6 Linear Time Series Analysis: Part I

6.1 Introduction

Studies have suggested that many financial return series are serially correlated. In this chapter, we look at a simple class of models that takes account of this correlation. To be specific, we look at models that attempt to capture the linear relationship between return at time t and information available prior to time t. Returns at different time in a series interact through autocovariance or autocorrelation.

We have encountered the concepts of autocovariance and autocorrelation in the previous chapter. There, we work with strictly stationary time series. Our framework in this chapter could be developed for a larger class of time series, namely the weakly stationary time series.

Definition 6.1. A sequence of random variables $\{X_t\}_{t\in\mathbb{Z}}$ is called weakly stationary if $\mathbb{E}X_t = \mathbb{E}X_{t+\tau}$ and $\mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{Cov}(X_{t_1+\tau}, X_{t_2+\tau})$ for all t, t_1, t_2 and $\tau \in \mathbb{Z}$.

Similarly as in the case of strictly stationary time series, we may define the autocovariance and autocorrelation functions for weakly stationary time series. These functions, again, are only functions of the time difference of the two random variables considered. All properties we introduced in the previous chapter for the autocovariance and autocorrelation functions continue to hold.

We note here that neither strict stationarity nor weak stationarity implies each other. For example, a strictly stationary time series may have infinite variance and therefore fails to be weakly stationary. A weakly stationary time series may not be identically distributed and therefore fails to be strictly stationary. However, if a time series is strictly stationary, and at the same time has finite mean and variance, then it is weakly stationary.

There is a special class of weakly stationary time series that is widely used in multiple disciplines.

Definition 6.2. A sequence of random variables $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is called white noise if $\mathbb{E}\varepsilon_t = 0$ for all t, and

$$\text{Cov}(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2, & \text{if } t = s, \\ 0, & \text{if } t \neq s. \end{cases}$$

for some $\sigma^2 > 0$.

We sometimes write $\varepsilon_t \sim \text{WN}(0, \sigma^2)$. If each ε_t is normally distributed, then this series is called Gaussian white noise. Notice that terms in Gaussian white noise are not only serially uncorrelated, but independent.

A time series $\{r_t\}$ is called linear if it can be written as

$$r_t = \mu + \sum_{t=0}^{\infty} \phi_i \varepsilon_{t-i}$$

where $\{\varepsilon_t\}$ is white noise, $\mu, \phi_0, \phi_1, \ldots$ are parameters and $\alpha_0 = 1$. Usually, ε_t represents new information at time t and is therefore often referred to as the innovation at time t.

6.2 The Autoregressive Models

6.2.1 The AR(1) Model

The simplest model that utilizes the first order autocorrelation is the autoregressive (AR) model of order one, or simply, the AR(1) model:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \varepsilon_t \tag{6.1}$$

where ε_t is white noise with variance σ^2.

We first look at under what conditions $\{\varepsilon_t\}$ is weakly stationary. Taking expectations on both sides, we have

$$\mathbb{E}r_t = \alpha_0 + \alpha_1 \mathbb{E}r_{t-1}.$$

If $\{r_t\}$ is weakly stationary, we have $\mathbb{E}r_t = \mathbb{E}r_{t-1}$. This implies that

$$\mathbb{E}r_t = \frac{\alpha_0}{1-\alpha_1} := \mu.$$

We may rewrite our original model as

$$r_t - \mu = \alpha_1(r_{t-1} - \mu) + \varepsilon_t. \tag{6.2}$$

This gives us a difference equation. There are multiple solutions $\{r_t\}$ to this equation. However, not all of them are weakly stationary.

Let us first try to solve it backwards. As a result,

$$r_t - \mu = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_1^2\varepsilon_{t-2} + \cdots = \sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}.$$

For this solution to make sense, we need to guarantee that the series in the above expression converges in some (probabilistic) sense. At this moment, we assume that this is the case. Then $r_t - \mu$ is a function of $\varepsilon_t, \varepsilon_{t-1}, \ldots$. Similarly, we should have that $r_{t-1} - \mu$ is a function of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. This implies that $\mathbb{E}(r_{t-1}-\mu)\varepsilon_t = 0$ since $\{\varepsilon_t\}$ is white noise and thus serially uncorrelated. Then we can take the variances of both sides of equation (6.2) and obtain

$$\text{Var}(r_t - \mu) = \alpha_1^2 \text{Var}(r_{t-1} - \mu) + \text{Var}(\varepsilon_t),$$

or equivalently,

$$\text{Var}(r_t) = \alpha_1^2 \text{Var}(r_{t-1}) + \sigma^2.$$

If $\{r_t\}$ is weakly stationary, then $\text{Var}(r_t) = \text{Var}(r_{t-1})$. Then we have

$$\text{Var}(r_t) = \frac{\sigma^2}{1-\alpha_1^2}.$$

In order for $\text{Var}(r_t)$ to be a variance, we need

$$|\alpha_1| < 1.$$

Otherwise, the variance is either infinity, or negative.

It turns out that if $|\alpha_1| < 1$, the series $\sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}$ converges almost surely. That is, if our sample space is given by $(\Omega, \mathcal{F}, \mathbb{P})$, then the collection of sample points $\omega \in \Omega$ such that $\sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}(\omega)$ does not converge has probability at most zero. That is, $\sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}$ converges almost in all cases.

Now we have found a condition, namely $|\alpha_1| < 1$, such that (6.2) has a weakly stationary solution, which is given by

$$r_t = \mu + \sum_{i=0}^{\infty} \alpha_1^i \varepsilon_{t-i}.$$

Actually, it could be proved that this is the only weakly stationary solution.

The case where $|\alpha_1| = 1$ is complicated. We shall consider this case in a later section.

In the case where $|\alpha_1| > 1$, the only weakly stationary solution is given by

$$r_t = \mu_t + \sum_{i=0}^{\infty} \left(\frac{1}{\alpha_1}\right)^i \varepsilon_{t+i}.$$

However, in this case, the current value of r_t depends on the future values of the innovations. (Not even the expectations of future innovations!) This is something counter-intuitive, and therefore we usually do not model financial time series in this way.

Now we assume that $|\alpha_1| < 1$. Multiply both sides of (6.2) by $r_{t-k} - \mu$ for any $k \geq 1$ and take expectations, we get

$$\mathbb{E}(r_t - \mu)(r_{t-k} - \mu) = \alpha_1 \mathbb{E}(r_{t-1} - \mu)(r_{t-k} - \mu) + \mathbb{E}\varepsilon_t(r_{t-k} - \mu),$$

or

$$\gamma(k) = \alpha_1 \gamma(k-1), \tag{6.3}$$

or

$$\rho(k) = \alpha_1 \rho(k-1), \tag{6.4}$$

where $\gamma(k)$ and $\rho(k)$ are the autocovariance and autocorrelation functions of the return series $\{r_t\}$. Using these relationships, starting from $\gamma(0)$ and $\rho(0)$, we can get $\gamma(k)$ and $\rho(k)$ for any k. Equations (6.3) or (6.4) is usually called the Yule-Walker equation of the AR(1) process (6.2).

6.2.2 The AR(p) Model

We may generalize our AR(1) model (6.1) so that $\{r_t\}$ is a p-th order autoregressive process:

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \cdots + \alpha_p r_{t-p} + \varepsilon_t \tag{6.5}$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2.

It could be shown that a necessary and sufficient condition for the difference equation (6.5) to have a weakly stationary solution that does not depend on the future values of $\{\varepsilon_t\}$ is that all the roots of the polynomial equation

$$x^p - \alpha_2 x^{p-1} - \alpha_2 x^{p-2} - \cdots - \alpha_p = 0$$

are smaller than one in modulus (or, all the roots lie in the unit circle on the complex plane). We assume that this is the case.

When $\{r_t\}$ is weakly stationary, using similar techniques, we may show that

$$\mathbb{E} r_t = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p}.$$

With a slight abuse of notation, we denote the expectation again by μ. Then we may write the AR(p) model (6.5) in its demeaned form:

$$r_t - \mu = \alpha_1(r_{t-1} - \mu) + \alpha_2(r_{t-2} - \mu) + \ldots + \alpha_p(r_{t-p} - \mu) + \varepsilon_t.$$

The Yule-Walker equation can be obtained by multiply both sides of the above equation by $r_{t-k} - \mu$ for some $k \geq 1$ and then take expectations:

$$\gamma(k) = \alpha_1\gamma(k-1) + \alpha_2\gamma(k-2) + \cdots + \alpha_p\gamma(k-p).$$

Together with the relationships $\gamma(k) = \gamma(-k)$, one is able to obtain $\rho(k)$ for all k.

6.3 Estimation of Autoregressive Models

Suppose that we have observations $r_1, r_2, \ldots, r_T$ and have determined to model the series of returns with an AR(p) model (6.5). We may estimate the model by ordinary least squares. For notation convenience, let $x_t = (1, r_{t-1}, r_{t-2}, \ldots, r_{t-p})'$ and $\beta = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_p)'$. Then we may write our model as

$$r_t = x_t'\beta + \varepsilon_t.$$

The OLS estimator of β is obtained by minimizing

$$L(\beta) = \frac{1}{T_p}\sum_{t=p+1}^{T}(r_t - x_t'\beta)^2.$$

where $T_p = T - p$. The first order condition, which is a necessary condition, is given by

$$\frac{\partial L(\beta)}{\partial \beta} = -\frac{1}{T_p}\sum_{t=p+1}^{T} x_t(r_t - x_t'\beta) = 0.$$

Solving the equation above, we obtain the OLS estimator of β:

$$\hat{\beta} = \left(\frac{1}{T_p} \sum_{t=p+1}^{T} x_t x_t' \right)^{-1} \left(\frac{1}{T_p} \sum_{t=p+1}^{T} x_t r_t \right).$$

The OLS estimator is well defined if and only if the matrix

$$\hat{\Sigma}_x = \frac{1}{T_p} \sum_{t=p+1}^{T} x_t x_t'$$

is invertible. That is, $\hat{\Sigma}_x$ is of full rank. This is usually the case in data since it is an estimator of the variance of the random vector x_t. As long as there is no multicollinearity issue in x_t, we shall see that $\hat{\Sigma}_x$ is of full rank in practice.

The OLS estimator could also be viewed as a sample analogue estimator. To see this, we multiply both sides of our model by x_t and take expectations. Notice that ε_t is uncorrelated with each of the terms in x_t, we have that

$$\mathbb{E} x_t r_t = \mathbb{E} x_t x_t' \beta.$$

If $\mathbb{E} x_t x_t'$ is of full rank, we can solve for β:

$$\beta = (\mathbb{E} x_t x_t')^{-1} \mathbb{E} x_t r_t.$$

Notice that we may estimate $\mathbb{E} x_t x_t'$ by its sample analogue estimator

$$\frac{1}{T_p} \sum_{t=p+1}^{T} x_t x_t'$$

and estimate $\mathbb{E} x_t r_t$ by its sample analogue estimator

$$\frac{1}{T_p} \sum_{t=p+1}^{T} x_t r_t.$$

Then $\hat{\beta}$ could be viewed as the sample analogue estimator of β.

To obtain the asymptotic distribution of the OLS estimator $\hat{\beta}$, write

$$\begin{aligned}
\sqrt{T_p}(\hat{\beta}-\beta) &= \sqrt{T_p}\left[\left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t r_t\right) - \beta\right] \\
&= \sqrt{T_p}\left[\left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t (x_t'\beta + \varepsilon_t)\right) - \beta\right] \\
&= \left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{T_p}}\sum_{t=p+1}^{T} x_t \varepsilon_t\right) \\
&= A^{-1}B.
\end{aligned}$$

Under some regularity conditions, we have that a Law of Large Numbers holds for A:

$$A = \frac{1}{T_p}\sum_{t=p+1}^{T} x_t x_t' \to_p \mathbb{E}x_t x_t'.$$

Let $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$. If $\{\varepsilon_t\}$ is a martingale difference sequence (note that this is stronger than serial uncorrelatedness), then $\mathbb{E}(x_t\varepsilon_t|\mathcal{F}_{t-1}) = x_t\mathbb{E}(\varepsilon_t|\mathcal{F}_t) = 0$. That is, $\{x_t\varepsilon_t\}$ is a martingale difference sequence. Then under some regularity conditions, a Central Limit Theorem holds for B:

$$B = \frac{1}{\sqrt{T_p}}\sum_{t=p+1}^{T} x_t\varepsilon_t \to_d \mathbb{N}\big(0, \mathbb{E}x_t x_t'\sigma^2\big).$$

Then

$$\sqrt{T_p}(\hat{\beta}-\beta) = A^{-1}B \to_d \mathbb{N}\big(0, (\mathbb{E}x_t x_t')^{-2}\mathbb{E}x_t x_t'\sigma^2\big) =_d \mathbb{N}\big(0, (\mathbb{E}x_t x_t')^{-1}\sigma^2\big).$$

The standard error of the estimates could be obtained as the square roots of the diagonal elements of $(\mathbb{E}x_t x_t')^{-1}\sigma^2$, which in turn could be estimated by

$$\left(\frac{1}{T_p}\sum_{t=p+1}^{T} x_t x_t'\right)^{-1}\frac{1}{T_p}\sum_{t=p+1}^{T}\hat{\varepsilon}_t^2$$

where the $\hat{\varepsilon}_t$'s are the fitted residuals obtain through

$$\hat{\varepsilon}_t = r_t - x_t'\hat{\beta}.$$

Note that we have estimated σ^2 using

$$\hat{\sigma}^2 = \frac{1}{T_p}\sum_{t=p+1}^{T}\hat{\varepsilon}_t^2.$$

We may also use the unbiased estimator for σ^2 given by

$$\frac{1}{T-2p-1}\sum_{t=p+1}^{T}\hat{\varepsilon}_t^2.$$

Asymptotically, they are equivalent.

6.4 Determining the Autoregressive Order

Before we start to estimate the model, we need to determine the autoregressive order p. There are two popular approaches. One is to utilize the partial autocorrelation function, and the other is to apply some information criteria.

6.4.1 Partial Autocorrelation Function

The lag-k partial autocorrelation function (PACF) of a weakly stationary time series measures the contribution of adding the term r_{t-k} over an AR($k-1$) model. The lag-k PACF could be estimated by the OLS estimator $\hat{\alpha}_{kk}$ of the AR(k) model

$$r_t = \alpha_{k0} + \alpha_{k1}r_{t-1} + \cdots + \alpha_{kk}r_{t-k} + \varepsilon_{kt}.$$

If the data is generated by an AR(p) model, $\hat{\alpha}_{pp}$ should not be zero, but $\hat{\alpha}_{kk}$ should be close to zero for all $k > p$. Therefore, we may obtain $\hat{\alpha}_{kk}$ for $k = 1, 2, \ldots, P$ where P is some reasonably large number, and test whether the PACFs are zero.

It has been shown that under regularity conditions, for a weakly stationary Gaussian AR(p) model, $\hat{\alpha}_{pp} \to_p \alpha_{pp}$, while $\sqrt{T}\hat{\alpha}_{kk} \to_d \mathbb{N}(0,1)$ for $k > p$ as $T \to \infty$.

6.4.2 Information Criteria

There are two popular information criteria for model selection: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

For a Gaussian AR(k) model, the AIC is defined to be

$$\text{AIC}(k) = \ln \tilde{\sigma}_k^2 + \frac{2k}{T}$$

where T is the sample size and

$$\tilde{\sigma}_k^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2.$$

The first term measures the goodness of fit of the model and the second term is a penalty term which increase as the number of parameters increase. The rule is to first calculate AIC(k) for $k = 0, 1, \ldots, P$ and then choose the k that minimizes the AIC.

The Bayesian information criteria uses a different penalty term:

$$\text{BIC}(k) = \ln \tilde{\sigma}_k^2 + \frac{k \ln T}{T}.$$

The selection rule is the same.

6.5 Forecasting

An important task in the study of financial time series is to make predictions. Suppose that we have observations $r_1, r_2, \ldots, r_T$ and we know that these observations are generated from an AR(p) model (6.5). Now we want to predict r_{T+h} for $h \geq 1$. Suppose that $\{\varepsilon_t\}$ is a martingale difference sequence and $\mathcal{G}_t = \sigma(r_t, r_{t-1}, r_{t-2}, \ldots)$. Given $\mathcal{G}_T$, the "best" forecast we can make for r_{T+1} is its conditional expectation with respect to all available information, namely $\mathcal{G}_T$. Notice that

$$r_{T+1} = \alpha_0 + \alpha_1 r_T + \cdots + \alpha_p r_{T-p+1} + \varepsilon_{T+1},$$

we have

$$\hat{r}_{T+1} = \mathbb{E}(r_{T+1}|\mathcal{G}_T) = \alpha_0 + \alpha_1 r_T + \alpha_2 r_{T-1} \cdots + \alpha_p r_{T-p+1}.$$

In practice, we replace the α_i's using their estimated values.

Similarly, we have

$$\hat{r}_{T+2} = \mathbb{E}(r_{T+2}|\mathcal{G}_T) = \alpha_0 + \alpha_1 \hat{r}_{T+1} + \alpha_2 r_T + \cdots + \alpha_p r_{T-p+2}.$$

We may repeat the above steps to get multi-step ahead forecasts.