Supplementary Material

# Supplement to "Estimating overidentified, nonrecursive, time-varying coefficients structural variable autoregressions"

(*Quantitative Economics*, Vol. 6, No. 2, July 2015, 359–384)

Fabio Canova
EUI and CEPR

Fernando J. Pérez Forero
BCRP

## Appendix A: Global identification of the constant coefficients SVAR

Consider the constant coefficients version of the SVAR model used in Section 5:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11} & 0 \\ \alpha_3 & \alpha_6 & 0 & \alpha_9 & 1 & 0 \\ \alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1 \end{bmatrix}}_{A(\alpha)} \times \begin{bmatrix} \mathrm{GDP}_t \\ P_t \\ U_t \\ R_t \\ M_t \\ P\mathrm{com}_t \end{bmatrix} = A^+(L) \begin{bmatrix} \mathrm{GDP}_{t-1} \\ P_{t-1} \\ U_{t-1} \\ R_{t-1} \\ M_{t-1} \\ P\mathrm{com}_{t-1} \end{bmatrix} + \Sigma \begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^p \\ \varepsilon_t^u \\ \varepsilon_t^{\mathrm{mp}} \\ \varepsilon_t^{\mathrm{md}} \\ \varepsilon_t^i \end{bmatrix}$$

with

$$\Sigma = \begin{bmatrix} \sigma^i & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^{\mathrm{md}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^{\mathrm{mp}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^y & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^p & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^u \end{bmatrix}.$$

To verify that the system is globally identified, we rewrite the model using the notation of Rubio Ramirez et al. (2010). Let $y_t \equiv (\mathrm{GDP}_t, P_t, U_t, R_t, M_t, P\mathrm{com}_t)'$ and $\varepsilon_t \equiv (\varepsilon_t^y, \varepsilon_t^p, \varepsilon_t^u, \varepsilon_t^{\mathrm{mp}}, \varepsilon_t^{\mathrm{md}}, \varepsilon_t^i)'$. Premultiplying by $\Sigma^{-1}$, we obtain

$$\Sigma^{-1} A(\alpha) y_t = \Sigma^{-1} A^+(L) y_{t-1} + \varepsilon_t$$

Fabio Canova: canova@eui.eu
Fernando J. Pérez Forero: fernando.perez@bcrp.gob.pe


DOI: 10.3982/QE305

with $\varepsilon_t \sim N(0, I_6)$. Define $\mathbf{A}_0' \equiv \Sigma^{-1} A(\alpha)$ and $\mathbf{A}'(L) \equiv \Sigma^{-1} A^{+}(L)$. Then

$$y_t'\mathbf{A}_0 = \sum_{L=1}^{p} y_{t-L}'\mathbf{A}_L + \varepsilon_t',$$

where

$$\mathbf{A}_0' = \begin{bmatrix} \sigma^y & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^p & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^u & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^{\mathrm{mp}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^{\mathrm{md}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^i \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 & 0 & 0 \\ \alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \alpha_{11} & 0 \\ \alpha_3 & \alpha_6 & 0 & \alpha_9 & 1 & 0 \\ \alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{\sigma^y} & 0 & 0 & 0 & 0 & 0 \\ \frac{\alpha_1}{\sigma^p} & \frac{1}{\sigma^p} & 0 & 0 & 0 & 0 \\ \frac{\alpha_2}{\sigma^u} & \frac{\alpha_5}{\sigma^u} & \frac{1}{\sigma^u} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma^{\mathrm{mp}}} & \frac{\alpha_{11}}{\sigma^{\mathrm{mp}}} & 0 \\ \frac{\alpha_3}{\sigma^{\mathrm{md}}} & \frac{\alpha_6}{\sigma^{\mathrm{md}}} & 0 & \frac{\alpha_9}{\sigma^{\mathrm{md}}} & \frac{1}{\sigma^{\mathrm{md}}} & 0 \\ \frac{\alpha_4}{\sigma^i} & \frac{\alpha_7}{\sigma^i} & \frac{\alpha_8}{\sigma^i} & \frac{\alpha_{10}}{\sigma^i} & \frac{\alpha_{12}}{\sigma^i} & \frac{1}{\sigma^i} \end{bmatrix}.$$

Denoting $\mathbf{A}_0 = [a_{kj}]$, we have

$$\mathbf{A}_0 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\ 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \end{bmatrix}.$$

The matrices $\mathbf{Q}_j$, $j = 1, \dots, 6$, present in Theorem 1 of Rubio Ramirez et al. (2010) are

$$\mathbf{Q}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{Q}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{Q}_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{Q}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{Q}_5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{Q}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Define the matrices

$$\mathbf{M}_j(\mathbf{A}_0) = \begin{bmatrix} \mathbf{Q}_j \mathbf{A}_0 \\ [\,\mathbf{I}_j \quad \mathbf{0}_{j\times(M-j)}\,] \end{bmatrix}, \quad j = 1, \ldots, M, \tag{A.1}$$

so that

$$\mathbf{M}_1 = \begin{bmatrix} 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_3 = \begin{bmatrix} 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{M}_4 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\ 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_5 = \begin{bmatrix} 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad \mathbf{M}_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since all $\mathbf{M}_j$ have full column rank, the model is globally identified.

## Appendix B: Single-move Metropolis for drawing $B_t$

Koop and Potter's (2011) approach for drawing the elements of the $B^T$ sequence separately works as follows. Given $(f^{i-1})^T, (\Sigma^{i-1})^T, Q^{i-1}, V^{i-1}, W^{i-1}$, the measurement equation is

$$y_t = X_t' B_t + A(\alpha_t)^{-1}\Sigma_t \varepsilon_t$$

and the transition equation is

$$B_t = B_{t-1} + v_t,$$

with $v_t \sim N(0, Q)$, $B_0$ given, and $A(\alpha_t)^{-1}\Sigma_t\varepsilon_t = u_t \sim N(0, \Omega_t)$. To sample the individual elements of $B^T$, all $t \geq 1$, the steps are as follows.

1. Draw a candidate $B_t^\dagger \sim N(\mu_t, \Psi_t)$, where

$$\mu_t = \begin{cases} \dfrac{B_{t-1}^i + B_{t+1}^{i-1}}{2} + G_t\left[y_t - X_t'\left(\dfrac{B_{t-1}^i + B_{t+1}^{i-1}}{2}\right)\right], & t < T, \\ B_{t-1}^i + G_t\left[y_t - X_t'\left(B_{t-1}^i\right)\right], & t = T, \end{cases}$$

$$G_t = \begin{cases} \dfrac{1}{2} Q^{i-1} X_t\left(X_t' Q^{i-1} X_t + \Omega_t\right)^{-1}, & t < T, \\ Q^{i-1} X_t\left(X_t' Q^{i-1} X_t + \Omega_t\right)^{-1}, & t = T, \end{cases}$$

$$\Psi_t = \begin{cases} \dfrac{1}{2}\left(I_K - G_t X_t'\right) Q^{i-1}, & t < T, \\ \left(I_K - G_t X_t'\right) Q^{i-1}, & t = T. \end{cases}$$

2. Construct the companion form matrix $\overline{\mathbf{B}}_t^\dagger$ and evaluate $\mathcal{I}(\max|\operatorname{eig}(\overline{\mathbf{B}}_t^\dagger)| < 1)$, where $\mathcal{I}(\cdot)$ is an indicator function taking the value of 1 if the condition within the parentheses is satisfied.

3. The acceptance rate of $B_t^\dagger$ is

$$\omega_{B,t} = \min\left\{ \frac{\dfrac{\mathcal{I}(\max|\operatorname{eig}(\overline{\mathbf{B}}_t^\dagger)| < 1)}{\lambda(B_t^\dagger, Q^{i-1})}}{\dfrac{1}{\lambda(B_t^{i-1}, Q^{i-1})}}, 1 \right\}$$

$$= \min\left\{ \frac{\mathcal{I}(\max|\operatorname{eig}(\overline{\mathbf{B}}_t^\dagger)| < 1)\lambda(B_t^{i-1}, Q^{i-1})}{\lambda(B_t^\dagger, Q^{i-1})}, 1 \right\},$$

where $\lambda(\cdot)$ is an integrating constant measuring the proportion of draws that satisfy the inequality constraint. To compute $\lambda(\cdot)$, first one draws $B_t^{\dagger,l} \sim N(B_t^\dagger, Q^{i-1})$ for $l = 1, \ldots, \overline{L}$, constructs the companion form matrix $\overline{\mathbf{B}}_t^{\dagger,l}$, and evaluates $\lambda_l =$

$\mathcal{I}(\max|\operatorname{eig}(\overline{\mathbf{B}}_t^{\dagger,l})| < 1)$. Second, one evaluates $\lambda(B_t^\dagger, Q^{i-1}) = \frac{\sum_{l=1}^{\overline{L}} \lambda_l}{\overline{L}}$ and $\lambda(B_{t,i-1}, Q^{i-1})$, and computes the acceptance probability. When $t = T$, this probability is

$$\omega_{B,T} = \mathcal{I}\big(\max\big|\operatorname{eig}\big(\overline{\mathbf{B}}_t^\dagger\big)\big| < 1\big).$$

4. Draw a $v \sim U(0,1)$. Set $B_t^i = B_t^c$ if $v < \omega_{B,t}$ and set $B_t^i = B_t^{i-1}$ otherwise.

Since $Q$ depends on $B_t$, we need to change the sampling scheme also for this matrix. Assume that $Q^{-1} \sim W(\underline{v}, \underline{Q}^{-1})$ so that the unrestricted posterior is $Q^{-1} \sim W(\overline{v}, \overline{Q}^{-1})$ with $\overline{v} = \underline{v} + T$ and $\overline{Q}^{-1} = [\underline{Q} + \sum_{t=1}^T (B_{t,i} - B_{t-1,i})(B_{t,i} - B_{t-1,i})']^{-1}$. Then draw a candidate $(Q^\dagger)^{-1} \sim W(\overline{v}, \overline{Q}^{-1})$, and for $t = 1, \dots, T$, evaluate $\lambda(B_t^i, Q^\dagger)$ and $\lambda(B_t^i, Q^{i-1})$ for a fixed $\overline{L}$, and calculate

$$\omega_Q = \min\left\{\prod_{t=1}^T \frac{\lambda(B_t^i, Q^{i-1})}{\lambda(B_t^i, Q^\dagger)}, 1\right\}.$$

Finally, we draw a $v \sim U(0,1)$, set $Q^i = Q^c$ if $v < \omega_Q$, and set $Q^i = Q^{i-1}$. In the exercise of Section 5, we set $\overline{L} = 25$ when evaluating the integrating constants $\lambda(\cdot)$ at each $t$.

Note that in a multi-move approach, $\lambda(\cdot) = 1$ when sampling both $B^T$ and $Q$. Therefore, Koop and Potter's approach generalizes the multi-move procedure at the cost of making convergence to the posterior, in general, much slower and, because $\lambda(\cdot)$ needs to be simulated at each $t$, of adding considerable computational time.

## Appendix C: A shrinkage approach to draw $B^T$ when $\Xi$ is known

The model still consists of

$$\begin{aligned}
y_t &= X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \\
\alpha_t &= \alpha_{t-1} + \zeta_t, \\
\log(\sigma_{m,t}) &= \log(\sigma_{m,t-1}) + \eta_{m,t},
\end{aligned}$$

but now

$$B_t = B_{t-1} + v_t$$

is substituted by

$$B_t = \Xi \theta_t + v_t, \quad v_t \sim N(0, I), \tag{C.1}$$

$$\theta_t = \theta_{t-1} + \rho_t, \quad \rho_t \sim N(0, Q), \tag{C.2}$$

where $\dim(\theta_t) \ll \dim(B_t)$ and where the matrix $\Xi$ is known, as in Canova and Ciccarelli (2009). Using (C.2) into (C.1), we have

$$y_t = X_t' \Xi \theta_t + A(\alpha_t)^{-1} \Sigma_t \varepsilon_t + X_t' v_t \equiv X_t' \Xi \theta_t + \psi_t, \tag{C.3}$$

where $\psi_t \sim N(0, H_t)$ with $H_t \equiv A(\alpha_t)^{-1} \Sigma_t \Sigma_t' (A(\alpha_t)^{-1})' + X_t' X_t$.

To estimate the unknowns, we do the following:

1. We sample $\theta^T$ with a multi-move routine using (C.3) and (C.2).

2. Given $\theta^T$, we compute $\widehat{y}_t = y_t - X_t'\Xi\theta_t$. Pre-multiplying by $A(\alpha_t)$, we get the concentrated structural model

$$A(\alpha_t)\widehat{y}_t = A(\alpha_t)\xi_t = \Sigma_t\varepsilon_t + A(\alpha_t)X_t'\upsilon_t.$$

As before,

$$\left(\widehat{y}_t' \otimes I_M\right)(S_A f_t + s_A) = \Sigma_t\varepsilon_t + A(\alpha_t)X_t'\upsilon_t$$

so that the second state space system is

$$\tilde{y}_t = Z_t f_t + \Sigma_t\varepsilon_t + A(\alpha_t)X_t'\upsilon_t, \tag{C.4}$$

$$f_t = f_{t-1} + \zeta_t \tag{C.5}$$

and we draw $f^T$ using our proposed Metropolis step. The variance of the measurement error is $\Sigma_t\Sigma_t' + A_t(\alpha_t)X_t'X_tA_t'(\alpha_t)$ and it is evaluated at $f_{t|t-1}$.

3. Given $(\theta^T, f^T)$,

$$\widehat{A}(\alpha_t)\widehat{y}_t = \Sigma_t\varepsilon_t + \widehat{A}(\alpha_t)X_t'\upsilon_t.$$

Since $\widehat{A}(\alpha_t)X_t'$ is known, let the lower triangular $P_t$ satisfy $P_t(\widehat{A}(\alpha_t)X_t'X_t\widehat{A}(\alpha_t)')P_t' = I$. Then

$$P_t\widehat{A}(\alpha_t)\widehat{y}_t = y_t^{**} = P_t\Sigma_t\varepsilon_t + P_t\widehat{A}(\alpha_t)X_t'\upsilon_t$$

with $\mathrm{var}(P_t\widehat{A}(\alpha_t)X_t'\upsilon_t) = I$ and where $P_t\Sigma_t\Sigma_t'P_t' + P_t(\widehat{A}(\alpha_t)X_t'X_t\widehat{A}(\alpha_t)')P_t'$ is a diagonal matrix. This transformation is similar to Cogley and Sargent (2005); however, since $\widehat{A}(\alpha_t)X_t'$ is known, we only need to sample the variances of $\epsilon_{m,t}$. We do this using the $\log(\chi^2)$ approximation of a mixture of $J$ normals.

4. Given $(\theta^T, f^T, \Sigma^T)$, sample $Q$, $V$, and $W$ from independent inverted Wishart distributions.

5. Given new values of $\sigma_{m,t}$, we construct $A(\alpha_t)^{-1}\Sigma_t\Sigma_t'(A(\alpha_t)^{-1})' + X_t'X_t$ and go back to step 1.

## Appendix D: A shrinkage approach to draw $B^T$ when $\Xi$ is unknown

When the $\Xi$'s are known, the algorithm needs to be modified as follows.

The TVC-SVAR model is

$$y_t = X_t'B_t + A(\alpha_t)^{-1}\Sigma_t\varepsilon_t,$$

where $X_t' = I_M \otimes [D_t', y_{t-1}', \dots, y_{t-k}']$, with

$$
\begin{aligned}
&B_t = \Xi \theta_t + \omega_t, \\
&\theta_t = \theta_{t-1} + \upsilon_t, \\
&f_t = f_{t-1} + \zeta_t, \\
&\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t, \\
&\mathrm{Var}\left(\begin{bmatrix} \varepsilon_t \\ \omega_t \\ \upsilon_t \\ \zeta_t \\ \eta_t \end{bmatrix}\right) = \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & Q & 0 & 0 & 0 \\ 0 & 0 & R & 0 & 0 \\ 0 & 0 & 0 & V & 0 \\ 0 & 0 & 0 & 0 & W \end{bmatrix},
\end{aligned}
$$

where $Q$ and $R$ are diagonal matrices. We exploit the hierarchical structure of the model to simulate the posterior distribution as in Chib and Greenberg (1995):

1. Given $(A(\alpha_t), \sigma_t Q)$, sample $B_t$ using

$$
y_t = X_t' B_t + A(\alpha_t)^{-1} \Sigma_t \varepsilon_t
$$

with $A(\alpha_t)^{-1} \Sigma_t \varepsilon_t = u_t \sim N(0, H_t)$. That is, for each $t = 1, \dots, T$, draw

$$
B_t \sim N(\overline{B}_t, \overline{VB}_t),
$$

where

$$
\begin{aligned}
&\overline{VB}_t = \left(\underline{VB}^{-1} + X_t H_t^{-1} X_t'\right)^{-1}, \\
&\overline{B}_t = \overline{VB}_t \left(\underline{VB}^{-1} \underline{B}_t + X_t H_t^{-1} y_t\right)
\end{aligned}
$$

and priors

$$
\underline{VB} = Q, \qquad \underline{B}_t = \Xi \theta_t.
$$

2. Given $(B_t, \theta_t)$, compute the residuals $(B_t - \Xi \theta_t)$ and sample $Q$ using an inverse Wishart distribution.
3. Given $B_t$, sample $\theta_t$ using the state space form

$$
\begin{aligned}
&B_t = \Xi \theta_t + \omega_t, \\
&\theta_t = \theta_{t-1} + \upsilon_t.
\end{aligned}
$$

4. Given $\theta_t$, sample $R$ using an inverse Wishart distribution.
5. Given $(B_t, \theta_t, Q)$, draw $\Xi$ using

$$
B_t = \Xi \theta_t + \omega_t, \quad t = 1, \dots, T,
$$

where, to achieve identification, we normalize the first upper block of $\Xi$ to be an identity matrix, as in Koop and Korobilis (2010). That is, denote $\mathcal{F} = \dim(\theta_t)$ and $K = \dim(B_t)$. Then $\Xi$ is a $K \times \mathcal{F}$ matrix. The first $\mathcal{F}$ rows of $\Xi$ are

$$\Xi_{(1:\mathcal{F})\times(1:\mathcal{F})} = I_{\mathcal{F}}.$$

Moreover, since $\omega_t \sim N(0, Q)$ and we have assumed that $Q$ is diagonal, we draw the loadings row by row for each element of $B_t$. That is, for each $f = \mathcal{F}+1, \dots, K$, draw

$$\Xi_{f\times(1:\mathcal{F})} \sim N(\overline{\Xi}_f, \overline{V\Xi}_f)$$

with

$$\overline{V\Xi}_f = \left(\underline{V\Xi}^{-1} + Q^{-1}_{(f,f)}\left(\theta^T\right)\left(\theta^T\right)'\right)^{-1},$$

$$\overline{\Xi}_f = \overline{V\Xi}_f\left(\underline{V\Xi}^{-1}\underline{\Xi}_f + Q^{-1}_{(f,f)}\theta^T B_f^T\right),$$

where $\theta^T$ is an $\mathcal{F} \times T$ matrix of explanatory variables, $B_f^T$ is a $T \times 1$ vector that contains the dependent variable, and $Q_{(f,f)}$ is the corresponding element of matrix $Q$ drawn previously. The priors are $\underline{\Xi}_f = \mathbf{0}_{\mathcal{F}\times 1}$ and $\underline{V\Xi} = k^2_{\Xi} I_{\mathcal{F}}$ with the hyperparameter $k^2_{\Xi} = 0.01$.

6. Given $(B_t, Q)$, sample $(A(\alpha_t), V, \sigma_t, W)$ as before. Then go back to step 1.

## Appendix E: Nonlinear models

### E.1 *The setup*

Consider the general nonlinear state space model

$$y_t = z_t(\beta_t, \alpha_t) + u_t(\sigma_t, \xi_{1t}), \tag{E.1}$$

$$\beta_t = w_t(\beta_{t-1}) + s_t(\beta_{t-1}, \xi_{2t}), \tag{E.2}$$

$$\alpha_t = t_t(\alpha_{t-1}) + r_t(\alpha_{t-1}, \xi_{3t}), \tag{E.3}$$

$$f_t(\sigma_t) = h_t(\sigma_{t-1}) + k_t\big(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})\big), \tag{E.4}$$

where $y_t$, and $\xi_{1t}$ are $M \times 1$ vectors, $\beta_t$ and $\xi_{2t}$ are $K_\beta \times 1$ vectors, $\alpha_t$ and $\xi_{3t}$ are $K_\alpha \times 1$ vectors, $\xi_{1t} \sim N(0, Q_{1t})$, $\xi_{2t} \sim N(0, Q_{2t})$, and $\xi_{3t} \sim N(0, Q_{3t})$. Assume that $z_t(\cdot)$, $u_t(\cdot)$, $w_t(\cdot)$, $s_t(\cdot)$, $t_t(\cdot)$, $r_t(\cdot)$, $f_t(\cdot)$, $h_t(\cdot)$, and $k_t(\cdot)$ are continuous and differentiable vector-valued functions. To estimate this system, we can linearize it around the previous forecast of the state vector, so that

$$z_t(\beta_t, \alpha_t) \simeq z_t(\widehat{b}_{t|t-1}, \widehat{a}_{t|t-1}) + \widehat{Z}_{1t}(\beta_t - \widehat{b}_{t|t-1}) + \widehat{Z}_{2t}(\alpha_t - \widehat{a}_{t|t-1}),$$

$$u_t(\sigma_t, \xi_{1t}) \simeq u_t(\widehat{\sigma}_{t|t-1}, 0) + \widehat{u}_{\sigma,t}(\sigma_t - \widehat{\sigma}_{t|t-1}) + \widehat{u}_{\xi_1,t}\xi_{1,t},$$

$$w_t(\beta_{t-1}) \simeq w_t(\widehat{b}_{t-1|t-1}) + \widehat{w}_t(\beta_{t-1} - \widehat{b}_{t-1|t-1}),$$

$$s_t(\beta_{t-1}, \xi_{2t}) \simeq s_t(\widehat{\beta}_{t-1|t-1}, 0) + \widehat{s}_{\beta,t}(\beta_{t-1} - \widehat{b}_{t-1|t-1}) + \widehat{s}_{\xi_2,t}\xi_{2,t},$$

$$t_t(\alpha_{t-1}) \simeq t_t(\widehat{a}_{t-1|t-1}) + \widehat{T}_t(\alpha_{t-1} - \widehat{a}_{t-1|t-1}),$$

$$r_t(\alpha_{t-1}, \xi_{3t}) \simeq r_t(\widehat{\alpha}_{t-1|t-1}, 0) + \widehat{r}_{\alpha,t}(\alpha_{t-1} - \widehat{a}_{t-1|t-1}) + \widehat{r}_{\xi_3,t}\xi_{3,t},$$

$$f_t(\sigma_t) \simeq f_t(\widehat{\sigma}_{t|t-1}) + \widehat{f}_t(\sigma_t - \widehat{\sigma}_{t|t-1}),$$

$$h_t(\sigma_{t-1}) \simeq h_t(\widehat{\sigma}_{t-1|t-1}) + \widehat{h}_t(\sigma_{t-1} - \widehat{\sigma}_{t-1|t-1}),$$

$$k_t\big(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})\big) \simeq k_t(\widehat{u}_{\xi_1,t-1}\xi_{1,t-1}),$$

where $\widehat{Z}_{i,t}$, $i = 1, 2$, and $\widehat{u}_{\sigma,t}$, $\widehat{u}_{\xi_1,t}$, $\widehat{w}_t$, $\widehat{T}_t$, $\widehat{s}_{\beta,t}$, $\widehat{s}_{\xi_2,t}$, $\widehat{r}_{\alpha,t}$, and $\widehat{r}_{\xi_3,t}$ are matrices corresponding to the Jacobian of $z_t(\cdot)$, $u_t(\cdot)$, $w_t(\cdot)$, $t_t(\cdot)$, $s_t(\cdot)$, and $r_t(\cdot)$, evaluated at $\beta_t = \widehat{b}_{t|t-1}$, $\alpha_t = \widehat{a}_{t|t-1}$, $\sigma_t = \widehat{\sigma}_{t|t-1}$, and $\xi_{1,t} = \xi_{2,t} = \xi_{3,t} = 0$. Thus, the approximated model is

$$\widehat{y}_t \simeq \widehat{Z}_{1t}\beta_t + \widehat{Z}_{2t}\alpha_t + \widehat{d}_t + \widehat{u}_{\xi_1,t}\xi_{1,t}, \tag{E.5}$$

$$\beta_t \simeq \widehat{w}_t\beta_{t-1} + \widehat{g}_t + \widehat{s}_{\xi_2,t}\xi_{2,t}, \tag{E.6}$$

$$\alpha_t \simeq \widehat{T}_t\alpha_{t-1} + \widehat{c}_t + \widehat{r}_{\xi_3,t}\xi_{3,t}, \tag{E.7}$$

$$\widehat{f}_t\sigma_t = \widehat{h}_t\sigma_{t-1} + k_t(\widehat{u}_{\xi_1,t-1}\xi_{1,t-1}), \tag{E.8}$$

where

$$\begin{aligned}\widehat{d}_t &= z_{1t}(\widehat{b}_{t|t-1}) - \widehat{Z}_{1t}\widehat{b}_{t|t-1} + z_{2t}(\widehat{a}_{t|t-1}) - \widehat{Z}_{2t}\widehat{a}_{t|t-1} \\ &\quad + u(\widehat{\sigma}_{t|t-1}, 0) - \widehat{u}_{\sigma,t}(\widehat{\sigma}_{t|t-1} - \sigma_t),\end{aligned} \tag{E.9}$$

$$\widehat{c}_t = t_t(\widehat{a}_{t-1|t-1}) - \widehat{T}_t\widehat{a}_{t-1|t-1} + r_t(\widehat{\alpha}_{t|t-1}, 0) - \widehat{r}_{\alpha,t}(\widehat{\alpha}_{t|t-1} - a_{t-1}), \tag{E.10}$$

$$\widehat{g}_t = w_t(\widehat{b}_{t-1|t-1}) - \widehat{W}_t\widehat{b}_{t-1|t-1} + s_t(\widehat{\beta}_{t|t-1}, 0) - \widehat{s}_{\beta,t}(\widehat{\beta}_{t|t-1} - b_{t-1}). \tag{E.11}$$

When (i) $z_t(\cdot)$, $w_t(\cdot)$, $t_t(\cdot)$, and $u_t(\cdot)$ are linear, (ii) $s_t(\cdot)$ is independent of $\beta_t$, (iii) $r_t(\cdot)$ is independent of $\alpha_t$, and (iv) $u_t(\cdot)$ is independent of $\sigma_t$, $\widehat{d}_t = \mathbf{0}$, $\widehat{c}_t = \mathbf{0}$, $\widehat{g}_t = \mathbf{0}$. In one of the cases considered by Rubio Ramírez et al. (2010), $\widehat{d}_t \neq \mathbf{0}$, while if the law of motion of the structural coefficient is nonlinear or there are nonlinear identification restrictions, $\widehat{c}_t \neq \mathbf{0}$ or $\widehat{g}_t \neq \mathbf{0}$.

### E.2 *Estimation*

Since (E.5)–(E.8) are linear, the algorithm described in Section 4 can now be applied. The only difference is that we now draw from distributions or proposals that are centered at the extended Kalman smoother estimates. For example, given $(f_0, y^T, \Sigma^T)$, we construct updated estimates according to

$$f_{t|t} = f_{t|t-1} + K_t[y_t - z_t f_{t|t-1}], \tag{E.12}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}\widehat{Z}_t'\Gamma_t^{-1}\widehat{Z}_t P_{t|t-1}', \tag{E.13}$$

where $f_{t|t-1} = t_t f_{t-1|t-1}$, $P_{t|t-1} = \widehat{T}_t P_{t-1|t-1}\widehat{T}_t' + \widehat{r}_{\xi_2,t}Q_{2t}\widehat{r}_{\xi_2,t}' K_t = P_{t|t-1}\widehat{Z}_t'\Gamma_t^{-1}$, and $\Gamma_t = \widehat{Z}_t' P_{t|t-1}\widehat{Z}_t + \widehat{u}_{\xi_1,t}Q_{2t}\widehat{u}_{\xi_1,t}'$.

Smoothed estimates are $f^*_{T|T} = f_{T|T}$, $P^*_{T|T} = P_{T|T}$, and

$$f^*_{t|t+1} = f_{t|t} + P_{t|t}\widehat{Z}'_t P^{-1}_{t+1|t}(f^*_{t+1|t+2} - t_t f(a_{t|t})), \tag{E.14}$$

$$P^*_{t|t+1} = P_{t|t} - P_{t|t}\widehat{Z}'_t\big[P_{t+1|t} + \widehat{r}_{\xi_2,t}Q_{2t}\widehat{r}'_{\xi_2,t}\big]^{-1}\widehat{Z}_t P'_{t|t-1} \tag{E.15}$$

for $t = T-1, \dots, 1$. Hence, when $f(\alpha_t)$ is nonlinear, we draw $f^T$ from a proposal centered at (E.14)–(E.15). Notice that the approximate model is used only in predicting and updating the mean squared error of $f(\alpha_t)$.

Depending on the exact specification of the nonlinear model, one or more steps in the algorithm may require some adjustments.

## E.3 *Sampling the GARCH model*

To sample volatilities when their law of motion is assumed to be a GARCH$(1, 1)$, we need to modify the transition and the measurement equations used in step 3 of the algorithm of Section 4. The $m$th equation of the model is

$$y^{**}_{m,t} = \sigma_{m,t}\varepsilon_{m,t}, \tag{E.16}$$

where $\sigma_{m,t}$ is the $m$th diagonal element of $\Sigma_t$. Assume

$$\sigma^2_{m,t} = \big(1 - \delta + \delta\sigma^2_{m,t-1} + \delta\big(y^{**}_{m,t-1}\big)^2\big) + \eta_{m,t} \tag{E.17}$$

with $\eta_t \sim N(0, W)$, where $\delta$ and $W$ are known parameters.

The system (E.16)–(E.17) is now nonlinear. Equation (E.16) can be written as

$$y^{**}_{m,t} = z(\sigma_{m,t}) + u_t(\sigma_{m,t}, \varepsilon_{m,t}).$$

Since $z(\sigma_{m,t}) = 0$, the linear approximation is

$$\sigma_{m,t}\varepsilon_{m,t} \simeq u_t(\widehat{\sigma}_{m,t|t-1}, 0) + \widehat{u}_{\sigma,t}(\sigma_{m,t} - \widehat{\sigma}_{m,t|t-1}) + \widehat{u}_{\varepsilon_{m,t}}\varepsilon_{m,t} = \widehat{\sigma}_{m,t|t-1}\varepsilon_{m,t}$$

because

- $u_t(\widehat{\sigma}_{m,t|t-1}, 0) = \widehat{\sigma}_{m,t|t-1} \times 0 = 0$,
- $\widehat{u}_{\sigma,t} = \frac{\partial u_t(\sigma_{m,t},\varepsilon_{m,t})}{\partial \sigma_{m,t}}|_{(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0)} = \varepsilon_{m,t}|_{(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0)} = 0$,
- $\widehat{u}_{\varepsilon_{m,t}} = \frac{\partial u_t(\sigma_{m,t},\varepsilon_{m,t})}{\partial \varepsilon_{m,t}}|_{(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0)} = \sigma_{m,t}|_{(\sigma_{m,t}=\widehat{\sigma}_{m,t|t-1},\varepsilon_{m,t}=0)} = \widehat{\sigma}_{m,t|t-1}$.

The transition equation (E.17) can be written as

$$\begin{aligned}\sigma^2_{m,t} &\equiv f_t(\sigma_{m,t}) = h_t(\sigma_{m,t-1}) + k_t(\sigma_{m,t-1}, \eta_{m,t}) \\ &\equiv \big(1 - \delta + \delta\sigma^2_{m,t-1} + \delta\big(y^{**}_{m,t-1}\big)^2\big) + \eta_{m,t}.\end{aligned}$$

Linearizing the two sides of the equation, we have

$$\begin{aligned}f_t(\sigma_{m,t}) &\simeq f_t(\widehat{\sigma}_{m,t|t-1}) + \widehat{f_t}(\widehat{\sigma}_{m,t-1|t-1})(\sigma_{m,t-1} - \widehat{\sigma}_{m,t-1|t-1}), \\ h_t(\sigma_{m,t-1}) &\simeq h_t(\widehat{\sigma}_{m,t-1|t-1}) + \widehat{h}_t(\widehat{\sigma}_{m,t-1|t-1})(\sigma_{m,t-1} - \widehat{\sigma}_{m,t-1|t-1}),\end{aligned}$$

where $\widehat{f_t}(\widehat{\sigma}_{m,t|t-1}, 0) = 2\sigma_{m,t}|_{(\widehat{\sigma}_{m,t|t-1},0)}$ and $\widehat{h}_t(\widehat{\sigma}_{m,t-1|t-1}, 0) = 2\delta\sigma_{m,t-1}|_{(\widehat{\sigma}_{m,t|t-1},0)}$.

E.4 *Long-run restrictions*

Long-run restrictions are nonlinear in the SVAR coefficients, but linear in the impulse responses. For the sake of presentation, we omit the intercept $B_{0,t}$. Let

$$y_t = B_{1,t} y_{t-1} + \cdots + B_{p,t} y_{t-p} + \left[A(\alpha_t)\right]^{-1} \Sigma_t \varepsilon_t.$$

Then we only need to modify how draws for the $B_t$ block are made, in particular, as follows:

1. At iteration $i$, given $A(\alpha_t)^{i-1}$ and $\Sigma_t^{i-1}$, sample $\{B_t^i\}_{t=1}^T$ using Carter and Kohn's routine or one of the other routines described in Section 5. With the sampled vector, compute the companion matrix

$$\mathbf{B}_t^i = \begin{bmatrix} B_{1,t}^i & \cdots & B_{p-1,t}^i & B_{p,t}^i \\ I_M & \cdots & \mathbf{0}_{M\times M} & \mathbf{0}_{M\times M} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{M\times M} & \cdots & I_M & \mathbf{0}_{M\times M} \end{bmatrix},$$

where $B_t^i = [\text{vec}(B_{1,t}^i)', \dots, \text{vec}(B_{p,t}^i)']'$.

2. Given $B_t^i$, $A(\alpha_t)^{i-1}$, and $\Sigma_t^{i-1}$, compute the long-run matrix for each $t$,

$$\begin{aligned} \mathbf{D}_t^i &= \mathbf{J}\big(I_{Mp} - \mathbf{B}_t^i\big)^{-1} \mathbf{J}' \big[A(\alpha_t)^{i-1}\big]^{-1} \Sigma_t^{i-1} \\ &= \big(I_M - B_{1t}^i - \cdots - B_{p,t}^i\big)^{-1} \big[A(\alpha_t)^{i-1}\big]^{-1} \Sigma_t^{i-1}, \end{aligned} \tag{E.18}$$

where $\mathbf{J} = [\, I_M \quad \mathbf{0}_{M\times M} \quad \cdots \quad \mathbf{0}_{M\times M} \,]$ is a selection matrix.

3. Impose long-run restrictions, i.e., construct $\tilde{\mathbf{D}}_t^i = R\mathbf{D}_t^i$, where $R$ is matrix restricting the entries of $\mathbf{D}_t^i$.

4. Given $\tilde{\mathbf{D}}_t^i$, then $A(\alpha_t)^{i-1}$, $\Sigma_t^{i-1}$, and $B_{j,t}^i$, $j = 1, \dots, p-1$, solve for $\tilde{B}_{p,t}^i$ using (E.18), so that

$$\tilde{B}_{p,t}^i = I_M - B_{1,t}^i - \cdots - B_{p-1,t}^i - \big[A(\alpha_t)^{i-1}\big]^{-1} \Sigma_t^{i-1} \big[\tilde{\mathbf{D}}_t^i\big]^{-1},$$

and with this construct the restricted draw $\tilde{B}_t^i = [\text{vec}(B_{1,t}^i)', \dots, \text{vec}(\tilde{B}_{p,t}^i)']'$.

5. Evaluate whether

$$\tilde{\mathbf{B}}_t^i = \begin{bmatrix} B_{1,t}^i & \cdots & B_{p-1,t}^i & \tilde{B}_{p,t}^i \\ I_M & \cdots & \mathbf{0}_{M\times M} & \mathbf{0}_{M\times M} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{M\times M} & \cdots & I_M & \mathbf{0}_{M\times M} \end{bmatrix}$$

has all its eigenvalues inside the unit circle. If so, we accept $\tilde{B}_t^i$; otherwise discard it.

Given a draw for $\tilde{B}_t$, the sampling of the remaining blocks $(A(\alpha_t), \Sigma_t, s, \mathcal{V})$ is unchanged.

## Additional reference

Chib, S. and E. Greenberg (1995), "Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models." *Journal of Econometrics*, 68, 339–360. [7]

---

Submitted September, 2012. Final version accepted July, 2014.