Long Run Effect in Heterogeneous Panel Data Model: A Bayesian Approach

CHEUNG Ying Lun*

Goethe University Frankfurt

Abstract

ABSTRACT HERE

*E-mail address: yinglun.cheung@stud.uni-frankfurt.de

1. Introduction

2. Literature Review

3. The Model

In this paper we consider the heterogeneous panel data model

$$y_{it} = \sum_{j=1}^{p_i} \psi_{ij} y_{i,t-j} + \sum_{j=0}^{q_i} \boldsymbol{\beta}_{ij}' \boldsymbol{x}_{i,t-j} + u_{it}, \qquad u_{it} = \boldsymbol{v}_i' \boldsymbol{f}_t + \varepsilon_{it}. \tag{3.1}$$

Here indices run for cross-sectional dimension from $i = 1, \ldots, N$, and for time-series dimension $t = 1, \ldots, T$. The disturbance term u_{it} is correlated with the k-dimension (weakly) exogenous variables $\boldsymbol{x}_{it}$, which is modeled by assuming $\boldsymbol{x}_{it}$ follows an m-factor model with the same factors of u_{it}, i.e.

$$\boldsymbol{x}_{it} = \boldsymbol{\Gamma}_i \boldsymbol{f}_t + \boldsymbol{e}_{it}, \qquad \boldsymbol{e}_{it} \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}_k, \boldsymbol{R}_{e_i}), \tag{3.2}$$

where $\boldsymbol{\Gamma}_i$ are $k \times m$ matrices and $\boldsymbol{0}_k$ indicates a k-dimension zero vector. $\boldsymbol{x}_{it}$ is correlated cross-sectionally, i.e. across i. Stacking $\boldsymbol{x}_{it}$ into a vector for all $i = 1, \ldots, N$, we model the cross-sectional correlation by

$$\boldsymbol{x}_t = \boldsymbol{\Gamma} \boldsymbol{f}_t + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}_{Nk}, \boldsymbol{R}), \tag{3.3}$$

where $\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{x}_{1t}' & \ldots & \boldsymbol{x}_{Nt}' \end{pmatrix}'$ is $Nk \times 1$, $\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_1' & \ldots & \boldsymbol{\Gamma}_N' \end{pmatrix}'$ is $Nk \times m$, $\boldsymbol{e}_t = \begin{pmatrix} \boldsymbol{e}_{1t}' & \ldots & \boldsymbol{e}_{Nt}' \end{pmatrix}'$ is $Nk \times 1$, and $\boldsymbol{R}$ is $Nk \times Nk$ with diagonal blocks of $\boldsymbol{R}_{e_i}$, and the off-diagonal elements, possibly non-zero, are not restricted.

We assume the common factors to follow a Gaussian VAR(p) model

$$\boldsymbol{f}_t = \sum_{j=1}^{p} \boldsymbol{\Phi}_j \boldsymbol{f}_{t-j} + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}_{Nk}, \boldsymbol{Q}_\nu). \tag{3.4}$$

Finally the disturbance term u_{it} is modeled similarly as $\boldsymbol{x}_{it}$. Stacking all N disturbance terms, we have

$$\boldsymbol{u}_t = \boldsymbol{\Upsilon} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{\Sigma}), \tag{3.5}$$

where $\boldsymbol{u}_t = \begin{pmatrix} u_{1t} & \ldots & u_{Nt} \end{pmatrix}'$ is a $N \times 1$ vector, $\boldsymbol{\Upsilon} = \begin{pmatrix} \boldsymbol{v}_1' & \ldots & \boldsymbol{v}_N' \end{pmatrix}'$ is $N \times m$, $\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1t} & \ldots & \varepsilon_{Nt} \end{pmatrix}'$ and $\boldsymbol{\Sigma}$ an $N \times N$ covariance matrix.

The vector of long-run coefficients for each i is defined by

$$\boldsymbol{\theta}_i = \frac{\sum_{j=0}^{q_i} \boldsymbol{\beta}_{ij}}{1 - \sum_{j=1}^{p_i} \psi_{ij}}. \tag{3.6}$$

This model, in particular the long-run effect in this model, is studied by Chudik et al. (2015) in a frequentist approach. They used the cross-sectional (weighted) average of y_{it} and $\boldsymbol{x}_{it}$ as instruments and suggested a mean-group and a pooled estimator of $\boldsymbol{\theta} = \mathbb{E}(\boldsymbol{\theta}_i)$. In this paper, we tackle the problem in a Bayesian approach.

4. Estimation

Notice that Eq.(3.1) reduces to a standard Bayesian VAR (BVAR) problem if $\boldsymbol{\upsilon}_i = \mathbf{0}$ for all i or when $\boldsymbol{f}_t$ is observed, we can break down the problem into two steps:

1. Estimate $\boldsymbol{f}_t$ from $\boldsymbol{X}_t$ by Gibbs sampling.
2. Conditional on the estimated $\widehat{\boldsymbol{f}_t}$, estimate the BVAR model.

Since the BVAR is standard in the second step, we may add identifying restrictions like sign restriction (e.g. ...) and long run restriction (e.g. ...) in the identification step.

4.1. State-Space Representation and Kalman Filter

Notice that the undesirable features of u_{it} that it is correlated with $\boldsymbol{x}_{it}$ and is autocorrelated are due to the assumption that it depends on the common factors $\boldsymbol{f}_t$. The problem can be circumvented if we can explicitly include $\boldsymbol{f}_t$ in the estimation of the BVAR model. If we treat $\boldsymbol{f}_t$ as mT additional parameters, we can in spirit estimate $\boldsymbol{f}_t$ together with other parameters, or in a sequential way, estimate $\boldsymbol{f}_t$ and then the remaining parameters conditional on the estimated $\widehat{\boldsymbol{f}_t}$.

Therefore, as a first step, we estimate $\boldsymbol{f}_t$ from the observed data. We notice that $\boldsymbol{x}_t$ and $\boldsymbol{f}_t$ can be written in a linear and Gaussian state-space representation. Thus, by treating it as a factor-augmented VAR (FAVAR) model, we follow the procedures of Bernanke et al. (2005) and obtain draws of $\boldsymbol{f}_t$ from the posterior distribution.

Appendix A. Derivations

A.1. Estimation of Common Factors

In this section we show the procedures of estimation of the common factors $\boldsymbol{f}_t$.

A.1.1. State-Space Representation

Notice that Eq.(3.3) and (3.4) are actually the state-space representation of a system. Writing Eq.(3.4) in companion form and expressing Eq.(3.3) in terms of $\widetilde{\boldsymbol{f}}_t = \begin{pmatrix} \boldsymbol{f}_t' & \dots & \boldsymbol{f}_{t-p+1}' \end{pmatrix}'$, we have

$$\boldsymbol{x}_t = \widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{f}}_t + \boldsymbol{e}_t \tag{A.1}$$

$$\widetilde{\boldsymbol{f}}_t = \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{f}}_{t-1} + \widetilde{\boldsymbol{\nu}}_t, \qquad \widetilde{\boldsymbol{\nu}}_t \sim \text{i.i.d. } \mathcal{N}(\boldsymbol{0}_{pm}, \boldsymbol{Q}), \tag{A.2}$$

where $\widetilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \boldsymbol{\Gamma} & \mathbf{O}_{Nk\times m} & \dots & \mathbf{O}_{Nk\times m} \end{pmatrix}'$ is $Nk \times pm$ and

$$\widetilde{\boldsymbol{\Phi}} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \boldsymbol{I}_m & \mathbf{O}_m & \dots & \mathbf{O}_m & \mathbf{O}_m \\ \mathbf{O}_m & \boldsymbol{I}_m & \dots & \mathbf{O}_m & \mathbf{O}_m \\ \vdots & & \ddots & & \vdots \\ \mathbf{O}_m & \dots & \dots & \boldsymbol{I}_m & \mathbf{O}_m \end{pmatrix}, \qquad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_\nu & \mathbf{O}_{p\times(m-1)p} \\ \mathbf{O}_{(m-1)p\times p} & \mathbf{O}_{(m-1)p} \end{pmatrix}$$

are both $pm \times pm$. Note that Eq.(A.1) is essentially the same as Eq.(3.3), and Eq.(A.2) is just the companion form of Eq.(3.4), as mentioned already above. Now the system is similar to that of Bernanke et al. (2005) and Amir-Ahmadi and Uhlig (2009) and we can estimate $\widetilde{\boldsymbol{f}}_f$ by the Kalman filter with the same procedures.

A.1.2. Gibbs Sampling and Kalman Smoothing

Since $\boldsymbol{e}_t$ and $\widetilde{\boldsymbol{\nu}}_t$ are independent and Normally distributed, Eq.(A.1) and (A.2) allow us to apply Gibbs sampling and Kalman smoothing in a standard way. The detailed steps can be found in Bernanke et al. (2005), who follows Kim and Nelson (1999). Nevertheless, it is still worthwhile to outline the procedures here to avoid confusions in notations.

Let $\boldsymbol{\Theta} = (\widetilde{\boldsymbol{\Gamma}}, \boldsymbol{R}, \widetilde{\boldsymbol{\Phi}}, \boldsymbol{Q})$ be the set of parameters and let $\boldsymbol{x}^T$ and $\widetilde{\boldsymbol{f}}^T$ denote the history of $\boldsymbol{x}_t$ and $\boldsymbol{f}_t$ for $t = 1, \dots, T$. The multi-move Gibbs sampling involves three steps:

1. *Choosing a set of starting value* $\boldsymbol{\Theta}_0$:
   A variety of $\boldsymbol{\Theta}_0$ should be chosen for robustness check. A 'meaningful' starting value,

e.g. the principal components estimation of Eq.(3.3) and a VAR estimation of Eq.(3.4), is suggested in Bernanke et al. (2005) to improve computational efficiency. Notice that for all $m \times m$ orthonormal matrix $\boldsymbol{A}$, we can re-write Eq.(3.3) as

$$\boldsymbol{x}_t = \boldsymbol{\Gamma}\boldsymbol{A}\boldsymbol{A}'\boldsymbol{f}_t + \boldsymbol{e}_t = \boldsymbol{\Gamma}^*\boldsymbol{f}_t^* + \boldsymbol{e}_t \tag{A.3}$$

without changing the likelihood functions, one may want to restrict the principal-component identification of $\boldsymbol{\Gamma}$ and $\boldsymbol{f}_t$ by for example fixing the upper $m \times m$ block of $\boldsymbol{\Gamma}$ to be equal to the identity matrix.

2. *Densities of $\boldsymbol{f}_t$ conditional on $\boldsymbol{\Theta}_i$ and $\boldsymbol{x}^T$*
Notice that due to the Markov property of Eq.(A.2), we can write the joint (conditional) density of $\widetilde{\boldsymbol{f}}^T$ as products of conditional densities

$$p\left(\widetilde{\boldsymbol{f}}^T \middle| \boldsymbol{\Theta}_i, \boldsymbol{x}^T\right) = p\left(\widetilde{\boldsymbol{f}}_1 \middle| \boldsymbol{\Theta}_i, \boldsymbol{x}^T\right) \prod_{t=2}^{T} p\left(\widetilde{\boldsymbol{f}}_t \middle| \widetilde{\boldsymbol{f}}_{t-1}, \boldsymbol{\Theta}_i, \boldsymbol{x}^T\right) \tag{A.4}$$

$$= p\left(\widetilde{\boldsymbol{f}}_T \middle| \boldsymbol{\Theta}_i, \boldsymbol{x}^T\right) \prod_{t=1}^{T-1} p\left(\widetilde{\boldsymbol{f}}_t \middle| \widetilde{\boldsymbol{f}}_{t+1}, \boldsymbol{\Theta}_i, \boldsymbol{x}^T\right). \tag{A.5}$$

Moreover, since the system is Gaussian and linear, we can apply the technique of Kalman smoothing to estimate $\widetilde{\boldsymbol{f}}_t$. We apply here the Rauch-Tung-Striebel (RTS) smoother, a forward-backward algorithm to obtain an estimate of $\widetilde{\boldsymbol{f}}_t$. The forward part is the same as the common Kalman filter. We define the usual notations of conditional expectations $\widetilde{\boldsymbol{f}}_{t|s} = \mathbb{E}\left(\widetilde{\boldsymbol{f}}_t \middle| \boldsymbol{\Theta}_i, \boldsymbol{x}^s\right)$ and $\widetilde{\boldsymbol{f}}_{t|s,\widetilde{\boldsymbol{f}}_\tau} = \mathbb{E}\left(\widetilde{\boldsymbol{f}}_t \middle| \widetilde{\boldsymbol{f}}_\tau, \boldsymbol{\Theta}_i, \boldsymbol{x}^s\right)$, as well as the conditional covariance matrices $\boldsymbol{Q}_{t|s} = \mathrm{cov}\left(\widetilde{\boldsymbol{f}}_{t|s}\right)$ and $\boldsymbol{Q}_{t|s,\widetilde{\boldsymbol{f}}_\tau} = \mathrm{cov}\left(\widetilde{\boldsymbol{f}}_{t|s,\widetilde{\boldsymbol{f}}_\tau}\right)$. Moreover, we let $\boldsymbol{\eta}_{t|t-1} = \boldsymbol{x}_t - \widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{f}}_{t|t-1}$ denote the prediction error and $\boldsymbol{H}_{t|t-1} = \widetilde{\boldsymbol{\Gamma}}\boldsymbol{Q}_{t|t-1}\widetilde{\boldsymbol{\Gamma}}' + \boldsymbol{R}$ its covariance matrix. Initializing with $\widetilde{\boldsymbol{f}}_{1|0} = \boldsymbol{0}_{mp}$ and $\boldsymbol{Q}_{1|0}$ be a diagonal matrix with only the first m diagonal elements being one, we have

$$\widetilde{\boldsymbol{f}}_{t|t} = \widetilde{\boldsymbol{f}}_{t|t-1} + \boldsymbol{Q}_{t|t-1}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{H}_{t|t-1}^{-1}\boldsymbol{\eta}_{t|t-1} \tag{A.6}$$

$$\boldsymbol{Q}_{t|t} = \boldsymbol{Q}_{t|t-1} - \boldsymbol{Q}_{t|t-1}\widetilde{\boldsymbol{\Phi}}'\boldsymbol{H}_{t|t-1}^{-1}\widetilde{\boldsymbol{\Phi}}\boldsymbol{Q}_{t|t-1}. \tag{A.7}$$

From Eq.(A.2), it is obvious that $\widetilde{\boldsymbol{f}}_{t|t-1} = \widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{f}}_{t-1|t-1}$ and $\boldsymbol{Q}_{t|t-1} = \widetilde{\boldsymbol{\Phi}}\boldsymbol{Q}_{t-1|t-1}\widetilde{\boldsymbol{\Phi}}' + \boldsymbol{Q}$. Since from Eq.(A.5) we have

$$\widetilde{\boldsymbol{f}}_T \middle| \boldsymbol{x}^T, \boldsymbol{\Theta}_i \sim \mathcal{N}\left(\widetilde{\boldsymbol{f}}_{T|T}, \boldsymbol{Q}_{T|T}\right), \tag{A.8}$$

$$\widetilde{\boldsymbol{f}}_t \middle| \widetilde{\boldsymbol{f}}_{t+1}, \boldsymbol{x}^T, \boldsymbol{\Theta}_i \sim \mathcal{N}\left(\widetilde{\boldsymbol{f}}_{t|t,\widetilde{\boldsymbol{f}}_{t+1}}, \boldsymbol{Q}_{T|T,\widetilde{\boldsymbol{f}}_{t+1}}\right), \tag{A.9}$$

we can draw $\widetilde{\boldsymbol{f}}_T$ from Eq.(A.8), in which the density parameters can be obtained from the last iteration of the Kalman filter described above. For the remaining periods, we can now apply the backward pass of the Kalman smoother according to

$$\boldsymbol{f}_{t|T} = \boldsymbol{f}_{t|t} + \boldsymbol{C}_t \left(\boldsymbol{f}_{t+1|T} - \boldsymbol{f}_{t+1|t} \right) \tag{A.10}$$

$$\boldsymbol{Q}_{t|T} = \boldsymbol{Q}_{t|t} + \boldsymbol{C}_t (\boldsymbol{Q}_{t+1|T} - \boldsymbol{Q}_{t+1|t}) \boldsymbol{C}_t' \tag{A.11}$$

$$\boldsymbol{C}_t = \boldsymbol{Q}_{t|t} \boldsymbol{\Phi}' \boldsymbol{Q}_{t+1|t}^{-1}. \tag{A.12}$$

However, notice that $\boldsymbol{Q}$ is singular if $p > 1$, Kim and Nelson (1999) and Amir-Ahmadi and Uhlig (2009) suggest that we only use the upper $m \times m$ block of all $\boldsymbol{Q}$ matrices and the first m rows of the corresponding vectors and matrices. The resulting $\boldsymbol{f}_{t|T}$ and $\boldsymbol{Q}_{t|T}$ will not contain lags.

3. *Densities of $\boldsymbol{\Theta}_{i+1}$ conditional on $\boldsymbol{f}^T$ and $\boldsymbol{x}^T$*
   Now given the observed data $\boldsymbol{x}^T$ and the estimated unobserved common factors $\widehat{\boldsymbol{f}}^T$ estimated through Kalman smoothing, we estimate a new set of parameters θ_{i+1}. The state and observation equations can be estimated separately.

   (a) *Observation equation*
   Notice that Eq.(3.3) is just a system of standard regression equations, we can apply OLS equation by equation to obtain estimates of parameter $\widehat{\boldsymbol{\gamma}}_i$ and of residuals $\widehat{\boldsymbol{e}}_i$ for the i-th equation. If we assume further that $\boldsymbol{R}$ is diagonal, as in Bernanke et al. (2005) and Amir-Ahmadi and Uhlig (2009), we can set $R_{ij} = 0$ for $i \neq j$, and impose an Inverse Gamma prior for R_{ii}, more specifically,

$$R_{ii} \sim \mathcal{IG}(\alpha, \beta). \tag{A.13}$$

   Here R_{ij} denotes the (i,j)-th element of $\boldsymbol{R}$. The posterior of R_{ii} is given by

$$R_{ii} | \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T \sim \mathcal{IG}(\alpha^*, \beta^*) \tag{A.14}$$

   where

$$\alpha^* = \alpha + \widehat{\boldsymbol{e}}_i' \widehat{\boldsymbol{e}}_i + \widehat{\boldsymbol{\gamma}}_i' \left[\boldsymbol{M}_0^{-1} + (\widehat{\boldsymbol{f}}^{T\prime} \widehat{\boldsymbol{f}}^T)^{-1} \right]^{-1} \widehat{\boldsymbol{\gamma}}_i \tag{A.15}$$

$$\beta^* = \beta + T. \tag{A.16}$$

   Bernanke et al. (2005) chooses $\alpha = 3$, $\beta = 0.001$ and $\boldsymbol{M}_0 = \boldsymbol{I}$ to obtain a diffuse prior. $\boldsymbol{M}_0^{-1}$ is the variance parameter in the prior on the coefficients $\boldsymbol{\gamma}_i$, which

has conditional prior distribution

$$\boldsymbol{\gamma}_i | R_{ii} \sim \mathcal{N}(\boldsymbol{\gamma}_0, R_{ii}\boldsymbol{M}_0^{-1}), \tag{A.17}$$

where we choose $\boldsymbol{\gamma}_0 = \mathbf{0}_m$. The posterior distribution of $\boldsymbol{\gamma}_i$ is then given by

$$\boldsymbol{\gamma}_i | \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T, \widehat{R}_{ii} \sim \mathcal{N}(\overline{\boldsymbol{\gamma}}_i, \widehat{R}_{ii}\boldsymbol{M}_i^{-1}), \tag{A.18}$$

where

$$\overline{\boldsymbol{\gamma}}_i = \boldsymbol{M}_i \left[\boldsymbol{M}_0^{-1}\boldsymbol{\gamma}_0 + \left(\widehat{\boldsymbol{f}}^{T\prime} \widehat{\boldsymbol{f}}^T \widehat{\boldsymbol{\gamma}}_i \right) \right] \tag{A.19}$$

$\widehat{R}_{ii}$ is drawn from Eq.(A.14), and $\boldsymbol{M}_i = \boldsymbol{M}_0 + \widehat{\boldsymbol{f}}^{T\prime} \widehat{\boldsymbol{f}}^T$. If we assume more generally that $\boldsymbol{R}$ is not diagonal, we can impose the Normal-Inverse Wishart prior, namely

$$\boldsymbol{R} \sim \mathcal{W}^{-1}(\boldsymbol{R}_0, Nk) \tag{A.20}$$

$$\mathrm{vec}(\boldsymbol{\Gamma}) | \boldsymbol{R} \sim \mathcal{N}(\mathbf{0}_{Nkm}, \boldsymbol{R} \otimes \boldsymbol{\Omega}_0). \tag{A.21}$$

$\boldsymbol{\Omega}_0$ is set as $\boldsymbol{I}_m$ so that each regressor is equally likely to be non-zero, and that coefficient of the i-th regressor is likely to be independent of that of the j-th regressor, where $i \neq j$. $\boldsymbol{R}_0$ is set as the covariance matrix of OSL residuals. The posterior distribution is given by

$$\boldsymbol{R} | \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T \sim \mathcal{W}^{-1}(\overline{\boldsymbol{R}}, Nk + T) \tag{A.22}$$

$$\mathrm{vec}(\boldsymbol{\Gamma}) | \boldsymbol{R}, \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T \sim \mathcal{N}\left(\mathrm{vec}(\overline{\boldsymbol{\Gamma}}), \boldsymbol{R} \otimes \overline{\boldsymbol{\Omega}} \right), \tag{A.23}$$

where

$$\overline{\boldsymbol{R}} = \boldsymbol{R}_0 + \widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}} + \widehat{\boldsymbol{\Gamma}}' \left[\boldsymbol{\Omega}_0 + (\boldsymbol{f}^{T\prime}\boldsymbol{f}^T)^{-1} \right]^{-1} \widehat{\boldsymbol{\Gamma}} \tag{A.24}$$

$$\overline{\boldsymbol{\Gamma}} = \overline{\boldsymbol{\Omega}}(\widehat{\boldsymbol{f}}^{T\prime}\widehat{\boldsymbol{f}}^T)\widehat{\boldsymbol{\Gamma}} \tag{A.25}$$

$$\overline{\boldsymbol{\Omega}} = (\boldsymbol{\Omega}_0^{-1} + \widehat{\boldsymbol{f}}^{T\prime}\widehat{\boldsymbol{f}}^T)^{-1} \tag{A.26}$$

and $\widehat{\boldsymbol{E}}$ is the residual matrix.

(b) *State equation*

Eq.(3.4) is a standard VAR model and sampling is done in a similar manner as

in the observation equation. We assume again the Normal-Inverse Wishart prior

$$\boldsymbol{Q}_\nu \sim \mathcal{W}^{-1}(\boldsymbol{Q}_{\nu 0}, m) \tag{A.27}$$

$$\text{vec}(\boldsymbol{\Phi})|\boldsymbol{Q}_\nu \sim \mathcal{N}(\boldsymbol{0}_{pm}, \boldsymbol{Q}_\nu \otimes \boldsymbol{\Omega}_{f0}), \tag{A.28}$$

where $\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \dots & \boldsymbol{\Phi}_p \end{pmatrix}$. $\boldsymbol{\Omega}_{f0}$ is set as in Bernanke et al. (2005), i.e. such that the prior variance of parameter on j lagged i_1-th variable in i_2-th equation equals $\sigma_{i_2}^2 / j\sigma_{i_1}^2$. The posterior distribution is given by

$$\boldsymbol{Q}_\nu | \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T \sim \mathcal{W}^{-1}(\overline{\boldsymbol{Q}}_\nu, m + T) \tag{A.29}$$

$$\text{vec}(\boldsymbol{\Phi})|\boldsymbol{Q}_\nu, \boldsymbol{x}^T, \widehat{\boldsymbol{f}}^T \sim \mathcal{N}\left(\text{vec}(\overline{\boldsymbol{\Phi}}), \boldsymbol{Q}_\nu \otimes \overline{\boldsymbol{\Omega}}_f\right), \tag{A.30}$$

where

$$\overline{\boldsymbol{Q}}_\nu = \boldsymbol{Q}_{\nu 0} + \widehat{\boldsymbol{V}}'\widehat{\boldsymbol{V}} + \widehat{\boldsymbol{\Phi}}' \left[\boldsymbol{\Omega}_{f0} + (\boldsymbol{f}^{T-1\prime}\boldsymbol{f}^{T-1})^{-1}\right]^{-1} \widehat{\boldsymbol{\Phi}} \tag{A.31}$$

$$\overline{\boldsymbol{\Phi}} = \overline{\boldsymbol{\Omega}}_f (\widehat{\boldsymbol{f}}^{T-1\prime} \widehat{\boldsymbol{f}}^{T-1}) \widehat{\boldsymbol{\Phi}} \tag{A.32}$$

$$\overline{\boldsymbol{\Omega}}_f = (\boldsymbol{\Omega}_{f0}^{-1} + \widehat{\boldsymbol{f}}^{T-1\prime} \widehat{\boldsymbol{f}}^{T-1})^{-1} \tag{A.33}$$

and $\widehat{\boldsymbol{V}}$ is the OLS residual matrix. As in Chudik et al. (2015), stationary of $\boldsymbol{f}_t$ is not enforced. Thus, we do not drop draws that suggest non-stationary like what is done in Bernanke et al. (2005) and Amir-Ahmadi and Uhlig (2009).

At this point, we have computed the posterior distribution of parameters required. We can then draw from it to obtain $\boldsymbol{\Theta}_{i+1}$ for the next run.

We record the draws from the second and third steps above after $\underline{I}$ iterations, when convergence is ensured after then.

References

Amir-Ahmadi, P., Uhlig, H., 2009. Measuring the effects of monetary policy: A Bayesian FAVAR approach with sign restriction, mimeo.

Bernanke, B. S., Boivin, J., Eliasz, P., 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. Quarterly Journal of Economics 120, 387–422.

Chudik, A., Mohaddes, K., Pesaran, M. H., Raissi, M., 2015. Long-run effects in large heterogenous panel data models with cross-sectionally correlated errors. USC INET Working Paper No. 15-05 .

Kim, C.-J., Nelson, C. R., 1999. State-Space Models with Regime Switching. MIT Press.