VAR Models

Some regerences:

- "Time Series Analysis by James Hamilton (1994)
- New Introduction to Multiple Time Series Analysis by Helmut Lutkepohl (2007)
- Methods for Applied Macroeconomic Research by Fabio Canova (2007)
- VARs and Cointegration in Handbook of Econometrics by M. Watson (1994)
- Lecture Notes on What is New in Econometrics at NBER by J. Stock and M. Watson (2008)

Goals of VAR models

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

(•) $$\boxed{\frac{\partial y_{t+h}}{\partial \varepsilon_t^r}, \quad h = 1, 2, 3} \quad \underline{\underline{\text{IRF}}}$$

given all the other possible intervention constant

Exercise: Calculate the IRF for a univariate AR(1) model $y_t = \phi y_{t-1} + \varepsilon_t$ $|\phi| \le 1$.

The challenge is to estimate $\left\{ \frac{\partial y_{t+h}}{\partial \varepsilon_t^r} \right\}$ from observational data

(••) Granger causality
Does $y_{2t}, \ldots y_{kt}$ 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)

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

with $\{e_t\}$ a vector white noise $E(e_t)=0$
$E(e_t e_{t-j}')=0 \quad j\neq 0$

<u>Remark</u>: Review univariate Wold Decomposition.

<u>Exercise</u>: Following the same "reasoning" of the univariate wold Decomposition obtain $\underline{\{e_t\}}$ and $\underline{C(L)}$.

$C(L)$ gives us the response of $\boxed{Y_t}$ to unit impulses to each of the elements of $\boxed{e_t}$

We could calculate instead the responses of Y_t to new shocks that are linear combination of the old shocks

$$\underset{2\times 1}{\varepsilon_t} = \underset{2\times 2}{Q}\,\underset{2\times 1}{e_t} = \begin{bmatrix} 1 & 0 \\ .5 & 1 \end{bmatrix}\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ .5e_{1t} + e_{2t} \end{bmatrix}$$

The MA representation can be written as

$$Y_t = C(L)Q^{-1}Q e_t = \underline{D(L)\,\varepsilon_t}$$

<u>Question</u>: Which linear combination of shocks we should look at?

<u>Answer</u>: It seems that the most interesting are the linear combinations that produce orthogonal shocks:

$$\underline{\text{orthogonal}}\ \underline{\text{shocks}} \overset{\Sigma_\varepsilon = \text{Diagonal}}{\underset{?}{=}} \underline{\text{structural}}\ \underline{\text{shocks}}$$

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

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

One way to construct such a Q is via choleski decomposition:
"The choleski decomposition of a Hermitian pd matrix A is a decomposition of the form

$$A = L L^*$$

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

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 and orthogonal matrix $RR' = I$

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

<u>Example</u>: Square roots of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$:

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

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

So Which $\underline{\underline{Q}}$ should we choose? <u>Problem</u>

An identification problem:

From <u>sample</u> <u>e_t</u> <u>shocks</u> $\longleftarrow$ many different STRUCTURAL SHOCKS

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

$$Y_{1t} = B_{0,12}Y_{2t} + B_{1,12}Y_{2t-1} + \dots + B_{p,12}Y_{2t-p} + B_{1,11}Y_{1t-1} + \dots + B_{p,11}Y_{1t-p} + \varepsilon_{1t}$$
$$Y_{2t} = B_{0,21}Y_{1t} + B_{1,21}Y_{1t-1} + \dots + B_{p,21}Y_{1t-p} + B_{1,22}Y_{2t-1} + \dots + B_{p,22}Y_{2t-p} + \varepsilon_{2t}$$

SVAR because $\varepsilon_{1t}, \varepsilon_{2t}$ are orthogonal shocks

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

In matrix format

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

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

$$B(L) = B_0 - B_1L - B_2L^2 - \dots B_pL^p$$

$$E(\varepsilon_t\varepsilon_t') = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & \cdot \\ 0 & \cdots & \sigma_k^2 \end{pmatrix}$$

This SVAR has a reduced form (Sims (1980)) which is identified:

Reduced form VAR(p): $\boxed{y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + e_t}$

or

$$A(L) y_t = e_t$$

where $A(L) = I - A_1 L - \dots A_p L^p$

innovations

$$e_t = y_t - \text{Proj}(y_t \mid y_{t-1}, \dots y_{t-p})$$

$$E(e_t e_t') = \Sigma_e$$

$\underline{\underline{K=2}}$ Reduced Form VAR:

$$y_{1t} = A_{1,12} y_{2t-1} + \dots + A_{p,12} y_{2t-p} + A_{1,11} y_{1t-1} + \dots + A_{p,11} y_{1t-p} + e_{1t}$$
$$y_{2t} = A_{1,21} y_{1t-1} + \dots + A_{p,21} y_{1t-p} + A_{1,22} y_{2t-1} + \dots + A_{p,22} y_{2t-p} + e_{2t}$$

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

$A(L)\,y_t = e_t$

$y_t = A(L)^{-1} e_t = C(L) e_t$

$A(L) = I - A_1 L - A_2 L^2 - \dots A_p L^p$

$E(e_t e_t') = \Sigma_e$ (unrestricted)

Structural VAR

$B(L)\,y_t = \varepsilon_t$

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

$B(L) = B_0 - B_1 L - \dots - B_p L^p$

$E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix}$

$$\boxed{\begin{gathered} Q e_t = \varepsilon_t \\ B(L) = Q\,A(L) \quad (B_0 = Q) \\ D(L) = C(L)\,Q^{-1} \end{gathered}}$$

IRF: $\dfrac{\partial y_{t+h}}{\partial \varepsilon_t'} = \underline{D_h}$

Some remarks:

(i) $A(L)$ is finite order p

(ii) $A(L)$, Σ_e, R are time invariant

(iii) e_t spans the space of structural shocks ε_t, that is, $\underline{\varepsilon_t = Q e_t}$

Question: When (iii) 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, $\boxed{\varepsilon_t^r}$

$$y_t = \begin{pmatrix} \underset{(k-1)\times 1}{X_t} \\ \underset{1\times 1}{r_t} \end{pmatrix}, \quad e_t = \begin{pmatrix} e_t^x \\ e_t^r \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{pmatrix}$$

The IRF/MA form is $y_t = D(L)\,\varepsilon_t$

$$y_t = \begin{pmatrix} D_{yx}(L) & D_{yr}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_t^x \\ \varepsilon_t^r \end{pmatrix} = D_{yr}(L)\,\varepsilon_t^r + v_t$$

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 $\underline{Q}$

$$Q e_t = \varepsilon_t$$

Then we can proceed as in (1)

(3) Suppose you have an IV $\underline{\underline{Z_t}}$ (not in Y_t) s.t.

$$\text{(i) } E(Z_t e_t^r) \neq 0 \quad \text{(relevance)}$$
$$\text{(ii) } E(Z_t \varepsilon_t^x) = 0 \quad \text{(exogeneity)}$$

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

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

so

$Q e_t = \varepsilon_t$ 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

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

$$e_t^r = -Q_{rr}^{-1} Q_{rx} e_t^x + Q_{rr}^{-1} \varepsilon_t^r \qquad (2)$$

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

(ii) Estimate $\tilde{\varepsilon}_t^x = Q_{xx}^{-1} \varepsilon_t^x$ as $\hat{\tilde{\varepsilon}}_t^x = e_t^x - \widehat{Q_{xx}^{-1} Q_{xr}} e_t^r$

(iii) Use $\hat{\tilde{\varepsilon}}_t^x$ as instrument for e_t^x in (2) to estimate $-Q_{rr}^{-1} Q_{rx}$

(iv) Estimate $\tilde{\varepsilon}_t^r = Q_{rr}^{-1} \varepsilon_t^r$ as $e_t^r + \widehat{Q_{rr}^{-1} Q_{rx}} e_t^x$

(v) IRF as in (1) by regressing Y_t on $\tilde{\varepsilon}_t^r, \tilde{\varepsilon}_{t-1}^r, \ldots$

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

(•) Short run restrictions

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

$$Y_t = Y_t$$

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

$$C_0 e_t = D_0 \varepsilon_t$$

$$e_t = D_0 \varepsilon_t \qquad (D_0 = Q^{-1})$$

or $\boxed{Q e_t = \varepsilon_t}$

So

$$\underset{\text{Unknown}}{Q}\ \underset{\text{Known}}{\Sigma_e}\ Q' = \underset{\text{Diagonal}}{\Sigma_\varepsilon} \qquad (*)$$

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

There are $\frac{K(K+1)}{2}$ distinct equations in $(*)$, 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

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

Example: If $K=2$, then we need to impose a single restriction on Q, usually that Q is lower or upper triangular.

Choleski

Instead of restrictions on Q you can think on restrictions on 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.

Partition $\varepsilon_t = Q e_t$ and Y_t so that

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

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 $\varepsilon_t^r, \varepsilon_{t-1}^r, \varepsilon_{t-2}^r + \dots$

Some extra comment:

The identification conditions discussed before (pure accounting) are "order conditions". We should not forget the rank conditions

$$\boxed{r(\Sigma_\varepsilon) = r(Q \Sigma_e Q')}$$ (see Hamilton)

Intuitively, this restriction rules out that any column of Q can be expressed 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: $A(L)y_t = e_t \quad (y_t = C(L)e_t)$

Structural VAR: $B(L)y_t = \varepsilon_t \quad (y_t = D(L)\varepsilon_t)$

LRV from VAR: $\Omega = A(1)^{-1}\Sigma_e A(1)^{-1'} = \underline{C(1)\Sigma_e C(1)'}$

LRV from SVAR: $\Omega = B(1)^{-1}\Sigma_\varepsilon B(1)^{-1'} = \underline{D(1)\Sigma_\varepsilon D(1)'}$

Notice that $D(1)$ is the long-run effect on y_t of ε_t:

$$y_t = D(L)\varepsilon_t = \big(D(1) + (1-L)\tilde{D}(L)\big)\varepsilon_t$$

Beveridge-Nelson Decomposition

$$\sum_1^T y_t = D(1)\sum_1^T \varepsilon_t + \tilde{D}(L)\varepsilon_t$$

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

$$\boxed{\begin{aligned} &A(1)^{-1}Q^{-1}\Sigma_\varepsilon Q^{-1'}A(1)^{-1'} = \underset{K\times K}{\Omega} \\ \text{or} \quad & \\ &D(1)\Sigma_\varepsilon D(1)' = \underset{K\times K}{\Omega} \end{aligned}} \quad (*)$$

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 (*), so the order conditions says 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 (*) 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 Cholesky factorization of Ω.

Example: Blanchard and Quah (1989).
Goal to decompose GNP into permanent and transitory shocks. They postulate demand side shocks have only temporary effects on GNP while supply side shocks have permanent effects

$$\begin{pmatrix} \Delta y_t \\ u_t \end{pmatrix} = \begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{st} \\ \varepsilon_{dt} \end{pmatrix} \qquad E(\varepsilon_t \varepsilon_t') = I$$

unemployment

- $\boxed{D_{12}(1) = 0}$

Estimate VAR(p)

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

From it get $\Omega = A(1)^{-1} \Sigma_e A(1)^{-1'}$

LRV from SVAR:

$$\Omega = D(1) \Sigma_\varepsilon D(1)'$$

$$\Omega_{11} = D_{11}^2(0) + \cancel{D_{12}^2(0)}$$
$$\Omega_{22} = D_{21}^2(0) + D_{21}^2(0)$$
$$\Omega_{12} = D_{11}(0) D_{21}(0) + \cancel{D_{12}(0) D_{22}(0)}$$

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 policy shock

- does not decrease FF rate for months 1, .., 6
- does not increase inflation for months 6, .., 12

These are restrictions on the sign of elements of $D(L)$.

Sign 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"

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

- Estimate $A(L)$ Σ_e VAR model
- Get orthogonal shocks without imposing zero restrictions

$$\Sigma_e = \underbrace{P}_{\text{eigenvectors}} \overbrace{V}^{\text{eigenvalue}} P' = \tilde{P}\tilde{P}'$$

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

(••••) Identification from Heteroskedasticity
(Rigobon (2003))

Suppose:

(a) The structural shock variance breaks at date "s": $\Sigma_{\varepsilon,1}$ before $\Sigma_{\varepsilon,2}$ after

(b) Q doesn't change between variance regimes

(c) Normalize Q to have 1's on the diagonal, but no other restrictions

Then unknowns are:

$$
\begin{aligned}
Q &\rightarrow k^2 - k \\
\Sigma_{\varepsilon,1} &\rightarrow k \\
\Sigma_{\varepsilon,2} &\rightarrow k \\
\hline
&\underline{\underline{k^2 + k}}
\end{aligned}
$$

First period: $Q \, \Sigma_{e,1} Q' = \Sigma_{\varepsilon,1}$ $\begin{cases} \frac{k(k+1)}{2} \text{ equations} \\ k^2 \text{ unknowns} \end{cases}$

Second period: $Q \, \Sigma_{e,2} Q' = \Sigma_{\varepsilon,2}$ $\begin{cases} \frac{k(k+1)}{2} \text{ equations} \\ k \text{ unknowns} \end{cases}$

$$\# \text{ equations} = \frac{k(k+1)}{2} + \frac{k(k+1)}{2} = k(k+1)$$

$$\# \text{ unknowns} = k^2 + k = k(k+1)$$

Question: Which is the strong assumption in this set-up??

What if $\Sigma_{e,1}$ is proportional to $\Sigma_{e,2}$??

Some Asymptotic Results

Stability

VAR(1): $y_t = A_1 y_{t-1} + e_t$

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

Result: If all the 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

$$\boxed{|I_k - A_1 z| \neq 0 \text{ for } |z| \leq 1}$$

VAR(p): Any VAR(p) can be written as a VAR(1)

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

$$\mathbf{Y}_t = \mathbf{A} \mathbf{Y}_{t-1} + \mathbf{e}_t \quad \text{companion form}$$

companion matrix

$$\mathbf{Y}_t = \underset{kp \times 1}{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}} ; \quad \mathbf{A} = \underset{kp \times kp}{\begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & & I_k & 0 \end{bmatrix}} ; \quad \mathbf{e}_t = \underset{kp \times 1}{\begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}}$$

$\mathbf{Y}_t$ is stable if $|I_{kp} - \mathbf{A} z| \neq 0$ for $|z| \leq 1$

Because

$$|I_{kp} - \mathbf{A} z| = |I_k - A_1 z - \cdots - A_p z^p|$$

then the stability condition can be written as

$$\boxed{|I_k - A_1 z - \cdots - A_p z^p| \neq 0 \text{ for } |z| \leq 1}$$

Example:

$$\underset{2\times 1}{y_t} = \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} y_{t-2} + e_t$$

$$\left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} z^2 \right| = 1 - z + .21 z^2 - .025 z^3$$

Roots: $z_1 = 1.3$; $z_2 = 3.55 + 4.26i$ and $z_3 = 3.55 - 4.26i$

so stable.

Exercise

(1) Find the MA representation of y_t

(2) Find the ARMA representation of y_{1t}.

Estimation (Least-Squares)

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

In simultaneous equation format

$$Y = BZ + e$$

$$Y = \underset{K\times T}{\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}} ; \quad B = \underset{K\times(Kp)}{(A_1, \dots A_p)} ; \quad Z_t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}$$

or

$$\operatorname{vec}(Y) = \operatorname{vec}(BZ) + \operatorname{vec}(e)$$

$$= (Z' \otimes I_K) \operatorname{vec}(B) + \operatorname{vec}(e)$$

or

$$y = (Z' \otimes I_K)\beta + e$$

with $\beta = \operatorname{vec}(B)$

$$\boxed{\hat{\beta} = ((Z'Z)^{-1} Z \otimes I_K)\, y}$$

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

with $\Gamma = \operatorname{plim} \dfrac{Z'Z}{T}$.

It can also be proved that

$$\operatorname{plim} \hat\Sigma_e = \operatorname{plim} \frac{ee'}{T} = \Sigma_e$$

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 $\hat\beta$ is straightforward to make inference:

$\boxed{H_0: \underset{(N\times K^2 p)}{R}\beta = c}$ vs $\boxed{H_1: R\beta \neq c}$

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

and hence

$$T(R\hat\beta - c)'\left[R(\Gamma^{-1}\otimes\Sigma_u R'\right]^{-1}(R\hat\beta - c) \xrightarrow{d} \chi^2(N)$$

WALD - STATISTIC:

$$(R\hat\beta - c)'\left[R((ZZ')^{-1}\otimes\hat\Sigma_e)R'\right]^{-1}(R\hat\beta - c) \xrightarrow{d} \chi^2(N)$$

Granger Causality (Granger (1969))

Let $y_t = \begin{pmatrix} z_t \\ x_t \end{pmatrix}$, $z_t(h|I_t)$ 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

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

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

Characterization of Granger-Causality:

VAR: $A(L)y_t = e_t \qquad A(0) = I \qquad E(e_t e_t') = \Sigma_e$

MA: $y_t = C(L)e_t \qquad C(0) = I$

and let's continue with $y_t = \begin{pmatrix} z_t \\ x_t \end{pmatrix}$.

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

$$\Updownarrow$$

$C_{12,i} = 0$ for $i = 1, 2 \ldots$

or equivalently

$A_{12,i} = 0$ for $i = 1, 2 \ldots$

In this situation we say x_t doesn't Granger cause z_t.

Think on how to test for Granger causality

Determining the VAR order

TESTING:

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

General to Particular:

H_0^1: $A_M = 0$ vs H_1^1: $A_M \neq 0$

H_0^2: $A_{M-1} = 0$ vs H_1^2: $A_{M-1} \neq 0 \mid A_M = 0$

$\vdots$

H_0^M: $A_1 = 0$ vs H_1^M: $A_1 \neq 0 \mid A_M = \dots = A_2 = 0$

In this scheme, each null hypothesis is tested conditionally on the previous ones being true. The procedure terminate and the VAR order is chosen accordingly, if one of the null hypotheses is rejected.

A big problem is how to calculate the type I error of the whole procedure.

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

- $$AIC(m) = \ln|\hat{\Sigma}_e(m)| + \frac{2}{T}(\#\text{ freely estimated parameters})$$
  $$= \ln|\hat{\Sigma}_e(m)| + 2m\frac{K^2}{T}$$
- $$SC(m) = \ln|\hat{\Sigma}_e(m)| + \frac{\ln T}{T} mK^2$$
- $$HQ(m) = \ln|\hat{\Sigma}_e(m)| + \frac{2\ln\ln T}{T} mK^2$$

Result: $\underset{K\times 1}{y_t} \sim VAR(p)$

$$M \geq p$$

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

$$IC(m) = \ln|\hat{\Sigma}_u(m)| + m\frac{c_T}{T}$$

over $m = 0, 1, \ldots M$.

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

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

strongly consistent iff

$$\frac{c_T}{2\ln\ln T} > 1$$

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

Result

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

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

In a great paper Gonzalo-Pitarakis (2002) "Lag length Estimation in Large Dimensional Systems" (JTSA) shows that the latter probability goes to zero as $K \to \infty$.

Impulse Response Function

Let's consider a bivariate system $Y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$.

The MA representation $\boxed{Y_t = C(L) e_t}$

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

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

IRF

$$\frac{\partial y_{1t+s}}{\partial e_{1t}} = \psi_{11,s}$$ effect of a unit change in e_{1t} in y_{1t+s} ≃ dynamic multiplier

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

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

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

There is a serious problem on interpreting these partial derivatives because e_{1t} & e_{2t} are correlated. This is one of the reasons to orthogonalize shocks

Exercise: In the bivariate case, using OLS, get orthogonal shocks from $\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$.

IRF with orthogonal shocks

$E(e_t e_t') = \Sigma_e$; pick a matrix Q s.t

$$Q \Sigma_e Q' = I$$

$$Y_t = C(L) Q^{-1} Q e_t = \underline{D(L) \varepsilon_t} \; ; \quad Q e_t = \varepsilon_t$$

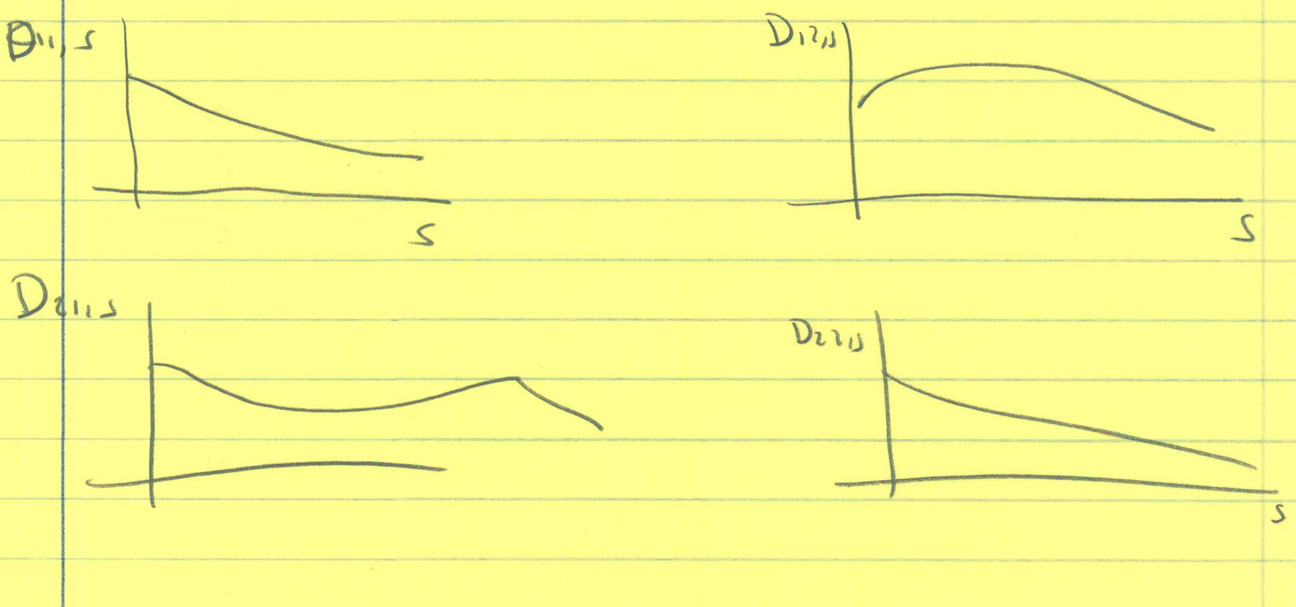
$$\begin{pmatrix} y_{1t+s} \\ y_{2t+s} \end{pmatrix} = \begin{pmatrix} D_{11,0} & D_{12,0} \\ D_{21,0} & D_{22,0} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t+s} \\ \varepsilon_{2t+s} \end{pmatrix} + \cdots + \begin{pmatrix} C_{11,s} & C_{12,s} \\ C_{21,s} & C_{21,s} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t}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

$$\frac{\partial y_{1t+s}}{\partial \varepsilon_{1t}} = D_{11,s} \; ; \quad \frac{\partial y_{1t+s}}{\partial \varepsilon_{2t}} = C_{12,s}$$

$$\frac{\partial y_{2t+s}}{\partial \varepsilon_{1t}} = D_{21,s} \; ; \quad \frac{\partial y_{2t+s}}{\partial \varepsilon_{2t}} = C_{22,s}$$

so we have four IRF for the bivariate case:

(1) Plot $D_{11,s}$ vs "s" (ε_{1t} shocks on y_{1t})
(2) Plot $D_{12,s}$ vs "s" (ε_{2t} " " ")
(3) Plot $D_{21,s}$ vs "s" (ε_{1t} shocks on y_{2t})
(4) Plot $D_{22,s}$ vs "s" (ε_{2t} " on y_{2t})



Long-run effect on each shock on y_{1t} and y_{2t} are:

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

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

Variance Decompositions

Goal: To determine the proportion of the variability of $\boxed{y_{1t+s}}$, $\boxed{y_{2t+s}}$ 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}

$$\text{Forecast error}(s) = y_{t+s} - E[y_{t+s} \mid I_t]$$
$$= D_0 \varepsilon_{t+s} + D_1 \varepsilon_{t+s-1} + \dots + D_{s-1}\varepsilon_{t+1}$$

or

$$\begin{pmatrix} y_{1t+s} \\ y_{2t+s} \end{pmatrix} - \begin{pmatrix} E(y_{1t+s} \mid I_t) \\ E(y_{2t+s} \mid I_t) \end{pmatrix} = \begin{pmatrix} D_{11,0} & D_{12,0} \\ D_{21,0} & D_{22,0} \end{pmatrix}\begin{pmatrix} \varepsilon_{1t+s} \\ \varepsilon_{2t+s} \end{pmatrix} + \begin{pmatrix} D_{11,1} & D_{12,1} \\ D_{21,1} & D_{22,1} \end{pmatrix}\begin{pmatrix} \varepsilon_{1t+s-1} \\ \varepsilon_{2t+s-1} \end{pmatrix}$$
$$+ \dots + \begin{pmatrix} D_{11,s-1} & D_{12,s-1} \\ D_{21,s-1} & D_{22,s-1} \end{pmatrix}\begin{pmatrix} \varepsilon_{1t+1} \\ \varepsilon_{2t+1} \end{pmatrix}$$

Focusing on the first equation

$$y_{1t+s} - E(y_{1t+s} \mid I_t) = D_{11,0}\varepsilon_{1t+s}, \dots + D_{11,s-1}\varepsilon_{1t+1}$$
$$+ D_{12,0}\varepsilon_{2t+s}, \dots + D_{12,s-1}\varepsilon_{2t+1}$$

$$MSE = E[y_{1t+s} - E[y_{1t+s} \mid I_t]]^2 =$$
$$= \sigma_1^2(s) = \sigma_1^2 (D_{11,0}^2 + D_{11,1}^2 + \dots + D_{11,s-1}^2)$$
$$+ \sigma_2^2 (D_{12,0}^2 + D_{12,1}^2 + \dots + D_{12,s-1}^2)$$

The proportion of $\sigma_1^2(s)$ due to shocks in ε_{1t} is

$$P_{11}(s) = \frac{\sigma_1^2 (D_{11,0}^2 + D_{11,1}^2 + \cdots + D_{11,s}^2)}{\sigma_1^2(s)}$$

due to ε_{2t} is

$$P_{12}(s) = \frac{\sigma_2^2 (D_{12,0}^2 + D_{12,1}^2 + \cdots + D_{12,s}^2)}{\sigma_1^2(s)}$$

and similarly

$$P_{21}(s) \ \& \ P_{22}(s)$$

These results are reported usually in the following way

Forecast Period s	y_1 MSE	Proportion due to shock ε_{1t} $P_{11}(s)$	Proportion due to ε_{2t} $P_{12}(s)$
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

(1) δ-method

Remember the "delta" method;

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

$$\sqrt{T}\left(g(\hat{\theta})-g(\theta_0)\right) \simeq \sqrt{T}\left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0}(\hat{\theta}-\theta_0)$$

$$\xrightarrow{d} N\left(0, \left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0} \Sigma_{\hat{\theta}} \left.\frac{\partial g}{\partial \hat{\theta}}\right|_{\theta_0}\right)$$

For SVAR IRFs:

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

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 $\hat{\beta} \sim N(.25, 1)$ which is the distribution of $\hat{\beta}^4$ or $\frac{1}{\hat{\beta}}$?

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

(2) Bootstrap method,

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

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

(iii) Estimate $\hat{A}^l(L)$ by using data constructed in (ii). 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 AR coeff usually are downward biased

All the methods produce POINT confidence intervals (at a single horizon "h") of the IRF when we are interested on its shape.

(+) & (−) of VAR models

(−) (1) Time aggregation

(2) Large dimension

(3) sometimes we have VAR(∞)

(4) Construction of confidence bands for the IRF

(+) They require very little to be used. This is just the opposite than DGSE models. People use VAR models to check DGSE results.