VAR Models

Jesús Gonzalo

Universidad Carlos III de Madrid

Time Series Econometrics

Some References

- Hamilton, J., Time Series Analysis. Princeton University Press, 1994.
- Canova, F.,Methods for Applied Macroeconomic Research. Princeton, 2007.
- Lutkepohl, H., New Inroduction to Multiple Time Series Analysis, second edition. Springer Verlag, 2007.
- Watson, M., VAR and Cointegration in Handbook of Econometrics, 1994.
- Stock, J., and M. Watson, Lecture Notes on What is New in Econometrics at NBER, 2008.

Goals of VAR models

- Let $\underset{K\times 1}{Y_t}$ be a vector of macro time series, and let ϵ_t^r denote an unanticipated (surprise, shock...) monetary policy intervention. We want to know the DYNAMIC CAUSAL EFFECT of ε_t^r on Y_t:

$$(*) \quad \frac{\partial Y_{t+h}}{\partial \epsilon_t^r}, h = 1, 2, 3, \dots \qquad \textbf{IRF} \tag{1}$$

  given all the other possible interventions constant.
- Exercise: Calculate the IRF for a univariable AR(1) model: $Y_t = \phi Y_{t-1} + \varepsilon_t$, $|\phi| \leq 1$
- The challenge is to estimate $\{\frac{\partial Y_{t+h}}{\partial \varepsilon_t^r}\}$ from observational data.
- (**) Granger causality: Does Y_{2t}, ..., Y_{kt} Granger cause Y_{1t}?
- (***) Do not forget prediction.

Wold Decomposition

- Everything starts from the Wold decomposition for $\underset{K\times 1}{Y_t}$ (weak stationary):

$$\underset{K\times 1}{Y_t} = \underset{(K\times K)}{C(L)} \underset{(K\times 1)}{e_t} \; ; \; C(0) = I; \Sigma_e \text{ unrestricted}$$

with $\{e_t\}$ a vector white noise $E(e_t) = 0$, $E(e_t e_{t-j}) = 0$, $j \neq 0$
- Remark: Review univariate Wold Decomposition
- Exercise: Following the same "reasoning" of the univariate Wold Decomposition, obtain $\{e_t\}$ and $C(L)$.
- $C(L)$ gives us the response of Y_t to unit impulses to each of the elements of e_t.
- We could calculate instead the responses of Y_t to new shocks that are linear combinations of the old shocks:

$$\underset{2\times 1}{\varepsilon} = \underset{(2\times 2)}{Q} \underset{(2\times 1)}{e_t} = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ 0.5e_{1t} + e_{2t} \end{bmatrix}$$

- The MA representation can be written as: $Y_t = C(L)Q^{-1}Qe_t = D(L)\varepsilon_t$

Wold Decomposition

- Question: Which linear combination of shocks should we look at?
- Answer: It seems that the most interesting are the linear combinations that produce orthogonal shocks: $\Sigma_\varepsilon =$ Diagonal
  Orthogonal shocks $\equiv$ Structural shocks
- We are going to pick a Q matrix s.t $E(\varepsilon_t \varepsilon_t') = I$. To do that choose a Q s.t.

$$Q^{-1}(Q^{-1})' = \Sigma_e$$

  Then

$$E(\varepsilon_t \varepsilon_t') = E(Q e_t e_t' Q') = Q \Sigma_e Q' = I$$

- One way to construct such a Q is via Choleski decomposition: "The Choleski decomposition of a Hermitian p.d matrix A is a decomposition of the form:

$$A = LL^*$$

  where L is a lower triangular matrix with real and positive diagonal entries and L^* is the conjugate transpose."

Wold Decomposition

- Unfortunately there are many different $Q's$ that act as "square root" matrices for Σ_e^{-1}. Given a Q we can form another $Q^* = RQ$ with R an orthogonal matrix:

$$RR' = I, \quad Q^* \Sigma_e Q^{*'} = RQ\Sigma_e Q' R' = RR' = I$$

- Example: Square roots of

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

:

$$\frac{1}{t}\begin{bmatrix} \mp s & \mp r \\ \mp r & \pm s \end{bmatrix}; \frac{1}{t}\begin{bmatrix} \pm s & \mp r \\ \mp r & \mp s \end{bmatrix}; \ldots; \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \text{ and } \begin{bmatrix} \pm 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where (r, s, t) is any set of positive integers such that $r^2 + s^2 = t^2$

- **So which Q should we choose? Problem**

Identification

- An identification problem:
  From sample e_t shocks, many different STRUCTURAL SHOCKS
- We solve this identification issue by imposing extra restrictions:
  - Short Run Restrictions
  - Long Run Restrictions
  - Sign Restrictions
  - Heterokedasticity
  - .....

Identification

- MA models are very nice representations to calculate IRF; BUT the models we estimate are VAR models.
- Assumption: The MA representation $Y_t = D(L)\varepsilon_t$ is invertible: roots of $D(z)$ are all greater than 1 in modulus.
- Exercise: Remember what invertibility is.
- With this assumption we can obtain a VAR(∞) for $\{Y_t\}$. Let's assume a finite VAR(p) is a good approximation.
- For $k = 2$ we will have:

$$Y_{1t} = B_{0,12}Y_{2t} + B_{1,12}Y_{2t-1} + \cdots + B_{p,12}Y_{2t-p} + B_{1,11}Y_{1t-1} + ... + B_{p,11}Y_{1t-p} + \varepsilon_{1t}$$

$$Y_{2t} = B_{0,21}Y_{1t} + B_{1,21}Y_{1t-1} + \cdots + B_{p,21}Y_{1t-p} + B_{1,22}Y_{2t-1} + ... + B_{p,22}Y_{2t-p} + \varepsilon_{2t}$$

- SVAR because ε_{1t}, ε_{2t} are orthogonal shocks

SVAR

- Exercise: Discuss the problems you encounter trying to estimate the above SVAR system by OLS

$$B(L)Y_t = \varepsilon_t \quad \underline{\text{Structural VAR}}$$

$$Y_t = B(L)^{-1}\varepsilon_t = D(L)\varepsilon_t$$

$$B(L) = B_0 - B_1 L - B_2 L^2 - \ldots - B_p L^P$$

$$E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_K^2 \end{bmatrix}$$

- This SVAR has a reduced form (Sims(1980)) which is identified: Reduced form VAR(p): $Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + e_t$

or $A(L)Y_t = e_t$

where $A(L) = 1 - A_1 L - \ldots - A_p L^P$

innovations: $e_t = Y_t - Proj(Y_t | Y_{t-1}, \ldots, Y_{t-p}), E(e_t e_t') = \Sigma_e$

Reduced form VAR

- $k = 2$, Reduced form VAR:

$$y_{1t} = A_{1,12} Y_{2t-1} + \ldots + A_{p,12} Y_{2t-p} + A_{1,11} Y_{1t-1} + \ldots + A_{p,11} Y_{1t-p} + e_{1t}$$

$$y_{2t} = A_{1,21} Y_{1t-1} + \ldots + A_{p,21} Y_{1t-p} + A_{1,22} Y_{2t-1} + \ldots + A_{p,22} Y_{2t-p} + e_{2t}$$

- From this VAR try to identify the parameters of the SVAR. What happens?
- Now is when we would wish the Σ_e not to be symmetric ha ha ha...

Summary of VAR and SVAR notation

Reduced form VAR	Structural VAR
$A(L)Y_t = e_t$	$B(L)Y_t = \varepsilon_t$
$Y_t = A(L)^{-1}e_t = C(L)e_t$	$Y_t = B(L)^{-1}\varepsilon_t = D(L)\varepsilon_t$
$A(L) = 1 - A_1L - A_2L^2 - \ldots - A_pL^p$	$B(L) = B_0 - B_1L - \ldots - B_pL^p$
$E(e_te_t') = \Sigma_e(\text{unrestricted})$	$E(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_K^2 \end{bmatrix}$

$Qe_t = \varepsilon_t,\ B(L) = QA(L),\ (B_0 = Q),\ D(L) = C(L)Q^{-1}$

- IRF: $\frac{\partial Y_{t+h}}{\partial \varepsilon_t} = D_h$
  Some remarks:
  1. $A(L)$ is finite order p
  2. $A(L)$, Σ_e, R are time invariant
  3. e_t spans the space of structural shocks ε_t, that is, $\varepsilon_t = Qe_t$
- Question: When 3 doesn't hold and how to solve the problem?

Identification of shocks

(*) Short run restrictions
(**) Long run restrictions
(***) Sign restrictions
(****) Identification via Heteroskedasticity

Before discussing these options, let's assume we have some extra knowledge: 1. We know one of the shocks, ε_t^r

$$Y_t = \begin{bmatrix} \underset{(K-1)\times 1}{X_t} \\ \underset{1\times 1}{r_t} \end{bmatrix}, e_t = \begin{bmatrix} e_t^x \\ e_t^r \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{bmatrix}$$

The IRF/MA form $Y_t = D(L)\varepsilon_t$

$$Y_t = \begin{bmatrix} D_{YX}(L) & D_{Yr}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{bmatrix} = D_{Yr}(L)\varepsilon_t^r + v_t$$

where $v_t = D_{YX}(L)\varepsilon_t^x$. Notice that $E(\varepsilon_t^r v_t) = 0$ then the IRF of Y_t w.r.t ε_t^r, $D_{Yr}(L)$ is identified by the population OLS regression Y_t onto ε_t^r.
2. Suppose we know Q, $Qe_t = \varepsilon_t$. Then we can proceed as in 1

3. Suppose you have an IV z_t (not in Y_t) s.t:
   i $E(z_t e_t^r) \neq 0$ (relevance)
   ii $E(z_t \varepsilon_t^x) = 0$ (exogeneity)

   Then you can estimate ε_t^r and act as in 1. To show this partion Y_t

$$Y_t = \begin{bmatrix} X_t \\ r_t^r \end{bmatrix}, e_t = \begin{bmatrix} e_t^x \\ e_t^r \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{bmatrix} \text{ and } Q = \begin{bmatrix} Q_{xx} & Q_{xr} \\ Q_{rx} & Q_{rr} \end{bmatrix}$$

   so $Qe_t = \varepsilon_t$ becomes:

$$Q_{xx} e_t^x = -Q_{xr} e_t^r + \varepsilon_t^x$$

$$Q_{rr} e_t^r = -Q_{rx} e_t^x + \varepsilon_t^r$$

   or

$$e_t^x = -Q_{xx}^{-1} Q_{xr} e_t^r + Q_{xx}^{-1} \varepsilon_t^x \quad (2)$$

$$e_t^r = -Q_{rr}^{-1} Q_{rx} e_t^x + Q_{rr}^{-1} \varepsilon_t^r \quad (3)$$

Identification of shocks

i Estimate $-Q_{xx}^{-1}Q_{xr}$ by IV estimation in (2)

ii Estimate $\widetilde{\varepsilon_t^x} = Q_{xx}^{-1}\varepsilon_t^x$ as $\widehat{\widetilde{\varepsilon_t^x}} = e_t^x + \widehat{Q_{xx}^{-1}Q_{xr}}e_t^r$

iii Use $\widehat{\widetilde{\varepsilon_t^x}}$ as instrument for e_t^x in (3) to estimate $-Q_{rr}^{-1}Q_{rx}$

iv Estimate $\widetilde{\varepsilon_t^r} = Q_{rr}^{-1}\varepsilon_t^r$ as $e_t^r + \widehat{Q_{rr}^{-1}Q_{rx}}e_t^x$

v IRF as in (2) by regressing Y_t on $\widetilde{\varepsilon_t^r}, \widetilde{\varepsilon_{t-1}^r}, \ldots$

- I don't know why this IV approach has not been used more??? Any answer or comments???

(*) **Short Run Restrictions:**

$\underset{K\times 1}{Y_t} = C(L) \underset{K\times 1}{e_t}$; $\underset{K\times 1}{Y_t} = D(L) \underset{K\times 1}{\varepsilon_t}$

$Y_t = Y_t$

$C(L)e_t = D(L)\varepsilon_t$

$C_0 e_t = D_0 \varepsilon_t$ or $Qe_t = \varepsilon_t$

$$\text{so } \boxed{\underbrace{Q}_{unknown} \underbrace{\Sigma_e}_{known} Q' = \underbrace{\Sigma_\varepsilon}_{Diagonal}} \tag{4}$$

or $\Sigma_e = D_0 \Sigma_\varepsilon D_0'$

There are $\frac{K(K+1)}{2}$ different equations in 4, so the order condition says that we can estimate at most $\frac{K(K+1)}{2}$ parameters. If we set $\Sigma_\varepsilon = I$ (a normalization), then we need:

$$K^2 - \frac{K(K+1)}{2} = \frac{K(K-1)}{2} \text{ restrictions on Q}$$

- Example: If K=2, then we need to impose a single restriction on Q, usually that Q is lower (Choleski) or upper triangular.
- Instead of restrictions on $\underline{Q}$ you can think on restrictions on $\underline{D_0}$ (this is why we call them short-run restrictions).
- We could also have PARTIAL IDENTIFICATION where only a row of Q is identified.

  Partion $\varepsilon_t = Qe_t$ and Y_t so that:

$$\begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{bmatrix} = \begin{bmatrix} Q_{xx} & Q_{xr} \\ Q_{rx} & Q_{rr} \end{bmatrix} \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{bmatrix}$$

  Suppose Q_{rx} and Q_{rr} are identified, then ε_t^r can be computed and $D_{yr}(L)$ can be computed by regressing Y_t on ε_t^r, ε_{t-1}^r,ε_{t-2}^r,...

- Some extra comment: The identification conditions discussed before (pure accounting) are "order conditions". We should not forget the rank conditions:
  $r(\Sigma_\varepsilon) = r(Q\Sigma_e Q')$ (see Hamilton).
  Intuitively this restriction rules out that any column of Q can be expressed as a linear combination of the others. While the rank condition is typically important in large-scale simultaneous equation systems, it is almost automatically satisfied in small scale VARs.

(**) **Long Run Restrictions:**

- Reduced form VAR: $A(L)Y_t = e_t$ ($Y_t = C(L)e_t$)
- Structural VAR: $B(L)Y_t = \varepsilon_t$ ($Y_t = D(L)\varepsilon_t$)
- LRV from VAR: $\Omega = A(1)^{-1}\Sigma_e(A(1)^{-1})' = C(1)\Sigma_e C(1)'$
- LRV from SVAR: $\Omega = B(1)^{-1}\Sigma_\varepsilon(B(1)^{-1})' = D(1)\Sigma_\varepsilon D(1)'$
- Notice that $D(1)$ is the long-run effect on Y_t of ε_t:

$$Y_t = D(L)\varepsilon_t = \underbrace{(D(1) + (1-L)D(\tilde{L}))\varepsilon_t}_{\text{Beveridge-Nelson decomposition}}$$

$$\sum_{t=1}^{T} Y_t = D(1)\sum_{t=1}^{T}\varepsilon_t + \tilde{\varepsilon_t} - \tilde{\varepsilon_0}$$

- System identification by long-run restrictions: The SVAR is identified if:

$$\boxed{\begin{array}{l} A(1)^{-1}Q^{-1}\Sigma_{\varepsilon}(Q^{-1})'A(1)^{-1} = \underset{K\times K}{\Omega} \\ \text{or} \\ D(1)\Sigma_{\varepsilon}D(1)' = \underset{K\times K}{\Omega} \end{array}} \tag{5}$$

  can be solved for the unknown elements of Q and Σ_{ε} (or $D(1)$ and Σ_{ε})
- Some accounting:
  There are $\frac{K(K+1)}{2}$ distinct equations in 5, so the order conditions say that you can estimate (at most) $\frac{K(K+1)}{2}$ parameters. If we set $\Sigma_{\varepsilon} = I$, it is clear that we need $K^2 - \frac{K(K+1)}{2} = \frac{K(K-1)}{2}$ restrictions on Q or $D(1)$.

- If $K = 2$, then $\frac{K(K-1)}{2} = 1$ which is delivered by imposing a single exclusion restriction on Q or $D(1)$ (for instance lower or upper triangular).
- If $\Sigma_\varepsilon = I$ then 5 can be rewritten:

$$\boxed{\Omega = D(1)D(1)'}$$

If the zero restrictions on $D(1)$ make $D(1)$ lower triangular, then $D(1)$ is the Choleski factorization for Ω.

Example: Blanchard and Quah (1984)
Goal to decompose GNP into permanent and transitory shocks. They postulate demand side shocks have only temporary effect on GNP while supply side shocks have permanent effect:

$$\begin{bmatrix} \Delta Y_t \\ \underbrace{u_t}_{\text{unemployment}} \end{bmatrix} = \begin{bmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_s \\ \varepsilon_d \end{bmatrix} \quad E(\varepsilon_t \varepsilon_t') = I$$

$$\bullet D_{12}(1) = 0$$

Estimate a VAR(p)

$$\begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \Delta Y_t \\ u_t \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

From it get $\Omega = A(1)^{-1}\Sigma_e(A(1)^{-1})'$
LRV from SVAR: $\Omega = D(1)\Sigma_\varepsilon D(1)'$

$$\begin{aligned}
\Omega_{11} &= D_{11}^2(0) + D_{12}^2(0) \\
\Omega_{22} &= D_{21}^2(0) + D_{21}^2(0) \\
\Omega_{12} &= D_{11}(0)D_{21}(0) + D_{12}(0)D_{22}(0)
\end{aligned}$$

and we only need to get $D_{11}(0)$, $D_{21}(0)$, $D_{22}(0)$

$$D(1)\varepsilon_t = C(1)e_t$$

$$\boxed{\varepsilon_t = D(1)^{-1}C(1)e_t}$$

(***) **Identification by Sign Restrictions:**

- Log-linearized version of DSGE models seldom deliver the whole set of zero restrictions needed to recover all economic shocks. Nevertheless, they contain a large number of sign restrictions usable for identification purposes. An example is: a monetary shock:
  - does not decrease FF rate for months 1,...,G.
  - does not increase inflation for months G,...,12

  These are restrictions on the signs of elements on $D(L)$.
- Signs restrictions can be used to set-identify $D(L)$. They are "weak" conditions and sometimes may be unable to distinguish shocks with somewhat similar features, i.e., labor supply and technology shocks. On the other side we have "strong" conditions that may fail to produce any meaningful economic shock.

"Weak" vs "strong"

Identification of shocks

By Sign Restrictions

- It is relatively complicated to impose sign restrictions on the coefficients of the VAR as this requieres maximum likelihood estimation of the full system under inequality constraints.
- However, it is relatively easy to do it ex-post on IRF. For instance, following Canova and De Nicolo (2002):

  1. Estimate $A(L)\Sigma_e$
  2. Get orthogonal shocks without imposing zero restrictions:

$$\Sigma_e = \underbrace{P}_{eigenvectors} \underbrace{V}_{eigenvalues} P' = \tilde{P}\tilde{P}' = \tilde{P}RR'\tilde{P}' \quad s.t \quad RR' = I$$

  3. For each of the orthogonalized shocks one can check whether the identifying restrictions are satisfied. If a shock is found the process terminates.
  4. If we find more than one we could impose stronger conditions or take the average of both shocks.

(****) **Heterokedasticity:**

Suppose

(a) The structural shock variance breaks at date "s": $\Sigma_{\epsilon,1}$ before, $\Sigma_{\epsilon,2}$ after
(b) Q does not change between variance regimes
(c) Normalize Q to have 1's on the diagonal, but no other restrictions

Then, unknowns are:

$$Q \to k^2 - k$$

$$\Sigma_{\epsilon,1} \to k$$

$$\Sigma_{\epsilon,2} \to k$$

$\Longrightarrow$ Summing up, we get: $k^2 + k$

Identification of shocks

From Heteroskedasticity, Rigobon (2003)

- First period: $Q\Sigma_{e,1}Q' = \Sigma_{\epsilon,1}$
  $\frac{k(k+1)}{2}$ equations and k^2 unknowns.
- Second period: $Q\Sigma_{e,2}Q' = \Sigma_{\epsilon,2}$
  $\frac{k(k+1)}{2}$ equations and k unknowns.

  Hence,

- Number of equations$= \frac{k(k+1)}{2} + \frac{k(k+1)}{2} = k(k+1)$
- Number of unknowns$= k^2 + k = k(k+1)$

Questions:

1. Which is the strong assumption in this set-up?
2. What if $\Sigma_{e,1}$ is proportional to $\Sigma_{e,2}$?

VAR(1): $y_t = A_1 y_{t-1} + e_t = (1 - A_1 L)^{-1} e_t = A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i e_{t-i}$

Result: If all eigenvalues of A_1 have modulus less than 1, then the sequence $A_1^i \; i = 0, 1, \ldots$ is absolutely summable and we call the VAR(1) STABLE. This condition is equivalent to:

$$|I_k - A_{1z}| \neq 0 \text{ for } |z| \leq 1$$

VAR(p): $y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + e_t$. Notice that any VAR(p) can be written as a VAR(1).

- Companion form

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{e}_t$$

- Matrix forms

$$\mathbf{y}_t = \underbrace{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}}_{kp \times 1} \quad \mathbf{A} = \underbrace{\begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{bmatrix}}_{kp \times kp} \quad \mathbf{e}_t = \underbrace{\begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{kp \times 1}$$

Thus, $\mathbf{y}_t$ is stable if $|\mathbf{I}_{kp} - \mathbf{A}_z| \neq 0$ for $|z| \leq 1$. Because,

$$|\mathbf{I}_{kp} - \mathbf{A}_z| = (I_k - A_{1z} - \cdots - A_p z^p)$$

Then, the stability condition can be written as:

$$|I_k - A_{1z} - \cdots - A_p z^p| \neq 0 \ \text{ for } \ |z| \leq 1$$

Example:

$$y_t = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} y_{t-2} + e_t$$

$$\left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ 0.25 & 0 \end{bmatrix} z^2 \right| = 1 - z + 0.21z^2 - 0.025z^3$$

Roots: $z_1 = 1.3$; $z_2 = 3.55 + 4.26i$ and $z_3 = 3.55 - 4.26i$. So, it is stable.

Exercise:

- Find the MA representation of y_t.
- Find the ARMA representation of y_t.

Estimation (Least-Squares)

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + e_t$$

In simultaneous equations format: $\mathbf{y} = \mathbf{B}\mathbf{z} + \mathbf{e}$

$$\mathbf{y} = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}}_{T \times K} \quad \mathbf{B} = \underbrace{\begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix}}_{K \times (Kp)} \quad z_t = \underbrace{\begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}}_{p \times K}$$

Or, $\text{vec}(\mathbf{y}) = \text{vec}(\mathbf{Bz}) + \text{vec}(\mathbf{e}) = (z' \otimes I_k)\text{vec}(B) + \text{vec}(e)$

Or, $y = (z' \otimes I_k)\beta + \text{vec}(e)$ with $\beta = \text{vec}(B)$.

$$\Longrightarrow \hat{\beta} = \left((z'z)^{-1} z \otimes I_k\right) y$$

$$\sqrt{T}(\hat{\beta} - \beta) = \sqrt{T}\text{vec}(\hat{B} - B) \xrightarrow{d} N(0, \Gamma^{-1} \otimes \Sigma_e)$$

with $\Gamma = p\lim \frac{z'z}{T}$.

It can also be proved that:

$$p\lim \hat{\Sigma}_e = p\lim \frac{ee'}{T} = \Sigma_e$$

Exercise: Show that if there are not restrictions on the VAR, OLS estimation of the parameters, equation by equation, is consistent and efficient.

From the asymptotic distribution of $\hat{\beta}$ it is straightforward to make inference:

$$H_0 : \underbrace{R}_{N \times k_p^2} \beta = c \quad \text{vs.} \quad H_1 : R\beta \neq c$$

$$\sqrt{T}(R\hat{\beta} - R\beta) \xrightarrow{d} N(0, R(\Gamma^{-1} \otimes \Sigma_e)R')$$

And hence,

$$T(R\hat{\beta} - c)'[R(\Gamma^{-1} \otimes \Sigma_e)R']^{-1}(R\hat{\beta} - c) \xrightarrow{d} \mathcal{X}^2(N)$$

WALD-STATISTIC:

$$(R\hat{\beta} - c)'[R((z'z)^{-1} \otimes \hat{\Sigma}_e)R']^{-1}(R\hat{\beta} - c) \xrightarrow{d} \mathcal{X}^2(N)$$

Let $y_t = (z_t' \ x_t')'$, $z_t(h|I_t)$ be the optimal (minimum MSE) h-step predictor of the process z_t given the information set I_t. The corresponding MSE will be denoted by $\Sigma_z(h|I_t)$. The process x_t is said to cause z_t in Granger sense if:

$$\Sigma_z(h|I_t) < \Sigma_z(h|I_t - \{x_t|s \leq t\})$$

with $I_t - \{x_t|s \leq t\}$ be all the information except the past and present of x_t.

Characterization of Granger Causality

- VAR: $A(L)y_t = e_t$; $A(0) = I$ and $\mathrm{E}(e_t e_t') = \Sigma_e$
- MA: $y_t = C(L)e_t$; $C(0) = I$
- And let's continue with $y_t = (z_t' \; x_t')'$. Thus,

$$z_t(1|\{y_s|s \leq t\}) = z_t(1|\{z_s|s \leq t\})$$

- iff $C_{12,i} = 0$ for $i = 1, 2, \ldots$ or, equivalently, $A_{12,i} = 0$ for $i = 1, 2, \ldots$
- In this situation, we say x_t does not Granger cause z_t.

Think on how to test for Granger causality.

Determining the VAR order

TESTING:

$$y_t = A_1 y_{t-1} + \cdots + A_M y_{t-M} + e_t$$

General to particular:

$$\begin{aligned}
&H_0^1 : A_M = 0 \;\text{ vs. }\; H_1^1 : A_M \neq 0 \\
&H_0^2 : A_{M-1} = 0 \;\text{ vs. }\; H_1^2 : A_{M-1} \neq 0 | A_M = 0 \\
&\vdots \\
&H_0^M : A_1 = 0 \;\text{ vs. }\; H_1^M : A_1 \neq 0 | A_M = \cdots = A_2 = 0
\end{aligned}$$

- In this scheme, each null hypothesis is tested conditionally on the previous ones being true. The procedure terminates and the VAR order is chosen accordingly, if one of the null hypothesis is rejected.
- A big problem is how to calculate the type I error of the whole procedure.

Determining the VAR order

Model Selection via information criteria:

An alternatie procedure abandoning testing is model selection via information criteria:

- $AIC(m) = \ln |\hat{\Sigma}_{e(m)}| + \frac{2}{T} mk^2$, where mk^2 is the number of freely estimated parameters.
- $SC(m) = \ln |\hat{\Sigma}_{e(m)}| + \frac{\ln T}{T} mk^2$
- $HQ(m) = \ln |\hat{\Sigma}_{e(m)}| + \frac{2 \ln(\ln T)}{T} mk^2$

Determining the VAR order

Result:

$$\underbrace{y_t}_{k\times 1} \sim \text{ VAR}(p); \quad M \geq p$$

And $\hat{p}$ is chosen so as to minimize a criterion:

$$IC(m) = \ln|\hat{\Sigma}_{u(m)}| + m\frac{C_T}{T} \text{ over } m = 0, 1, \ldots, M$$

The estimate $\hat{p}$ is consistent iff:

$$C_T \to \infty \text{ and } \frac{C_T}{T} \to 0 \text{ as } T \to \infty$$

And strongly consistent iff

$$\frac{C_T}{2\ln(\ln T)} > 1$$

Exercise: Which IC is consistent and which one is not?

Result:

$$\lim_{T \to \infty} \text{Prob}(\hat{p}(AIC) < p) = 0$$

and,

$$\lim_{T \to \infty} \text{Prob}(\hat{p}(AIC) > p) > 0$$

In a great paper, “Lag Length estimation in Large Dimensional Systems, JTSA”, Gonzalo and Pitarakis (2002) show that the latter probability goes to zero as $k \to \infty$.

Impulse Response Function

Let's consider a bivariate system $y_t = (y'_{1t} \ y'_{2t})'$. Thus, the MA representation $y_t = C(L)e_t$ is given by:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} + \begin{bmatrix} C_{11,1} & C_{12,1} \\ C_{21,1} & C_{22,1} \end{bmatrix} \begin{bmatrix} e_{1t-1} \\ e_{2t-1} \end{bmatrix} + \begin{bmatrix} C_{11,2} & C_{12,2} \\ C_{21,2} & C_{22,s} \end{bmatrix} \begin{bmatrix} e_{1t-2} \\ e_{2t-2} \end{bmatrix} + \cdots$$

Impulse Response Function: For $i = 1, 2$, it is the effect of a unit change in e_{it} in $y_{it+s} \approx$ dynamic multiplier. That is,

$$\frac{\partial y_{1t+s}}{\partial e_{1t}} = \psi_{11,s} \qquad \frac{\partial y_{1t+s}}{\partial e_{2t}} = \psi_{12,s}$$

$$\frac{\partial y_{2t+s}}{\partial e_{1t}} = \psi_{21,s} \qquad \frac{\partial y_{2t+s}}{\partial e_{2t}} = \psi_{22,s}$$

Notice that there is a serious problem on interpreting these partial derivatives because e_{1t} and e_{2_t} are correlated. This is one of the reasons to orthogonalize shocks.

Exercise: In the bivariate case, using OLS, get orthogonal shocks from $(e'_{1t}\ e'_{2t})'$.

IRF with orthogonal shocks

$\mathrm{E}(e_t e_t') = \Sigma_e$; pick a matrix Q such that: $Q\Sigma_e Q' = I$. Thus,

$$y_t = C(L)Q^{-1}Qe_t = D(L)\epsilon_t; \quad Qe_t = \epsilon_t$$

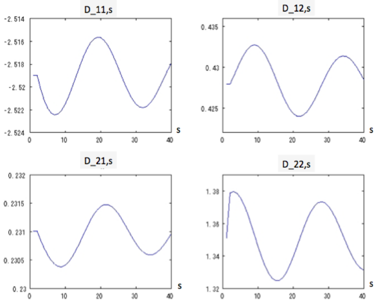
$$\begin{bmatrix} y_{1t+s} \\ y_{2t+s} \end{bmatrix} = \begin{bmatrix} D_{11,0} & D_{12,0} \\ D_{21,0} & D_{22,0} \end{bmatrix} \begin{bmatrix} \epsilon_{1t+s} \\ \epsilon_{2t+s} \end{bmatrix} + \cdots + \begin{bmatrix} D_{11,s} & D_{12,s} \\ D_{21,s} & D_{22,s} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \cdots +$$

$$\frac{\partial y_{1t+s}}{\partial \epsilon_{1t}} = D_{11,s} \qquad \frac{\partial y_{1t+s}}{\partial \epsilon_{2t}} = D_{12,s}$$

$$\frac{\partial y_{1t+s}}{\partial \epsilon_{1t}} = D_{11,s} \qquad \frac{\partial y_{1t+s}}{\partial \epsilon_{2t}} = D_{12,s}$$

IRF with orthogonal shocks

So, we have four IRF for the bivariate case:



- Plot $D_{11,s}$ vs. "s" (ϵ_{1t} shocks on y_{1t})
- Plot $D_{12,s}$ vs. "s" (ϵ_{2t} shocks on y_{1t})
- Plot $D_{21,s}$ vs. "s" (ϵ_{1t} shocks on y_{2t})
- Plot $D_{22,s}$ vs. "s" (ϵ_{2t} shocks on y_{2t})
- Long-run effects on each shock on y_{1t} and y_{2t} are:

$$(1) \sum_{s=0}^{\infty} D_{11,s} \qquad (2) \sum_{s=0}^{\infty} D_{12,s}$$

$$(3) \sum_{s=0}^{\infty} D_{21,s} \qquad (4) \sum_{s=0}^{\infty} D_{22,s}$$

Variance decompositions

Goal: To determine the proportion of the variability of y_{1t+s}, y_{2t+s} that is due to the shocks ϵ_{1t} and ϵ_{2t}. This allows us to determine the relative importance of the exogenous shocks to the evolution of y_{1t} and y_{2t}.

The Forecast error (FE) is given by:

$$FE(s) = y_{t+s} - \mathrm{E}[y_{t+s}|I_t] = D_0\epsilon_{t+s} + D_1\epsilon_{t+s-1} + \cdots + D_{s-1}\epsilon_{t+1}$$

Or,

$$\begin{bmatrix} y_{1t+s} \\ y_{2t+s} \end{bmatrix} - \begin{bmatrix} \mathrm{E}(y_{1t+s}|I_t) \\ \mathrm{E}(y_{2t+s}|I_t) \end{bmatrix} = \begin{bmatrix} D_{11,0} & D_{12,0} \\ D_{21,0} & D_{22,0} \end{bmatrix} \begin{bmatrix} \epsilon_{1t+s} \\ \epsilon_{2t+s} \end{bmatrix} +$$

$$+ \begin{bmatrix} D_{11,1} & D_{12,1} \\ D_{21,1} & D_{22,1} \end{bmatrix} \begin{bmatrix} \epsilon_{1t+s-1} \\ \epsilon_{2t+s-1} \end{bmatrix} + \cdots + \begin{bmatrix} D_{11,s-1} & D_{12,s-1} \\ D_{21,s-1} & D_{22,s-1} \end{bmatrix} \begin{bmatrix} \epsilon_{1t+1} \\ \epsilon_{2t+1} \end{bmatrix}$$

Variance decompositions

Focusing on the first equation:

$$\begin{aligned} y_{1t+s} - \mathrm{E}(y_{1t+s}|I_t) &= D_{11,0}\epsilon_{1t+s} + \cdots + D_{11,s-1}\epsilon_{1t+1} \\ &+ D_{12,0}\epsilon_{2t+s} + \cdots + D_{12,s-1}\epsilon_{2t+1} \end{aligned}$$

Thus,

$$\begin{aligned} MSE = \mathrm{E}\left[y_{1t+s} - \mathrm{E}(y_{1t+s}|I_t)\right]^2 &= \delta_1^2(s) \\ &= \delta_1^2(D_{11,0}^2 + D_{11,1}^2 + \cdots + D_{11,s-1}^2) \\ &+ \delta_2^2(D_{12,0}^2 + D_{12,1}^2 + \cdots + D_{12,s-1}^2) \end{aligned}$$

Variance decompositions

The proportion of $\delta_1^2(s)$ due to shocks in ϵ_{1t} is:

$$P_{11}(s) = \frac{\delta_1^2(D_{11,0}^2 + D_{11,1}^2 + \cdots + D_{11,s-1}^2)}{\delta_1^2(s)}$$

due to ϵ_{2t} is:

$$P_{12}(s) = \frac{\delta_2^2(D_{12,0}^2 + D_{12,1}^2 + \cdots + D_{12,s-1}^2)}{\delta_1^2(s)}$$

and similarly, for $P_{21}(s)$ and $P_{22}(s)$

Variance decompositions

The previous results are reported usually in the following way:

s	MSE	$P_{11}(s)$	$P_{12}(s)$
1	.0084	100%	0%
2	.0089	99%	1%
3	.0092	98.5%	1.5%
4	.0093	98.1%	1.9%

Confidence intervals for IRF

1. δ-method
2. Bootstrap methods
3. Monte-Carlo methods
4. Bayesian Methods

We will only discuss the first two.

Confidence intervals for IRF

δ-method

Remember the "delta" method:

If $\sqrt{T}(\hat{\theta}-\theta_0) \overset{d}{\sim} N(0,\Sigma_{\hat{\theta}})$ and if $g(\cdot)$ has continuous derivatives then

$$\sqrt{T}\left(g(\hat{\theta})-g(\theta_0)\right) \approx \sqrt{T}\left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0}(\hat{\theta}-\theta_0) \overset{d}{\sim} N\left(0,\left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0}\Sigma_{\hat{\theta}}\left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0}\right)$$

For SVAR IRFs:

$$\hat{\theta}=(\hat{A}(L),Q) \quad \text{and} \quad g(\hat{\theta})=\hat{D}(L)=\hat{A}(L)^{-1}\hat{Q}$$

Problems:

(i) $g(\cdot)$ is very non-linear so then even if $\hat{A}(L)$ were exactly normally distributed the IRF may not be. Let $\hat{\beta} \sim N(0.25, 1)$, which is the distribution of $\hat{\beta}^4$ or $\frac{1}{\hat{\beta}}$?

(ii) $\hat{A}(L)$ is not real approximated by a normal if roots are large.

Algorithm:

(i.) Obtain VAR estimates $\hat{A}(L)$, $\hat{e}_t$.

(ii.) Obtain $\hat{e}^l$ via bootstrap and construct $\hat{A}(L)y_t^l = \hat{e}_t^l$, for $l = 1, \ldots, L$.

(iii.) Estimate $\hat{A}^l(L)$ by using data constructed in the previous apart. Compute $\hat{D}^l(L)$.

(iv.) Report percentiles of the distribution of D_j.

Remarks:

- $\hat{e}_t$ should be white noise. Serious problems when it shows correlations and/or heteroskedasticity.
- Problem when there is a large persistence because then VAR coefficients usually are downward biased.

(+) & (-) of VAR models

(-)

- Time aggregation
- Large dimension
- Sometimes we have VAR(∞)
- Construction of confidence bands for the IRF

(+)

- They require very little to be used. This is just the opposite than DSGE models. Notice that people use VAR models to check DSGE results.