Structural VAR

Spring 2022

Recommended Readings

Canova (2011): Ch4
Kilian and Lütkepohl (2017): Ch4, Ch8, Ch10, Ch12, Ch13

1 Reduced Form vs Structural Form

Reduced form

$$y_t = b + B_1 y_{t-1} + \cdots + B_p y_{t-p} + e_t,\ e_t \sim N(0, \Sigma_e) \tag{1}$$

or

$$y_t = B x_t + e_t,\ e_t \sim N(0, \Sigma_e) \tag{2}$$

where $B = [B_1 \dots B_p\ b]$ and $x_t' = [y_{t-1}' \dots y_{t-p}'\ 1]$.

Structural form

$$A(L) y_t = \varepsilon_t,\ \varepsilon_t \sim N(0, I_n) \tag{3}$$

or

$$A_0 y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t,\ \varepsilon_t \sim N(0, I_n) \tag{4}$$

or

$$A_0 y_t = A_+ x_t + \varepsilon_t,\ \varepsilon_t \sim N(0, I_n) \tag{5}$$

where $A_+ = [A_1 \dots A_p\ c]$ and $x_t' = [y_{t-1}' \dots y_{t-p}'\ 1]$.

Note: the linkage between structural form and reduced form

2 Estimation

Maximum likelihood estimation for unrestricted reduced-form VAR(p)
The density for y_t is

$$f(y_t | y_{t-1}, \dots, B, \Sigma_e) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_e|}} e^{-\frac{1}{2}(y_t - B x_t)' \Sigma_e^{-1} (y_t - B x_t)} \tag{6}$$

The density for $Y = \{y_1, \ldots, y_T\}$ is

$$f(Y|B, \Sigma_e) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^n |\Sigma_e|}} e^{-\frac{1}{2}(y_t - Bx_t)'\Sigma_e^{-1}(y_t - Bx_t)} \tag{7}$$

The log likelihood

$$L(B, \Sigma_e|Y) = -\frac{1}{2}T(n \ln(2\pi) - \ln|\Sigma_e^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(y_t - Bx_t)'\Sigma_e^{-1}(y_t - Bx_t) \tag{8}$$

The first-order conditions w.r.t. B

$$\hat{B}' = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t y_t'\right) \tag{9}$$

Note: the ML estimator is the same as OLS; seemingly unrelated regression (SUR); only for unrestricted VAR model.

The first-order conditions w.r.t. Σ_e

$$\hat{\Sigma}_e = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t \hat{e}_t' \tag{10}$$

where $\hat{e}_t = y_t - \hat{B}x_t$. Note: $\hat{\Sigma}_e$ is biased but consistent. ($\tilde{\Sigma}_e = \frac{1}{T-np-1}\sum_{t=1}^{T} \hat{e}_t \hat{e}_t'$)

The log likelihood under estimates

$$\begin{aligned} L(\hat{B}, \hat{\Sigma}_e|Y) &= -\frac{1}{2}T(n \ln(2\pi) - \ln|\hat{\Sigma}_e^{-1}|) - \frac{1}{2}\sum_{t=1}^{T}(y_t - \hat{B}x_t)'\hat{\Sigma}_e^{-1}(y_t - \hat{B}x_t) \\ &= -\frac{Tn}{2}\ln(2\pi) - \frac{T}{2}\ln|\hat{\Sigma}_e| - \frac{1}{2}\sum_{t=1}^{T} \hat{e}_t'\hat{\Sigma}_e^{-1}\hat{e}_t \\ &= -\frac{Tn}{2}\ln(2\pi) - \frac{T}{2}\ln|\hat{\Sigma}_e| - \frac{Tn}{2} \end{aligned} \tag{11}$$

where the third equality comes from

$$\begin{aligned} \sum_{t=1}^{T} \hat{e}_t'\hat{\Sigma}_e^{-1}\hat{e}_t = \text{trace}\left[\sum_{t=1}^{T} \hat{e}_t'\hat{\Sigma}_e^{-1}\hat{e}_t\right] = \text{trace}\left[\sum_{t=1}^{T} \hat{\Sigma}_e^{-1}\hat{e}_t\hat{e}_t'\right] \\ = \text{trace}\left[\hat{\Sigma}_e^{-1}(T\hat{\Sigma}_e)\right] = \text{trace}(T \cdot I_n) = Tn \end{aligned} \tag{12}$$

Determine number of lags

- Akaike information criterion (AIC): $\min_p AIC(p) = \ln|\Sigma_e| + \frac{2(pn^2+n)}{T}$
- Hannan-Quinn criterion (HQC): $\min_p HQC(p) = \ln|\Sigma_e| + \frac{2(pn^2+n)}{T}\ln(\ln T)$
- Schwarz criterion (SIC): $\min_p SIC(p) = \ln|\Sigma_e| + \frac{(pn^2+n)}{T}\ln T$

3 Identification

Definition: A parameter point (A_0, A_+) is globally identified if and only if there is no other parameter point that is observationally equivalent.

- (A_0, A_+) and $(\tilde{A}_0, \tilde{A}_+)$ are observationally equivalent
- $\Leftrightarrow$ they imply the same distribution of y_t for $1 \leq t \leq T$
- $\Leftrightarrow$ they have the same reduced-form presentation (B, Σ_e)
- $\Leftrightarrow$ there is an orthogonal matrix P such that $\tilde{A}_0 = PA_0$ and $\tilde{A}_+ = PA_+$

3.1 Short-run Restrictions

Short-run restrictions: no contemporary effects among variables

Recursively identified models (Cholesky decomposition)
Example 1: stylized model of monetary policy
Let $y_t = (\Delta gdp_t, \pi_t, i_t)'$, where Δgdp_t is the real GDP growth rate, π_t is the inflation rate, and i_t is the federal funds rate.

$$A_0 = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}$$

Example 2: Christiano, Eichenbaum and Evans (1999)
Let $y_t = (gdp_t, p_t, pcom_t, i_t, tr_t, nbr_t, m_t)'$, where gdp_t is real GDP, p_t is GDP deflator, $pcom_t$ is commodity price index, i_t is the federal funds rate, tr_t is the total reserves, nbr_t is the nonborrowed reserves, and m_t is M1 or M2.

$$A_0 = \begin{bmatrix} * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 & 0 \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \end{bmatrix}$$

Only monetary policy shock is identified.

Non-recursively identified models
Example 3: Sims and Zha (2006)
Let $y_t = (pcom_t, m_t, i_t, gdp_t, p_t, u_t)'$, where $pcom_t$ is commodity price index, m_t is M2, i_t is

the federal funds rate, gdp_t is real GDP, p_t is CPI, and u_t is unemployment rate.

$$A_0 = \begin{bmatrix} * & * & * & * & * & * \\ 0 & * & * & 0 & 0 & 0 \\ 0 & * & * & * & * & 0 \\ 0 & 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & * & * & 0 \\ 0 & 0 & 0 & * & * & * \end{bmatrix}$$

3.2 Long-run Restrictions

Long-run restrictions: no long-run effects for shocks

Structural form

$$A(L)y_t = \varepsilon_t,\ \varepsilon_t \sim N(0, I_n) \tag{13}$$

Reduced form

$$B(L)y_t = e_t,\ e_t \sim N(0, \Sigma_e) \tag{14}$$

where $A(L) = A_0 B(L)$.

Structural MA representation

$$y_t = A(L)^{-1}\varepsilon_t = \Theta(L)\varepsilon_t \tag{15}$$

The long-run *cumulative* effects are $\Theta(1) = \sum_{i=0}^{\infty} \Theta_i = A(1)^{-1} = B(1)^{-1}A_0^{-1}$.

Example 4: Blanchard and Quah (1989)
Let $y_t = (\Delta gdp_t, u_t)'$, where Δgdp_t is real GDP growth, and u_t is unemployment rate. The long-run restriction is that aggregate demand shock has no long-run effect on the level of real GDP.

$$A(L)\begin{bmatrix} \Delta gdp_t \\ u_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t^{AS} \\ \varepsilon_t^{AD} \end{bmatrix} \tag{16}$$

From the long-run restriction,

$$\Theta(1) = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix} \tag{17}$$

Procedure

(1) estimate B and Σ_e;

(2) calculate $B(1)$, and $B(1)^{-1}\Sigma_e B(1)^{-1\prime}$;

(3) calculate $\Theta(1)$ using $\Theta(1)\Theta(1)' = [B(1)^{-1}A_0^{-1}][B(1)^{-1}A_0^{-1}]' = B(1)^{-1}\Sigma_e B(1)^{-1\prime}$ and long-run restrictions;

(4) calculate $A_0 = \Theta(1)^{-1}B(1)^{-1}$, and A_+.

3.3 Sign Restrictions

Sign restrictions: restrict the sign (and/or shape) of structural responses. Set identified.

Procedure:

- OLS estimation: B and Σ_e.
- Eigenvalue-eigenvector decomposition: $\Sigma_e = PVP' = \tilde{P}\tilde{P}'$.
  For any H with $HH' = I$, $\Sigma_e = \tilde{P}\tilde{P}' = \tilde{P}HH'\tilde{P}'$ (Note $A_0 = (\tilde{P}H)^{-1}$).
- Choose $H = H_{j,k}(\omega), \omega \in (0, 2\pi)$ (rotate columns j and k of $\tilde{P}$ by an angle ω)

$$H_{j,k}(\omega) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \cos(\omega) & \dots & -\sin(\omega) & 0 \\ \vdots & \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & 0 & \sin(\omega) & \dots & \cos(\omega) & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

- Algorithm
  (1) Draw ω^l from $(0, 2\pi)$. Draw j and k from $1, \dots, n$.
  (2) Use $H_{j,k}(\omega)$ to compute $A(L)$. Check whether, in response to ε_{it}, $i = 1, \dots, n$, sign restrictions are satisfied. If they are, keep the draw; if they are not, drop the draw.
  (3) Repeat (1) and (2) until L draws satisfying the restrictions are found.

3.4 Narrative Sign Restrictions

Antolín-Díaz and Rubio-Ramírez (2018): "Narrative sign restrictions constrain the structural shocks and/or the historical decomposition around key historical events, ensuring that they agree with the established narrative account of these episodes."

Algorithm

(1) Randomly draws for A_0;

(2) Check whether the sign of ε_t or historical decomposition in certain episode is consistent with the narrative. If they are, keep the draw; if they are not, drop the draw.

(3) Repeat (1) and (2) until enough draws satisfying the restrictions are found.

3.5 Identification via heteroskedasticity

Identification via heteroskedasticity: identify shocks through a Markov regime switching in heteroskedasticity (Brunnermeier et al., 2019; Lanne et al., 2010)

Example:
Regime 1

$$A(L)y_t = \varepsilon_t,\ \varepsilon_t \sim N(0, I_n)$$

Regime 2

$$A(L)y_t = \varepsilon_t,\ \varepsilon_t \sim N(0, \Lambda_n)$$

where $\Lambda_n = diag(\lambda_1, \dots, \lambda_n)$. Shocks are identified as long as $\lambda_i \neq \lambda_j, \forall i \neq j$.

4 VAR Results

4.1 Impulse Response

From the structural form

$$y_t = A(L)^{-1}\varepsilon_t = \Theta(L)\varepsilon_t \tag{18}$$

The responses of each element of $y_t = [y_{1t}, \dots, y_{nt}]'$ to a one-time impulse in $\varepsilon_t = [\varepsilon_{1t}, \dots, \varepsilon_{nt}]'$

$$\frac{\partial y_{t+i}}{\partial \varepsilon_t'} = \Theta_i \tag{19}$$

where Θ_i is a $n \times n$ matrix. The elements of this matrix for given i are denoted as

$$\theta_{kj,i} = \frac{\partial y_{k,t+i}}{\partial \varepsilon_{jt}} \tag{20}$$

The VAR(1) representation of the VAR(p) process

$$Y_t = \mathbf{B}Y_{t-1} + U_t \tag{21}$$

where

$$Y_t \equiv \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \mathbf{B} \equiv \begin{bmatrix} B_1 & B_2 & \dots & B_{p-1} & B_p \\ I_n & 0 & & 0 & 0 \\ 0 & I_n & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & & I_n & 0 \end{bmatrix}, \quad \text{and } U_t \equiv \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Then

$$Y_t = \mathbf{B}^\tau Y_{t-\tau} + \sum_{i=0}^{\tau-1} \mathbf{B}^i U_{t-i} \tag{22}$$

Left-multiplying this equation by $J = [I_n, 0_{n\times n(p-1)}]$ yields

$$y_t = J\mathbf{B}^\tau Y_{t-\tau} + \sum_{i=0}^{\tau-1} J\mathbf{B}^i J' e_{t-i} \tag{23}$$

Define $\Phi_i \equiv J\mathbf{B}^i J'$. From the MA representation,

$$y_t = \sum_{i=0}^{\infty} \Phi_i e_{t-i} = \sum_{t=0}^{\infty} \Phi_i A_0^{-1} \varepsilon_{t-i} = \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} \tag{24}$$

Thus, $\Theta_i = \Phi_i A_0^{-1}$.

4.2 Distribution of Impulse Responses

Calculate the standard errors of impulse response: assess the statistical significance of the dynamics induced by certain shocks

Standard bootstrap method

- Algorithm

  (1) Obtain $\hat{A}_0$, $\hat{A}_+$ and $\hat{\varepsilon}_t$ from OLS.

  (2) Obtain $\hat{\varepsilon}_t^l$ via bootstrap and construct $\hat{y}_t^l$ from $\hat{y}_t^l = \hat{A}_0^{-1}\hat{A}_+\hat{x}_t^l + \hat{A}_0^{-1}\hat{\varepsilon}_t^l$, $l = 1, \ldots, L$.

  (3) Estimate $\hat{A}_0^l$, $\hat{A}_+^l$ by using data constructed in (2). Compute impulse responses.

  (4) Report percentile of the distribution of impulse responses (i.e., 16-84% or 2.5-97.5%).

Bias-adjusted bootstrap method

- Algorithm

  (1) Given $\hat{A}_0$, $\hat{A}_+$ from OLS, obtain $\hat{\varepsilon}_t^l$ and construct $\hat{y}_t^l$ from $\hat{y}_t^l = \hat{A}_0^{-1}\hat{A}_+\hat{x}_t^l + \hat{A}_0^{-1}\hat{\varepsilon}_t^l$, $l = 1, \ldots, L$.

  (2) Estimate $\hat{A}_0^l$, $\hat{A}_+^l$.

  (3) Calculate the largest root of the system $(\hat{A}_0, \hat{A}_+)$. If it is larger than or equal to 1, set $\tilde{A}_0 = \hat{A}_0$, $\tilde{A}_+ = \hat{A}_+$. Otherwise, set $\tilde{A}_0 = \hat{A}_0 - \hat{A}_0^{bias}$, $\tilde{A}_+ = \hat{A}_+ - \hat{A}_+^{bias}$, where $\hat{A}_0^{bias} = (1/L)\sum_{l=1}^{L}[\hat{A}_0^l - \hat{A}_0]$, $\hat{A}_+^{bias} = (1/L)\sum_{l=1}^{L}[\hat{A}_+^l - \hat{A}_+]$.

  (4) Repeat standard bootstrap method L_1 times by using $\tilde{A}_0$, $\tilde{A}_+$ in place of $\hat{A}_0$, $\hat{A}_+$.

Note: roots of the VAR can be calculated with $|A_0 - A_1 z - \cdots - A_p z^p| = 0$ or $|I_n - B_1 z - \cdots - B_p z^p| = 0$.

4.3 Forecast Error Variance Decomposition

The h-step ahead forecast error is

$$y_{t+h} - y_{t+h|t} = \sum_{i=0}^{h-1} \Phi_i e_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i \varepsilon_{t+h-i} \tag{25}$$

The variance of y_{t+h} is

$$var(y_{t+h}) = E[(y_{t+h} - y_{t+h|t})(y_{t+h} - y_{t+h|t})'] = \sum_{i=0}^{h-1} \Theta_i \Theta_i' \tag{26}$$

The variance of y_{kt} at horizon h is

$$var(y_{k,t+h}) = \sum_{i=0}^{h-1} \sum_{j=1}^{n} \theta_{kj,i}^2 = \sum_{j=1}^{n} (\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2)$$

The fraction of the contribution of shock ε_j to the forecast error variance of variable y_k in horizon h is

$$R_{kj,h} = \frac{\theta_{kj,0}^2 + \cdots + \theta_{kj,h-1}^2}{\sum_{i=1}^{n} (\theta_{ki,0}^2 + \cdots + \theta_{ki,h-1}^2)} \tag{27}$$

4.4 Historical Decomposition

For any t,

$$y_t = \sum_{i=0}^{t-1} \Theta_i \varepsilon_{t-i} + \sum_{i=t}^{\infty} \Theta_i \varepsilon_{t-i} = \sum_{i=0}^{t-1} \Theta_i \varepsilon_{t-i} + y_0 \tag{28}$$

The cumulative effect of structural shock ε_j on variable y_k at time t is

$$\hat{y}_{kt}^{(j)} = \sum_{i=0}^{t-1} \hat{\theta}_{kj,i} \hat{\varepsilon}_{j,t-i} \tag{29}$$

and note

$$y_t - y_0 = \sum_{j=1}^{n} \hat{y}_{kt}^{(j)}$$

References

Antolín-Díaz, Juan and Juan F. Rubio-Ramírez, "Narrative Sign Restrictions for SVARs," *The American Economic Review*, 2018, *108* (10), 2802–2829.

Blanchard, Olivier J and Danny Quah, "The dynamic effects of aggregate demand and aggregate supply," *The American Economic Review*, 1989, *79* (4), 655–73.

Brunnermeier, Markus Konrad, Darius Palia, Karthik A Sastry, and Christopher A Sims, *Feedbacks: financial markets and economic activity*, Benjamin H. Griswold III, Class of 1933, Center for Economic Policy Studies ..., 2019.

Canova, Fabio, *Methods for applied macroeconomic research*, Princeton university press, 2011.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans, "Monetary policy shocks: What have we learned and to what end?," *Handbook of macroeconomics*, 1999, *1*, 65–148.

Kilian, Lutz and Helmut Lütkepohl, *Structural vector autoregressive analysis*, Cambridge University Press, 2017.

Lanne, Markku, Helmut Lütkepohl, and Katarzyna Maciejowska, "Structural vector autoregressions with Markov switching," *Journal of Economic Dynamics and Control*, 2010, *34* (2), 121–131.

Sims, Christopher A and Tao Zha, "Were there regime switches in US monetary policy?," *American Economic Review*, 2006, *96* (1), 54–81.

Bayesian VAR

Spring 2022

Recommended Readings

Canova (2011): Ch9, Ch10
Kilian and Lütkepohl (2017): Ch5
Sims and Zha (1998)
Waggoner and Zha (2003)
Robertson and Tallman (1999)

1 Brief Introduction to Bayesian Methods

Bayesian approach

- Data $Y = (y_1', \ldots, y_T')'$: given
- Parameter θ: unknown
- Inference about θ conditional on data

Formula

- Prior: $g(\theta)$
- Likelihood function: $L(\theta|Y) = f(Y|\theta)$
- Bayes' theorem

$$g(\theta|Y) = \frac{f(Y|\theta)g(\theta)}{f(Y)}$$

- Posterior kernel

$$g(\theta|Y) \propto f(Y|\theta)g(\theta) = L(\theta|Y)g(\theta)$$

Point estimators

- Some function $h(\theta)$
- Loss function $\mathcal{L}(h^+, h(\theta))$

- Estimate

$$\hat{h}^+ = \arg\min_{h^+} \int \mathcal{L}(h^+, h(\theta))g(\theta|Y)d\theta$$

- Example
  - $\mathcal{L}(h^+, h(\theta)) = (h(\theta) - h^+)^2$: mean
  - $\mathcal{L}(h^+, h(\theta)) = |h(\theta) - h^+|$: median
  - $\mathcal{L}(h^+, h(\theta)) = I_{\{\theta \neq h^+\}}$: mode

Credit sets

- Set Ω: a $(1-\gamma)100\%$ credible set for θ w.r.t. posterior $g(\theta|Y)$

$$P(\theta \in \Omega|Y) = \int_\Omega g(\theta|Y)d\theta = 1 - \gamma$$

- Not unique: choose a highest posterior density interval
- In practice, replace with equal-tail-probability sets

Model comparison

- The posterior odds ratio

$$\frac{g(M_1|Y)}{g(M_2|Y)} = \frac{g(M_1)}{g(M_2)} \times \frac{f(Y|M1)}{f(Y|M2)}$$

  where

$$f(Y|M_j) = \int f(Y|M_j, \theta)g(\theta)d\theta$$

- Marginal data density incorporates a penalty for estimated model parameters

Posterior simulators

- Direct sampling
- Importance sampling

  (1) For $i = 1$ to N, draw θ^i from $\pi(\theta)$ and compute the unnormalized importance weights

$$w^i = w(\theta^i) = \frac{g(\theta^i)}{\pi(\theta^i)}$$

  (2) Compute the normalized importance weights

$$W^i = \frac{w^i}{\frac{1}{N}\sum_{j=1}^N w^i}$$

(3) An approximation of $E_g[h(\theta)]$ is given by

$$\bar{h}_N = \sum_{j=1}^{N} W^i h(\theta^i)$$

- Markov Chain Monte Carlo methods (MCMC)
  - Gibbs sampler
    (1) Choose initial values $(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$.
    (2) Draw θ_1^l from $g(\theta_1|\theta_2^{l-1}, \ldots, \theta_k^{l-1}, y)$, θ_2^l from $g(\theta_2|\theta_1^l, \theta_3^{l-1}, \ldots, \theta_k^{l-1}, y)$, $\ldots$, θ_k^l from $g(\theta_k|\theta_1^l, \ldots, \theta_{k-1}^l, y)$.
    (3) Repeat step (2) N times.
  - Metropolis-Hastings Algorithm:
    For $i = 1$ to N,
    (1) Draw ϑ from a density $q(\vartheta|\theta^{i-1})$.
    (2) Set $\theta^i = \vartheta$ with probability

$$\alpha(\vartheta|\theta^{i-1}) = \min\left\{1, \frac{f(Y|\vartheta)g(\vartheta)/q(\vartheta|\theta^{i-1})}{f(Y|\theta^{i-1})g(\theta^{i-1})/q(\theta^{i-1}|\vartheta)}\right\}$$

    and $\theta^i = \theta^{i-1}$ otherwise.

2 The Minnesota Prior

Reduced form

$$y_t = b + B_1 y_{t-1} + \cdots + B_p y_{t-p} + e_t,\ e_t \sim N(0, \Sigma) \tag{1}$$

Rewrite

$$Y = XB + u \tag{2}$$

where $Y \equiv [y_1 \ldots y_T]'$, $x_t \equiv [y_{t-1}' \ldots y_{t-p}'\ 1]'$, $X \equiv [x_1 \ldots x_T]'$, $u \equiv [e_1 \ldots e_T]'$, and $B = [B_1 \ldots B_p\ b]'$.

2.1 General Bayesian Update Rule

Normal-Inverse-Wishart prior on (β, Σ) [1]

$$\begin{aligned} \Sigma &\sim IW(\Psi, d) \\ \beta|\Sigma &\sim N(\mathtt{b}, \Sigma \otimes \Omega) \end{aligned} \tag{3}$$

[1] The probability density function of the inverse Wishart distribution is

$$f(\Sigma; \Psi, d) = \frac{|\Psi|^{d/2}}{2^{dn/2}\Gamma_n(d/2)} |\Sigma|^{-(d+n+1)/2} e^{-\frac{1}{2} trace(\Psi\Sigma^{-1})}$$

where Σ and Ψ are $n \times n$ positive definite matrices, and $\Gamma_n(\cdot)$ is the multivariate gamma function.

where $\beta = vec(B)$, and $\mathtt{b} = vec(\mathcal{B})$.

Posterior distribution

$$\begin{aligned} \Sigma|Y &\sim IW(\Psi + \hat{u}'\hat{u} + (\hat{B} - \mathcal{B})'\Omega^{-1}(\hat{B} - \mathcal{B}), T + d) \\ \beta|\Sigma, Y &\sim N(\hat{\beta}, \Sigma \otimes (X'X + \Omega^{-1})^{-1}) \end{aligned} \tag{4}$$

where $\hat{B} \equiv (X'X + \Omega^{-1})^{-1}(X'Y + \Omega^{-1}\mathcal{B})$, $\hat{\beta} \equiv vec(\hat{B})$, and $\hat{u} = Y - X\hat{B}$.

2.2 The Minnesota Prior Specification (Reduced-Form BVAR)

Inverse-Wishart prior (Σ)

- $d = n + 2$: existence of mean ($\Psi/(d - n - 1)$)
- Ψ:
  - Diagonal: $diag([\psi_1 \dots \psi_n])$
  - $\sqrt{\psi_j}$: low-order univariate AR regression and std of residual; sample standard deviation of the initial conditions; a priori reasoning about the likely scale of variation
  - Only the order of magnitude matters

Conditional normal prior ($B|\Sigma$)

$$\begin{aligned} E\left[(B_s)_{ij}|\Sigma\right] &= \begin{cases} 1 & \text{if } i = j \text{ and } s = 1 \\ 0 & \text{otherwise} \end{cases} \\ cov\left((B_s)_{ij}, (B_r)_{hm}|\Sigma\right) &= \begin{cases} \frac{\lambda^2}{s^2}\frac{\Sigma_{ih}}{\psi_j/(d-n-1)} & \text{if } m = j \text{ and } r = s \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5}$$

- Random walk
- Lower variance (tighter prior) for more distant lags
- Coefficients associated with the same variable and lag in different equations are allowed to be correlated
- Hyperparameter λ determines the overall tightness of prior

Dummy observation
$\bar{y}_0$ is an $n \times 1$ vector containing the average of the first p observations for each variable.

- "sum-of-coefficients" prior

$$\begin{aligned} \underset{n\times n}{y^+} &= diag(\frac{\bar{y}_0}{\mu}) \\ \underset{n\times(np+1)}{x^+} &= \left[y^+ \dots y^+ \; 0\right] \end{aligned} \tag{6}$$

  - No-change forecast
  - $\mu \to \infty$: uninformative
  - $\mu \to 0$: a unit root in each equation, no integration

- "dummy-initial-observation" prior

$$\begin{aligned} \underset{1\times n}{y^{++}} &= \frac{\bar{y}_0'}{\delta} \\ \underset{1\times(np+1)}{x^{++}} &= \left[y^{++} \dots y^{++} \ \frac{1}{\delta} \right] \end{aligned} \tag{7}$$

  - All lagged y_t's are at some level $\bar{y}_0$, y_t tends to persist at that level
  - $\delta \to \infty$: uninformative
  - $\delta \to 0$: stationary, and initial state close to unconditional mean.

Hyperparameters selection

- $\lambda = 0.2$, $\mu = 1$, $\delta = 1$
- Hierarchical models: hyperprior

3 Structural BVAR

3.1 Mathematical Derivation

The structural form

$$A(L)y_t + c = \varepsilon_t \tag{8}$$

The structural form in matrix

$$\underset{T\times n}{Y} \underset{n\times n}{A_0} - \underset{T\times k}{X} \underset{k\times n}{A_+} = \underset{T\times n}{E} \tag{9}$$

where $k = np + 1$.
Let

$$Z = [Y \ \ -X], \quad \text{and } A = \begin{bmatrix} A_0 \\ A_+ \end{bmatrix}. \tag{10}$$

Denote $a \equiv vec(A)$, $a_0 \equiv vec(A_0)$, and $a_+ \equiv vec(A_+)$.

The (conditional) likelihood function

$$\begin{aligned} L(A|Y) &\propto |A_0|^T \exp\left[-\frac{1}{2} \sum_t (A(L)y_t + c)'(A(L)y_t + c) \right] \\ &\propto |A_0|^T \exp[-0.5 trace(ZA)'(ZA)] \\ &\propto |A_0|^T \exp[-0.5a'(I \otimes Z'Z)a] \end{aligned} \tag{11}$$

The prior for a

$$\pi(a) = \pi_0(a_0)\varphi(a_+ - \mu(a_0); H(a_0)) \tag{12}$$

where φ is the standard normal p.d.f.

The posterior density function

$$\begin{aligned} q(a) \propto\ & \pi_0(a_0)|A_0|^T |H(a_0)|^{-1/2} \\ & \times \exp\big[-0.5(a_0'(I \otimes Y'Y)a_0 - 2a_+'(I \otimes X'Y)a_0 + a_+'(I \otimes X'X)a_+ \\ & \qquad + (a_+ - \mu(a_0))' H(a_0)^{-1}(a_+ - \mu(a_0)))\big] \end{aligned} \tag{13}$$

We can derive

$$q(a_+|a_0) = \varphi(a_0^*; (I \otimes X'X + H(a_0)^{-1})^{-1}) \tag{14}$$

$$\begin{aligned} q(a_0) \propto\ & \pi_0(a_0)|A_0|^T |(I \otimes X'X)H(a_0) + I|^{-1/2} \\ & \times \exp\big[-0.5(a_0'(I \otimes Y'Y)a_0 + \mu(a_0)'H(a_0)^{-1}\mu(a_0) - a_0^{*\prime}(I \otimes X'X + H(a_0)^{-1})a_0^*)\big] \end{aligned} \tag{15}$$

where

$$a_0^{*\prime} = (I \otimes X'X + H(a_0)^{-1})^{-1}((I \otimes X'Y)a_0 + H(a_0)^{-1}\mu(a_0))$$

3.2 Prior Formulation

- Prior for A_0
  Non-zero coefficient in A_0 for variable $i \sim N(0, (\lambda_0/\hat{\sigma}_i)^2)$

- Prior for A_+
  Mean

$$E[A_+|A_0] = \begin{bmatrix} A_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  Std of coefficient on lag l of variable j in equation i

$$\begin{cases} \frac{\lambda_0\lambda_1}{\hat{\sigma}_j l^{\lambda_3}} & \text{if } i = j \\ \frac{\lambda_0\lambda_1\lambda_2}{\hat{\sigma}_j l^{\lambda_3}} & \text{if } i \neq j \end{cases}$$

  Constant term: $N(0, (\lambda_0\lambda_4)^2)$

- Dummy observation
  "sum-of-coefficients" prior
  "dummy-initial-observation" prior

Common values for hyperparameters

Overall tightness:	λ_0	$= 1$
Lag tightness (own):	λ_1	$= 0.2$
Lag tightness (other):	λ_2	$= 1$
Lag decay:	λ_3	$= 1$
Constant:	λ_4	$= 1$
“sum-of-coefficients”:	μ	$= 1$
“dummy-initial-observation”:	δ	$= 1$

4 BVAR Result Analyses

- Conduct MCMC (Gibbs and Metropolis-Hastings) simulation
- Use posterior mode/mean for impulse response, variance decomposition, and historical decomposition
- Use posterior draws for distribution of impulse response

References

Canova, Fabio, *Methods for applied macroeconomic research*, Princeton university press, 2011.

Kilian, Lutz and Helmut Lütkepohl, *Structural vector autoregressive analysis*, Cambridge University Press, 2017.

Robertson, John C and Ellis W Tallman, "Vector autoregressions: forecasting and reality," *Economic Review (Atlanta, Ga.)*, 1999, *84* (1), 4–5.

Sims, Christopher A and Tao Zha, "Bayesian methods for dynamic multivariate models," *International Economic Review*, 1998, pp. 949–968.

Waggoner, Daniel F and Tao Zha, "A Gibbs sampler for structural vector autoregressions," *Journal of Economic Dynamics and Control*, 2003, *28* (2), 349–366.