Time Series Analysis

Lecture 9

Review

1. Long Memory Models
2. Seasonal Models

Today's Topics

1. Multivariate time series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models

Weak stationarity

Consider a k-dimensional time series $\mathbf{r}_t = (r_{1t}, ..., r_{kt})$. The series $\mathbf{r}_t$ is weakly stationary if its first and second moments are time invariant.

A simplest type of weakly stationary vector processes is *vector white noise* denoted by $\mathrm{WN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$. We say $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ if $E\boldsymbol{\varepsilon}_t = \mathbf{0}$, $\mathrm{var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$, and $\mathrm{cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s) = \mathbf{0}$ for any $t \neq s$. Hence there exists no serial correlation across all the components of $\boldsymbol{\varepsilon}_t$. However different components of $\boldsymbol{\varepsilon}_t$ may be correlated with each contemporaneously as $\boldsymbol{\Sigma}_\varepsilon$ is not necessarily a diagonal matrix.

Vector white noise processes serve as building blocks for constructing vector stationary processes.

Mean and Covariance

For a weakly stationary time series $\mathbf{r}_t$, we define its mean vector and covariance matrix as

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t), \quad \boldsymbol{\Gamma}_0 = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_t - \boldsymbol{\mu})'],$$

where the expectation is taken element by element over the joint distribution of $\mathbf{r}_t$. We write $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)'$ and $\boldsymbol{\Gamma}_0 = [\Gamma_{ij}(0)]$ when the elements are needed.

The ith diagonal element of $\boldsymbol{\Gamma}_0$ is the variance of r_{it}, whereas the (i, j)th element of $\boldsymbol{\Gamma}_0$ is the covariance between r_{it} and r_{jt} .

Cross-Correlation Matrices

Let $\mathbf{D} = diag\{\sqrt{\Gamma_{11}(0)}, \cdots, \sqrt{\Gamma_{kk}(0)}\}$. The concurrent, or lag-zero, cross-correlation matrix of $\mathbf{r}_t$ is defined as

$$\boldsymbol{\rho}_0 \equiv [\rho_{ij}(0)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1}.$$

More specifically, the (i,j)th element of $\boldsymbol{\rho}_0$ is

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

which is the concurrent correlation coefficient between r_{it} and r_{jt}.

It is easy to see that

$$\rho_{ij}(0) = \rho_{ji}(0),\ -1 \le \rho_{ij}(0) \le 1, \text{ and } \rho_{ii}(0) = 1 \text{ for } 1 \le i, j \le k.$$

Thus, $\boldsymbol{\rho}_0$ is a symmetric matrix with unit diagonal elements.

The lag-ℓ cross-covariance matrix of $\mathbf{r}_t$ is defined as

$$\boldsymbol{\Gamma}_\ell \equiv [\Gamma_{ij}(\ell)] = E[(\boldsymbol{r}_t - \boldsymbol{\mu})(\boldsymbol{r}_{t-\ell} - \boldsymbol{\mu})'],$$

Therefore, the (i, j)th element of $\boldsymbol{\Gamma}_\ell$ is the covariance between r_{it} and $r_{j,t-\ell}$. For a weakly stationary series, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ is a function of ℓ, not the time index t.

The lag-ℓ cross-correlation matrix (CCM) of $\mathbf{r}_t$ is defined as

$$\boldsymbol{\rho}_\ell \equiv [\rho_{ij}(\ell)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_\ell \boldsymbol{D}^{-1},$$

with

$$\rho_{ij}(\ell) = \frac{\Gamma_{ij}(\ell)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\mathrm{Cov}(r_{it}, r_{j,t-\ell})}{\mathrm{std}(r_{it})\mathrm{std}(r_{jt})},$$

which is the correlation coefficient between r_{it} and $r_{j,t-\ell}$.

When $\ell > 0$, $\rho_{ij}(\ell)$ measures the linear dependence of r_{it} on $r_{j,t-\ell}$, which occurred prior to time t. If $\rho_{ij}(\ell) \neq 0$ and $\ell > 0$, we say r_{jt} leads r_{it} at lag ℓ.

Properties of the cross correlations when $\ell > 0$

First, in general, $\rho_{ij}(\ell) \neq \rho_{ji}(\ell)$ for $i \neq j$. Therefore, $\mathbf{\Gamma}_\ell$ and $\boldsymbol{\rho}_\ell$ are in general not symmetric.

Second, since we have

$$\mathrm{Cov}(r_{it}, r_{j,t-\ell}) = \mathrm{Cov}(r_{j,t-\ell}, r_{it}) = \mathrm{Cov}(r_{jt}, r_{i,t+\ell}) = \mathrm{Cov}(r_{jt}, r_{i,t-(-\ell)}),$$

so $\Gamma_{ij}(\ell) = \Gamma_{ji}(-\ell)$. Then we have $\mathbf{\Gamma}_\ell = \mathbf{\Gamma}'_{-\ell}$ and $\boldsymbol{\rho}_\ell = \boldsymbol{\rho}'_{-\ell}$. Thus it suffices to consider $\boldsymbol{\rho}_\ell$ for $\ell \geq 0$.

Sample Cross-Correlation Matrices

Given the data $\mathbf{r}_t | t = 1, ..., T$, the cross-covariance matrix $\boldsymbol{\Gamma}_\ell$ can be estimated by

$$\widehat{\boldsymbol{\Gamma}}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (\boldsymbol{r}_t - \bar{\boldsymbol{r}})(\boldsymbol{r}_{t-\ell} - \bar{\boldsymbol{r}})', \qquad \ell \geq 0,$$

where $\bar{\mathbf{r}} = (\sum_{t=1}^{T} \mathbf{r}_t)/T$ is the vector of sample means.

The cross-correlation matrix $\boldsymbol{\rho}_\ell$ is estimated by

$$\widehat{\boldsymbol{\rho}}_\ell = \widehat{\boldsymbol{D}}^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{D}}^{-1}, \qquad \ell \geq 0,$$

where $\hat{\mathbf{D}}$ is the $k \times k$ diagonal matrix of the sample standard deviations of the component series.

Simplified CCM (Tiao and Box, 1981)

Empirical experience indicates that it is rather hard to absorb simultaneously many cross-correlation matrices, especially when the dimension k is greater than 3.

1. Plus sign (+) means that the corresponding correlation coefficient is greater than or equal to $2/\sqrt{T}$.
2. Minus sign (−) means that the corresponding correlation coefficient is less than or equal to $-2/\sqrt{T}$.
3. Period (.) means that the corresponding correlation coefficient is between $-2/\sqrt{T}$ and $2/\sqrt{T}$.

And $1/\sqrt{T}$ is the asymptotic standard deviation of the sample correlation under the assumption that $\mathbf{r}_t$ is a white noise series.

Multivariate Portmanteau (Ljung-Box) Tests

$H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_m = 0 \;\; v.s. \quad H_1 : \boldsymbol{\rho}_i \neq 0 \textit{ for some } i \in \{1, \cdots, m\}$

The statistic is used to test that there are no auto- and cross correlations in the vector series $\mathbf{r}_t$.

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} \mathrm{tr}(\widehat{\boldsymbol{\Gamma}}_\ell' \widehat{\boldsymbol{\Gamma}}_0^{-1} \widehat{\boldsymbol{\Gamma}}_\ell \widehat{\boldsymbol{\Gamma}}_0^{-1}),$$

where T is the sample size, k is the dimension of $\mathbf{r}_t$.

Under the null hypothesis and some regularity conditions, $Q_k(m)$ follows asymptotically a chi squared distribution with k^2m degrees of freedom.
See Hosking (1980, 1981) and Li and Mcleod (1981) for more details.

The $Q_k(m)$ statistic is a joint test for checking the first m cross-correlation matrices of $\mathbf{r}_t$ being zero.

If it rejects the null hypothesis, then we build a multivariate model for the series to study the lead-lag relationships between the component series.

Example

Consider the monthly log returns of IBM stock
and the S&P 500 index from January 1926 to December 2008 with 996 observations. Denote the returns of IBM stock and the S&P 500 index by
r_{1t} and r_{2t} , respectively. These two returns form a bivariate time series $\mathbf{r_t} = (r_{1t}, r_{2t})'$.

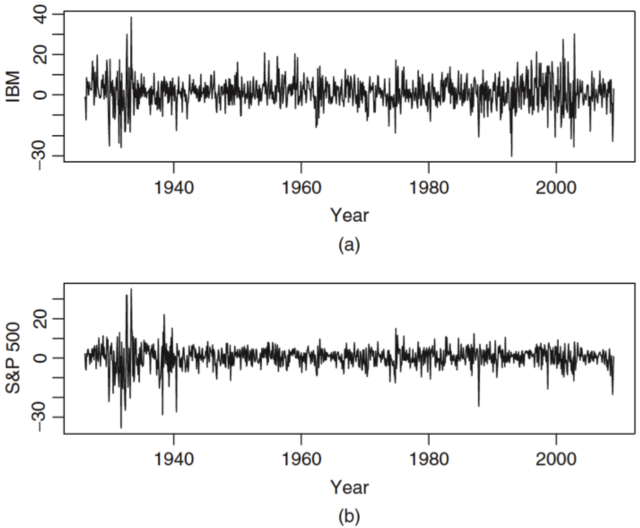


Figure 8.1 Time plots of monthly log returns, in percentages, for (a) IBM stock and (b) the S&P 500 index from January 1926 to December 2008.

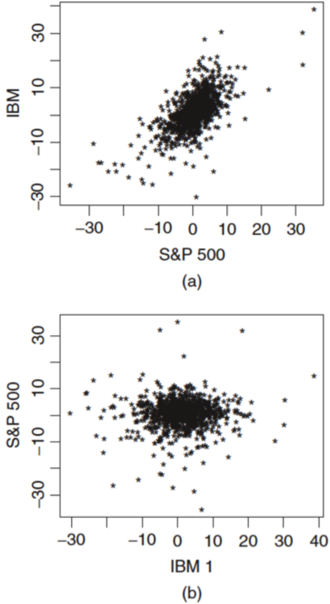


Figure 8.2 Some scatterplots for monthly log returns of IBM stock and S&P 500 index: (a) concurrent plot of IBM vs. S&P 500, (b) S&P 500 vs. lag-1 IBM, (c) IBM vs. lag-1 S&P 500, and (d) S&P 500 vs. lag-1 S&P 500.

TABLE 8.1 Summary Statistics and Cross-Correlation Matrices of Monthly Log Returns of IBM Stock and S&P 500 Index: January 1926 to December 2008

(a) Summary Statistics

Ticker	Mean	Standard Error	Skewness	Excess Kurtosis	Minimum	Maximum
IBM	1.089	7.033	−0.068	2.622	−30.37	38.57
SP5	0.430	5.537	−0.521	7.927	−35.59	35.22

(b) Cross-Correlation Matrices

Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
0.04	0.10	0.00	−0.08	−0.01	−0.06	−0.03	−0.03	0.02	0.08
0.04	0.08	0.02	−0.02	−0.06	−0.10	0.04	0.03	0.00	0.09

TABLE 8.1 Summary Statistics and Cross-Correlation Matrices of Monthly Log Returns of IBM Stock and S&P 500 Index: January 1926 to December 2008

(a) Summary Statistics

Ticker	Mean	Standard Error	Skewness	Excess Kurtosis	Minimum	Maximum
IBM	1.089	7.033	−0.068	2.622	−30.37	38.57
SP5	0.430	5.537	−0.521	7.927	−35.59	35.22

(b) Cross-Correlation Matrices

Lag 1		Lag 2		Lag 3		Lag 4		Lag 5	
0.04	0.10	0.00	−0.08	−0.01	−0.06	−0.03	−0.03	0.02	0.08
0.04	0.08	0.02	−0.02	−0.06	−0.10	0.04	0.03	0.00	0.09

(c) Simplified notation

$$\begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix} \quad \begin{bmatrix} \cdot & - \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & - \end{bmatrix} \quad \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \quad \begin{bmatrix} \cdot & + \\ \cdot & + \end{bmatrix}$$

Applying the $Q_k(m)$ statistics to the bivariate monthly log returns of IBM stock and the S&P 500 index, we have $Q_2(1) = 9.81$, $Q_2(5) = 47.06$, and $Q_2(10) = 71.65$. Based on asymptotic chi-squared distributions with degrees of freedom 4, 20, and 40, the p values of these $Q_2(m)$ statistics are 0.044, 0.001, and 0.002, respectively. The portmanteau tests thus confirm the existence of serial dependence in the bivariate return series at the 5% significance level.

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models

Today's Topics

1. Multivariate Time Series

2. Vector AR models
   - VAR(1)
   - VAR(p)
   - Building a VAR(p) Model
   - Granger causality

3. Vector MA models

4. Vector ARMA models

VAR(1)

A multivariate time series $\mathbf{r}_t$ is a VAR(1) process if it follows the model

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi} \boldsymbol{r}_{t-1} + \boldsymbol{a}_t,$$

where ϕ_0 is a k-dimensional vector, $\boldsymbol{\Phi}$ is a $k \times k$ matrix, and $\{\mathbf{a}_t\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$.

In application, the covariance matrix $\boldsymbol{\Sigma}$ is required to be positive definite; otherwise, the dimension of $\mathbf{r}_t$ can be reduced. In the literature, it is often assumed that $\mathbf{a}_t$ is multivariate normal.

Consider the bivariate case [i.e., $k = 2$, $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$, and $\boldsymbol{a}_t = (a_{1t}, a_{2t})'$]. The VAR(1) model consists of the following two equations:

$$r_{1t} = \phi_{10} + \Phi_{11} r_{1,t-1} + \Phi_{12} r_{2,t-1} + a_{1t},$$

$$r_{2t} = \phi_{20} + \Phi_{21} r_{1,t-1} + \Phi_{22} r_{2,t-1} + a_{2t},$$

where Φ_{ij} is the (i,j)th element of $\mathbf{\Phi}$ and ϕ_{i0} is the ith element of ϕ_0.

Consider the two equations jointly. If $\Phi_{12} = 0$ and $\Phi_{21} \neq 0$, then there is a unidirectional relationship from r_{1t} to r_{2t}. If $\Phi_{12} = \Phi_{21} = 0$, then r_{1t} and r_{2t} are uncoupled. If $\Phi_{12} \neq 0$ and $\Phi_{21} \neq 0$, then there is a feedback relationship between the two series.

Stationarity condition and moments

Taking expectation on both sides of the VAR(1) gives

$$E(\mathbf{r}_t) = \phi_0 + \mathbf{\Phi} E(\mathbf{r}_{t-1}).$$

Since $E(\mathbf{r}_t)$ is time invariant, we have

$$\boldsymbol{\mu} = E(\mathbf{r}_t) = (\mathbf{I} - \mathbf{\Phi})^{-1}\phi_0$$

Let $\tilde{\mathbf{r}}_t = \mathbf{r}_t - \boldsymbol{\mu}$ be the mean-corrected time series. Then the VAR(1) model becomes

$$\tilde{\boldsymbol{r}}_t = \mathbf{\Phi}\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t.$$

By repeated substitutions, we can rewrite

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \mathbf{\Phi}\boldsymbol{a}_{t-1} + \mathbf{\Phi}^2\boldsymbol{a}_{t-2} + \mathbf{\Phi}^3\boldsymbol{a}_{t-3} + \cdots.$$

$\mathbf{r}_t$ depends on the past innovation $\mathbf{a}_{t-j}$ with coefficient matrix $\mathbf{\Phi}^j$. For such dependence to be meaningful, $\mathbf{\Phi}^j$ must converge to zero as $j \to \infty$. This means that the k eigenvalues of $\mathbf{\Phi}$ must be less than 1 in modulus; otherwise, $\mathbf{\Phi}^j$ will either explode or converge to a nonzero matrix as $j \to \infty$.

All eigenvalues of $\mathbf{\Phi}$ are less than 1 in modulus is the necessary and sufficient condition for weak stationarity of $\mathbf{r}_t$ provided that the covariance matrix of $\mathbf{a}_t$ exists.

Furthermore, because

$$|\lambda \boldsymbol{I} - \boldsymbol{\Phi}| = \lambda^k \mid \boldsymbol{I} - \boldsymbol{\Phi}\frac{1}{\lambda} \mid,$$

the eigenvalues of $\boldsymbol{\Phi}$ are the inverses of the zeros of the determinant $|\mathbf{I} - \boldsymbol{\Phi}B|$. Thus, an equivalent sufficient and necessary condition for stationarity of $\mathbf{r}_t$ is that all zeros of the determinant $|\boldsymbol{\Phi}(B)|$ are greater than one in modulus; that is, all zeros are outside the unit circle in the complex plane.

CCM for VAR(1)

Postmultiplying $\tilde{\mathbf{r}}'_{t-\ell}$ to

$$\tilde{\mathbf{r}}_t = \mathbf{\Phi}\tilde{\mathbf{r}}_{t-1} + \mathbf{a}_t, \tag{1}$$

taking expectation, we obtain

$$E(\tilde{\boldsymbol{r}}_t \tilde{\boldsymbol{r}}'_{t-\ell}) = \mathbf{\Phi} E(\tilde{\boldsymbol{r}}_{t-1}\tilde{\boldsymbol{r}}_{t-\ell})', \qquad \ell > 0.$$

Therefore,

$$\mathbf{\Gamma}_\ell = \mathbf{\Phi}\mathbf{\Gamma}_{\ell-1}, \ \ell > 0, \tag{2}$$

$$\mathbf{\Gamma}_\ell = \mathbf{\Phi}^\ell \mathbf{\Gamma}_0, \ \ell > 0.$$

Pre- and postmultiplying (2) by $\mathbf{D}^{-1}$, we obtain

$$\boldsymbol{\rho}_\ell = \mathbf{D}^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}_{\ell-1}\mathbf{D}^{-1} = \mathbf{D}^{-1}\boldsymbol{\Phi}\mathbf{D}\mathbf{D}^{-1}\boldsymbol{\Gamma}_{\ell-1}\mathbf{D}^{-1} = \boldsymbol{\Upsilon}\boldsymbol{\rho}_{\ell-1},$$

where $\boldsymbol{\Upsilon} = \mathbf{D}^{-1}\boldsymbol{\Phi}\mathbf{D}$. Then we have

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}^\ell\boldsymbol{\rho}_0, \ \ell > 0.$$

Today's Topics

The time series $\mathbf{r}_t$ follows a VAR(p) model if it satisfies

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \qquad p>0,$$

where $\boldsymbol{\phi}_0$ and $\boldsymbol{a}_t$ are defined as before, and $\boldsymbol{\Phi}_j$ are $k \times k$ matrices. Using the back-shift operator B, the VAR(p) model can be written as

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{I}$ is the $k \times k$ identity matrix. This representation can be written in a compact form as

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t,$$

where $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ is a matrix polynomial. If $\boldsymbol{r}_t$ is weakly stationary, then we have

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$$

provided that the inverse exists.

Let $\tilde{\mathbf{r}}_t = \mathbf{r}_t - \boldsymbol{\mu}$. The VAR(p) model becomes

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}_1 \tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Phi}_p \tilde{\boldsymbol{r}}_{t-p} + \boldsymbol{a}_t.$$

Similar to VAR(1), we can obtain

- $\mathrm{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{a}_t$.
- $\mathrm{Cov}(\boldsymbol{r}_{t-\ell}, \boldsymbol{a}_t) = \boldsymbol{0}$ for $\ell > 0$.
- $\boldsymbol{\Gamma}_\ell = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{\ell-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}_{\ell-p}$ for $\ell > 0$.

In terms of CCM, the moment equations become

$$\boldsymbol{\rho}_\ell = \boldsymbol{\Upsilon}_1 \boldsymbol{\rho}_{\ell-1} + \cdots + \boldsymbol{\Upsilon}_p \boldsymbol{\rho}_{\ell-p} \qquad \text{for } \ell > 0,$$

where $\boldsymbol{\Upsilon}_i = \mathbf{D}^{-1} \boldsymbol{\Phi}_i \mathbf{D}$.

Today's Topics

Step 1. Order specification

Consider the following consecutive VAR models:

$$
\begin{aligned}
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{a}_t \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{r}_{t-2} + \boldsymbol{a}_t \\
&\vdots \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_i \boldsymbol{r}_{t-i} + \boldsymbol{a}_t \\
&\vdots
\end{aligned}
$$

Parameters of these models can be estimated by the ordinary least-squares (OLS) method.

In general, we use the ith and $(i-1)$th equations to test $H_0 : \mathbf{\Phi}_i = 0$ versus $H_1 : \mathbf{\Phi}_i \neq 0$; that is, testing a VAR(i) model versus a VAR(i-1) model. The test statistic is

$$M(i) = -\left(T - k - i - \frac{3}{2}\right) \ln\left(\frac{|\widehat{\mathbf{\Sigma}}_i|}{|\widehat{\mathbf{\Sigma}}_{i-1}|}\right).$$

where

$$\widehat{\mathbf{\Sigma}}_i = \frac{1}{T - 2i - 1} \sum_{t=i+1}^{T} \widehat{\boldsymbol{a}}_t^{(i)} (\widehat{\boldsymbol{a}}_t^{(i)})', \qquad i \geq 0.$$

$$\widehat{\boldsymbol{a}}_t^{(i)} = \boldsymbol{r}_t - \widehat{\boldsymbol{\phi}}_0^{(i)} - \widehat{\mathbf{\Phi}}_1^{(i)} \boldsymbol{r}_{t-1} - \cdots - \widehat{\mathbf{\Phi}}_i^{(i)} \boldsymbol{r}_{t-i}.$$

Asymptotically, $M(i)$ is distributed as a chi-squared distribution with k^2 degrees of freedom; see Tiao and Box (1981).

Information Criteria

Alternatively, one can use the Akaike information criterion (AIC) or its variants to select the order p.

$$AIC(i) = \ln(|\hat{\Sigma}_i|) + \frac{2k^2 i}{T}$$

$$BIC(i) = \ln(|\hat{\Sigma}_i|) + \frac{k^2 i \ln(T)}{T}$$

$$HQ(i) = \ln(|\hat{\Sigma}_i|) + \frac{k^2 i \ln[\ln(T)]}{T}$$

Step 2. Estimation

For a specified VAR model, one can estimate the parameters using either the OLS method or the ML method. The two methods are equivalent (under normality).

For AR models, the OLS estimates are equivalent to the (conditional) ML estimates.

Under some regularity conditions, the estimates are asymptotically normal; see Reinsel (1993).

Step 3. Model Checking

A fitted model should then be checked carefully for any possible inadequacy. The $Q_k(m)$ statistic can be applied to the residual series to check the assumption that there are no serial or cross correlations in the residuals.

For a fitted VAR(p) model, the $Q_k(m)$ statistic of the residuals is asymptotically a chi-squared distribution with $k^2m - g$ degrees of freedom, where g is the number of estimated parameters in the AR coefficient matrices; see Lütkepohl (2005).

Forecasting

For a VAR(p) model, the 1-step-ahead forecast at the time origin h is

$$\boldsymbol{r}_h(1) = \boldsymbol{\phi}_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+1-i},$$

and the associated forecast error is $\mathbf{e}_h(1) = \mathbf{a}_{h+1}$. The covariance matrix of the forecast error is $\boldsymbol{\Sigma}$.

For 2-step-ahead forecasts, we substitute $\mathbf{r}_{h+1}$ by its forecast to obtain

$$\boldsymbol{r}_h(2) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{r}_h(1) + \sum_{i=2}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{h+2-i},$$

and the associated forecast error is

$$\boldsymbol{e}_h(2) = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1[\boldsymbol{r}_t - \boldsymbol{r}_h(1)] = \boldsymbol{a}_{h+2} + \boldsymbol{\Phi}_1 \boldsymbol{a}_{h+1}.$$

The covariance matrix of the forecast error is $\boldsymbol{\Sigma} + \boldsymbol{\Phi}_1 \boldsymbol{\Sigma} \boldsymbol{\Phi}_1'$.

If $\mathbf{r}_t$ is weakly stationary, then the ℓ-step-ahead forecast $\mathbf{r}_h(\ell)$ converges to its mean vector $\boldsymbol{\mu}$ as the forecast horizon ℓ increases and the covariance matrix of its forecast error converges to the covariance matrix of $\mathbf{r}_t$.

Example

R-package **vars**

Suppose X is the matrix with 3 columns containing the log daily returns in 2011 of FTSE 100, FTSE MidCap, FTSE SmallCap. To fit a vector AR(p) model with p determined by BIC (i.e. the Schwartz Criterion),

```
> FTSEvar = VAR(X, lag.max=3, ic="SC"); summary(FTSEvar)
```

```
Estimation results for equation FTSE100:
========================================
FTSE100 = FTSE100.l1 + FTSE.MidCap.l1 + FTSE.SmallCap.l1
          + const
                   Estimate Std. Error t value Pr(>|t|)
FTSE100.l1        0.4623839  0.1822584   2.537   0.0118 *
FTSE.MidCap.l1   -0.5277461  0.2154000  -2.450   0.0150 *
FTSE.SmallCap.l1  0.2438484  0.1981300   1.231   0.2196
const            -0.0003319  0.0008497  -0.391   0.6964
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01329 on 244 degrees of freedom
Multiple R-Squared: 0.03624,Adjusted R-squared: 0.02439
F-statistic: 3.058 on 3 and 244 DF,  p-value: 0.02896
```

```
Estimation results for equation FTSE.MidCap:
============================================
FTSE.MidCap = FTSE100.l1 + FTSE.MidCap.l1 + FTSE.SmallCap.l1
              + const
                   Estimate Std. Error t value Pr(>|t|)
FTSE100.l1        0.6041161  0.1689675   3.575 0.000422 ***
FTSE.MidCap.l1   -0.5709383  0.1996923  -2.859 0.004616 **
FTSE.SmallCap.l1  0.2651595  0.1836817   1.444 0.150139
const            -0.0005873  0.0007877  -0.746 0.456663
---
Residual standard error: 0.01232 on 244 degrees of freedom
Multiple R-Squared: 0.08665,Adjusted R-squared: 0.07542
F-statistic: 7.716 on 3 and 244 DF,  p-value: 6.044e-05
```

```
Estimation results for equation FTSE.SmallCap:
==============================================
FTSE.SmallCap = FTSE100.l1 + FTSE.MidCap.l1 + FTSE.SmallCap.l1
                + const


                    Estimate Std. Error t value Pr(>|t|)
FTSE100.l1         0.2692036  0.1076958   2.500   0.0131 *
FTSE.MidCap.l1    -0.1276323  0.1272791  -1.003   0.3170
FTSE.SmallCap.l1   0.0778511  0.1170743   0.665   0.5067
const             -0.0006338  0.0005021  -1.262   0.2080
---
Residual standard error: 0.007855 on 244 degrees of freedom
Multiple R-Squared: 0.1037,Adjusted R-squared: 0.09271
F-statistic: 9.414 on 3 and 244 DF,  p-value: 6.557e-06

Covariance matrix of residuals:
                FTSE100 FTSE.MidCap FTSE.SmallCap
FTSE100       1.767e-04   1.545e-04     8.617e-05
FTSE.MidCap   1.545e-04   1.519e-04     8.237e-05
FTSE.SmallCap 8.617e-05   8.237e-05     6.170e-05
```

```
Covariance matrix of residuals:
              FTSE100 FTSE.MidCap FTSE.SmallCap
FTSE100     1.767e-04   1.545e-04     8.617e-05
FTSE.MidCap 1.545e-04   1.519e-04     8.237e-05
FTSE.SmallCap 8.617e-05 8.237e-05     6.170e-05

Correlation matrix of residuals:
            FTSE100 FTSE.MidCap FTSE.SmallCap
FTSE100      1.0000      0.9430        0.8252
FTSE.MidCap  0.9430      1.0000        0.8508
FTSE.SmallCap 0.8252     0.8508        1.0000
```

To refit the model by leaving out insignificant terms, run **FTSEvarR=restrict(FTSEvar)** which leads to

$$\begin{cases} X_{t1} = 0.469X_{t-1,1} - 0.398X_{t-1,2} + \varepsilon_{t1}, \\ X_{t2} = 0.608X_{t-1,1} - 0.427X_{t-1,2} + \varepsilon_{t2}, \\ X_{t3} = 0.195X_{t-1,1} + \varepsilon_{t3}. \end{cases}$$

where X_{t1}, X_{t2}, X_{t3} denote the log returns on day t of, respectively, FTSE 100, FTSE MidCap, FTSE SmallCap.

The cross-correlations of the residuals from the above fitted vector AR(1) model, produced by **acf(residuals(FTSEvarR))**, is

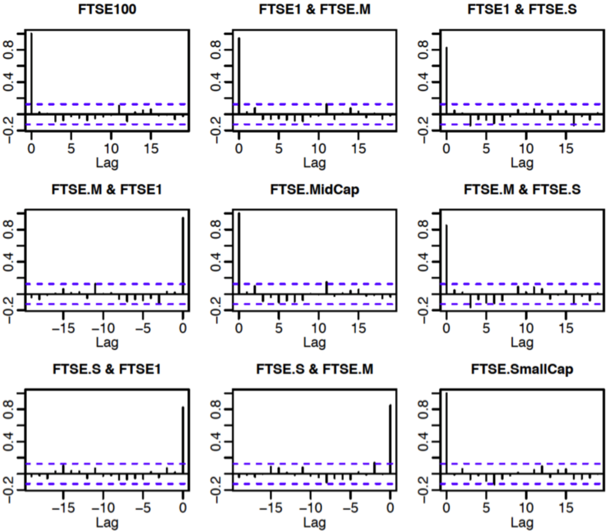
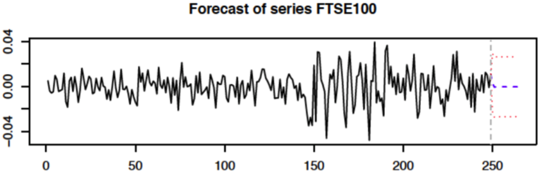


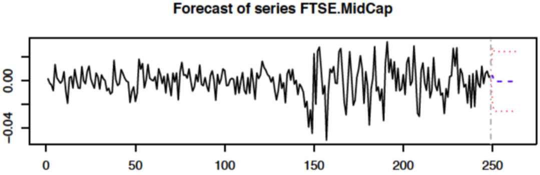
FIGURE 4.3. Sample cross-correlations of the residuals resulted from the fitted vector AR(1) model (4.23) for the log returns of the daily prices of FTSE 100 index, FTSE MidCap index, and FTSE SmallCap index in 2011-2012.

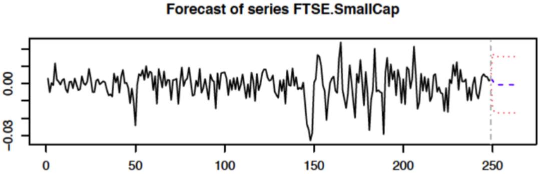
More diagnostic plots can be produced by calling the following R-functions:
FTSEdiag = serial.test(FTSEvarR); plot(FTSEdiag)

To perform the portmanteau test for the residual with, for example, $m = 6$, run
serial.test(FTSEvar, lags.pt=6, type ="PT.adjusted").

We may forecast next 15 returns of those three FTSE indices with the predictive boundarys with the coverary probability 0.95 by calling
FTSEpred = predict(FTSEvarR, n.ahead=15, ci=0.95)
plot(FTSEpred)


Forecast of series FTSE100
0.04
0.00
−0.04
0
50
100
150
200
250


Forecast of series FTSE.MidCap
0.00
−0.04
0
50
100
150
200
250


Forecast of series FTSE.SmallCap
0.00
−0.03
0
50
100
150
200
250

Today's Topics

Granger causality

The Granger causality (Granger, 1969) is an important concept in econometrics. Let Z_t and Y_t be two univariate time series. Let $\mathcal{L}(U|V)$ denotes the conditional distribution of U given V. Time series Z_t is said to Granger cause time series Y_t if

$$\mathcal{L}(Y_t|Y_{t-1}, Z_{t-1}, Y_{t-2}, Z_{t-2} \cdots) \neq \mathcal{L}(Y_t|Y_{t-1}, Y_{t-2}, \cdots),$$

Obviously the Granger causality would present if the changes in the lagged values of Z_t indeed cause the changes in Y_t.

Granger causality in mean

While the above definition is general, practical applications often narrow down to the so-called the Granger causality in mean, for which the above condition is replace by

$$E(Y_t|Y_{t-1}, Z_{t-1}, Y_{t-2}, Z_{t-2} \cdots) \neq E(Y_t|Y_{t-1}, Y_{t-2}, \cdots).$$

The causality can easily be verified for the VAR models.

Let $\mathbf{X}_t = (Z_t, Y_t)'$, and assume $\mathbf{X}_t \sim AR(p)$, i.e.

$$\mathbf{X}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim WN(0, \boldsymbol{\Sigma}_\varepsilon)$. Under the additional assumption that $\mathbf{X}_t$ is a Guassian process, the above Granger causality condition is equivalent to the condition that at least one of $a_{21}^{(k)} \neq 0$ for $1 \leq k \leq p$, where $a_{ij}^{(k)}$ denotes the (i, j)th element in the coefficient matrix $\mathbf{A}_k$.

The so-called Granger causality test is to test the hypothesis

$$H_0 : a_{21}^{(1)} = \cdots = a_{21}^{(p)} = 0. \tag{3}$$

When H_0 is rejected, Z_t is regarded as Grange causing Y_t.

To test hypothesis H_0, we apply the standard F-test statistic:

$$F = \frac{(\mathrm{RSS}_r - \mathrm{RSS})/p}{\mathrm{RSS}/(2T - 4p - 2)},$$

where the residual sum squares is defined as

$$\mathrm{RSS} = \sum_{t=p+1}^{T} \left\| \mathbf{X}_t - \widehat{c} - \sum_{j=1}^{p} \widehat{\mathbf{A}}_j \mathbf{X}_{t-j} \right\|^2,$$

and the restrictive residual sum squares RSS_r is defined in the similar manner but with the constraints specified in H_0.

Under H_0, pF is asymptotically chi-squared distributed with p degrees of freedom. Alternatively F may be approximately regarded as obeying F-distribution with $(p, 2T - 4p - 2)$ degrees of freedom.

Instantaneous causality

For the bivariate AR model

$$\mathbf{X}_t = \mathbf{c} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j} + \boldsymbol{\varepsilon}_t,$$

if the off-diagonal element $\boldsymbol{\Sigma}_\varepsilon$ is not zero, $Corr(Z_t, Y_t) \neq 0$. In this case, Z_t and Y_t have instantaneous Granger causality.

To test the instantaneous Granger causality is to test $H_0 : \sigma_{21} = 0$, where σ_{21} denotes the off-diagonal element of $\boldsymbol{\Sigma}_\varepsilon$. A Wald test statistic can be contructed based on the estimate for σ_{21} and its asymtotic normality. See, for example, §3.6.3 of Lütkepohl (2006).

Example

The function **causality** in the package **vars** implement the above two tests.

Let $X12$ be a data matrix containing the daily log returns of FTSE 100 and FTSE Midcap as its two columns, we run

```
> m12 = VAR(X12, ic="SC")
> causality(m12)

H0: FTSE100 do not Granger-cause FTSE.MidCap
F-Test = 13.3076, df1 = 1, df2 = 490, p-value = 0.0002927

H0: No instantaneous causality between FTSE100 and FTSE.MidCap
Chi-squared = 116.7708, df = 1, p-value < 2.2e-16
```

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models

VMA(q)

A vector moving-average model of order q, or VMA(q), is in the form

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1 \boldsymbol{a}_{t-1} - \cdots - \boldsymbol{\Theta}_q \boldsymbol{a}_{t-q} \quad \text{or} \quad \boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t, \tag{8.23}$$

where $\boldsymbol{\theta}_0$ is a k-dimensional vector, $\boldsymbol{\Theta}_i$ are $k \times k$ matrices, and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ is the MA matrix polynomial in the back-shift operator B.

Similar to the univariate case, VMA(q) processes are weakly stationary provided that the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{a}_t$ exists.

It is easy to see that $\boldsymbol{\mu} = E(\mathbf{r}_t) = \boldsymbol{\theta}_0$.

Let $\tilde{\boldsymbol{r}}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0$ be the mean-corrected VAR(q) process. Then using Eq. (8.23) and the fact that $\{\boldsymbol{a}_t\}$ has no serial correlations, we have

1. $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t) = \boldsymbol{\Sigma}$.
2. $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1 \boldsymbol{\Sigma} \boldsymbol{\Theta}_1' + \cdots + \boldsymbol{\Theta}_q \boldsymbol{\Sigma} \boldsymbol{\Theta}_q'$.
3. $\boldsymbol{\Gamma}_\ell = \mathbf{0}$ if $\ell > q$.
4. $\boldsymbol{\Gamma}_\ell = \sum_{j=\ell}^{q} \boldsymbol{\Theta}_j \boldsymbol{\Sigma} \boldsymbol{\Theta}_{j-\ell}'$ if $1 \leq \ell \leq q$, where $\boldsymbol{\Theta}_0 = -\boldsymbol{I}$.

Since $\boldsymbol{\Gamma}_\ell = \mathbf{0}$ for $\ell > q$, the cross-correlation matrices (CCMs) of a VMA(q) process $\boldsymbol{r}_t$ satisfy

$$\boldsymbol{\rho}_\ell = \mathbf{0}, \qquad \ell > q. \tag{8.24}$$

Therefore, similar to the univariate case, the sample CCMs can be used to identify the order of a VMA process.

To better understand the VMA processes, let us consider the bivariate MA(1) model

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}\boldsymbol{a}_{t-1} = \boldsymbol{\mu} + \boldsymbol{a}_t - \boldsymbol{\Theta}\boldsymbol{a}_{t-1}, \tag{8.25}$$

where, for simplicity, the subscript of $\boldsymbol{\Theta}_1$ is removed. This model can be written explicitly as

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}. \tag{8.26}$$

It says that the current return series $\boldsymbol{r}_t$ only depends on the current and past shocks. Therefore, the model is a finite-memory model.

Estimation

Unlike the VAR models, estimation of VMA models is much more involved; see Hillmer and Tiao (1979), Lütkepohl (2005), and the references therein.

For the likelihood approach, there are two methods available. The first is the conditional-likelihood method that assumes that $\mathbf{a}_t = 0$ for $t \leq 0$. The second is the exact-likelihood method that treats $\mathbf{a}_t$ with $t \leq 0$ as additional parameters of the model.

Conditional MLE: take VMA(1) as an example

The conditional-likelihood method assumes that $\boldsymbol{a}_0 = \boldsymbol{0}$. Under such an assumption and rewriting the model as $\boldsymbol{a}_t = \boldsymbol{r}_t - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_{t-1}$, we can compute the shock $\boldsymbol{a}_t$ recursively as

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0, \qquad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}_1\boldsymbol{a}_1, \quad \ldots.$$

Consequently, the likelihood function of the data becomes

$$f(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T | \boldsymbol{\theta}_0, \boldsymbol{\Theta}_1, \boldsymbol{\Sigma}) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{a}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{a}_t\right),$$

which can be evaluated to obtain the parameter estimates.

In summary, building a VMA model involves three steps:

(a) Use the sample cross-correlation matrices to specify the order q–for a VMA(q) model, $\boldsymbol{\rho}_\ell = 0$ for $\ell > q$;

(b) Estimate the specified model by using either the conditional- or exact likelihood method–the exact method is preferred when the sample size is not large;

(c) the fitted model should be checked for adequacy [e.g., applying the $Q_k(m)$ statistics to the residual series].

Finally, forecasts of a VMA model can be obtained by using the same procedure as a univariate MA model.

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models
   - Identifiability Problem
   - VARMA(p,q)

Example 1

Unlike the univariate ARMA models, VARMA models may not be uniquely defined.

For example, the VMA(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

is identical to the VAR(1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

The equivalence of the two models can easily be seen by examining their component models. For the VMA(1) model, we have

$$r_{1t} = a_{1t} - 2a_{2,t-1}, \qquad r_{2t} = a_{2t}.$$

For the VAR(1) model, the equations are

$$r_{1t} + 2r_{2,t-1} = a_{1t}, \qquad r_{2t} = a_{2t}.$$

From the model for r_{2t}, we have $r_{2,t-1} = a_{2,t-1}$. Therefore, the models for r_{1t} are identical. This type of identifiability problem is harmless because either model can be used in a real application.

Example 2

Consider the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

This model is identical to the VARMA(1,1) model

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix},$$

for any nonzero ω and η. In this particular instance, the equivalence occurs because we have $r_{2t} = a_{2t}$ in both models.

Such an identifiability problem is serious because, without proper constraints, the likelihood function of a vector ARMA(1,1) model for the data is not uniquely defined.

In the time series literature, methods of structural specification have been proposed to overcome the identifiability problem; see Tiao and Tsay (1989), Tsay (1991), and the references therein.

Today's Topics

1. Multivariate Time Series
2. Vector AR models
3. Vector MA models
4. Vector ARMA models
   - Identifiability Problem
   - VARMA(p,q)

VARMA(p,q)

A VARMA(p, q) model can be written as

$$\mathbf{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \mathbf{\Theta}(B)\boldsymbol{a}_t,$$

where $\mathbf{\Phi}(B) = \boldsymbol{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p$ and $\mathbf{\Theta}(B) = \boldsymbol{I} - \mathbf{\Theta}_1 B - \cdots - \mathbf{\Theta}_q B^q$ are

two $k \times k$ matrix polynomials.

The necessary and sufficient condition of weak stationarity for $\mathbf{r}_t$ is the same as that for the VAR(p) model with matrix polynomial $\mathbf{\Phi}(B)$.

Estimation and Checking

Estimation of a VARMA model can be carried out by either the conditional or exact maximum-likelihood method.

For computation, use the 'varma' function in the R package **MTS**.

The $Q_k(m)$ statistic continues to apply to the residual series of a fitted model, but the degrees of freedom of its asymptotic chi-squared distribution are $k^2m - g$, where g is the number of estimated parameters in both the AR and MA coefficient matrices.