Vector Autoregressions

Series de Tiempo

Birkbeck

January 2022

VAR

Uses of Vector Autoregressions:

- Forecast
- Testing Linear Rational Expectations Models.
- Granger Causality.
- Impulse Response Analysis.
- Variance Decomposition.

STRUCTURE OF THE LECTURE

- Definition and Properties.
- Estimation
- VAR analysis.

VAR

Definition

A vector autoregressive (VAR) is simply an autoregressive process for a vector of variables.

Let us define $W_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix}$, *a matrix* A_{2X2} *and* $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$.

where

$$E(\varepsilon_t) = 0, \qquad E(\varepsilon_t \varepsilon_s') = \begin{cases} \Omega & t = s \ (\Omega = \Omega',\ c'\Omega c > 0,\ c \neq 0), \\ 0 & \textit{otherwise} \end{cases}$$

VAR (1)

- Then a **VAR(1)** may be written as

$$W_t = AW_{t-1} + \varepsilon_t$$

or

$$x_t = a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_{1t},$$
$$y_t = a_{21}x_{t-1} + a_{22}y_{t-1} + \varepsilon_{2t},$$

VAR(p)

- A VAR of order p can be written as

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

  Using the lag operator

$$(I - A_1 L - A_2 L^2 - ... - A_p L^p) W_t = \varepsilon_t$$

- The VAR is covariance stationary if all the values of L satisfying $|I - A_1 L - A_2 L^2 - ... - A_p L^p| = 0$ lie outside the unit circle.

The Autocovariance Matrix

- For a covariance stationary n dimensional vector process we may define the *autocovariance function* for a VAR in a way similar to the univariate case:

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) \quad \textit{where} \quad \Gamma_{k(ij)} = cov(W_{i,t}, W_{j,t-k})$$

Example

Using the above two variables VAR we get the following:

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) = \begin{bmatrix} E(x_t x_{t-k}) & E(x_t y_{t-k}) \\ E(y_t x_{t-k}) & E(y_t y_{t-k}) \end{bmatrix}$$

Comments

- Contrary to the univariate case $\Gamma_k \neq \Gamma_{-k}$, instead the correct relationship is:

$$\Gamma_k' = \Gamma_{-k}$$

Proof.

Leading $E(W_t W_{t-k}')$ k periods we get $\Gamma_{k\ (\mathrm{nxn})} = E(W_{t+k} W_t')$. Then, transposing, we obtain:

$$\Gamma_{k\ (\mathrm{nxn})}' = E(W_t W_{t+k}') = \Gamma_{-k}$$

$\square$

Intuition There is no reason why $E(x_t y_{t-1})$ should be equal to $E(x_{t-1} y_t)$.

Example

VAR (1) *autocovariance function*

$$\Gamma_{k\ (nxn)} = E(W_t W'_{t-k}) = AE(W_{t-1} W'_{t-k}) + E(\varepsilon_t W'_{t-k})$$

Thus, for $k \geq 1$:

$$\Gamma_{k\ (nxn)} = A\Gamma_{k-1}.$$

Example

(*continues*) *For* $k = 0$:

$$E(W_t W_t') = AE(W_{t-1} W_{t-1}')A' + E(\varepsilon_t \varepsilon_t')$$

or

$$\Gamma_0 = A\Gamma_0 A' + \Omega$$

In order to obtain Γ_0 *we use the vec operator*

$$vec(\Gamma_0) = vec(A\Gamma_0 A') + vec(\Omega) = (A \otimes A)vec(\Gamma_0) + vec(\Omega),$$

Using that $vec(ABC) = (C' \otimes A)vec(B)$:

$$vec(\Gamma_0) = (I_{(n)^2} - (A \otimes A))^{-1} vec(\Omega).$$

The companion form

- Notice that a VAR(p) may always be re-written as a VAR(1) by defining a vector H_t such that:

$$H_t = FH_{t-1} + \nu_t$$

where

$$H_t = \begin{bmatrix} x_t \\ y_t \\ . \\ x_{t-i} \\ y_{t-i} \\ . \\ x_{t-(p-1)} \\ y_{t-(p-1)} \end{bmatrix} \; F = \begin{bmatrix} A_1 & A_2 & | & ... & A_p \\ I_{2\times 2} & 0 & | & ... & 0 \\ ... & ... & | & I & 0 \end{bmatrix} \; \nu_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ .. \\ 0 \end{bmatrix}$$

The companion form

- Then, the VAR in the Companion form can be expressed in the following way

$$H_t = FH_{t-1} + \nu_t \qquad E(\nu_t \nu_t') = \begin{cases} Q & t = s \\ 0 & \text{otherwise} \end{cases}$$

where

$$Q_{(\text{np x np})} = \begin{bmatrix} \Omega & 0 & 0 & \ldots & 0 \\ 0 & . & & & \\ . & . & & & \\ 0 & 0 & & \ldots & 0 \end{bmatrix}$$

- The variance covariance matrix can be found noticing that

$$E(H_t H_t') = FE(H_{t-1} H_{t-1}')F' + Q$$

or

$$\Sigma = F \Sigma F' + Q \qquad \qquad \textit{where } \Sigma = E(H_t H_t').$$

$$\Sigma = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & & \Gamma_{p-2} \\ & & & \\ \Gamma_{p-1}' & \Gamma_{p-2}' & & \Gamma_0 \end{bmatrix}$$

with Γ_p the autocovariance of the original process

- If the process is covariance stationary, then the unconditional variance can be calculated simply using vec operators, i.e.,

$$vec(\Sigma) = vec(F\Sigma F') + vec(Q) = (F \otimes F)vec(\Sigma) + vec(Q),$$

  Then the unconditional variance can be obtained as

$$vec(\Sigma) = (I_{(np)^2} - (F \otimes F))^{-1} vec(Q).$$

- Notice as well that the j^{th} autocovariance function of H (denoted Σ_j) can be found by post-multiplying by H'_{t-j} and taking expectations.

$$E(H_t H'_{t-j}) = FE(H_{t-1} H'_{t-j}) + E(\nu_t H'_{t-j})$$

- Thus,

$$\Sigma_k = F\Sigma_{k-1} \quad para\ k = 1, 2, ...$$

or

$$\Sigma_k = F^k \Sigma.$$

- The k^{th} autocovariance Γ_k of the original process W_t is given by the n first rows and n columns of $\Sigma_k = F\Sigma_{k-1}$:

$$\Gamma_k = A_1\Gamma_{k-1} + A_2\Gamma_{k-2} + ... + A_p\Gamma_{k-p} \qquad k = p, p+1, p+2, ...$$

The Conditional likelihood for a vector autoregression.

- Let W_t denote an $(n_{\times}1)$ vector which we assume follows a p^{th} order Gaussian VAR.

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t \qquad \varepsilon_t \tilde{} N(0, \Omega),$$

- The approach is to condition on the first p observations $(W_0, ...W_{-p+1}\ )$ and to base the estimation on the last T observations $(W_T, ...W_1)$.

$$f(W_T, W_{T-1}, W_{T-2}....W_1 | W_0, ., W_{-p+1}; \Theta)$$

and maximize with respect to Θ, where Θ is a vector that contains the elements of $A_1, A_2, A_3, ...A_p$ and Ω.

- Then,

$$W_t | W_{t-1}, ..., W_{-p+1} \tilde{} N \left(A_1 W_{t-1} + ... + A_p W_{t-p}, \Omega_{n_\times n} \right)$$

- It will be convenient to stack the p lags in a vector x_t,.

$$x_t = \begin{bmatrix} \underbrace{W_{t-1}}_{nx1} \\ \underbrace{W_{t-2}}_{nx1} \\ . \\ \underbrace{W_{t-p}}_{nx1} \end{bmatrix}_{npx1}$$

- let Π' denote the following $n_{\times} np$ matrix :

$$\Pi' = [A_1, A_2, A_3, ...A_p]_{n\times np}$$

- Thus, the conditional mean is just

$$\Pi' x_t$$

- Therefore,

$$W_t | W_{t-1,}...., W_{-p+1} \tilde{} N(\Pi' x_t, \Omega)$$

or

$$\begin{aligned} & f(W_t | W_{t-1}, .., W_{-p+1}; \Theta) \\ = \quad & (2\pi)^{-n/2} |\Omega^{-1}|^{.5} \exp[(-1/2)(W_t - \Pi' x_t)' \Omega^{-1} (W_t - \Pi' x_t)] \end{aligned}$$

- The joint density conditional on the first p observations can be written as:

$$
\begin{aligned}
& f(W_T, W_{T-1}, W_{T-2} \ldots . W_1 | W_0, ., W_{-p+1}; \Theta) \\
= & \prod_{t=1}^{T} f(W_t | W_{t-1}, .., W_{-p+1}; \Theta)
\end{aligned}
$$

- Taking logs

$$
\begin{aligned}
L(\Theta) &= \sum_{t=1}^{T} \ln[f(W_t | W_{t-1}, .., W_{-p+1}; \Theta)] \\
&= -(Tn/2) log(2\pi) + (T/2) log|\Omega^{-1}| \\
&\quad -(1/2) \sum_{t=1}^{T} (W_t - \Pi' x_t)' \Omega^{-1} (W_t - \Pi' x_t)
\end{aligned}
$$

Estimating the parameters in A

It turns out to be that the maximum likelihood estimator is

$$\hat{\Pi}' = [\sum_{t=1}^{T} W_t x_t'][\sum_{t=1}^{T} x_t x_t']^{-1}$$

where the j column is just

$$\hat{\pi}_{j\ (1\times \mathrm{np})} = [\sum_{t=1}^{T} W_{jt} x_t'][\sum_{t=1}^{T} x_t x_t']^{-1}$$

The Maximum likelihood estimator of

Ω

- We can now "concentrate" the likelihood using the previous results to find the MLE estimator of Ω
- Evaluate the likelihood at the estimate of Π

$$L(\Omega, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\Omega^{-1}| - (1/2)\sum_{t=1}^{T} \hat{\varepsilon}_t' \Omega^{-1} \hat{\varepsilon}_t$$

- Using the differentation rules $\frac{\partial x' A x}{\partial A} = xx'$, $\frac{\partial \log |A|}{\partial A} = \left(A^{-1}\right)'$and taking the derivative of $L(\Omega, \hat{\Pi})$ with respect to Ω^{-1}

$$\frac{\partial L(\Omega, \hat{\Pi})}{\partial \Omega^{-1}} = (T/2)\Omega' - (1/2)\sum_{t=1}^{T} (\hat{\varepsilon}_t \hat{\varepsilon}_t').$$

- Equating this expression to zero we obtain the MLE of the variance-covariance matrix.

$$\hat{\Omega}' = (1/T)\sum_{t=1}^{T} (\hat{\varepsilon}_t \hat{\varepsilon}_t')$$

Comments

Row i, column i of $\hat{\Omega}$ is given by

$$\hat{\sigma}_i^2 = (1/T) \sum_{t=1}^{T} (\hat{\varepsilon}_{it}^2)$$

which is just the average squared residual from a regression of a variable of the VAR on the p lags of all variables
Therefore I can use OLS results to construct both $\hat{\Omega}$ and $\hat{\Pi}$.

Choosing the order of VAR

1. The validity of the tests we carry out depend on having identified the order of the VAR correctly
   - A simple way to do so is comparing likelihood ratios
     - These can be easily computed because we are using OLS
   - Consider the likelihood function at is Maximum value of a VAR with p_0 lags:

$$L_0(\hat{\Omega}, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_0^{-1}| - (1/2)\sum_{t=1}^{T} \hat{\varepsilon}_t' \hat{\Omega}_0^{-1} \hat{\varepsilon}_t.$$

Consider now the last term of this equation,

$$
\begin{aligned}
(1/2)\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}_0^{-1}\hat{\varepsilon}_t \underbrace{=}_{(\text{ a scalar})} & TR((1/2)\sum_{t=1}^{T}\hat{\varepsilon}_t'\hat{\Omega}_0^{-1}\hat{\varepsilon}_t) \\
\underbrace{=}_{TR(A.B)=TR(B.A)} & (1/2)TR(\sum_{t=1}^{T}\hat{\Omega}_0^{-1}\hat{\varepsilon}_t\hat{\varepsilon}_t') \\
\underbrace{=}_{\hat{\Omega}_0=\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'/T)} & (1/2)TR(\hat{\Omega}_0^{-1}T\hat{\Omega}_0) \\
= & (T/2)TR(I) = (nT)/2.
\end{aligned}
$$

Thus,

$$L_0(\hat{\Omega},\hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_0^{-1}| - (nT)/2$$

- If we want to test the Hypothesis that the VAR has p lags against p_0 lags we calculate the likelihood for the VAR with p_1 lags ($p_1 > p_0$)

$$L_1(\hat{\Omega}, \hat{\Pi}) = -(Tn/2)log(2\pi) + (T/2)log|\hat{\Omega}_1^{-1}| - (nT)/2$$

and compute the likelihood ratio which is

$$\begin{aligned} & 2(L_1(\hat{\Omega}, \hat{\Pi}) - L_0(\hat{\Omega}, \hat{\Pi})) \\ = \quad & T(\log|\hat{\Omega}_1^{-1}| - \log|\hat{\Omega}_0^{-1}|)\tilde{\ }\chi^2(n^2(p_1 - p_0)) \\ & \text{under } H_0 \end{aligned}$$

- Sims (1980) proposed the following for small samples:

$$(T-k)(\log|\hat{\Omega}_1^{-1}| - \log|\hat{\Omega}_0^{-1}|)$$

where $k = np_1 = \max\{\text{number of parameters estimated per equation}\}$

Goodness of Fit Criteria

- Measures how good a model is relative to others
- Balance between fit and complexity
- Typically, we would like to minimize:

$$C(p) = -2max(logL) + \beta(number\ of\ freely\ estimated\ parameters)$$

- For Gaussian models, the maximized log-likelihood is proportional to

$$-(T/2)log|\Omega| \qquad\qquad (since|\Omega^{-1}| = 1/|\Omega|)$$

Goodness of Fit Criteria

- Hence, we choose p to minimize:

$$C(p) = Tlog|\Omega| + \beta(n^2 p)$$

- For example

$$\begin{array}{ll} \text{AIC} & \beta = 2 \text{ (Akaike information criterion)} \\ \text{SBC} & \beta = log(T) \text{ (Scharz Bayesian criterion)} \\ \text{HQ} & \beta = 2log(log(T)) \text{ (Hannan-Quin criterion)} \end{array}$$

- Alternatively the Akaike's prediction error (FPE) criterion chooses p so that to minimize the expected one -step ahead squared forecast error:

$$FPE = [\frac{T + np + 1}{T - np - 1}]^n |\Omega|$$

Asymptotic Distribution of the VAR estimators

- MLE will give consistent estimators of Π and Ω
- Standard errors of $\hat{\Pi}$ are given by standard OLS formulas
- Let $\hat{\pi}_T = vec(\hat{\Pi}_T)$ denote the $nk_{\times}1$ vector of coefficients resulting from OLS. Then

$$\sqrt{T}(\hat{\pi}_T - \pi) \underrightarrow{L} N(0, \Omega \otimes Q^{-1})$$

  where $Q = E(x_t x_t')$.
- Standard OLS t and F statistics applied to the coefficients of any single equation in the VAR are asymptotically valid.

Testing Rational Expectations Hypothesis

- These models usually impose non-linear cross equation restrictions between the parameters of the model which are tested using a likelihood ratio test
- Consider a first order bivariate VAR:

$$\begin{aligned} x_t &= a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_t \\ y_t &= a_{21}x_{t-1} + a_{22}y_{t-1} + \nu_t \end{aligned}$$

  where x_t is the interest rates differential and y_t is the first difference of the logs of the spot exchange rate
- Then uncovered interest parity can be written as

$$x_t = E_t y_{t+1}.$$

Testing Rational Expectations Hypothesis

- Condition on both sides of the previous equation on $t-1$,we get the following restrictions

$$\begin{bmatrix} 1 & 0 \end{bmatrix} A = \begin{bmatrix} 0 & 1 \end{bmatrix} A^2$$

which can be expressed as:

$$a_{11} = a_{22}a_{21}/(1-a_{21}) \qquad a_{12} = a_{22}^2/(1-a_{21})$$

- Estimate the unrestricted and the restricted model and perform a likelihood ratio test:

$$2(L_u - L_r)\tilde{}\, asymptotically \chi^2$$

Granger Causality

Definition

y fails to Granger-cause x if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...)$ is the same as the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, ...)$ and $(y_t, y_{t-1}, ...)$. For linear functions

$$MSE[E(x_{t+s}|x_t, x_{t-1}, ...)] = MSE[E(x_{t+s}|x_t, x_{t-1}, ..., y_t, y_{t-1}, ...)]$$

Remark Granger's reason for proposing this definition was that if an event Y is the cause of another event X, then the event Y should precede the event X.

Granger Causality

- The null hypothesis is that y fails to Granger - cause x
- We just regress both the general model

$$x_t = a_{11}x_{t-1} + a_{12}y_{t-1} + \varepsilon_{1t}$$

and the restricted model

$$x_t = a_{11}x_{t-1} + \varepsilon'_{1t}$$

and compare the residuals sum squares
$T(RRS(\varepsilon') - RRS(\varepsilon))/RRS(\varepsilon)\tilde{}\chi^2(1)$ (asymptotically)

Conditioning on the correct information set

- Omitting a relevant variable in the information set may give spurious results, since a variable that is thought to be useful for forecasting others, may be not longer useful once you condition on the right information.
- The question of whether a scalar y can help forecast another scalar x needs to be accommodated considering the information about z.
- Then, y fails to Granger-cause x if for all $s > 0$ the mean squared error of a forecast of x_{t+s} based on $(x_t, x_{t-1}, .z_t, z_{t-1}...)$ is the same as the MSE of a forecast of x_{t+s} based on $(x_t, x_{t-1}, .z_t, z_{t-1}...)$ and $(y_t, y_{t-1}, ...)$. For linear functions

$$
\begin{aligned}
& MSE[E(x_{t+s}|x_t, x_{t-1}, .z_t, z_{t-1}, .)] \\
= \quad & MSE[E(x_{t+s}|x_t, x_{t-1}, ..z_t, z_{t-1}.., y_t, y_{t-1}, .)]
\end{aligned}
$$

Granger Causality

Example

The Market efficiency hypothesis yields prices as a function of dividends:

$$P_t = \sum_{i=1}^{\infty} (1/(1+r))^i E(D_{t+i}|I_t)$$

Suppose

$$D_t = d + u_t + \delta u_{t-1} + \nu_t$$

where u_t and ν_t are independent white noise processes, then

$$E_t D_{t+i} = \begin{cases} d + \delta u_t & \text{for } i = 1 \\ d & \text{for } i = 2, 3, \ldots \end{cases}$$

Granger Causality

Example

The stock prices will be given by

$$P_t = d/r + \delta u_t/(1+r)$$

Thus, P_t *is a white noise: no series should granger cause stock prices. Nevertheless, notice that:*

$$\delta u_{t-1} = (1+r)P_{t-1} - (1+r)d/r$$

Substituting back in the D_t*:*

$$D_t = d + u_t + (1+r)P_{t-1} - (1+r)d/r + \nu_t$$

Thus stock prices Granger cause dividends

Granger Causality

Example

The bivariate VAR takes the form

$$\begin{bmatrix} P_t \\ D_t \end{bmatrix} = \begin{bmatrix} d/r \\ -d/r \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ (1+r) & 0 \end{bmatrix} \begin{bmatrix} P_{t-1} \\ D_{t-1} \end{bmatrix} + \begin{bmatrix} \delta u_t/(1+r) \\ u_t + \nu_t \end{bmatrix}$$

Hence in this model, Granger causation runs in the opposite direction from the true causation.

Impulse Response Functions

- Recall a stationary VAR has a VMA(∞) representation:

$$W_t = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t-z} \quad , \psi_0 = I$$

- Lead the above expression s periods:

$$W_{t+s} = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t+s-z}$$

- Evaluate the above expression at $z = s$. Then

$$\psi_s = \frac{\partial W_{t+s}}{\partial \varepsilon_t'}$$

has the interpretation of a dynamic multiplier

Impulse Response Functions

- $(\psi_s)_{ij}$ = effect of a one unit increase in the j^{th} variable's innovation at time t (ε_{jt}) for the value of the i^{th} variable at time $t+s$ ($W_{i,t+s}$), holding all other innovations at all dates constant
- You can find these multipliers numerically by simulation:
  - set $W_t = ... = W_{t-p} = 0$, then set $\varepsilon_{jt} = 1$ and all the other terms to zero, and simulate the system

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

  for $t, t+1, t+s$, with $\varepsilon_{t+1}, \varepsilon_{t+2}, ... = 0$ This simulation corresponds to the J column of the matrix ψ_s . By doing this for other values of j we get the whole matrix.

Impulse Response Functions

Definition

A plot of $(\psi_s)_{ij}$, that is row i column j of ψ_s , as a function of s is called the impulse response function. It describes the response of $W_{i,t+s}$ to a one time impulse in W_{jt} with all other variables dated t or earlier held constant.

Impulse Response Functions

- Define interim multipliers:

$$\sum_{j=1}^{m} \psi_j$$

- and the long run multiplier:

$$\sum_{j=1}^{\infty} \psi_j.$$

Impulse Response Function

- The assumption that a shock in one innovation does not affect others is problematic since

$$E(\varepsilon_t \varepsilon_t') = \Omega \neq \text{a diagonal matrix}$$

- Since Ω is symmetric and positive definite, it can be expressed as

$$\Omega = ADA'$$

where A is a lower triangular matrix and D is a diagonal Matrix.

Impulse Response Function

- Let $u_t = A^{-1}\varepsilon_t$, then

$$W_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j A A^{-1} \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j^* u_{t-j}$$

where

$$\psi_j^* = \psi_j A$$

$$E(u_t u_t') = E(A^{-1}\varepsilon_t \varepsilon_t' (A^{-1})') = A^{-1}\Omega(A^{-1})' = A^{-1}ADA'(A^{-1})' = D$$

- The matrix D gives the variance of u_{jt}

Impulse Response Function

- A plot of ψ_s^* as a function of s is known as an orthogonalized impulse response function.
- The matrix
$$\psi_s^* = \frac{\partial W_{t+s}}{\partial u_t'}$$
gives the consequences of an increase in W_{jt} by a unit impulse in u_t .
- Notice that
$$\psi_0^* = \psi_0 A = IA$$
is lower triangular. This implies that the ordering of variables is of importance.
- The ordering cannot be determined with statistical methods.

Variance Decomposition

- Consider the error in forecasting a VAR s periods ahead:

$$W_{t+S} - \widehat{W}_{t+S|t} = \sum_{j=0}^{s-1} \psi_j \varepsilon_{t+s-j}, \qquad \psi_0 = I$$

- The mean squared error of this s-period ahead forecast is thus

$$MSE(\widehat{W}_{t+s|t}) = \Omega + \psi_1 \Omega \psi_1' + ... + \psi_{s-1} \Omega \psi_{s-1}'$$

- Let us now consider how each of the orthogonalized disturbances $(u_{1t}, ..u_{nt})$ contributes to this MSE.

Variance Decomposition

- Lets write

$$\varepsilon_t = Au_t = a_1 u_{1t} + ...a_n u_{nt},$$

where a_j denotes the j^{th} column of the matrix A.

- Recalling that the u's are uncorrelated, we get

$$\Omega = a_1 a'_{1(n\times n)} Var(u_{1t}) + ... + a_n a'_{n(n\times n)} Var(u_{nt})$$

Variance Decomposition

- Substituting this in the MSE of the s period ahead forecast we get

$$MSE(\widehat{W}_{t+s|t}) = \sum_{j=1}^{n} Var(u_{jt})(a_j a_j' + \psi_1 a_j a_j' \psi_1' + ... + \psi_{s-1} a_j a_j' \psi_{s-1}')$$

- With this expression we can calculate the contribution of the j^{th} orthogonalized innovation to the MSE of the s-period ahead forecast.

$$Var(u_{jt})(a_j a_j' + \psi_1 a_j a_j' \psi_1' + \psi_2 a_j a_j' \psi_2' ... + \psi_{s-1} a_j a_j' \psi_{s-1}')$$

- Magnitude in general depends on the ordering of the variables

Structural VARs

- Blanchard (1989) considers the following structure

$$\varepsilon_{1t} = e u_{2t} + u_{1t}$$

$$\varepsilon_{2t} = c_{21}\varepsilon_{1t} + u_{2t}$$

where u_{1t} and u_{2t} are regarded as demand and supply shocks, while ε_{1t} and ε_{2t} are output and unemployment innovations respectively

Structural VARs

- Blanchard and Quah (1989) :
  - They argue that a demand shock should have a zero long-run effect while a supply shock will not.

$$\varepsilon_{1t} = a_1 u_{2t} + u_{1t}$$

$$\varepsilon_{2t} = a_2 u_{1t} + u_{2t}$$

  where the covariance of u_{jt} is assumed to be zero.

- Consider

$$W_t = A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p} + \varepsilon_t$$

is estimated and the implied MA representation is

$$W_t = \sum_{z=0}^{\infty} \psi_z \varepsilon_{t-z} \qquad , \psi_0 = I.$$

Structural VARs

- In terms of the shocks of interest, we will write

$$\varepsilon_t = Au_t$$

where A is now defined as

$$A = \begin{bmatrix} 1 & a_1 \\ a_2 & 1 \end{bmatrix}.$$

- The MA representation in terms of the u_t shocks becomes:

$$W_t = \sum_{z=0}^{\infty} \psi_z A u_{t-z} \qquad , \text{ where } \psi_0 = I$$

- If the long run effect of a demand shock upon output say, W_{1t} , is to be zero:

$$[\sum_{z=0}^{\infty} \psi_z A]_{[1,1]} = 0$$

Structural VARs

- To see how we compute this restriction we first notice that

$$\sum_{z=0}^{\infty} \psi_z$$

is a 2x2 Matix. Let the first row of this matrix be

$$[\delta_1, \delta_2]$$

- Then, the restriction is just

$$\delta_1 + a_2\delta_2 = 0 \text{ or } a_2 = -\delta_1/\delta_2.$$

- Thus one parameter can be found from this restriction

- The other three come from the fact that:

$$V(\varepsilon_t) = A \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} A'$$

since there are three unknowns in $V(\varepsilon_t)$ to determine a_1, σ_1^2 and σ_2^2.

- All that is needed is to estimate δ_1, δ_2 .

- Notice that the long run multiplier is easy to compute, since we know that

$$\sum_{z=0}^{\infty} \psi_z = \psi(1)$$

$$\sum_{z=0}^{\infty} \psi_z L^i = \psi(L)$$

$$\sum_{z=0}^{\infty} \psi_z L^i = (I - A_1 L - A_2 L^2 - ... + A_p L^p)^{-1} = A(L)^{-1}$$

$$\psi(1) = (I - A_1 - A_2 - .... + A_p)^{-1} = A(1)^{-1}$$

- All the information that is needed for the impulse response function, is obtained from the estimated parameters in the VAR.

Impulse Resonponse Functions using Local Projections

- It has been proposed in the literature an alternative way of carrying out impulse response functions by doing local projections. Assume you know you have a VAR(1) (which can be the companion form of a VAR(p)) of the type.

$$Y_t = AY_{t-1} + \varepsilon_t$$

Now instead of analysing the $MA(\infty)$ we substitute backwards τ times to get

$$Y_{t+\tau} = A\varepsilon_{t+\tau-1} + A^2\varepsilon_{t+\tau-2} + ... + A^\tau\varepsilon_t + A^{\tau+1}Y_{t-1} + \varepsilon_{t+\tau}$$

or

$$Y_{t+\tau} = A^{\tau+1}Y_{t-1} + u_{t+\tau}$$

where

$$u_{t+\tau} = A\varepsilon_{t+\tau-1} + A^2\varepsilon_{t+\tau-2} + ... + A^\tau\varepsilon_t + \varepsilon_{t+\tau}$$

- Then the coefficient of a regression of regression of $Y_{t+\tau}$ on Y_{t-1} has the interpretation of $\frac{\partial Y_{t+\tau}}{\partial \varepsilon_{t-1}}$, that is the the $\tau + 1$ impulse response

- Points to consider

1) The shock needs to be identify. We can simply do an orthogonalization.
$\frac{\partial Y_{t+\tau}}{\partial B\varepsilon_{t-1}} = A^{\tau+1}B^{-1}$.

where $B\varepsilon_{t-1}$ is the orthogonal shock.
2) If It is a VAR(1), then estimating

$$Y_{t+\tau} = A^{\tau+1}Y_{t-1} + u_{t+\tau}$$

give unbiased but probably very imprecise estimates of $A^{\tau+1}$, with standard errors that need to be corrected (all that can be done using the theoretical structure of $u_{t+\tau}$).