Multivariate Time Series Models

The real data Z_t is a vector:

$$Z_t = \begin{bmatrix} Z_{1t} \\ Z_{2t} \\ \vdots \\ Z_{kt} \end{bmatrix}.$$

Z_t is called an k-dimensional vector time series.

1. Covariance and Correlation Matrix Functions

A. Mean

$$EZ_t = \begin{bmatrix} EZ_{1t} \\ EZ_{2t} \\ \vdots \\ EZ_{kt} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{kt} \end{bmatrix} \equiv \mu.$$

B. Covariance

$$\begin{aligned} \Gamma(0) &= \text{Cov}(Z_t, Z_t) = E[(Z_t - \mu)(Z_t - \mu)'] \\ &= \begin{bmatrix} \gamma_{11}(0) & \gamma_{12}(0) & \cdots & \gamma_{1k}(0) \\ \gamma_{21}(0) & \gamma_{22}(0) & \cdots & \gamma_{2k}(0) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{k1}(0) & \gamma_{k2}(0) & \cdots & \gamma_{kk}(0) \end{bmatrix}. \end{aligned}$$

where $\gamma_{ii}(0) = E[(Z_{it} - \mu_i)^2]$ and $\gamma_{ij}(0) = E[(Z_{it} - \mu_i)(Z_{jt} - \mu_j)]$.

C. Cross-Correlation Matrices

Let

$$D = diag\{\sqrt{\gamma_{11}(0)}, \ldots, \sqrt{\gamma_{kk}(0)}\}.$$

The *concurrent*, or lag-zero, *cross-correlation matrix* is defined as

$$\begin{aligned} \rho(0) \equiv [\rho_{ij}(0)] &= D^{-1}\Gamma(0)D^{-1} \\ &= \begin{bmatrix} \rho_{11}(0) & \rho_{12}(0) & \cdots & \rho_{1k}(0) \\ \rho_{21}(0) & \rho_{22}(0) & \cdots & \rho_{2k}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1}(0) & \rho_{k2}(0) & \cdots & \rho_{kk}(0) \end{bmatrix}. \end{aligned}$$

where

$$\rho_{ij}(0) = \text{correlation of } Z_{it} \text{ and } Z_{jt} = \frac{\gamma_{ij}(0)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}.$$

$$-1 \leq \rho_{ij}(0) \leq 1,$$

$$\text{and} \quad \rho_{ii}(0) = 1$$

for $1 \leq i, j \leq k$.

Thus, $\rho(0)$ is a symmetric matrix with unit *diagonal* elements,

$$\rho(0) = \begin{bmatrix} 1 & \rho_{21}(0) & \cdots & \rho_{k1}(0) \\ \rho_{21}(0) & 1 & \cdots & \rho_{k2}(0) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1}(0) & \rho_2(0) & \cdots & 1 \end{bmatrix}.$$

D. The lag-l *cross-correlation matrix* (CCM)

$$\begin{aligned} \Gamma(l) &= \text{Cov}(Z_t, Z_{t-l}) = E[(Z_t - \mu)(Z_{t-l} - \mu)'] \\ &= \begin{bmatrix} Cov(Z_{1t}, Z_{1t-l}) & \cdots & Cov(Z_{1t}, Z_{kt-l}) \\ Cov(Z_{2t}, Z_{1t-l}) & \cdots & Cov(Z_{2t}, Z_{kt-l}) \\ \vdots & \ddots & \vdots \\ Cov(Z_{kt}, Z_{1t-l}) & \cdots & Cov(Z_{kt}, Z_{kt-l}) \end{bmatrix} \\ &= \begin{bmatrix} \gamma_{11}(l) & \gamma_{12}(l) & \cdots & \gamma_{1k}(l) \\ \gamma_{21}(l) & \gamma_{22}(l) & \cdots & \gamma_{2k}(l) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{k1}(l) & \gamma_{k2}(l) & \cdots & \gamma_{kk}(l) \end{bmatrix}. \end{aligned}$$

where $l \neq 0$ and $\gamma_{ij}(l) = E[(Z_{it} - \mu_i)(Z_{jt-l} - \mu_j)]$.

The lag-l *cross-correlation matrix* (CCM)

$$\rho(l) = D^{-1}\Gamma(l)D^{-1} = [\rho_{ij}(l)]_{k\times k},$$

where

$$\rho_{ij}(l) = \frac{\gamma_{ij}(l)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}$$

E. Properties

$$\begin{aligned} \Gamma(l) &= \Gamma'(-l) \geq 0 \\ \rho(l) &= \rho'(-l) \geq 0 \text{ (positive semidefinite).} \end{aligned}$$

Important case: when $\rho(l) = 0$ for all $l > 0$,

Z_t is called the l-dimensional white noise, denoted by a_t.

$$E[a_t a'_{t-l}] = \begin{cases} \Sigma > 0, & \text{if } l = 0 \\ 0, & \text{if } l \neq 0. \end{cases}$$

2. Sample Cross-Correlation Matrices

Given the data $\{Z_t\}_{t=1}^T$, the *cross-covariance matrix* $\Gamma(l)$ can be estimated by

$$\widehat{\Gamma}(l) = \frac{1}{T}\sum_{t=1}^{T-l}(Z_t - \bar{Z})(Z_{t-l} - \bar{Z})', \quad l \geq 0,$$

where $\bar{Z}$ is the vector of *sample means* given by

$$\bar{Z} = \frac{1}{T}\sum_{t=1}^{T} Z_t.$$

The *cross-correlation matrix* $\rho(l)$ is estimated by

$$\widehat{\rho}(l) = \widehat{D}^{-1}\widehat{\Gamma}(l)\widehat{D}^{-1}, \quad l \geq 0, \tag{1}$$

where

$$\widehat{D} = \mathbf{diag}\{\widehat{\gamma}_{11}^{1/2}(0), \ldots, \widehat{\gamma}_{kk}^{1/2}(0)\}$$

is the $k \times k$ *diagonal* matrix of the *sample standard deviations* of the component series.

Multivariate Portmanteau Tests

The univariate Ljung−Box statistic $Q(m)$ has been generalized to the multivariate case by Hosking (1980, 1981) and Li and McLeod (1981).

The null hypothesis is

$$H_0 : \rho(1) = \cdots = \rho(m) = 0,$$

and the alternative hypothesis $H_a : \rho(i) \neq 0$ for some $i \in \{1, \ldots, m\}$. Thus, the statistic is used to test that there are *no auto- and cross-correlations* in the vector series $\{Z_t\}$.

The test statistic assumes the form

$$\begin{aligned} Q_k(m) &= T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr(\hat{\Gamma}'(l)\hat{\Gamma}^{-1}(0)\hat{\Gamma}(l)\hat{\Gamma}^{-1}(0)) \\ &\sim \chi^2(k^2 m), \end{aligned}$$

where T is the sample size, m is the dimension of Z_t, and $tr(A)$ is the trace of the matrix A, which is the sum of the diagonal elements of A.

3. Moving average and autoregressive representation of vector processes

A. MA(∞) model

$$Z_t = \sum_{s=1}^{\infty} \Psi_s a_{t-s} + a_t,$$

where $\Psi_s = (\Psi_{ij,\,s})_{k\times k}$ and $\sum_{s=0}^{\infty} |\Psi_{ij,\,s}|^2 < \infty$.

B. AR(∞) model

$$Z_t = \sum_{s=1}^{\infty} \Pi_s Z_{t-s} + a_t,$$

where $\Pi_s = (\Pi_{ij,\,s})_{k\times k}$ and $\sum_{s=0}^{\infty} |\Pi_{ij,\,s}|^2 < \infty$.

4. Vector AR model

The vector AR(1) model

A. Model

$$(I - \Phi_1 B) Z_t = a_t, \quad \text{or } Z_t = \Phi_1 Z_{t-1} + a_t.$$

when $k = 2$,

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

or

$$\begin{cases} Z_{1t} &= \phi_{11} Z_{1,t-1} + \phi_{12} Z_{2,t-1} + a_{1t}, \\ Z_{2t} &= \phi_{21} Z_{1,t-1} + \phi_{22} Z_{2,t-1} + a_{2t}. \end{cases}$$

B. Stationarity conditions

All the roots of the determinant $|I - \Phi_1 z| = 0$ lie outside the unite circle.

In this case,

$$(I - \Phi_1 B)^{-1} = I + \Phi_1 B + \Phi_1^2 B^2 + \cdots$$

and

$$Z_t = a_t + \Phi_1 a_{t-1} + \Phi_1^2 a_{t-2} + \cdots .$$

C. Covariance matrix function

$$\Gamma(l) = \begin{cases} \Gamma(1)\Phi_1' + \Sigma, & \text{if } l = 0 \\ \Phi_1\Gamma(l-1) = \Phi_1^l\Gamma(0), & \text{if } l \geq 1. \end{cases}$$

The vector AR(p) model

A. Model

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \cdots + \Phi_p Z_{t-p} + a_t.$$

B. Stationarity conditions

All the roots of the determinant $|I - \Phi_1 z - \cdots - \Phi_p z^p| = 0$ lie outside the unite circle.

In this case,

$$Z_t = a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots .$$

where Ψ_s satisfies the matrix equation:

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

The covariance matrix function can be found by direct calculation, although it is quite complicated.

5. Fitting VAR(p) Model

Assume that $\{Z_1, \cdots, Z_T\}$ are from the k-dimensional VAR(p) model:

$$Z_t = \Phi_0 + \Phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + a_t,$$

where $a_t \sim IIDN(0, \Sigma)$. We could rewrite it more concisely as

$$Z_t = \Pi' X_t + a_t,$$

where

$$\underset{k\times(kp+1)}{\Pi'} \equiv \begin{bmatrix} \Phi_0 & \Phi_1 & \Phi_2 & \cdots & \Phi_p \end{bmatrix}$$

$$\underset{(kp+1)\times 1}{X_t} \equiv \begin{bmatrix} 1 \\ Z_{t-1} \\ Z_{t-2} \\ \vdots \\ Z_{t-p} \end{bmatrix}.$$

Let $\theta \equiv (\Phi_0 \quad \Phi_1 \quad \Phi_2 \quad \cdots \quad \Phi_p \quad \Sigma)$. The conditional log-likelihood function

$$\begin{aligned}
&\prod_{t=1}^{T} f(Z_t | Z_{t-1}, Z_{t-2}, \ldots, Z_{-p+1}; \theta) \\
&= \prod_{t=1}^{T} 2\pi)^{-k/2} |\Sigma|^{-1/2} \\
&\quad \times \exp\left\{ -\frac{1}{2} (Z_t - \Pi' X_t)' \Sigma^{-1} (Z_t - \Pi' X_t) \right\}.
\end{aligned}$$

Then the log-likelihood function is *(given that* Z_0 $, \ldots, Z_{1-p}$ *is observed)*

$$\begin{aligned}
\ln L(\theta) &= \sum_{t=1}^{T} \ln f(Z_t | Z_{t-1}, Z_{t-2}, \ldots, Z_{-p+1}; \theta) \\
&= -\frac{Tk}{2} \ln(2\pi) - \frac{T}{2} \ln |\Sigma| \\
&\quad - \frac{1}{2} \sum_{t=1}^{T} \left[(Z_t - \Pi' X_t)' \Sigma^{-1} (Z_t - \Pi' X_t) \right].
\end{aligned}$$

Taking first derivative with respect to Π and Σ, we have that

$$\widehat{\Pi}' = \left(\sum_{t=1}^{T} Z_t X_t' \right) \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1}.$$

Here, $\widehat{\Pi}$ is the MLE of Π. The MLE of Σ is

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \widehat{a}_t \widehat{a}_t' = \begin{bmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} & \cdots & \widehat{\sigma}_{1n} \\ \widehat{\sigma}_{21} & \widehat{\sigma}_2^2 & \cdots & \widehat{\sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\sigma}_{n1} & \widehat{\sigma}_{n2} & \cdots & \widehat{\sigma}_n^2 \end{bmatrix},$$

where $\widehat{a}_t' = Z_t - \widehat{\Pi}' X_t, \widehat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{a}_{it}^2$ and $\widehat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \widehat{a}_{it}\widehat{a}_{jt}$ for $i, j = 1, 2, \ldots, n$.

Denote

$$S_k(\theta) = \sum_{t=1}^{T} \left[(Z_t - \Pi' X_t)' \Sigma^{-1} (Z_t - \Pi' X_t)\right].$$

The minimizer of $S_k(\theta)$ is called the general LSE of Π, which is the same as the MLE. Denote

$$\begin{aligned} S_k(\theta) = & \sum_{t=1}^{T} \left[(Z_t - \Pi' X_t)'(Z_t - \Pi' X_t)\right] \\ = & \sum_{t=1}^{T} \|Z_t - \Pi' X_t\|^2. \end{aligned}$$

The minimizer of $S_k(\theta)$ is called the ordinary LSE of Π, which is the same as the MLE.

Math: The chain rule and product rules of vector differentiation

$$(a) \qquad \frac{\partial[\alpha(\beta)' A c(\beta)]}{\partial \beta'} = c(\beta)' A' \frac{\partial[\alpha(\beta)]}{\partial \beta'} + \alpha(\beta)' A \frac{\partial[c(\beta)]}{\partial \beta'}.$$

$$(b) \qquad \frac{\partial AB}{\partial \beta} = \frac{\partial A}{\partial \beta} B + A \frac{\partial B}{\partial \beta}.$$

$$\alpha(\beta) = c(\beta) = Z_t - \Pi' X_t = Z_t - (X_t' \otimes I_k) vec(\Pi').$$

$$\frac{\partial[\alpha(\beta)]}{\partial \beta'} = -\frac{\partial (X_t' \otimes I_k) vec(\Pi')}{\partial vec'(\Pi')} = -(X_t' \otimes I_k)$$

Thus,

$$\frac{\partial S_k(\theta)}{\partial vec'(\Pi')} = -2 \sum_{t=1}^{T} [Z_t - (X_t' \otimes I_k) vec(\Pi')]' \Sigma^{-1} (X_t' \otimes I_k).$$

Let

$$\sum_{t=1}^{T} [Z_t - (X_t' \otimes I_k) vec(\Pi')]' \Sigma^{-1} (X_t' \otimes I_k) = 0.$$

Then

$$\begin{aligned}
&\sum_{t=1}^{T} Z_t' \Sigma^{-1}(X_t' \otimes I_k) \\
&= \sum_{t=1}^{T} vec'(\Pi')(X_t \otimes I_k)\Sigma^{-1}(X_t' \otimes I_k) \\
&= vec'(\Pi')(\sum_{t=1}^{T} X_t X_t' \otimes \Sigma^{-1})
\end{aligned}$$

Taking transposition,

$$\sum_{t=1}^{T}(X_t \otimes I_k)\Sigma^{-1} Z_t = (\sum_{t=1}^{T} X_t X_t' \otimes \Sigma^{-1}) vec(\Pi')$$

$$\begin{aligned}
vec(\Sigma^{-1}\sum_{t=1}^{T} Z_t X_t') &= (\sum_{t=1}^{T} X_t X_t' \otimes \Sigma^{-1}) vec(\Pi') \\
&= vec(\Sigma^{-1}\Pi' \sum_{t=1}^{T} X_t X_t').
\end{aligned}$$

$$\sum_{t=1}^{T} Z_t X_t' = \Pi' \sum_{t=1}^{T} X_t X_t'.$$

Thus, the minimizer (LSE) is

$$\hat{\Pi}' = (\sum_{t=1}^{T} Z_t X_t')(\sum_{t=1}^{T} X_t X_t')^{-1}.$$

Lag p Length Selection

The general approach is fit VAR(p) models with orders $p = 0, 1, \ldots, p_{\text{max}}$ and choose the value of p which minimizes some model selection criteria.

The three most common information criteria are the Akaike (AIC), Schwarz-Bayesian (BIC) and Hannan-Quinn (HQ):

$$\text{AIC}(p) = \ln|\widehat{\Sigma}(p)| + \frac{2}{T}pk^2,$$
$$\text{BIC}(p) = \ln|\widehat{\Sigma}(p)| + \frac{\ln(T)}{T}pk^2,$$
$$\text{HQ}(p) = \ln|\widehat{\Sigma}(p)| + \frac{2\ln\ln(T)}{T}pk^2,$$

where $\widehat{\Sigma}(p) = T^{-1}\sum_{t=1}^{T}\hat{a}_t\hat{a}_t'$ is the residual covariance matrix without a degrees of freedom correction from a VAR(p) model.

Sequential Likelihood Ratio Tests

$$H_0 : \Phi_l = 0 \; v.s. \; H_a : \phi_l \neq 0.$$

Test statistic:

$$M(l) = (T - l - 1.5 - kl)\log\frac{|\hat{\Sigma}_{a,l}|}{|\hat{\Sigma}_{a,l-1}|},$$

where $\hat{\Sigma}_{a,l}$ is the MLE of Σ_a based the residuals of MLE of AR(p) model.

6. Building a VAR(p) Model

- Model Selection: Use the AIC or its variants to select the order p.

- Estimation: For a specified VAR model, one can estimate the parameters using either the OLS method or the ML method.

- Model Checking:

$$H_0 : R_1 = \cdots = R_m = 0$$

$$v.s.\ H_a : R_j \neq 0\ for\ some\ j,$$

where R_j is the CCM of a_t. If model is correct, then the CCM of residual $\widehat{a}_t$ should be close to R_l, i.e. accept H_0. Test statistic:

$$Q_k(m) = T^2 \sum_{l=1}^{m} \frac{1}{T-l} tr(\widehat{C}_l' \widehat{C}_0^{-1} \widehat{C}_l \widehat{C}_0^{-1}),$$

where $\widehat{C}_l = \frac{1}{T-l} \sum_{t=l+1}^{T} \widehat{a}_t \widehat{a}_{t-l}'$. Then, the test statistic is asymptotically distributed as a chi-square distribution with with $mk^2 - g$ degrees of freedom, g is the number of parameters estimated.

- Refine AR(p)Mode–Testing Zero Parameters

$$H_0 : \omega = 0 \ v.s. \ H_a : \omega \neq 0,$$

where ω is a $\nu-$dimensional vector consisting of some elements from β. Let $\widehat{\omega}$ be the MLE of ω. Then

$$(\widehat{\omega} - 0)' \widehat{\Omega}^{-1} (\widehat{\omega} - 0) \sim \chi^2_{\nu},$$

where $\widehat{\Omega}$ is covariance matrix of $\widehat{\omega}$. This is called Wald test.

When $\nu = 1$, we used type I error $\alpha = 0.05$ and $\alpha = 0.1$, respectively, to identify target parameters (i.e. not zero parameter). The corresponding critical values are 1.96 and 1.645, respectively.

7. Forecasting Based on VAR Model

Given data $F_h = \{Z_1, \cdots, Z_h\}$, the best linear predictor, in terms of minimum mean squared error (MSE), of Z_{h+1} or 1-step forecast based on information available at time h is

$$\hat{Z}_h(1) = E(Z_{h+1}|F_h) = \Phi_0 + \Phi_1 Z_h + \cdots + \Phi_p Z_{h-p+1}.$$

Forecasts for longer horizons l (l-step forecasts) may be obtained using the chain-rule of forecasting as

$$\begin{aligned} \hat{Z}_h(l) &= E(Z_{h+l}|F_h) \\ &= \Phi_0 + \Phi_1 \hat{Z}_n(l-1) + \cdots + \Phi_p \hat{Z}_n(l-p), \end{aligned}$$

where $\hat{Z}_h(j) = Z_{h+j}$ if $j < 0$.

The l-step forecast errors may be expressed as

$$\hat{e}_h(l) = Z_{h+l} - \hat{Z}_h(l) = \sum_{s=0}^{l-1} \Psi_s a_{h+l-s}$$

where the matrices Ψ_s are determined by recursive substitution

$$\Psi_s = \Phi_1 \Psi_{s-1} + \Phi_2 \Psi_{s-2} + \cdots + \Phi_p \Psi_{s-p},$$

with $\Psi_0 = I_k$ and $\Psi_j = 0$ for $j < 0$.

The forecasts are unbiased since all of the forecast errors have expectation zero and the MSE matrix of $\hat{e}_h(l)$ is

$$\Sigma(l) = \text{COV}(\hat{e}_h(l)) = \sum_{s=0}^{l-1} \Psi_s \Sigma \Psi_s'.$$

Asymptotic $(1-\alpha)$ C.I.s for the individual elements of $\hat{Z}_h(l)$ are

$$\left[\hat{Z}_{ih}(l) - z_{\alpha/2}\hat{\sigma}_i(l), \hat{Z}_{ih}(l) + z_{\alpha/2}\hat{\sigma}_i(l)\right],$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}_i(l)$ denotes the square root of the diagonal element of $\widehat{\Sigma}(l)$, called the standard errors of prediction.

When Φ_i is estimated by $\hat{\Phi}_i$, we need to take into account of estimator effect.

$$\begin{aligned} \tilde{Z}_h(l) &= E(Z_{h+l}|F_h) \\ &= \hat{\Phi}_0 + \hat{\Phi}_1 \tilde{Z}_n(l-1) + \cdots + \hat{\Phi}_p \tilde{Z}_n(l-p), \end{aligned}$$

where $\hat{Z}_h(j) = Z_{h+j}$ if $j < 0$. The l-step forecast errors may be expressed as

$$\begin{aligned} \tilde{e}_h(l) &= Z_{h+l} - \tilde{Z}_h(l) \\ &= Z_{h+l} - \hat{Z}_h(l) + \hat{Z}_{h+l} - \tilde{Z}_h(l) \\ &= \hat{e}_h(l) + \hat{Z}_{h+l} - \tilde{Z}_h(l). \end{aligned}$$

We can show that $\sqrt{T-p}\tilde{e}_h(l) \sim N(0, \Omega_l)$. The square root of the diagonal element of Ω_l is called the root mean squared errors of prediction.

Example 2.3. Consider the quarterly growth rates, in percentages, of real gross domestic product (GDP) of United Kingdom, Canada, and United States from the second quarter of 1980 to the second quarter of 2011. The data were seasonally adjusted and downloaded from the database of Federal Reserve Bank at St. Louis. The GDP were in millions of local currency, and the growth rate denotes the differenced series of log GDP

$$
\begin{aligned}
Z_t \ &= \ \begin{bmatrix} 0.16 \\ - \\ 0.28 \end{bmatrix} + \begin{bmatrix} 0.47 & 0.21 & - \\ 0.33 & 0.27 & 0.50 \\ 0.47 & 0.23 & 0.23 \end{bmatrix} Z_{t-1} \\
&\quad + \begin{bmatrix} - & - & - \\ -0.20 & - & - \\ -0.30 & - & - \end{bmatrix} Z_{t-2} + a_t.
\end{aligned}
$$

$$
\hat{\Sigma}_a = \begin{bmatrix} 0.29 & 0.02 & 0.07 \\ 0.02 & 0.31 & 0.15 \\ 0.07 & 0.15 & 0.36 \end{bmatrix}.
$$

8. Impulse Response Function

We know that a VAR model has the representation:

$$Z_t = a_t + \Psi_1 a_{t-1} + \cdots \Psi_i a_{t-i} + \cdots .$$

In practice, we have

$$Z_n = a_n + \Psi_1 a_{n-1} + \cdots \Psi_{n-1} a_1 + \Psi_n a_0.$$

Thus, Ψ_n presents the effect of shock a_0 to Z_n (i.e. the stock in n days). If $a_0 = (1, 0 \cdots, 0)'$, then it is the effect of a unit shock of the first component to a stock in n days.

$$\underline{\Psi}_n = \sum_{i=1}^{n} \Psi_i$$

is called the accumulated responses over periods to a unit shock to Z_t.

The total accumulated responses for all future periods are defined as

$$\underline{\Psi}_\infty = \sum_{i=1}^{\infty} \Psi_i$$

is called the total multipliers or long-run effects.

9. Vector MA Model

VMA(1) model

A. Model

$$Z_t = (I - \Theta_1 B) a_t, \quad \text{or } Z_t = a_t - \Theta_1 a_{t-1}.$$

When $k = 2$,

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

B. Invertibility conditions

All the roots of the determinant $|I - \Theta_1 z| = 0$ lie outside the unit circle.

In this case,

$$(I - \Theta_1 B)^{-1} = I + \Theta_1 B + \Theta_1^2 B^2 + \cdots$$

$$a_t = Z_t + \Theta_1 Z_{t-1} + \Theta_1^2 Z_{t-2} + \cdots .$$

C. Covariance matrix function

$$\Gamma(0) \;=\; \Sigma + \Theta_1 \Sigma \Theta_1'.$$

$$\Gamma(k) \;=\; \begin{cases} -\Sigma\Theta_1', & \text{if } k = -1 \\ -\Theta_1\Sigma, & \text{if } k = 1. \\ 0, & \text{if } |k| > 1. \end{cases}$$

VMA(q) model

A. Model

$$Z_t = a_t - \Theta_1 a_{t-1} - \cdots - \Theta_q a_{t-q}.$$

B. Invertibility conditions

All the roots of $|I - \Theta_1 z - \cdots - \Theta_q z^q| = 0$ lie outside the unit circle. In this case,

$$\begin{aligned} a_t &= Z_t - \Pi_1 Z_{t-1} - \Pi_2 Z_{t-2} - \cdots, \\ &\text{or} \\ Z_t &= \Pi_1 Z_{t-1} + \Pi_2 Z_{t-2} + \cdots + a_t, \end{aligned}$$

where Π_s satisfies the matrix equation:

$$(I - \Theta_1 z - \cdots - \Theta_q z^q)(I - \Pi_1 z - \Pi_2 z^2 - \cdots) = I.$$

The covariance matrix function can be found by direct calculation.

10. Specifying q of VMA Model

The order of a VMA process can be easily identified via the cross-correlation matrices. For a VAM(q) model, the cross-correlation matrices satisfy $\rho_j = 0$ for $j \geq q + 1$. Therefore, for a given j, one can consider the null hypothesis

$$H_0 : \rho_j = \rho_{j+1} = \cdots = \rho_m = 0 \; v.s. \; H_a : \rho_i \neq 0$$

for some i between j and m, where m is a prespecified positive integer. A simple test statistic to use

The test statistic assumes the form

$$\begin{aligned} Q_k(j,m) &= T^2 \sum_{j=j}^{m} \frac{1}{T-l} tr(\hat{\Gamma}'(l)\hat{\Gamma}^{-1}(0)\hat{\Gamma}(l)\hat{\Gamma}^{-1}(0)) \\ &\sim \chi^2(k^2(m-j+1)). \end{aligned}$$

11. Estimation of VMA Model

Assume that $\{Z_1, \cdots, Z_n\}$ is from the MA(1) model:

$$Z_t = -\Theta_0 a_{t-1} + a_t.$$

Conditional LSE– minimizer of

$$S_n(\Theta) = \sum_{t=1}^{n} \|Z_t - (-\Theta a_{t-1})\|^2,$$

where Θ is the unknown parameter of Θ_0. Note that

$$a_1 = Z_1 + \Theta_0 a_0$$

$$a_2 = Z_2 + \Theta_0 a_1$$

$$\cdots$$

$$a_t = Z_t + \Theta_0 a_{t-1}.$$

Let $a_0 = 0$ and replace Θ_0 by Θ. Denote

$$a_1(\Theta) = Z_1 + \Theta \times 0$$

$$a_2(\Theta) = Z_2 + \Theta a_1(\Theta)$$

$$\cdots$$

$$a_t(\Theta) = Z_t + \Theta a_{t-1}(\Theta).$$

Thus, conditional LSE– minimizer of

$$S_n(\Theta) = \sum_{t=1}^{n} \|a_t(\Theta)\|^2.$$

Note that

$$a_t = Z_t + \sum_{i=1}^{t-1} \Theta_0^i Z_{t-i} + \Theta_0^t a_0,$$

$$a_t(\Theta) = Z_t + \sum_{i=1}^{t-1} \Theta^i Z_{t-i} + \Theta^t a_0.$$

The initial value a_0 does not affect the estimator, asymptotically.

Similarly, for MA(q) model:

$$a_t(\theta) = Z_t - \mu + \Theta_1 a_{t-1}(\theta) + \Theta_2 a_{t-2}(\theta) + \cdots + \Theta_q a_{t-q}(\theta),$$

where $\theta = vec[\mu, \Theta_1, \cdots, \Theta_q]$. Conditional LSE– minimizer of

$$S_n(\theta) = \sum_{t=1}^{n} \|a_t(\theta)\|^2.$$

Conditional on $Z_t = 0$ and $a_t = 0$ for $t \leq 0$. Denote the minimizer by $\widehat{\theta}$. Then $a_t(\widehat{\theta})$ is called residual.

Model checking: Ljung-Box text to examine residuals (to be white noise)

$$a_t(\widehat{\theta}) = Z_t - \widehat{\mu} + \widehat{\Theta}_1 a_{t-1}(\widehat{\theta}) + \widehat{\Theta}_2 a_{t-2}(\widehat{\theta}) + \cdots + \widehat{\Theta}_q a_{t-q}(\widehat{\theta}),$$

where $\widehat{\theta}$ is the LSE.

Test statistic $Q_k(m) \sim \chi^2(mk^2 - g)$, where g is the number of estimated parameters.

12. Forecast based VMA Model

Assume that $F_t \equiv \{Z_1, \cdots, Z_t\}$ from VAM(1) model.

$$
\begin{aligned}
\hat{Z}_t(1) &= E(Z_{t+1}|F_t) \\
&= E(\mu + a_{t+1} - \Theta_1 a_t | F_t) \\
&= \mu - \Theta_1 a_t.
\end{aligned}
$$

The associated forecast error and its covariance matrix are

$$e_t(1) = Z_{t+1} - \hat{Z}_t(1) = a_{t+1}$$

$$Cov[e_t(1)] = \Sigma_a.$$

$$a_t = Z_t + \Theta Z_{t-1} + \cdots + \Theta^{t-1} Z_1.$$

Example 3.1 Consider the monthly log returns, in percentages, of some capitalization-based portfolios obtained from the Center for Research in Security Prices (CRSP). The data span is from January 1961 to December 2011 with sample size T =612. The portfolios consist of stocks listed on NYSE, AMEX, and NASDAQ. The two portfolios used are decile 5 and decile 8 and the returns are total returns, which include capital appreciation and dividends. Figure 3.1 shows the time plots of the two log return series, whereas Figure 3.2 shows the plots of sample crosscorrelation matrices. From the sample cross-correlation plots, the two return series have significant lag-1 dynamic dependence, indicating that a VMA(1) model might be appropriate for the series. Some further details of the portfolio returns are given in the following R demonstration

$$Z_t = \begin{bmatrix} 0.92 \\ 0.98 \end{bmatrix} + \begin{bmatrix} -0.43 & 0.23 \\ -0.60 & 0.31 \end{bmatrix} a_{t-1} + a_t.$$

$$\hat{\Sigma}_a = \begin{bmatrix} 29.6 & 32.8 \\ 32.8 & 39.1 \end{bmatrix}.$$

13. Vector ARMA models

VARMA(1, 1) models

A. Model

$$(1-\Phi_1 B)Z_t = (1-\Theta_1 B)a_t.$$

B. Stationarity condition:

All the roots of $|I-\Phi_1 z| = 0$ lie outside the unit circle.

In this case, Z_t has an MA(∞) representation.

$$Z_t = w_t + \Phi_1 w_{t-1} + \cdots + \Phi_1^i w_{t-i} + \cdots,$$

where $w_t = a_t - \Theta_1 a_{t-1}$.

We can get that

$$\Gamma_0 = \Phi_1\Gamma_{-1} + \Sigma_a - \Theta_1\Sigma_a(\Phi_1 - \Theta_1)',$$
$$\Gamma_1 = \Phi_1\Gamma_0 - \Theta_1\Sigma_a,$$
$$\Gamma_j = \Phi_1\Gamma_{j-1} \; if \; j > 1.$$

C. Invertibility condition:

All the roots of $|I - \Theta_1 z| = 0$ lie outside the unit circle.

In this case, Z_t has an AR(∞) representation.

$$a_t = \tilde{Z}_t + \Theta_1 \tilde{Z}_{t-1} + \cdots + \Theta_1^i \tilde{Z}_{t-i} + \cdots ,$$

where $\tilde{Z}_t = Z_t - \Phi_1 Z_{t-1}$, or

$$\tilde{Z}_t = a_t - \Theta_1 \tilde{Z}_{t-1} - \cdots + \Theta_1^i \tilde{Z}_{t-i} - \cdots .$$

D. Constant Term

$$Z_t = \Phi_0 + \Phi_1 Z_{t-1} - \Theta_1 a_{t-1} + a_t.$$

Let $\mu = EZ_t$. Then

$$\mu = \Phi_0 + \Phi_1 \mu.$$
$$\mu = (I_k - \Phi_1)^{-1} \Phi_0.$$

VARMA(p, q) models

A. Model

$$\Phi_p(B)Z_t = \Phi_0 + \Theta_q(B)a_t,$$

where $\Phi_p(z) = I - \Phi_1 z - \cdots - \Phi_p z^p$ and $\Theta_q(z) = I - \Theta_1 z - \cdots - \Theta_q z^q$.

B. Stationarity condition:

All the roots of $|\Phi_p(z)| = 0$ lie outside the unit circle.

In this case, Z_t has an MA(∞) representation.

C. Invertibility condition:

All the roots of $|\Theta_q(z)| = 0$ lie outside the unit circle.

In this case, Z_t has an AR(∞) representation.

The covariance matrix function can be found by direct calculation, but it is quite complicated.

D. Identifiability condition

Hannan (1969, 1970, 1976, 1979):

The only common left divisors of $\Phi_p(B)$ and $\Theta_q(B)$ are unimodular ones, i.e., if $\Phi_p(z) = C(z)H(z)$ and $\Theta(z) = C(z)K(z)$, then the determinant $|C(z)|$ is a constant.

Example:

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & 2+\omega \\ 0 & \beta \end{bmatrix} \begin{bmatrix} Z_{1t-1} \\ Z_{2t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0.3 & \omega \\ 0 & \beta \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

This model is the same as:

$$\begin{aligned} Z_{1t} &= 0.8z_{1t-1} + 2z_{2t-1} + a_{1t} - 0.3a_{1t-1}, \\ Z_{2t} &= a_{2t} \end{aligned}$$

Two polynomial matrices are not left co-prime, because

$$\begin{bmatrix} 1-0.8z & -(2+\omega)z \\ 0 & 1-\beta z \end{bmatrix} = \begin{bmatrix} 1 & -\omega z \\ 0 & 1-\beta z \end{bmatrix} \begin{bmatrix} 1-0.8z & -2z \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1-0.3z & -\omega z \\ 0 & 1-\beta z \end{bmatrix} = \begin{bmatrix} 1 & -\omega z \\ 0 & 1-\beta z \end{bmatrix} \begin{bmatrix} 1-0.3z & 0 \\ 0 & 1 \end{bmatrix}$$

14. Estimation of Vector ARMA(p,q) models

Assume that $\{Z_1, \cdots, Z_n\}$ is from the ARMA(p, q) model:

$$Z_t = \Phi_{10} Z_{t-1} - \Theta_{10} a_{t-1} + a_t.$$

Conditional LSE– minimizer of

$$S_n(\Theta) = \sum_{t=1}^{n} \|Z_t - \Phi Z_{t-1} + \Theta a_{t-1}\|^2,$$

where (Φ, Θ) is the unknown parameter of (Φ_{10}, Θ_{10}). Note that

$$a_1 = Z_1 - \Phi_{10} Z_0 + \Theta_{10} a_0$$

$$a_2 = Z_2 - \Phi_{10} Z_1 + \Theta_{10} a_1$$

$$\cdots$$

$$a_t = Z_t - \Phi_{10} Z_{t-1} + \Theta_{10} a_{t-1}.$$

Let $a_0 = 0$ and replace Θ_{10} by Θ. Denote

$$a_1(\Theta) = Z_1 - \Phi Z_0 + \Theta \times 0$$

$$a_2(\Theta) = Z_2 - \Phi Z_1 + \Theta a_1(\Theta)$$

$$\cdots$$

$$a_t(\Theta) = Z_t - \Phi Z_{t-1} + \Theta a_{t-1}(\Theta).$$

Thus, conditional LSE– minimizer of

$$S_n(\Theta) = \sum_{t=1}^{n} \|a_t(\Theta)\|^2.$$

Note that

$$a_t = Z_t + \sum_{i=1}^{t-1} \Theta_0^i (Z_{t-i} - \Phi_{10} Z_{t-i-1}) + \Theta_0^t a_0,$$

$$a_t(\Theta) = Z_t + \sum_{i=1}^{t-1} \Theta^i (Z_{t-i} - \Phi Z_{t-i-1}) + \Theta^t a_0.$$

The initial value a_0 does not affect the estimator, asymptotically.

Similarly, for ARMA(p, q) model:

$$\begin{aligned} a_t(\theta) \;=\; & Z_t - \Phi_0 - \Phi_1 Z_{t-1} - \cdots - \Phi_p Z_{t-p} \\ & + \Theta_1 a_{t-1}(\theta) + \Theta_2 a_{t-2}(\theta) + \cdots + \Theta_q a_{t-q}(\theta), \end{aligned}$$

where $\theta = vec[\Phi_0, \Phi_1, \cdots, \Phi_p, \Theta_1, \cdots, \Theta_q]$. Conditional LSE– minimizer of

$$S_n(\theta) = \sum_{t=1}^{n} \|a_t(\theta)\|^2.$$

Conditional on $Z_t = 0$ and $a_t = 0$ for $t \leq 0$. Denote the minimizer by $\widehat{\theta}$. Then $a_t(\widehat{\theta})$ is called residual.

Model checking: Ljung-Box text to examine residuals (to be white noise)

$$
\begin{aligned}
a_t(\widehat{\theta}) \;=\;& Z_t - \widehat{\Phi}_0 - \widehat{\Phi}_1 Z_{t-1} - \cdots - \widehat{\Phi}_p Z_{t-p} \\
& + \widehat{\Theta}_1 a_{t-1}(\widehat{\theta}) + \widehat{\Theta}_2 a_{t-2}(\widehat{\theta}) + \cdots + \widehat{\Theta}_q a_{t-q}(\widehat{\theta}),
\end{aligned}
$$

where $\widehat{\theta}$ is the LSE.

Test statistic $Q_k(m) \sim \chi^2(mk^2 - g)$, where g is the number of estimated parameters.

15. Identification of vector ARMA model

A Summary Two-Way Table via Multivariate Q-Statistic $Q^{(m)}_{(j+1):l}$ (Page: 172).

Example 3.8. Consider the VARMA(2,1) model

$$
Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + a_t - \Theta_1 a_{t-1},
$$

where the parameters are

$$
\Phi_1 = \begin{bmatrix} 0.816 & -0.623 \\ -1.116 & 1.074 \end{bmatrix}, \; \Phi_2 = \begin{bmatrix} -0.643 & 0.592 \\ 0.615 & -0.133 \end{bmatrix}
$$

$$
\Theta_1 = \begin{bmatrix} 0 & -1.248 \\ -0.801 & 0 \end{bmatrix}, \; \Sigma_a = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}
$$

Example 3.9. Personal Income and Expenditures Consider first the monthly personal consumption expenditure (PCE) and disposable personal income (DSPI) of the United States from January 1959 to March 2012 for 639 observations. The data are in billions of dollars and seasonally adjusted. The two original series are nonstationary for obvious reasons so that we focus on the growth rates of PCE and DSPI, that is, the difference of logged data. Let z_{1t} and z_{2t} denote the growth rates, in percentages, of PCE and DSPI, respectively (PP. 176)

AR(3)

AR(8)

ARMA(3,1)

16. Nonstationary Vector ARMA(p, q) models

$$\Phi_p(B)(1-B)^d Y_t = \Theta_q(B) a_t.$$

where $\Phi_p(B)$ and $\Theta_q(B)$ are the same as those before.

General case:

$$\Phi_p(B) D(B) Y_t = \Theta_q(B) a_t.$$

where

$$D(B) = \begin{bmatrix} (1-B)^{d_1} & & & O \\ & (1-B)^{d_2} & & \\ & & \ddots & \\ O & & & (1-B)^{d_k} \end{bmatrix}.$$

Denote $Z_t = D(B) Y_t$. Then

$$\Phi_p(B) Z_t = \Theta_q(B) a_t.$$

We use this stationary and invertible ARMA(p,q) to identify (p, q) and estimate the parameters as well as model checking.

In most of cases, $d = 1$ in practice, i.e. $Z_t = Y_t - Y_{t-1}$ and $Y_t = \log P_t$. In this case, Z_t is called the log return.