

14.384 Time Series Analysis, Fall 2007  
 Professor Anna Mikusheva  
 Paul Schrimpf, scribe  
 September 27, 2007  
 revised October 5, 2009

Lecture 11

VARs

Notation and some Linear Algebra

Let

$$y_t = \sum_{j=1}^p a_j y_{t-j} + e_t \tag{1}$$

where  $y_t$  and  $e_t$  are  $k \times 1$ , and  $a_j$  is  $k \times k$ .  $e_t$  is white noise with  $Ee_t e_t' = \Omega$  and  $Ee_t e_s' = 0$

**Lemma 1.**  $y_t$  is stationary if  $\det \left( I_k - \sum_{j=1}^p a_j z^j \right) \neq 0$  for all  $|z| \leq 1$ , i.e. all roots of  $\det \left( I_k - \sum_{j=1}^p a_j z^j \right)$  are outside the unit circle.

**Definition 2.** Companion form of (1) is:

$$\mathcal{Y}_t = \mathcal{A} \mathcal{Y}_{t-1} + \mathcal{E}_t,$$

where

$$\mathcal{Y}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \mathcal{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_p \\ I & 0 & \dots & 0 \\ 0 & I & 0 & \dots \\ \vdots & \ddots & & \vdots \\ & & I & 0 \end{bmatrix}, \mathcal{E}_t = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

so  $\mathcal{Y}_t$  and  $\mathcal{E}_t$  are  $kp \times 1$  and  $\mathcal{A}$  is  $kp \times kp$ .

So, any VAR(p) can be wrote as a multi-dimensional VAR(1). From a companion form one can note that

$$\Sigma_{\mathcal{Y}} = \mathcal{A} \Sigma_{\mathcal{Y}} \mathcal{A}' + \Sigma_{\mathcal{E}}$$

This may help to calculate variance-covariance structure of VAR. In particular, we may use the following formula from linear algebra:

$$\text{vec}(ABC) = (C' \otimes A) \text{vec}(B),$$

here  $\otimes$  stays for tensor product and  $\text{vec}(A)$  transforms a matrix to a vector according to the following

semantic rule: if  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ , then  $\text{vec}(A) = [a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32}]'$

$$\begin{aligned} \text{vec}(\Sigma_{\mathcal{Y}}) &= \text{vec}(\mathcal{A} \Sigma_{\mathcal{Y}} \mathcal{A}' + \Sigma_{\mathcal{E}}) \\ &= (\mathcal{A} \otimes \mathcal{A}) \text{vec}(\Sigma_{\mathcal{Y}}) + \text{vec}(\Sigma_{\mathcal{E}}) \\ &\Rightarrow \\ \text{vec}(\Sigma_{\mathcal{Y}}) &= (I - (\mathcal{A} \otimes \mathcal{A}))^{-1} \text{vec}(\Sigma_{\mathcal{E}}) \end{aligned}$$

## Estimation

**Lemma 3.** *MLE (with normal error assumption) = OLS equation-by-equation with  $\hat{\Omega} = \frac{1}{T} \sum \hat{e}_t \hat{e}_t'$*

Intuition: all variables are included in all equations, so there is nothing gained by doing SUR. This also implies that OLS equation by equation is asymptotically efficient. The usual statements of consistency and asymptotic normality hold, as well as OLS formulas for standard errors.

## Granger Causality

Granger Causality is a misleading name. It would be better called Granger predictability.

**Definition 4.** *y fails to Granger cause x if it's not helpful in linear predicting x (in MSE sense). More formally,*

$$\text{MSE} \left[ \hat{E}(x_{t+s} | x_t, x_{t-1}, \dots) \right] = \text{MSE} \left[ \hat{E}(x_{t+s} | x_t, \dots, y_t, y_{t-1}, \dots) \right], \forall s > 0$$

where  $\hat{E}(x_t | z)$  denotes the best linear prediction of  $x_t$  given  $z$

A test of Granger causality is to run OLS:

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + e_t$$

and test  $H_0 : \beta_1 = \beta_2 = \dots = 0$ .

Note that:

- Granger causality is not related to economic causality, it's more about predictability.
- There could be simultaneous casualty or omitted variable problems. For example, there may be a variable  $z$  that causes both  $y$  and  $x$  but with the different lag (sluggish response). If one does not include  $z$  (omitted variable), it may look like  $x$  causes  $y$ .
- Forward looking (rational) expectations may even reverse the causality. For example, suppose analysts rationally predict that a stock is going to pay high dividends tomorrow. That will provoke people to buy the stock today, and the price will rise. In the data you would observe that the price rise is followed by high dividends. So, we would find that prices Granger cause dividends, even though it was really that anticipated high dividends caused high prices. Or increase in orange juice price Granger causes bad weather in Florida.

How to do Granger causality in multivariate case?

Assume  $y_{1t}$  is  $k_1 \times 1$  vector and  $y_{2t}$  is  $k_2 \times 1$ . Assume that we have VAR system

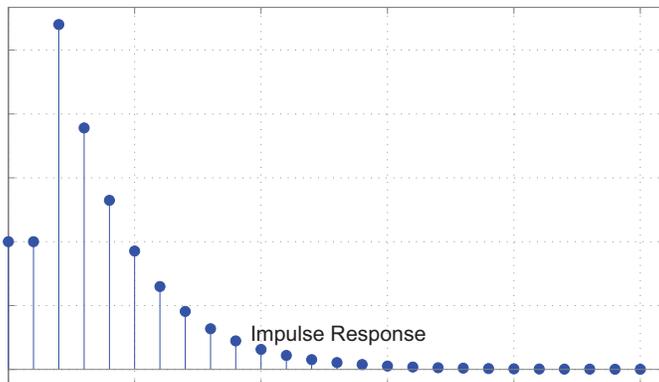
$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} A_1(L)y_{1t-1} + A_2 y_{2t-1} \\ B_1(L)y_{1t-1} + B_2 y_{2t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

Group of variables  $y_2$  fails to *Granger cause*  $y_1$  if  $A_2 = 0$ . To perform this test we have to run unrestricted regression  $y_{1t} = A_1(L)y_{1t-1} + A_2 y_{2t-1} + e_t^u$  and restricted regression  $y_{1t} = A_1(L)y_{1t-1} + e_t^r$ . Then we estimate the corresponding variance-covariance matrix  $\Omega^u = \frac{1}{T} \sum_{t=1}^T e_t^u e_t^{u'}$  and  $\Omega^r = \frac{1}{T} \sum_{t=1}^T e_t^r e_t^{r'}$ . The test statistic compares these matrices:

$$LR = T (\log |\Omega^r| - \log |\Omega^u|)$$

Under the null (absence of Granger Causality)  $LR$  statistic is asymptotically  $\chi^2$  with the degrees of freedom equal to the number of restrictions imposed.

Figure 1: Impulse-response function



## Reporting Results

Reporting the matrix of coefficients is not very informative. There are too many of them, and the coefficients are difficult to interpret anyway. Instead, people present impulse-response functions and variance decompositions.

### Impulse-response

Suppose

$$y_t = a(L)y_{t-1} + e_t$$

with MA representation

$$y_t = c(L)e_t$$

and  $Du_t = e_t$  such that  $u_t$  are orthonormal, i.e.  $E u_t^k u_t^j = I$ . Let  $\tilde{c}(L) = c(L)D$ , so

$$y_t = \tilde{c}(L)u_t$$

**Definition 5.** The *impulse-response* function is  $\frac{\partial y_t^i}{\partial u_{t-j}^k}$ . It is the change in  $y_t^i$  in response to a unit change in  $u_{t-j}^k$  holding all other shocks constant. We can plot the impulse-response function as in figure 1.

To estimate an impulse-response, we would

1. Estimate VAR by OLS –  $\hat{a}$
2. Invert to MA
3. Find and apply rotation  $D$  to get orthonormal shocks – the impulse response is given by  $\hat{c}$

**Standard Errors**

**Delta-method** To calculate standard errors, we can apply the delta-method to  $\hat{a}$  – the  $\hat{c}$  are just some complicated function of  $\hat{a}$ . In practice, we can do this recursively:

$$\begin{aligned} y_t &= a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t \\ &= a_1 (a_1 y_{t-2} + \dots + a_p y_{t-p-1} + e_{t-1}) + a_2 y_{t-2} + \dots + y_{t-p} + e_t \\ &\vdots \end{aligned}$$

so,  $c_1 = a_1$ ,  $c_2 = a_2 + a_1^2$ , etc. We can apply the delta-method to each of these coefficients. We'd also need to apply the delta-method to our estimate of  $D$ . Sometimes, this is done in practice. However, it is not really the best way, for two reasons:

- We estimate many  $a_j$  from not all that big of a sample, so our asymptotics may not be very good.
- This is made even worse by the fact that  $c_k$  are highly non-linear transformations of  $a_j$

Instead of the delta-method, we can use the bootstrap.

**Bootstrap** The typical construction of bootstrap confidence sets would be the following:

1. run regression  $y_t = c + a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t$  to get  $\hat{c}, \hat{a}_1, \dots, \hat{a}_p$  and residuals  $\hat{e}_t$
2. Invert the estimated AR process to get the estimates of impulse response  $\hat{c}_j$  from  $\hat{a}_1, \dots, \hat{a}_p$
3. For  $b = 1..B$ 
  - (a) Form  $y_{t,b}^* = \hat{c} + \hat{a}_1 y_{t-1,b}^* + \dots + \hat{a}_p y_{t-p,b}^* + e_{t,b}^*$ , where  $e_{t,b}^*$  is sampled randomly with replacement from  $\{\hat{e}_t\}$
  - (b) Run regression  $y_{t,b}^* = c + a_1 y_{t-1,b}^* + \dots + a_p y_{t-p,b}^* + e_t$  to get  $\hat{a}_{1,b}^*, \dots, \hat{a}_{p,b}^*$
  - (c) Invert the estimated AR process to get the estimates of impulse response  $\hat{c}_{j,b}^*$  from  $\hat{a}_{1,b}^*, \dots, \hat{a}_{p,b}^*$
4. sort  $\hat{c}_{j,b}^*$  in ascending order  $\hat{c}_{j,(1)}^* \leq \dots \leq \hat{c}_{j,(B)}^*$
5. Form  $\alpha$  confidence interval.

There are at least three way's of forming a confidence set:

1. Efron's interval (percentile bootstrap): uses  $\hat{c}_j$  as a test statistics. The interval is  $[\hat{c}_{j,([B\alpha/2])}^*, \hat{c}_{j,([B(1-\alpha)/2])}^*]$
2. Hall's interval ("turned around" bootstrap) : uses  $\hat{c}_j - c_j$  as a test statistics. It employs the idea of bias correction. The interval is a solution to inequalities

$$\begin{aligned} &\hat{c}_{j,([B\alpha/2])}^* - \hat{c}_j \leq \hat{c}_j - c_j \leq \hat{c}_{j,([B(1-\alpha)/2])}^* - \hat{c}_j \\ \text{or } &[2\hat{c}_j - \hat{c}_{j,([B(1-\alpha)/2])}^*, 2\hat{c}_j - \hat{c}_{j,([B\alpha/2])}^*] \end{aligned}$$

3. studentized bootstrap : uses t-statistics  $t_j = \frac{\hat{c}_j - c_j}{s.e.(\hat{c}_j)}$ . The interval is a solution to inequalities

$$\begin{aligned} &t_{j,([B\alpha/2])}^* \leq \frac{\hat{c}_j - c_j}{s.e.(\hat{c}_j)} \leq t_{j,([B(1-\alpha)/2])}^* \\ \text{or } &[\hat{c}_j - t_{j,([B(1-\alpha)/2])}^* s.e.(\hat{c}_j), \hat{c}_j - t_{j,([B\alpha/2])}^* s.e.(\hat{c}_j)] \end{aligned}$$

*Remark 6.* The bootstrap is still an asymptotic procedure. One advantage of the bootstrap is its simplicity. There is no need to apply the delta-method.

*Remark 7.* There are variations of the bootstrap that also work. For example, you could sample the errors from a normal distribution with variance  $\hat{\Omega}$ . This would be called a parametric bootstrap because we'd be relying on a parametric assumption to create our simulated samples.

**Bootstrap-after-bootstrap** Simulations show that bootstrap works for impulse responses somewhat better than asymptotic (delta-method). This is due to final sample correction - remember that the dependence between  $\{a_j\}$  and  $\{c_j\}$  is non-linear. However, the coverage of these intervals is still very far from ideal, especially for very persistent processes. The main reason for that is  $\hat{a}_j$  are very biased estimates of  $a_j$ . To correct this a bootstrap-after-bootstrap was suggested.

1. run regression  $y_t = c + a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t$  to get  $\hat{c}, \hat{a}_1, \dots, \hat{a}_p$  and residuals  $\hat{e}_t$
2. Invert the estimated AR process to get the estimates of impulse response  $\hat{c}_j$  from  $\hat{a}_1, \dots, \hat{a}_p$
3. For  $b = 1..B$ 
  - (a) Form  $y_{t,b}^* = \hat{c} + \hat{a}_1 y_{t-1,b}^* + \dots + \hat{a}_p y_{t-p,b}^* + e_{t,b}^*$ , where  $e_{t,b}^*$  is sampled randomly with replacement from  $\{\hat{e}_t\}$
  - (b) Run regression  $y_{t,b}^* = c + a_1 y_{t-1,b}^* + \dots + a_p y_{t-p,b}^* + e_t$  to get  $\hat{a}_{1,b}^*, \dots, \hat{a}_{p,b}^*$
4. calculate bias corrected estimates of  $a_j$ :  $\tilde{a}_j = 2\hat{a}_j - \frac{1}{B} \sum_{b=1}^B \hat{a}_{j,b}^*$
5. For  $b = 1..B$ 
  - (a) Form  $\tilde{y}_{t,b}^* = \tilde{a}_1 \tilde{y}_{t-1,b}^* + \dots + \tilde{a}_p \tilde{y}_{t-p,b}^* + \tilde{e}_{t,b}^*$ , where  $e_{t,b}^*$  is sampled randomly with replacement from  $\{\hat{e}_t\}$
  - (b) Run regression  $\tilde{y}_{t,b}^* = c + a_1 \tilde{y}_{t-1,b}^* + \dots + a_p \tilde{y}_{t-p,b}^* + e_t$  to get  $\tilde{a}_{1,b}^*, \dots, \tilde{a}_{p,b}^*$
  - (c) Invert the estimated AR process to get the estimates of impulse response  $\tilde{c}_{j,b}^*$  from  $\tilde{a}_{1,b}^*, \dots, \tilde{a}_{p,b}^*$
6. Form  $\alpha$  confidence interval.

MIT OpenCourseWare  
<http://ocw.mit.edu>

14.384 Time Series Analysis  
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.