Vector Autoregressions

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 $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$.

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

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

or

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

where

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

VAR(p)

$$W_t = A_1W_{t-1} + A_2W_{t-2} + ... + A_pW_{t-p} + \varepsilon_t$$

or

$$(I - A_1L - A_2L^2 - ... - A_pL^p)W_t = \varepsilon_t$$

The VAR is covariance stationary if all the values of L satisfying $|I - A_1L - A_2L^2 - ... - A_pL^p| = 0$ lie outside the unit circle[1].

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

$$\Gamma_{k \ (\text{nxn})} = E(W_tW_{t-k}') \quad where \quad \Gamma_{k(ij)} = cov(W_{i,t}, W_{j,t-k})$$

Using the above two variables VAR we get the following

[1] For the VAR(1) example with 2 variables this is equivalent to say that the real part both roots in L are greater than 1.

$$\text{Det}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} L \right] = 0$$

or

$$1 - (a_{11} + a_{22})L + (a_{11}a_{22} - a_{12}a_{21})L^2 = 0$$

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

Contrary to the univariate case $\Gamma_k \neq \Gamma_{-k}$, instead the correct relationship is $\Gamma'_k = \Gamma_{-k}$. This can be easily understood looking at the simple bivariate case where there is no reason why $E(x_t y_{t-1})$ should be equal to $E(x_{t-1} y_t)$.

Proof

i) Leading the vector k times we obtain

$$\Gamma_{k\ (\text{nxn})} = E(W_{t+k} W'_t)$$

ii) Taking transposes

$$\Gamma'_{k\ (\text{nxn})} = E(W_t W'_{t+k}) = \Gamma_{-k}$$

In terms of our two variables example

$$\Gamma_{k\ (\text{nxn})} = \begin{bmatrix} E(x_t x_{t-k}) & E(x_t y_{t-k}) \\ E(y_t x_{t-k}) & E(y_t y_{t-k}) \end{bmatrix};$$

$$\Gamma'_{k\ (\text{nxn})} = \begin{bmatrix} E(x_t x_{t-k}) & E(y_t x_{t-k}) \\ E(x_t y_{t-k}) & E(y_t y_{t-k}) \end{bmatrix} \overbrace{=}^{\textit{Leading } k \textit{ periods}} \begin{bmatrix} E(x_{t+k} x_t) & E(y_{t+k} x_t) \\ E(x_{t+k} y_t) & E(y_{t+k} y_t) \end{bmatrix};$$

$$\Gamma_{-k\ (\text{nxn})} = \begin{bmatrix} E(x_t x_{t+k}) & E(x_t y_{t+k}) \\ E(y_t x_{t+k}) & E(y_t y_{t+k}) \end{bmatrix}.$$

Autocovariance Function

We can now calculate the autocovariance function for a VAR more or less in the same way we did for the univariate process. Assume for simplicity that we have a first order VAR.

Then the autocovariance function can be derived as

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

then we obtain that for $k \geq 1$

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

When $k = 0$ we can simply notice that

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

or

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

To obtain Γ_0 we use the fact that $vec(ABC) = (C' \otimes A)vec(B)$, then

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

or

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

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

$$H_t = \begin{bmatrix} x_t \\ y_t \\ . \\ x_{t-i} \\ y_{t-i} \\ . \\ x_{t-(p-1)} \\ y_{t-(p-1)} \end{bmatrix}, \quad F = \begin{bmatrix} \mathrm{A}_1 & \mathrm{A}_2 & | & & \mathrm{A}_p \\ \mathrm{I}_{2\times 2} & 0 & | & & 0 \\ & \mathrm{I}_{2\times 2} & & & \\ 0 & 0 & \mathrm{I}_{2\times 2} & & 0 \\ & & | & \mathrm{I} & 0 \end{bmatrix} \quad and\, \nu_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ . \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

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

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

where

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

Notice also that the following relationships holds:

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

or

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

where

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

and Γ_p is 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.,

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

(Since $vec(ABC) = (C' \otimes A)vec(B)$)

Then the unconditional variance can be obtained as

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

(NB $F \otimes F$ has no unit eigen values since the eigen values of $F \otimes F$ are of the form $\lambda_j \lambda_i$, and we knew that all $|\lambda_i| < 1$)

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

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

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 $\Sigma_k = F\Sigma_{k-1}$:

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

Maximum likelihood Estimation.
The Conditional likelihood for a vector autoregression.

Let W_t denote an $(n_\times 1)$ vector containing the values that n variables take at time t. W_t is assumed to follow a p^{th} order gaussian VAR.

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

where both Ω and A_i are $n_\times n$ matrices of parameters.

Suppose that we observe these variables for $T + p$ time periods. 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

$$W_t | W_{t-1}, W_{t-2}, ...W_{-p+1} \tilde{} N((A_1 W_{t-1} + A_2 W_{t-2} + ... + A_p W_{t-p})_{n\times 1}, \Omega_{n\times n})$$

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

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

and let Π' denote the following $n_\times np$ matrix ; $\Pi' = [A_1, A_2, A_3, ...A_p]_{n\times np}$, then we can write the conditional mean as $\Pi' x_t$.

Using this notation, we can write the conditional distribution of W_t as $W_t | W_{t-1}, W_{t-2}, ...W_{-p+1} \tilde{} N(\Pi^{'} x_t, \Omega)$

$$\begin{aligned} & f(W_t | W_{t-1}, W_{t-2}....W_1, W_0, .., W_{-p+1}; \Theta) \\ = \quad & (2\pi)^{-n/2} |\Omega^{-1}|^{.5} \exp[(-1/2)(W_t - \Pi' x_t)' \Omega^{-1} (W_t - \Pi' x_t)] \end{aligned}$$

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

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

and taking logs,

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

It turns out to be (for a proof see textbook) 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 (evaluated at the estimate of Π).

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

taking the derivative with respect to Ω^{-1}

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

and using the following results from matrix algebra: $\frac{\partial(x'Ax)}{\partial A} = xx'$ and $\frac{\partial log|A|}{\partial A} = (A')^{-1}$,we can differentiate the concentrated likelihood with respect to Ω^{-1}

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

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

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

A very important result is that the 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. Analogously the row i column

j of $\hat{\Omega}'$ is given by $\hat{\sigma}_{ij}^2 = (1/T)\sum_{t=1}^{T}(\hat{\varepsilon}_{it}\hat{\varepsilon}_{jt})$ which is the average product of the OLS residual for variable i and the OLS residual foe variable j. Therefore I can use OLS results to construct both $\hat{\Omega}$ and $\hat{\Pi}$.

How to choose the order of a VAR

The results of any test that we carry out using a VAR crucially depend on identifying correctly the order of that VAR. An easy way to attempt to identify the order of a VAR is to perform likelihood ratio tests. To do this turns out to be computationally very simple since the test can be constructed using OLS results.

Consider the likelihood function at is Maximum value of a VAR with p_0 lags, denoted

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

Consider now the last term of this equation,

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

then

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

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

$$2(L_1(\hat{\Omega}, \hat{\Pi}) - L_0(\hat{\Omega}, \hat{\Pi})) = T(\log|\hat{\Omega}_1^{-1}| - \log|\hat{\Omega}_0^{-1}|)$$

which is distributed under the Null χ^2 with degrees of freedom equal to the number of restrictions imposed under H_0, $n^2(p_1 - p_0)$.

Sims (1980) proposed a modification of the likelihood ratio test to take into account small sample bias

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

where $k = np_1$ = number of parameters estimated per equation.

Goodness of fit Criteria

The goodness of fit criteria are measures of how good a model is relative to others. They reflect a balance between the model's goodness of fit and the complexity of the model.

Typically, we want to minimize a scalar measure such as

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

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

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

Hence, we choose p to minimize:

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

AIC	$\beta = 2$ (Akaike information criterion)
SBC	$\beta = log(T)$ (Scharz Bayesian criterion)
HQ	$\beta = 2log(log(T))$ (Hannan-Quin criterion)

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

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

AIC and FPE are not consistent, so that asymptotically they overestimate p with positive probability. SBC and HQ are consistent in the sense that $\hat{p} \rightarrow p$.

Asymptotic Distribution of $\hat{\Pi}$

The maximum likelihood estimates of $\hat{\mathbf{\Pi}}$ and $\hat{\Omega}$ will give consistent estimates of the population parameters. Standard errors for $\hat{\mathbf{\Pi}}$ can be based on the usual OLS formulas.

Let $\hat{\pi}_T = vec(\hat{\mathbf{\Pi}}_T)$ denote the $nk_{\times}1$ vector of coefficients resulting from OLS regressions of each of the elements of W_t on x_t

Then

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

where $Q = E(x_t x_t')$.

This establishes that the standard OLS t and F statistics applied to the coefficients of any single equation in the VAR are asymptotically valid.

$$\sqrt{T}(\hat{\pi}_{iT} - \pi_i) \underrightarrow{L} N(0, (\sigma_i^2 Q^{-1}))$$

where $\sigma_i^2 = E(\varepsilon_{it}^2)$.

Main uses of Vector Autoregressions

i) Forecasting
ii) Testing Hypothesis
iii) Granger Causality
iv) Use of Impulse Response Functions
v) Use of the variance decomposition.

ii) Testing Rational expectations Hypothesis

The VAR methodology is very useful to test linear 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 which is distributed under the Null (that the model is correct) as a χ^2 distribution with degrees of freedom equal to the number of restrictions imposed by the model.

Consider a first order bivariate VAR

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

Assume that x_t is the interest rates differential, say $(i_t - i_t^*)$, and that y_t is the first difference of the logs of the spot exchange rate $(e_t - e_{t-1})$;

Then uncovered interest parity can be written as

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

Then if we condition on both sides of the previous equation on information available at $t-1$ we get the following set of non linear restrictions.

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

The above equation can be easily solved and yields the following nonlinear restrictions on the parameters

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

Then we simply estimate the unrestricted model and the restricted (a function of only two parameters (and the variance-covariance)), and perform a likelihood ratio test, where $2(L_u - L_r)$~asymptotically χ^2 (number of restrictions=2).

iii)Granger Causality

One of the key questions that can be addressed with vector autoregressions is how useful some variables are for forecasting others.

Definition

The question investigated is whether a scalar y can help forecast another scalar x. If it cannot, then we say that y does not Granger-cause x

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 $(x_t, x_{t-1}, ...)$ 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

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

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.

Testing for Granger causality

Ho) y does not cause x or $a_{12} = 0$

We just regress both the general model

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

and the restricted model

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

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

Granger-Causality Tests and Forward-Looking Behaviour

Let us assume risk neutral agents such that stock prices may be written as

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

suppose

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

where u_t and ν_t are independent white noise processes, then

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

The stock prices will be given by

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

Thus for this example the stock price is a white noise and could not be forecast on the basis of lagged prices or dividends. No series should granger cause stock prices.

Nevertheless, notice that using the stock price equation and rearranging terms, I might express

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

Then substituting back in the Dividend process we get the following expression for dividends

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

Thus stock prices Granger cause dividends

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. Dividends fail to G-C prices even though investors' perceptions of dividends are the sole determinant of stock prices. On the other hand, prices do "granger-cause" dividends, even though the market's evaluation of the stock in reality has no effect on the dividend process.

iv) The Impulse - Response Function

If a VAR is stationary it can always be written as an infinite vector moving average. Consider the following vector infinite moving average representation of W_t

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

Analogously, if we lead the above expression s periods we get

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

Therefore we can easily see from the above expression (evaluated at $z = s$) that matrix ψ_s has the interpretation of a dynamic multiplier

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

(dynamic multiplier or impulse response) where $(\psi_s)_{ij}$ = 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[2].

A simple way of finding these multipliers numerically is by simulation. To implement the 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.

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.

[2] Consider a first order two-variable Var. The relevant term of the infinite moving average is

$$\psi_s \varepsilon_t = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

Now assume a unit increase in ε_1 (i.e. $\varepsilon_1 = 1$ and $\varepsilon_2 = 0$), then the effect on W_{t+s} is $\begin{bmatrix} a \\ c \end{bmatrix}$. (that is **a** on the first variable and **c** on the second). Now varying s we can get a plot of the effects of a shock in the innovation of variable 1 on both variables. A similar argument can be constructed for a schock in ε_2, therefor the columns, first or second, of ψ_s represent the effect on each variable of W at time $t + s$ of a unit shock to the first or second variable innovation, keeping the other constant.

We can also define the Interim multipliers, which are given by the accumulated responses over m periods

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

and the long run multiplier which give the total accumulated effects for all future time periods:

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

The assumption that a shock in one innovation does not affect others is problematic since $E(\varepsilon_t \varepsilon_t') = \Omega \neq$ a diagonal matrix . This means that a shock in one variable is likely to be accompanied by a shock in another variable in the same period.

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.

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}

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 .

In the new MA representation, it is reasonable to assume that a change in one component of u_t has no effect on the other components because the components are orthogonal.

Notice that $\psi_0^* = \psi_0 A = IA$ is lower triangular. This implies that the ordering of variables is of importance. The ordering has to be such that W_{lt} is the only one with a potential immediate impact on all other variables. W_{2t} may have an immediate impact on the last $n-2$ components but not on W_{1t} , and so on. The ordering cannot be determined with statistical methods.

v) 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

$$MSE(\widehat{W}_{t+s|t}) = \Omega + \psi_1 \Omega \psi_1' + \psi_2 \Omega \psi_2' ..... + \psi_{s-1} \Omega \psi_{s-1}'$$

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

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

$$\Omega = a_1 a_{1(n\times n)}' Var(u_{1t}) + a_2 a_{2(n\times n)}' Var(u_{2t}) + ...a_n a_{n(n\times n)}' Var(u_{\rm nt})$$

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

$$MSE(\widehat{W}_{t+s|t}) = \sum_{j=1}^{n} Var(u_{jt})(a_j a_j' + \psi_1 a_j a_j' \psi_1' + \psi_2 a_j a_j' \psi_2' ... + \psi_{s-1} a_j a_j' \psi_{s-1}')$$

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

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

Again the magnitude in general depends on the ordering of the variables

Structural VAR's

Blanchard (1989)

To introduce structural information in a VAR there are several ways to proceed. Probably the most popular is to try to impose restrictions in the covariance matrix. For a VAR(p), there are $p(p+1)/2$ elements in the covariance matrix of the elements of p so we might consider models of the innovations of the form

$C\varepsilon_t = Du_t$ where the u_t's are uncorrelated with variances $\sigma_1^2, ..., \sigma_p^2$. Then, we can have $p(p+1)/2$ - p unknown terms in C and D.

How exactly one would want to specify C and D is left to structural reasoning, for example Blanchard (1989) considers the following structure

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

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. If $e = 0$ output just respond to demand shocks.

Blanchard and Quah (1989).

Probably the most well known approach is Blanchard and Quah (1989). They have a bivariate system with demand and supply shocks but they do not impose the assumption that the shocks are uncorrelated. Rather they argue that a demand shock should have a zero long-run effect while a supply shock will not. Hence they will have

$$\varepsilon_{1t} = a_1 u_{2t} + u_{1t}$$
$$\varepsilon_{2t} = a_2 u_{1t} + u_{2t}$$

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

To see how it works 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

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

In terms of the shocks of interest, we will write $\varepsilon_t = Au_t$ where A is now defined as $A = \begin{bmatrix} 1 & a_1 \\ a_2 & 1 \end{bmatrix}$.

Then the moving average representation becomes,

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

But we want the long run effect of a demand shock, taken to be u_{1t} , upon output say, W_{1t} , to be zero. The long run effect of u_{1t} on W_{1t} is just obtained by summing

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

Let the first row of $\sum_{z=0}^{\infty} \psi_z$, be $[\delta_1, \delta_2]$, then the restriction is just $\delta_1 + a_2\delta_2 = 0$ or $a_2 = -\delta_1/\delta_2$.

Thus one parameter can be found from this restriction and 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. Then all that is needed is to estimate δ_1, δ_2 . Now since we know that

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

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

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

iv) $\psi(1) = (I - A_1 - A_2 - ..... + A_p)^{-1} = A(1)^{-1}$, then all the information that is needed for the impulse response function, is obtained from the estimated parameters in the VAR.