Econometrics for Financial Time Series

Chapter 3: Multiple Time Series Analysis

In Choi

Sogang University

- Reference:
  Chapter 8 of Tsay.
  Kilian, L. and H. Lütkepohl (2017). "Structural Vector Autoregressive Analysis," Cambridge University Press.
  Lütkepohl, H. (1991) "Introduction to Multiple Time Series Analysis," Springer-Verlag: New York.
  Hamilton, J.D. (1994) "Time Series Analysis," Princeton University Press: New York.
  Reinsel, G.C. (1997) "Elements of Multivariate Time Series Analysis," Springer-Verlag: New York.
  Sims, C. A. (1980). Macroeconomics and reality. Econometrica, 1-48.

Weak stationarity and cross-correlation matrix

- Let $r_t = \begin{pmatrix} r_{1t} \\ \vdots \\ r_{Kt} \end{pmatrix}$.
- Mean vector:

$$\mu_t = E(r_t) = \begin{pmatrix} E(r_{1t}) \\ \vdots \\ E(r_{Kt}) \end{pmatrix} = \begin{pmatrix} \mu_{1t} \\ \vdots \\ \mu_{Kt} \end{pmatrix}$$

- Covariance matrices

$$\Gamma_{tl} = Cov(r_t, r_{t-l}) = E[(r_t - \mu_t)(r_{t-l} - \mu_t)'] = [\Gamma_{tij}(l)].$$

Weak stationarity and cross-correlation matrix

- Notice that Γ_{tl} is not a symmetric matrix when $l \neq 0$. When $l = 0$,

$$
\begin{aligned}
\Gamma_{t0} &= E[(r_t - \mu_t)(r_t - \mu_t)'] \\
&= \begin{bmatrix} E[(r_{1t} - \mu_{1t})(r_{1t} - \mu_{1t})] & \cdots & E[(r_{1t} - \mu_{1t})(r_{kt} - \mu_{kt})] \\ \vdots & \ddots & \vdots \\ E[(r_{kt} - \mu_{1t})(r_{1t} - \mu_{1t})] & \cdots & E[(r_{kt} - \mu_{kt})(r_{kt} - \mu_{kt})] \end{bmatrix} \\
&= [\Gamma_{tij}(0)].
\end{aligned}
$$

  The diagonal elements are variances and off-diagonal elements covariances.

- The multivariate time series $\{r_t\}$ is said to be (weakly) stationary if μ_t and Γ_{tl} are independent of the time index t.

Weak stationarity and cross-correlation matrix

- Assume $\{r_t\}$ is stationary. Let

$$D = diag[\sqrt{\Gamma_{11}(0)}, ..., \sqrt{\Gamma_{kk}(0)}].$$

The concurrent cross-correlation matrix (CCM) of r_t is defined as

$$\rho_0 = D^{-1}\Gamma_0 D^{-1} = [\rho_{ij}(0)].$$

The (i,j)th elements of ρ_0 is the correlation between r_{it} and r_{jt}.

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)}\sqrt{\Gamma_{jj}(0)}} = \frac{Cov(r_{it}, r_{jt})}{std(r_{it})std(r_{jt})}.$$

Weak stationarity and cross-correlation matrix

- The lag$-l$ cross-correlation matrix of r_t is defined by

$$\rho_l = D^{-1}\Gamma_l D^{-1} = [\rho_{ij}(l)].$$

$\rho_{ij}(l)$ is the correlation between r_{it} and $r_{j,t-l}$.
Since

$$\begin{aligned}
\Gamma_{ij}(l) &= Cov(r_{it}, r_{j,t-l}) \\
&= Cov(r_{j,t-l}, r_{it}) \\
&= Cov(r_{j,t}, r_{i,t+l}) \text{ (stationarity)} \\
&= Cov(r_{j,t}, r_{i,t-(-l)}) \\
&= \Gamma_{ji}(-l),
\end{aligned}$$

we have

$$\Gamma_l = \Gamma'_{-l}.$$

Weak stationarity and cross-correlation matrix

1. r_{it} and r_{jt} have no linear relationship if $\rho_{ij}(l) = \rho_{ji}(l) = 0$ for all $l \geq 0$.
2. r_{it} and r_{jt} are concurrently correlated if $\rho_{ij}(0) \neq 0$.
3. r_{it} and r_{jt} have no lead-lag relationship if $\rho_{ij}(l) = \rho_{ji}(l) = 0$ for all $l > 0$.
4. There is a unidirectional relationship from r_{it} to r_{jt} if $\rho_{ij}(l) = 0$ for all $l > 0$, but $\rho_{ji}(v) \neq 0$ for some $v > 0$. (r_{jt} depends on some past values of r_{it}).
5. There is a feedback relationship between r_{it} and r_{jt} if $\rho_{ij}(l) \neq 0$ for some $l > 0$ and $\rho_{ji}(v) \neq 0$ for some $v > 0$.

Weak stationarity and cross-correlation matrix

- Sample cross-correlation matrixes

$$\hat{\Gamma}_l = \frac{1}{T}\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})', \ l \geq 0,$$

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t.$$

$$\hat{\rho}_l = \hat{D}^{-1}\hat{\Gamma}_l\hat{D}^{-1}, \ l \ \geq 0.$$

- Multivariate Ljung–Box test

$$Q_K(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr(\hat{\Gamma}_l' \hat{\Gamma}_0^{-1} \hat{\Gamma}_l \hat{\Gamma}_0^{-1}) \sim \chi^2(K^2 m).$$

- VAR(1) model

$$r_t = \phi_0 + \Phi r_{t-1} + a_t,$$

where ϕ_0 a k-dimensional vector, Φ is a $K \times K$ matrix, and $\{a_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix Σ.

- Bivariate case

$$\begin{aligned} r_{1t} &= \phi_{10} + \Phi_{11} r_{1,t-1} + \Phi_{12} r_{2,t-1} + a_{1t} \\ r_{2t} &= \phi_{20} + \Phi_{21} r_{1,t-1} + \Phi_{22} r_{2,t-1} + a_{2t} \end{aligned}$$

  Φ_{12} : linear dependence of r_{1t} on $r_{2,t-1}$ in the presence of $r_{1,t-1}$
  Φ_{21} : linear dependence of r_{2t} on $r_{1,t-1}$ in the presence of $r_{2,t-1}$
  $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$: a unidirectional relationship from r_{1t} to r_{2t}
  $\Phi_{12} = 0$ and $\Phi_{21} = 0$: r_{1t} and r_{2t} are uncoupled.
  $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$: a feedback relationship between r_{1t} and r_{2t}

- The concurrent relationship between r_{1t} and r_{2t} is shown by the off-diagonal element σ_{12} of the covariance matrix Σ.

- There exists a lower triangular matrix L with all of its diagonal elements being equal to one such that $\Sigma = LGL'$ where G is a diagonal matrix.
  Define $b_t = L^{-1}a_t$. Then,

$$E(b_t) = 0, \ Cov(b_t) = L^{-1}\Sigma(L^{-1})' = G$$

and

$$\begin{aligned} L^{-1}r_t &= L^{-1}\phi_0 + L^{-1}\Phi r_{t-1} + b_t \\ &= \phi_0^* + \Phi^* r_{t-1} + b_t. \end{aligned}$$

- The $j-$th equation of this model is

$$r_{jt} + \sum_{i=1}^{j-1} \omega_{ji} r_{it} = \phi^*_{j,0} + \sum_{i=1}^{j} \Phi^*_{ji} r_{i,t-1} + b_{jt},$$

where ω_{ji} are the elements of the j-th row of L. This shows explicitly the concurrent linear dependence of r_{jt} on $r_{1t}, \dots, r_{j-1,t}$.

- Assume that the VAR(1) model is weakly stationary. Since

$$E(r_t) = \phi_0 + \Phi E(r_{t-1}),$$

$$\mu = E(r_t) = (I - \Phi)^{-1}\phi_0.$$

Using $\phi_0 = (I - \Phi)\mu$, write

$$r_t - \mu = \Phi(r_{t-1} - \mu) + a_t$$

or

$$\tilde{r}_t = \Phi\tilde{r}_{t-1} + a_t.$$

Repeated substitutions give

$$\tilde{r}_t = a_t + \Phi a_{t-1} + \Phi^2 a_{t-2} + \ldots.$$

1.

$$Cov(a_t, r_{t-1}) = 0.$$

2.

$$Cov(a_t, r_t) = \Sigma.$$

3. Φ^j must converge to zero as $j \longrightarrow \infty$. Otherwise, Φ^j will either explode or converge to a nonzero matrix as $j \longrightarrow \infty$.

4. For Φ^j to converge to zero as $j \to \infty$, all eigenvalues of Φ should be less than 1 in modulus. In fact, this is the condition for the stationarity of r_t.
5.

$$E(\tilde{r}_t \tilde{r}'_{t-l}) = \Phi E(\tilde{r}_{t-1} \tilde{r}'_{t-l})$$

or

$$\Gamma_l = \Phi \Gamma_{l-1}, \ \ l > 0.$$

This gives

$$\Gamma_l = \Phi^l \Gamma_0, \ \ l > 0.$$

- VAR(p) model

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \dots + \Phi_p r_{t-p} + a_t.$$

Assume that the VAR(p) model is weakly stationary. Since

$$E(r_t) = \phi_0 + \Phi_1 E(r_{t-1}) + \dots + \Phi_p E(r_{t-p}),$$

$$\mu = E(r_t) = (I - \Phi_1 - \dots - \Phi_p)^{-1}\phi_0.$$

Using $\phi_0 = (I - \Phi_1 - \dots - \Phi_p)\mu$, write

$$r_t - \mu = \Phi_1(r_{t-1} - \mu) + \dots + \Phi_p(r_{t-p} - \mu) + a_t$$

or

$$\tilde{r}_t = \Phi_1 \tilde{r}_{t-1} + \dots + \Phi_p \tilde{r}_{t-p} + a_t.$$

1. $$Cov(a_t, r_{t-l}) = 0 \text{ for } l > 0.$$
2. $$Cov(a_t, r_t) = \Sigma.$$
3. $$\Gamma_l = \Phi_1 \Gamma_{l-1} + ... + \Phi_p \Gamma_{l-p}, \;\; l > 0.$$

- The VAR(p) model can be written as the VAR(1) model
  Let

$$x_t = \begin{bmatrix} \tilde{r}_{t-p+1} \\ \tilde{r}_{t-p+2} \\ \vdots \\ \tilde{r}_t \end{bmatrix} \text{ and } b_t = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ a_t \end{bmatrix}.$$

- Then, the VAR(p) model can be written as

$$x_t = \Phi^* x_{t-1} + b_t,$$

where

$$\Phi^* = \begin{bmatrix} 0 & I & 0 & 0 & \cdots & 0 \\ 0 & 0 & I & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & I \\ \Phi_p & \Phi_{p-1} & \Phi_{p-2} & \Phi_{p-3} & \cdots & \Phi_1 \end{bmatrix}.$$

- Note that the last row of Φ^*signifies the VAR(p) model and that the rest are identity relations. This representation tells that if all eigenvalues of Φ^* are less than 1 in modulus, r_t is weakly stationary. But this is equivalent to

$$|I - \Phi_1 z - \cdots - \Phi_p z^p| \neq 0 \text{ for } |z| \leq 1.$$

Estimating the VAR(p) model

- $vec(\cdot)$ operator: Let $\underset{m\times n}{A} = (a_1 \cdots a_n)$. Then,

$$vec(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \, mn\times 1 \ vector$$

Estimating the VAR(p) model

Example

If

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

$$vec(A) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}.$$

Estimating the VAR(p) model

Definition

The Kronecker product

Let

$$\underset{m \times n}{A} = (a_{ij}) \text{ and } \underset{p \times q}{B} = (b_{ij}).$$

The $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

is the Kronecker product of A and B.

Estimating the VAR(p) model

Example

Let

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 7 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 4 & 5 \end{bmatrix}.$$

Then,

$$A \otimes B = \begin{bmatrix} 3[4\ 5] & 2[4\ 5] \\ 1[4\ 5] & 7[4\ 5] \end{bmatrix} = \begin{bmatrix} 12 & 15 & 8 & 10 \\ 4 & 5 & 28 & 35 \end{bmatrix}.$$

Estimating the VAR(p) model

- The following property of the $vec(\cdot)$ operator will be useful.

$$vec(AB) = (B' \otimes I)vec(A).$$

- Write the $VAR(p)$ model

$$r_t = \mu + \Phi_1 r_{t-1} + \cdots + \Phi_p r_{t-p} + a_t$$

as a multivariate linear regression model

$$Y = BW + U$$

Estimating the VAR(p) model

where

$$Y = (r_1, \cdots, r_n)$$

$$B = (\mu, \Phi_1, \cdots, \Phi_p)$$

$$\begin{aligned} W &= (W_0, \cdots, W_{n-1}) \\ U &= (a_1, \cdots, a_n) \end{aligned}$$

and

$$W_t = \begin{bmatrix} \mathbf{1} \\ r_t \\ \vdots \\ r_{t-p+1} \end{bmatrix},$$

where $\mathbf{1} = [1, ..., 1]'$.

Estimating the VAR(p) model

- Using the $vec(\cdot)$ operator, the $VAR(p)$ model can be written compactly as

$$\begin{aligned} vec(Y) &= vec(BW) + vec(U) \\ &= (W' \otimes I)vec(B) + vec(U) \end{aligned}$$

  or

$$y = (W' \otimes I)\beta + u.$$

This is a linear regression model! Thus[1],

$$\hat{\beta} = [(W' \otimes I)'(W \otimes I)]^{-1}(W' \otimes I)'y.$$

[1] Recall that the OLS estimator of β in the linear regression model $y = X\beta + u$ is $\hat{\beta} = (X'X)^{-1}X'y$.

Estimating the VAR(p) model

But

$$\begin{aligned}(A \otimes B)' &= A' \otimes B' \\ (A \otimes B)(C \otimes D) &= AC \otimes BD \\ (A \otimes B)^{-1} &= A^{-1} \otimes B^{-1}.\end{aligned}$$

Thus

$$\begin{aligned}\hat{\beta} &= (WW' \otimes I)^{-1}(W \otimes I)y \\ &= [(WW')^{-1} \otimes I][W \otimes I]y \\ &= [(WW')^{-1}W \otimes I]y.\end{aligned}$$

Estimating the VAR(p) model

- This can be rewritten as

$$vec(\hat{B}) = \hat{\beta} = vec(YW'(WW')^{-1})$$

Thus

$$\hat{B} = YW'(WW')^{-1}.$$

- For the VAR(1) model,

$$\hat{\Phi} = \left(\sum r_t r_{t-1}'\right)\left(\sum r_{t-1} r_{t-1}'\right)^{-1}.$$

- We use information criteria to select the VAR order p.

Granger-causality

- Main idea: If a variable x affects a variable z, the former should help improving the predictions of the latter variables.
- To formalize the idea, let
  - Ω_t : the information set containing all the relevant information in the universe available up to and including period t.
  - $z_t(h \mid \Omega_t)$: the optimal (minimum MSE) h-step predictor of the process z_t at origin t, based on the information in Ω_t.
  - $\Sigma_z(h \mid \Omega_t) = E(z_t(h \mid \Omega_t) - z_{t+h})^2$: the forecast MSE.

Granger-causality

- The process x_t is said to cause z_t in Granger's sense if

$$\Sigma_z(h \mid \Omega_t) < \Sigma_z\left(h \mid \Omega_t \backslash \{x_s \mid s \leq t\}\right)$$

  for at least one $h = 1, 2, \ldots$.
- $\Omega_t \backslash \{x_s \mid s \leq t\}$: all the relevant information in the universe except for the information in the past and present of the x_t process.
- In practice, we use

$$\Omega_t = \{z_s, x_s \mid s \leq t\}$$

  as an information set.

Characterization of 1-step ahead Granger-Causality

- For a stationary VAR process,

$$r_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{t-1} \end{bmatrix} + \ldots$$
$$+ \begin{bmatrix} \Phi_{11,p} & \Phi_{12,p} \\ \Phi_{21,p} & \Phi_{22,p} \end{bmatrix} \begin{bmatrix} z_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

  if $\Phi_{12,i} = 0$ for $i = 1, 2, \ldots,$ x_t does not help predicting z_t.

- Therefore,

$$z_t\left(1 \mid |\{r_s \mid s \leq t\}\right) = z_t\left(1 \mid \{z_s \mid s \leq t\}\right)$$

$$\Leftrightarrow \Phi_{12,i} = 0 \text{ for } i = 1, \ldots, p.$$

Granger noncausality test for stationary VAR

- Consider a stationary VAR model

$$r_t = \begin{pmatrix} z_t \\ x_{1t} \\ x_{2t} \end{pmatrix} \begin{matrix} n \\ m \\ l \end{matrix} = \sum_{i=1}^{p} \begin{bmatrix} \Phi_{11i} & \Phi_{12i} & \Phi_{13i} \\ \Phi_{21i} & \Phi_{22i} & \Phi_{23i} \\ \Phi_{31i} & \Phi_{32i} & \Phi_{33i} \end{bmatrix} \begin{bmatrix} z_{t-i} \\ x_{1(t-i)} \\ x_{2(t-i)} \end{bmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \\ a_{3t} \end{pmatrix}$$

- The null hypothesis that x_{2t} does not Granger-cause z_t at the horizon 1 can be written as

$$H_0 : \Phi_{13i} = 0 \ (i = 1, 2, ..., p).$$

- The Wald test for this null hypothesis is

$$W = vec\left(\hat{\theta}\right)'(s \otimes s_1)\left[\left(s' \otimes s_1'\right)\left[\left(v'v\right)^{-1} \otimes \hat{\Sigma}_a\right](s \otimes s_1)\right]^{-1} \times \left(s' \otimes s_1'\right) vec\left(\hat{\theta}\right)$$

where

$$s_1 = \begin{bmatrix} I_n \\ 0 \end{bmatrix}_{m+l},$$

$$s = I_p \otimes s_3 \text{ with } s_3 = \begin{bmatrix} 0 \\ I_l \end{bmatrix}_{n+m}$$

$$\hat{\theta} = \left(\sum_{t=1}^{T} r_t v_t'\right)\left(\sum_{t=1}^{T} v_t v_t'\right)^{-1}, v_t = \left[r_{t-1}', \ldots, r_{t-p}'\right]',$$

$$v = \left[v_1, \ldots, v_T\right]' \ \& \ \hat{\Sigma}_a = \frac{1}{T}\sum_{t=1}^{T}\left(r_t - \hat{\theta}v_t\right)\left(r_t - \hat{\theta}v_t\right)'.$$

Granger noncausality test for stationary VAR

- As $T \to \infty$,

$$W \xrightarrow{d} \chi^2_{nlp}.$$

Impulse response function

- A stationary VAR(p) model $r_t = \mu + \Phi_1 r_{t-1} + \cdots + \Phi_p r_{t-p} + a_t$ can be written as

$$r_t = \mu' + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + ...$$

where the coefficient matrices $\{\Psi_i\}$ satisfy the relation

$$(I - \Phi_1 z - \Phi_1 z^2 - ... - \Phi_p z^p)(I + \Psi_1 z + \Psi_2 z^2 + ...) = I.$$

Impulse response function

- The matrix Ψ_s has the interpretation

$$\frac{\partial r_{t+s}}{\partial a_t'} = \Psi_s.$$

  Namely, $[\Psi_s]_{ij}$ denotes the effect of a one unit increase in a_{jt} on the value of $r_{t+s,i}$.
- A plot of $[\Psi_s]_{ij}$ as a function of s is called the impulse response function. It describes the response of $r_{t+s,i}$ to a one-time impulse in r_{tj} with all other variables dated t or earlier held constant. $\left([\Psi_s]_{ij} = \frac{\partial r_{t+s,i}}{\partial a_{t,j}}\right)$.

Impulse response function

- When all other variables dated t or earlier are held constant,

$$[\Psi_s]_{ij} = \frac{\partial r_{t+s,i}}{\partial a_{t,j}} = \frac{\partial r_{t+s,i}}{\partial r_{t,j}} \frac{\partial r_{t,j}}{\partial a_{t,j}} = \frac{\partial r_{t+s,i}}{\partial r_{t,j}}.$$

- But if $a_{t,j}$ and $a_{t,k}$ $(j \neq k)$ are correlated, $[\Psi_s]_{ij}$ does not capture the effect of $a_{t,j}$ on $r_{t+s,i}$ correctly since $a_{t,k}$ would also affect $r_{t+s,i}$ indirectly. That is,

$$\begin{tikzcd}
a_{t,j} \arrow[dr] \arrow[dd] & \\
& r_{t+s,i} \\
a_{t,k} \arrow[ur] &
\end{tikzcd}$$

Orthogonalized impulse response function

- Consider a decomposition of $\Sigma = E(a_t a_t')$

$$\Sigma = LGL'$$

where L is a lower triangular matrix with its diagonal elements being equal to one and G a diagonal matrix.

- Rewrite the original MA(∞) model such that

$$\begin{aligned} r_t &= \mu' + LL^{-1}a_t + \Psi_1 LL^{-1}a_{t-1} + \Psi_2 LL^{-1}a_{t-2} + \ldots \\ &= \mu' + \Psi_0^* b_t + \Psi_1^* b_{t-1} + \Psi_2^* b_{t-2} + \ldots \end{aligned}$$

Then,

$$E(b_t b_t') = E(L^{-1}a_t a_t L^{'-1}) = L^{-1}\Sigma_a L^{'-1} = L^{-1}LGL'L^{'-1} = G.$$

That is, the variance-covaraince matrix of b_t is diagonal. Thus, $[\Psi_s^*]_{ij}$ measure the effect of $a_{t,j}$ on $r_{t+s,i}$ correctly.

Orthogonalized impulse response function

- The plot of $[\Psi_s^*]_{ij}$ as a function of s is called the orthogonalized impulse response function.

Example

$r_t = \begin{pmatrix} \text{\# of Hyundai cars sold in the US} \\ \text{\# of Nissan, Honda, Toyota cars sold in the US} \end{pmatrix}$. The orthogonalized impulse response function $[\Psi_s^*]_{12}$ shows how the sales of Nissan, Honda, Toyota cars affect those of Hyundai cars over time.

Orthogonalized impulse response function

- A major drawback of the orthogonalized impulse response function is that it depends on the ordering of the variables involved. The orthogonalized impulse response function changes as the ordering changes.
- When K is large, trying every ordering is practically difficult.
- Even when the results are robust to different orderings, it does not mean that the recursive system is correct.

Orthogonalized impulse response function

- The reason for this is that L and Ψ change as the ordering changes.
- Consider the simple case $K = 3$ and calculate $[\Psi_s^*]_{12}$ for the original and changed orderings. Note that

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \sigma_{21}\sigma_{11}^{-1} & 1 & 0 \\ \sigma_{31}\sigma_{11}^{-1} & h_{32}h_{22}^{-1} & 1 \end{bmatrix}$$

where $h_{22} = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$, $h_{32} = \sigma_{32} - \sigma_{21}\sigma_{11}^{-1}\sigma_{13}$ and $\Sigma = [\sigma_{ij}]$ (cf. Hamilton, 1994, p.91).

Generalized impulse response function

See Pesaran, H.H. and Y. Shin (1998) "Generalized impulse response analysis in linear multivariate models," Economics Letters, 58, 17-29.

- Write

$$
\begin{aligned}
\frac{dr_{t+s,i}}{da_{t,j}} &= \frac{\partial r_{t+s,i}}{\partial a_{t,1}} \frac{\partial a_{t,1}}{\partial a_{t,j}} + ... + \frac{\partial r_{t+s,i}}{\partial a_{t,K}} \frac{\partial a_{t,K}}{\partial a_{t,j}} \\
&= \sum_{m=1}^{K} \frac{\partial r_{t+s,i}}{\partial a_{t,m}} \frac{\partial a_{t,m}}{\partial a_{t,j}} \\
&= \sum_{m=1}^{K} [\Psi_s]_{im} \frac{\partial a_{t,m}}{\partial a_{t,j}}.
\end{aligned}
$$

Generalized impulse response function

- Assume

$$\begin{aligned} a_{t,m} &= \delta_{m,j} a_{t,j} + \varepsilon_{t,m,j}, \ \varepsilon_{t,m,j} \sim iid(0, \sigma_\varepsilon^2), \\ & \{\varepsilon_t\} \text{ and } \{a_t\} \text{ are independent.} \end{aligned}$$

Then, since $E(a_{t,m} a_{t,j}) = \sigma_{mj}$,

$$E(a_{t,m} a_{t,j}) = \delta_{m,j} Var(a_{t,j})$$

which gives

$$\delta_{m,j} = \frac{\sigma_{mj}}{\sigma_{jj}}.$$

Generalized impulse response function

- Since $\frac{\partial a_{t,m}}{\partial a_{t,j}} = \delta_{m,j}$, the generalized impulse response function can be written as
$$\sum_{m=1}^{K} [\Psi_s]_{im} \frac{\sigma_{mj}}{\sigma_{jj}}.$$
The parameter $\frac{\sigma_{mj}}{\sigma_{jj}}$ can be estimated by using the sample variance-covariance matrix from the VAR analysis.
- Some authors prefer using
$$\frac{\partial r_{t+s,i}}{\partial (a_{t,j}/\sqrt{\sigma_{jj}})}.$$
This denotes the change in $r_{t+s,i}$ per one standard deviation change in $a_{t,j}$.

Generalized impulse response function

- The scaled generalized impulse response function is written as

$$\sum_{m=1}^{K} [\Psi_s]_{im} \frac{\sigma_{mj}}{\sqrt{\sigma_{jj}}}.$$

Forecast error variance decomposition

- Suppose that $\{r_t\}$ is a $K \times 1$ vector linear process written as

$$\begin{aligned} r_t &= \mu + \sum_{i=0}^{\infty} \Psi_i P P^{-1} a_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \end{aligned}$$

  where $\Theta_i = \Phi_i P$, $w_t = P^{-1} a_t$ and $E(w_t w_t') = I$ for all t.
- If $\Sigma > 0$, we can find P^{-1} such that $P^{-1} \Sigma P^{-1\prime} = I$.

Forecast error variance decomposition

- Assume $E(a_{t+h} \mid r_t, r_{t-1}, \ldots) = 0$ for $h > 0$.
- The optimal h-step forecast is

$$r_t(h) = E(r_{t+h} \mid r_t, r_{t-1}, \ldots) = \mu + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i}.$$

The forecast error is

$$r_{t+h} - r_t(h) = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}.$$

- The mn-th element of Θ_i is denoted as $\theta_{mn,i}$, and the h-step forecast error of the j-th component of r_t is

$$\begin{aligned} r_{j,t+h} - r_{j,t}(h) &= \sum_{i=0}^{h-1} \left(\theta_{j1,i} w_{1,t+h-i} + \ldots + \theta_{jK,i} w_{K,t+h-i}\right) \\ &= \sum_{k=1}^{K} \left(\theta_{jk,0} w_{k,t+h} + \ldots + \theta_{jk,h-1} w_{k,t+1}\right). \end{aligned}$$

Forecast error variance decomposition

- The MSE of the forecast error is

$$E(r_{j,t+h} - r_{j,t}(h))^2 = \sum_{k=1}^{K} \left(\theta^2_{jk,0} + ... + \theta^2_{jk,h-1}\right).$$

  Here $\theta^2_{jk,0} + ... + \theta^2_{jk,h-1}$ is the contribution of the k-the variable to the MSE.

- The quantity $\omega_{jk,h} = \left(\theta^2_{jk,0} + ... + \theta^2_{jk,h-1}\right) / \sum_{k=1}^{K} \left(\theta^2_{jk,0} + ... + \theta^2_{jk,h-1}\right)$ is the proportion of the h-step forecast error variance of variable j accounted for by the k-th variable. The quantities $\{\omega_{jk,h}\}$ constitute the forecast error variance decomposition.