10 Multivariate Time Series Analysis

10.1 Introduction

In this section we look at modeling multiple time series together. This helps us to understand the dependence structure among different assets and different financial markets.

Let $r_t = (r_{1t}, r_{2t}, \ldots, r_{Nt})'$ be a random vector of the returns (or log returns) of N assets at time t. The series r_t is called weakly stationary if its mean vector

$$\mu = \mathbb{E} r_t$$

exists and is independent of time t, and its autocovariance function, which is defined by

$$\Gamma(k) = \mathbb{E}(r_t - \mu)(r_{t-k} - \mu)' = [\mathrm{Cov}(r_{it}, r_{j,t-k)})]_{ij}$$

exists, and is independent of time t. We define the lag-k autocorrelation between assets i and j by

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

and define the autocorrelation function of r_t by

$$\rho(k) = [\rho_{ij}(k)].$$

We may show that $\rho(k) = D^{-1}\Gamma(k)D^{-1}$, where D is the diagnonal matrix of the standard deviations of the individual series. Also, we have $\Gamma(-k) = \Gamma(k)'$ and $\rho(-k) = \rho(k)'$.

The diagonal elements $\rho_{ii}(k)$ are the autocorrelation function of r_{it}. The off-diagonal elements $\rho_{ij}(0)$ measures the concurrent linear relationship between r_{it} and r_{jt}, and the off-diagonal element $\rho_{ij}(k)$ measures the linear dependence of r_{it} on the past value $r_{j,t-k}$.

To estimate the autocorrelation function, we first estimate the autocovariance function by

$$\hat{\Gamma}(k) = \frac{1}{T} \sum_{t=k+1}^{T} (r_t - \hat{\mu})(r_{t-k} - \hat{\mu})',$$

where $\hat{\mu}$ is the sample analogue estimator of the mean μ:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t.$$

Then the autocorrelation function could be estimated by

$$\hat{\rho}(k) = \hat{D}^{-1} \hat{\Gamma}(k) \hat{D}^{-1}$$

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

Vector white noise is a time series of random vectors which is mean zero and serially uncorrelated with a constant zero-order autocovariance. If ε_t is white noise with $\mathrm{Cov}(\varepsilon_t) = \Sigma$, then we write $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$.

10.2 Multivariate Portmanteau Tests

The univariate Ljung-Box Q-statistic could be generalized to the multivariate case. We consider the null hypothesis

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0$$

against the alternative

$$H_1 : \rho(\ell) \neq 0 \quad \text{for some } \ell \in \{1, 2, \ldots, m\}.$$

The test statistic is given by

$$Q(m) = T^2 \sum_{k=1}^{m} \frac{1}{T-k} \operatorname{tr}(\hat{\Gamma}(k)'\hat{\Gamma}(0)^{-1}\hat{\Gamma}(k)\hat{\Gamma}(0)^{-1}).$$

Under the null hypothesis and some regularity conditions,

$$Q(m) \to_d \chi^2_{N^2 m}$$

where N is the dimension of the random vector r_t.

10.3 Vector Autoregressive Models

Just as in the univariate case, we may develop autoregressive models for multivariate time series. The simplest vector autoregressive model is given by the VAR(1) model:

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

where r_t is an N-dimensional vector, Φ is an $N \times N$ matrix, and $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$ is N-dimensional white noise.

In the case when $N = 2$, we have

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

Based on the first equation, Φ_{12} is the linear dependence of r_{1t} on $r_{2,t-1}$ in the presence of $r_{1,t-1}$. Therefore, Φ_{12} may be interpreted as the conditional effect of $r_{2,t-1}$ on r_{1t} given $r_{1,t-1}$.

Assuming that the solution to the VAR(1) model is weakly stationary. Taking expectations on both sides and utilize the properties of weak stationarity, we obtain that

$$\mathbb{E}r_t = (I - \Phi)^{-1}\phi_0,$$

providing that $I - \Phi$ is invertible. And using this relationship, we may write our model as

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

We may iterate backwards the above equation and obtain

$$r_t - \mu = \varepsilon_t + \Phi\varepsilon_{t-1} + \Phi^2\varepsilon_{t-2} + \cdots .$$

For the right hand side of the above equation to converge, we need all the eigenvalues of the matrix Φ to be smaller than one in modulus. Actually, this condition is both necessary and sufficient for the VAR(1) difference equation to have a weakly stationary solution which does not depend on future innovations.

Similarly as in the univariate case, we may obtain the Yule-Walker equation for the VAR(1) by right multiplying $r_{t-k} - \mu'$ on both side of the demeaned VAR(1) model and then take expectations. We obtain

$$\Gamma(k) = \Phi\Gamma(k-1), \quad k \geq 1.$$

Consequently, we have

$$\rho(k) = \Theta\rho(k-1)$$

where $\Theta = D^{-1}\Phi D$.

Now for a general VAR(p) model

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + \cdots + \Phi_p r_{t-p} + \varepsilon_t$$

where $\varepsilon_t \sim \text{WN}(0, \sigma^2)$, if a solution to it is weakly stationary, then the mean of this solution is given by

$$\mu = \mathbb{E}r_t = (I - \Phi_1 - \Phi_2 - \cdots - \Phi_p)^{-1}\phi_0,$$

and the Yule-Walker equations are given by

$$\Gamma(k) = \Phi_1\Gamma(k-1) + \Phi_2\Gamma(k-2) + \cdots + \Phi_p\Gamma(k-p)$$

for $k > 0$, or

$$\rho(k) = \Theta_1\rho(k-1) + \Theta_2\rho(k-2) + \cdots + \Theta_p\rho(k-p),$$

where $\Theta_i = D^{-1}\Phi_i D$.

A VAR(p) model always has a VAR(1) representation:

$$\begin{bmatrix} r_t - \mu \\ r_{t-1} - \mu \\ \vdots \\ r_{t-p+1} - \mu \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} - \mu \\ r_{t-2} - \mu \\ \cdots \\ r_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

For notation convenience, we write the above model as

$$x_t = \Psi x_{t-1} + u_t,$$

where the meaning of x_t, u_t and Ψ should be clear from the context. We may use this representation to derive properties of a VAR(p) model from the corresponding VAR(1) model. For example, a necessary and sufficient condition for the VAR(p) model to have a unique weakly stationary solution that does not depend on the future innovations is that the eigenvalues of the matrix Ψ are all smaller than one in modulus. Some tedious algebra shows that this is also equivalent to that the roots of the equation

$$\det(x^p - x^{p-1}\Phi_1 - x^{p-2}\Phi_2 - \cdots - x\Phi_{p-1} - \Phi_p) = 0$$

lie inside the unit circle on the complex plane.

We may use OLS to estimate the VAR(p) model. To estimate, we write the model as

$$r_t = \Upsilon \tilde{x}_t + \varepsilon_t$$

where $\Upsilon = [\phi_0, \Phi_1, \Phi_2, \ldots, \Phi_p]$ and $\tilde{x}_t = (1, r'_{t-1}, r'_{t-2}, \ldots, r'_{t-p})'$. The OLS estimator of Υ is given by

$$\hat{\Upsilon} = \left(\frac{1}{T} \sum_{t=p+1}^{T} r_t \tilde{x}'_t \right) \left(\frac{1}{T} \sum_{t=p+1}^{T} \tilde{x}_t \tilde{x}'_t \right)^{-1}.$$

We may estimate the residuals by

$$\hat{\varepsilon}_t = r_t - \hat{\phi}_0 - \hat{\Phi}_1 r_{t-1} - \cdots - \hat{\Phi}_p r_{t-p}$$

and estimate the covariance matrix of the innovations by

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=p+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t.$$

The AIC and BIC of the model are given respectively by

$$\text{AIC} = \ln \det(\hat{\Sigma}) + \frac{k^2 p}{T}$$

and

$$\text{BIC} = \ln \det(\hat{\Sigma}) + \frac{2k^2 p \ln T}{T}.$$

We may use AIC or BIC to choose the order p of the model.

10.4 Impulse Response Analysis of VAR Models

A VAR(p) model could be written as a linear function of the past innovations:

$$r_t = \mu + \varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \cdots .$$

This is the moving-average representation of the VAR model. We could easily see that A_i is the impact of the innovation ε_{t-i} on r_t. Therefore, A_i, viewed as a function of the time lag i, is called the impulse response function of r_t.

However, since the components of ε_t are often correlated, the interpretation of elements in A_i is not straightforward. Therefore, we use Cholesky decomposition to construct "structural" innovations such that the components in the innovations are uncorrelated. To be specific, we may uniquely decompose Σ as

$$\Sigma = LL'$$

where L is a lower triangular matrix. We have that $L^{-1}\Sigma L^{-1} = I$. Now we may write

$$r_t = \mu + \sum_{i=1}^{\infty} A_i\varepsilon_{t-i} = \mu + \sum_{i=1}^{\infty} A_i L L^{-1}\varepsilon_{t-i} = \mu + \sum_{i=1}^{\infty} \Pi_i e_{t-i}$$

where $\Pi_i = A_i L$ and $e_t = L^{-1}\varepsilon_t$. It is easy to see that $\mathrm{Var}(e_t) = I$. Therefore, we may interpret $[\Pi_i]_{mn}$ as the impulse of the m-th component of r_t to one standard deviation *structural* shock in the n-th component of r_{t-i}.

10.5 Vector Moving-Average Models

Similarly, we have moving-average models for vector processes. A q-th order moving average model, or VMA(q) for r_t, is given by

$$r_t = \theta_0 + \varepsilon_t + \Theta_1\varepsilon_{t-1} + \cdots + \Theta_q\varepsilon_{t-q},$$

where $\varepsilon_t \sim \mathrm{WN}(0, \Sigma)$. Similarly as in the univariate case, $\mu = \mathbb{E}r_t = \theta_0$, and r_t is always weakly stationary as long as its covariance matrix is well defined.

Once again, this model could be estimated by maximum likelihood estimation, and it should be not difficult to write down the likelihood function of $r_t - \mu$ given that $\varepsilon_t \sim \mathrm{GWN}(0, \Sigma)$. And we may choose the order q by the information criteria.

10.6 Vector Autoregressive Moving-Average Models

Just as in the univariate case, we may combine VAR and VMA to get VARMA model. A general VARMA(p, q) model goes like this:

$$r_t = \phi_0 + \Phi_1 r_{t-1} + \Phi_2 r_{t-2} + \cdots + \Phi_p r_{t-p} + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \cdots + \Theta_q \varepsilon_{t-q}.$$

However, new issues occur when we combine VAR and VMA. One of the most important issues is the identifiability problem. The following is an example of two identical models from Tsay (2010). The first VARMA(1, 1) model is

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

and the second VARMA(1, 1) model is

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

It is easy to show that the two different VARMA models actually give the same data generating process.

In most financial applications, VAR and VMA models are usually sufficient. Due to the issues introduced above, VARMA models are not very popular.