Large BVARs with Stochastic Volatility

Marco Del Negro, Domenico Giannone — *Federal Reserve Bank of New York*

Daniel L. Greenwald — *NYU*

January 26, 2015

1 A VAR with Factor Innovations and Stochastic Volatility

The model is a vector autoregression:

$$y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t, \tag{1}$$

where $t = 1, \dots T$, y_t, and u_t are $n \times 1$ vectors of observables and innovation c and $\Phi_1, \dots \Phi_p$, are $n \times 1$ vector of intercepts and $n \times n$ matrices respectively.

The innovations u_t follow a factor model

$$u_t = Bf_t + \eta_t, \tag{2}$$

where B is a $n \times k$ matrix of loading f_t and η_t are $k \times 1$ and $n \times 1$ vectors of common and idiosyncratic shocks. We assume that each element $e_{i,t}$ of the vector of shocks $[\eta_t', f_t']'$ can be written as

$$e_{i,t} = \sigma_{i,0} e^{\tilde{\sigma}_{i,t}} \tilde{h}_{i,t}^{-1/2} \varepsilon_{i,t}, \tag{3}$$

for $i = 1, .., n, .., n+k$, where i) $\varepsilon_{i,t}$ is a standard Gaussian innovation i.i.d. across i, t; ii) $\sigma_{i,0}$ captures the average volatility for innovation i; iii) $e^{\tilde{\sigma}_{i,t}}$ captures changes in volatility, whose law of motion we discuss below ; iv) $\tilde{h}_{i,t}$ is distributed to a Chi square distribution:

$$\lambda_i \tilde{h}_{i,t} \sim \chi^2(\lambda_i), \text{ i.i.d. across } i,\ t, \tag{4}$$

and captures fat tail with the fatness of the tails being potentially different across regimes.[1] These assumptions imply that

$$\eta_t|\sigma^2_{1:n,0}, \tilde{\sigma}_{1:n,t}, \tilde{h}_{1:n,t} \sim \mathcal{N}(0, \Sigma_{\eta,t}), \text{ i.i.d. across } t, \tag{5}$$

and

$$f_t|\sigma^2_{n+1:n+k,0}, \tilde{\sigma}_{n+1:n+k,t}, \tilde{h}_{n+1:n+k,t} \sim \mathcal{N}(0, \Sigma_{f,t}), \text{ i.i.d. across } t, \tag{6}$$

where $\Sigma_{f,t}$ and $\Sigma_{\eta,t}$ are diagonal matrices with elements $\sigma^2_{i,t} = \sigma^2_{i,0}e^{2\tilde{\sigma}_{it}}\tilde{h}^{-1}_{i,t}$ for $i = 1,..,n$ and $i = n+1,...,n+k$, respectively, and where we we use the notation $x_{1:n,1:T}$ to denote the sequence $\{x_{1,1},..,x_{i,t},..,x_{n,T}\}$ for a generic variable x.

We can thus rewrite the VAR as

$$y_t = \Phi' x_t + Bf_t + \eta_t = X'_t\varphi + F'_t b + \eta_t \tag{7}$$

where $x_t = [1, y'_{t-1}, \ldots, y'_{t-p}]'$, $\Phi' = [c, \Phi_1, \ldots, \Phi_p]$, $X'_t = I_n \otimes x'_t$, $\varphi = vec(\Phi)$, $F'_t = I_n \otimes f'_t$, and $b = vec(B')$.[2]

We assume that the innovations to the law of motion of the stochastic volatilities also follow a factor structure. Specifically, we let innovations in the volatility of the factors $\tilde{\sigma}_{n+1:n+k,t}$ affect the volatility of the idiosyncratic shocks $\tilde{\sigma}_{1:n,t}$, which otherwise evolve independently from one another. The law of motion of the factor volatilities follows a VAR(1)

$$\tilde{\sigma}_{f,t} = \mathrm{P}'_f\tilde{\sigma}_{f,t-1} + \zeta_{f,t}, \; \zeta_{f,t} \sim \mathcal{N}(0, W_f), \text{ i.i.d. across } t, \tag{8}$$

where $\tilde{\sigma}_{f,t} = [\tilde{\sigma}_{n+1,t}, \ldots, \tilde{\sigma}_{n+k,t}]'$, P_f and W_f are $k \times k$ matrice the latter also symmetric and positive definite matrix. Conditional on the factor innovation the idiosyncratic shocks volatilities evolve according to independent AR(1) processes:

$$\tilde{\sigma}_{\eta,t} = \mathrm{P}'_\eta\tilde{\sigma}_{\eta,t-1} + \Xi'\zeta_{f,t} + \zeta_{\eta,t}, \; \zeta_{\eta,t} \sim \mathcal{N}(0, W_\eta), \text{ i.i.d. across } t, \tag{9}$$

where $\tilde{\sigma}_{\eta,t} = [\tilde{\sigma}_{1,t}, \ldots, \tilde{\sigma}_{n,t}]'$, and P^η and W_η are both diagonal $n \times n$ matrices.

[1] These assumption imply that, $e_{i,t}$ for $s = 1, 2$ are drawn from a multivariate Student's t distribution, conditional on $\tilde{\sigma}_{i,t}$.

[2] Note $\Phi' x_t = vec(x'_t\Phi I_n) = (I_n \otimes x'_t)vec(\Phi)$.

1.1 Identification, Reparameterization and Restrictions

Expression (3) distinguishes between non time-varying and time-varying component of the volatilities. As a consequence the scale of the latter cannot be separately estimated, which is why there is no constant in the law of motion of the volatilities (8) and (9). Moreover in the case of the factors even the non time-varying component of the volatility cannot be separately identified from the loadings B. Two are the solutions used in the literature: setting one of the loadings B to one, or setting the variance of the factor to one. We choose the latter, hence $\sigma_{i,0} = 1,\ i = n+1, .., n+k$. Rotational indeterminacy arises if we have more than one factor. We address that by setting some of the loadings B to zero.

1.2 Priors

This section describes the functional form of the priors. The actual choice of the prior parameters is discussed in the application. We use a conjugate prior for the VAR parameters φ, the factor loadings b, and the variances of the idiosyncratic innovations σ_0. This prior is a implemented using T^* dummy observations (see Sims and Zha (1998) or the review of BVARs in Del Negro and Schorfheide (2010)), which are generated by the constant volatility/Gaussian version of the VAR in (7):

$$y^*_{it} = x^*_{it}{}'\varphi_i + f^*_{it}{}'b_i + \eta_{it},\ \eta_t \sim \mathcal{N}(0, \sigma^2_{i,0}),\ \text{for } t = 1, .., T^* \tag{10}$$

where φ_i and b_i the elements of φ and b corresponding to the i^th equation of the VAR. Note that the dummy observations are potentially different across equations. In addition to the prior implicitly defined by (10), we also use a Jeffreys prior for the non time-varying components of the variances σ_0, independent across i. Therefore the joint prior on ϕ, b, and σ_0 is given by the likelihood of the dummy observations (10) times the Jeffreys prior:

$$p(\varphi, b, \sigma_0|\mu) \propto \Big(\Pi^n_{i=1}\sigma^{-2}_{i,0}\Big)\Big(c(\mu)^{-1}\prod_{t=1}^{T^*} p(y^*_t|x^*_t, f^*_t, \varphi, b, \sigma_0)\Big). \tag{11}$$

where the constant of integration $c(\mu)$ is such that prior integrates to one, that is:

$$c(\mu) = \prod_{i=1}^{n} \left((2\pi)^{-(T^* - \kappa_i)/2} |\underline{V}_i|^{\frac{1}{2}} |\underline{s}_i|^{-\frac{T^*-\kappa_i}{2}} 2^{\frac{(T^*-\kappa_i)}{2}} \Gamma[\frac{(T^* - \kappa_i)}{2}] \right) \tag{12}$$

where

$$\underline{V}_i = \left(\sum_{t=1}^{T^*} \tilde{x}^*_{i,t} \tilde{x}^{*\,\prime}_{i,t} \right)^{-1}, \ \underline{\beta}_i = \underline{V}_i \left(\sum_{t=1}^{T^*} \tilde{x}^*_{i,t} \tilde{y}^*_{i,t} \right), \underline{s}_i = \sum_{t=1}^{T^*} \tilde{y}^{*\,2}_{i,t} - \underline{\beta}_i' \underline{V}_i^{-1} \underline{\beta}_i, \tag{13}$$

$\tilde{x}^*_{i,t} = \left[x^{*\prime}_t f^{*\prime}_{i,t} \right]'$, $f^*_{i,t}$ is the subset of factors for which unit i has non-zero loading and κ_i is the size of $\tilde{x}^*_{i,t}$. Note that $p(\varphi, b, \sigma_0|\mu)$ depends on the vector μ of hyper-parameters describing the tightness of the Minnesota prior. Section **??** in the appendix provides details on the mapping between μ and the dummy observations $\{y^*_t, x^*_t, f^*_t\}_{t=1}^{T^*}$. Following Giannone et al. (2011), we conduct inference on the vector μ assuming independent Gamma priors over the elements of μ:

$$p(\mu) = \prod_{j=1}^{4} \frac{\left(\underline{\mu}_j / \underline{\nu}_{\mu,j} \right)^{-\underline{\nu}_{\mu,j}}}{\Gamma(\underline{\nu}_{\mu,j})} \lambda_j^{\underline{\nu}_{\mu,j}-1} \exp(-\underline{\nu}_{\mu,j} \frac{\lambda_j}{\underline{\mu}_j}). \tag{14}$$

where $\underline{\mu}_j$ is the mean and $\underline{\nu}_{\mu,j}$ is the number of degrees of freedom.

For the prior on the parameters λ_i we assume a Gamma distributions with parameters $\underline{\lambda}/\underline{\nu}$ and $\underline{\nu}$:

$$p(\lambda_i) = \frac{(\underline{\lambda}/\underline{\nu})^{-\underline{\nu}}}{\Gamma(\underline{\nu})} \lambda_i^{\underline{\nu}-1} exp(-\underline{\nu} \frac{\lambda_i}{\underline{\lambda}}), \text{ i.i.d. across } i, s. \tag{15}$$

where $\underline{\lambda}$ is the mean and $\underline{\nu}$ is the number of degrees of freedom (Geweke (2005) assumes a Gamma with one degree of freedom). The prior on P_f, Ξ, and the diagonal elements of P_η are Gaussian, times an indicator function that imposes stationarity:

$$p(vec(\mathrm{P}_f)|W_f) = \mathcal{N}\left(vec(\underline{\mathrm{P}}_f), W_f \otimes \underline{V}_f \right) \mathcal{I}(\mathrm{P}_f), \tag{16}$$

$$p(\mathrm{P}_\eta|W_\eta) = \prod_{i=1}^{n} \mathcal{N}\left(\underline{\rho}_\eta, w_i \underline{v}_\eta \right) \mathcal{I}(\rho_i), \tag{17}$$

where $\mathcal{I}(\mathrm{P}_f)$ and $\mathcal{I}(\rho_i)$ are zero if any of the roots of P_f and any of the ρ_is are larger than one, respectively. The prior on Ξ is also Gaussian:

$$p(vec(\Xi)|W_\eta) = \mathcal{N}\left(vec(\underline{\Xi}), W_\eta \otimes \underline{V}_\Xi \right). \tag{18}$$

The prior on W_f is an inverse-Wishart distribution with parameters $\underline{W}$ and ν_w:

$$p(W_f) = \frac{|\underline{W}|^{\nu_w/2}}{2^{k\nu_w/2}\Gamma(\nu_w/2)}|W_f|^{-(k+\nu_w+1)/2}\exp\left(-\frac{1}{2}\,\mathrm{tr}(W_f^{-1}\underline{W})\right), \tag{19}$$

where $\underline{W}$ is diagonal with identical elements $\nu_w\underline{w}$. Finally, the prior on $w_{1:n}$, the diagonal elements of W_η, is given by the product of independent inverse gamma distributions with parameters $\nu_w\underline{w}$ and ν_w:

$$p(W_\eta) = \prod_{i=1}^{n}\frac{(\nu_w\underline{w})^{\nu_w/2}}{2^{\nu_w/2}\Gamma(\nu_w/2)}w_i^{-(\nu_w+2)/2}\exp\left(-\frac{1}{2}w_i^{-1}\nu_w\underline{w}\right). \tag{20}$$

1.3 The Gibbs Sampler

The joint distribution of data and unobservables conditional on the initial observations $y_{-p+1:0}$, is given by:

$$\Bigg(\prod_{t=1}^{T}p(y_t|y_{t-p:t-1},f_t,\varphi,b,\sigma_0,\tilde{\sigma}_{1:n,t},\tilde{h}_{1:n,t})p(f_t|\tilde{\sigma}_{n+1:n+k,t},\tilde{h}_{n+1:n+k,t})$$
$$p(\tilde{h}_{1:n+k,t}|\lambda_{1:n+k})\Bigg)p(\tilde{\sigma}_{n+1:n+k,1:T}|\mathrm{P}_f,W_f)p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta,\Xi,W_\eta,\tilde{\sigma}_{n+1:n+k,1:T})$$
$$p(\varphi,b,\sigma_0|\mu)p(\mu)p(\lambda_{1:n+k})p(\mathrm{P}_\eta|W_\eta)p(\Xi|W_\eta)p(W_\eta)p(\mathrm{P}_f|W_f)p(W_f). \tag{21}$$

The term in the first row of expression (21) is the likelihood of y_t conditional on the $y_{t-p:t-1}$, the parameters, and the factors, which obtains from (7). Expression $p(f_t|,\tilde{\sigma}_{n+1:n+k,1:T},\tilde{h}_{n+1:n+k,t})$ is the distribution of the factors (6), $p(\tilde{h}_{1:n+k,t}|\lambda_{1:n+k})$ comes from (4), and the distributions of the stochastic volatilities of the factors and the idiosyncratic component $p(\tilde{\sigma}_{n+1:n+k,1:T}|\mathrm{P}_f,W_f)$ and $p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta,\Xi,W_\eta,\tilde{\sigma}_{n+1:n+k,1:T})$, respectively, are obtained from expressions (8) and (9). The prior distributions $p(\varphi,b,\sigma_0|\mu)$, $p(\mu)$, $p(\lambda_{1:n+k})$, $p(\mathrm{P}_\eta|W_\eta)$, $p(\Xi|W_\eta)$, $p(W_\eta)$, $p(\mathrm{P}_f|W_f)$, and $p(W_f)$ come from (11), (14), (15), (17), (18), (20), (16), and (19), respectively.

We conduct inference on the time varying volatilities using the method of Kim et al. (1998). This method involves approximating the distribution of $\log(\varepsilon_{i,t}^2)$ in expression (3) using a mixture of normal distribution as described in detail in appendix I.B. Call $\varsigma_{i,t}$ the indicators associated with these mixture and $\vartheta_1 =$

$\{\varphi, b, \sigma_0, \mu, \lambda_{1:n+k}, f_{1:T}, \tilde{h}_{1:n+k,1:T}, \tilde{h}_{1:n+k,1:T}\}$ all the remaining parameters/latent variables other than the stochastic volatilities $\tilde{\sigma}_{1:n+k,1:T}$ and the parameters entering the law of motion of $\tilde{\sigma}_{1:n+k,1:T}$, which we group under the label ϑ_2 ($\vartheta_2 = \{\Xi, \mathrm{P}_\eta, W_\eta, \mathrm{P}_f, W_f\}$). Our Gibbs sampler takes the following structure, which borrows from Del Negro and Primiceri (2013):[3]

I) draw $\vartheta_1, \varsigma_{1:T}$ from $y_{-p+1:T}, \vartheta_1, \varsigma_{1:T} | \vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$, which can be divided into two substeps:

   I.A) draw ϑ_1 from the marginal $\vartheta_1 | \vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$;

   I.B) draw $\varsigma_{1:T}$ from the conditional $\varsigma_{1:T} | \vartheta_1, \vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$.

II) draw $\tilde{\sigma}_{1:n+k,1:T}$ from $\tilde{\sigma}_{1:n+k,1:T} | \varsigma_{1:T}, \vartheta_1, \vartheta_2, y_{-p+1:T}$;

III) draw ϑ_2 from $\vartheta_2 | \tilde{\sigma}_{1:n+k,1:T}, \vartheta_1, \varsigma_{1:T}, y_{-p+1:T}$;

Draws of the elements of ϑ_1 conditional on $\vartheta_2, \tilde{\sigma}_{1:n+k,1:T}$, and $y_{-p+1:T}$ are obtained using a Gibbs-within-Gibbs approach, that is, drawing each element of ϑ_1 conditional on the others. Our Gibbs sampler steps borrow extensively from the existing literature, and in particular from Geweke (1993) and Geweke (2005) for inference with regard to Student's t shocks. We leave the details to appendix I.A. Steps I.B, II, and III are fairly standard, as they build on Kim et al. (1998), and are also described in the appendix (sections I.B, II, and III, respectively).

We conclude the section with a discussion of the computational advantages deriving from assuming a factor structure for the reduced form shocks u_t in the VAR (1). Notice that after conditioning on f_t the variance covariance matrix of the VAR innovations in (7) is diagonal, implying that the VAR can be estimated equation by equation. This makes it possible to efficiently estimate VARs with time-variation in volatility even for large n. Recall in fact that the posterior variance of the

[3] The alternative would be a four-block sampler with blocks $\tilde{\sigma}_{1:n+k,1:T}$, $(y_{-p+1:T}, \vartheta_1)$, ϑ_2, and $\varsigma_{1:T}$. However, in standard macro models it is often difficult to draw from $p(y_{-p+1:T}, \vartheta_1 | \tilde{\sigma}_{1:n+k,1:T}, \varsigma_{1:T}, \vartheta_2, y_{1:T})$ and consequently this sampler is difficult to implement.

VAR parameters (with a flat prior) is given by the expression $\left(\sum_{t=1}^{T} \Sigma_t^{-1} \otimes x_t x_t'\right)^{-1}$, where Σ_t is the covariance matrix of the innovations. When there is no time variation $\Sigma_t = \Sigma$ can be taken out of the summation and each of the elements of the Kronecker product can be inverted separately, leading to the standard formula $\Sigma \otimes \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}$. When Σ_t varies over time this is no longer possible, and one has to invert the $n(np+1) \times n(np+1)$ matrix $\left(\sum_{t=1}^{T} \Sigma_t^{-1} \otimes x_t x_t'\right)$. In our case, after conditioning on the factor the matrix $\Sigma_{\eta,t}$ is diagonal. Therefore $\sum_{t=1}^{T} \Sigma_{\eta,t}^{-1} \otimes x_t x_t'$ becomes a block diagonal matrix, and its inversion trivial. In fact, in this case the variance matrix of the VAR parameters can be computed using the standard GLS formula $\left(\sum_{t=1}^{T} \frac{x_t x_t'}{\sigma_{i,t}^2}\right)^{-1}$, equation by equation, making the computational cost of this step linear in n (and amenable to parallelization).

Of course, the additional cost relative to VARs with a standard innovation covariance matrix comes from computing the smoothed estimates of $f_t, t = 1, .., T$. The fact that the f_t is not autocorrelated makes this task fairly simple because it can be performed for each t separately, as we discuss later. Moreover, the application of standard factor analysis formulas imply that the size of n is not an hindrance.

2 Application

2.1 The Data

2.2 Prior Parameterization

.

....[NEED TO FILL UP PARAMS FOR OTHER PRIORS] ...

References

Del Negro, Marco and Frank Schorfheide, "Bayesian Macroeconometrics," in Herman K. van Dijk, Gary Koop, and John Geweke, eds., *Handbook of Bayesian Econometrics*, Oxford University Press, 2010.

Durbin, James and Siem Jan Koopman, "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 2002, *89* (3), 603–616.

Geweke, John, "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics,*, 1993, *8*, S19–S40.

_ , *Contemporary Bayesian Econometrics and Statistics*, Wiley, 2005.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri, "Prior Selection for Vector Autoregressions," *Mimeo, ULB*, 2011.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 1998, *65* (3), 361–393.

Lehmann, B. and D. Modest, "The Empirical Foundations of Arbitrage Pricing Theory I: The Empirical Tests," *NBER Working Paper*, 1985, *1725.*

Negro, Marco Del and Giorgio Primiceri, "Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum," *Federal Reserve Bank of New York Staff Reports*, 2013, *619.*

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima, "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 2007, *140*, 425449.

Sims, Christopher A. and Tao Zha, "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 1998, *39* (4), 949–968.

A Appendix

I.A Drawing $\vartheta_1|\vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$

Conditional on $\vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$, the elements of ϑ_1, where $\vartheta_1 = \{\varphi, b, \sigma_0, \mu, \lambda_{1:n+k}, f_{1:T}, \tilde{h}_{1:n+k,1:T}\}$, can be drawn using the following Gibbs-within-Gibbs steps:

I.A.1 $\mathbf{f_{1:T}}|\varphi, ..., \boldsymbol{\vartheta_2}, \tilde{\boldsymbol{\sigma}}_{\mathbf{1:n+k,1:T}}, \mathbf{y_{-p+1:T}}$: This is accomplished by drawing for each t from

$$p(y_t|y_{t-p:t-1}, \varphi, b, f_t, \sigma_0, \tilde{\sigma}_{1:n,t}, \tilde{h}_{1:n,t})p(f_t|\tilde{\sigma}_{n+1:n+k,t}, \tilde{h}_{n+1:n+k,t}).$$

This is the standard Gaussian filtering problem encountered in factor analysis, where $p(y_t|...)$ is the measurement equation and $p(f_t|...)$ the distribution of the unobservable leading to the formulas:

$$\begin{aligned}\bar{f}_t &= I\!\!E[f_t|y_t, \tilde{h}_{1:n+k,1:T}, \tilde{\sigma}_{1:n+k,1:T}] \\ &= \Sigma_{f,t}B'\left(B\Sigma_{f,t}B' + \Sigma_{\eta,t}^{-1}\right)^{-1}u_t \\ &= \left(\Sigma_{f,t}^{-1} + B'\Sigma_{\eta,t}^{-1}B\right)^{-1}B'\Sigma_{\eta,t}^{-1}u_t, \text{ and}\end{aligned}$$

$$\begin{aligned}\bar{V}_t^f &= var[f_t|y_t, \tilde{h}_{1:n+k,1:T}, \tilde{\sigma}_{1:n+k,1:T}] \\ &= \Sigma_{f,t} - \Sigma_{f,t}B'\left(B\Sigma_{f,t}B' + \Sigma_{\eta,t}^{-1}\right)^{-1}B\Sigma_{f,t} \\ &= \Sigma_{f,t} - \left(\Sigma_{f,t}^{-1} + B'\Sigma_{\eta,t}^{-1}B\right)^{-1}B'\Sigma_{\eta,t}^{-1}B\Sigma_{f,t},\end{aligned}$$

where $u_t = y_t - X_t'\varphi$ is the vector of residuals in equation (1), and where the last equality in each of the two equations turn the inversion of an $n \times n$ matrix into the inversion of a (smaller) $k \times k$ matrix, using the formulas popularized by Lehmann and Modest (1985). Therefore the distribution of the factors is given by:

$$f_t|u_t, \tilde{h}_{1:n+k,1:T}, \tilde{\sigma}_{1:n+k,1:T} \sim \mathcal{N}(\bar{f}_t, \bar{V}_t^f). \tag{A.2}$$

I.A.2 $\varphi, \mathbf{b}, \boldsymbol{\sigma_0}, \mu|\mathbf{f_{1:T}}, \boldsymbol{\lambda}_{\mathbf{1:n+k}}, \boldsymbol{\sigma_0}, ..., \mathbf{y_{-p+1:T}}$: The relevant part of the joint distribution of data and paramters (21) is given by:

$$\left(\prod_t p(y_t|y_{t-p:t-1}, \varphi, b, f_t, \sigma_0, \tilde{\sigma}_{1:n,t}, \tilde{h}_{1:n,t})\right)p(\varphi, b, \sigma_0|\mu)p(\mu). \tag{A.3}$$

As discussed above, the advantage of conditioning on the factors f_t is that since the $\eta_{i,t}$ are uncorrelated across i, ϕ_i and b_i – the i^{th} row of Φ' and B, respectively – can be drawn equation by equation, using standard GLS formulas. The VAR in (7) is composed by n regressions (one for each $i = 1,..,n$) of the form:

$$\begin{aligned} y_{i,t} &= x_t'\varphi_i + f_t'b_i + \eta_{1,i,t}, \\ &\quad \eta_{i,t} \sim N(0, \sigma_{i,0}^2 e^{2\tilde{\sigma}_{it}} \tilde{h}_{i,t}^{-1}), \quad \text{for } t = 1,..,T,..,T+T^*. \end{aligned} \tag{A.4}$$

where we used the notation $f_{i,t}$ to denote the subset of factors for which unit i has non-zero loadings, and b_i is the vector of these loadings.

II.A) $\varphi, \mathbf{b}, \sigma_{\mathbf{0}}, \mu|..., \mathbf{y}_{-\mathbf{p}+\mathbf{1}:\mathbf{T}}$: We draw from the joint distribution of $\{\varphi, b, \sigma_0, \mu\}$ conditional on the other parameters by first drawing from the marginal $p(\mu|...)$, then from the conditional $p(\sigma_0|\mu,...)$, and finally from the conditional $p(\varphi, b|\mu, \sigma_0, ...)$.

I.A.2.i.a $\mu|..., \mathbf{y}_{-\mathbf{p}+\mathbf{1}:\mathbf{T}}$: Call

$$\begin{aligned} \tilde{x}_{i,t} &= \begin{bmatrix} x_t \\ f_{i,t} \end{bmatrix} \text{ for } t = 1,..,T,..,T+T^*, \\ \breve{\sigma}_{i,t}^2 &= \begin{cases} e^{2\tilde{\sigma}_{it}} \tilde{h}_{i,t}^{-1} \text{ for } t = 1,..,T, \\ 1 \text{ for } t = T+1,..,T+T^*, \end{cases} \end{aligned}$$

and

$$\bar{V}_i = \left(\sum_{t=1}^{\bar{T}} \frac{\tilde{x}_{i,t}\tilde{x}_{i,t}'}{\breve{\sigma}_{i,t}^2} \right)^{-1}, \ \bar{\beta}_i = \bar{V}_i \left(\sum_{t=1}^{\bar{T}} \frac{\tilde{x}_{i,t} y_{i,t}}{\breve{\sigma}_{i,t}^2} \right),$$

$$\bar{s}_i = \sum_{t=1}^{\bar{T}} \frac{y_{i,t}^2}{\breve{\sigma}_{i,t}^2} - \bar{\beta}_i' \bar{V}_i^{-1} \bar{\beta}_i, \ \bar{T} = T + T^*. \tag{A.5}$$

If we integrate out expression A.3 with respect to ϕ, b, and σ_0 we obtain the (conditional) posterior distribution of μ:

$$\left(\prod_{i=1}^{n} \left((2\pi)^{-(\bar{T}-\kappa_i)/2} |\bar{V}_i|^{\frac{1}{2}} |\bar{s}_i|^{-\frac{\bar{T}-\kappa_i}{2}} 2^{\frac{(\bar{T}-\kappa_i)}{2}} \Gamma[\frac{(\bar{T}-\kappa_i)}{2}] \right) \right) c(\mu)^{-1} p(\mu), \tag{A.6}$$

which we draw from using a Metropolis-Hastings step. Note that

$$\left(\prod_{i=1}^{n}\left((2\pi)^{-(\bar{T}-\kappa_i)/2}|\bar{V}_i|^{\frac{1}{2}}|\bar{s}_i|^{-\frac{\bar{T}-\kappa_i}{2}}2^{\frac{(\bar{T}-\kappa_i)}{2}}\Gamma[\frac{(\bar{T}-\kappa_i)}{2}]\right)\right)c(\mu)^{-1} = \frac{\prod_{i=1}^{n}\left((2\pi)^{-(\bar{T}-\kappa_i)/2}|\bar{V}_i|^{\frac{1}{2}}|\bar{s}_i|^{-\frac{\bar{T}-\kappa_i}{2}}2^{\frac{(\bar{T}-\kappa_i)}{2}}\Gamma[\frac{(\bar{T}-\kappa_i)}{2}]\right)}{\prod_{i=1}^{n}\left((2\pi)^{-(T^*-\kappa_i)/2}|\underline{V}_i|^{\frac{1}{2}}|\underline{s}_i|^{-\frac{T^*-\kappa_i}{2}}2^{\frac{(T^*-\kappa_i)}{2}}\Gamma[\frac{(T^*-\kappa_i)}{2}]\right)}$$

which is the standard marginal likelihood formula (see Del Negro and Schorfheide (2010)) for BVARs where the prior is implemented via dummy observation except that in this case the VAR innovations are cross-sectionally independent.

I.A.2.i.b $\sigma_{\mathbf{0}}|\mu, ..., \mathbf{y_{-p+1:T}}$: Conditional on the μ draw obtained from the previous step, compute $\tilde{x}_{i,t}$, $\bar{V}_i$, $\bar{\beta}_i$, and $\bar{s}_i$ as described above and draw $\sigma_{i,0}^2$ for $i = 1, .., n$ from

$$\sigma_{i,0}^2|\ldots, s \sim \mathcal{IG}(\bar{s}_i, \bar{T} - \kappa_i).$$

I.A.2.i.c $\varphi, \mathbf{b}|\mu, \sigma_{\mathbf{0}}, ..., \mathbf{y_{-p+1:T}}$: Draw $\{\varphi_i, b_i\}$ from

$$\begin{bmatrix} \varphi_i \\ b_i \end{bmatrix} \sim \mathcal{N}(\bar{\beta}_i, \sigma_{0,i}^2\bar{V}_i), \text{ for } i = 1, .., n, \tag{A.7}$$

where $\bar{\beta}_i$ and $\bar{V}_i$ are also defined as above.

I.A.3 $\tilde{\mathbf{h}}_{\mathbf{1:n+k,1:T}}|\varphi, \mathbf{b}, \mathbf{f_{1:T}}, \lambda_{\mathbf{1:n+k}}, ..., \mathbf{y_{-p+1:T}}$: The posterior of the latent variables $\tilde{h}_{1:n+k,1:T}$, conditional on all other parameter is proportional to

$$\left(\prod_{t=1}^{T} p(y_t|y_{t-p:t-1}, \varphi, b, f_t, \sigma_0, \tilde{\sigma}_{1:n,t}, \tilde{h}_{1:n,t}) \right. \left. p(f_t|\tilde{\sigma}_{n+1:n+k,t}, \tilde{h}_{n+1:n+k,t})p(\tilde{h}_{1:n+k,t}|\lambda_{1:n+k})\right).$$

Expression (4) implies that the prior on $\tilde{h}_{i,t}$ has the form

$$p(\tilde{h}_{i,t}|\lambda_i) = \left(2^{\lambda_i/2}\Gamma(\lambda_i/2)\right)^{-1}\lambda_i^{\lambda_i/2}\tilde{h}_{i,t}^{(\lambda_i-2)/2}\exp(-\lambda_i\tilde{h}_{i,t}/2).$$

Conditional on φ, B, and f_t, each element $e_{i,t}$ of the vector of shocks $[\eta_t', f_t']'$ is known. From (3), its distribution is:

$$p(e_{i,t}|\sigma_{i,0}, \tilde{\sigma}_{i,t}, \tilde{h}_{i,t}) \propto \frac{\tilde{h}_{i,t}^{1/2}}{\sigma_{i,0}e^{\tilde{\sigma}_{i,t}}}\exp\left(-e_{i,t}^2\frac{\tilde{h}_{i,t}}{\sigma_{i,0}^2e^{2\tilde{\sigma}_{i,t}}}\right), \tag{A.8}$$

where $\sigma_{i,0} = 1$ for the factors ($i = n+1, .., n+k$). Therefore

$$p(e_{i,t}|\sigma_{i,0}, \tilde{\sigma}_{i,t}, \tilde{h}_{i,t})p(\tilde{h}_{i,t}|\lambda_i) \propto \tilde{h}_{i,t}^{(\lambda_i^- 1)/2} \exp(-\left[\lambda_i + \frac{e_{i,t}^2}{\sigma_{i,0}^2 e^{2\tilde{\sigma}_{i,t}}}\right] \tilde{h}_{i,t}/2),$$

which implies

$$\left[\lambda_i + \frac{e_{i,t}^2}{\sigma_{i,0}^2 e^{2\tilde{\sigma}_{i,t}}}\right] \tilde{h}_{i,t}|\tilde{\sigma}_{i,t}, \lambda_i, ... \sim \chi^2(\lambda_i + 1),$$

independently across i and t.

I.A.4 $\lambda_{\mathbf{1:n+k}}|\tilde{\mathbf{h}}_{\mathbf{1:n+k,1:T}}, \varphi, \mathbf{b}, \mathbf{f}_{\mathbf{1:T}}, ..., \mathbf{y}_{\mathbf{-p+1:T}}$: The product of the relevant elements of (21) amounts to:

$$p(\tilde{h}_{1:T}|\lambda_{1:n+k})p(\lambda_{1:n+k}) \quad \propto \quad ((\underline{\lambda}/\underline{\nu})^{\underline{\nu}}\Gamma(\underline{\nu}))^{-1} \left[2^{\lambda_i/2}\Gamma(\lambda_i/2)\right]^{-T} \lambda_i^{T\lambda_i/2+\underline{\nu}-1} \left(\prod_{t\in\mathcal{T}} \tilde{h}_{i,t}^{(\lambda_i-2)/2}\right) \exp\left[-\left(\frac{\underline{\nu}}{\underline{\lambda}} + \frac{1}{2}\sum_{t\in\mathcal{T}} \tilde{h}_{i,t}\right)\lambda_i\right].$$

This is a non-standard distribution, hence the draw is obtained from a Metropolis-Hastings step. Following Geweke (2005), we use a log-normal proposal density.

I.B Drawing $\varsigma_{1:n+k,1:T}|\vartheta_1, \vartheta_2, \tilde{\sigma}_{1:n+k,1:T}, y_{-p+1:T}$

Recall that $e_{\bar{s}_t,i,t}$ is an element of the vector $[\eta_t', f_t']'$, which we can treat as known conditional on $\vartheta_1, y_{-p+1:T}$ (note that we only consider those innovations associated with the realized state $\bar{s}_t$ at time t). Taking squares and then logs of (3) one obtains:

$$e_{i,t}^* = 2\tilde{\sigma}_{i,t} + \varepsilon_{i,t}^* \tag{A.9}$$

where $e_{i,t}^* = \log(\frac{\tilde{h}_{i,t}}{\sigma_{\bar{s}_t,i,0}^2} e_{\bar{s}_t,i,t}^2 + c)$, $c = .001$ being an offset constant, and $\varepsilon_{i,t}^* = \log(\varepsilon_{\bar{s}_t,i,t}^2)$ is distributed as a $\log(\chi_1^2)$. Kim et al. (1998) approximate the $\log(\chi_1^2)$ with a mixture of normal that i expressing the distribution of $\varepsilon_{i,t}^*$ as:

$$p(\varepsilon_{i,t}^*) = \sum_{k=1}^{K} \pi_k^* \mathcal{N}(m_k^* - 1.2704, \nu_k^{*\,2}) \tag{A.10}$$

We follow Omori et al. (2007) in using a 10-mixture approximation, as opposed to the 7-mixture approximation adopted in Kim et al. (1998) The parameters that optimize this approximation, namely $\{\pi_k^*, m_k^*, \nu_k^*\}_{k=1}^K$ and K, are given inOmori et al. (2007), and are independent of the specific application. The mixture of normals can be equivalently expressed as:

$$\varepsilon_{i,t}^*|\varsigma_{i,t} = k \sim \mathcal{N}(m_k^* - 1.2704, \nu_k^{*\,2}), \; Pr(s_{i,t} = k) = \pi_k^*. \tag{A.11}$$

Hence we implement $\varsigma_{1:n+k,1:T}|f_{1:T}, \varphi, b, \varsigma_{1:n+k,1:T}, y_{-p+1:T}$ by drawing from

$$Pr\{\varsigma_{i,t} = k|\tilde{\sigma}_{i,1:T}, e_{i,1:T}^*\} \propto \pi_k^* \nu_k^{-1} \exp\left[-\frac{1}{2\nu_k^{*\,2}}(\varepsilon_{i,t}^* - m_k^* + 1.2704)^2\right]. \tag{A.12}$$

where from (A.9) $\varepsilon_{i,t}^* = e_{i,t}^* - 2\tilde{\sigma}_{i,t}$.

II Drawing $\tilde{\sigma}_{1:n+k,1:T}|\vartheta_1, \vartheta_2, \varsigma_{1:n+k,1:T}, y_{-p+1:T}$

Because of the factor structure of the law of motion of the stochastic volatilitie here we also adopt a Gibbs-within-Gibbs approach. We draw the factor volatilities $\tilde{\sigma}_{n+1:n+k,1:T}$ conditional on the idiosyncratic volatilities $\tilde{\sigma}_{1:n,1:T}$, and viceversa.

II.1 The law of motion of $\tilde{\sigma}_{n+1:n+k,1:T}$ (8) can be rewritten as:

$$\begin{bmatrix} \tilde{\sigma}_{f,t} \\ \tilde{\sigma}_{f,t-1} \end{bmatrix} = \begin{bmatrix} \mathrm{P}_f' & 0 \\ I & 0 \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_{f,t-1} \\ \tilde{\sigma}_{f,t-2} \end{bmatrix} + \begin{bmatrix} \zeta_{f,t} \\ 0 \end{bmatrix} \tag{A.13}$$

where recall that $\tilde{\sigma}_{f,t} = [\tilde{\sigma}_{n+1,t}, \ldots, \tilde{\sigma}_{n+k,t}]'$ and $\tilde{\sigma}_{\eta,t} = [\tilde{\sigma}_{1,t}, \ldots, \tilde{\sigma}_{n,t}]'$. To this transition equations is associated a system of measurement equations composed by (A.9) for $e_{n+1:n+k,1:T}^*$ and the law of motion of the idiosyncratic volatiles (9) , which we rewrite as

$$\tilde{\sigma}_{\eta,t} = \mathrm{P}_\eta' \tilde{\sigma}_{\eta,t-1} + \begin{bmatrix} \Xi' & -\Xi'\mathrm{P}_f' \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_{f,t} \\ \tilde{\sigma}_{f,t-1} \end{bmatrix} + \zeta_{\eta,t}. \tag{A.14}$$

Draws of $\xi_t = \left[\tilde{\sigma}_{f,t}', \tilde{\sigma}_{f,t-1}'\right]'$ are obtained using state-space methods as described in Durbin and Koopman (2002).

II.2 Drawing the idiosyncratic volatilities is much simpler, and can be done separately for each $i = 1, .., n$ using the appropriate row of (9) as the transition equation and (A.9) for $e_{i,1:T}^*$ as the measurement equation.

III Drawing $\vartheta_2|\tilde{\sigma}_{1:n+k,1:T}, \vartheta_1, \varsigma_{1:n+k,1:T}, y_{-p+1:T}$

Drawing ϑ_2, the parameters of the law of motion of the stochastic volatilities is also accomplished using a sequence of Gibbs-within-Gibbs steps.

III.1 $\mathrm{P}_{\mathbf{f}}|\mathbf{W_f}, \mathrm{P}_\eta, \mathbf{\Xi}, \mathbf{W}_\eta, \tilde{\sigma}_{\mathbf{1:n+k,1:T}}, \vartheta_{\mathbf{1}}, \varsigma_{\mathbf{1:n+k,1:T}}, \mathbf{y_{-p+1:T}}$: Drawing P_f is complicated by the fact that it not only enters the law of motion of the factor stochastic volatilities (8), which we can rewrite as

$$p(\tilde{\sigma}_{n+1:n+k,1:T}|\mathrm{P}_f, W_f) \propto$$
$$|W_f|^{-T/2} exp\left[-\sum_{t=1}^{T}(\tilde{\sigma}_{f,t} - \mathrm{P}'_f\tilde{\sigma}_{f,t-1})'W_f^{-1}(\tilde{\sigma}_{f,t} - \mathrm{P}'_f\tilde{\sigma}_{f,t-1})/2\right], \quad \text{(A.15)}$$

but also the law of motion of the idiosyncratic stochastic volatilities (9):

$$p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta, W_\eta, \Xi, \mathrm{P}_f, \tilde{\sigma}_{n+1:n+k,1:T}) \propto$$
$$|W_\eta|^{-T/2} exp\Bigg[-\sum_{t=1}^{T}(\tilde{\sigma}_{\eta,t} - \mathrm{P}'_\eta\tilde{\sigma}_{\eta,t-1} - \Xi'(\tilde{\sigma}_{f,t} - \mathrm{P}'_f\tilde{\sigma}_{f,t-1}))'$$
$$W_\eta^{-1}(\tilde{\sigma}_{\eta,t} - \mathrm{P}'_\eta\tilde{\sigma}_{\eta,t-1} - \Xi'(\tilde{\sigma}_{f,t} - \mathrm{P}'_f\tilde{\sigma}_{f,t-1}))/2\Bigg]. \quad \text{(A.16)}$$

Since P_f enters this last expression in a non standard way, we use a Metropolis-within-Gibbs step, where the product of (8) and the Gaussian prat of the prior (16) form the proposal density, which is therefore:

$$p(vec(\mathrm{P}_f)|W_f, \tilde{\sigma}_{n+1:n+k,1:T}) \propto \mathcal{N}\left(vec(\bar{\mathrm{P}}_f), W_f \otimes \bar{V}_f\right), \quad \text{(A.17)}$$

where

$$\bar{V}_f = \left(\sum_{t=1}^{T}\tilde{\sigma}_{f,t-1}\tilde{\sigma}'_{f,t-1} + \underline{V}_f^{-1}\right)^{-1}, \ \bar{\mathrm{P}}_f = \bar{V}_f\left(\sum_{t=1}^{T}\tilde{\sigma}_{f,t-1}\tilde{\sigma}'_{f,t} + \underline{V}_f^{-1}\underline{P}_f\right).$$

The acceptance probability is given by[4]

$$\alpha = \frac{p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta, W_\eta, \Xi, \mathrm{P}^*_f, \tilde{\sigma}_{n+1:n+k,1:T})\mathcal{I}(\mathrm{P}^*_f)}{p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta, W_\eta, \Xi, \mathrm{P}_f^{(j-1)}, \tilde{\sigma}_{n+1:n+k,1:T})\mathcal{I}(\mathrm{P}_f^{(j-1)})},$$

where P^*_f and $\mathrm{P}_f^{(j-1)}$ represent the candidate and the previous draw of P_f, respectively.

[4]Note that the other terms cancel out.

III.2 $\mathbf{W_f}|\mathrm{P}_\mathbf{f}, \mathrm{P}_\eta, \mathbf{W}_\eta, \tilde{\sigma}_{\mathbf{1:n+k,1:T}}, \vartheta_\mathbf{1}, \varsigma_{\mathbf{1:n+k,1:T}}, \mathbf{y}_{\mathbf{-p+1:T}}$: The product of expressions (8) and (19), the prior on W_f, yields the posterior:

$$p(W_f|\mathrm{P}_f, \tilde{\sigma}_{n+1:n+k,1:T}) \propto |W|^{-(k+\nu_w+T+1)/2} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[W^{-1}\left(\underline{W} + \bar{S}_f\right)\right]\right\}, \quad \text{(A.18)}$$

where $\bar{S}_f = \sum_{t=1}^{T} (\tilde{\sigma}_{f,t} - \mathrm{P}'_f \tilde{\sigma}_{f,t-1})(\tilde{\sigma}_{f,t} - \mathrm{P}'_f \tilde{\sigma}_{f,t-1})'$, implying that $W_f|P_f, \tilde{\sigma}_{n+1:n+k,1:T} \sim \mathcal{IW}(\underline{W} + \bar{S}_f, \nu_w + T)$, where $\mathcal{IW}$ is an inverse Wishart distribution.

III.3 $\mathrm{P}_\eta, \mathbf{\Xi}, \mathbf{W}_\eta|\tilde{\sigma}_{\mathbf{1:n+k,1:T}}, \mathrm{P}_\mathbf{f}, \mathbf{W_f}, \vartheta_\mathbf{1}, \varsigma_{\mathbf{1:n+k,1:T}}, \mathbf{y}_{\mathbf{-p+1:T}}$: The law of motion of the idiosyncratic volatilities (9) can be further rewritten as

$$p(\tilde{\sigma}_{1:n,1:T}|\mathrm{P}_\eta, \Xi, w_i, \mathrm{P}_f, \tilde{\sigma}_{n+1:n+k,1:T}) \propto \prod_i w_i^{-T/2} exp\left[-\sum_{t=1}^{T} (\tilde{\sigma}_{\eta,t} - \gamma'_i \tilde{x}_{i,t})^2 / 2w_i\right].$$

where $\gamma_i = [\rho_i \ \Xi'_i]$ stacks ρ_i and the i^{th} column of Ξ and $\tilde{x}_{i,t} = [\tilde{\sigma}_{\eta,t-1} \ (\tilde{\sigma}_{f,t} - \mathrm{P}_f \tilde{\sigma}_{f,t-1})']'$. The product of this distribution and the priors on ρ_i and Ξ (expressions (16) and (19)) yield standard Gaussian/Inverse Gamma posteriors for the regression coefficients and the variance of the innovations. Specifically we obtain

$$p(\gamma_i|w_i, \mathrm{P}_f, \tilde{\sigma}_{1:n+k,1:T}) \propto \mathcal{N}\left(\bar{\gamma}_i), w_i \bar{V}_{\gamma,i}\right), \quad \text{(A.19)}$$

where

$$\bar{V}_{\gamma,i} = \left(\sum_{t=1}^{T} \tilde{x}_{i,t} \tilde{x}'_{i,t} + \underline{V}_\gamma^{-1}\right)^{-1}, \ \bar{\gamma}_i = \bar{V}_f \left(\sum_{t=1}^{T} \tilde{x}_{i,t} \tilde{\sigma}_{i,t} + \underline{V}_\gamma^{-1} \underline{\gamma}\right),$$

$$\underline{\gamma} = \begin{bmatrix} \underline{\rho} \\ \underline{\Xi}_i \end{bmatrix}, \ \underline{V}_\gamma = \begin{bmatrix} \underline{v}_\eta & 0 \\ 0 & \underline{V}_\Xi \end{bmatrix}$$

where $\underline{\Xi}_i$ is the i^{th} column of $\underline{\Xi}$, and

$$p(w_i|\mathrm{P}_f, \tilde{\sigma}_{1:n+k,1:T}) \propto w_i^{-(\nu_w+T+2)/2} \exp\left(-\frac{1}{2} w_i^{-1}\left(\bar{s}_i + \nu_w \underline{w}\right)\right), \quad \text{(A.20)}$$

where $\bar{s}_i = \sum_{t=1}^{T} (\tilde{\sigma}_{\eta,t} - \gamma'_i \tilde{x}_{i,t})^2$, implying that $w_i|\mathrm{P}_f, \tilde{\sigma}_{1:n+k,1:T} \sim \mathcal{IG}(\underline{W} + \bar{s}_i, \nu_w + T)$, where $\mathcal{IG}$ is an inverse Gamma distribution.