Supplemental Appendix to "Bayesian Dynamic Factor Models for High-Dimensional Matrix-Valued Time Series"

Wei Zhang

Purdue University

zhan3721@purdue.edu

January 2025

Abstract

This appendix provides additional details on proof of identification, priors, the posterior sampler, marginal likelihood estimation, simulation results and data.

Contents

1 Proof of Propositions

1.1 Proof of *Proposition 1*

Proof of Proposition 1: Without the loss of generality, we prove one of the two cases in Proposition 1. That is, we assume $\text{Var}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$ and $\mathbf{A}$ is a lower-triangular matrix with ones on the diagonal, while $\mathbf{B}$ is a lower-triangular matrix with strictly positive diagonal elements.

As shown in (1.1), we may identify a rotation of $\mathbf{F}_t$, given by $\mathbf{C}\mathbf{F}_t\mathbf{D}'$.

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1}\mathbf{C}\mathbf{F}_t\mathbf{D}'(\mathbf{D}')^{-1}\mathbf{B}' + \mathbf{E}_t, \tag{1.1}$$

where $\mathbf{C}$ and $\mathbf{D}$ are $p_1 \times p_1$ and $p_2 \times p_2$ invertible matrices.

We use $\widetilde{\mathbf{F}}_t$ to denote the rotated factor matrix: $\widetilde{\mathbf{F}}_t \equiv \mathbf{C}\mathbf{F}_t\mathbf{D}'$, and we use $\widetilde{\mathbf{f}}_t$ to denote the vectorized $\mathbf{F}_t$: $\widetilde{\mathbf{f}}_t \equiv (\mathbf{D} \otimes \mathbf{C})\mathbf{f}_t$.

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & ... & 0 \\ a_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1} & a_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & ... & a_{np_1} \end{bmatrix}, \mathbf{C}^{-1} = \begin{bmatrix} c_{11} & ... & c_{1p_1} \\ \vdots & \ddots & \vdots \\ c_{p_1 1} & ... & c_{p_1 p_1} \end{bmatrix}$$

Then the rotated factor loadings $\mathbf{A}\mathbf{C}^{-1}$ needs to be a lower triangular matrix with ones on the diagonal as well, that is,

$$\begin{bmatrix} 1 & 0 & ... & 0 \\ a_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{p_1 1} & a_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & ... & a_{np_1} \end{bmatrix} \begin{bmatrix} c_{11} & ... & c_{1p_1} \\ \vdots & \ddots & \vdots \\ c_{p_1 1} & ... & c_{p_1 p_1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... & 0 \\ a^*_{21} & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a^*_{p_1 1} & a^*_{p_1 2} & ... & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a^*_{n1} & a^*_{n2} & ... & a^*_{np_1} \end{bmatrix} \tag{1.2}$$

For (1.2) to hold, we must have $c_{i,j} = 0$ for any i, j such that $i < j$ and $c_{i,i} = 1$, or $\mathbf{C}^{-1}$ is lower triangular with ones on the diagonal.

Similarly,

$$\begin{bmatrix} b_{11} & 0 & ... & 0 \\ b_{21} & b_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p_2 1} & b_{p_2 2} & ... & b_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & ... & b_{np_2} \end{bmatrix} \begin{bmatrix} d_{11} & ... & d_{1p_2} \\ \vdots & \ddots & \vdots \\ d_{p_2 1} & ... & d_{p_2 p_2} \end{bmatrix} = \begin{bmatrix} b^*_{11} & 0 & ... & 0 \\ b^*_{21} & b^*_{22} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{p_1 1} & b^*_{p_1 2} & ... & b^*_{p_2 p_2} \\ \vdots & \vdots & \ddots & \vdots \\ b^*_{n1} & b^*_{n2} & ... & b^*_{np_2} \end{bmatrix} \tag{1.3}$$

For (1.3) to hold, we must have $d_{ij} = 0$ for any i, j such that $i < j$, or $\mathbf{D}^{-1}$ is lower triangular given the assumption that $b_{ii} \neq 0$, $b^*_{ii} \neq 0$, for $i = 1, ..., p_2$.

Define $\mathbf{f}_t \equiv \text{vec}(\mathbf{F}_t)$. Consider the case $q = 1$, we rewrite (2) as follows

$$\mathbf{f}_t = \mathbf{H}_{\boldsymbol{\rho}} \mathbf{f}_{t-1} + \mathbf{u}_t, \tag{1.4}$$

where $\mathbf{H}_{\boldsymbol{\rho}}$ is a diagonal matrix with $\boldsymbol{\rho} = (\rho_{1,1,t}, ..., \rho_{p_1,p_2,t})'$ on the diagonal. $\mathbf{u}_t = (u_{1,1,t}, ..., u_{p_1,p_2,t})'$, $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \Lambda_t)$, where $\Lambda_1 = \text{diag}(\lambda^2_{1,1}/(1-\rho^2_{1,1}), ..., \lambda^2_{p_1,p_2}/(1-\rho^2_{p_1,p_2}))$ for $t = 1$, and $\Lambda_t = \text{diag}(\lambda^2_{1,1}, ..., \lambda^2_{p_1,p_2})$ for $t = 2, ..., T$.

Define $\mathbf{M} \equiv \mathbf{D} \otimes \mathbf{C}$, multiply (1.4) by $\mathbf{M}$ on both side, we have

$$\mathbf{M}\mathbf{f}_t = \mathbf{M}\mathbf{H}_{\boldsymbol{\rho}} \mathbf{f}_{t-1} + \mathbf{M}\mathbf{u}_t. \tag{1.5}$$

Therefore

$$\widetilde{\mathbf{f}}_t = \mathbf{M}\mathbf{H}_{\boldsymbol{\rho}}\mathbf{M}^{-1}\widetilde{\mathbf{f}}_{t-1} + \mathbf{M}\mathbf{u}_t. \tag{1.6}$$

The observation equation after the rotation becomes

$$\mathbf{Y}_t = \mathbf{A}\mathbf{C}^{-1}\widetilde{\mathbf{F}}_t(\mathbf{D}')^{-1}\mathbf{B}' + \mathbf{E}_t. \tag{1.7}$$

Given the condition that $\text{Var}(\mathbf{u}_t) = \mathbf{I}_{p_1 p_2}$, $\text{Var}(\mathbf{M}\mathbf{u}_t)$ should be an identity matrix as well.

That is, $\mathbf{M}\text{Var}(\mathbf{u}_t)\mathbf{M}' = \mathbf{I}_{p_1p_2}$. Therefore, we have $\mathbf{M}\mathbf{M}' = \mathbf{I}_{p_1p_2}$. Or $\mathbf{M}$ is an orthogonal matrix. Therefore, we have

$$\mathbf{M}\mathbf{M}' = \mathbf{I} \Leftrightarrow (\mathbf{D} \otimes \mathbf{C})(\mathbf{D} \otimes \mathbf{C})' = \mathbf{I} \Leftrightarrow (\mathbf{D}\mathbf{D}') \otimes (\mathbf{C}\mathbf{C}') = \mathbf{I},$$

which holds if and only if $\mathbf{D}\mathbf{D}' = \mathbf{I}_{p_2}$ and $\mathbf{C}\mathbf{C}' = \mathbf{I}_{p_1}$, given that C and D are lower triangular matrices and the diagonal elements of $\mathbf{C}$ is ones.

This means that $\mathbf{C}$ and $\mathbf{D}$ are orthogonal matrices. An orthogonal matrix that is lower triangular must be diagonal. Therefore, the rotation matrix $\mathbf{C}$ is an identity matrix. Given that $b_{ii} > 0$ for $i = 1, ..., p_2$, we must have that the rotation matrix $\mathbf{D}$ is also an identity matrix.

This proves that the proposed assumptions in *MDFM1* fully identify the factor matrix and the factor loading matrices.

1.2 Proof of *Proposition 2*

Proof of Proposition 2: Similar to the proof of proposition 1, the rotated factor loadings $\mathbf{C}^{-1}$ needs to be a lower triangular matrix, as shown in (1.2). Additionally, given we have ones on the diagonal of $\mathbf{A}$, $\mathbf{C}^{-1}$ needs to have ones on its diagonal as well. Similarly, $\mathbf{D}^{-1}$ needs to be a lower triangular matrix with ones on its diagonal. Therefore, the matrix $\mathbf{M}$ is a lower triangular matrix with ones on its diagonal.

Again, we need $\text{Cov}(\mathbf{M}\mathbf{u}_t) = \text{Cov}(\mathbf{u}_t)$, i.e., $\mathbf{M}\mathbf{\Lambda}_t\mathbf{M}' = \mathbf{\Lambda}_t$, where $\mathbf{\Lambda}_t$ is a diagonal matrix. Given that the diagonal elements in $\mathbf{\Lambda}_t$ must be larger than 0, this requires that $m_{i,j}$ for all $i > j$ must be zero, for $\mathbf{M}\mathbf{\Lambda}_t\mathbf{M}'$ to only have non-zero terms on its diagonal and match $\mathbf{\Lambda}_t$. Therefore, $\mathbf{M}$ must be identity matrix.

This proves that assumptions 1, 4 and 5 fully identifies the factor matrix and the factor loading matrices.

2 Bayesian Estimation for MDFM with Stochastic Volatility

Recall the dynamic factor model for matrix-valued time series with stochastic volatility

$$\mathbf{Y}_t = \mathbf{A}\mathbf{F}_t\mathbf{B}' + \mathbf{E}_t, \quad \text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}, \omega_t \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r), \tag{2.8}$$

$$\text{vec}(\mathbf{F}_t) = \mathbf{H}_{\boldsymbol{\rho}_1}\text{vec}(\mathbf{F}_{t-1}) + \ldots + \mathbf{H}_{\boldsymbol{\rho}_q}\text{vec}(\mathbf{F}_{t-q}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t), \tag{2.9}$$

where $\mathbf{A}$ is a $n \times p_1$ matrix of factor loadings, $\mathbf{B}$ is a $k \times p_2$ matrix of factor loadings, $\mathbf{F}_t$ is a $p_1 \times p_2$ latent matrix-valued time series of common factors, $\mathbf{E}_t$ is a $n \times k$ idiosyncratic component, $\text{vec}(\,\cdot\,)$ is a vectorizing function, $\mathbf{H}_{\boldsymbol{\rho}_l}$ is a diagonal matrix of autoregressive coefficients $(\rho_{1,l}, \ldots, \rho_{p_1 p_2,l})'$, $l = 1, \ldots, q$, and $\boldsymbol{\Lambda}_t$ is a covariance matrix for the error in factor evolution process.

We use a natural conjugate prior for the transpose of factor loadings: $\mathbf{A}'$ and $\mathbf{B}'$. In addition, we use inverse-Wishart prior for $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_c$:

$$\begin{aligned} \boldsymbol{\Sigma}_r &\sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\text{vec}(\mathbf{A}')|\boldsymbol{\Sigma}_r) \sim \mathcal{N}(\text{vec}(\mathbf{A}_0'), \boldsymbol{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}'}), \\ \boldsymbol{\Sigma}_c &\sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\text{vec}(\mathbf{B}')|\boldsymbol{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\mathbf{B}_0'), \boldsymbol{\Sigma}_c \otimes \mathbf{V}_{\mathbf{B}'}). \end{aligned} \tag{2.10}$$

The autoregressive coefficient $\rho_{j,k,l}$ is assumed to have a truncated normal prior on the interval $(-1, 1)$:

$$\rho_{j,k,l} \sim \mathcal{TN}(\rho_{j,k,l,0}, V_{\rho_{j,k,l}}), \quad j = 1, ..., p_1, \quad k = 1, ..., p_2, \quad l = 1, \ldots, q.$$

The prior variance $\lambda_{j,k}^2$ is assumed to have a inverse-gamma prior: $\mathcal{IG}(\nu_{\lambda_{j,k}}, S_{\lambda_{j,k}})$. We also treat the first q factors as unknown, and use the following prior

$$f_{j,k,l} \sim \mathcal{N}\left(0, \frac{\lambda_{j,k}^2}{1 - \sum_{m=1}^{q} \rho_{j,k,m}^2}\right), \quad l = 1, \ldots, q.$$

For identification, we use assumptions 1, 4 and 5. We employ Markov Chain Monte Carlo (MCMC) methods to obtain a draw from the joint posterior of the latent factors and parameters of the model. Specifically, the following steps are carried out:

1. Sampling from $(\mathbf{A}', \boldsymbol{\Sigma}_r | \mathbf{Y}, \mathbf{B}, \mathbf{F}, \boldsymbol{\Sigma}_c)$

We sample $(\mathbf{A}', \boldsymbol{\Sigma}_r)$ conditional on the latent factors and other parameters from a normal-inverse-Wishart distribution:

$$(\mathbf{A}', \boldsymbol{\Sigma}_r | \cdot\,) \sim \mathcal{NIW}(\widehat{\mathbf{A}}', \mathbf{K}_{\mathbf{A}'}^{-1}, \widehat{\nu}_r, \widehat{\mathbf{S}}_r),$$

where

$$\mathbf{K}_{\mathbf{A}'} = \mathbf{V}_{\mathbf{A}'}^{-1} + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t \mathbf{B}' \boldsymbol{\Sigma}_c^{-1} \mathbf{B} \mathbf{F}_t', \quad \widehat{\mathbf{A}}' = \mathbf{K}_{\mathbf{A}'}^{-1} \left(\mathbf{V}_{\mathbf{A}'}^{-1} \mathbf{A}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t \mathbf{B}' \boldsymbol{\Sigma}_c^{-1} \mathbf{Y}_t' \right)$$

$$\widehat{\nu}_r = \nu_r + Tk, \quad \widehat{\mathbf{S}}_r = \mathbf{S}_r + \mathbf{A}_0 \mathbf{V}_{\mathbf{A}'}^{-1} \mathbf{A}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{Y}_t \boldsymbol{\Sigma}_c^{-1} \mathbf{Y}_t' - \widehat{\mathbf{A}} \mathbf{K}_{\mathbf{A}'} \widehat{\mathbf{A}}'.$$

With the constraints for identification, we cannot directly sample from the above normal-inverse-Wishart distribution. Here we outline the sampling scheme for $\mathbf{A}'$ with the structure constraints. To that end, we first represent the restrictions as a system of linear restrictions. For example, for $\mathbf{A}'$, we represent the restrictions that $\mathbf{A}$ is a lower triangular matrix with ones on the diagonal using $\mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}') = \mathbf{a}_0$. Assuming $n > p_1$, $\mathbf{M}_{\mathbf{A}'} = (m_{i,j})$ is a $p_1(p_1+1)/2 \times np_1$ selection matrix, and $\mathbf{a}_0$ is a $p_1(p_1+1)/2 \times 1$ vector consisting of ones and zeros. Then we apply Algorithm 2 in Cong et al. (2004) or Algorithm 1 in Chan and Qi (2024) to efficiently sample $(\text{vec}(\mathbf{A}')| \cdot\,) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}'), \boldsymbol{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})$ such that $\mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}') = \mathbf{a}_0$. In particular, one can first sample $\text{vec}(\mathbf{A}_u')$ from the unconstrained conditional posterior distribution in Step 1, and then return

$$\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}_u') + (\boldsymbol{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1})\mathbf{M}_{\mathbf{A}'}' \left(\mathbf{M}_{\mathbf{A}'} (\boldsymbol{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}'}^{-1}) \mathbf{M}_{\mathbf{A}'}' \right)^{-1} (\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}_u')),$$

which can be realized by the following four steps:

(1) Compute $\mathbf{C} = \mathbf{C}_{\boldsymbol{\Sigma}_r^{-1}} \otimes \mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$, where $\mathbf{C}_{\boldsymbol{\Sigma}_r^{-1}}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}_r^{-1}$, and $\mathbf{C}_{\mathbf{K}_{\mathbf{A}'}}$ is the lower Cholesky factor of $\mathbf{K}_{\mathbf{A}'}$;

(2) Solve $\mathbf{C}\mathbf{C}'\mathbf{U} = \mathbf{M}_{\mathbf{A}'}'$ for $\mathbf{U}$;

(3) Solve $\mathbf{M}_{\mathbf{A}'}\mathbf{U}\mathbf{V} = \mathbf{U}'$ for $\mathbf{V}$;

(4) Return $\text{vec}(\mathbf{A}') = \text{vec}(\mathbf{A}_u') + \mathbf{V}'(\mathbf{a}_0 - \mathbf{M}_{\mathbf{A}'}\text{vec}(\mathbf{A}_u'))$.

2. Sampling from $(\mathbf{B}', \boldsymbol{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \boldsymbol{\Sigma}_r)$

Similar to step 1, $(\mathbf{B}, \mathbf{\Sigma}_c)$ are drawn from a normal-inverse-Wishart distribution:

$$(\mathbf{B}, \mathbf{\Sigma}_c | \cdot) \mathcal{NIW}(\widehat{\mathbf{B}}', \mathbf{K}_{\mathbf{B}'}^{-1}, \widehat{\nu}_c, \widehat{\mathbf{S}}_c),$$

where

$$\mathbf{K}_{\mathbf{B}'} = \mathbf{V}_{\mathbf{B}'}^{-1} + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \mathbf{\Sigma}_r^{-1} \mathbf{A} \mathbf{F}_t, \quad \widehat{\mathbf{B}}' = \mathbf{K}_{\mathbf{B}'}^{-1} \left(\mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{F}_t' \mathbf{A}' \mathbf{\Sigma}_r^{-1} \mathbf{Y}_t \right)$$

$$\widehat{\nu}_c = \nu_c + Tn, \quad \widehat{\mathbf{S}}_c = \mathbf{S}_c + \mathbf{B}_0 \mathbf{V}_{\mathbf{B}'}^{-1} \mathbf{B}_0' + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{Y}_t' \mathbf{\Sigma}_r^{-1} \mathbf{Y}_t - \widehat{\mathbf{B}} \mathbf{K}_{\mathbf{B}'} \widehat{\mathbf{B}}'.$$

We sample $(\mathbf{B}, \mathbf{\Sigma}_c | \cdot)$ in two steps. First, we sample $\mathbf{\Sigma}_c$ marginally from $(\mathbf{\Sigma}_c | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r) \sim \mathcal{IW}(\widehat{\mathbf{S}}_c, \nu_c + Tn)$ with the normalization restriction that $\sigma_{c,1,1} = 1$. This can be done using the algorithm in Nobile (2000) described below. Then we simulate $(\text{vec}(\mathbf{B}') | \mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \mathbf{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}'}^{-1})$, which can be done using the algorithm described in step 1.

The algorithm in Nobile (2000) can be realized by the following steps:

(1) Exchange row/column 1 and n in the matrix $\widehat{\mathbf{S}}_c$. Denote this matrix as $\widehat{\mathbf{S}}_c^{Trans}$.
(2) Construct a lower triangular matrix $\mathbf{\Delta}$ such that
- δ_{ii} equal to the square root of $\chi^2_{\widehat{\nu}_c+1-i}$ for $i = 1, \ldots, n-1$;
- $\delta_{nn} = (l_{nn})^{-1}$, where l_{nn} is the (n, n)-th element in the Cholesky decomposition of $(\widehat{\mathbf{S}}_c^{Trans})^{-1}$, denoted as $\boldsymbol{L}$
- δ_{ij} equal to $\mathcal{N}(0, 1)$ random variates, $i > j$.

(3) Set $\mathbf{\Sigma}_c = (\boldsymbol{L}^{-1})' (\mathbf{\Delta}^{-1})' \mathbf{\Delta}^{-1} \boldsymbol{L}^{-1}$.
(4) Exchange the row/column 1 and n of $\mathbf{\Sigma}_c$ back.

3. Sampling from $(\text{vec}(\mathbf{F}_t) | \mathbf{Y}_t, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, \boldsymbol{\omega}^2, \boldsymbol{\rho})$, $t = 1, \ldots, T$
We sample the factors by t. Specifically, conditional on parameters, $\text{vec}(\mathbf{F}_t)$ from a normal distribution:

$$(\text{vec}(\mathbf{F}_t) | \cdot) \sim \mathcal{N}(\widehat{\mathbf{f}}_t, \mathbf{K}_{\mathbf{f}_t}^{-1}),$$

where

$$\mathbf{K}_{\mathbf{f}_t} = \omega_t^{-1} \mathbf{B}' \mathbf{\Sigma}_c^{-1} \mathbf{B} \otimes \mathbf{A}' \mathbf{\Sigma}_r^{-1} \mathbf{A} + \mathbf{\Lambda}^{-1}, \quad \widehat{\mathbf{f}}_t = \mathbf{K}_{\mathbf{f}_t}^{-1} \left[\omega_t^{-1} (\mathbf{B}' \mathbf{\Sigma}_c^{-1} \otimes \mathbf{A}' \mathbf{\Sigma}_r^{-1}) \text{vec}(\mathbf{Y}_t) + \mathbf{\Lambda}^{-1} \mathbf{H}_{\boldsymbol{\rho}} \mathbf{f}_{t-1} \right].$$

Step 4. Sampling from $(\lambda^2_{j,k}|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}), \quad j = 1, ..., p_1, k = 1, ..., p_2$

It is clear that $(\lambda^2_{j,k}|\mathbf{f}_{j,k}, \boldsymbol{\rho}_{j,k}) \sim \mathcal{IG}(\widehat{\nu}_{\lambda_{j,k}}, \widehat{S}_{\lambda_{j,k}})$, where $\widehat{\nu}_{\lambda_{j,k}} = \nu_{\lambda_{j,k}} + \frac{T}{2}$, and $\widehat{S}_{\lambda_{j,k}} = S_{\lambda_{j,k}} + \frac{1}{2}\left[\sum_{t=1}^{q} f^2_{j,k,t}(1 - \sum_m \rho^2_{j,k,m}) + \sum_{t=q+1}^{T}(f_{j,k,t} - \rho_{j,k,1} f_{j,k,t-1} - ... - \rho_{j,k,q} f_{j,k,q})^2\right]$.

Step 5. Sampling from $(\boldsymbol{\rho}_{j,k}|\mathbf{f}_{j,k}, \lambda^2_{j,k}), \quad j = 1, ..., p_1, k = 1, ..., p_2$

Note that $\boldsymbol{\rho}_{j,k}$ is a $q \times 1$ vector: $\boldsymbol{\rho}_{j,k} = (\rho_{j,k,1}, \ldots, \rho_{j,k,q})'$. We rewrite (2) as follows:

$$\widetilde{\mathbf{f}}_{j,k} = \widetilde{\mathbf{F}}_{j,k}\boldsymbol{\rho}_{j,k} + \mathbf{u}_{j,k}, \quad \mathbf{u}_{j,k} \sim \mathcal{N}(\mathbf{0}, \lambda_{j,k}\mathbf{I}_{T-q}), \tag{2.11}$$

where $\widetilde{\mathbf{f}}_{j,k} = (f_{j,k,q+1}, \ldots, f_{j,k,T})'$, and

$$\widetilde{\mathbf{F}}_{j,k} = \begin{bmatrix} f_{j,k,1} & f_{j,k,2} & \cdots & f_{j,k,q} \\ f_{j,k,2} & f_{j,k,3} & \cdots & f_{j,k,q+1} \\ \vdots & \cdots & \cdots & \vdots \\ f_{j,k,T-q} & f_{j,k,T-q+1} & \cdots & f_{j,k,T} \end{bmatrix}.$$

Following Chib and Greenberg (1994) and Chan and Jeliazkov (2009), we design an Metropolis-Hastings algorithm with proposal $\boldsymbol{\rho}^*_{j,k} \sim \mathcal{N}(\hat{\boldsymbol{\rho}}_{j,k}, \mathbf{K}^{-1}_{\boldsymbol{\rho}_{j,k}})$, where $\mathbf{K}_{\boldsymbol{\rho}_{j,k}} = \mathbf{V}^{-1}_{\boldsymbol{\rho}_{j,k}} + \widetilde{\mathbf{F}}'_{j,k}\widetilde{\mathbf{F}}_{j,k}/\lambda^2_{j,k}$, $\hat{\boldsymbol{\rho}}_{j,k} = \mathbf{K}^{-1}_{\boldsymbol{\rho}_{j,k}}(\mathbf{V}^{-1}_{\boldsymbol{\rho}_{j,k}}\boldsymbol{\rho}_{j,k,0} + \widetilde{\mathbf{F}}'_{j,k}\widetilde{\mathbf{f}}_{j,k}/\lambda^2_{j,k})$. The proposed value $\boldsymbol{\rho}^*_{j,k}$ is accepted with probablity

$$\alpha_{MH}(\boldsymbol{\rho}_{j,k}, \boldsymbol{\rho}^*_{j,k}) = \min\left\{1, \frac{f_{\mathcal{N}}(\mathbf{f}_{j,k,1:q}|\mathbf{0}, \lambda^2_{j,k}/(1 - \sum_m \rho^{*2}_{j,k,m})\mathbf{I}_q)}{f_{\mathcal{N}}(\mathbf{f}_{j,k,1:q}|\mathbf{0}, \lambda^2_{j,k}/(1 - \sum_m \rho^2_{j,k,m})\mathbf{I}_q)}\right\}.$$

5. Sampling the time-varying volatility

For clearer illustration, assume that we have only one type of time-varying volatility. The following three steps correspond to each type.

5.1 Common stochastic volatility: sampling from $(\mathbf{h}|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r)$

The conditional posterior for $\mathbf{h}$ is not a standard distribution. In this paper, we follow Chan (2017) for this purpose. In particular, we first obtain the mode of the log density of $(\mathbf{h}|\cdot)$ as well as the negative Hessian evaluated at the mode, denoted as $\widehat{\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}}$, respectively. Then we use $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}^{-1}_{\mathbf{h}})$ as the proposal distribution, and sample $\mathbf{h}$ using an acceptance-rejection Metropolis-Hasting step. Samplers for ϕ and σ^2_h are standard and we omit the details in this paper.

5.2 The explicit outlier components: sampling from $(\mathbf{o}, p_{\mathbf{o}}|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$

We follow Stock and Watson (2016) to discretize the support of o_t to simplify estimation. Specifically, we use a grid with points at $1, 2, 3, ..., 20$. The likelihood can be easily evaluated at these grid points. Finally, a draw from the full conditional posterior distribution of o_t can be obtained using the inverse transform method.

The conditional distribution of p_{o_i} is a Beta distribution:

$$(p_{o_i}|\mathbf{o}_i) \sim \mathcal{B}(a_{p_{o_i}} + n_2, b_{p_{o_i}} + n_1),$$

where $n_1 = \sum_{t=1}^{T} \mathbf{I}(o_{i,t} = 1)$ is the number of "regular" periods, and $n_2 = T - \sum_{t=1}^{T} \mathbf{I}(o_{i,t} = 1)$ is the number of "outlier" periods.

5.3 Fat-tailed innovations: sampling from $(q_t^2|\mathbf{Y}, \mathbf{A}, \mathbf{F}, \mathbf{B}, \mathbf{\Sigma}_c, \mathbf{\Sigma}_r)$, $t = 1, \ldots, T$

Conditional on the factors and parameters, the posterior for q_t^2 has an inverse-gamma distribution:

$$(q_t^2| \cdot) \sim \mathcal{IG}((nk + l)/2, (s_t^2 + l)/2),$$

where $s_t^2 = \text{tr}\,[\mathbf{\Sigma}_c^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{F}_t\mathbf{B})'\mathbf{\Sigma}_r^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{F}_t\mathbf{B})]$.

3 Estimating Marginal Likelihoods

This section describes the method we use to obtain integrated likelihood and the importance-sampling densities. For illustration, we consider $q = 1$.

3.1 Integrated Likelihood

Model (1)(2) can be rewritten as follows

$$\begin{aligned} &\mathbf{y}_t = (\mathbf{B} \otimes \mathbf{A})\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r), \\ &\mathbf{f}|\boldsymbol{\rho}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\mathbf{H}_{\boldsymbol{\rho}}'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_{\boldsymbol{\rho}}\right]^{-1}\right), \end{aligned} \tag{3.12}$$

System (3.12) can be rewritten as follows

$$\begin{aligned} &\mathbf{y} = (\mathbf{I}_T \otimes \mathbf{A} \otimes \mathbf{B})\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T \otimes (\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r)), \\ &\mathbf{f}|\boldsymbol{\rho}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\mathbf{0}, \left[\mathbf{H}_{\boldsymbol{\rho}}'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_{\boldsymbol{\rho}}\right]^{-1}\right). \end{aligned} \tag{3.13}$$

It is easy to integrate out $\mathbf{f}$ and we can get the following likelihood

$$\mathbf{y}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho} \sim \mathcal{N}(\bar{\mathbf{y}}, \overline{\mathbf{D}}_{\mathbf{y}}), \tag{3.14}$$

where

$$\begin{aligned} \bar{\mathbf{y}} &= \mathbb{E}\left[\mathbb{E}\left(\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right)|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right] \\ &= \mathbb{E}\left[(\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})\mathbf{f}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right] \\ &= (\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})\mathbb{E}[\mathbf{f}|\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}] \\ &= \mathbf{0}, \end{aligned}$$

and

$$\begin{aligned} \overline{\mathbf{D}}_{\mathbf{y}} &= \mathbb{E}\left\{[\mathrm{Var}\left(\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}, \boldsymbol{\rho}\right)|\cdot\right\} + \mathrm{Var}\left(\mathbb{E}[\mathbf{y}|\mathbf{f}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\Omega}]|\cdot\right)] \\ &= \mathbf{I}_T \otimes \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{I}_T \otimes \mathbf{B} \otimes \mathbf{A})[\mathbf{H}_\rho'(\mathbf{I}_T \otimes \boldsymbol{\Omega})^{-1}\mathbf{H}_\rho]^{-1}(\mathbf{I}_T \otimes \mathbf{B}' \otimes \mathbf{A}'). \end{aligned}$$

It can be very costly to compute the inverse of the covariance matrix $\overline{\mathbf{D}}_{\mathbf{y}}$. Therefore, here we use Kalman filter. In particular, it is not difficult to show that the marginal distribution for $\mathbf{f}_t \equiv \text{vec}(\mathbf{F}_t)$ is as follows:

$$
\begin{aligned}
(\mathbf{f}_1 \mid \boldsymbol{\rho}, \boldsymbol{\lambda}) &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_1) \\
(\mathbf{f}_t \mid \boldsymbol{\rho}, \boldsymbol{\lambda}) &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t + \mathbf{H}_{\boldsymbol{\rho}} \boldsymbol{\Lambda}_{t-1} \mathbf{H}'_{\boldsymbol{\rho}}), \quad t = 2, \ldots, T,
\end{aligned}
$$

where for $t = 2, \ldots, T$, $\boldsymbol{\Lambda}_t = \text{diag}(\lambda^2_{1,1}, \lambda^2_{2,1}, ..., \lambda^2_{p_1,p_2})$, and for $t = 1$, $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda^2_{1,1}/(1 - \rho^2_{1,1}), \lambda^2_{2,1}/(1 - \rho^2_{2,1}), ..., \lambda^2_{p_1,p_2}/(1 - \rho^2_{p_1,p_2}))$. $\mathbf{H}_{\boldsymbol{\rho}} = \text{diag}(\rho_{1,1}, \rho_{2,1}, ..., \rho_{p_1,p_2})$

Therefore, the integrated likelihood at time t is:

$$
(\mathbf{y}_t \mid \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r) \sim \mathcal{N}(\mathbf{0}, \overline{\mathbf{D}}_{\mathbf{y}_t}),
$$

where

$$
\begin{aligned}
\overline{\mathbf{D}}_{\mathbf{y}_1} &= \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{B} \otimes \mathbf{A}) \boldsymbol{\Lambda}_1 (\mathbf{B}' \otimes \mathbf{A}') \\
\overline{\mathbf{D}}_{\mathbf{y}_t} &= \boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r + (\mathbf{B} \otimes \mathbf{A})(\boldsymbol{\Lambda}_t + \mathbf{H}_{\boldsymbol{\rho}} \boldsymbol{\Lambda}_{t-1} \mathbf{H}'_{\boldsymbol{\rho}})(\mathbf{B}' \otimes \mathbf{A}'), \quad t = 2, \ldots, T.
\end{aligned}
$$

3.2 Finding the Optimal Importance-sampling Densities

The next step is to find the maximum likelihood estimators for the hyperparameters in the importance-sampling density. The importance-sampling density is denoted as

$$
\begin{aligned}
f(\boldsymbol{\theta}; \mathbf{v}) &= f(\mathbf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}, \boldsymbol{\rho}; \mathbf{v}) \\
&= f(\mathbf{A}; \overline{\mathbf{A}}, \overline{\mathbf{D}}_{\mathbf{A}}) \cdot f(\boldsymbol{\Sigma}_c; \Psi_c, \nu_c) \cdot f(\boldsymbol{\Sigma}_r; \Psi_r, \nu_r) \cdot f(\boldsymbol{\lambda}; \nu_\lambda, S_\lambda) \cdot f(\boldsymbol{\rho}; \overline{\boldsymbol{\rho}}, \overline{\mathbf{D}}_{\boldsymbol{\rho}}).
\end{aligned} \tag{3.15}
$$

In terms of the parameteric family, we use Gaussian density for $f(\mathbf{A}; \overline{\mathbf{A}}, \overline{\mathbf{D}}_{\mathbf{A}})$, where $\overline{\mathbf{A}}$ and $\overline{\mathbf{D}}$ are the corresponding mean and covariance matrix. We use inverse Wishart densities for $f(\boldsymbol{\Sigma}_c; \nu_c, \Psi_c)$ as well as $f(\boldsymbol{\Sigma}_r; \nu_r, \Psi_r)$. We use inverse gamma density for We use the truncated normal density on the interval $(-1, 1)$ for $f(\boldsymbol{\rho}; \overline{\boldsymbol{\rho}}, \overline{\mathbf{D}}_{\boldsymbol{\rho}})$, where $\overline{\boldsymbol{\rho}}$ and $\overline{\mathbf{D}}_{\boldsymbol{\rho}}$ are the corresponding mean and covariance matrix. we use inverse-gamma distribution for $f(\boldsymbol{\lambda}; \nu_\lambda, S_\lambda)$.

In order to obtain the maximum likelihood estimators for the parameters in inverse

Wishart distribution, we first use maximum likelihood estimation on the Wishart distribution given the posterior samples, and then compute the degree of freedom and scale matrix of the inverse Wishart distribution using *Lemma 1*.

Lemma 1: $\mathbf{\Sigma}$ follows an inverse Wishart distribution if $\mathbf{K} \equiv \mathbf{\Sigma}^{-1}$ follows a Wishart distribution, formally expressed as

$$\mathbf{\Sigma} \sim \mathcal{IW}_d(\delta - d + 1, \mathbf{\Psi}^{-1}) \Leftrightarrow \mathbf{K} = \mathbf{\Sigma}^{-1} \sim \mathcal{W}_d(\delta, \mathbf{\Psi}), \tag{3.16}$$

where d is the dimension of the matrix $\mathbf{\Sigma}$, δ is the degree of freedom of the Wishart distribution, and $\mathbf{\Psi}$ is the scale matrix.

A Wishart distribution is defined as:

$$f(\mathbf{K}|\mathbf{\Psi}, \delta) = \frac{|\mathbf{K}|^{\frac{\delta-d-1}{2}}}{2^{\frac{\delta d}{2}}|\mathbf{\Psi}|^{\frac{\delta}{2}}\Gamma_d\left(\frac{\delta}{2}\right)} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{K}\mathbf{\Psi}^{-1})\right\}.$$

We assume that each matrix is drawn independently from the same Wishart distribution $\mathcal{W}(\mathbf{\Psi}, \delta)$. Therefore, we can model the joint distribution as:

$$f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta) = \prod_{m=1}^{M} \frac{|\mathbf{K}_m|^{\frac{\delta-d-1}{2}}}{2^{\frac{\delta d}{2}}|\mathbf{\Psi}|^{\frac{\delta}{2}}\Gamma_d\left(\frac{\delta}{2}\right)} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{K}_m\mathbf{\Psi}^{-1})\right\}.$$

The log-likelihood function is therefore

$$\log f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta) = -\frac{\delta d M}{2}\log 2 - \frac{\delta M}{2}\log|\mathbf{\Psi}| - M\log\Gamma_d\left(\frac{\delta}{2}\right) + \frac{\delta - d - 1}{2}\sum_{m=1}^{M}\log|\mathbf{K}_m| - \frac{1}{2}\mathrm{tr}\left(\sum_{m=1}^{M}\mathbf{K}_m\mathbf{\Psi}^{-1}\right).$$

The first derivative of the log-likelihood function with respect to the scale matrix $\mathbf{\Psi}$ is equal to

$$\frac{\mathrm{d}\log(f(\mathbf{K}_1, ..., \mathbf{K}_M|\mathbf{\Psi}, \delta))}{\mathrm{d}\mathbf{\Psi}} = -\frac{M\delta}{2}\mathbf{\Psi}^{-1} + \frac{1}{2}\mathbf{\Psi}^{-1}\sum_{m=1}^{M}\mathbf{K}_m\mathbf{\Psi}^{-1}, \tag{3.17}$$

where two results are used

1. $\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}} = |\mathbf{X}|\mathbf{X}^{-1}$;
2. $\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^{-1})}{\partial\mathbf{X}} = -\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}$.

From equation (3.17) we obtain a function of the MLE of $\mathbf{\Psi}$ with respect to the degree of freedom δ

$$\widehat{\mathbf{\Psi}}^{mle} = \frac{1}{M\delta}\sum_{m=1}^{M}\mathbf{K}_m. \tag{3.18}$$

In order to obtain the MLE for the degree of freedom, a straightforward way is to find the first order condition and second order condition to maximize the log-likelihood function with respect to δ. We then use the Newton-type methods to find the estimate for $\widehat{\delta}$.

In particular, the first derivative of the log-likelihood function after we plug in (3.18) is

$$\begin{aligned}\frac{\partial \log f(\mathbf{K}_1,...,\mathbf{K}_M|\delta)}{\partial\delta} =& -\frac{dM}{2}(\log 2+1) + \frac{Md}{2}\log\delta - \frac{M}{2}\log\left|M^{-1}\sum_m \mathbf{K}_m\right| \\ & - \frac{M}{2}\psi_d\left(\frac{\delta}{2}\right) + \frac{1}{2}\sum_m \log|\mathbf{K}_m|.\end{aligned} \tag{3.19}$$

The second derivative is

$$\frac{\partial^2 \log f(\mathbf{K}_1,...,\mathbf{K}_M|\delta)}{\partial\delta^2} = -\frac{Md}{2\delta} - \frac{M}{4}\psi_d^{(2)}\left(\frac{1}{2}\delta\right).$$

Maximum likelihood estimators for parameters for normal distributions and inverse gamma distributions are straightforward to obtain so that we omit the details here.

4 Additional Simulation Results

Table 4.1: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 3$, $p_2 = 2$

(n, k)	T = 200		T = 500		T = 1000	
	0.98	0.98	0.99	0.97	0.98	0.99
(10, 10)	0.98	0.96	0.98	0.98	0.98	0.99
	0.96	0.96	0.99	0.99	0.98	0.99
Average	0.97		0.98		0.98	
	1.00	0.98	1.00	0.99	1.00	0.99
(20, 15)	0.97	0.98	0.99	0.98	1.00	0.99
	0.99	0.99	0.99	0.97	0.98	0.98
Average	0.98		0.99		0.99	
	1.00	0.98	1.00	1.00	1.00	0.99
(30, 20)	0.99	0.98	0.99	0.99	0.99	0.99
	0.99	0.98	0.98	0.98	0.99	0.99
Average	0.99		0.99		0.99	

Table 4.2: Adjusted R^2 from regressing the true factors on the estimates: $p_1 = 5$, $p_2 = 5$

(n, k)			$T = 200$					$T = 500$					$T = 1000$		
	0.97	0.98	0.95	0.97	0.97	0.99	0.98	0.97	0.95	0.97	0.99	0.99	0.99	0.98	0.98
	0.96	0.97	0.94	0.97	0.97	0.98	0.97	0.97	0.93	0.97	0.98	0.97	0.98	0.97	0.98
(10, 10)	0.97	0.96	0.97	0.97	0.97	0.97	0.96	0.95	0.92	0.98	0.98	0.97	0.98	0.98	0.98
	0.96	0.96	0.93	0.95	0.96	0.97	0.98	0.96	0.94	0.98	0.96	0.96	0.95	0.95	0.97
	0.95	0.96	0.95	0.93	0.95	0.98	0.97	0.96	0.91	0.96	0.99	0.98	0.98	0.97	0.98
Average			0.96					0.96					0.98		
	0.99	0.99	0.99	0.99	0.96	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.97
	0.99	0.99	0.99	0.98	0.94	0.99	0.97	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.98
(20, 15)	0.99	0.99	0.99	0.98	0.96	0.99	0.98	0.98	0.97	0.99	0.99	0.99	0.99	0.99	0.97
	0.98	0.98	0.98	0.98	0.98	0.99	0.95	0.98	0.97	0.98	0.99	0.99	0.99	0.99	0.97
	0.98	0.97	0.98	0.97	0.95	0.98	0.97	0.96	0.96	0.96	0.99	0.98	0.99	0.98	0.97
Average			0.98					0.98					0.99		
	1.00	1.00	0.99	0.99	0.92	1.00	0.99	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	0.99	0.99	0.99	0.99	0.97	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99
(30, 20)	0.99	0.99	0.99	0.97	0.96	1.00	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.98	0.99
	0.97	0.98	0.99	0.99	0.91	0.98	0.99	0.99	0.97	0.99	1.00	0.99	0.99	0.99	0.99
	0.99	0.99	0.98	0.99	0.97	0.97	0.99	0.98	0.97	0.98	0.99	0.99	0.99	0.99	0.99
Average			0.98					0.99					0.99		

5 Data: Multinational Macroeconomic Panel

Table 5.3 describes the list of variables we use for the first application. We attach the link of the website we downloaded the specific variable to the variable name in the table. The second column of 5.3 is the stationarity transformation for each variable.

Table 5.3: List of variables

Variable	Transformation
Real GDP	No transformation
Consumption	$\Delta \log(x)$
Labor unit costs	Δx
Unemployment	Δx
Headline CPI	Δx
Energy CPI	Δx
Food CPI	Δx
Core CPI	Δx
Imports	$\Delta \log(x)$
Exports	$\Delta \log(x)$

References

Chan, J. C. (2017): "The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling," *Journal of Business & Economic Statistics*, 35, 17–28.

Chan, J. C. and I. Jeliazkov (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

Chan, J. C. and Y. Qi (2024): "Large Bayesian Matrix Autoregressions," *Available at SSRN 4855762*.

Chib, S. and E. Greenberg (1994): "Bayes inference in regression models with ARMA (p, q) errors," *Journal of Econometrics*, 64, 183–206.

Cong, Y., B. Chen, and M. Zhou (2004): "Fast Simulation of Hyperplane-Truncated Multivariate Normal Distributions," *Bayesian Analysis*, 1.

Nobile, A. (2000): "Comment: Bayesian multinomial probit models with a normalization constraint," *Journal of Econometrics*, 99, 335–345.

Stock, J. H. and M. W. Watson (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98, 770–784.