Ch2 Linear Time Series Analysis

The contents of this manuscript are summarized from "Tsay, R. S. (2010). Analysis of Finacial Time Series (3rd edition). John Wiley & Sons, Inc, New Jersey."

- In this chapter, we discuss basic theories of linear time series analysis, introduce some simple econometric models useful for analyzing financial data, and apply the models to financial time series such as asset returns.

- Treating an asset return (e.g., log return r_t of a stock) as a collection of random variables over time, we have a time series $\{r_t\}$.

- Linear time series analysis provides a natural framework to study the dynamic structure of such a series.

- The theories of linear time series discussed include stationary, dynamic dependence, autocorrelation function, modeling, and forecasting.

- For an asset return r_t, simple models attempt to capture the linear relationship between r_t and information available prior to time t.

- The information may contain the historical values of r_t and a random vector Y, which is a state vector consisting of variables that summarize the environment in which asset returns are determined.

- As such, correlation plays an important role in understanding these models.

- In particular, correlations between the variable of interest and its past values become the focus of linear time series analysis.

- These correlations are referred to as serial correlations or autocorrelations. They are the basic tool for studying a stationary time series.

- **Stationarity**
  - The foundation of time series analysis is stationarity.
  - A time series $\{r_t\}$ is said to be strictly stationary if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers.
  - In other words, strict stationarity requires that the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is invariant under time shift. This is a very strong condition that is hard to verify empirically.
  - A time series $\{r_t\}$ is weakly stationary if both the mean of r_t and the covariance between r_t and $r_{t-\ell}$ are time invariant, where ℓ is an arbitrary integer.
  - More specifically, $\{r_t\}$ is weakly stationary if
    (a) $\mathrm{E}(r_t) = \mu$, which is a constant, and
    (b) $\mathrm{Cov}(r_t, r_{t-\ell}) = \gamma_\ell$, which only depends on ℓ.
  - In practice, suppose that we have observed T data points $\{r_t \mid t = 1, \ldots, T\}$. The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level.
  - In applications, weak stationary enables one to make inference concerning future observations.
  - In this book, we are mainly concerned with weakly stationary series.
  - The covariance $\gamma_\ell = \mathrm{Cov}(r_t, r_{t-\ell})$ is called the lag-ℓ autocovariance of r_t. It has two properties:
    (a) $\gamma_0 = \mathrm{Var}(r_t)$ and
    (b) $\gamma_{-\ell} = \gamma_\ell$.

- In the finance literature, it is common to assume that an asset return series is weakly stationary.

- This assumption can be checked empirically provided that a sufficient number of historical returns are available. For example, one can divide the data into subsamples and check the consistency of the results obtained across the subsamples.

- **Autocorrelation Function (ACF)**

  - The correlation coefficient between r_t and $r_{t-\ell}$ is called the lag-ℓ autocorrelation of r_t and is commonly denoted by ρ_ℓ, which under the weak stationarity assumption is a function of ℓ only.

  - Specifically, we define
$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0} \tag{1}$$

  - From the definition, we have $\rho_0 = 1$, $\rho_\ell = \rho_{-\ell}$, and $-1 \leq \rho_\ell \leq 1$.

  - In addition, a weakly stationary series r_t is not serially correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.

  - For a given sample of returns $\{r_t\}_{t=1}^T$, let $\bar{r}$ be the sample mean. Then the lag-ℓ sample autocorrelation function of r_t is defined as
$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq \ell < T - 1 \tag{2}$$

  - If $\{r_t\}$ is an iid sequence satisfying $\mathrm{E}(r_t^2) < \infty$, then $\hat{\rho}_\ell$ is asymptotically normal with mean zero and variance $1/T$ for any fixed positive integer ℓ.

  - More generally, if r_t is weakly stationary time series satisfying $r_t = \mu + \sum_{i=0}^q \psi_i a_{t-i}$, where $\psi_0 = 1$ and $\{a_j\}$ is a sequence of iid random variables with mean zero, then $\hat{\rho}_\ell$ is asymptotically normal with mean zero and variance $(1 + 2\sum_{i=1}^q \rho_i^2)/T$ for $\ell > q$.

- **Testing Individual ACF**

  - For a given positive integer ℓ, the previous result can be used to test $H_0 : \rho_\ell = 0$ vs. $H_1 : \rho_\ell \neq 0$. The test statistic is

$$t \text{ ratio} = \frac{\hat{\rho}_\ell}{\sqrt{(1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_i^2)/T}}$$

  - If $\{r_t\}$ is a stationary Gaussian series satisfying $\rho_j = 0$ for $j > \ell$, the t ratio is asymptotically distributed as a standard normal random variable.

  - Hence, the decision rule of the test is to reject H_0 if t ratio $> Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard normal distribution.

  - For simplicity, many software packages use $1/T$ as the asymptotic variance of $\hat{\rho}_\ell$ for all $\ell \neq 0$. They essentially assume that the underlying time series is an iid sequence.

  - In finite samples, $\hat{\rho}_\ell$ is a biased estimator of ρ_ℓ. The bias is in the order of $1/T$, which can be substantial when the sample size T is small. In most financial applications, T is relatively large so that the bias is not serious.

- **Portmanteau Test**

  - Financial applications often require to test jointly that several autocorrelations of r_t are zero.

  - Box and Pierce (1970) propose the Portmanteau statistic

$$Q^*(m) = T \sum_{\ell=1}^{m} \hat{\rho}_\ell^2$$

    as a test statistic for the null hypothesis $H_0 : \rho_1 = \cdots = \rho_m = 0$ against alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$.

- Under the assumption that $\{r_t\}$ is an iid sequence with certain moment conditions, $Q^*(m)$ is asymptotically a chi-squared random variable with m degrees of freedom.

- Ljung and Box (1978) modify the $Q^*(m)$ statistic as below to increase the power of the test in finite samples,

$$Q(m) = T(T+2)\sum_{\ell=1}^{m}\frac{\hat{\rho}_\ell^2}{T-\ell} \tag{3}$$

- The decision rule is to reject H_0 if $Q(m) > \chi^2_\alpha$, where χ^2_α denotes the $100(1-\alpha)$th percentile of a chi-squared distribution with m degrees of freedom.

- In practice, the choice of m may affect the performance of the $Q(m)$ statistic. Several values of m are often used. Simulation studies suggest that the choice of $m \approx \ln T$ provides better power performance.

- This general rule needs modification in analysis of seasonal time series for which autocorrelations with lags at multiples of the seasonality are more important.

- The statistics $\hat{\rho}_1, \hat{\rho}_2, \ldots$ defined in (2) is called the sample ACF of r_t. It plays an important role in linear time series analysis.

- As a matter of fact, a linear time series model can be characterized by its ACF, and linear time series modeling makes use of the sample ACF to capture the linear dynamic of the data.

- See Figures 2.1 and 2.2 on pages 33-34.

- In the finance literature, a version of the capital asset pricing model (CAPM) theory is that the return $\{r_t\}$ of an asset is not predictable and should have no autocorrelations.

- Testing for zero autocorrelations has been used as a tool to check the efficient market assumption.

- However, the way by which stock prices are determined and index returns are calculated might introduce autocorrelations in the observed return data. This is particularly so in analysis of high-frequency financial data.

- **White Noise**
  - A time series $\{r_t\}$ is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance.

  - In particular, if r_t is normally distributed with mean zero and variance σ^2, the series is called a Gaussian white noise.

  - For a white noise series, all the ACFs are zero.

  - In practice, if all sample ACFs are close to zero, then the series is a white noise series.

- **Linear Time Series**
  - A time series r_t is said to be linear if it can be written as

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{4}$$

    where μ is the mean of r_t, $\psi_0 = 1$, and $\{a_t\}$ is a sequence of iid random variables with mean zero and a well-defined distribution (i.e., $\{a_t\}$ is a white noise series).

  - a_t denotes the new information at time t of the time series and is often referred to as the innovation or shock at time t.

– For a linear time series in (4), the dynamic structure of r_t is governed by the coefficients ψ_i, which are called the ψ weights of r_t in the time series literature.

– If r_t is weakly stationary, we can obtain its mean and variance easily by using the independence of $\{a_t\}$ as

$$\mathrm{E}(r_t) = \mu, \quad \mathrm{Var}(r_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 \tag{5}$$

where σ_a^2 is the variance of a_t.

– Because $\mathrm{Var}(r_t) < \infty$, $\{\psi_i^2\}$ must be a convergent sequence, that is, $\psi_i^2 \to 0$ as $i \to \infty$.

– Consequently, for a stationary series, impact of the remote shock a_{t-i} on the return r_t vanishes as i increases.

– The lag-ℓ autocovariance of r_t is

$$\begin{aligned}
\gamma_\ell &= \mathrm{Cov}(r_t, r_{t-\ell}) = \mathrm{E}\Big\{\Big(\sum_{i=0}^{\infty}\psi_i a_{t-i}\Big)\Big(\sum_{j=0}^{\infty}\psi_j a_{t-\ell-j}\Big)\Big\} \\
&= \mathrm{E}\Big\{\Big(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j a_{t-i}a_{t-\ell-j}\Big)\Big\} \\
&= \sum_{j=0}^{\infty}\psi_{j+\ell}\psi_j \mathrm{E}(a_{t-\ell-j}^2) = \sigma_a^2\sum_{j=0}^{\infty}\psi_j\psi_{j+\ell}
\end{aligned} \tag{6}$$

– Consequently, the ψ weights are related to the autocorrelations of r_t as follows:

$$\rho_\ell = \frac{\sum_{i=0}^{\infty}\psi_i\psi_{i+\ell}}{1+\sum_{i=1}^{\infty}\psi_i^2}, \quad \ell \geq 0 \tag{7}$$

– Linear time series models are econometric and statistical models used to describe the pattern of the ψ weights of r_t.

- For a weakly stationary time series, $\psi_i \to 0$ as $i \to \infty$ and, hence, ρ_ℓ converges to zero as ℓ increases.

- For asset returns, this means that, as expected, the linear dependence of current return r_t on the remote past return $r_{t-\ell}$ diminishes for large ℓ.

- **Simple AR Models**

  - If the return series $\{r_t\}$ has a statistically significant lag-1 autocorrelation, then the lagged return r_{t-1} might be useful in predicting r_t.

  - A simple model that makes use of such predictive power is

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \tag{8}$$

    where $\{a_t\}$ is assumed to be a white noise series with mean zero and variance σ_a^2.

  - This model is in the same form as the well-known simple linear regression model in which r_t is the dependent variable and r_{t-1} is the explanatory variable.

  - In the time series literature, model (8) is referred to as an autoregressive (AR) model of order 1 or simply an AR(1) model.

  - The AR(1) model has several properties similar to those of the simple linear regression model. However, there are some significant differences between the two models.

  - Note that an AR(1) model implies that, conditional on the past return r_{t-1}, we have

$$\mathrm{E}(r_t \mid r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \quad \mathrm{Var}(r_t \mid r_{t-1}) = \mathrm{Var}(a_t) = \sigma_a^2$$

    That is, given the past return r_{t-1}, the current return is centered around $\phi_0 + \phi_1 r_{t-1}$ with standard deviation σ_a.

- This is a Markov property such that conditional on r_{t-1}, the return r_t is not correlated with r_{t-i} for $i > 1$.

- Obviously, there are situations in which r_{t-1} alone cannot determine the conditional expectation of r_t and a more flexible model must be sought.

- A straightforward generalization of the AR(1) model is the AR(p) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_o r_{t-p} + a_t \tag{9}$$

  where p is a nonnegative integer and $\{a_t\}$ is defined in (8).

- The AR(p) model is in the same form as a multiple linear regression model with lagged values serving as the explanatory variables.

• **Properties of AR (1) Model**

- We begin with the sufficient and necessary condition for weak stationarity of the AR(1) model in (8).

- Mean:

$$\mathrm{E}(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}$$

  This result has two implications for r_t. First, the mean of r_t exists if $\phi_1 \neq 1$. Second, the mean of r_t is zero if and only if $\phi_0 = 0$.

- Next, using $\phi_0 = (1 - \phi_1)\mu$, the AR(1) model can be rewritten as

$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i a_{t-i} \tag{10}$$

  This equation expresses an AR(1) model in the form of (4) with $\psi_i = \phi_1^i$. Thus, $r_t - \mu$ is a linear function of a_{t-i} for $i \geq 0$.

- Variance:

$$\mathrm{Var}(r_t) = \frac{\sigma_a^2}{1 - \phi_1^2}$$

  provided that $\phi_1^2 < 1$. Consequently, the weak stationarity of an AR(1) model implies that $|\phi_1| < 1$.

– Autocovariance:

$$\gamma_\ell = \begin{cases} \phi_1\gamma_1 + \sigma_a^2 & \text{if } \ell = 0 \\ \phi_1\gamma_{\ell-1} & \text{if } \ell > 0 \end{cases}$$

– ACF:

$$\rho_\ell = \phi_1\rho_{\ell-1} = \phi_1^\ell, \quad \text{for } \ell > 0$$

This result says that the ACF of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$. For a positive ϕ_1, the plot of ACF of an AR(1) model shows a nice exponential decay. For a negative ϕ_1, the plot consists of two alternating exponential decays with rate ϕ_1^2.

- **Properties of AR(2) Model**

– An AR(2) model assumes the form

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t \tag{11}$$

– Mean:

$$\mathrm{E}(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

provided that $\phi_1 + \phi_2 \neq 1$.

– Autocovariance: the moment equation of a stationary AR(2) model

$$\gamma_\ell = \phi_1\gamma_{\ell-1} + \phi_2\gamma_{\ell-2}, \quad \text{for } \ell > 0$$

Dividing the above equation by γ_0, we have the property

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2} \tag{12}$$

for the ACF of r_t.

– ACF:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_\ell = \phi_1\rho_{\ell-1} + \phi_2\rho_{\ell-2}, \quad \ell \geq 2$$

- The result of (12) says that the ACF of a stationary AR(2) series satisfies the second-order difference equation

$$(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0$$

where B is called the back-shift operator such that $B\rho_\ell = \rho_{\ell-1}$.

- Corresponding to the prior difference equation, there is a second-order polynomial equation:

$$1 - \phi_1 x - \phi_2 x^2 = 0 \tag{13}$$

Solutions of this equation are

$$x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

- In the time series literature, inverses of the two solutions are referred to as the characteristic roots of the AR(2) model.

- Denote the two characteristic roots by ω_1 and ω_2. If both ω_i are real valued, then the second-order difference equation of the model can be factored as $(1 - \omega_1 B)(1 - \omega_2 B)$ and the AR(2) model can be regarded as an AR(1) model operators on top of another AR(1) model. The ACF of r_t is then a mixture of two exponential decays.

- If $\phi_1^2 + 4\phi_2 < 0$, then ω_1 and ω_2 are complex numbers, and the plot of ACF of r_t would show a picture of damping sine and cosine waves.

- In business and economic applications, complex characteristic roots are important. They give rise to the behavior of business cycles.

- For an AR(2) model in (11) with a pair of complex characteristic roots, the average length of the stochastic cycles is

$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]}$$

where the cosine inverse is stated in radians. If one writes the complex solutions as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$, $\phi_2 = -(a^2+b^2)$, and

$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2+b^2})}$$

where $\sqrt{a^2+b^2}$ is the absolute value of $a \pm bi$.

– Stationarity: the stationarity condition of an AR(2) time series is that the absolute values of its two characteristic roots are less than 1. Equivalently, the two solutions of the characteristic equation are greater than 1 in modulus.

- **AR(p) Model**
  – The results of the AR(1) and AR(2) models can readily be generalized to the general AR(p) model in (9).

  – Mean:

  $$\mathrm{E}(r_t) = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

  provided that the denominator is not zero.

  – The associated characteristic equation of the model is

  $$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0$$

  If all the solutions of this equation are greater than 1 in modulus, then the series r_t is stationary. Again, inverses of the solutions are the characteristic roots of the model. Thus, stationarity requires that all characteristic roots are less than 1 in modulus.

  – For a stationary AR(p) series, the ACF satisfies the difference equation

  $$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\rho_\ell = 0, \quad \text{for } \ell > 0$$

  The plot of ACF of a stationary AR(p) model would then show a mixture of damping sine and cosine patterns and exponential decays depending on the nature of its characteristic roots.

- **Identifying AR Models in Practice**
  - In application, the order p of an AR time series is unknown.
  - Two general approaches are available for determining the value of p.
  - The first approach is to use the partial autocorrelation function (PACF), and the second approach uses some information criteria.
  - **PACF**
    * The PACF of a stationary time series is a function of its ACF and is a useful tool for determining the order p of an AR model.
    * A simple, yet effective way to introduce PACF is to consider the following AR models in consecutive orders:

$$\begin{aligned} r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t} \\ r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t} \\ r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t} \\ &\vdots \end{aligned}$$

      where $\phi_{0,j}$, $\phi_{i,j}$, and $\{e_{jt}\}$ are, respectively, the constant term, the coefficient of r_{t-i}, and the error term of an AR(j) model.
    * These models are in the form of a multiple linear regression and can be estimated by the least-squares method.
    * As a matter of fact, they are arranged in a sequential order that enables us to apply the idea of partial F test in multiple linear regression analysis.
    * The estimate $\hat{\phi}_{1,1}$ of the first equation is called the lag-1 sample PACF of r_t. The estimate $\hat{\phi}_{2,2}$ of the second equation is called the lag-2 sample PACF of r_t. The estimate $\hat{\phi}_{3,3}$ of the third equation is called the lag-3 sample PACF of r_t, and so on.

* From the definition, the lag-2 PACF $\hat{\phi}_{2,2}$ shows the added contribution of r_{t-2} to r_t over the AR(1) model $r_t = \phi_0 + \phi_1 r_{t-1} + e_{1t}$. The lag-3 PACF $\hat{\phi}_{3,3}$ shows the added contribution of r_{t-3} to r_t over the AR(2) model, and so on.

* Therefore, for an AR(p) model, the lag-p sample PACF should not be zero, but $\hat{\phi}_{j,j}$ should be close to zero for all $j > p$. We use this property to determine the order p.

* For a stationary Gaussian AR(p) model, it can be shown that the sample PACF has the following properties:

  · $\hat{\phi}_{p,p}$ converges to ϕ_p as the sample size T goes to infinity.

  · $\hat{\phi}_{\ell,\ell}$ converges to zero for all $\ell > p$.

  · The asymptotic variance of $\hat{\phi}_{\ell,\ell}$ is $1/T$ for $\ell > p$.

* These results say that, for an AR(p) series, the sample PACF cuts off at lag p.

– **Information Criteria**

* There are several information criteria available to determine the order p of an AR(p) process. All of them are likelihood based.

* The well-known Akaike information criterion (AIC) (Akaike, 1973) is defined as

$$\text{AIC} = \frac{-2}{T}\ln(\text{likelihood}) + \frac{2}{T}(\text{number of parameters}) \tag{14}$$

  where the likelihood function is evaluated at the maximum-likelihood estimates and T is the sample size.

* For a Gaussian AR(ℓ) model, AIC reduces to

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

  where $\tilde{\sigma}_\ell^2$ is the maximum-likelihood estimate of σ_a^2, which is the variance of a_t.

* The first term of the AIC in (14) measures the goodness of fit of the AR(ℓ) model to the data, whereas the second term is called the penalty function of the criterion because it penalizes a candidate model by the number of parameters used.

* Different penalty functions result in different information criteria.

* Another commonly used criterion function is the Schwarz-Bayesian information criterion (BIC).

* For a Gaussian AR(ℓ) model, the criterion is

$$\text{BIC} = \ln(\tilde{\sigma}_\ell^2) + \frac{\ell \ln T}{T}$$

* The penalty for each parameter used is 2 for AIC and $\ln T$ for BIC. Thus, compared with AIC, BIC tends to select a lower AR model when the sample size is moderate or large.

* To use AIC to select an AR model in practice, one computes AIC(ℓ) for $\ell = 0, 1, \ldots, P$, where P is a prespecified positive integer and selects the order k that has the minimum AIC value. The same rule applies to BIC.

- **Parameter Estimation**
  - For a specified AR(p) model in (9), the conditional least-squares method, which starts with the $(p+1)$th observation, is often used to estimate the parameters.
  - Specifically, conditioning on the first p observations, we have

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t, \; t = p+1, \ldots, T$$

which is in the form of multiple linear regression and can be estimated by the least-squares method.

- Denote the estimate of ϕ_i by $\hat{\phi}_i$. The fitted model is

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \cdots + \hat{\phi}_p r_{t-p}$$

  and hence the residual is

$$\hat{a}_t = r_t - \hat{r}_t$$

- The series $\{\hat{a}_t\}$ is called the residual series, from which we obtain

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T - 2p - 1}$$

- If the conditional likelihood method is used, the estimates of ϕ_i remain unchanged, but the estimate of σ_a^2 becomes $\tilde{\sigma}_a^2 = \hat{\sigma}_a^2 \times (T-2p-1)/(T-p)$

- **Model Checking**

  - If the model is adequate, then the residual series should behave as a white noise.

  - The ACF and the Ljung-Box statistics in (3) of the residuals can be used to check the closeness of $\hat{a}_t$ to a white noise.

  - For an AR(p) model, the Ljung-Box statistic $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR coefficients used in the model.

- **Goodness of Fit**

  - A commonly used statistic to measure goodness of fit of a stationary model is the R square (R^2) defined as

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}}$$

– For a stationary AR(p) time series model with T observations, the measure becomes

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (r_t - \bar{r})^2}$$

where $\bar{r} = \sum_{t=p+1}^{T} r_t/(T-p)$.

– Typically, a larger R^2 indicates that the model provides a closer fit to the data.

– For a given data set, it is well known that R^2 is a nondecreasing function of the number of parameters used. To overcome this weakness, an adjusted R^2 is proposed, which is defined as

$$\text{Adj } R^2 = 1 - \frac{\text{variance of residuals}}{\text{variance of } r_t} = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2}$$

where $\hat{\sigma}_r^2$ is the sample variance of r_t.

- **Forecasting**

  – Forecasting is an important application of time series analysis.

  – For the AR(p) model in (9), suppose that we are at the time index h and are interested in forecasting $r_{h+\ell}$, where $\ell \geq 1$. The time index h is called the forecast origin and the positive integer ℓ is the forecast horizon.

  – Let $\hat{r}_h(\ell)$ be the forecast of $r_{h+\ell}$ using the minimum squared error loss function. That is,

$$\mathrm{E}\{[r_{h+\ell} - \hat{r}_h(\ell)]^2 \mid F_h\} \leq \min_g \mathrm{E}[(r_{h+\ell} - g)^2 \mid F_h]$$

  where g is a function of the information available at time h.

  – We referred to $\hat{r}_h(\ell)$ as the ℓ-step ahead forecast of r_t at the forecast origin h.

- **1-Step-Ahead Forecast**
  * Under the minimum squared error loss function, the point forecast of r_{h+1} given F_h is the conditional expectation

    $$\hat{r}_h(1) = \mathrm{E}(r_{h+1} \mid F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i}$$

    and the associated forecast error is

    $$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$$

  * Consequently, the variance of the 1-step-ahead forecast error is

    $$\mathrm{Var}[e_h(1)] = \mathrm{Var}(a_{h+1}) = \sigma_a^2$$

  * In the econometric literature, a_{t+1} is referred to as the shock to the series at time $t+1$.

- **2-Step-Ahead Forecast**
  * Next consider the forecast of r_{h+2} at the forecast origin h. From the AR(p) model, we have

    $$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \phi_2 r_h + \cdots + \phi_p r_{h+2-p} + a_{h+2}$$

  * Taking conditional expectation, we have

    $$\hat{r}_h(2) = \mathrm{E}(r_{h+2} \mid F_h) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \cdots + \phi_p r_{h+2-p}$$

    and the associated error

    $$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} + \phi_1 a_{h+1}$$

  * The variance of the forecast error is

    $$\mathrm{Var}[e_h(2)] = (1 + \phi_1^2)\sigma_a^2$$

- It is interesting to see that $\text{Var}[e_h(2)] \geq \text{Var}[e_h(1)]$, meaning that as the forecast horizon increases the uncertainty in forecast also increases.

- This is in agreement with common sense that we are more uncertain about r_{h+2} than r_{h+1} at the time index h for a linear time series.

- **Multistep-Ahead Forecast**

  - The ℓ-step-ahead forecast based on the minimum squared error loss function is the conditional expectation of $r_{h+\ell}$ given F_h, which can be obtained as

$$\hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i)$$

    where it is understood that $\hat{r}_h(i) = r_{h+i}$ if $i \leq 0$.

  - The ℓ-step-ahead forecast error is $e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$.

  - It can be shown that for a stationary AR(p) model, $\hat{r}_h(\ell)$ converges to $\text{E}(r_t)$ as $\ell \to \infty$, meaning that for such a series long-term point forecast approaches its unconditional mean. This property is referred to as the mean reversion in the finance literature.

  - The variance of the forecast error then approaches the unconditional variance of r_t.

- **Simple MA Models**

  - Another class of simple models that are also useful in modeling return series in finance is the moving-average (MA) model.

  - In this book, we use the approach which treats the model as an infinite-order AR model with some parameter constraints to introduce MA models.

  - There is no particular reason, but simplicity, to assume a priori that the order of an AR model is finite.

- We may entertain, at least in theory, an AR model with infinite order as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + a_t$$

- However, such an AR model is not realistic because it has infinite many parameters.

- One way to make the model practical is to assume that the coefficients ϕ_i's satisfy some constraints so that they are determined by a finite number of parameters. A special case of this idea is

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t \tag{15}$$

  where the coefficients depend on a single parameter θ_1 via $\phi_i = \theta_1^i$ for $i \geq 1$.

- For the model in (15) to be stationary, θ_1 must be less than 1 in absolute value; otherwise, θ_1^i and the series will explode.

- Because $|\theta_1| < 1$, we have $\theta_1^i \to 0$ as $i \to \infty$. Thus, the contribution of r_{t-i} to r_t decays exponentially as i increases.

- This is reasonable as the dependence of a stationary series r_t on its lagged value r_{t-i}, if any, should decay over time.

- The model in (15) can be rewritten in a rather compact form:

$$r_t = c_0 + (1 - \theta_1 B) a_t \tag{16}$$

  where $c_0 = \phi_0(1 - \theta_1)$ is a constant.

- The model in (16) is called an MA model of order 1 or MA(1) model for short.

– Similarly, an MA(2) model is in the form

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \tag{17}$$

and an MA(q) model is

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{18}$$

where $q > 0$.

- **Properties of MA Models**
  - **Stationarity**
    - MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant.
    - Consider the MA(1) model in (16). Taking expectation of the model, we have

      $$\mathrm{E}(r_t) = c_0$$

      which is time invariant. Taking the variance of (16), we have

      $$\mathrm{Var}(r_t) = (1 + \theta_1^2)\sigma_a^2$$

      which is also time invariant.
    - The prior discussion applies to general MA(q) models, and we obtain two general properties. First, the constant term of an MA model is the mean of the series, i.e., $\mathrm{E}(r_t) = c_0$. Second, the variance of an MA(q) model is

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

  - **ACF**
    - Assume for simplicity that $c_0 = 0$ for an MA(1) model. We have

      $$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1}{1 + \theta_1^2}, \quad \rho_\ell = 0, \quad \text{for} \quad \ell > 1$$

      Thus, for an MA(1) model, the lag-1 ACF is not zero, but all higher order ACFs are zero. In other words, the ACF of an MA(1) model cuts off at lag 1.

* For the MA(2) model in (17), the ACFs are

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_\ell = 0, \quad \text{for} \quad \ell > 2$$

  Hence, the ACF cuts off at lag 2.

* This property generalizes to other MA models. For an MA(q) model, the lag-q ACF is not zero, but $\rho_\ell = 0$ for $\ell > q$. Consequently, an MA(q) series is only linearly related to its first q-lagged values and hence is a "finite-memory" model.

– **Invertibility**

  * Rewriting a zero-mean MA(1) model as $a_t = r_t + \theta_1 a_{t-1}$, one can use repeated substitutions to obtain

$$a_t = r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \theta_1^3 r_{t-3} + \cdots$$

  * This equation expresses the current shock a_t as a linear combination of the present and past returns.

  * Intuitively, θ_1^j should go to zero as j increases because the remote return r_{t-j} should have very little impact on the current shock, if any.

  * Consequently, for an MA(1) model to be plausible we require $\theta_1 < 1$. Such an MA(1) model is said to be invertible. If $|\theta_1| = 1$, then the MA(1) model is noninvertible.

– **Identifying MA Order**

  * The ACF is useful in identifying the order of an MA model.

  * For a time series r_t with ACF ρ_ℓ, if $\rho_\ell = 0$ for $\ell > q$, then r_t follows an MA(q) model.

– **Estimation**

* Maximum-likelihood estimation is commonly used to estimate MA models.

* There are two approaches for evaluating the likelihood function of an MA model. The first approach assumes that the initial shocks (i.e., a_t for $t \leq 0$) are zero.

* As such, the shocks needed in likelihood function calculation are obtained recursively form the model, starting with $a_1 = r_1 - c_0$ and $a_2 = r_2 - c_0 + \theta_1 a_1$. This approach is referred to as the conditional-likelihood method and the resulting estimates the conditional MLEs.

* The second approach treats the initial shocks as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the exact-likelihood method.

* The exact-likelihood estimates are preferred over the conditional ones, especially when the MA model is close to being noninvertible.

* The exact method, however, requires more intensive computation. If the sample size is large, then the two types of MLEs are close to each other.

- **Forecasting Using MA Models**

  * Forecasts of an MA model can easily be obtained. Because the model has finite memory, its point forecasts go to the mean of the series quickly.

  * For the 1-step-ahead forecast of an MA(1) process, the model says

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h$$

Taking the conditional expectation, we have

$$\begin{aligned} \hat{r}_h(1) &= \mathrm{E}(r_{h+1} \mid F_h) &= c_0 - \theta_1 a_h \\ e_h(1) &= r_{h+1} - \hat{r}_h(1) &= a_{h+1} \end{aligned}$$

The variance of the 1-step-ahead forecast error is $\mathrm{Var}[e_h(1)] = \sigma_a^2$.

* For the 2-step-ahead forecast, we have

$$\begin{aligned}\hat{r}_h(2) &= \mathrm{E}(r_{h+2} \mid F_h) = c_0 \\ e_h(2) &= r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}\end{aligned}$$

  The variance of the 2-step-ahead forecast error is $\mathrm{Var}[e_h(2)] = (1 + \theta_1^2)\sigma_a^2$, which is the variance of the model and is greater than or equal to that of the 1-step-ahead forecast error.

* More generally, $\hat{r}_h(\ell) = c_0$ for $\ell \geq 2$.

* Similarly, for an MA(2) model, we have

$$\begin{aligned}\hat{r}_h(1) &= c_0 - \theta_1 a_h - \theta_2 a_{h-1} \\ \hat{r}_h(2) &= c_0 - \theta_2 a_h \\ \hat{r}_h(\ell) &= c_0, \quad \text{for} \quad \ell > 2\end{aligned}$$

  Thus, the multistep-ahead forecasts of an MA(2) model go to the mean of the series after two steps. The variance of forecast errors go to the variance of the series after two steps.

* In general, for an MA(q) model, multistep-ahead forecasts go to the mean after the first q steps.

– **Summary**

  * For MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series.

  * For AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process.

  * An MA series is always stationary, but for an AR series to be stationary, all of its characteristic roots must be less than 1 in modulus.

  * For a stationary series, the multistep-ahead forecasts converge to the mean of the series, and the variances of forecast errors converge to the variance of the series as the forecast horizon increases.

- **Simple ARMA Models**
  - In some applications, the AR or MA models discussed in the previous sections become cumbersome because one may need a high-order model with many parameters to adequately describe the dynamic structure of the data.
  - To overcome this difficulty, the autoregressive moving-average (ARMA) models are introduced.
  - Basically, an ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization.
  - In this section, we study the simplest ARMA(1,1) model. A time series r_t follows an ARMA(1,1) model if it satisfies

    $$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1} \tag{19}$$

    where a_t is a white noise series. For this model to be meaningful, we need $\phi_1 \neq \theta_1$; otherwise, there is a cancelation in the equation and the process reduces to a white noise series.
  - **Properties of ARMA(1,1) Models**
    * We start with the stationary condition. The mean of r_t is

      $$\mathrm{E}(r_t) = \frac{\phi_0}{1 - \phi_1}$$

      provided that the series is weakly stationary.
    * Next, assuming for simplicity that $\phi_0 = 0$, we consider the autocovariance function of r_t. If the series r_t is weakly stationary, then we have

      $$\mathrm{Var}(r_t) = \frac{(1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2}{1 - \phi_1^2}$$

      Because the variance is positive, we need $\phi_1^2 < 1$ (i.e., $|\phi_1| < 1$).

* In addition, we have

$$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma_a^2$$

and

$$\gamma_\ell - \phi_1 \gamma_{\ell-1} = 0, \quad \text{for} \quad \ell > 1 \tag{20}$$

In terms of ACF, the previous results show that for a stationary ARMA(1,1) model

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma_a^2}{\gamma_0}, \quad \rho_\ell = \phi_1 \rho_{\ell-1}, \quad \text{for} \quad \ell > 1$$

* Thus, the ACF of an ARMA(1,1) model behaves very much like that of an AR(1) model except that the exponential decay starts with lag 2. Consequently, the ACF of an ARMA(1,1) model does not cut off at any finite lag.

* Turning to PACF, one can show that the PACF of an ARMA(1,1) model does not cut off at any finite lag either. It behaves very much like that of an MA(1) model except that the exponential decay starts with lag 2 instead of lag 1.

– **General ARMA Models**

* A general ARMA(p,q) model is in the form

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i}$$

where $\{a_t\}$ is a white noise series and p and q are nonnegative integers. Using the back-shift operator, the model can be written as

$$(1 - \phi_1 B - \cdots - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t \tag{21}$$

The polynomial $1 - \phi_1 B - \cdots - \phi_p B^p$ is the AR polynomial of the model. Similarly, $1 - \theta_1 B - \cdots - \theta_q B^q$ is the MA polynomial.

* We require that there are no common factors between these two polynomials; otherwise the order (p, q) of the model can be reduced.

* Like a pure AR model, the AR polynomial introduces the characteristic equation of an ARMA model. If all the solutions of the characteristic equation are greater than 1 in absolute value, then the ARMA model is weakly stationary. In this case, the unconditional mean of the model is $\mathrm{E}(r_t) = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$.

– **Identifying ARMA Models**

  * The ACF and PACF are not informative in determining the order of an ARMA model.

  * Tsay and Tiao (1984) propose a new approach that uses the extended autocorrelation function (EACF) to specify the order of an ARMA process.

  * The information criteria discussed earlier can also be used to select ARMA(p, q) models.

  * Typically, for some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p, q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC).

  * In addition, the Ljung-Box statistics of the residuals can be used to check the adequacy of a fitted model. If the model is correctly specified, then $Q(m)$ follows asymptotically a chi-squared distribution with $m - g$ degrees of freedom, where g denotes the number of AR or MA coefficients fitted in the model.

– **Forecasting Using an ARMA Model**

  * Denote the forecast origin by h and the available information by F_h. The 1-step-ahead forecast of r_{h+1} can be easily obtained from the model as

$$\hat{r}_h(1) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i a_{h+1-i}$$

    and the associated forecast error is $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$. The variance of 1-step-ahead forecast error is $\mathrm{Var}[e_h(1)] = \sigma_a^2$.

* For the ℓ-step-ahead forecast, we have

$$\hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i a_h(\ell - i)$$

  where it is understood that $\hat{r}_h(\ell-i) = r_{h+\ell-i}$ if $\ell-i \leq 0$ and $a_h(\ell-i) = 0$ if $\ell - i > 0$ and $a_h(\ell - i) = a_{h+\ell-i}$ if $\ell - i \leq 0$.

* Thus, the multistep-ahead forecasts of an ARMA model can be computed recursively. The associated forecast error is

$$e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$$

– **Three Model Representations for an ARMA Model**

* We briefly discuss three model representations for a stationary ARMA(p, q) model. The three representations serve three different purposes.

* The first representation is the ARMA(p, q) model in (21). This representation is compact and useful in parameter estimation. It is also useful in computing recursively multistep-ahead forecasts of r_t.

* For the other two representations, we use long division of two polynomials. Given two polynomials $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{i=1}^{1} \theta_i B^i$, we can obtain that

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + \cdots \equiv \psi(B) \qquad (22)$$

  and

$$\frac{\phi(B)}{\theta(B)} = 1 - \pi_1 B - \pi_2 B^2 + \cdots \equiv \pi(B) \qquad (23)$$

* **AR Representation**
  · Using the result of long division of (23), the ARMA(p, q) model can be written as

$$r_t = \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \cdots + a_t \qquad (24)$$

- This representation shows the dependence of the current return r_t on the past returns r_{t-i}, where $i > 0$. The coefficients $\{\pi_i\}$ are referred to as the π weights of an ARMA model.

- To show that the contribution of the lagged value r_{t-i} to r_t is diminishing as i increases, the π_i coefficient should decay to zero as i increases. An ARMA(p, q) model that has this property is said to be invertible.

- A sufficient condition for invertibility is that all the zeros of the polynomial $\theta(B)$ are greater than unity in modulus.

- From the AR representation in (24), an invertible ARMA(p, q) series r_t is a linear combination of the current shock a_t and a weighted average of the past values. The weights decay exponentially for more remote past values.

∗ **MA Representation**

- Using the result of long division in (22), an ARMA(p, q) model can also be written as

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \psi(B) a_t \tag{25}$$

where $\mu = \mathrm{E}(r_t) = \phi_0/(1 - \phi_1 - \cdots - \phi_p)$.

- This representation shows explicitly the impact of the past shock a_{t-i} $(i > 0)$ on the current return r_t. The coefficients $\{\psi_i\}$ are referred to as the impulse response function of the ARMA model.

- For a weakly stationary series, the ψ_i coefficients decay exponentially as i increases. This is understandable as the effect of shock a_{t-i} on the return r_t should diminish over time.

- Thus, for a stationary ARMA model, the shock a_{t-i} does not have a permanent impact on the series.

- The MA representation in (25) is also useful in computing the variance of a forecast error. At the forecast origin h, we have the shocks $a_h, a_{h-1}, \ldots$ Therefore, the ℓ-step-ahead point forecast is

$$\hat{r}_h(\ell) = \mu + \psi_\ell a_h + \psi_{\ell+1} a_{h-1} + \cdots \tag{26}$$

and the associated forecast error is

$$e_h(\ell) = a_{h+\ell} + \psi_1 a_{h+\ell-1} + \cdots + \psi_{\ell-1} a_{h+1}$$

Consequently, the variance of ℓ-step-ahead forecast error is

$$\text{Var}[e_h(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2 \tag{27}$$

which, as expected, is a nondecreasing function of the forecast horizon ℓ.

· The MA representation provides a simple proof of mean reversion of a stationary time series. The stationarity implies that ψ_i approaches zero as $i \to \infty$. Therefore, by (26), we have $\hat{r}_h(\ell) \to \mu$ as $\ell \to \infty$.

· Furthermore, we have $\text{Var}(r_t) = (1 + \sum_{i=1}^{\infty} \psi_i^2)\sigma_a^2$. Consequently, by (27), we have $\text{Var}[e_h(\ell)] \to \text{Var}(r_t)$ as $\ell \to \infty$.

- **Unit-Root Nonstationarity**
  - So far we have focused on return series that are stationary. In some studies, interest rates, foreign exchange rates, or the prices series of an asset are of interest.
  - These series tend to be nonstationary. For a price series, the nonstationarity is mainly due to the fact that there is no fixed level for the price.
  - In time series literature, such a nonstationary series is called unit-root nonstationary time series.
  - The best known example of unit-root nonstationary time series is the random-walk model.
  - **Random Walk**
    - A time series $\{p_t\}$ is a random walk if it satisfies

$$p_t = p_{t-1} + a_t \tag{28}$$

where p_0 is a real number denoting the starting value of the process and $\{a_t\}$ is a white noise series.

* If we treat the random-walk model as a special AR(1) model, then the coefficient of p_{t-1} is unity, which does not satisfy the weak stationarity condition of an AR(1) model.

* A random-walk series is, therefore, not weakly stationary, and we call it a unit-root nonstationary time series.

* The 1-step-ahead forecast of model (28) at the forecast origin h is

$$\hat{p}_h(1) = \mathrm{E}(p_{h+1} \mid p_h, p_{h-1}, \ldots) = p_h$$

Such a forecast has no practical value. The 2-step-ahead forecast is

$$\hat{p}_h(2) = \mathrm{E}(p_{h+2} \mid p_h, p_{h-1}, \ldots) = p_h$$

In fact, for any forecast horizon $\ell > 0$, we have

$$\hat{p}_h(\ell) = p_h$$

Thus, for all forecast horizons, point forecasts of a random-walk model are simply the value of the series at the forecast origin. Therefore, the process is not mean reverting.

* The MA representation of the random-walk model in (28) is

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

This representation has several important practical implications.

* First, the ℓ-step-ahead forecast error is

$$e_h(\ell) = a_{h+\ell} + \cdots + a_{h+1}$$

so that $\mathrm{Var}[e_h(\ell)] = \ell\sigma_a^2$, which diverges to infinity as $\ell \to \infty$. The length of an interval forecast of $p_{t+\ell}$ will approach infinity as the forecast horizon increases. This result says that the usefulness of point forecast $\hat{p}_h(\ell)$ diminishes as ℓ increases, which implies that the model is not predictable.

* Second, the unconditional variance of p_t is unbounded because $\mathrm{Var}[e_h(\ell)]$ approaches infinity as ℓ increases.

* Third, for the representation, $\psi_i = 1$ for all i. Thus, the impact of any past shock a_{t-i} on p_t does not decay over time. Consequently, the series has a strong memory as it remembers all of the past shocks. In economics, the shocks are said to have a permanent effect on the series. The strong memory of a unit-root time series can be seen from the sample ACF of the observed series. The sample ACFs are all approaching 1 as the sample size increases.

- **Random Walk with Drift**
  * As shown by empirical examples considered so far, the log return series of a market index tends to have a small and positive mean. This implies that the model for the log price is

$$p_t = \mu + p_{t-1} + a_t \tag{29}$$

    where $\mu = \mathrm{E}(p_t - p_{t-1})$ and $\{a_t\}$ is a zero-mean white noise series.

  * The constant term μ of model (29) is very important in financial study. It represents the time trend of the log price p_t and is often referred to as the drift of the model. To see this, assume that the initial log price is p_0. Then we have

$$p_t = t\mu + p_0 + a_t + a_{t-1} + \cdots + a_1$$

    which shows that the log price consists of a time trend $t\mu$ and a pure random-walk process $\sum_{i=1}^{t} a_i$.

  * Because $\mathrm{Var}(\sum_{i=1}^{t} a_i) = t\sigma_a^2$, the conditional standard deviation of p_t is $\sqrt{t}\sigma_a$, which grows at a slower rate than the conditional expectation of p_t.

  * Therefore, if we graph p_t against the time index t, we have a time trend with slope μ.

- **Interpretation of the Constant Term**
  * From the previous discussions, it is important to understand the meaning of a constant term in a time series model.

* First, for an MA(q) model in (18), the constant term is simply the mean of the series.

* Second, for a stationary AR(p) model in (9) or ARMA(p, q) in (21), the constant term is related to the mean via $\mu = \phi_0/(1-\phi_1-\cdots-\phi_p)$.

* Third, for a random walk with drift, the constant term becomes the time slope of the series.

* These different interpretations for the constant term in a time series model clearly highlight the difference between dynamic and usual linear regression models.

* Another important difference between dynamic and regression models is shown by an AR(1) model and a simple linear regression model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \quad \text{and} \quad y_t = \beta_0 + \beta_1 x_t + a_t$$

  For the AR(1) model to be meaningful, the coefficient ϕ_1 must satisfy $|\phi_1| < 1$. However, the coefficient β_1 can assume any fixed real number.

– **Trend Stationary Time Series**

  * A closely related model that exhibits linear trend is the trend-stationary time series model
$$p_t = \beta_0 + \beta_1 t + r_t$$
    where r_t is a stationary time series.

  * Here p_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to that of a random-walk model with drift.

  * However, there is a major difference between the two models.

  * To see this, suppose that p_0 is fixed. The random-walk model with drift assumes the mean $\mathrm{E}(p_t) = p_0 + \mu t$ and variance $\mathrm{Var}(p_t) = t\sigma_a^2$, both of them are time dependent.

* On the other hand, the trend-stationary model assumes the mean $\mathrm{E}(p_t) = \beta_0 + \beta_1 t$, which depends on time, and variance $\mathrm{Var}(p_t) = \mathrm{Var}(r_t)$, which is finite and time invariant.

* The trend-stationary series can be transformed into a stationary one by removing the time trend via a simple linear regression analysis.

- **General Unit-Root Nonstationary Models**
  * Consider an ARMA model. If one extends the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the well-known autoregressive integrated moving-average (ARIMA) model.

  * An ARIMA model is said to be unit-root nonstationary because its AR polynomial has a unit root.

  * Like a random walk model, an ARIMA model has strong memory because the ψ_i coefficients in its MA representation do not decay over time to zero, implying that the past shock a_{t-i} of the model has a permanent effect on the series.

  * A conventional approach for handling unit-root nonstationary is to use differencing.

- **Differencing**
  * A time series is said to be an ARIMA$(p, 1, q)$ process if the change series $c_t = y_t - y_{t-1} = (1 - B)y_t$ follows a stationary and invertible ARMA(p, q) model.

  * In finance, price series are commonly believed to be nonstationary, but the log return series, $r_t = \ln(P_t) - \ln(P_{t-1})$, is stationary. In this case, the log price series is unit-root nonstationary and hence can be treated as an ARIMA process.

  * The idea of transforming a nonstationary series into a stationary one by considering its change series is called differencing in the time series literature.

– **Unit-Root Test**

* To test whether the log price p_t of an asset follows a random walk or a random walk with drift, we employ the models

$$p_t = \phi_1 p_{t-1} + e_t \quad (30)$$

$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t \quad (31)$$

where e_t denotes the error term, and consider the null hypothesis $H_0 : \phi_1 = 1$ versus the alternative hypothesis $H_1 : |\phi_1| < 1$.

* This is the well-known unit-root testing problem; see Dickey and Fuller (1979).

* A convenient test statistic is the t ratio of the least-squares (LS) estimate of ϕ_1 under the null hypothesis.

* Case 1: For (30), the LS method gives

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} p_{t-1} p_t}{\sum_{t=1}^{T} p_{t-1}^2}, \quad \sigma_e^2 = \frac{\sum_{t=1}^{T} (p_t - \hat{\phi}_1 p_{t-1})^2}{T-1}$$

where $p_0 = 0$ and T is the sample size. The t ratio is

$$\mathrm{DF} \equiv t \text{ ratio} = \frac{\hat{\phi}_1 - 1}{\mathrm{std}(\hat{\phi}_1)} = \frac{\sum_{t=1}^{T} p_{t-1} e_t}{\hat{\sigma}_e \sqrt{\sum_{t=1}^{T} p_{t-1}^2}}$$

which is commonly referred to as the Dickey-Fuller (DF) test. If $\{e_t\}$ is a white noise series with finite moments of order slightly greater than 2, then the DF statistic converges to a function of the standard Brownian motion as $T \to \infty$; see Chan and Wei (1988) and Phillips (1987) for more information.

* Case 2: If ϕ_0 is zero but (31) is employed anyway, then the resulting t ratio for testing $\phi_1 = 1$ will converge to another nonstandard asymptotic distribution.

* In either case, simulation is used to obtain critical values of the test statistics.

* Case 3: Yet if $\phi_0 \neq 0$ and (31) is used, then the t ratio for testing $\phi_1 = 1$ is asymptotically normal. However, large sample sizes are needed for the asymptotic normal distribution to hold.

* For many economic time series, ARIMA(p, d, q) models might be more appropriate than the simple model in (31). In the econometric literature, AR(p) models are often used. Denote the series by x_t. To verify the existence of a unit root in an AR(p) process, one may perform the test $H_0 : \beta = 1$ vs. $H_1 : \beta < 1$ using the regression

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \triangle x_{t-i} + e_t \tag{32}$$

  where c_t is a deterministic function of the time index t and $\triangle x_j = x_j - x_{j-1}$ is the differenced series of x_t.

* In practice, c_t can be zero or a constant or $c_t = \omega_0 + \omega_1 t$. The t ratio of $\hat{\beta} - 1$,

$$\mathrm{ADF} - \mathrm{test} = \frac{\hat{\beta} - 1}{\mathrm{std}(\hat{\beta})}$$

  where $\hat{\beta}$ denotes the LS estimate of β, is the well-known augmented Dickey-Fuller (ADF) unit-root test.

* Note that because of the first differencing, (32) can also be rewritten as

$$\triangle x_t = c_t + \beta_c x_{t-1} + \sum_{i=1}^{p-1} \phi_i \triangle x_{t-i} + e_t$$

  where $\beta_c = \beta - 1$. One can then test the equivalent hypothesis $H_0 : \beta_c = 0$ vs. $H_1 : \beta_c < 0$.

- **Seasonal Models**
  - Some financial time series such as quarterly earnings per share of a company exhibits certain cyclical or periodic behavior. Such a time series is called a seasonal time series.

– Analysis of seasonal time series has a long history.

1. In some applications, seasonality is of secondary importance and is removed from the data, resulting in a seasonally adjusted time series that is then used to make inference. The procedure to remove seasonality from a time series is referred to as seasonal adjustment.

2. In other applications such as forecasting, seasonality is as important as other characteristics of the data and must be handled accordingly.

– Because forecasting is a major objective of financial time series analysis, we focus on the latter approach and discuss some econometric models that are useful in modeling seasonal time series.

– **Seasonal Differencing**

  * In general, for a seasonal time series y_t with periodicity s, seasonal differencing means

$$\triangle_s y_t = y_t - y_{t-s} = (1 - B^s) y_t$$

– **Multiplicative Seasonal Models**

  * A special seasonal time series model:

$$(1 - B^s)(1 - B)x_t = (1 - \theta B)(1 - \Theta B^s)a_t \tag{33}$$

where s is the periodicity of the series, a_t is a white noise, $|\theta| < 1$, and $|\Theta| < 1$.

  * This is referred to as the airline model in the literature. It has been found to be widely applicable in modeling seasonal time series.

  * The AR part of the model simply consists of the regular, $(1 - B)$, and seasonal, $(1 - B^s)$, differences, whereas the MA part involves two parameters.

  * ACF of $w_t = (1 - B^s)(1 - B)x_t$ has a nice symmetric structure (see the text on pages 84-85).

* In particular,

$$\rho_1 = \frac{-\theta}{1+\theta^2}, \quad \rho_s = \frac{-\Theta}{1+\Theta^2}, \quad \text{and} \quad \rho_{s-1} = \rho_{s+1} = \rho_1 \rho_s,$$

  which can be regarded as the interaction between lag-1 and lag-s serial dependence, and the model of w_t is called a multiplicative seasonal MA model.

* In practice, a multiplicative seasonal model says that the dynamics of the regular and seasonal components of the series are approximately orthogonal.

* Alternatively, the model

$$w_t = (1 - \theta B - \Theta B^s) a_t \tag{34}$$

  where $|\theta| < 1$, and $|\Theta| < 1$, is a nonmultiplicative seasonal MA model.

* A multiplicative model is more parsimonious than the corresponding nonmultiplicative model because both models use the same number of parameters, but the multiplicative model has more nonzero ACFs.

- **Regression Models with Time Series Errors**
  - In many applications, the relationship between two time series is of major interest.
  - Consider a linear regression in the form

$$y_t = \alpha + \beta x_t + e_t \tag{35}$$

    where y_t and x_t are two time series and e_t denotes the error term.
  - The LS method is often used to estimate model (35). If $\{e_t\}$ is a white noise series, then the LS method produces consistent estimates.
  - In practice, however, it is commonly to see that the error term e_t is serially correlated. In this case, we have a regression model with time series errors, and the LS estimates of α and β may not be consistent.

- A regression model with time series errors is widely applicable in economics and finance, but it is one of the most commonly misused econometric models because the serial dependence in e_t is often overlooked. It pays to study the model carefully.

- We outline a general procedure for analyzing linear regression models with time series errors:
  1. Fit the linear regression model and check serial correlations of the residuals.
  2. If the residual series is unit-root nonstationary, take the first difference of both the dependent and explanatory variables. Go to step 1. If the residual series appears to be stationary, identify an ARMA model for the residuals and modify the linear regression model accordingly.
  3. Perform a joint estimation via the maximum-likelihood method and check the fitted model for further improvement.

- **Long-Memory Models**
  - We have discussed that for a stationary time series with ACF decays exponentially to zero as lag increases.
  - Yet for a unit-root nonstationary time series, it can be shown that the sample ACF converges to 1 for all fixed lags as the sample size increases.
  - There exist some time series whose ACF decays slowly to zero at a polynomial rate as the lag increases. These processes are referred to as long-memory time series.
  - One such example is the fractionally differenced process defined by

$$(1-B)^d x_t = a_t, \quad -0.5 < d < 0.5 \tag{36}$$

    where $\{a_t\}$ is a white noise series.

– We summarize some of the properties of model (36):

1. If $d < 0.5$, then x_t is a weakly stationary process and has the infinite MA representation

$$x_t = a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i}$$

where

$$\psi_k = \frac{d(d+1)\cdots(d+k-1)}{k!} = \frac{(k+d-1)!}{k!(d-1)!} = O(k^{d-1})$$

2. If $d > -0.5$, then x_t is invertible and has the infinite AR representation

$$x_t = \sum_{i=1}^{\infty} \pi_i x_{t-i} + a_t$$

where

$$\pi_k = \frac{d(1-d)\cdots(k-1-d)}{k!} = -\frac{(k-d-1)!}{k!(-d-1)!} = O(k^{-d-1})$$

3. For $-0.5 < d < 0.5$, the ACF of x_t is

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)}, \; k = 1, 2, \ldots$$

In particular, $\rho_1 = d/(1-d)$ and

$$\rho_k \approx \frac{(-d)!}{(d-1)!} k^{2d-1} \quad \text{as } k \to \infty$$

4. For $-0.5 < d < 0.5$, the PACF of x_t is

$$\phi_{k,k} = d/(k-d) \quad \text{for } k = 1, 2, \ldots$$

5. For $-0.5 < d < 0.5$, the spectral density function $f(\omega)$ of x_t, which is the Fourier transform of the ACF of x_t, satisfies

$$f(\omega) \sim \omega^{-2d}, \quad \text{as } \omega \to 0 \tag{37}$$

where $\omega \in [0, 2\pi]$ denotes the frequency.

– Of particular interest here is the behavior of ACF of x_t when $d < 0.5$. The property says that $\rho_k \sim k^{2d-1}$, which decays at a polynomial, instead of exponential, rate.

– A special characteristic of the spectral density function in (37) is that the spectrum diverges to infinity as $\omega \to 0$. However, the spectral density function of a stationary ARMA process is bounded for all $\omega \in [0, 2\pi]$.

– If the fractionally differenced series $(1 - B)^d x_t$ follows an ARMA(p, q) model, then x_t is called an ARFIMA(p, d, q) process, which is a generalized ARIMA model by allowing for noninteger d.

– In practice, if the sample ACF of a time series is not large in magnitude, but decays slowly, then the series may have long memory.

– For the pure fractionally differenced model in (36), one can estimate d using either a ML method or a regression method with logged periodogram at the lower frequencies.