A Global Identification of the constant coefficient SVAR

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

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\alpha_1 & 1 & 0 & 0 & 0 & 0 \\
\alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \alpha_{11} & 0 \\
\alpha_3 & \alpha_6 & 0 & \alpha_9 & 1 & 0 \\
\alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1
\end{bmatrix}
A(\alpha)
\begin{bmatrix}
GDP_t \\
P_t \\
U_t \\
R_t \\
M_t \\
Pcom_t
\end{bmatrix}
= A^+(L)
\begin{bmatrix}
GDP_{t-1} \\
P_{t-1} \\
U_{t-1} \\
R_{t-1} \\
M_{t-1} \\
P_{com_{t-1}}
\end{bmatrix}
+ \Sigma
\begin{bmatrix}
\varepsilon^y_t \\
\varepsilon^p_t \\
\varepsilon^u_t \\
\varepsilon^{mp}_t \\
\varepsilon^{md}_t \\
\varepsilon^i_t
\end{bmatrix}
\]

with

\[
\Sigma =
\begin{bmatrix}
\sigma^i & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^{md} & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^{mp} & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^y & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^p & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^u
\end{bmatrix}
\]

To verify that the system is globally identified, we rewrite the model using the notation of Rubio Ramirez et. al. (2010). Let \( y_t \equiv (GDP_t, P_t, U_t, R_t, M_t, Pcom_t)' \)

*EUI and CEPR; and BCRP.
and \( \varepsilon_t \equiv (\varepsilon_{t}^y, \varepsilon_{t}^p, \varepsilon_{t}^{mp}, \varepsilon_{t}^{md}, \varepsilon_{t}^i)' \). Pre-multiplying by \( \Sigma^{-1} \), we obtain

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

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

\[
y_t'A_0 = \sum_{L=1}^{p} y'_{t-L}A_L + \varepsilon'_t
\]

where

\[
A'_0 = \begin{bmatrix}
\sigma^y & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^p & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^n & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^{mp} & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^{md} & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^i
\end{bmatrix}^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\alpha_1 & 1 & 0 & 0 & 0 & 0 \\
\alpha_2 & \alpha_5 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \alpha_{11} & 0 \\
\alpha_3 & \alpha_6 & \alpha_9 & 1 & 0 & 0 \\
\alpha_4 & \alpha_7 & \alpha_8 & \alpha_{10} & \alpha_{12} & 1
\end{bmatrix}
\]

Denoting \( A_0 = [a_{ij}] \) we have

\[
A_0 = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\
a_{22} & a_{23} & 0 & a_{25} & a_{26} \\
0 & a_{33} & 0 & 0 & a_{36} \\
0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\
0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\
0 & 0 & 0 & 0 & 0 & a_{66}
\end{bmatrix}
\]

The matrices \( Q_j \), \( j = 1, \ldots, 6 \), present in Theorem 1 of Rubio Ramirez et al. (2010) are:

\[
Q_1 = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}; \quad Q_2 = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
Define the matrices

\[ M_j(A_0) = \begin{bmatrix} Q_j A_0 \\ I_j \end{bmatrix} ; \quad j = 1, \ldots, M \]  

so that

\[ M_1 = \begin{bmatrix} 0 & a_{22} & a_{23} & 0 & a_{25} & a_{26} \\ 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad ; \quad M_2 = \begin{bmatrix} 0 & 0 & a_{33} & 0 & 0 & a_{36} \\ 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ M_3 = \begin{bmatrix} 0 & 0 & 0 & a_{44} & a_{45} & a_{46} \\ 0 & 0 & 0 & a_{54} & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad ; \quad M_4 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 & a_{15} & a_{16} \\ a_{22} & a_{23} & 0 & a_{25} & a_{26} & 0 \\ a_{33} & 0 & 0 & a_{36} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{66} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \]

4
Since all $M_j$ have full column rank, the model is globally identified.

B Single-move Metropolis for drawing $B_t$

The 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

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

and the transition equation is

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

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

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

$$
\mu_t = \begin{cases} 
\frac{B^i_{t-1} + B^i_{t+1}}{2} + G_t \left[ y_t - X_t^i \left( \frac{B^i_{t-1} + B^i_{t+1}}{2} \right) \right], & t < T \\
B^i_t + G_t \left[ y_t - X_t^i \left( B^i_{t-1} \right) \right], & t = T 
\end{cases}
$$

$$G_t = \begin{cases} 
\frac{1}{2} Q^{i-1} X_t (X_t Q^{i-1} X_t + \Omega_t)^{-1}, & t < T \\
Q^{i-1} X_t (X_t Q^{i-1} X_t + \Omega_t)^{-1}, & t = T 
\end{cases}
$$

$$\Psi_t = \begin{cases} 
\frac{1}{2} (I_K - G_t X_t^i) Q^{i-1}, & t < T \\
(I_K - G_t X_t^i) Q^{i-1}, & t = T 
\end{cases}
$$
2. Construct the companion form matrix $\mathbf{B}_t^i$ and evaluate $I \left( \max \left| \text{eig} \left( \mathbf{B}_t^i \right) \right| < 1 \right)$, where $I \left( . \right)$ is an indicator function taking the value of 1 if the condition within the parenthesis is satisfied.

3. The acceptance rate of $B_t^i$ is

$$
\omega_{B,T} = \min \left\{ \frac{I \left( \max \left| \text{eig} \left( \mathbf{B}_t^i \right) \right| < 1 \right)}{\lambda \left( B_t^i, Q_t^{i-1} \right)}, 1 \right\} = \min \left\{ \frac{I \left( \max \left| \text{eig} \left( \mathbf{B}_t^i \right) \right| < 1 \right) \lambda \left( B_t^{i-1}, Q_t^{i-1} \right)}{\lambda \left( B_t^i, Q_t^{i-1} \right)}, 1 \right\}
$$

where $\lambda \left( . \right)$ is an integrating constant, measuring the proportion of draws that satisfy the inequality constraint. To compute $\lambda$, one first draws $\mathbf{B}_t^{i,l} \sim N \left( \mathbf{B}_t^i, Q_t^{i-1} \right)$, for $l = 1, \ldots, L$, constructs the companion form matrix $\mathbf{B}_t^{i,l}$ and evaluates $\lambda_l = I \left( \max \left| \text{eig} \left( \mathbf{B}_t^{i,l} \right) \right| < 1 \right)$. Second, one evaluates $\lambda \left( B_t^i, Q_t^{i-1} \right) = \sum_{l=1}^L \lambda_l$ and $\lambda \left( B_t^{i-1}, Q_t^{i-1} \right)$ and compute the acceptance probability. When $t = T$, this probability is

$$
\omega_{B,T} = I \left( \max \left| \text{eig} \left( \mathbf{B}_T^i \right) \right| < 1 \right)
$$

4. Draw a $v \sim U \left( 0, 1 \right)$. 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 $Q^{-1} \sim W \left( \mathbf{v}, \mathbf{Q}^{-1} \right)$ so that the unrestricted posterior is $Q^{-1} \sim W \left( \mathbf{v}, \mathbf{Q}^{-1} \right)$ with $\mathbf{v} = \mathbf{v} + \mathbf{T}$ and $\mathbf{Q}^{-1} = \left[ Q + \sum_{t=1}^T (B_{t,i} - B_{t-1,i}) (B_{t,i} - B_{t-1,i})' \right]^{-1}$.

Then draw a candidate $\left( Q_t^i \right)^{-1} \sim W \left( \mathbf{v}, \mathbf{Q}^{-1} \right)$ and for $t = 1, \ldots, T$, evaluate $\lambda \left( B_t^i, Q_t^i \right)$ and $\lambda \left( B_t^i, Q_t^{i-1} \right)$, for a fixed $L$, and calculate

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

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

Note that in a multi-move approach $\lambda \left( . \right) = 1$, 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 $\lambda \left( . \right)$ needs to be simulated at each $t$, of adding considerable computational time.
C  A shrinkage approach to draw $B^T$ when $\Xi$ is known

The model is still consists of

\[
y_t = X_t' B_t + A_t^{-1} \Sigma_t \xi_t \\
\alpha_t = \alpha_{t-1} + \zeta_t \\
\log(\sigma_{m,t}) = \log(\sigma_{m,t-1}) + \eta_{m,t}
\]

but now

\[
B_t = B_{t-1} + \nu_t
\]

is substituted by

\[
B_t = \Xi \theta_t + \nu_t \quad \nu_t \sim N(0, I) \quad (C.1) \\
\theta_t = \theta_{t-1} + \rho_t \quad \rho_t \sim N(0, Q) \quad (C.2)
\]

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

\[
y_t = X_t' \Xi \theta_t + A(\alpha_t)^{-1} \Sigma_t \xi_t + X_t' \nu_t \equiv X_t' \Xi \theta_t + \psi_t \quad (C.3)
\]

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

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

\[
A(\alpha_t) \hat{y}_t = A(\alpha_t) \xi_t = \Sigma_t \xi_t + A(\alpha_t) X_t' \nu_t
\]

As before

\[
(\hat{y}_t' \otimes I_M)(S_A f_t + s_A) = \Sigma_t \xi_t + A(\alpha_t) X_t' \nu_t
\]

so that the second state-space system is

\[
\hat{y}_t = Z_t f_t + \Sigma_t \xi_t + A(\alpha_t) X_t' \nu_t \quad (C.4) \\
f_t = f_{t-1} + \zeta_t \quad (C.5)
\]

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

\[
\hat{A}(\alpha_t)\hat{y}_t = \Sigma_t \varepsilon_t + \hat{A}(\alpha_t)X'_t v_t
\]

Since \(\hat{A}(\alpha_t)X'_t\) is known, let the lower-triangular \(P_t\) satisfy \(P_t \left( \hat{A}(\alpha_t)X'_tX_t\hat{A}(\alpha_t)' \right) P_t' = I\). Then

\[
P_t\hat{A}(\alpha_t)\hat{y}_t = y_t^* = P_t\Sigma_t \varepsilon_t + P_t\hat{A}(\alpha_t)X'_t v_t
\]

with \(\text{var} \left( P_t\hat{A}(\alpha_t)X'_t v_t \right) = I\) and where \(P_t\Sigma_t \Sigma_t' P_t' + P_t \left( \hat{A}(\alpha_t)X'_tX_t\hat{A}(\alpha_t)' \right) P_t'\) is a diagonal matrix. This transformation is similar to Cogley and Sargent (2005); however, since \(\hat{A}(\alpha_t)X'_t\) is known, we only need to sample the variances of \(\varepsilon_{m,t}\). We do this using the \(\log(\chi^2)\) approximation of a mixture of \(J\) normals.

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

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

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 \(X'_t = I_M \otimes [D'_t, y'_{t-1}, \ldots, y'_{t-k}]\), with

\[
B_t = \Xi \theta_t + \omega_t
\]

\[
\theta_t = \theta_{t-1} + \nu_t
\]

\[
f_t = f_{t-1} + \zeta_t
\]

\[
\log (\sigma_t) = \log (\sigma_{t-1}) + \eta_t
\]

\[
\text{Var} \left( \begin{bmatrix} \varepsilon_t \\ \omega_t \\ \nu_t \\ \zeta_t \end{bmatrix} \right) = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & R & 0 \\ 0 & 0 & 0 & V \end{bmatrix}
\]

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 \(t = 1, \ldots, T\) draw:

\[
B_t \sim N(\overline{B}_t, \overline{V}B_t)
\]

where

\[
\overline{V}B_t = (VB^{-1} + X_t H_t^{-1} X_t')^{-1} \\
\overline{B}_t = \overline{V}B_t (VB^{-1}B_t + X_t H_t^{-1} y_t)
\]

and priors

\[
VB = Q; \quad B_t = \Xi \theta_t
\]

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:

\[
B_t = \Xi \theta_t + \omega_t \\
\theta_t = \theta_{t-1} + v_t
\]

4. Given \(\theta_t\), sample \(R\) using an inverse Wishart distribution.

5. Given \((B_t, \theta_t, Q)\) draw \(\Xi\) using:

\[
B_t = \Xi \theta_t + \omega_t; \quad t = 1, \ldots, T
\]

where, in order 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 \(\mathcal{F} = \text{dim}(\theta_t)\) and \(K = \text{dim}(B_t)\), then \(\Xi\) is a \(K \times \mathcal{F}\) matrix. The first \(\mathcal{F}\) rows of \(\Xi\) are:

\[
\Xi_{(1: \mathcal{F})x(1: \mathcal{F})} = I_{\mathcal{F}}
\]

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 \(f = \mathcal{F} + 1, \ldots, K\) draw:

\[
\Xi_{f \times (1: \mathcal{F})} \sim N(\Xi_f, \overline{V} \Xi_f)
\]
Consider the general non-linear state space model:

\[ y_t = z_t(\beta_t, \alpha_t) + u_t(\sigma_t, \xi_{1t}) \]  
(E.1) 
\[ \beta_t = w_t(\beta_{t-1}) + s_t(\beta_{t-1}, \xi_{2t}) \]  
(E.2) 
\[ \alpha_t = t_t(\alpha_{t-1}) + r_t(\alpha_{t-1}, \xi_{3t}) \]  
(E.3) 
\[ f_t(\sigma_t) = h_t(\sigma_{t-1}) + k_t(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})) \]  
(E.4)

where \( \theta^T \) is a \( F \times T \) 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 \( \Xi_f = 0_{F \times 1} \); \( V \Xi = k^2_{\Xi} \mathbb{I}_F \) 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 1.

## E Non-linear models

### E.1 The setup

Consider the general non-linear state space model:

\[ y_t = z_t(\beta_t, \alpha_t) + u_t(\sigma_t, \xi_{1t}) \]  
(E.1) 
\[ \beta_t = w_t(\beta_{t-1}) + s_t(\beta_{t-1}, \xi_{2t}) \]  
(E.2) 
\[ \alpha_t = t_t(\alpha_{t-1}) + r_t(\alpha_{t-1}, \xi_{3t}) \]  
(E.3) 
\[ f_t(\sigma_t) = h_t(\sigma_{t-1}) + k_t(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})) \]  
(E.4)

where \( y_t, \xi_{1t} \) 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; \( \xi_{1t} \sim N(0, Q_{1t}), \xi_{2t} \sim N(0, Q_{2t}), \xi_{3t} \sim N(0, Q_{3t}) \). 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

\[ z_t(\beta_t, \alpha_t) \approx z_t(\hat{b}_{t|t-1}, \hat{a}_{t|t-1}) + \hat{Z}_t(\beta_t - \hat{b}_{t|t-1}) + \hat{Z}_t(\alpha_t - \hat{a}_{t|t-1}) \]
\[ u_t(\sigma_t, \xi_{1t}) \approx u_t(\hat{\sigma}_{t|t-1}, 0) + \hat{u}_{\sigma,t}(\sigma_t - \hat{\sigma}_{t|t-1}) + \hat{u}_{\xi_{1,t}} \xi_{1,t} \]
\[ w_t(\beta_{t-1}) \approx w_t(\hat{b}_{t-1|t-1}) + \hat{w}_t(\beta_{t-1} - \hat{b}_{t-1|t-1}) \]
\[ s_t(\beta_{t-1}, \xi_{2t}) \approx s_t(\hat{\beta}_{t-1|t-1}, 0) + \hat{s}_{\beta,t}(\beta_{t-1} - \hat{b}_{t-1|t-1}) + \hat{s}_{\xi_{2,t}} \xi_{2,t} \]
\[ t_t(\alpha_{t-1}) \approx t_t(\hat{a}_{t-1|t-1}) + \hat{T}_t(\alpha_{t-1} - \hat{a}_{t-1|t-1}) \]
\[ r_t(\alpha_{t-1}, \xi_{3t}) \approx r_t(\hat{\alpha}_{t-1|t-1}, 0) + \hat{r}_{\alpha,t}(\alpha_{t-1} - \hat{a}_{t-1|t-1}) + \hat{r}_{\xi_{3,t}} \xi_{3,t} \]
\[ f_t(\sigma_t) \approx f_t(\hat{\sigma}_{t|t-1}) + \hat{f}_t(\sigma_t - \hat{\sigma}_{t|t-1}) \]
\[ h_t(\sigma_{t-1}) \approx h_t(\hat{\sigma}_{t-1|t-1}) + \hat{h}_t(\sigma_{t-1} - \hat{\sigma}_{t-1|t-1}) \]
\[ k_t(u_{t-1}(\sigma_{t-1}, \xi_{1t-1})) \approx k_t(\hat{u}_{\xi_{1,t-1}} \xi_{1,t-1}) \]
where $\hat{Z}_{i,t}, i = 1, 2; \hat{u}_{\sigma,t}, \hat{u}_{\xi_1,t}, \hat{w}_t, \hat{T}_t, \hat{s}_{\beta,t}, \hat{s}_{\xi_2,t} \hat{r}_{\alpha,t}, \hat{r}_{\xi_3,t}$ are matrices corresponding to the Jacobian of $z_t(\cdot), u_t(\cdot), w_t(\cdot), t_l(\cdot), s_l(\cdot), r_l(\cdot)$, evaluated at $\beta_t = \hat{b}_{l|t-1}, \alpha_t = \hat{a}_{l|t-1}, \sigma_t = \hat{\sigma}_{l|t-1}, \xi_{1,t} = \xi_{2,t} = \xi_{3,t} = 0$. Thus, the approximated model is

$$y_t \simeq \hat{Z}_{l|t} \beta_t + \hat{Z}_{2l} \alpha_t + \hat{d}_t + \hat{u}_{\xi_1,t} \xi_{1,t}$$  \hspace{1cm} (E.5)

$$\beta_t \simeq \hat{w}_t \beta_{t-1} + \hat{g}_t + \hat{s}_{\xi_2,t} \xi_{2,t}$$  \hspace{1cm} (E.6)

$$\alpha_t \simeq \hat{T}_t \alpha_{t-1} + \hat{c}_t + \hat{r}_{\xi_3,t} \xi_{3,t}$$  \hspace{1cm} (E.7)

$$\hat{f}_t \sigma_t \simeq \hat{h}_t \sigma_{t-1} + k_t (\hat{u}_{\xi_1,t-1} \xi_{1,t-1})$$  \hspace{1cm} (E.8)

where

$$\hat{d}_t = z_t (\hat{b}_{l|t-1}) - \hat{Z}_{l|t} \hat{b}_{l|t-1} + z_{2t} (\hat{a}_{l|t-1}) - \hat{Z}_{2l} \hat{a}_{l|t-1} + u(\hat{\sigma}_{l|t-1}, 0) - \hat{u}_{\sigma,t} \hat{\sigma}_{l|t-1} - \sigma_t$$  \hspace{1cm} (E.9)

$$\hat{c}_t = t_t (\hat{a}_{l|t-1}) - \hat{T}_t \hat{a}_{l-1|t-1} + r_t (\hat{\alpha}_{l|t-1}, 0) - \hat{r}_{t,t} \hat{\alpha}_{l|t-1} - \alpha_{t-1}$$  \hspace{1cm} (E.10)

$$\hat{g}_t = w_t (\hat{b}_{l|t-1}) - \hat{W}_t \hat{b}_{l-1|t-1} + s_t (\hat{\beta}_{l|t-1}, 0) - \hat{s}_{\beta,t} \hat{\beta}_{l|t-1} - \beta_{t-1}$$  \hspace{1cm} (E.11)

When i) $z_t(\cdot), w_t(\cdot), t_l(\cdot), u_t(\cdot)$ 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 $\sigma_t$, $\hat{d}_t = 0$, $\hat{c}_t = 0$, $\hat{g}_t = 0$. In one of the cases considered by Rubio Ramírez et al. (2010) $\hat{d}_t \neq 0$, while if the law of motion of the structural coefficient is non-linear or there are non-linear identification restrictions, $\hat{c}_t \neq 0$ or $\hat{g}_t \neq 0$.

**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 which are centered at the Extended Kalman Smoother estimates. For example, given $(f_0, y^T, \Sigma^T)$, we construct updated estimates according to

$$f_{l|t} = f_{l|t-1} + K_t \left[y_t - z_t f_{l|t-1}\right]$$  \hspace{1cm} (E.12)

$$P_{l|t} = P_{l|t-1} - P_{l|t-1} \hat{Z}_t\Gamma_t^{-1} \hat{Z}_t P_{l|t-1}'$$  \hspace{1cm} (E.13)

where $f_{l|t-1} = t_l f_{l-1|t-1}, P_{l|t-1} = \hat{T}_t P_{l-1|t-1} \hat{T}_t' + \hat{r}_{\xi_2,t} Q_2 \hat{r}_{\xi_2,t} K_t = P_{l|t-1} \hat{Z}_t\Gamma_t^{-1}$, and $\Gamma_t = \hat{Z}_t P_{l|t-1} \hat{Z}_t + \hat{u}_{\xi_1,t} Q_2 \hat{u}_{\xi_1,t}$.

Smoothed estimates are $f_{l|T} = f_{T|T}, P_{l|T} = P_{T|T}$ and

$$f_{l+1|T} = f_{l|T} + P_{l|T} \hat{Z}_t P_{l+1|T}^{-1} \left(f_{l+1|T} - t_l f(a_{l|T})\right)$$  \hspace{1cm} (E.14)
Linearizing the two sides of the equation we have:

\[ P'_{t+1} = P_{t} - P_{t} \tilde{Z}'_{t} \left[ P_{t+1|t} + \tilde{\hat{\xi}}_{t|t} Q 2n \tilde{\hat{\xi}}'_{t} \right]^{-1} \tilde{Z}_{t} P'_{t-1} \quad (E.15) \]

for \( t = T - 1, \ldots, 1 \). 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 non-linear 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:

\[ y_{m,t}^{**} = \sigma_{m,t} \varepsilon_{m,t} \quad (E.16) \]

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

\[ \sigma_{m,t}^2 = (1 - \delta + \delta \sigma_{m,t-1}^2 + \delta \left( y_{m,t-1}^{**} \right)^2) + \eta_{m,t} \quad (E.17) \]

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

The system (E.16) – (E.17) is now non-linear. Equation (E.16) can be written as:

\[ y_{m,t}^{**} = z(\sigma_{m,t}) + u(\sigma_{m,t}, \varepsilon_{m,t}) \]

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

\[ \sigma_{m,t} \varepsilon_{m,t} \simeq u_t(\tilde{\sigma}_{m,t}[t-1,0]) + \tilde{u}_u(\sigma_{m,t} - \tilde{\sigma}_{m,t}[t-1]) + \tilde{u}_\varepsilon(\varepsilon_{m,t} - \tilde{\varepsilon}_{m,t}) = \tilde{\sigma}_{m,t}[t-1] \varepsilon_{m,t} \]

because:

- \( u_t(\tilde{\sigma}_{m,t}[t-1,0]) = \tilde{\sigma}_{m,t}[t-1] \times 0 = 0 \)
- \( \tilde{u}_u(\sigma_{m,t} - \tilde{\sigma}_{m,t}[t-1]) = \varepsilon_{m,t} \bigm| (\sigma_{m,t} = \tilde{\sigma}_{m,t}[t-1], \varepsilon_{m,t} = 0) = 0 \)
- \( \tilde{u}_\varepsilon(\varepsilon_{m,t} - \tilde{\varepsilon}_{m,t}) = \sigma_{m,t} \bigm| (\sigma_{m,t} = \tilde{\sigma}_{m,t}[t-1], \varepsilon_{m,t} = 0) = \tilde{\sigma}_{m,t}[t-1] \)

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

\[ \sigma_{m,t}^2 \equiv f_t(\sigma_{m,t}) = h_t(\sigma_{m,t-1}) + k_t(\sigma_{m,t-1}, \eta_{m,t}) \equiv \left( 1 - \delta + \delta \sigma_{m,t-1}^2 + \delta \left( y_{m,t-1}^{**} \right)^2 \right) + \eta_{m,t} \]

Linearizing the two sides of the equation we have:

\[ f_t(\sigma_{m,t}) \simeq f_t(\tilde{\sigma}_{m,t}[t-1]) + \tilde{f}_t(\tilde{\sigma}_{m,t}[t-1])(\sigma_{m,t-1} - \tilde{\sigma}_{m,t-1}[t-1]) \]
\[ h_t(\sigma_{m,t-1}) \simeq h_t(\tilde{\sigma}_{m,t-1}[t-1]) + \tilde{h}_t(\tilde{\sigma}_{m,t-1}[t-1])(\sigma_{m,t-1} - \tilde{\sigma}_{m,t-1}[t-1]) \]

where \( \tilde{f}_t(\tilde{\sigma}_{m,t}[t-1],0) = 2\sigma_{m,t}[\tilde{\sigma}_{m,t}[t-1],0] \) and \( \tilde{h}_t(\tilde{\sigma}_{m,t-1}[t-1],0) = 2\delta \sigma_{m,t-1} \bigm| (\tilde{\sigma}_{m,t-1}[t-1],0) \)
E.4 Long-run restrictions

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

$$y_t = B_{1,t}y_{t-1} + \ldots + B_{p,t}y_{t-p} + [A(\alpha_t)]^{-1} \Sigma_t \xi_t$$

Then, we only need to modify how draws for the $B_t$ block are made. In particular,

1. At iteration $i$, given $A(\alpha_t)^{-1}, \Sigma_t^{-1}$ 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

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

where $B_t^i = \left[ vec \left( B_{1,t}^i \right)^\prime, \ldots, vec \left( B_{p,t}^i \right)^\prime \right]^\prime$.

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

$$D_t^i = J \left( I_{Mp} - B_t^i \right)^{-1} J^\prime \left[ A(\alpha_t)^{-1} \right]^{-1} \Sigma_t^{-1} \quad \text{(E.18)}$$

$$= \left( I_M - B_{1,t}^i - \ldots - B_{p,t}^i \right)^{-1} \left[ A(\alpha_t)^{-1} \right]^{-1} \Sigma_t^{-1}$$

where $J = \begin{bmatrix} I_M & 0_{M \times M} & \cdots & 0_{M \times M} \end{bmatrix}$ is a selection matrix.

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

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

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

and with this construct the restricted draw $\tilde{B}_t^i = \left[ vec \left( B_{1,t}^i \right)^\prime, \ldots, vec \left( \tilde{B}_{p,t}^i \right)^\prime \right]^\prime$. 

12
5. Evaluate whether

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

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.