Vector autoregresivo bayesiano

Alan Ledesma Arista[1]

Índice

[1]Email: alan.ledesma@bcrp.gob.pe

1. BVAR Vs VAR

Principal referencia de esta nota de clase: Karlsson (2013)

- En la práctica, los VARs generan pronósticos pobres a pesar de toda la información insumida
- Las limitaciones de los VARs son generalmente consecuencia de:
  - Problemas de dimensionalidad: $\#(\mathbf{B}) >> \#(\mathbf{y})$.
  - Un VAR estimado suele tener pocos parámetros significativos.
  - Sólo se considera una fuente de información.
- **Propuestas comunes para salvar estas dificultades**:
  - **Estructurar** (o añadir restricciones):
    - Cointegración.
    - Modelos Microfundados (DSGEs).
    - Modelos semi-estructurales (e.g., MPT)
  - FAVAR
  - **BVAR**
  - **BVAR**+FAVAR
  - **BVAR**+DSGE

2. VAR en formato regresión

- La forma reducida VAR es

$$\mathbf{x}_t = \mathbf{b}_d \mathbf{d}_t + \sum_{i=1}^{p} \mathbf{b}_i \mathbf{x}_{t-i} + \mathbf{a}_t \text{ con } \mathbf{a}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.1}$$

con $\mathbf{x}_t, \mathbf{a}_t \in \mathbb{R}^{k\times 1}$, $\mathbf{d}_t \in \mathbb{R}^{k_d\times 1}$; por lo tanto, $\mathbf{b}_d \in \mathbb{R}^{k\times k_d}$ y $\{\mathbf{b}_i\}_{i=1}^{p}, \boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$. El vector de variables endógenas es $\mathbf{x}_t$ mientras que $\mathbf{d}_t$ es un vector de k_d variables determinísticas (por ejemplo, el intercepto)

- La ecuación (2.1) se puede escribir de forma compacta (o en formato regresión) después de tomar las siguientes definiciones
  - $\mathbf{y} = [\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_T]'$ of order $T \times k$
  - $\mathbf{a} = [\mathbf{a}_1\ \mathbf{a}_2\ \dots\ \mathbf{a}_T]'$ of order $T \times k$
  - $n = kp + k_d$
  - $\mathbf{X}_t = [\mathbf{x}'_{t-1}\ \mathbf{x}'_{t-2}\ \dots\ \mathbf{x}_{t-p+1}\ \mathbf{x}_{t-p}\ \mathbf{d}_t]$ of order $1 \times n$
  - $\mathbf{X} = [\mathbf{X}'_1\ \mathbf{X}'_2\ \dots\ \mathbf{X}'_T]'$ of order $T \times n$
  - $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \dots\ \mathbf{b}_{p-1}\ \mathbf{b}_p\ \mathbf{b}_d]'$ of order $n \times k$

- Como resultado, la ecuación (2.1) puede expresarse de la siguiente manera

$$\mathbf{y} = \mathbf{X}\mathbf{B} + \mathbf{a} \text{ con } \mathbf{a} \overset{iid}{\sim} \mathcal{MN}(\mathbf{0}, \mathbf{I}_T, \boldsymbol{\Sigma}) \tag{2.2}$$

- A veces será conveniente utilizar el operador vec en (2.2) para obtener una expresión incluso más parecida a una regresión

$$y = \mathbb{X}\boldsymbol{\beta} + a \text{ con } a \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_T) \tag{2.3}$$

donde $y = \mathrm{vec}(\mathbf{y})$, $a = \mathrm{vec}(\mathbf{a})$, $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$ y $\mathbb{X} = \mathbf{X} \otimes \mathbf{I}_k$

- Ver funciones `get_yXform(·)` y `get_MCO_VAR(·)` del modulo creado para este curso.

3. Función de verosimilitud y estimación MCO

- De la ecuación (2.2) usando el PDF de la normal-matricial, la función de verosimilitud es

$$\mathcal{L}(\mathbf{B},\boldsymbol{\Sigma}) \propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{XB})'(\mathbf{y}-\mathbf{XB})\right]\right\}}{|\boldsymbol{\Sigma}|^{T/2}}$$

- Tome las estimaciones MCO como definiciones

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y} \text{ and } \widehat{\mathbf{S}} = (T-k)\widehat{\boldsymbol{\Sigma}} = (\mathbf{y}-\mathbf{X}\widehat{\mathbf{B}})'(\mathbf{y}-\mathbf{X}\widehat{\mathbf{B}}) \tag{3.1}$$

- Luego, con un álgebra similar a la clase de ayer

$$\begin{aligned}\mathcal{L}(\mathbf{B},\boldsymbol{\Sigma}) &\propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\widehat{\mathbf{S}}+(\widehat{\mathbf{B}}-\mathbf{B})'\mathrm{X}'\mathrm{X}(\widehat{\mathbf{B}}-\mathbf{B}))\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \\ &\propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\widehat{\mathbf{B}}-\mathbf{B})'\mathrm{X}'\mathrm{X}(\widehat{\mathbf{B}}-\mathbf{B})\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{S}}\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{T-n}{2}}}\end{aligned} \tag{3.2}$$

- De manera equivalente, si usamos el operador vec:

$$\mathcal{L}(\mathbf{B},\boldsymbol{\Sigma}) \propto \frac{\exp\left\{-\frac{1}{2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1}\otimes\mathrm{X}'\mathrm{X})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{S}}\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{T-n}{2}}} \tag{3.3}$$

donde $\widehat{\boldsymbol{\beta}} = \mathrm{vec}(\widehat{\mathbf{B}}) = [\mathbf{I}\otimes(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$

4. Distribuciones prior y posterior

Prior no informativo (impropio) de Jeffrey

- Un *prior* no informativo sobre **B**: $p(\mathbf{B}) = 1$, con un *prior* impropio para $\boldsymbol{\Sigma}$: $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{k+1}{2}}$
- La distribución *posterior* is

$$
\begin{aligned}
p(\mathbf{B}, \boldsymbol{\Sigma}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{B}, \boldsymbol{\Sigma}) p(\mathbf{B}) p(\boldsymbol{\Sigma}) = \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Sigma}) p(\boldsymbol{\Sigma}) \\
&\propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\widehat{\mathbf{B}} - \mathbf{B})'\mathbf{X}'\mathbf{X}(\widehat{\mathbf{B}} - \mathbf{B})\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{S}}\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{T+k+1-n}{2}}}
\end{aligned} \tag{4.1}
$$

- Como resultado, la distribución *posterior* de **B** dado $\boldsymbol{\Sigma}$ es

$$
p(\mathbf{B}|\boldsymbol{\Sigma}, \mathbf{y}) \propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\widehat{\mathbf{B}} - \mathbf{B})'\mathbf{X}'\mathbf{X}(\widehat{\mathbf{B}} - \mathbf{B})\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \tag{4.2}
$$

  de donde $\mathbf{B}|\boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{MN}\left(\widehat{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}, \boldsymbol{\Sigma}\right)$ or de manera equivalente $\beta|\boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{N}\left(\widehat{\beta}, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}\right)$

- La distribución *posterior* marginal de $\boldsymbol{\Sigma}$ es

$$
p(\boldsymbol{\Sigma}|\mathbf{y}) \propto \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{S}}\right]\right\}}{|\boldsymbol{\Sigma}|^{\frac{T+k-n+1}{2}}}. \tag{4.3}
$$

  then $\boldsymbol{\Sigma}|\mathbf{y} \sim \mathcal{W}^{-1}\left(\widehat{\mathbf{S}}, T - n\right)$

Resumen de priors y posteriors

- *Prior* no informativo (impropio) de Jeffrey
  - *Prior*: $\Sigma \sim |\Sigma|^{-\frac{k+1}{2}}$ y $p(\mathbf{B}) = 1$
  - Momentos *Posterior*: $\overline{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ y $\overline{\mathbf{S}} = (\mathbf{y} - \mathbf{X}\overline{\mathbf{B}})'(\mathbf{y} - \mathbf{X}\overline{\mathbf{B}})$
  - Distribución *Posterior*: $\Sigma|\mathbf{y} \sim \mathcal{W}^{-1}\left(\overline{\mathbf{S}}, T - n\right)$ y $\mathbf{B}|\Sigma, \mathbf{y} \sim \mathcal{MN}\left(\overline{\mathbf{B}}, (\mathbf{X}'\mathbf{X})^{-1}, \Sigma\right)$ o $\beta|\Sigma, \mathbf{y} \sim \mathcal{N}\left(\overline{\beta}, \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1}\right)$
- Normal-Wishart (Conjugado)
  - *Prior*: $\Sigma \sim \mathcal{W}^{-1}\left(\underline{\mathbf{S}}, \underline{v}\right)$ and $\mathbf{B}|\Sigma \sim \mathcal{MN}(\underline{\mathbf{B}}, \underline{\mathbf{\Omega}}, \Sigma)$ o $\beta|\Sigma \sim \mathcal{N}(\underline{\beta}, \Sigma \otimes \underline{\mathbf{\Omega}})$
  - Momentos *Posterior*: $\overline{\mathbf{\Omega}} = (\underline{\mathbf{\Omega}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$, $\overline{\mathbf{B}} = \overline{\mathbf{\Omega}}(\underline{\mathbf{\Omega}}^{-1}\underline{\mathbf{B}} + \mathbf{X}'\mathbf{y})$ y $\overline{\mathbf{S}} = \underline{\mathbf{S}} + (\mathbf{y} - \mathbf{X}\overline{\mathbf{B}})'(\mathbf{y} - \mathbf{X}\overline{\mathbf{B}}) - (\underline{\mathbf{B}} - \overline{\mathbf{B}})'\underline{\mathbf{\Omega}}^{-1}(\underline{\mathbf{B}} - \overline{\mathbf{B}})$
  - Distribución *Posterior*: $\Sigma|\mathbf{y} \sim \mathcal{W}^{-1}\left(\overline{\mathbf{S}}, T + \underline{\nu}\right)$ y $\mathbf{B}|\Sigma, \mathbf{y} \sim \mathcal{MN}\left(\overline{\mathbf{B}}, \overline{\mathbf{\Omega}}, \Sigma\right)$ o $\beta|\Sigma, \mathbf{y} \sim \mathcal{N}\left(\overline{\beta}, \Sigma \otimes \overline{\mathbf{\Omega}}\right)$
- Normal-impropio (Zellner)
  - *Prior*: $\Sigma \sim |\Sigma|^{-\frac{k+1}{2}}$ y $\beta \sim \mathcal{N}(\underline{\beta}, \underline{\mathbf{\Psi}})$
  - Momentos *Posterior*: $\overline{\mathbf{\Psi}} = (\underline{\mathbf{\Psi}}^{-1} + \Sigma^{-1} \otimes (\mathbf{X}'\mathbf{X}))^{-1}$, $\overline{\beta} = \overline{\mathbf{\Psi}}(\underline{\mathbf{\Psi}}^{-1}\underline{\beta} + \mathrm{vec}(\mathbf{X}'\mathbf{y}\Sigma^{-1}))$ y $\overline{\mathbf{S}} = (\mathbf{y} - \mathbf{X}\mathbf{B})'(\mathbf{y} - \mathbf{X}\mathbf{B})$
    Note que $\overline{\mathbf{\Psi}}$ y $\overline{\beta}$ son función de Σ mientras $\overline{\mathbf{S}}$ es función de $\mathbf{B}$
  - Distribución *Posterior*: $\Sigma|\mathbf{B}, \mathbf{y} \sim \mathcal{W}^{-1}\left(\overline{\mathbf{S}}, T\right)$ y $\beta|\Sigma, \mathbf{y} \sim \mathcal{N}\left(\overline{\beta}, \overline{\mathbf{\Psi}}\right)$
- Normal-Wishart (Independent)
  - *Prior*: $\Sigma \sim \mathcal{W}^{-1}\left(\underline{\mathbf{S}}, \underline{v}\right)$ and $\beta \sim \mathcal{N}(\underline{\beta}, \underline{\mathbf{\Psi}})$
  - Momentos *Posterior*: $\overline{\mathbf{\Psi}} = (\underline{\mathbf{\Psi}}^{-1} + \Sigma^{-1} \otimes (\mathbf{X}'\mathbf{X}))^{-1}$, $\overline{\beta} = \overline{\mathbf{\Psi}}(\underline{\mathbf{\Psi}}^{-1}\underline{\beta} + \mathrm{vec}(\mathbf{X}'\mathbf{y}\Sigma^{-1}))$ y $\overline{\mathbf{S}} = \underline{\mathbf{S}} + (\mathbf{y} - \mathbf{X}\mathbf{B})'(\mathbf{y} - \mathbf{X}\mathbf{B})$
    Note que $\overline{\mathbf{\Psi}}$ y $\overline{\beta}$ son función de Σ mientras $\overline{\mathbf{S}}$ es función de $\mathbf{B}$
  - Distribución *Posterior*: $\Sigma|\mathbf{B}, \mathbf{y} \sim \mathcal{W}^{-1}\left(\overline{\mathbf{S}}, T + \underline{\nu}\right)$ y $\beta|\Sigma, \mathbf{y} \sim \mathcal{N}\left(\overline{\beta}, \overline{\mathbf{\Psi}}\right)$

Simulaciones de $\mathcal{W}^{-1}$ y $\mathcal{MN}$

- **Inverse Wishart**: $\mathcal{W}^{-1}(\Omega, v)$
  - Construir matriz triangular inferior Q de orden $\dim(\Omega)$ con elementos $q_{ii}^2 \sim \chi^2(v - \dim(\Omega) + i)$ y $q_{ij} \sim \mathcal{Z}$ entonces $QQ' \sim \mathtt{IW}(I, v)$
  - Con $L = \mathrm{chol}(\Omega)$ (tal que $LL' = \Omega$) calcular $C = L(Q')^{-1}$ entonces $\Omega^{draw} = CC'$
- **Matricvariate Normal**: $\mathcal{MN}_{pk+1,k}(\mu, \Omega, \Sigma)$
  - Calcular $L_\Omega = \mathrm{chol}(\Omega)$ y $L_\Sigma = \mathrm{chol}(\Sigma)$ (tal que $L_\Omega L'_\Omega = \Omega$ y $L_\Sigma L'_\Sigma = \Sigma$)
  - Simular la matriz Z de $\mathrm{orden}(\mu)$ con elementos $z_{ij} \sim \mathcal{Z}$
  - Entonces $\mu^{draw} = \mu + L_\Sigma Z L'_\Omega$
- Ver funciones `LT_IWrnd`$(\cdot)$ y `MatNormrnd`$(\cdot)$ del modulo `VARstuff.py`

Proyección

La proyección se computa vía Gibbs-sampling:

1. Seleccionar $(\mathbf{B}^{(0)}, \mathbf{\Sigma}^{(0)}) = (\underline{\mathbf{B}}, \underline{\mathbf{\Sigma}})$
2. Calcular $\overline{\mathbf{S}}^{(r-1)}$ (que podría depender de $\mathbf{B}^{(r-1)}$) y simular $\mathbf{\Sigma}^{(r)}$ de su distribución correspondiente
3. Calcular $\overline{\beta}^{(r)}$ y $\overline{\mathbf{\Omega}}^{(r)}$ o $\overline{\mathbf{\Psi}}^{(r)}$ (que podría depender de $\mathbf{\Sigma}^{(r)}$) y simular $\beta^{(r)}$ de su distribución correspondiente
4. Simular $\{\mathbf{a}_t^{(r)}\}_{t=T+1}^{T+F}$ de su distribución (i.e., $\mathcal{MN}(\mathbf{0}, \mathbf{I}_T, \mathbf{\Sigma}^{(r)})$)
5. Realizar un pronóstico $\{\mathbf{x}_t^{(r)}\}_{t=T+1}^{T+F}$:

$$\mathbf{x}_{T+h}^{(r)} = \mathbf{b}_d \mathbf{d}_{T+h} + \sum_{i=1}^{p} \mathbf{b}_i \mathbf{x}_{T+h-i}^{(r)} + \mathbf{a}_{T+h}^{(r)} \text{ with } \mathbf{x}_{T+h-i}^{(r)} = \begin{cases} \mathbf{x}_{T+h-i}, & \text{if } h \leq i \\ \mathbf{x}_{T+h-i}^{(r)}, & \text{if } h > i \end{cases}$$

6. Repetir $B+R$ veces los pasos 2 a 5 para obtener la secuencia $\{\{\mathbf{x}_t^{(r)}\}_{t=T+1}^{T+F}\}_{r=1}^{B+R}$ y eliminar las primeras B iteraciones

Ver función `get_BVARforvast(·)` del modulo `VARstuff.py` y `Lab3-VAR.ipynb`

5. Formulación del prior

El prior de Minnesota (o Litterman)

- Publicado en Litterman (1986)
- La mayoría de las variables macroeconómicas exhiben un comportamiento similar al de un paseo aleatorio. Por tanto, el *prior* podría formularse de modo que el VAR corresponda a un vector de paseos aleatorios independientes
  - Sea $b_{m,j}$ representando el coeficiente en la fila j y columna i de la matriz en el rezago ℓ si $m = (\ell - 1)k + i$ o coeficiente en la fila j de componente determinista i si $m = pk + i$
  - Entonces $b_{m,j} \sim \mathcal{N}(\cdot)$ con

$$\mathrm{E}b_{m,j} = \begin{cases} \delta_i, & \text{si } m = i \\ 0, & \text{en caso contrario} \end{cases} \text{ y } \mathrm{var}(b_{m,j}) = \begin{cases} \frac{\pi_1^2}{\ell^{2\pi_3}} & \text{if } j = i \text{ and } m = (\ell-1)k + i \\ \frac{\pi_1^2 \pi_2^2}{\ell^{2\pi_3}} \frac{\sigma_j^2}{\sigma_i^2} & \text{if } j \neq i \text{ and } m = (\ell-1)k + i \\ \infty & \text{variables determinísticas (i.e., } m > pk) \end{cases}$$

Notas:

- $\delta_i = 0$ ($\delta_i = 1$) si x_i se cree que es estacionario (no estacionario)
- $\sigma_i^2 = \widehat{\boldsymbol{\Sigma}}_{[i,i]}$
- π_1 es un hiperparámetro al que se hace referencia como la rigidez general del anterior: si $\pi_1 \to 0$ ($\pi_1 \to \infty$) la media posterior coincide con la media anterior (estimación de MCO)
- π_2 es un hiperparámetro al que se hace referencia como la rigidez relativa de otros parámetros en el anterior (normalmente establecido en 1)
- π_3 es un hiperparámetro al que se hace referencia como la tasa de decaimiento del retardo en el anterior (normalmente se establece en 1)
- Estos hiperparámetros se establecen comúnmente como aquellos que maximizan la probabilidad del modelo o aquellos que minimizan el error de pronóstico.

- Visualicemos el prior en un caso simplificado correspondiente a $k=2$, $p=2$, $k_d=1$: $\mathbf{B}=\begin{bmatrix} b_{11}(1) & b_{21}(1) \\ b_{12}(1) & b_{22}(1) \\ b_{11}(2) & b_{21}(2) \\ b_{12}(2) & b_{22}(2) \\ c_1 & c_2 \end{bmatrix}$

$$\boldsymbol{\beta}=\operatorname{vec}(\mathbf{B})=\begin{bmatrix} b_{11}(1) \\ b_{12}(1) \\ b_{11}(2) \\ b_{12}(2) \\ c_1 \\ b_{21}(1) \\ b_{22}(1) \\ b_{21}(2) \\ b_{22}(2) \\ c_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \delta_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \delta_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \pi_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \pi_1^2\pi_2^2\frac{\sigma_1^2}{\sigma_2^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\pi_1^2}{2^{2\pi_3}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\pi_1^2\pi_2^2}{2^{2\pi_3}}\frac{\sigma_1^2}{\sigma_2^2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \infty & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_1^2\pi_2^2\frac{\sigma_2^2}{\sigma_1^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \pi_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\pi_1^2\pi_2^2}{2^{2\pi_3}}\frac{\sigma_2^2}{\sigma_1^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\pi_1^2}{2^{2\pi_3}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \infty \end{bmatrix}\right) \tag{5.1}$$

- La mayoría de las veces, los momentos posteriores son computacionalmente costosos. Por ejemplo, para calcular $overline\boldsymbol{\Psi}$ se requiere la inversa de una matriz cuadrada con kn filas. En Bańbura et al. (2010), $p=13$, $k=113$ and $k_d=1$; por lo tanto, $kn=166110$.

- Considerando que un VAR es un SUR, una estimación ecuación por ecuación es factible y genera cierta eficiencia computacional
  - La columna j de **B** (i.e., coeficientes de la ecuación para la variable j) podría estimarse en la siguiente regresión

$$\mathbf{y}_j=\mathbf{X}\mathbf{B}_j+\mathbf{a}_j \text{ con } \mathbf{a}_j\sim\mathcal{N}(0,\sigma_j^2) \text{ donde } \mathbf{y}_j \text{ es la columna } j \text{ de } \mathbf{y}\text{; por lo tanto, } \hat{\mathbf{B}}_j=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j \tag{5.2}$$

  - Además, se podrían diseñar datos artificiales para implementar el anterior en una sola regresión

El prior de Litterman con data extendida

- Para realizar una estimación bayesiana de $\mathbf{B}_j$ con el *prior* de Litterman, se puede añadir a la regresión (5.2) lo siguiente

$$\mathbf{r}_j = \mathbf{R}_j \mathbf{B}_j + \widetilde{\mathbf{a}}_j; \text{ por lo tanto, } \begin{bmatrix} \mathbf{r}_j \\ \mathbf{y}_j \end{bmatrix} = \begin{bmatrix} \mathbf{R}_j \\ \mathbf{X} \end{bmatrix} \mathbf{B}_j + \begin{bmatrix} \widetilde{\mathbf{a}}_j \\ \mathbf{a}_j \end{bmatrix} \tag{5.3}$$

donde $\mathbf{r}_j = \begin{bmatrix} \left\{ \mathrm{E} b_{m,j} \left(\frac{\sigma_j}{\mathrm{var}(b_{m,j})^{1/2}} \right) \right\}_{m=1}^{n} \\ \underline{\mathbf{S}}_{[:,j]}/n \end{bmatrix}$ (con $\underline{\mathbf{S}}_{[j,:]}$ corresponde a la j^{th} columna de $\underline{\mathbf{S}}$) mientras que $\mathbf{R}_j = \begin{bmatrix} \mathrm{diag}\left(\left\{ \frac{\sigma_j}{\mathrm{var}(b_{m,j})^{1/2}} \right\}_{m=1}^{n} \right) \\ \mathbf{0}_{k,n} \end{bmatrix}$

- Por lo tanto, la media y varianza posterior de $\mathbf{B}_j$ corresponde a la estimación MCO de (5.3)

$$\overline{\mathbf{B}}_j = \left(\mathbf{R}_j' \mathbf{R}_j + \mathbf{X}'\mathbf{X} \right)^{-1} \left(\mathbf{R}_j' \mathbf{r}_j + \mathbf{X}' \mathbf{y}_j \right) \text{ y } \overline{\mathbf{V}}_j = \sigma_j^2 \left(\mathbf{R}_j' \mathbf{R}_j + \mathbf{X}'\mathbf{X} \right)^{-1} \tag{5.4}$$

- En nuestro ejemplo

$$\mathbf{r}_1 = \begin{bmatrix} \delta_1 \frac{\sigma_1}{\pi_1} \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_1 \\ 0 \end{bmatrix}, \ \mathbf{R}_1 = \begin{bmatrix} \frac{\sigma_1}{\pi_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sigma_2}{\pi_1 \pi_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{2^{\pi_3}\sigma_1}{\pi_1} & 0 & 0 \\ 0 & 0 & 0 & \frac{2^{\pi_3}\sigma_2}{\pi_1\pi_2} & 0 \\ 0 & 0 & 0 & 0 & \epsilon \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ y } \mathbf{r}_2 = \begin{bmatrix} 0 \\ \delta_2 \frac{\sigma_2}{\pi_1} \\ 0 \\ 0 \\ 0 \\ 0 \\ \sigma_2 \end{bmatrix}, \ \mathbf{R}_2 = \begin{bmatrix} \frac{\sigma_1}{\pi_1 \pi_2} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sigma_2}{\pi_1} & 0 & 0 & 0 \\ 0 & 0 & \frac{2^{\pi_3}\sigma_1}{\pi_1\pi_2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2^{\pi_3}\sigma_2}{\pi_1} & 0 \\ 0 & 0 & 0 & 0 & \epsilon \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.5}$$

con $\epsilon > 0$ pero $\epsilon \to 0$

Variaciones comunes en el prior Litterman

- **Variables determinísticas**: Configurar las desviaciones estándar *prior* como $\pi_1\pi_4\sigma_j$ (en lugar de ∞), esto tiene la ventaja de conducir a un *prior* adecuado para los coeficientes en variables deterministas sin dejar de ser informativo al establecer π_4 grande.

- **Variables "exógenas"**: Configurar las desviaciones estándar *prior* como $(\pi_1\pi_5/\ell^{\pi_3})/(\sigma_j/\sigma_r)$, para el rezago ℓ de las variables "endógenas" r en la ecuación de la variable "exógena" j, $i=(\ell-1)m+r$: Esto es útil, por ejemplo, al modelar una pequeña economía abierta con variables del "resto del mundo" incluidas en el modelo. Establecer π_5 pequeño reduce el coeficiente agresivamente hacia cero y nos permite expresar que el resto de las variables del mundo son esencialmente exógenas a la economía doméstica.

- **Prior de suma de coeficientes**: Este *prior* expresa la noción a *priori* de que la suma de los coeficientes de los rezagos propios es 1 y la suma de los coeficientes de los rezagos de cada una de las otras variables es 0. Para implementar esto agregue k filas a $\mathbf{R}_j$ que son cero excepto en las posiciones p de la i^{th} fila correspondiente a la variable $i=1,...,k$; es decir, la fila i viene dada por $(\overline{\mathbf{w}}_i'\otimes\mathbf{j}_p',\mathbf{0}_{n,k_d})$ donde los ceros corresponden a las variables deterministas, el elemento i de $\overline{\mathbf{w}}$ es $\bar{y}_{0,i}\sigma_i/(\pi_1\pi_6\sigma_j)$ para $\bar{y}_{0,i}=p^{-1}\sum_{t=1-p}^{0}y_{t,i}$ el promedio de las condiciones iniciales para la variable i y los elementos restantes $k-1$ cero, $\mathbf{j}_p$ es un vector $p\times 1$ de unos. Además, agregue k elementos a $\mathbf{r}_j$ con el elemento j^{th} igual a $\bar{y}_{0,i}/(\pi_1\pi_6)$ El previo induce una correlación entre los coeficientes de la misma variable y fuerza al modelo a caminar al azar con una posible deriva para la variable j como $\pi_6\to 0$

  En nuestro ejemplo, las filas de k adicionales para la primera ecuación tienen el siguiente aspecto:

  $$\mathbf{r}_1^{sc}=\begin{bmatrix}\bar{y}_{0,1}/(\pi_1\pi_6)\\ 0\end{bmatrix},\ \mathbf{R}_1^{sc}=\begin{bmatrix}\bar{y}_{0,1}/(\pi_1\pi_6) & \bar{y}_{0,1}/(\pi_1\pi_6) & 0 & 0 & 0\\ 0 & 0 & \bar{y}_{0,1}/(\pi_1\pi_6)\frac{\sigma_2}{\sigma_1} & \bar{y}_{0,1}/(\pi_1\pi_6)\frac{\sigma_2}{\sigma_1} & 0\end{bmatrix}$$

- **Prior de observaciones iniciales ficticias**: agregue una fila $(\overline{y}_0\otimes\mathbf{j}_p',\mathbf{0}_{1,k_d})/(\pi_1\pi_7\sigma_j)$ a $\mathbf{R}_j$ y $\overline{y}_{0,j}/(\pi_1\pi_7\sigma_j)$ a $\mathbf{r}_j$. Este *prior* también implica que las observaciones iniciales son un buen pronóstico sin imponer valores de parámetros específicos e induce una correlación previa entre todos los parámetros de la ecuación. Cuando $\pi_7\to 0$, lo anterior implica que todas las variables son estacionarias con media $\overline{\mathbf{y}}_0$ o que hay componentes de raíz unitaria sin intercepto

  En nuestro ejemplo, las filas de k adicionales para la primera ecuación tienen el siguiente aspecto:

  $$\mathbf{r}_1^{dio}=\overline{y}_{0,1}/(\pi_1\pi_7\sigma_1),\ \mathbf{R}_1^{dio}=\begin{bmatrix}\bar{y}_{0,1}/(\pi_1\pi_7\sigma_1) & \bar{y}_{0,1}/(\pi_1\pi_7\sigma_1) & \bar{y}_{0,2}/(\pi_1\pi_7\sigma_1) & \bar{y}_{0,2}/(\pi_1\pi_7\sigma_1) & 0\end{bmatrix}$$

6. BVAR de gran escala

- Los *priors* no informativo (impropio) de Jeffrey y el Normal-Wishart (conjugado) son adecuados para VAR con un gran número de variables. Esto se debe a que, la matriz cuadrada más grande para invertir tiene solo n filas (frente a las filas kn en los otros dos casos).

- El *prior* de Litterman puede ser implementado con el prior conjugado Normal-Wishart si $\pi_2 = 1$ ya que la varianza anterior podría escribirse como $\text{var}(\boldsymbol{\beta}) = \underline{\boldsymbol{\Sigma}} \otimes \underline{\boldsymbol{\Omega}}$. En este caso $\underline{\boldsymbol{\Sigma}} = \text{diag}(\{\sigma_i^2\}_{i=1}^k)$ y $\underline{\boldsymbol{\Omega}} = \text{diag}([1/\sigma_1^2, 1/\sigma_2^2, ..., 1/\sigma_k^2] \otimes \mathbf{J}_p^{2\pi_3}, \infty)\pi_1^2$, donde $\mathbf{J}_p^{2\pi_3}$ es un abuso de notación, ya que se refiere a una potencia elemento por elemento. En cuanto a la matriz de covarianza, sus momentos *prior* son $\underline{v} = k$ y $\underline{\mathbf{S}} = k\underline{\boldsymbol{\Sigma}}$. Por tanto, los momentos posteriores son
  - *Prior*: $\Sigma \sim \mathcal{W}^{-1}(\underline{\mathbf{S}}, k)$ y $\mathbf{B}|\Sigma \sim \mathcal{MN}(\underline{\mathbf{B}}, \underline{\boldsymbol{\Omega}}, \Sigma)$
  - Momentos posteriores: $\overline{\boldsymbol{\Omega}} = (\underline{\boldsymbol{\Omega}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$, $\overline{\mathbf{B}} = \overline{\boldsymbol{\Omega}}(\underline{\boldsymbol{\Omega}}^{-1}\underline{\mathbf{B}} + \mathbf{X}'\mathbf{y})$ y $\overline{\mathbf{S}} = (\mathbf{y} - \mathbf{X}\overline{\mathbf{B}})'(\mathbf{y} - \mathbf{X}\overline{\mathbf{B}}) - (\underline{\mathbf{B}} - \overline{\mathbf{B}})'\underline{\boldsymbol{\Omega}}^{-1}(\underline{\mathbf{B}} - \overline{\mathbf{B}})$
  - Distribuciones posteriores: $\Sigma|\mathbf{y} \sim \mathcal{W}^{-1}\left(\overline{\mathbf{S}}, T + k + 2\right)$ y $\mathbf{B}|\Sigma, \mathbf{y} \sim \mathcal{MN}\left(\overline{\mathbf{B}}, \overline{\boldsymbol{\Omega}}, \Sigma\right)$

- Sin embargo, más eficiente desde el punto de vista computacional es implementar Litterman antes (con $\pi_2 = 1$) con datos artificiales. Para este caso, los datos podrían extenderse a todo el sistema VAR en lugar de ecuación por ecuación, ya que $\mathbf{R}_j$ resulta independiente de j.

  $$\mathbf{r} = [\mathbf{r}_1 \ \ldots \ \mathbf{r}_k] \text{ and } \mathbf{R}_j = \mathbf{R}_i = \mathbf{R}. \tag{6.1}$$

  Note que $\underline{\mathbf{B}} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{r}$, $\underline{\boldsymbol{\Omega}} = (\mathbf{R}'\mathbf{R})^{-1}$ Por lo tanto, el *prior* no informative (impropio) de Jeffreys puede aplicarse a siguiente data extendida:

  $$\begin{bmatrix} \mathbf{r} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ \mathbf{X} \end{bmatrix} \mathbf{B} + \begin{bmatrix} \widetilde{\mathbf{a}} \\ \mathbf{a} \end{bmatrix} \tag{6.2}$$

  - *Prior*: $\Sigma \sim |\Sigma|^{-\frac{k+1}{2}}$ y $p(\mathbf{B}) = 1$
  - Los momentos y la distribución posteriores replican los del *prior* Normal-Wishart (conjugado)

- Pro: k puede ser 'grande'

- Con: Se pierde algo de flexibilidad debido a la configuración del *prior*

En Bańbura et al. (2010)

- Observaciones 'dummy':

$$\mathbf{y}_d^{lit} = \begin{bmatrix} \mathrm{diag}(\delta_1\sigma_1 \ \ldots \ \delta_k\sigma_k)/\lambda \\ \mathbf{0}_{(p-1)k\times k} \\ \mathrm{diag}(\sigma_1 \ \ldots \ \sigma_k) \\ \mathbf{0}_{1\times k} \end{bmatrix} \text{ y } \mathbf{X}_d^{lit} = \begin{bmatrix} \mathbf{J}_p \otimes \mathrm{diag}(\sigma_1 \ \ldots \ \sigma_k)/\lambda & \mathbf{0}_{pk\times 1} \\ \mathbf{0}_{k\times pk} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times pk} & \varepsilon \end{bmatrix}$$

  donde $\mathbf{J}_p = \mathrm{diag}(1 \ 2 \ \ldots \ p)$ y ε es un número positivo 'muy pequeño'

- También se muestra que el prior de suma de coeficientes se puede implementar con

$$\mathbf{y}_d^{sc} = \mathrm{diag}(\mu_1\sigma_1 \ \ldots \ \mu_k\sigma_k)/\tau \text{ y } \mathbf{X}_d^{sc} = [\mathbf{1}_{1\times p} \otimes \mathbf{y}_d^{sc} \quad \mathbf{0}_{k\times 1}]$$

  donde τ es el hiperparámetro que controla la fuerza del prior: si $\tau \to 0$ ($\tau \to \infty$) la posterior es igual al prior (coincide con el estimador MCO)

- Ver funciones `get_DumObsLitterman`$(\cdot)$ y `get_DumObsSumCoef`$(\cdot)$

7. Un prior jerárquico para los hiperparámetros

- Selección de hiperparámetros: i. Valores predeterminados similares a los utilizados por Litterman, ii. minimizar los errores de proyección sobre una muestra de entrenamiento o iii. de una forma empírica, maximizando la probabilidad marginal con respecto a los hiperparámetros.

- Alternativamente Giannone et al. (2015) sugieren un enfoque más flexible en el que se agrega una capa más a la estructura *prior* colocando un *prior* sobre los hiperparámetros de manera jerárquica. Para ello se apilan los hiperparámetros en el vector $\boldsymbol{\delta}$ y se trabaja con la familia Normal-Wishart de distribuciones *prior*, la estructura *prior* se convierte en

$$p(\mathbf{B}|\Sigma, \boldsymbol{\delta})p(\Sigma, |\boldsymbol{\delta})p(\boldsymbol{\delta}) \tag{7.1}$$

cuya posterior es

$$\mathcal{L}(\mathbf{B}, \Sigma, \boldsymbol{\delta})p(\mathbf{B}|\Sigma, \boldsymbol{\delta})p(\Sigma, |\boldsymbol{\delta})p(\boldsymbol{\delta}) \tag{7.2}$$

- Simulación: Metropolis-Hastings-within-Gibbs sampling

  - Simular candidato, $\boldsymbol{\delta}^*$ desde $\boldsymbol{\delta}^* \sim \mathcal{N}(\boldsymbol{\delta}^{(r-1)}, c\mathbf{H}^{-1})$ donde $\mathbf{H}$ es el hessiano del negativo del logposterior para $\boldsymbol{\delta}$. Fijar $\boldsymbol{\delta}^{(r)} = \boldsymbol{\delta}^*$ con probabilidad $\alpha(\cdot)$, de lo contrario, fijar $\boldsymbol{\delta}^{(r)} = \boldsymbol{\delta}^{(r-1)}$ (repetir este paso B veces)
  - Simular $\boldsymbol{\Sigma}^{(r)}$ desde el condicional posterior $\boldsymbol{\Sigma}^{(j)}|\mathbf{y}, \boldsymbol{\delta}^{(r)}$
  - Simular $\mathbf{B}^{(r)}$ desde el condicional posterior $\mathbf{B}^{(r)}|\mathbf{y}, \boldsymbol{\Sigma}^{(j)}, \boldsymbol{\delta}^{(r)}$
  - Simular la medida de interés (previsión, IRF, ...)
  - Calcule diferentes momentos de la medida de interés vía Monte Carlo

8. Structural BVAR

Cholesky

- Dentro del enfoque de simulación, en cualquier momento que $\boldsymbol{\Sigma}^{(r)}$ es simulado, la descomposición triangular inferior de Cholesky podría calcularse $\mathbf{C}^{(r)} = \mathrm{chol}(\boldsymbol{\Sigma}^{(r)})$ que identifica el VAR estructural
  - Considere el $\mathrm{SVAR(p)}$: $\mathbf{x}_t = \mathbf{b}_0 + \sum_{i=1}^{p} \mathbf{b}_i \mathbf{x}_{t-i} + \mathbf{C}\mathbf{u}_t$
  - Su forma '*companion*' es:
    - $\boldsymbol{\mathcal{X}}_t \equiv [\mathbf{x}_t' \ \mathbf{x}_{t-1}' \ \dots \ \mathbf{x}_{t-p+1}']'$ de dimensiones $kp \times 1$
    - $\boldsymbol{\mathcal{B}} \equiv \begin{bmatrix} \mathbf{B}'_{[1:(p-1)k,:]} & \mathbf{B}'_{[((p-1)k+1):pk,:]} \\ \mathbf{I}_{(p-1)k} & \mathbf{0}_{(p-1)k,k} \end{bmatrix}$, $\boldsymbol{\mathcal{B}}_d \equiv \begin{bmatrix} \mathbf{b}_d \\ \mathbf{0}_{(p-1)k,k} \end{bmatrix}$ y $\boldsymbol{\mathcal{C}} \equiv \begin{bmatrix} \mathbf{C} \\ \mathbf{0}_{(p-1)k,k} \end{bmatrix}$ de dimensiones $kp \times kp$, $kp \times k_d$ y $kp \times k$, respectivamente
  - Por lo tanto, el $\mathrm{SVAR(p)}$ se puede escribir como un $\mathrm{SVAR(1)}$: $\boldsymbol{\mathcal{X}}_t = \boldsymbol{\mathcal{B}}\boldsymbol{\mathcal{X}}_{t-1} + \boldsymbol{\mathcal{B}}_d \mathbf{d}_t + \boldsymbol{\mathcal{C}}\mathbf{u}_t$. Se puede demostrar que

$$\boldsymbol{\mathcal{X}}_t = \sum_{k=0}^{\ell} \boldsymbol{\mathcal{B}}^k \boldsymbol{\mathcal{B}}_0 + \boldsymbol{\mathcal{B}}^{\ell+1} \boldsymbol{\mathcal{X}}_{t-\ell-1} + \sum_{k=0}^{\ell-1} \boldsymbol{\mathcal{B}}^k \boldsymbol{\mathcal{C}} \mathbf{u}_{t-k} + \boldsymbol{\mathcal{B}}^{\ell} \boldsymbol{\mathcal{C}} \mathbf{u}_{t-\ell};$$

    por lo tanto,

$$\frac{\partial \boldsymbol{\mathcal{X}}_t}{\partial \mathbf{u}'_{t-\ell}} = \boldsymbol{\mathcal{B}}^{\ell} \boldsymbol{\mathcal{C}}$$

  - Sea $\mathbf{J}_{n,k}$ ser un vector de ceros de dimensión $n \times 1$ con un 1 en la posición k entonces $x_{it} = \mathbf{J}'_{kp,i} \boldsymbol{\mathcal{X}}_t$ y $u_{jt-\ell} = \mathbf{J}'_{k,j} \mathbf{u}_{t-\ell}$; por lo tanto:

$$\mathrm{IRF}_{ij}(\ell) = \frac{\partial x_{it}}{\partial u_{j,t-\ell}} = \mathbf{J}'_{kp,i} \boldsymbol{\mathcal{B}}^{\ell} \boldsymbol{\mathcal{C}} \mathbf{J}_{k,j}$$

- Como resultado, se simulan muchos respuestas a los impulsos $\{\mathrm{IRF}^{(r)}\}_{r=B+1}^{B+R}$ que se integran vía Monte Carlo
- Ver función `get_CompanionB(·)` y `get_IRF_Chol_BVAR(·)` del módulo `VARstuff.py`

Restricciones de signos, ceros y narrativas

- Dentro del enfoque de simulación, en cualquier momento que $\Sigma^{(r)}$ es simulado, la descomposición triangular inferior de Cholesky podría calcularse $\mathbf{L}^{(r)} = \mathrm{chol}(\Sigma^{(r)})$ y una matriz normalizada ortogonal $\mathbf{Q}^{(r)}$ también puede simularse de manera que $\mathbf{Q}^{(r)\prime}\mathbf{Q}^{(r)} = \mathbf{I}_k$. Por tanto, la matriz de identificación es $\mathbf{C}^{(r)} = \mathbf{Q}^{(r)}\mathbf{L}^{(r)}$

- Simulando **Q**
  - Si solo hay restricciones de signos
    - Simular $\mathbf{Q}^*$ usando $\mathcal{MN}(\mathbf{0}_{k,k}, \mathbf{I}_k, \mathbf{I}_k)$
    - Computar la descomposición QR de $\mathbf{Q}^*$ y la matriz unitaria resultante se denomina **Q**
  - Si hay restricciones de signos y ceros, **Q** debe simularse tal que las restricciones cero siempre se cumplan (ver teorema 4 en Arias et al. (2018))

- Se simulan muchas respuestas a impulsos $\{\mathrm{IRF}^{(r)}\}_{r=B+1}^{B+\overline{R}}$, de los cuales solo se mantienen aquellos que satisfacen las restricciones de signos $\{\mathrm{IRF}_Q^{(r)}\}_{r=B+1}^{B+R}$, luego se integra vía Monte Carlo

- **Nota**: **Restricciones narrativas** (Antolín-Díaz and Rubio-Ramírez (2018))
  - Estos podrían añadirse además de las restricciones de signos y ceros
  - Idea simple: Los choques estructurales históricos deben ser consistentes con eventos conocidos en la historia. Por ejemplo, en el caso de los datos de EE. UU., si los precios del petróleo son parte del sistema: el choque de oferta de petróleo debe tomar valores negativos en diciembre de 1978-enero de 1979, septiembre-octubre de 1980, agosto de 1990, diciembre de 2002, marzo de 2003 y febrero de 2011
  - Tipo de restricciones narrativas
    - Restricciones sobre **signos de los choques estructurales**: similar al ejemplo anterior
    - **Tipo A**: restricciones en la **descomposición histórica**: se puede especificar que un choque dado fue el **más importante (o menos importante)** impulsor del cambio inesperado en una variable durante algunos períodos
    - **Tipo B**: restricciones en la **descomposición histórica**: se puede especificar que un impacto dado fue el **factor abrumador (o insignificante)** del cambio inesperado en una variable durante algunos períodos

9. Extensiones

Un BVAR de volatilidad estocástica

- El supuesto de errores homocedásticos puede ser cuestionable
- Cogley and Sargent (2005) y Primiceri (2005) elaboran el marco para un BVAR con volatilidad estocástica (dentro de una configuración más general a la que se revisa más adelante)
- La idea es sencilla e implica cambios menores en el marco BVAR estudiado
- El VAR se especifica de la siguiente manera

$$\mathbf{x}_t = \mathbf{b}_d \mathbf{d}_t + \sum_{i=1}^{p} \mathbf{b}_i \mathbf{x}_{t-i} + \mathbf{a}_t \text{ con } \mathbf{a}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{9.1}$$

  Observe que la matriz de covarianza ahora varía en el tiempo

- Siguiendo a Cogley and Sargent (2005), la matriz de covarianza se puede descomponer de la siguiente manera

$$\boldsymbol{\Sigma}_t = \mathsf{F} \boldsymbol{\Lambda}_t \mathsf{F}' \tag{9.2}$$

  donde **F** es una matriz triangular inferior con unos en la diagonal principal, mientras que $\boldsymbol{\Lambda}_t = \operatorname{diag}(\{\bar{s}_i \exp(\lambda_{i,t})\}_{i=1}^{k})$ y

$$\lambda_{i,t} = \gamma_i \lambda_{i,t-1} + v_{i,t} \text{ con } v_{i,t} \sim \mathcal{N}(0, \phi_i) \tag{9.3}$$

- Parámetros a estimar: $\mathbf{b}_d$,$\{\mathbf{b}_i\}_{i=1}^{p}$, **F**, $\{\gamma_i\}_{i=1}^{k}$, $\{\lambda_{i,t}\}_{i=1,t=1}^{i=k,t=T}$ y ϕ.

BVAR con parámetros cambiantes en el tiempo (TVP-BVAR)

- ¿Cómo podríamos estimar rápidamente un TVP-VAR?
  Use el filtro de Kalman en el sistema de espacio de estado a continuación

$$\mathbf{x}_t = (\mathbf{I}_k \otimes \mathbf{X}_t)\boldsymbol{\beta}_t + \mathbf{a}_t \text{ with } \mathbf{a}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \text{(Space)}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{v}_t \text{ with } \mathbf{v}_t \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \qquad \text{(State)}$$

  donde $\mathbf{X}_t = [\mathbf{x}'_{t-1}\ \mathbf{x}'_{t-2}\ ...\ \mathbf{x}'_{t-p+1}\ \mathbf{x}'_{t-p}\ \mathbf{d}'_t]$. Así, el TVP-VAR podría entenderse como una reinterpretación de un modelo de espacio de estados donde $\boldsymbol{\beta}_t$ son los estados y se pueden obtener a través del filtro de Kalman

- Para un contexto bayesiano con homocedasticidad (TVP-BVAR), véase la sección 6.1. de la guía técnica BEAR

- Para un contexto bayesiano con heterocedasticidad (TVP-SV-BVAR), véase la sección 6.2. de la guía técnica BEAR

- Sims (1993): Un TVP-SV-BVAR produjo mejoras en la precisión de proyecciones para la variable de precios de activos, precisiones comparables o ligeramente mejores para las proyecciones de variables reales y preciciones ligeramente peores para la proyección de tasas de interés en comparación con el modelo BVAR original (Litterman).

- Canova (2007) proyecta la tasa de inflación de los países del G7 utilizando una variedad de modelos. En general, el modelo con el conjunto de información más grande y la especificación más general, el VAR de panel bayesiano funcionó mejor. Comparando modelos con conjuntos de información similares, un BVAR mejora un VAR estimado con OLS y los parámetros variables en el tiempo mejoran los pronósticos para modelos univariados pero no para los BVAR.

- Clark and McCracken (2010) utilizan un conjunto de datos en tiempo real y proyectan la inflación, la tasa de interés y la producción de EEUU utilizando una amplia gama de modelos VAR trivariados basados en diferentes enfoques para permitir el cambio estructural. Si bien la atención se centra en diferentes métodos para combinar proyecciones y qué tan bien se desempeñan frente a cambios estructurales, los autores reportan algunos resultados para modelos BVAR. De estos, para el caso de la inflación un BVAR funciona mejor y, aunque no es directamente comparable, considerablemente mejor que un TVP-BVAR.

El paquete BEAR desarrollado en Matlab

- El paquete *Bayesian Estimation, Analysis and Regression* (BEAR desarrollado por Dieppe et al. (2016)) implementa éstas y otras extensiones del BVAR
- El SV-BVAR descrito aquí se llama "SV-BVAR con inercia aleatoria" en BEAR. Dicha inercia se refiere a la persistencia de la volatilidad estocástica (γ_i), que se trata como un coeficiente aleatorio para estimar
- En el paquete BEAR, el modelo SV estándar es equivalente al modelo descrito aquí, pero con el supuesto simplificador de que $\gamma_i = \gamma$ para todo $i \in \{1, ..., k\}$, donde γ está calibrado en lugar de estimado
- En el paquete BEAR, el SV-BVAR de gran escala (LS-SV-BVAR) corresponde al mismo modelo que el aquí descrito. Sin embargo, además de tener $\gamma_i = \gamma$ para todo $i \in \{1, ..., k\}$ (con γ calibrado), también se supone que $\lambda_i = \lambda$ para todo $i \in \{1, ..., k\}$. Estos supuestos reducen en gran medida la dimensionalidad del modelo.

Desempeño

- Clark (2011) estudia los proyecciones puntuales y por intervalos utilizando datos en tiempo real sobre el crecimiento de la producción, desempleo, inflación y de la tasa FED. Se utiliza un enfoque de la colatilidad cambiante motivado por la gran moderación. La calidad de la proyección del SV-BVAR es significativamente mejor en comparación de la proyección BVAR estándar
- D'Agostino et al. (2013) proyecta el desempleo en EEUU, la inflación y una tasa de interés a corto plazo. El objetivo del ejercicio de proyección es investigar qué tan importante es tener en cuenta los parámetros cambiantes en el tiempo y la volatilidad estocástica. Los autores encuentran que un TVP-BVAR supera a SV-BVAR, el que a su vez supera al BVAR estándar.

Otras extensiones

- Panel BVAR: Ver apartado 4 de la guía técnica BEAR y ver Liu et al. (2020*a*) para una actualización reciente (con códigos de replicación aquí)
- BVAR de frecuencia mixta: Schorfheide and Song (2015) (códigos aquí) y Brave et al. (2019) (códigos aquí)
- BVAR en tiempos de Covid-19: Primiceri and Tambalotti (2020), Lenza and Primiceri (2020) y Liu et al. (2020*b*). El último documento muestra actualizaciones semanales de su pronóstico de varios países aquí (en ese vínculo también se encuentran los códigos de replicación). Los códigos que replican los dos primeros documentos están aquí.

Bibliografía

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018), 'Narrative sign restrictions for svars', *American Economic Review* **108**(10), 2802–29.

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018), 'Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications', *Econometrica* **86**(2), 685–720.

Bańbura, M., Giannone, D. and Reichlin, L. (2010), 'Large bayesian vector auto regressions', *Journal of applied Econometrics* **25**(1), 71–92.

Brave, S. A., Butters, R. A. and Justiniano, A. (2019), 'Forecasting economic activity with mixed frequency bvars', *International Journal of Forecasting* **35**(4), 1692–1707.

Canova, F. (2007), 'G-7 inflation forecasts: Random walk, phillips curve or what else?', *Macroeconomic Dynamics* **11**(1), 1–30.

Clark, T. E. (2011), 'Real-time density forecasts from bayesian vector autoregressions with stochastic volatility', *Journal of Business & Economic Statistics* **29**(3), 327–341.

Clark, T. E. and McCracken, M. W. (2010), 'Averaging forecasts from vars with uncertain instabilities', *Journal of Applied Econometrics* **25**(1), 5–29.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: monetary policies and outcomes in the post wwii us', *Review of Economic dynamics* **8**(2), 262–302.

D'Agostino, A., Gambetti, L. and Giannone, D. (2013), 'Macroeconomic forecasting and structural change', *Journal of applied econometrics* **28**(1), 82–101.

Dieppe, A., Legrand, R. and Van Roye, B. (2016), 'The bear toolbox'.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Karlsson, S. (2013), Forecasting with bayesian vector autoregression, *in* 'Handbook of economic forecasting', Vol. 2, Elsevier, pp. 791–897.

Lenza, M. and Primiceri, G. E. (2020), How to estimate a var after march 2020, Technical report, National Bureau of Economic Research.

Litterman, R. B. (1986), 'Forecasting with bayesian vector autoregressions—five years of experience', *Journal of Business & Economic Statistics* **4**(1), 25–38.

Liu, L., Moon, H. R. and Schorfheide, F. (2020*a*), 'Forecasting with dynamic panel data models', *Econometrica* **88**(1), 171–201.

Liu, L., Moon, H. R. and Schorfheide, F. (2020*b*), 'Panel forecasts of country-level covid-19 infectionsliu'.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Primiceri, G. E. and Tambalotti, A. (2020), 'Macroeconomic forecasting in the time of covid-19', *Manuscript, Northwestern University* .

Schorfheide, F. and Song, D. (2015), 'Real-time forecasting with a mixed-frequency var', *Journal of Business & Economic Statistics* **33**(3), 366–380.

Sims, C. A. (1993), A nine-variable probabilistic macroeconomic forecasting model, *in* 'Business cycles, indicators, and forecasting', University of Chicago press, pp. 179–212.