7 Linear Time Series Analysis: Part II

7.1 The Moving-Average Models

Another set of simple models frequently used in financial econometrics is the moving-average (MA) models. We may model the return r_t as an MA(q) process given by

$$r_t = \mu_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

It is easy to see that

$$\mathbb{E} r_t = \mu_0,$$

and

$$\mathrm{Var}(r_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2.$$

Therefore, an MA(q) process is always weakly stationary.

It is also easy to obtain the autocorrelation function $\rho(k)$ of any MA(q) process. We note that for any MA(q) process, $\rho(k) = 0$ for all $k > q$. This is opposed to the AR cases. The MA models are "finite-memory" models.

7.2 Estimation of Moving-Average Models

We usually use the maximum likelihood estimation (MLE) to estimate MA models. Suppose that we observe $r_1, r_2, \ldots, r_T$. To obtain the likelihood function, we need to first impose some assumptions on the distribution of the innovations. For the MA models, we usually assume that the innovations are Gaussian, that is $\varepsilon_t \sim \mathrm{GWN}(0, \sigma^2)$.

We write

$$r = \mu + A\varepsilon$$

where

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_T \end{bmatrix},$$

$$\mu = \begin{bmatrix} \mu_0 \\ \mu_0 \\ \vdots \\ \mu_0 \end{bmatrix},$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{1-q} \\ \varepsilon_{2-q} \\ \vdots \\ \varepsilon_T \end{bmatrix},$$

and

$$A = \begin{bmatrix} \theta_q & \theta_{q-1} & \cdots & \theta_1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & \theta_q & \theta_{q-1} & \cdots & \theta_1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \theta_q & \theta_{q-1} & \cdots & \theta_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \theta_1 & 1 \end{bmatrix}.$$

Since the ε_t's are Gaussian and uncorrelated, they are independent. Then ε is a jointly normal random vector with mean zero and diagonal covariance matrix $\Sigma = \sigma^2 I$, where I is a $(T+q) \times (T+q)$ identity matrix. The random vector, as a linear transformation of ε, is also jointly normal, with mean μ and variance $\Omega = A\Sigma A'$. Now we are able to write down the density function of r:

$$f_r = \frac{1}{(2\pi)^{T/2} \det(\Omega)^{1/2}} \exp\left(-\frac{1}{2}(r-\mu)'\Omega^{-1}(r-\mu) \right).$$

The likelihood function $\mathcal{L}(r;\theta_1,\theta_2,\ldots,\theta_q,\mu,\sigma^2)$ of r, is the density f_r viewed as a function of the parameters $\theta_1,\theta_2,\ldots,\theta_q,\mu,\sigma^2$:

$$\mathcal{L}(r;\theta_1,\theta_2,\ldots,\theta_q,\mu,\sigma^2) = \frac{1}{(2\pi)^{T/2}\det(\Omega)^{1/2}}\exp\left(-\frac{1}{2}(r-\mu)'\Omega^{-1}(r-\mu)\right).$$

The maximum likelihood estimator, is the values $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_q,\hat{\mu},\hat{\sigma}^2$ that maximizes the likelihood function, or the log-likelihood function:

$$\ell(r;\theta_1,\theta_2,\ldots,\theta_q,\mu,\sigma^2) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\ln\det(\Omega) - \frac{1}{2}(r-\mu)'\Omega^{-1}(r-\mu).$$

For algorithms that computes the estimators using computers, see Hamilton (1994, Chapter 5).

7.3 Determining the Moving-Average Order

Since the autocorrelation function $\rho(k)$ for an MA(q) process is zero for all $k > q$, we usually use the sample autocorrelation function (ACF) to determine the order q of the MA model.

7.4 Forecasting

As an exercise, think about how you can forecast an MA(q) process given data $r_1, r_2, \ldots, r_T$.

7.5 The Autoregressive Moving-Average Models

We may combine the autoregressive models and the moving-average models to obtain the autoregressive moving-average (ARMA) models. If we model the time series $\{r_t\}$ as an ARMA(p,q) process, then

$$r_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i r_{t-i} + \varepsilon_t + \sum_{i=1}^{q}\theta_i\varepsilon_{t-i},$$

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

For determinacy, we require that the two polynomial equations

$$x^p - \alpha_1 x^{p-1} - \cdots - \alpha_{p-1} x - \alpha_p = 0$$

and

$$x^q + \theta_1 x^{q-1} + \cdots + \theta_{q-1} x + \theta_q = 0$$

do not have common roots. Otherwise, p and q could be reduced. And similarly as in the AR models, a necessary and sufficient condition for an ARMA(p, q) model to have a unique weakly stationary solution is that each of the roots of the polynomial equation

$$x^p - \alpha_1 x^{p-1} - \cdots - \alpha_{p-1} x - \alpha_p = 0$$

is less than one in modulus.

We usually use the maximum likelihood estimation to estimate weakly stationary ARMA models. The simplest way to calculate the likelihood function of an ARMA process is through the Kalman filter. We shall not go into detail here.

The PACF and ACF does not provide too much information for determining the orders p and q in an ARMA(p, q) model. Actually the identification of the orders is a difficult problem. In practice, information criteria such as AIC and BIC are often used to identify the orders.

7.6 Unit Roots

We now turn to nonstationary time series. The simplest example of a nonstationary time series is a random walk:

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

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$.

It is obvious that a random walk process is not mean-reverting. It contains a stochastic trend. And if we represent r_t using only past innovations, we have

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

Therefore, the effect of any innovation ε_t to r_t is permanent: it not only affects r_t, but will continue to affect $r_{t+1}, r_{t+2}, \ldots$.

Now we consider unit roots in a more general ARMA context. We know that the MA part is always weakly stationary, while the weak stationarity of the AR part depends on the modulus of the roots of the polynomial equation corresponding to the AR part. Suppose that the AR polynomial equation has roots with modulus less than one, and also has roots one. Now this difference equation does not have a weakly stationary solution. However, it has a nonstationary solution that can be made weakly stationary by differencing. That is, there is a solution $\{r_t\}$ such that $\{(1-L)^{d-1}r_t\}$ is not weakly stationary but $\{(1-L)^d r_t\}$ is weakly stationary for some positive integer d, where L is the lag operator defined by

$$Lr_t = r_{t-1},$$

and

$$(1-L)^d = (1-L)[(1-L)^{d-1}].$$

Such a process is called an autoregressive integrated moving-average (ARIMA(p,d,q)) process.

A common way to deal with such a process is to difference the original time series. If r_t is an ARIMA(p,d,q) process, then $(1-L)^d r_t$ is an ARMA(p,q) process. Therefore, once we have a nonstationary process, we may try to difference it first, and then model it using an ARMA(p,q) model.

Although there are processes in reality that are integrated of higher orders, most of the time the nonstationary time series we model in finance are integrated of order one. That is, $d = 1$. For these time series, a first difference is enough. For example, prices $\{P_t\}$ of financial assets are usually believed to be nonstationary, while the log returns $\{\ln P_t - \ln P_{t-1}\}$ are modeled as weakly stationary. Here we mainly consider the case in which $d = 1$.

Before we start to model a time series, we need to test whether a time series contains a unit root. We first look at a simple case. Suppose that we have a series $r_0, r_1, \ldots, r_T$ that is generated from

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

where $\varepsilon_t \sim \text{WN}(0, \sigma^2)$. To detect nonstationarity in the series, we run the regression

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

and test the null hypothesis $H_0 : \beta = 1$ against the alternative $H_1 : |\beta| < 1$. The OLS estimator of β is given by

$$\hat{\beta} = \left(\frac{1}{T} \sum_{t=1}^{T} r_{t-1}^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} r_{t-1} r_t \right).$$

We have that

$$\hat{\beta} - 1 = \left(\frac{1}{T} \sum_{t=1}^{T} r_{t-1}^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} r_{t-1} \varepsilon_t \right).$$

It is easy to see that if we treat r_0 to be deterministic, we have that

$$\text{Var}(r_t) = t\sigma^2.$$

That is, the variance of r_t is getting larger as t gets larger. As a consequence, we do not have the usual LLN and CLT for

$$\frac{1}{T} \sum_{t=1}^{T} r_{t-1}^2$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} r_{t-1} \varepsilon_t$$

respectively any more. Instead, we have

$$\frac{1}{T^2} \sum_{t=1}^{T} r_{t-1}^2 \to_d \int_0^1 W^2(r) \mathrm{d}r,$$

$$\frac{1}{T} \sum_{t=1}^{T} r_{t-1}^2 \to_d \int_0^1 W(r) \mathrm{d}W(r),$$

and

$$T(\hat{\beta} - \beta) \to_d \frac{\int_0^1 W(r) \mathrm{d}W(r)}{\int_0^1 W^2(r) \mathrm{d}r},$$

where $W(r)$ is Brownian motion on $[0, 1]$ with $\mathrm{Var}(W(1)) = \sigma^2$, and the second integral is the Ito integral. Critical values of the asymptotic distribution could be obtained by Monte Carlo simulations. We may also get the asymptotic distribution of the t-statistic:

$$t(\beta) = \frac{\hat{\beta} - 1}{\mathrm{s.e.}(\hat{\beta})} \to_d \frac{\int_0^1 W(r) \mathrm{d}W(r)}{\left(\int_0^1 W^2(r) \mathrm{d}r \right)^{1/2}}.$$

Note here that the convergence rate of the OLS estimator in this non-stationary case is T, which is faster than the usual $\sqrt{T}$ rate in the iid or the stationary case. And the limit distribution is also different.

Now we consider a more general case. Suppose that

$$\Delta r_t = \alpha_0 + \alpha_1 \Delta r_{t-1} + \alpha_2 \Delta r_{t-2} + \cdots + \alpha_p \Delta r_{t-p} + \varepsilon_t$$

where $\Delta = 1 - L$ is the difference operator, and $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$. We may run the regression

$$r_t = \alpha_0 + \beta r_{t-1} + \alpha_1 \Delta r_{t-1} + \alpha_2 \Delta r_{t-2} + \cdots + \alpha_p \Delta r_{t-p} + \varepsilon_t$$

and test the null hypothesis $H_0 : \beta = 1$ against the alternative $H_1 : |\beta| < 1$. It turns out that the asymptotic distribution of the t-statistic for the OLS estimator of β is still given by the Dickey-Fuller distribution:

$$t(\beta) = \frac{\hat{\beta} - 1}{\text{s.e.}(\hat{\beta})} \to_d \frac{\int_0^1 W(r)\mathrm{d}W(r)}{\left(\int_0^1 W^2(r)\mathrm{d}r\right)^{1/2}}.$$

We may further generalize our setting to allow the error term to be serially correlated instead of being a white noise. Interested readers may refer to Hamilton (1994, Chapter 17).

7.7 Trend and Seasonal Components

Besides stochastic trends that comes from unit roots, there could be deterministic trends and seasonal components that makes the series non-stationary. In either case, we first need to transform our series to stationary ones.

7.7.1 Deterministic Trends

If we know the form of the trend, we may use OLS to eliminate the trend component. Suppose we want to study Y_t and

$$Y_t = m_t + X_t$$

where X_t is a weakly stationary process and m_t represents a deterministic trend. If we know that $m_t = \alpha_0 + \alpha_1 t$, we may use OLS to estimate

$$Y_t = \alpha_0 + \alpha_1 t + u_t,$$

and recover X_t by

$$Y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t$$

where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the OLS estimated coefficients.

In the above example, we could also do first differencing. Notice that $\Delta Y_t = \alpha_1 + \Delta X_t$, which is weakly stationary. However, we have to bear in mind that when we difference Y_t, we at the same time difference X_t, and information about X_t could be lost in this differencing procedure.

Another way to estimate the trend component m_t is to use two sided moving averages. We may consider to estimate m_t by

$$\hat{m}_t = \sum_{k=-q}^{q} w_k Y_{t+k}$$

for some positive integer q and weights w_k. For example, one may choose $w_k = \frac{1}{2q+1}$ for all k. Then one may recover the stationary component X_t by $Y_t - \hat{m}_t$.

7.7.2 Seasonality

Suppose we want to study Y_t and

$$Y_t = s_t + X_t$$

where X_t is a weakly stationary process and s_t is a seasonal component. One way to deseasonalize is to first estimate the seasonal component using season-wise means of the data and then subtract the seasonal component from the series. For example, if you have data that is at monthly frequency, you may estimate the January mean as the mean of all the January observations, the February mean as the mean of all the February observations.

Another approach to eliminate the seasonal component is through differencing at lag-s where s is the period. The lag-s difference operator Δ_s is defined by

$$\Delta_s Y_t = (1 - L^s) Y_t = Y_t - Y_{t-s}.$$

For example, if we have monthly data, $s = 12$. Note also that by lag-s differencing, we not only eliminate the seasonal component, but also difference the stationary component $\{X_t\}$, in which process information could be lost.