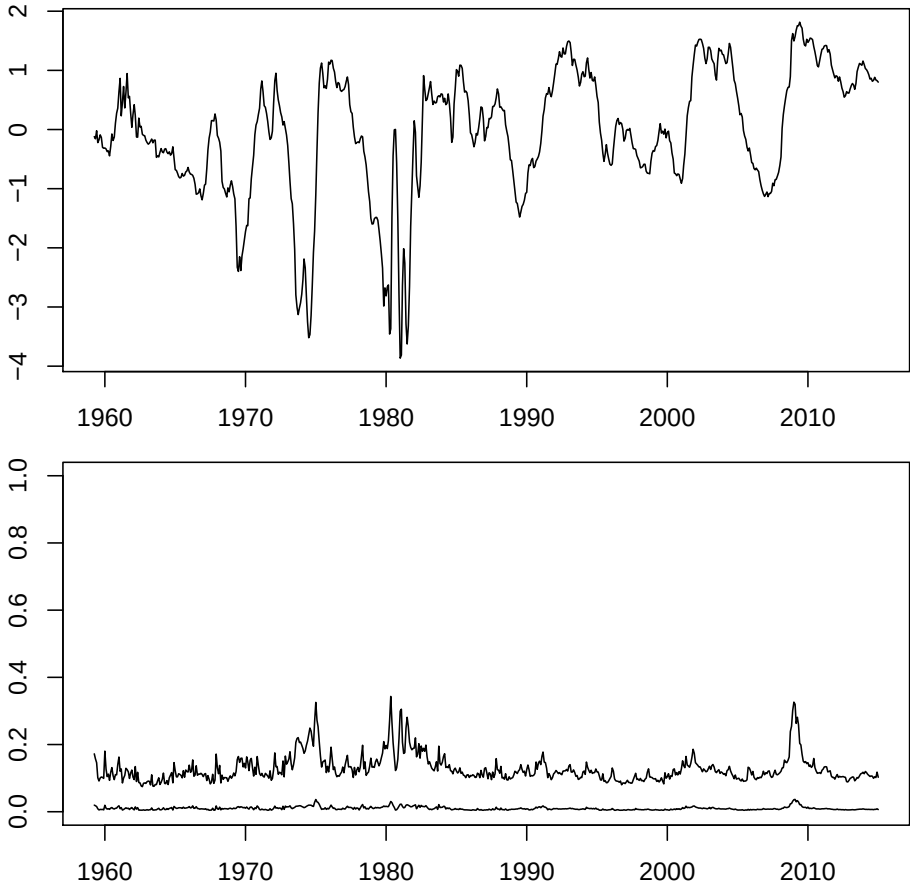


Table 10 shows some summary statistics for the ten currency series. Table 11 show the sample correlations among them.

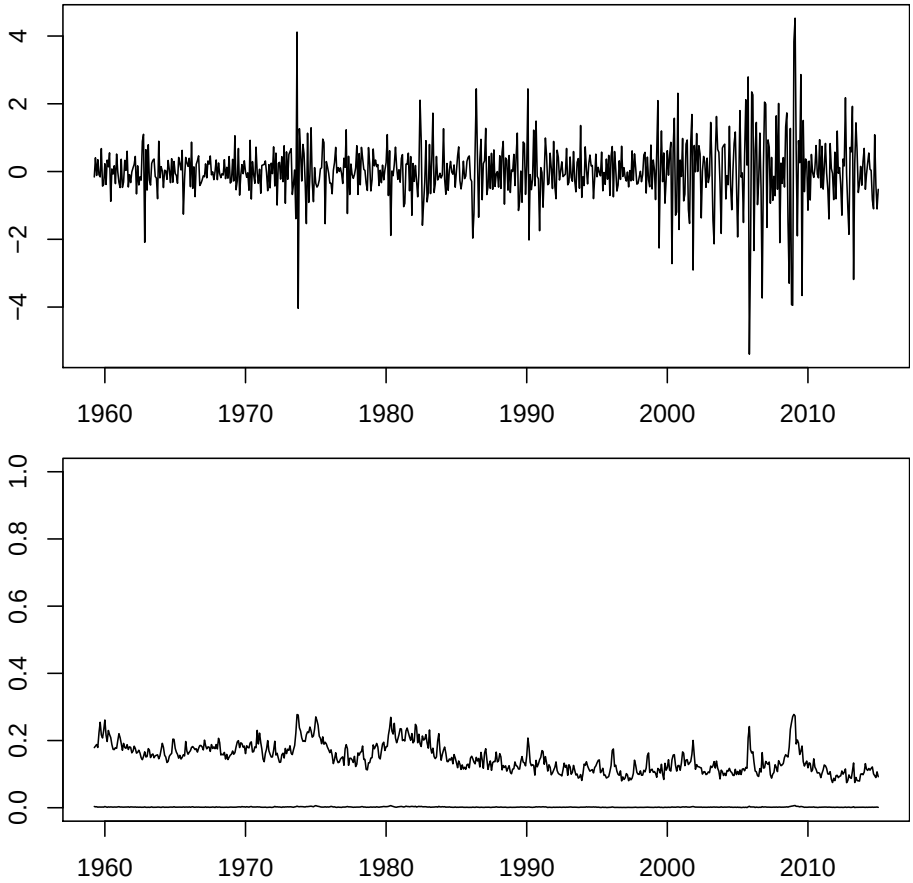
FIGURE 3. FRED panel, mean factor 2. Upper panel: posterior mean of $F_{\mu,t2}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t2}$ and numerical standard error for the posterior mean



	Mean	SD	log variance	squared return autocorrelation
AUD	1.91e-06	9.11e-03	-9.40	0.27
BRL	-1.24e-04	1.02e-02	-9.16	0.40
EUR	-2.99e-05	6.36e-03	-10.12	0.05
JPY	-2.88e-05	6.62e-03	-10.03	0.10
MXN	-1.54e-04	7.16e-03	-9.88	0.54
NZD	1.14e-06	9.22e-03	-9.37	0.12
SGD	9.20e-05	3.57e-03	-11.27	0.11
CHF	1.28e-04	7.52e-03	-9.78	0.13
GBP	-4.90e-05	6.16e-03	-10.18	0.12
CAD	-1.65e-05	6.38e-03	-10.11	0.13

TABLE 10. Summary statistics for Bank of Canada currency panel

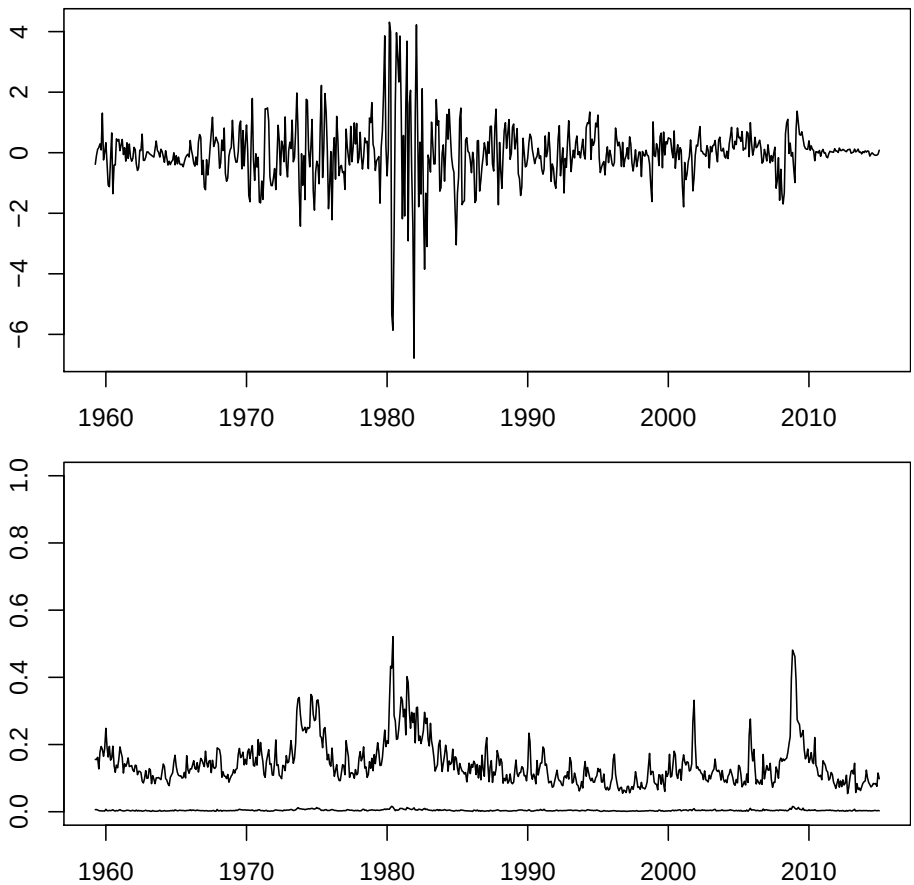
FIGURE 4. FRED panel, mean factor 3. Upper panel: posterior mean of $F_{\mu,t3}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t3}$ and numerical standard error for the posterior mean



	AUD	BRL	EUR	JPY	MXN	NZD	SGD	CHF	GBP	CAD
AUD	1.00	0.59	0.60	-0.07	0.60	0.84	0.65	0.37	0.59	0.68
BRL	0.59	1.00	0.39	-0.15	0.66	0.50	0.46	0.19	0.38	0.50
EUR	0.60	0.39	1.00	0.21	0.39	0.59	0.65	0.70	0.66	0.53
JPY	-0.07	-0.15	0.21	1.00	-0.16	-0.04	0.18	0.36	0.07	-0.08
MXN	0.60	0.66	0.39	-0.16	1.00	0.52	0.50	0.19	0.39	0.53
NZD	0.84	0.50	0.59	-0.04	0.52	1.00	0.61	0.38	0.58	0.62
SGD	0.65	0.46	0.65	0.18	0.50	0.61	1.00	0.47	0.53	0.55
CHF	0.37	0.19	0.70	0.36	0.19	0.38	0.47	1.00	0.45	0.32
GBP	0.59	0.38	0.66	0.07	0.39	0.58	0.53	0.45	1.00	0.52
CAD	0.68	0.50	0.53	-0.08	0.53	0.62	0.55	0.32	0.52	1.00

TABLE 11. Sample correlation matrix for Bank of Canada currency panel

FIGURE 5. FRED panel, mean factor 4. Upper panel: posterior mean of $F_{\mu,t4}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t4}$ and numerical standard error for the posterior mean

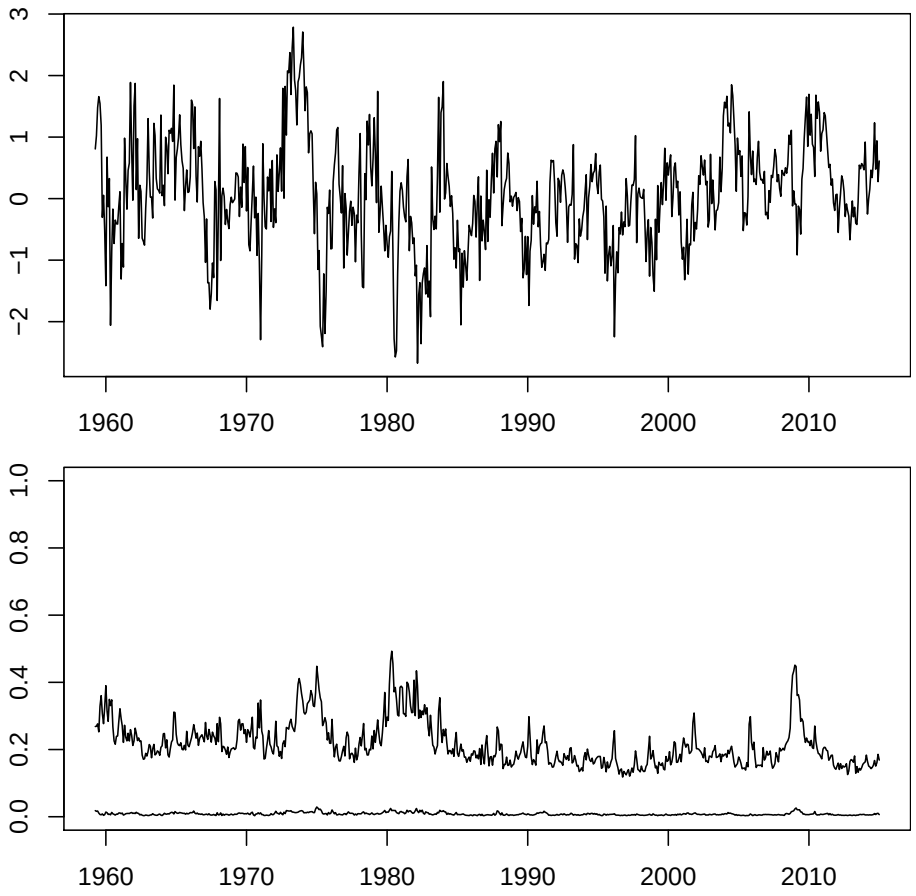


I simulate the posterior distribution for a model with $K_\mu = 2$ mean factors and $K_\sigma = 2$ variance factors. I draw a posterior sample of size 20100, drop the first 100 draws and keep every 10th draw, for a sample of 2000 retained draws.

Tables 12 through 15 show simulation consistent approximations of the posterior mean and standard deviation of various parameter values. They also report numerical standard errors for the posterior mean, which measure simulation noise associated with the reported approximation of the posterior mean. We use the R package coda to compute numerical standard errors, which uses a time series method.

Figures 11 and 12 plot approximations of the posterior mean and standard deviation of the mean factors $F_{\mu,t1}$ and $F_{\mu,t2}$ as a function of the observation time t. They also show

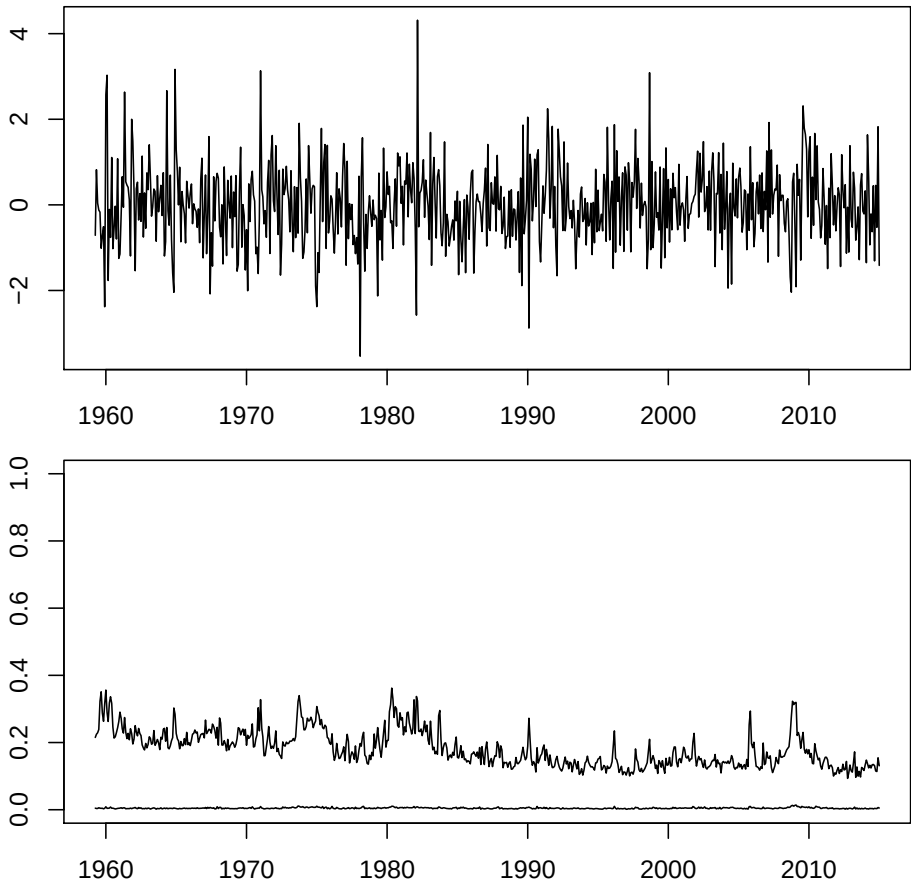
FIGURE 6. FRED panel, mean factor 5. Upper panel: posterior mean of $F_{\mu,t5}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t5}$ and numerical standard error for the posterior mean



the numerical standard error of the posterior mean approximation. Figures 13 and 14 do the same for the variance factors $F_{\sigma,t1}$ and $F_{\sigma,t2}$. Bank of Canada

Not surprising for returns data, the mean factors have low persistence (and therefore low predictability), although the diagonal autoregressive coefficient for the second mean factor is greater than zero with very high posterior probability: its posterior mean, 0.061, while close to zero, is nearly three posterior standard deviations from zero. The factors are identified by the exclusion restriction that the factor loading of the Euro on the second factor is zero. The Euro and the Swiss Franc have particularly high factor loadings on the first factor, and the Mexican Peso and the Brazilian Real have the lowest loadings on this

FIGURE 7. FRED panel, mean factor 6. Upper panel: posterior mean of $F_{\mu,t6}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t6}$ and numerical standard error for the posterior mean

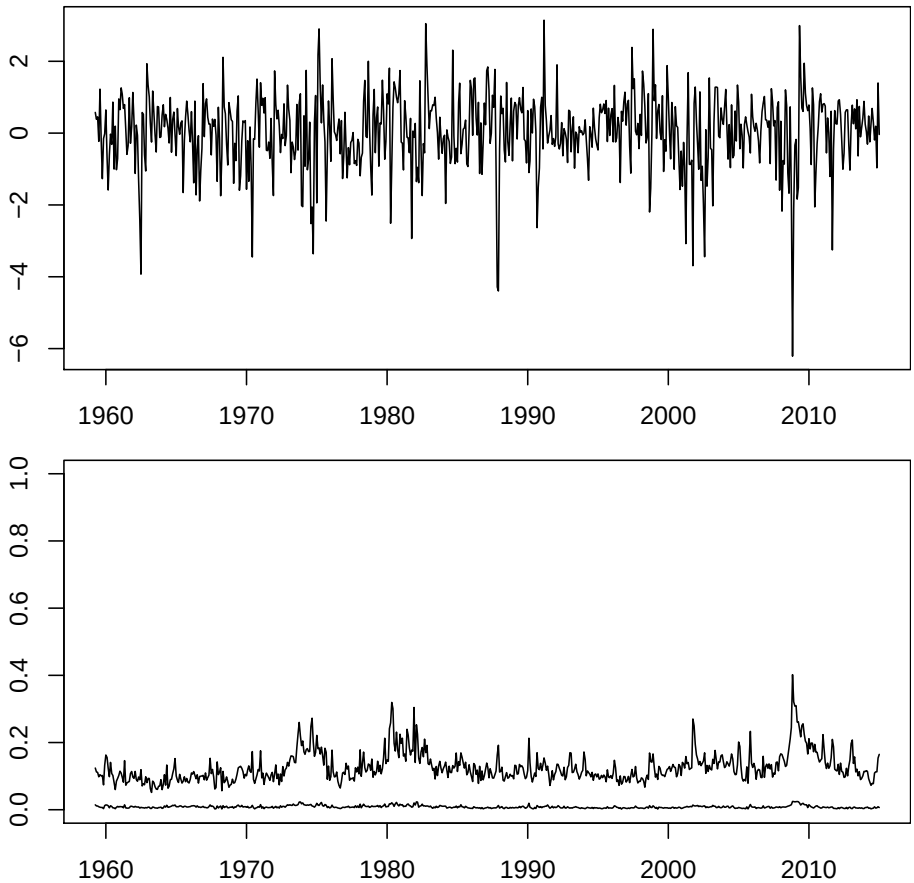


factor. On the second factor, the Australian and New Zealand dollars and the Real have the highest loadings and the Yen has the lowest.

The variance factors are much more persistent, with the second factor being much more persistent than the first. The Australian dollar and Swiss Franc have high loadings on the first variance factor; the Real and Singapore dollar have low loadings. On the second, highly persistent factor, the Real, Singapore dollar and Peso have high loadings and the Pound Sterling and Canadian dollar have low loadings.

The constant terms $B_{\sigma,i}$ in the volatility have posterior means that are much lower than the corresponding log sample variances in Table 10. This is an indication that the factors are capturing much of the variation in returns.

FIGURE 8. FRED panel, mean factor 7. Upper panel: posterior mean of $F_{\mu,t7}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t7}$ and numerical standard error for the posterior mean

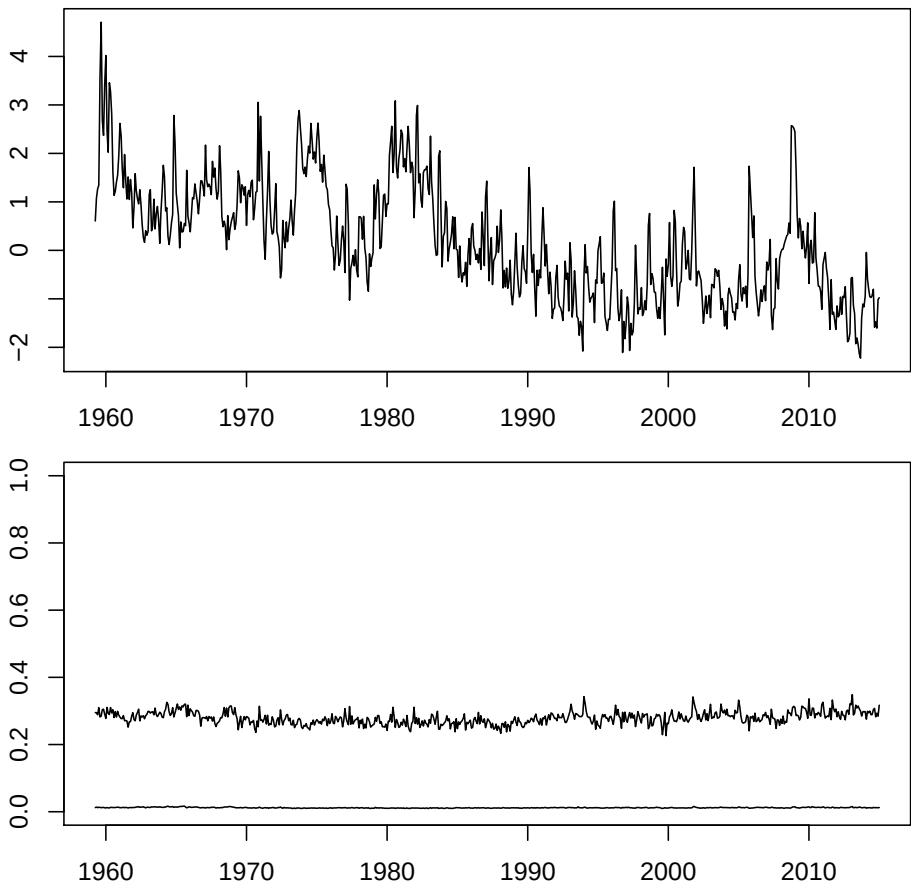


6. CONCLUSIONS

I have provided posterior simulation methods for a factor model with both mean and variance factors. These methods are particularly well suited to data-rich environments, where N is large, because they use block band matrix operations, rather than the Kalman filter, to draw factors, and to integrate them out. The Kalman filter requires solving T systems of N equations; the block band matrix operations involve solving T systems of K equations, where K is the number of factors.

Adding variance factors helps account for the changing volatility of macroeconomic and financial variables, such as seen before, during and after the Great Moderation. In data-rich environments, the dimension reduction achieved using mean factor models has proved

FIGURE 9. FRED panel, variance factor 1. Upper panel: posterior mean of $F_{\sigma,t1}$, against t; Lower panel: posterior standard deviation of $F_{\sigma,t1}$ and numerical standard error for the posterior mean

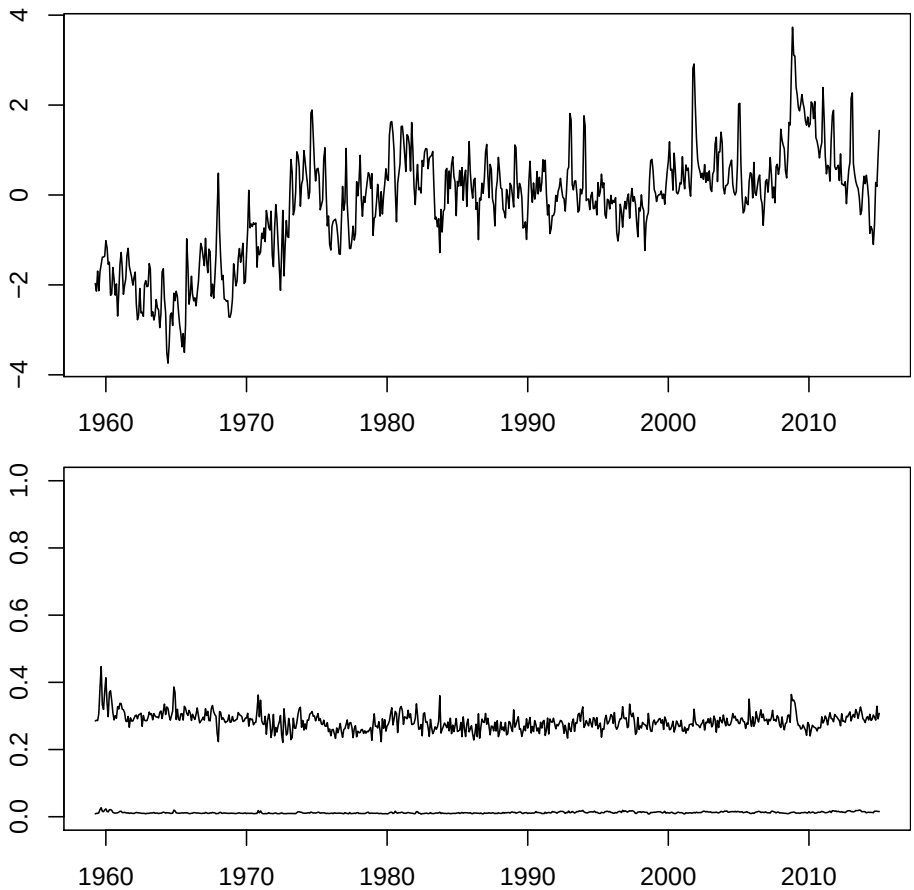


extremely useful. It is natural to do the same thing for variances using a variance factor model, using a small number of factors to account for common variation in idiosyncratic variances across many different series.

My methods do not require variance factors to be *a priori* independent. In the application using the FRED macroeconomic panel, I find strong evidence against the independence of the variance factors; off diagonal elements of the first order autocorrelation matrix of variance factors are positive with very high posterior probability.

Computation for the mean and variance factor models are easily decoupled using Gibbs sampling. In the mean factor model, factors and observed values are jointly Gaussian, with a block band precision matrix, and so block band matrix techniques can be applied fairly easily

FIGURE 10. FRED panel, variance factor 2. Upper panel: posterior mean of $F_{\sigma,t2}$, against t; Lower panel: posterior standard deviation of $F_{\sigma,t2}$ and numerical standard error for the posterior mean



to update the mean factors F_μ. I also use these techniques to integrate out F_μ analytically, and draw the VAR coefficient matrix Φ_μ from its conditional posterior distribution given all unknowns *except* F_μ. While this distribution is not a known distribution, generic methods such as random walk Metropolis or slice sampling are computationally cheap to apply multiple times, with the marginal cost of additional iterations not depending on N. All this allows us to perform joint draws of Φ_μ and F_μ from their joint conditional posterior distribution that are computationally efficient in two ways: computations are linear in N, and the dependence between the old (Φ_μ, F_μ) state and the updated state (Φ'_μ, F'_μ) is weak.

The variance factor model has Gaussian factors, but a non-linear and non-Gaussian measurement equation of high dimension. However, while the conditional posterior distribution

	$E[\Lambda_{\mu,i1}\|y]$	$\sigma[\Lambda_{\mu,i1}\|y]$	$\hat{\sigma}_{\text{nse},i1}$	$E[\Lambda_{\mu,i2}\|y]$	$\sigma[\Lambda_{\mu,i2}\|y]$	$\hat{\sigma}_{\text{nse},i2}$
AUD	7.25e-03	1.41e-04	2.96e-06	0.00e+00	0.00e+00	0.00e+00
BRL	4.91e-03	1.77e-04	3.39e-06	-2.23e-03	1.89e-04	4.79e-06
EUR	3.92e-03	1.19e-04	2.77e-06	2.49e-03	1.59e-04	6.60e-06
JPY	8.35e-04	1.40e-04	2.45e-06	3.07e-03	1.38e-04	2.42e-06
MXN	3.84e-03	1.27e-04	2.57e-06	-1.88e-03	1.39e-04	3.33e-06
NZD	7.21e-03	1.61e-04	3.03e-06	6.43e-05	1.19e-04	1.98e-06
SGD	2.57e-03	6.61e-05	1.36e-06	3.08e-04	5.91e-05	1.41e-06
CHF	3.58e-03	1.29e-04	2.96e-06	3.53e-03	1.56e-04	6.61e-06
GBP	3.53e-03	1.16e-04	2.26e-06	1.40e-03	1.22e-04	3.65e-06
CAD	4.08e-03	1.16e-04	2.22e-06	-6.70e-04	1.12e-04	2.81e-06

TABLE 12. Bank of Canada currency panel, posterior moments of mean factor loadings. The table shows the posterior mean, posterior standard deviation, and the numerical standard error for the mean, for loading on the first ($\Lambda_{\mu,i1}$) and second ($\Lambda_{\mu,i2}$) mean factors.

	$E[\Lambda_{\sigma,i1}\|y]$	$\sigma[\Lambda_{\sigma,i1}\|y]$	$\hat{\sigma}_{\text{nse},i1}$	$E[\Lambda_{\sigma,i2}\|y]$	$\sigma[\Lambda_{\sigma,i2}\|y]$	$\hat{\sigma}_{\text{nse},i2}$
AUD	1.090	0.083	0.0033	0.218	0.070	0.0014
BRL	0.733	0.040	0.0008	-0.217	0.038	0.0007
EUR	0.805	0.053	0.0010	0.714	0.052	0.0012
JPY	0.548	0.039	0.0008	0.362	0.033	0.0006
MXN	0.829	0.049	0.0010	-0.282	0.045	0.0008
NZD	0.476	0.037	0.0007	0.242	0.034	0.0005
SGD	0.598	0.038	0.0008	0.104	0.034	0.0006
CHF	1.107	0.072	0.0016	1.245	0.077	0.0022
GBP	0.687	0.038	0.0008	0.000	0.000	0.0000
CAD	0.569	0.037	0.0007	-0.166	0.032	0.0006

TABLE 13. Bank of Canada currency panel, posterior moments of variance factor loadings. The table shows the posterior mean, posterior standard deviation, and the numerical standard error for the mean, for loading on the first ($\Lambda_{\sigma,i1}$) and second ($\Lambda_{\sigma,i2}$) variance factors.

of factors is not Gaussian, the log density does have a Hessian matrix that is block band diagonal. We update jointly the k'th factor series $F_{\sigma,k}$ and the k'th row of Φ_σ from its conditional posterior distribution using the HESSIAN method of McCausland (2012). The HESSIAN method has several advantages. Unlike auxiliary mixture model approaches, the method is generic, not model specific, and there are no mixture approximations of transformed distributions. The HESSIAN method approximation is close enough that entire latent state sequences can be jointly and efficiently sampled together with their associated parameters.

	$E[B_{\mu,i}\|y]$	$\sigma[B_{\mu,i}\|y]$	$\hat{\sigma}_{\text{nse},\mu}$	$E[B_{\sigma,i}\|y]$	$\sigma[B_{\sigma,i}\|y]$	$\hat{\sigma}_{\text{nse},\sigma}$
AUD	4.91e-05	1.45e-04	3.42e-06	-12.13	0.17	0.008
BRL	3.19e-05	1.61e-04	3.24e-06	-10.22	0.06	0.002
EUR	-8.76e-06	1.10e-04	2.83e-06	-12.50	0.09	0.003
JPY	-1.17e-04	1.11e-04	1.99e-06	-10.55	0.05	0.002
MXN	-9.76e-05	1.17e-04	2.29e-06	-11.22	0.09	0.003
NZD	5.07e-05	1.60e-04	3.70e-06	-10.75	0.05	0.002
SGD	8.69e-05	6.38e-05	1.53e-06	-12.37	0.05	0.002
CHF	9.12e-06	1.17e-04	3.14e-06	-12.58	0.12	0.005
GBP	3.78e-05	1.03e-04	2.47e-06	-11.13	0.06	0.002
CAD	-3.56e-05	1.08e-04	2.23e-06	-11.10	0.05	0.002

TABLE 14. Bank of Canada currency panel, posterior moments of mean and variance constants. The table shows the posterior mean, posterior standard deviation, and the numerical standard error for the mean constant ($B_{\mu,i1}$) and variance constant ($B_{\sigma,i1}$).

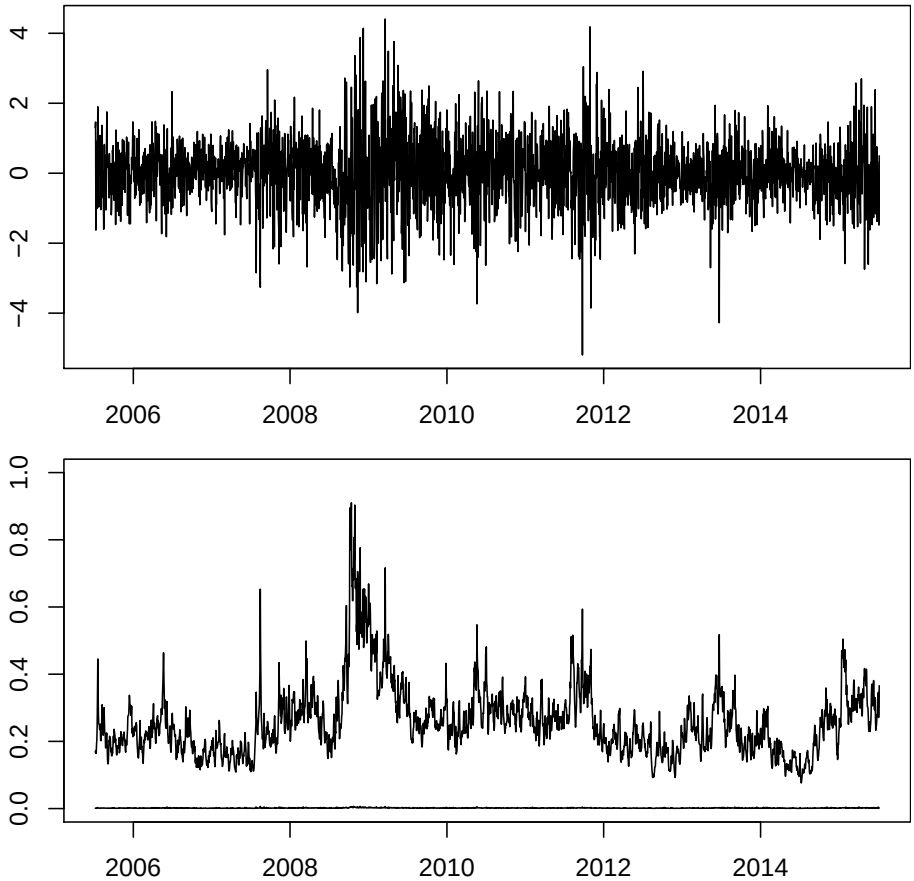
	$E[\Phi_{ij}\|y]$	$\sigma[\Phi_{ij}\|y]$	$\hat{\sigma}_{\text{nse}}$
$\Phi_{\mu,11}$	0.004	0.021	0.0003
$\Phi_{\mu,12}$	-0.056	0.024	0.0004
$\Phi_{\mu,21}$	-0.029	0.024	0.0004
$\Phi_{\mu,22}$	0.007	0.025	0.0004
$\Phi_{\sigma,11}$	0.852	0.017	0.0005
$\Phi_{\sigma,12}$	0.143	0.019	0.0004
$\Phi_{\sigma,21}$	0.129	0.019	0.0004
$\Phi_{\sigma,22}$	0.201	0.018	0.0003

TABLE 15. Bank of Canada currency panel, posterior moments of mean and variance autoregressive coefficients. The table shows the posterior mean, posterior standard deviation, and the numerical standard error for the mean factor autoregressive coefficients ($\Phi_{\mu,kl}$) and variance factor autoregressive coefficients ($\Phi_{\sigma,kl}$).

I have used the "Getting it Right" procedure of Geweke (2004) to demonstrate the conceptual soundness and correct implementation of my posterior simulation methods. I have applied the methods to a foreign exchange panel of ten currencies and 2503 daily observations, and a macroeconomic panel of 134 indicators and 670 monthly observations.

Several extensions of the methods described here to more flexible models are possible. Many features could be added to the mean factor model to better complement the variance factor model. We have seen that the Gibbs sampling approach decouples the computational aspects associated with the mean and variance factor models, so it would be relatively easy to slot in alternative mean factor models. Promising extensions include factors with time-varying variance, factors that are higher order vector autoregressions, factor loadings with

FIGURE 11. Bank of Canada currency panel, mean factor 1. Upper panel: posterior mean of $F_{\mu,t1}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t1}$ and numerical standard error for the posterior mean

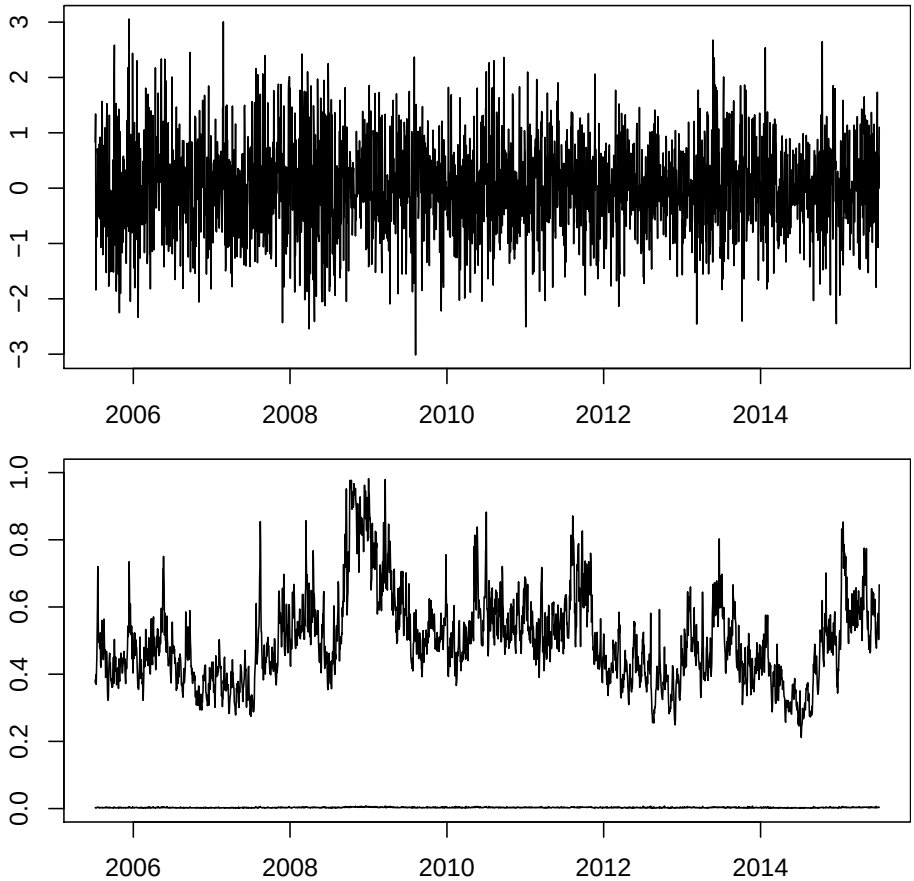


parsimonious structure or hierarchical factors, and factor loadings on lagged factors. It would also be possible to model mean and variance factors as a joint vector autoregression.

APPENDIX A. MISSING DATA, FACTOR FOUNDERS AND INITIAL VALUES

A.1. **Missing data.** In many applications, including the macro panel application described below, one does not observe all the series in exactly the same periods. Fortunately it is straightforward to accommodate unbalanced panels; the y_{ti} are conditionally independent given unknown factors and parameters, and so we can simply delete likelihood factors for missing data. The sums in (6), (7), (8), (9) should run for all t in which y_{ti} is observed. The sums in (15), (16), (17) should run for all i for which y_{ti} is observed in period t.

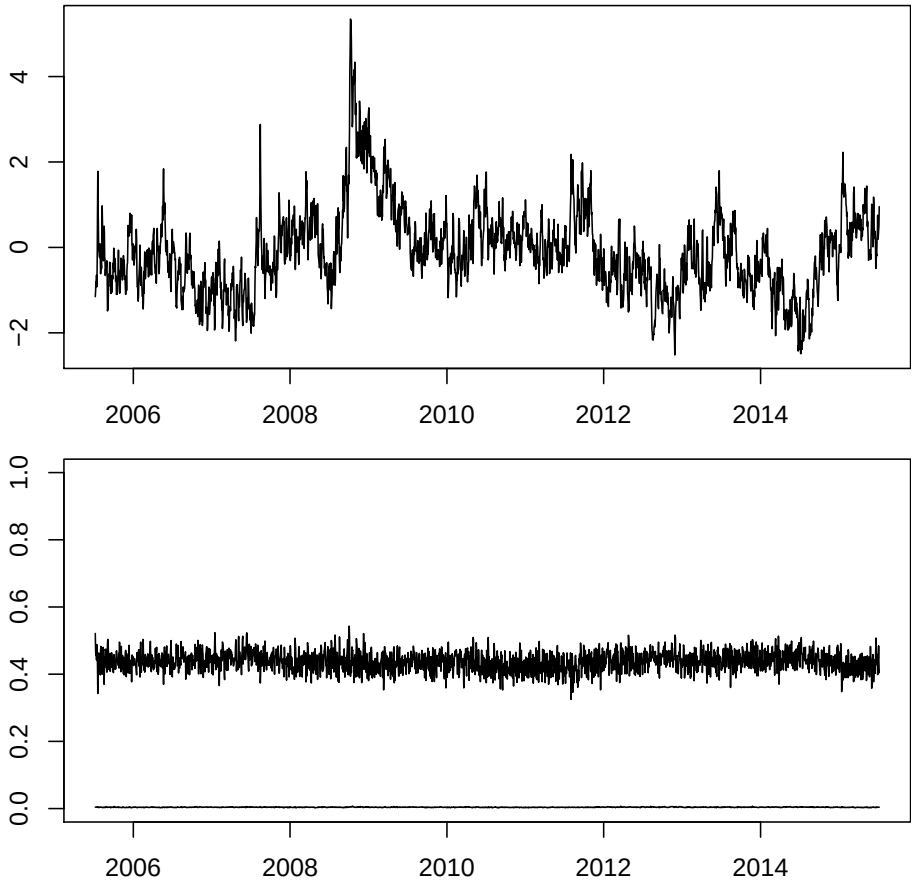
FIGURE 12. Bank of Canada currency panel, mean factor 2. Upper panel: posterior mean of $F_{\mu,t2}$, against t; Lower panel: posterior standard deviation of $F_{\mu,t2}$ and numerical standard error for the posterior mean



A.2. **Selection of factor founders.** I selected factor founders by trial and error with the objective that factors would resemble the static factors obtained in a principal components analysis. I selected the mean factor founders in order; I chose the i'th mean factor founder by choosing the index k, among the series not already chosen as mean factor founders, minimizing $(\sum_{j=i+1}^{K_\mu} L_{kj}^2)/|L_{ki}|^{15}$, where L is the static loading matrix obtained in a principal components analysis. I selected the two variance factor founders by unsystematic trial and error.

A.3. **Initial values.** To avoid slow convergence of the distribution of MCMC draws to the posterior target distribution, I choose initial values in the following way.

FIGURE 13. Bank of Canada currency panel, variance factor 1. Upper panel: posterior mean of $F_{\sigma,t1}$, against t; Lower panel: posterior standard deviation of $F_{\sigma,t1}$ and numerical standard error for the posterior mean



For B_μ, I use Ordinary Least Squares (OLS) to compute each row $B_{\mu,i}$, $i = 1, \ldots, N$, then compute $\tilde{Y}$, the matrix of OLS residuals, organized as a $T \times N$ matrix:

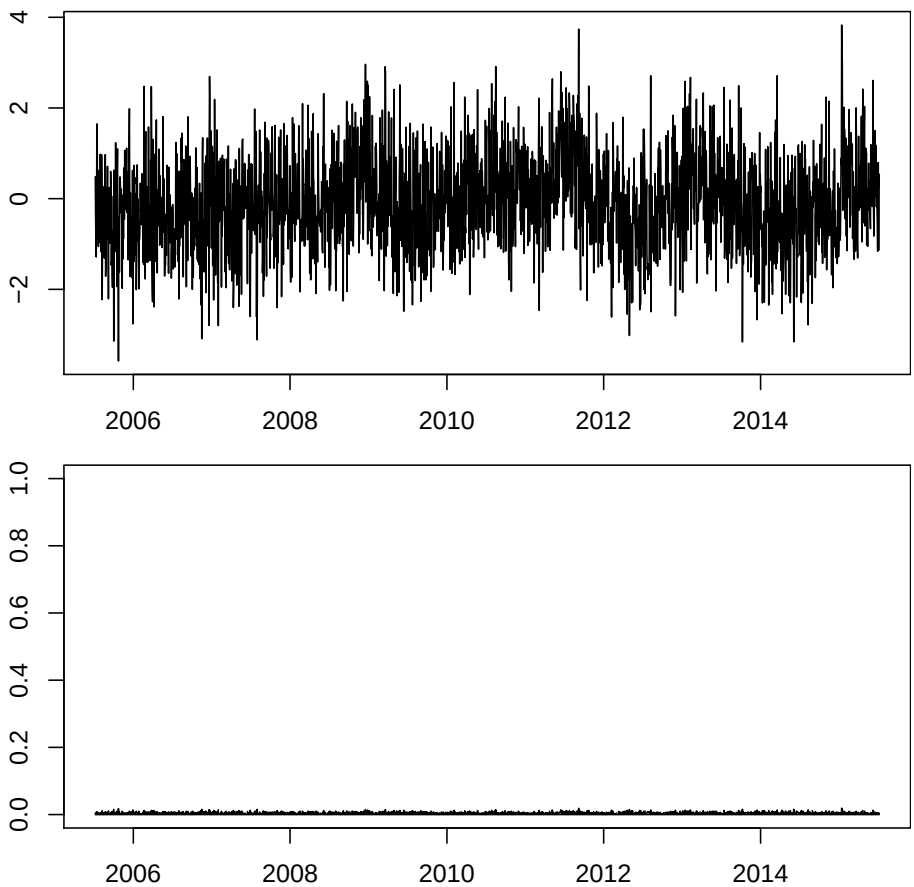
$$B_{\mu,i}^\top = \left(\sum_{t=1}^{T} x_t x_t^\top \right)^{-1} \sum_{t=1}^{T} x_t y_{ti}, \qquad \tilde{Y}_{ti} = y_{ti} - B_{\mu,i} x_t.$$

Then I compute the sample variance of each residual series and construct the $N \times N$ diagonal matrix D whose diagonal elements are these sample variances. Thus the N columns of $D^{-1/2}\tilde{Y}$ each have zero sample mean and unit sample variance.

For Λ_μ and F_μ, I first compute the usual "thin" version of the singular value decomposition of $D^{-1/2}\tilde{Y}$:

$$D^{-1/2}\tilde{Y} = USV^\top,$$

FIGURE 14. Bank of Canada currency panel, variance factor 2. Upper panel: posterior mean of $F_{\sigma,t2}$, against t; Lower panel: posterior standard deviation of $F_{\sigma,t2}$ and numerical standard error for the posterior mean



where U is $T \times N$ and has orthogonal columns; S is $N \times N$ and diagonal; and V is $N \times N$ and orthogonal. The singular values on the diagonal of S are arranged in decreasing order. Partition

$$U = \begin{bmatrix} U_{11} & U_{12} \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where U_{11} is $T \times K_\mu$, and S_{11} and V_{11} are $K_\mu \times K_\mu$. Setting $U_{12} = 0$, $S_{22} = 0$, $V_{12} = 0$ and $V_{22} = 0$ gives the approximation

$$\tilde{Y} \approx U_{12} S_{11} \begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix}^\top . \tag{18}$$

Next, I compute the QR decomposition of $S_{11}[V_{11}^\top V_{21}^\top]$, which will allow us to rotate $S_{11}[V_{11}^\top V_{21}^\top]$ so that it becomes right triangular. The decomposition is

$$S_{11}\begin{bmatrix} V_{11} \\ V_{21} \end{bmatrix}^\top = QR,$$

where Q is $K_\mu \times K_\mu$ and orthogonal and R is $K_\mu \times N$ and right triangular. Then write the approximation in (18) as

$$\tilde{Y} \approx U_{12}QR = (T^{1/2}U_{12}Q) \cdot (T^{-1/2}R).$$

The initial values of F_μ and Λ_μ are $F_\mu = T^{1/2}\,\mathrm{vec}((U_{12}Q)^\top)$, and $\Lambda_\mu = T^{-1/2}R^\top$.

I set Φ_μ to the first order sample autocovariance of the initial value of F_μ:

$$\Phi_\mu = \frac{1}{T}\sum_{t=2}^{T} F_{\mu,t}F_{\mu,t-1}^\top.$$

Since the initial value of $F_\mu^\top F_\mu$ equals I_{K_μ} by construction, the sample autocovariance is also the sample autocorrelation.

For B_σ, I first compute the $T \times N$ residual matrix $E = \tilde{Y} - U_{12}QR$. Then I set $B_{\sigma,i1} = \ln \frac{1}{T}\sum_{t=1}^{T} E_{ti}^2$, $i = 1, \ldots, N$. Note that in the empirical applications of this paper, there is only an intercept and B_σ is a column vector.

For Λ_σ, F_σ, I first compute the smoothed squared residual series σ_{ti}^2 defined by the following GARCH(1,1)—see Bollerslev (1986)—smoothing of squared residuals E_{ti}^2:

$$\sigma_{1i} = \exp(B_{i1}), \quad \sigma_{ti} = \omega + \alpha E_{ti}^2 + \beta\sigma_{t-1,i},$$

with $\alpha = 0.2$, $\beta = 0.6$, $\omega = \exp(B_{i1})(1 - \alpha - \beta)$. Then I construct the $T \times N$ matrix $[\ln \sigma_{ti}^2 - \ln\exp(B_{i1})]$ and compute the initial factor matrix F_σ and initial factor loading matrix Λ_σ using the same procedure used to compute initial values of F_μ and Λ_μ above: $[\ln \sigma_{ti}^2 - \ln\exp(B_{i1})]$, K_σ, F_σ and Λ_σ replace $\tilde{Y}$, K_μ, F_μ and Λ_μ.

I set Φ_σ to the first order sample autocorrelation of the initial value of F_σ:

$$\Phi_\sigma = \frac{1}{T}\sum_{t=2}^{T} F_{\sigma,t}F_{\sigma,t-1}^\top.$$

Appendix B. Details on drawing Φ_μ and F_μ

Define the stacked mean factor vector $F_\mu = (F_{\mu,1}^\top, \ldots, F_{\mu,T}^\top)^\top$ and the stacked measurement equation shocks $\epsilon = (\epsilon_1^\top, \ldots, \epsilon_T^\top)^\top$.

Since the variance of $F_{\mu t}$ is I and the variance of the mean factor innovation u_t is $I - \Phi_\mu \Phi_\mu^\top$, I can write $\mathbf{\Phi} F_\mu \sim N(0, \mathbf{\Omega})$, where

$$\mathbf{\Phi} \equiv \begin{bmatrix} I & & & & \\ -\Phi_\mu & I & & & \\ & -\Phi_\mu & \ddots & & \\ & & \ddots & I & \\ & & & -\Phi_\mu & I \end{bmatrix} \quad \text{and} \quad \mathbf{\Omega} \equiv \begin{bmatrix} I & & & \\ & I - \Phi_\mu \Phi_\mu^\top & & \\ & & \ddots & \\ & & & I - \Phi_\mu \Phi_\mu^\top \end{bmatrix}.$$

Therefore $F_\mu | \Phi \sim N(0, \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top})$.

Now define $\tilde{y}_t \equiv y_t - Bx_t = \Lambda_\mu F_t + \epsilon_t$ and let $\tilde{y} = (\tilde{y}_1^\top, \ldots, \tilde{y}_T^\top)$. The variance of $\tilde{y}$ is $\mathrm{diag}(\sigma_1^\top, \ldots, \sigma_T^\top)^\top$, which I denote $\mathbf{D}$.

Since F_μ and $\tilde{y}$ are jointly Gaussian, and since

$$\begin{bmatrix} F_\mu \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathbf{\Lambda} & I \end{bmatrix} \begin{bmatrix} F_\mu \\ \epsilon \end{bmatrix},$$

where $\mathbf{\Lambda} = I_T \otimes \Lambda_\mu$, I can write the variance of the joint Gaussian distribution as

$$\mathrm{Var}\left(\begin{bmatrix} F_\mu \\ \tilde{y} \end{bmatrix} \middle| \Lambda_\mu, \Phi_\mu \right) = \begin{bmatrix} I & 0 \\ \mathbf{\Lambda} & I \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top} & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & \mathbf{\Lambda}^\top \\ 0 & I \end{bmatrix}.$$

WJM: next equation for my benefit, hide from paper

$$\mathrm{Var}\left(\begin{bmatrix} F_\mu \\ \tilde{y} \end{bmatrix} \middle| \Lambda_\mu, \Phi_\mu \right) = \begin{bmatrix} \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top} & \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top} \mathbf{\Lambda}^\top \\ \mathbf{\Lambda} \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top} & \mathbf{\Lambda} \mathbf{\Phi}^{-1} \mathbf{\Omega} \mathbf{\Phi}^{-\top} \mathbf{\Lambda}^\top + D \end{bmatrix}$$

The precision matrix for the same distribution is the block band matrix

$$\mathbf{H} \equiv \mathrm{Var}\left(\begin{bmatrix} F_\mu \\ \tilde{y} \end{bmatrix} \middle| \Lambda_\mu, \Phi_\mu \right)^{-1} = \begin{bmatrix} I & -\mathbf{\Lambda}^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}^\top \mathbf{\Omega}^{-1} \mathbf{\Phi} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\mathbf{\Lambda} & I \end{bmatrix} \tag{19}$$

$$= \begin{bmatrix} \mathbf{\Phi}^\top \mathbf{\Omega}^{-1} \mathbf{\Phi} + \mathbf{\Lambda}^\top D^{-1} \mathbf{\Lambda} & -\mathbf{\Lambda}^\top D^{-1} \\ -D^{-1} \mathbf{\Lambda} & D^{-1} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{H}_{FF} & \mathbf{H}_{Fy} \\ \mathbf{H}_{yF} & \mathbf{H}_{yy} \end{bmatrix},$$

where the last equation defines the partition blocks $\mathbf{H}_{FF}$, $\mathbf{H}_{Fy}$, $\mathbf{H}_{yF}$ and $\mathbf{H}_{yy}$.

Then the conditional mean and variance of $\tilde{y}$ given Λ_μ and Φ_μ are, respectively, zero and $(\mathbf{H}_{yy} - \mathbf{H}_{yF}\mathbf{H}_{FF}^{-1}\mathbf{H}_{Fy})^{-1}$. We can write

$$f(\tilde{y}|\Lambda_\mu, \Phi_\mu) = \frac{1}{(2\pi)^{NT/2}} \frac{|\mathbf{H}|^{1/2}}{|\mathbf{H}_{FF}|^{1/2}} \exp\left[-\frac{1}{2}\tilde{y}^\top(\mathbf{H}_{yy} - \mathbf{H}_{yF}\mathbf{H}_{FF}^{-1}\mathbf{H}_{Fy})\tilde{y}\right],$$

where I use the result $|\mathbf{H}_{yy} - \mathbf{H}_{yF}\mathbf{H}_{FF}^{-1}\mathbf{H}_{Fy}| = |\mathbf{H}|/|\mathbf{H}_{FF}|$, which follows directly from Proposition 2.30 of in Dhrymes (2000).

We can also use standard formulas for conditional Gaussian distributions to derive the conditional distribution of F_μ given $\mathbf{\Phi}$:

$$F_\mu|\tilde{y}, \Lambda_\mu, \Phi_\mu \sim N(-\mathbf{H}_{FF}^{-1}\mathbf{H}_{Fy}\tilde{y}, \mathbf{H}_{FF}^{-1}).$$

In the rest of this appendix, I discuss some implementation details that make computations particularly efficient.

B.1. **Computing** $\log f(\tilde{y}|\Lambda_\mu, \Phi_\mu)$**.** For our purposes—evaluating the Hastings ratio, equation (10)—we only need to evaluate $\log f(\tilde{y}|\Lambda_\mu, \Phi_\mu)$ up to an additive term not depending on Φ_μ. We can write

$$\log f(\tilde{y}|\Lambda_\mu, \Phi_\mu) = k + \frac{1}{2}\left[\log|\mathbf{H}| - \log|\mathbf{H}_{FF}| - ||\mathbf{L}_{FF}^{-1}\mathbf{H}_{Fy}\tilde{y}||^2\right],$$

where k does not depend on Φ_μ, and $\mathbf{L}_{FF}$ is the lower Cholesky factor of $\mathbf{H}_{FF}$. Appendix D describes how to compute $\mathbf{L}_{FF}$ efficiently. The quantities $-\mathbf{\Lambda}^\top\mathbf{D}^{-1}$ and $\mathbf{\Lambda}^\top\mathbf{D}^{-1}\mathbf{\Lambda}$ are products of block band matrices and can be computed and stored efficiently using well known block band matrix operations and representations. See for example, Golub and Van Loan (1996), Sections 1.3, 4.3, 4.5. Furthermore, they do not depend on Φ_μ, so they can be precomputed and used for multiple evaluations of $\log f(\tilde{y}|\Lambda_\mu, \Phi_\mu)$ at different values of Φ_μ.

The determinants are easily and cheaply computed. From the representation of $\mathbf{H}$ in (19) as a product of three matrices, it is easy to see that

$$|\mathbf{H}| = |\mathbf{\Omega}|^{-1}|D|^{-1} = |I - \Phi_\mu\Phi_\mu^\top|^{-(T-1)}\prod_{t=1}^{T}\prod_{i=1}^{N}\sigma_{ti}^{-2}.$$

Now $|\mathbf{H}_{FF}|^{1/2} = |\mathbf{L}_{FF}|$ and since $\mathbf{L}_{FF}$ is lower triangular, its determinant is simply the product of its diagonal elements.

Using the precomputed $\mathbf{H}_{Fy}\tilde{y}$, the vector $\mathbf{L}_{FF}^{-1}[\mathbf{H}_{Fy}\tilde{y}]$ is efficiently computed using block band forward substitution, as described in Appendix D.

B.2. **Drawing** $F_\mu|\Lambda_\mu, \Phi_\mu, \tilde{y}$**.** To draw $F_\mu|\Lambda_\mu, \Phi_\mu, \tilde{y}$, draw a $K_\mu T \times 1$ vector $u \sim N(0, I)$ and compute $F_\mu = -\mathbf{L}_{FF}^{-\top}(\mathbf{L}_{FF}^{-1}\mathbf{H}_{Fy}\tilde{y} + u)$. The mean of the result is $-\mathbf{H}_{FF}^{-\top}\mathbf{H}_{Fy}\tilde{y}$ and the variance is $\mathbf{L}_{FF}^{-\top}\mathbf{L}_{FF}^{-1} = \mathbf{H}_{FF}^{-1}$, as required. Note that the vector $\mathbf{L}_{FF}^{-1}\mathbf{H}_{Fy}\tilde{y}$ is also part of the computation of $\log f(\tilde{y}|\mathbf{\Phi})$, and so its value will be available whenever we need to draw F_μ. Premultiplication by $\mathbf{L}_{FF}^{-\top}$ is efficiently performed by block band backward substitution, also described in Appendix D.

Appendix C. Details on drawing Φ

This appendix gives details on how I draw Φ_μ and Φ_σ from their conditional posterior distributions. The description below applies equally well for $(\Phi, F) = (\Phi_\mu, F_\mu)$ and $(\Phi, F) = (\Phi_\sigma, F_\sigma)$, and so I omit the subscript.

We can write the conditional density of F given Φ as

$$f(F|\Phi) = \frac{|I - \Phi\Phi^\top|^{-(T-1)/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{1}{2}\left[F_1^\top F_1 + \sum_{t=2}^{T}(F_t - \Phi F_{t-1})^\top (I - \Phi\Phi^\top)^{-1}(F_t - \Phi F_{t-1})\right]\right\}.$$

Then the log conditional posterior density of Φ can be expressed as

$$\begin{aligned}\ln f(\Phi|F) = k + \ln f(\Phi) - \frac{T-1}{2}\ln|I - \Phi\Phi^\top| - \frac{1}{2}\operatorname{tr}\left[(I - \Phi\Phi^\top)^{-1}\sum_{t=2}^{T} F_t F_t^\top\right] \\ + \operatorname{tr}\left[(I - \Phi\Phi^\top)^{-1}\Phi\sum_{t=2}^{T-1} F_{t-1}F_t^\top\right] - \frac{1}{2}\operatorname{tr}\left[\Phi^\top(I - \Phi\Phi^\top)^{-1}\Phi\sum_{t=1}^{T-1} F_t F_t^\top\right],\end{aligned} \tag{20}$$

where k does not depend on Φ. The sufficient statistics $\sum_{t=2}^{T} F_t F_t^\top$, $\sum_{t=2}^{T-1} F_{t-1}F_t^\top$ and $\sum_{t=1}^{T-1} F_t F_t^\top$ are computed once. Each new proposal of Φ requires evaluating the log density in (20) at that value, a computation that does not depend on N or T.

As I do when I update Φ with F integrated out, I use a sequence of random walk Metropolis steps. I update each column Φ_k multiple times, using a proposal $\Phi_k^* \sim \mathrm{N}(\Phi_k, \Omega_k)$. Here, the proposal Φ^* is accepted with probability

$$\min\left[\frac{f(\Phi^*|F)}{f(\Phi|F)}\right].$$

Let Φ_k be the k'th column of Φ and $\Phi_{-k} \equiv \Phi - \Phi_k$. Thus Φ_{-k} is Φ with its k'th column set to zero. Let $\Sigma_{-k} = I - \Phi_{-k}\Phi_{-k}^\top$ and note that $\Sigma \equiv (I - \Phi\Phi^\top) = \Sigma_{-k} - \Phi_k\Phi_k^\top$. Now write the period t innovation to the factor as

$$F_t - \Phi F_{t-1} = F_t - \Phi_{-k}F_{t-1} - \Phi_k F_{t-1,k} = v_t - F_{t-1,k}\Phi_k,$$

where $v_t \equiv F_t - \Phi_{-k}F_{t-1}$. By the matrix determinant lemma,

$$|\Sigma| = |\Sigma_{-k} - \Phi_k\Phi_k^\top| = (1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)|\Sigma_{-k}|.$$

This allows us to write the determinant term of (20) as

$$-\frac{T-1}{2}\ln|\Sigma| = -\frac{T-1}{2}[\ln(1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k) + \ln|\Sigma_{-k}|]$$

The gradient and Hessian of this term are

$$\frac{\partial}{\partial\Phi_k^\top}\left[-\frac{T-1}{2}\ln|\Sigma|\right] = (T-1)\frac{\Sigma_{-k}^{-1}\Phi_k}{1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k},$$

$$\frac{\partial^2}{\partial\Phi_k\partial\Phi_k^\top}\left[-\frac{T-1}{2}\ln|\Sigma|\right] = (T-1)\left[\frac{\Sigma_{-k}^{-1}}{1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k} + \frac{2\Sigma_{-k}^{-1}\Phi_k\Phi_k^\top\Sigma_{-k}^{-1}}{(1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)^2}\right].$$

Now I compute the gradient and Hessian of the sum-of-squares term. By the Woodbury matrix identity,

$$(\Sigma_{-k} - \Phi_k\Phi_k^\top)^{-1} = \Sigma_{-k}^{-1} + \Sigma_{-k}^{-1}\Phi_k(1 - \Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)^{-1}\Phi_k^\top\Sigma_{-k}^{-1}$$

Then each quadratic term can be written

$$
\begin{aligned}
&(F_t - \Phi F_{t-1})^\top \Sigma^{-1} (F_t - \Phi F_{t-1}) \\
&\quad = (v_t - F_{t-1,k}\Phi_k)^\top (\Sigma_{-k} - \Phi_k \Phi_k^\top)^{-1} (v_t - F_{t-1,k}\Phi_k) \\
&\quad = v_t^\top \Sigma_{-k}^{-1} v_t + \frac{v_t^\top \Sigma_{-k}^{-1} \Phi_k \Phi_k^\top \Sigma_{-k}^{-1} v_t}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} + F_{t-1,k}^2 \left[\Phi_k^\top \Sigma_{-k}^{-1} \Phi_k + \frac{(\Phi_k^\top \Sigma_{-k}^{-1} \Phi_k)^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} \right] \\
&\qquad - 2F_{t-1,k} \left[\Phi_k^\top \Sigma_{-k}^{-1} v_t + \frac{\Phi_k^\top \Sigma_{-k}^{-1} \Phi_k \Phi_k^\top \Sigma_{-k}^{-1} v_t}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} \right] \\
&\quad = v_t^\top \Sigma_{-k}^{-1} v_t + \frac{(v_t^\top \Sigma_{-k}^{-1} \Phi_k)^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} + F_{t-1,k}^2 \frac{\Phi_k^\top \Sigma_{-k}^{-1} \Phi_k}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} - 2F_{t-1,k} \frac{\Phi_k^\top \Sigma_{-k}^{-1} v_t}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} \\
&\quad = v_t^\top \Sigma_{-k}^{-1} v_t + \frac{(v_t^\top \Sigma_{-k}^{-1} \Phi_k)^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} + F_{t-1,k}^2 \left(\frac{1}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} - 1 \right) - 2F_{t-1,k} \frac{\Phi_k^\top \Sigma_{-k}^{-1} v_t}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} \\
&\quad = \frac{(v_t^\top \Sigma_{-k}^{-1} \Phi_k - F_{t-1,k})^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} + \text{terms not depending on } \Phi_k.
\end{aligned}
$$

Now write the full sum-of-squares term as

$$
-\frac{1}{2} \sum_{t=2}^{T} \frac{(v_t^\top \Sigma_{-k}^{-1} \Phi_k - F_{t-1,k})^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} = -\frac{1}{2} \frac{\Phi_k^\top A \Phi_k - 2b^\top \Phi_k + c}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k},
$$

where $A \equiv \Sigma_{-k}^{-1} \sum_{t=2}^{T} v_t v_t^\top \Sigma_{-k}^{-1}$, $b^\top \equiv \sum_{t=2}^{T} F_{t-1,k} v_t^\top \Sigma_{-k}^{-1}$, and $c \equiv \sum_{t=2}^{T} F_{t-1,k}^2$.

$$
\sum_{t=2}^{T} v_t v_t^\top = \sum_{t=2}^{T} F_t F_t^\top - \Phi_{-k} \sum_{t=2}^{T} F_{t-1} F_t^\top - \left(\sum_{t=2}^{T} F_{t-1} F_t^\top \right)^\top \Phi_{-k}^\top + \Phi_{-k} \left(\sum_{t=2}^{T} F_{t-1} F_{t-1}^\top \right) \Phi_{-k}^\top.
$$

$$
\sum_{t=2}^{T} F_{t-1,k} v_t^\top = \sum_{t=2}^{T} F_{t-1,k} F_t^\top - \left(\sum_{t=2}^{T} F_{t-1,k} F_{t-1}^\top \right) \Phi_{-k}^\top
$$

$$
\sum_{t=2}^{T} F_{t-1,k}^2 = \left(\sum_{t=2}^{T} F_{t-1} F_{t-1}^\top \right)_{kk}.
$$

Its gradient and Hessian are

$$
\frac{\partial}{\partial \Phi_k^\top} \left[-\frac{1}{2} \sum_{t=2}^{T} \frac{(v_t^\top \Sigma_{-k}^{-1} \Phi_k - F_{t-1,k})^2}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} \right] = -\frac{A\Phi_k - b}{1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k} - \frac{(\Phi_k^\top A \Phi_k - 2b^\top \Phi_k + c) \Sigma_{-k}^{-1} \Phi_k}{(1 - \Phi_k^\top \Sigma_{-k}^{-1} \Phi_k)^2}.
$$

$$\frac{\partial^2}{\partial\Phi_k\partial\Phi_k^\top}\left[-\frac{1}{2}\sum_{t=2}^{T}\frac{(v_t^\top\Sigma_{-k}^{-1}\Phi_k - F_{t-1,k})^2}{1-\Phi_k^\top\Sigma_{-k}^{-1}\Phi_k}\right] = -\frac{A}{1-\Phi_k^\top\Sigma_{-k}^{-1}\Phi_k}$$
$$-\frac{(\Phi_k^\top A\Phi_k - 2b^\top\Phi_k + c)\Sigma_{-k}^{-1}}{(1-\Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)^2}$$
$$-2\frac{(A\Phi_k - b)\Phi_k^\top\Sigma_{-k}^{-1} + \Sigma_{-k}^{-1}\Phi_k(A\Phi_k - b)^\top}{(1-\Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)^2}$$
$$-4\frac{(\Phi_k^\top A\Phi_k - 2b^\top\Phi_k + c)\Sigma_{-k}^{-1}\Phi_k\Phi_k^\top\Sigma_{-k}^{-1}}{(1-\Phi_k^\top\Sigma_{-k}^{-1}\Phi_k)^3}.$$

Appendix D. Cholesky decomposition, forward and backward substitution for block tridiagonal systems

Let $\mathbf{H}$ be a symmetric positive definite tridiagonal block band matrix with $T\times T$ blocks, each block of dimension $K\times K$. Denote by $\mathbf{H}_{st}$ the $K\times K$ block at position s,t. Tridiagonal refers to the block structure here; it means that $\mathbf{H}_{st}=0$ for $|t-s|<1$.

Because of the tridiagonal block band structure of $\mathbf{H}$, its lower Cholesky factor $\mathbf{L}$ is a lower triangular block band matrix, with lower block bandwidth equal to one. Thus, $\mathbf{L}_{st}=0$ except for $s=t$ and $s=t-1$.

We can compute $\mathbf{L}$ using $\mathbf{L}_{11}=\mathrm{chol}(\mathbf{H}_{11})$ and for $t=2,\ldots,T$,

$$\mathbf{L}_{t,t-1}=\mathbf{H}_{t,t-1}\mathbf{L}_{t-1,t-1}^{-\top},\qquad \mathbf{L}_{tt}=\mathrm{chol}(\mathbf{H}_{tt}-\mathbf{L}_{t,t-1}\mathbf{L}_{t,t-1}^\top),$$

where $\mathrm{chol}(A)$ refers to the lower Cholesky factor of any positive definite matrix A.

We can use forward substitution to compute $x=\mathbf{L}^{-1}y$, for any $KT\times 1$ vectors x and y. Denoting by x_t the t'th $K\times 1$ subvector of x and similarly for y,

$$x_1=\mathbf{L}_{11}^{-1}y_1,\quad x_t=\mathbf{L}_{tt}^{-1}(y_t-\mathbf{L}_{t,t-1}x_{t-1}).$$

Likewise, we can use backward substitution to compute $z=\mathbf{L}^{-\top}x$, for any $KT\times 1$ vectors x and z:

$$z_T=\mathbf{L}_{TT}^{-\top}x_T\quad z_t=\mathbf{L}_{tt}^{-\top}(x_t-\mathbf{L}_{t+1,t}^\top z_{t+1}).$$

References

Aguilar, O., and West, M. (2000). 'Bayesian dynamic factor models and variance matrix discounting for portfolio allocation', *Journal of Business and Economic Statistics*, 18: 338–357.

Bai, J., and Ng, S. (2008). 'Large Dimensional Factor Analysis', *Foundations and Trends in Econometrics*, 3: 89–163.

Bollerslev, T. (1986). 'Generalized Autoregressive Conditonal Heteroskedasticity', *Journal of Econometrics*, 31(3): 307–327.

Carter, C. K., and Kohn, R. (1994). 'On Gibbs Sampling for State Space Models', *Biometrika*, 81(3): 541–553.

Carvalho, C. M., Chang, J., Lucas, J. E., Wang, Q., and West, M. (2008). 'High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics', *Journal of the American Statistical Association*, 103(484): 1438–1456.

Chan, J. C. C., and Jeliazkov, I. (2009). 'Efficient Simulation and Integrated Likelihood Estimation in State Space Models', Working paper.

Chib, S., Omori, Y., and Asai, M. (2009). 'Multivariate Stochastic Volatility', in T. G. Anderson (ed.), *Handbook of Financial Time Series*. Springer–Verlag.

de Jong, P., and Shephard, N. (1995). 'The Simulation Smoother for Time Series Models', *Biometrika*, 82(1): 339–350.

Del Negro, M., and Otrok, C. (2008). 'Dynamic factor models with time-varying parameters: measuring changes in international business cycles', Discussion Paper Staff Reports 326, Federal Reserve Bank of New York.

Dhrymes, P. J. (2000). *Mathematics for Econometrics, Third Edition*. Springer–Verlag, New York.

Durbin, J., and Koopman, S. J. (2002). 'A Simple and Efficient Simulation Smoother for State Space Time Series Analysis', *Biometrika*, 89(3): 603–615.

Frühwirth-Schnatter, S. (1994). 'Data augmentation and Dynamic Linear Models', *Journal of Time Series Analysis*, 15: 183–202.

Geweke, J. (2004). 'Getting it Right: Joint Distribution Tests of Posterior Simulators', *Journal of the American Statistical Association*, 99: 799–804.

Geweke, J. F. (1977). 'The Dynamic Factor Analysis of Economic Time Series', in D. J. Aigner and A. S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*. North–Holland.

Golub, G. H., and Van Loan, C. F. (1996). *Matrix Computations*. Johns Hopkins University Press, 3rd edn.

Kim, C.-J., and Nelson, C. (2000). *State Space Models with Regime Switching*. MIT Press.

McCausland, W. J. (2012). 'The HESSIAN Method: Highly Efficient State Smoothing, in a Nutshell', *Journal of Econometrics*, 168: 189–206.

McCausland, W. J., Miller, S., and Pelletier, D. (2011). 'Simulation smoothing for state-space models: A computational efficiency analysis', *Computational Statistics and Data Analysis*, 55: 199–212.

McCracken, M. W., and Ng, S. (2014). 'FRED-MD: A Monthly Database for Macroeconomic Research', (unpublished manuscript).

Nakajima, J., and West, M. (2013). 'Dynamic Factor Volatility Modeling: A Bayesian Latent Threshold Approach', *Journal of Financial Econometrics*, 11: 116–153.

Rue, H. (2001). 'Fast Sampling of Gaussian Markov Random Fields with Applications', *Journal of the Royal Statistical Society Series B*, 63: 325–338.

Stock, J. H., and Watson, M. W. (2011). 'Dynamic Factor Models', in M. P. Clements (ed.), *The Oxford Handbook of Economic Forecasting*. Oxford University Press.