Structural VARs

Frank Schorfheide
University of Pennsylvania
Econ 722 – Part 1

February 7, 2019

Reduced-Form and Structural VARs

- So far, we considered reduced form VARs, say,

$$y_t = \Phi_1 y_{t-1} + u_t, \quad \mathbb{E}[u_t u_t'] = \Sigma \tag{1}$$

- Error terms u_t have the interpretation of one-step ahead forecast errors.
- If the eigenvalues of Φ_1 are inside the unit-circle then y_t has the following moving-average (MA) representation in terms of u_t:

$$y_t = (I - \Phi_1 L)^{-1} u_t = \sum_{j=0}^{\infty} \Phi_1^j u_{t-j} = \sum_{j=0}^{\infty} C_j u_{t-j} \tag{2}$$

- Dynamic macroeconomic theory suggest that the one-step ahead forecast errors are functions of some fundamental shocks, such as technology shocks, preference shocks, or monetary policy shocks.

Parameterization I

- Let ϵ_t a vector of such fundamental shocks and assume that $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathcal{I}$. Moreover, assume that

$$u_t = \Phi_\epsilon \epsilon_t. \tag{3}$$

- Then we can express the VAR in structural form as follows

$$\begin{aligned} y_t &= \Phi_1 y_{t-1} + \Phi_\epsilon \epsilon_t \\ \Phi_\epsilon^{-1} y_t &= \Phi_\epsilon^{-1} \Phi_1 y_{t-1} + \epsilon_t \end{aligned} \tag{4}$$

- The moving-average representation of y_t in terms of the structural shocks is given by

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Phi_\epsilon \epsilon_{t-j} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \epsilon_{t-j}. \tag{5}$$

Parameterization II

- Sims and Zha (IER, 1998) use the following parameterization

$$A_0 y_t = A_1 y_{t-1} + \epsilon_t, \tag{6}$$

- which can be rewritten as

$$y_t = A_0^{-1} A_1 y_{t-1} + A_0^{-1}\epsilon_t = \Phi_1 y_{t-1} + \Phi_\epsilon \epsilon_t. \tag{7}$$

- This will matter for prior elicitation and posterior sampling.

Objects of Interest: Impulse Responses

- Impulse responses are defined as

$$
\begin{aligned}
IRF(h) &= \Big\{ \mathbb{E}[y_{t+h}|\epsilon_{t,1}=1,\mathcal{F}_{t-1}] - \mathbb{E}[y_{t+h}|\mathcal{F}_{t-1}],\ldots, \\
&\quad \mathbb{E}[y_{t+h}|\epsilon_{t,n}=1,\mathcal{F}_{t-1}] - \mathbb{E}[y_{t+h}|\mathcal{F}_{t-1}] \Big\} \\
&= \frac{\partial y_{t+h}}{\partial \epsilon_t'} \\
&= C_h \Phi_\epsilon
\end{aligned} \tag{8}
$$

and correspond to the MA coefficient matrices in the moving average representation of y_t in terms of structural shocks.

Objects of Interest: Variance Decompositions

- The covariance matrix of y_t is given by

$$\Gamma_{yy,0} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \mathcal{I} \Phi'_\epsilon C'_j \tag{9}$$

Let $\mathcal{I}^i$ be matrix for which element i, i is equal to one and all other elements are equal to zero. Then we can define the contribution of the i'th structural shock to the variance of y_t as

$$\Gamma^{(i)}_{yy,0} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \mathcal{I}^{(i)} \Phi'_\epsilon C'_j \tag{10}$$

Thus the fraction of the variance of $y_{l,t}$ explained by shock i is

$$\frac{[\Gamma^{(i)}_{yy,0}]_{ll}}{[\Gamma_{yy,0}]_{ll}}. \tag{11}$$

Identification

- We begin with Parameterization I
- For (1) and (4) the matrix Φ_ϵ has to satisfy the restriction

$$\Phi_\epsilon \Phi_\epsilon' = \Sigma \tag{12}$$

Notice that the matrix Φ_ϵ has n^2 elements but Σ only $n(n+1)/2$.
- Take a "square root," e.g. Cholesky, decomposition of

$$\Sigma = \Sigma_{tr} \Sigma_{tr}', \tag{13}$$

Σ_{tr} is lower triangular. If Σ is non-singular the decomp. is unique.
- Let Ω be an orthogonal matrix, meaning that $\Omega\Omega' = \Omega'\Omega = \mathcal{I}$.
- Then

$$u_t = \Sigma_{tr} \Omega \epsilon_t, \tag{14}$$

where Σ_{tr} is identifiable and Ω is not, because:

$$\mathbb{E}[u_t u_t'] = A\Omega \mathbb{E}[\epsilon_t \epsilon_t'] \Omega' \Sigma_{tr}' = \Sigma_{tr} \Omega \Omega' \Sigma_{tr}' = \Sigma_{tr} \Sigma_{tr}' = \Sigma.$$

Consequences – Or, Bayesian Inference with Non-Identified Parameters

- The joint distribution of data and parameters is given by

$$p(Y, \Phi, \Sigma, \Omega) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma)p(\Omega|\Phi, \Sigma). \tag{15}$$

- Integrating the joint density with respect to Ω yields

$$p(Y, \Phi, \Sigma) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma). \tag{16}$$

  Thus, the calculation of the posterior distribution of the reduced form parameters is not affected by the presence of the non-identifiable matrix Ω.

- The conditional posterior density of Ω can be calculated as follows:

$$p(\Omega|Y, \Phi, \Sigma) = \frac{p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)}{\int p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)d\Omega} = p(\Omega|\Phi, \Sigma). \tag{17}$$

Algorithm

For $s = 1, \ldots, n_{sim}$:

1. Draw $(\Phi^{(s)}, \Sigma^{(s)})$ from the posterior $p(\Phi, \Sigma | Y)$.
2. Draw $\Omega^{(s)}$ from the conditional prior distribution $p(\Omega | \Phi^{(s)}, \Sigma^{(s)})$. $\square$

Orthogonal Matrices

- For $n = 2$ we obtain a nice representation of the space of orthogonal matrices:

$$\Omega(\varphi, \xi) = \begin{bmatrix} \cos\varphi & -\xi \sin\varphi \\ \sin\varphi & \xi \cos\varphi \end{bmatrix} \tag{18}$$

  where $\varphi \in (-\pi, \pi]$ and $\xi \in \{-1, 1\}$
- Normalization issues: notice that, for instance,

$$\Omega(\pi/2, \xi) = -\Omega(-\pi/2, \xi) \tag{19}$$

  which means that only the signs of the impulse responses change but not the shape.
- Identification schemes impose restrictions hard or soft restrictions on φ: can be represented by $p(\Omega|\Phi, \Sigma)$.
- Two surveys: Stock and Watson (2001, JEP) and Ramey (2016, Handbook of Macroeconomics).

Point Identification: Short-run Restrictions

- Reference: Sims (1980, Ecta).
- Assumption: the central bank does not react contemporaneously to technology shocks because data on aggregate output only become available with a one-quarter lag.
- Suppose that

$$y_t = \begin{bmatrix} \text{Fed Funds} \\ \text{Output} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} = \begin{bmatrix} \text{Tech. Shock} \\ \text{Mon. Pol. Shock} \end{bmatrix}.$$

- Formalization in our framework: (i) $\varphi = 0$ and $\xi = 1$; (ii) $\varphi = 0$ and $\xi = -1$, (iii) $\varphi = \pi$ and $\xi = 1$, (iv) $\varphi = \pi$ and $\xi = -1$, e.g. under (i) we have:

$$u_t = \begin{bmatrix} \Sigma^{tr}_{11} & 0 \\ \Sigma^{tr}_{21} & \Sigma^{tr}_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix}. \tag{20}$$

- Choosing between (i)-(iv) normalizes the direction of responses.

Point Identification: Long-run Restrictions

- Reference: Blanchard and Quah (1989, AER).
- Assumption: monetary policy shocks do not raise output in the long-run.
- Suppose that

$$y_t = \left[\begin{array}{c} \text{Inflation} \\ \text{Output Growth} \end{array} \right], \quad \epsilon_t = \left[\begin{array}{c} \epsilon_{R,t} \\ \epsilon_{z,t} \end{array} \right] = \left[\begin{array}{c} \text{Tech. Shock} \\ \text{Mon. Pol. Shock} \end{array} \right]$$

Moreover,

$$y_t = (\sum_{j=0}^{\infty} C_j L^j) u_t = C(L) u_t. \quad (21)$$

- Notice that ξ only affects the response to the second shock. Hence we set it to $\xi = 1$.

Point Identification: Long-run Restrictions

Let's examine the moving average representation of y_t in terms of the structural shocks

$$
\begin{aligned}
y_t &= \left[\begin{array}{cc} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{array}\right]\left[\begin{array}{cc} \Sigma_{11}^{tr} & 0 \\ \Sigma_{21}^{tr} & \Sigma_{22}^{tr} \end{array}\right]\left[\begin{array}{cc} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{array}\right]\left[\begin{array}{c} \epsilon_{R,t} \\ \epsilon_{z,t} \end{array}\right] \\
&= \left[\begin{array}{cc} \cdot & \cdot \\ \Sigma_{11}^{tr}\cos\varphi c_{21}(L) + (\Sigma_{21}^{tr}\cos\varphi + \Sigma_{22}^{tr}\sin\varphi)c_{22}(L) & \cdot \end{array}\right]\left[\begin{array}{c} \epsilon_{R,t} \\ \epsilon_{z,t} \end{array}\right] \\
&= \left[\begin{array}{cc} d_{11}(L) & d_{12}(L) \\ d_{21}(L) & d_{22}(L) \end{array}\right]\left[\begin{array}{c} \epsilon_{R,t} \\ \epsilon_{z,t} \end{array}\right]
\end{aligned}
$$

Suppose that in period $t = 0$ log output and log prices are equal to zero. Then the log-level of output and prices in period $t = T > 0$ is given by

$$
y_T^c = \sum_{t=1}^{T} y_t = \sum_{t=1}^{T}\sum_{j=0}^{\infty} D_j \epsilon_{t-j} \qquad (22)
$$

Point Identification: Long-run Restrictions

- Now consider the derivative

$$\frac{\partial y_T^c}{\partial \epsilon_1'} = \sum_{j=0}^{T-1} D_j \tag{23}$$

- Letting $T \longrightarrow \infty$ gives us the long-run response of the level of prices and output to the shock ϵ_1:

$$\frac{\partial y_\infty^c}{\partial \epsilon_1'} = \sum_{j=0}^{\infty} D_j = D(1) \tag{24}$$

- Here, we want to restrict the long-run effect of monetary policy shocks on output:

$$d_{21}(1) = 0 \tag{25}$$

Point Identification: Long-run Restrictions

- This leads us to the equation

$$[\Sigma_{11}^{tr} c_{21}(1) + \Sigma_{21}^{tr} c_{22}(1)] \cos\varphi + \Sigma_{22}^{tr} c_{22}(1) \sin\varphi = 0. \tag{26}$$

- Notice that the equation has two solutions for $\varphi \in (-\pi, \pi]$. Under one solution a positive monetary policy shock is contractionary, under the other solution it is expansionary. The shape of the responses is, of course, the same.

Remarks

- Short-run restrictions: rely on often dubious assumptions about decision or informational lags. Many of these assumptions are more plausible in settings with "high" frequency observations, e.g. monthly, weekly, daily.
- Long-run restrictions suffer from the problem that the long-run multiplier matrices for the structural shocks are very imprecisely estimated, see Faust and Leeper (1997, JBES).

Implementation of Bayesian Inference

- Consider a simple VAR of the form $y_t = \Phi_1 y_{t-1} + u_t$, $u_t = \Sigma_{tr}\Omega(\varphi)\epsilon_t$, $\Phi = \Phi_1'$. For $s = 1, \dots, n_{sim}$:

  1. Generate a draw from the posterior distribution of (Φ, Σ), e.g., using sampling techniques for the $\mathcal{IW} - \mathcal{N}$ distribution. Let $\Sigma_{tr} = chol(\Sigma)$.
  2. Compute $\varphi = f(\Phi, \Sigma_{tr})$.
  3. Once $(\Phi, \Sigma_{tr}, \varphi)$ is determined, compute impulse responses and variance decompositions.

Implementation of Bayesian Inference

- This algorithm leaves you with n_{sim} draws from the posterior of the impulse responses and variance decompositions. You can now compute summary statistics for this posterior, such as means, medians, standard deviations, and (pointwise) credible sets.
- Sims and Zha (1999, Ecta) discuss how to construct alternative "error" bands for impulse response.

Partial Identification: Sign Restrictions

- References: Canova and De Nicolo (2002, JME), Uhlig (2005, JME).
- Assumption: upon impact, a monetary policy shock raises both prices and output.
- Again consider

$$y_t = \begin{bmatrix} \text{Inflation} \\ \text{Output Growth} \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \epsilon_{R,t} \\ \epsilon_{z,t} \end{bmatrix} = \begin{bmatrix} \text{Mon. Pol. Shock} \\ \text{Tech. Shock} \end{bmatrix}$$

- Our identification assumption implies:

$$\theta = \frac{\partial y_t}{\partial \epsilon_{R,t}} = \begin{bmatrix} \Sigma_{11}^{tr} \cos\varphi \\ \Sigma_{21}^{tr} \cos\varphi + \Sigma_{22}^{tr} \sin\varphi \end{bmatrix}. \tag{27}$$

- Let $\phi = [\Sigma_{11}^{tr}, \Sigma_{21}^{tr}, \Sigma_{22}^{tr}]'$.
- Define $q = [q_1, q_2]' = [\cos(\varphi), \sin(\varphi)]'$

Compact Representation

- Structural parameter:

$$\theta = q_1 \phi_1$$

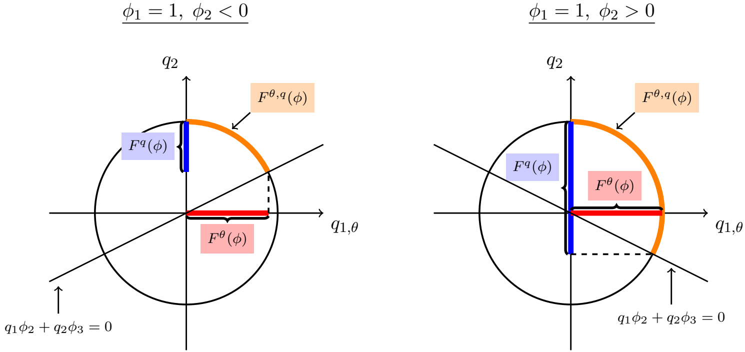
- Inequality restrictions:

$$q_1 \phi_1 \geq 0, \quad q_1 \phi_2 + q_2 \phi_3 \geq 0.$$

- Identified Sets:

$$F^{\theta,q}(\phi), \quad F^{q}(\phi), \quad F^{\theta}(\phi)$$

Identified Sets



Identified Sets

- The identified set $F^{\theta,q}(\phi)$ can be represented as the argmin of the following non-negative objective function:

$$G(\theta, q; \phi, W) = \min_{\mu \geq 0} \left\| \begin{bmatrix} q_1\phi_1 - \theta \\ q_1\phi_2 + q_2\phi_3 - \mu \end{bmatrix} \right\|_W^2$$

- The identified set $F^{\theta}(\phi)$ can be represented as the argmin of

$$G^{\theta}(\theta; \phi, W) = \min_{\|q\|=1} G(\theta, q; \phi, W).$$

- The identified set $F^{q}(\phi)$ can be represented as the argmin of the function

$$G^{q}(q; \phi, W) = \min_{\theta \geq 0} G(\theta, q; \phi, W).$$

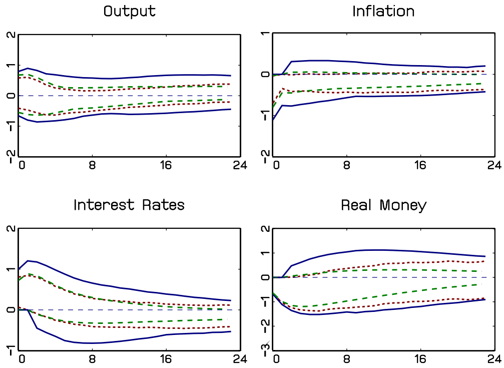
Computation of Identified Set

- Let $\mathcal{Q}$ be a grid for unit length vector q. You can generate this grid by:
  - Let $q = [\cos\varphi, \sin\varphi]'$ and use an equally-spaced grid for φ on $[-\pi, \pi]$.
  - Generate grid points randomly by letting $\tilde{q}^{(s)} \sim iidN(0, I)$ and $q^{(s)} = \tilde{q}^{(s)}/\|\tilde{q}^{(s)}\|$.
- Fix ϕ.
- For each $q^{(s)} \in \mathcal{Q}$ check whether $\left(\min_{\theta \geq 0}\; G(\theta, q^{(s)}; \phi, W)\right) = 0$. If it is, then $q^{(s)}$ belongs to the identified set.
- Compute $\theta^{(s)} = q_2^{(s)}\phi_1$ for each $q^{(s)} \in \hat{F}^q(\phi)$ to obtain $\hat{F}^{\theta,q}(\phi)$ and $\hat{F}^\theta(\phi)$.

Bayesian Inference for VAR(0)

- Popular choice: $p(\Sigma) \propto |\Sigma|^{-(n+1)/2}$, $\phi = vech(chol(\Sigma))$, and q is uniformly distributed on the arc $F^q(\phi)$.
- Factorize posterior: $p(\Sigma, q|Y) = p(\Sigma|Y)p(q|\Sigma, Y)$
- Posterior of Σ has Inverted Wishart distribution; $p(q|\Sigma, Y) = p(q|\Sigma)$.
- Posterior sampling: generate draws from $p(\Sigma|Y)$ and $p(q|\Sigma)$ and compute θ for each draw (q, Σ).

Moon and Schorfheide (2013): Impulse Responses Based on Pure Sign Restrictions



Notes: The figure depicts projection-based 90% frequentist confidence sets $CS_P^\theta(I)$ (blue, solid); 90% Bayesian credible intervals (red, short dashes); and estimated sets $F^\theta(\hat{\phi})$ (green, long dashes).

Alternative SVAR Representation

- Reference: Sims and Zha (1998, IER)
- Write VAR as

$$A_0 y_t = A_1 y_{t-1} + \dots A_p y_{t-p} + A_c + \epsilon_t \tag{28}$$

- In matrix form

$$YA_0' = XA + E. \tag{29}$$

- Likelihood function

$$\begin{aligned} & p(Y|A_0, A) \\ \propto \quad & |A_0'|^{-T/2} \exp\left\{ -\frac{1}{2} tr[(YA_0' - XA)'(YA_0' - XA)] \right\}. \end{aligned} \tag{30}$$

Alternative SVAR Representation

- Sims tends to place exclusion (zero) restrictions on A_0 instead of A_0^{-1}.
- Reparameterization affects prior: If a scalar A is $\mathcal{U}[-a, a]$ then A^{-1} has support on $(-\infty, -1/a)$ and $[1/a, +\infty)$ with density proportional to $|A^{-2}|$.
- The posterior of $A|A_0$ is normal, but the posterior of $A_0|A$ is non-normal and draws can be obtained with a Metropolis-Hastings step.
- Sims and Zha (1998, IER) consider prior distributions on $A|A_0$ that preserve the trace structure of the exponential term that appears in the likelihood. Thus, conditional on A_0 the posterior of A can be evaluated equation by equation.

Alternative SVAR Representation

- Waggoner and Zha (2003, JEDC):
  - Consider restrictions on A_0' and A of the form $Q_i a_i^0 = 0$ and $R_i a_i$, $1 \leq i \leq n$.
  - Restrictions on the structural coefficients are linear, which implies that the restrictions on the reduced form coefficients may be nonlinear.
  - The authors use priors of the form

    $$p(a_i^0, a_i | Q_i a_i^0 = 0, R_i a_i = 0)$$

    and develop a Gibbs sampler (that does not involve Metropolis-Hastings steps).
- Rubio-Ramirez, Waggoner, and Zha (2010, REStud):
  - provide conditions for global identification of VAR models;
  - sampling from columns of orthogonal matrix under "uniform" (Haar) prior.
- Baumeister and Hamilton (2015, Ecta):
  - emphasize prior elicitation on interpretable A_0 elements, e.g., demand and supply elasticities.

Other Strands of VAR Literature

(See Kilian and Lütkepohl's book for an introduction to these approaches)

- Identification via external instruments:
  - Recall that $u_t = \Phi_\epsilon \epsilon_t$.
  - Suppose we have a scalar instrument correlated with the first structural shock $\epsilon_{1,t}$ but uncorrelated with all other shocks.
  - Then

$$\begin{aligned} \mathbb{E}[u_t z_t] &= \Sigma_{tr}[q,\ \Omega_{(-1)}] \left[\begin{array}{c} \mathbb{E}[\epsilon_{1,t} z_t] \\ \mathbb{E}[\epsilon_{(-1),t} z_t] \end{array} \right] \\ &= \Sigma_{tr} q \mathbb{E}[\epsilon_{1,t} z_t] \end{aligned}$$

  - This provides a system of n equations with $n-1$ unknown elements in the vector q and the unknown scalar $\mathbb{E}[\epsilon_{1,t} z_t]$.
- Identification via heteroskedasticity:
  - Pre-break: $u_t = \Phi_\epsilon \epsilon_t$.
  - Post-break: $u_t = \Phi_\epsilon \Lambda \epsilon_t$, where Λ is diagonal.
  - We have $n^2 + n$ unknowns in Φ_ϵ and Λ; we have $2n(n+1)/2$ restrictions based on the pre- and post-break covariance matrices of the u_t's.