Times Series and Forecasting (V)

Chapter 5. Models for Nonstationary Time Series

Jianhua Hu

hu.jianhua@mail.shufe.edu.cn

Spring 2013, SHUFE

5.1 Introduction

- In chapter 4, considered the large class of ARMA models

$$\phi(B)Y_t = \theta(B)e_t.$$

- A process $\{Y_t\}$ in this class is stationary if and only if the roots of the AR characteristic polynomial $\phi(x)$ all exceed 1 in absolute value (or modulus).
- In chapter 5, we extend this class of models to handle nonstationary processes. We broaden the class of ARMA models by differencing.
- This gives rise to the autoregressive integrated moving average, or ARIMA, class of models. This class incorporates a wide range of nonstationary time series processes.

The dth difference

- Suppose that $\{Y_t\}$ is a non-stationary stochastic process.
- The first difference process $\{\nabla Y_t\}$ consists of values of

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - B)Y_t.$$

- The second difference process $\{\nabla^2 Y_t\}$ consists of values of

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = \nabla Y_t - \nabla Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}.$$

- Higher order differences are formed similarly. In general,

$$\nabla^d Y_t = \nabla(\nabla^{d-1} Y_t) = \nabla^{d-1} Y_t - \nabla^{d-1} Y_{t-1},$$

for $d = 1, 2, \cdots ,$.

Example 5.1. A simulated series from the random walk

Example

- Suppose that $\{Y_t\}$ is a random walk process $Y_t = Y_{t-1} + e_t$.
- We know that $\{Y_t\}$ is not stationary because its autocovariance depends on t (see Chapter 2).
- However, the first difference process

$$\nabla Y_t = Y_t - Y_{t-1} = e_t$$

is stationary (because white noise processes are stationary).

Performance of ACF of sample from a random walk

- In Figure 5.1, we display (top) a sample realization of a random walk process with $n = 200$ and $e_t \sim N(0, 1)$.
- The sample ACF of the series decays very, very slowly over time. This is a characteristic of a nonstationary series.
- The first difference (white noise) process also appears in Figure 5.1 (bottom), along with its sample ACF. As we would expect from a white noise process, the sample autocorrelations r_k are within the $\pm 2/\sqrt{n}$ bounds.

Figure 5.1 Simulated series from a random walk

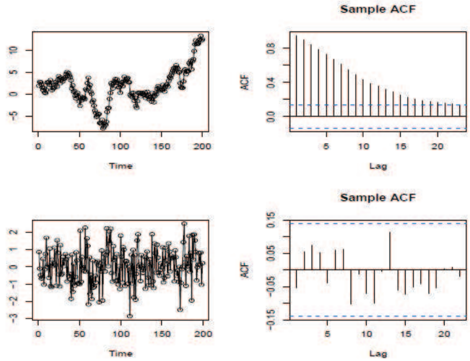

Figure 5.1: *Top: A simulated random walk process, with* $n = 200$ *and* $\sigma_e^2 = 1$, *and its sample ACF. Bottom: The first differences series with sample ACF.*

Deterministic trend models

- In Chapter 3, we talked about how to use regression methods to fit models of the form

$$Y_t = \mu_t + X_t,$$

  where μ_t is a deterministic function and where X_t is a stochastic process with $\mathrm{E}(X_t) = 0$.

- We expect that X_t is stationary. If so, we will use ARMA model to investigate $\{X_t\}$.

- For nonstationary Y_t, there are two ways to remove nonstationarity. Fitting the model or differencing data.

- Here, we discuss differencing data.

Taking first difference removes linear trend

- If $Y_t = \mu_t + X_t = \beta_0 + \beta_1 t + X_t$, where X_t is stationary and μ_t is a linear function of time, $\{Y_t\}$ is not a stationary process because $\mathsf{E}(Y_t)$ depends on t.
- The first differences are given by

$$\nabla Y_t = Y_t - Y_{t-1} = \beta_1 + X_t - X_{t-1}.$$

- Therefore, $\{\nabla Y_t\}$ is a stationary process because $\mathsf{E}(\nabla Y_t) = \beta_1$ and $\mathsf{Cov}(\nabla Y_t, \nabla Y_{t-k})$ is free of t (verify!).
- Taking first differences removes the linear trend.

Taking second difference removes quadratic trend

- If $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + X_t$, where X_t is stationary and μ_t is a quadratic function, $\{Y_t\}$ is not stationary and

$$\begin{aligned}\nabla Y_t =& Y_t - Y_{t+1} = (\mu_t + X_t) - (\mu_{t-1} + X_{t-1}) \\ =& (\beta_1 - \beta_2) + 2\beta_2 t + X_t - X_{t-1}\end{aligned}$$

  is not stationary either since $\mathsf{E}(\nabla Y_t)$ depends on t.

- However,

$$\begin{aligned}\nabla^2 Y_t =& (2\beta_2 t + X_t - X_{t-1}) - [2\beta_2(t-1) + X_{t-1} - X_{t-2}] \\ =& 2\beta_2 + X_t - 2X_{t-1} + X_{t-2}\end{aligned}$$

  is stationary (verify!).

- Taking second difference removes the quadratic trend.

Taking dth difference removes polynomial trend of degree d

- $Y_t = \mu_t + X_t$, where μ_t is a deterministic function and where X_t is a stationary stochastic process with $\mathrm{E}(X_t) = 0$.
- In general, if

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_d t^d$$

  is a polynomial in t of degree d, then the dth difference process $\{\nabla^d Y_t\}$ is stationary.
- The result is stated in Exercise 2.9.

Example 5.2. Gold price data

Example

Figure 5.2 contains a time series of $n = 254$ *daily observations of the price of gold (per troy ounce) in US dollars during the year 2005. Also depicted is the sample autocorrelation function* r_k*, for* $k = 1, 2, \cdots, 25$ *(i.e., out to 25 lags).*

Figure 5.2. Gold price data and sample ACF

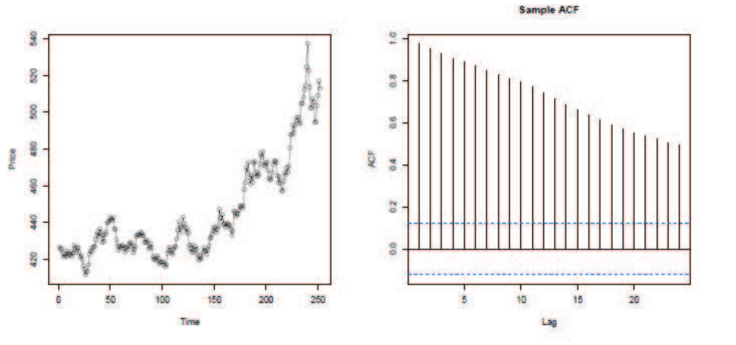

Figure 5.2: *Gold price data. Left: Daily price in US dollars per troy ounce: 1/4/05 through 12/30/05. Right: Sample autocorrelation function for the gold series.*

Observation of gold price data

- The gold time series does not resemble a realization of a stationary process. There is a pronounced change in mean over time.
- This observation is also seen in the sample ACF for the series, in particular, the sample ACF decays very, very slowly (a sign of nonstationarity).

- In Chapter 3, we used regression methods to detrend this series by fitting a quadratic regression model.
- We also examine first and second differences. Figure 5.3 depicts the first and second difference data differences, along with their sample autocorrelation functions.

Figure 5.3 First and second differences for Gold price data

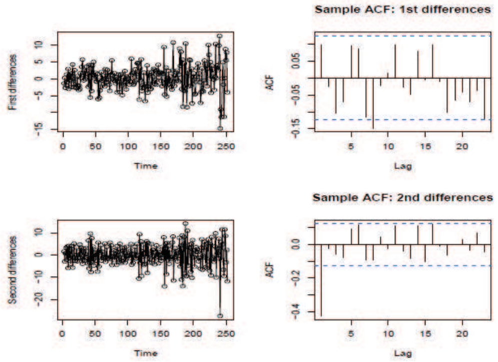

Figure 5.3: *Gold price data. Top: First difference series with sample ACF. Bottom: Second difference series with sample ACF.*

Analysis for gold price data

- Note that in the first difference series, the mean looks to be somewhat constant, but there looks to be a slight amount of increasing variation over time (particularly towards the end of the series). This violates the stationarity requirements.
- The second difference series looks to be fairly stationary (at least in the mean, which remains constant, but there is still a slight hint of increasing variance over time). Note that the sample ACF of the second difference series looks like what we would expect of an MA(1) process.

Analysis for gold price data (continued)

- There is somewhat of an outlier late in the series (December 10, 2005). This spike, along with the pronounced upward trend, prompted Jack Speer (at NPR) to write

  "Gold prices are up sharply this year, closing Friday at $537 an ounce. That's the highest value for gold in nearly 25 years and market observers expect the price may rise even further in the months ahead. But just why gold prices are up so much is something of a mystery."

- The prediction that prices would rise turned out to be correct (at the time of this writing, the price of gold is over $1,000 per ounce).

5.2. ARIMA models

- A time series $\{Y_t\}$ is said to follow an autoregressive integrated moving average (ARIMA(p, q)) model if the dth difference $\nabla^d Y_t$ is a stationary ARMA process.
- There are three important values which characterize an ARIMA process:
  - p, the order of the autoregressive component
  - d, the number of differences needed to arrive at a stationary ARMA(p, q) process
  - q, the order of the moving average component.
- In particular, we have the general relationship:

$$Y_t \text{ is ARIMA}(p, d, q) \Longleftrightarrow \nabla^d Y_t \text{ is ARMA}(p, q).$$

Expressions form for ARIMA(p,d,q)

- A stationary ARMA(p, q) process can be represented as

$$\phi(B)Y_t = \theta(B)e_t.$$

- An ARIMA($p, 1, q$) process can be written succinctly as

$$\phi(B)(1 - B)Y_t = \theta(B)e_t$$

  because $\nabla Y_t = (1 - B)Y_t$ follows an ARMA(p, q).

- An ARIMA($p, 2, q$) process can be written as

$$\phi(B)(1 - B)^2 Y_t = \theta(B)e_t$$

  because $\nabla^2 Y_t = (1 - B)^2 Y_t$ follows an ARMA(p, q).

- In general, an ARIMA(p, d, q) process can be written as

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t.$$

Special cases of ARIMA(p,d,q)

- In practice, there will rarely be a need to consider values of the differencing order $d > 2$.
- Most real data sets can be coerced into a stationarity ARMA process by taking one or two differences.
- AR models, MA models, and ARMA models all are members of the ARIMA(p, d, q) family, e.g.,
  - AR$(p) \longleftrightarrow$ ARIMA$(p, 0, 0)$
  - MA$(q) \longleftrightarrow$ ARIMA$(0, 0, q)$
  - ARMA$(p, q) \longleftrightarrow$ ARIMA$(p, 0, q)$
  - ARI$(p, d) \longleftrightarrow$ ARIMA$(p, d, 0)$
  - IMA$(d, q) \longleftrightarrow$ ARIMA$(0, d, q)$.

Example 5.3. Two ARIMA(p, d, q) processes

Example

Identify each process

(a) $Y_t = 1.7Y_{t-1} - 0.7Y_{t-2} + e_t$

(b) $Y_t = 1.5Y_{t-1} - 0.5Y_{t-2} + e_t - e_{t-1} + 0.25e_{t-2}$

as an ARIMA(p, d, q) process. That is, specify p, d, and q.

Solutions for Example 5.3 (a)

- Upon first glance, this looks like an AR(2) process with $\phi_1 = 1.7$ and $\phi_2 = -0.7$.
- Upon closer inspection, we see that this process is not stationary because the AR(2) stationary conditions

$$\phi_1 + \phi_2 < 1 \qquad \phi_2 - \phi_1 < 1 \qquad |\phi_2| < 1$$

are not met with $\phi_1 = 1.7$ and $\phi_2 = -0.7$.
- We can write this process as

$$\begin{aligned} Y_t - 1.7Y_{t-1} + 0.7Y_{t-2} = e_t &\iff \left(1 - 1.7B + 0.7B^2\right) Y_t = e_t \\ &\iff (1 - 0.7B)(1 - B) Y_t = e_t, \end{aligned}$$

- $\{\nabla_t = Y_t - Y_{t-1}\}$ is an AR(1) process with $\phi = 0.7$.
- $\{Y_t\}$ is an ARIMA(1,1,0), or ARI(1,1), process with $\phi = 0.7$.
- A realization of this ARI(1,1) process is shown in Figure 5.4.

Figure 5.4. Simulated ARI(1,1) and its first difference

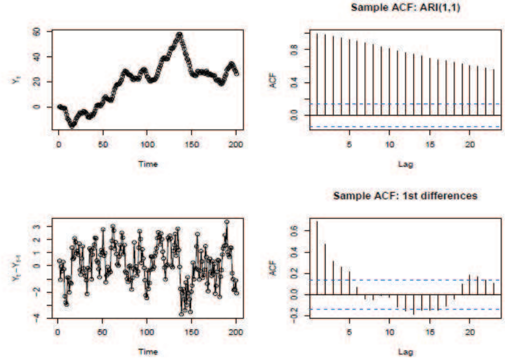

Figure 5.4: *Top: Simulated ARI(1,1) realization, with* $\phi = 0.7$, $n = 200$, *and* $\sigma_e^2 = 1$, *and the sample ACF. Bottom: First difference process with sample ACF.*

Solutions for Example 5.3 (b)

- Upon first glance, this looks like an ARMA(2,2) process.
- But this process is not stationary either. To see why, note that we can write this process as

$$
\begin{aligned}
& Y_t - 1.5Y_{t-1} + 0.5Y_{t-2} = e_t - e_{t-1} + 0.25e_{t-2} \\
\Longleftrightarrow\ & \left(1 - 1.5B + 0.5B^2\right) Y_t = \left(1 - B + 0.25B^2\right) e_t \\
\Longleftrightarrow\ & \left(1 - 0.5B\right) \left(1 - B\right) Y_t = \left(1 - 0.5B\right)^2 e_t \\
\Longleftrightarrow\ & \left(1 - B\right) Y_t = \left(1 - 0.5B\right) e_t
\end{aligned}
$$

- $\{\nabla_t = Y_t - Y_{t-1}\}$ is an MA(1) process with $\theta = 0.5$.
- $\{Y_t\}$ is an ARIMA(0,1,1), or IMA(1,1), process with $\theta = 0.5$.
- A realization of this IMA(1,1) process is shown in Figure 5.5.

Figure 5.5 Simulated IMA(1,1) series and its first difference

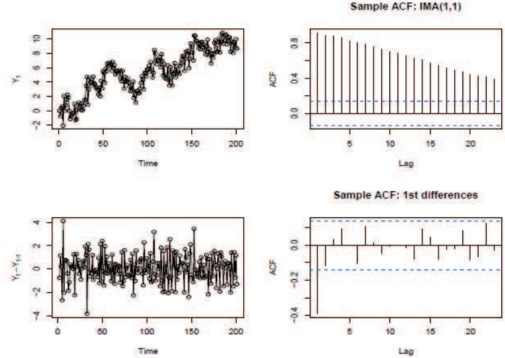

Figure 5.5: *Top: A simulated IMA(1, 1) realization, with $\theta = 0.5$, $n = 200$, and $\sigma_e^2 = 1$, and the sample ACF. Bottom: First difference process with sample ACF.*

5.2.1. IMA(1,1) process

- An IMA(1,1) process is an ARIMA$(0, 1, 1)$ process and is given by

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1}.$$

- This process is of importance because of its popularity in modeling time series in economics applications. Note that if $\theta = 0$, the IMA(1,1) reduces to a random walk.
- An IMA(1,1) process can be written as

$$(1 - B)Y_t = (1 - \theta B)e_t.$$

- This process is not stationary since the root of the AR characteristic polynomial $\phi(x) = 1 - x$ does not exceed 1 in absolute value.
- The root of $\phi(x)$ is $x = 1$, a so-called unit root. It is clear that $\{\nabla Y_t = Y_t - Y_{t-1}\}$ follows an MA(1) model with parameter θ.
- As usual, this ∇Y_t process is invertible if $|\theta| < 1$.

ACF of IMA(1,1) process

- Suppose that we fit an IMA(1,1) model to a series of n_0 observations $Y_1, Y_2, \cdots, Y_{n_0}$ (we'll learn how to do this later).
- With reference to this time origin n_0, for $t > n_0$, we can write by successive substitutions

$$\begin{aligned} Y_t &= Y_{t-1} + e_t - \theta e_{t-1} = Y_{t-2} + e_t + (1-\theta)e_{t-1} - \theta e_{t-2} \\ &\cdots \\ &= Y_{n_0} + e_t + (1-\theta)e_{t-1} + \cdots + (1-\theta)e_{n_0+1} - \theta e_{n_0}. \end{aligned}$$

- Similarly, for $t - k > n_0$,

$$Y_t = Y_{n_0} + e_{t-k} + (1-\theta)e_{t-k-1} + \cdots + (1-\theta)e_{n_0+1} - \theta e_{n_0}.$$

ACF of IMA(1,1) process (continued)

- Treating Y_{n_0} and e_{n_0} as constants (they are known values with respect to the fixed origin n_0), direct calculation shows that

$$\text{var}(Y_t) = \left[1 + (t - 1 - n_0)(1-\theta)^2\right] \sigma_e^2$$
$$\text{var}(Y_{t-k}) = \left[1 + (t - k - 1 - n_0)(1-\theta)^2\right] \sigma_e^2.$$

- It can also be shown that

$$\text{Cov}(Y_t, Y_{t-k}) = \left[(1-\theta) + (t - k - 1 - n_0)(1-\theta)^2\right] \sigma_e^2$$

and that

$$\text{corr}(Y_t, Y_{t-k}) = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t-k})}}$$
$$= \frac{\left[(1-\theta) + (t - k - 1 - n_0)(1-\theta)^2\right]}{\sqrt{\left[1 + (t - 1 - n_0)(1-\theta)^2\right]\left[1 + (t - k - 1 - n_0)(1-\theta)^2\right]}}.$$

ACF of IMA(1,1) process (continued)

- The variance var(Y_t) depends on time t. In fact, it is unbounded; i.e., as t gets large, so does var(Y_t).
- The ACF depend on t.
- If n_0 is large (i.e., much larger than k), then

$$\text{corr}(Y_t, Y_{t-k}) \approx 1.$$

- This implies that the ACF will decay slowly as k increases.

5.2.2. IMA(2,2) process

- An ARIMA(p, d, q) process with $p = 0$, $d = 2$, and $q = 2$ is called an IMA(2,2) process and is given by

$$(1 - B)^2 Y_t = \left(1 - \theta_1 B - \theta_2 B^2\right) e_t,$$

or, equivalently,

$$\nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

- In terms of the original process values, an IMA(2,2) process is

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}.$$

ACF of IMA(2,2)

The calculations for variance, autocovariance, and autocorrelation are even more difficult for an IMA(2,2) process than for an IMA(1,1) process! It suffices to note that

- the variance of Y_t increases as t does (perhaps very quickly).
- the autocorrelation $\text{corr}(Y_t, Y_{t-k}) \approx 1$, for most small k and
- the ACF decays very slowly as k increases.

Figure 5.6. Simulated IMA(2,2) and its second difference

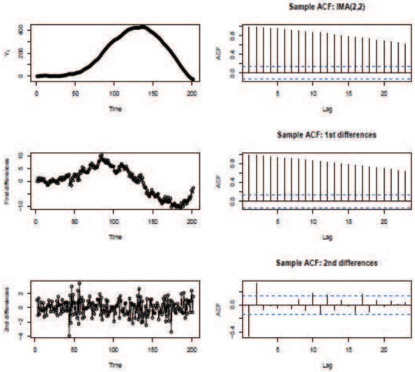

Figure 5.6: *Top: An IMA(2,2) realization with* $n = 200$, $\theta_1 = 0.3$, $\theta_2 = -0.3$, *and* $\sigma_e^2 = 1$. *Middle: First difference process. Bottom: Second difference process.*

- The defining characteristic of the simulated IMA(2,2) process is its very strong autocorrelation at early lags. This is also seen in the sample ACF.
- The process $\{\nabla Y_t\}$, which is that of an IMA(1,2), is also clearly nonstationary to the naked eye. This is also seen in the sample ACF.
- The process $\{\nabla^2 Y_t\}$, in theory, is an invertible MA(2) process. This is seen in the sample ACF for the second differences, that is, note how r_1 and r_2 are both outside the white noise bounds and how r_k, for $k > 2$, are mostly negligible.

5.2.3. API(1,1) process

- An ARIMA(p, d, q) process with $p = 1$, $d = 1$, and $q = 0$ is called an ARI(1,1) process and is given by

$$(1 - \phi B)(1 - B)Y_t = e_t,$$

or, equivalently,

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + e_t.$$

- $\nabla Y_t = (1 - B)Y_t$ is a stationary AR(1) process if $|\phi| < 1$.

Remark for API(1,1) process

- Upon first glance, the process

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t$$

  looks like an AR(2) model.

- This process is not stationary since the coefficients satisfy $(1+\phi) - \phi = 1$; this violates the stationarity requirements for the AR(2) model.

- A simulated ARI(1,1) process appeared in Figure 5.4.

5.2.4. ARIMA(1,1,1) process

- An ARIMA(p, d, q) process with $p = 1$, $d = 1$, and $q = 1$ is called an ARIMA(1,1,1) process and is given by

$$(1 - \phi B)(1 - B)Y_t = (1 - \theta B)e_t,$$

or, equivalently,

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + e_t - \theta e_{t-1}.$$

- $\nabla Y_t = (1 - B)Y_t$ follows an ARMA(1,1) process which is stationary and invertible if and only if $|\phi| < 1$ and $|\theta| < 1$.
- A simulated ARI(1,1) process appears in Figure 5.7.

Figure 5.7. Simulated ARIMA(1,1,1) and its first difference

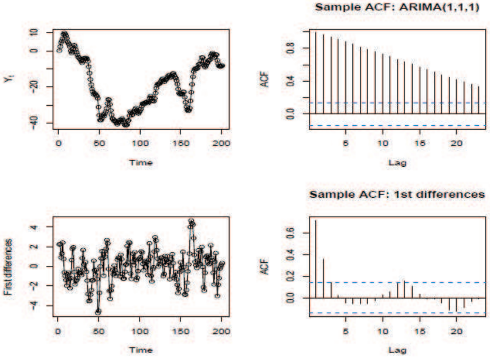

Figure 5.7: *Top: An ARIMA*(1, 1, 1) *realization and sample ACF with* $n = 200$, $\phi = 0.5$, $\theta = -0.5$, *and* $\sigma_e^2 = 1$. *Bottom: First difference process with sample ACF.*

5.3. Constant terms in ARIMA models

- In general, an ARIMA(p, d, q) process can be written as

$$\phi(B)(1-B)^d Y_t = \theta(B)e_t.$$

- An extension of this model is

$$\phi(B)(1-B)^d Y_t = \theta_0 + \theta(B)e_t,$$

  where the parameter θ_0 is a constant term.
- The parameter θ_0 plays very different roles for the cases:
  - $d = 0$ (a stationary ARMA model)
  - $d > 0$ (a nonstationary model).

Stationary case: ARIMA$(p, 0, q)$

- Suppose that $d = 0$, in which case the model becomes

$$\phi(B)Y_t = \theta(B)e_t,$$

a stationary ARMA model, where the AR and MA characteristic polynomials are given by

$$\begin{aligned}\phi(B) &= \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)\\ \theta(B) &= \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right).\end{aligned}$$

Stationary case: ARIMA$(p, 0, q)$ (continued)

- To examine the effects of adding a constant term, replace Y_t with $Y_t - \mu$ ($\mu = \mathsf{E}(Y_t)$). Due to $B^k \mu = \mu$ for all k, we get

$$\begin{aligned}\phi(B)(Y_t - \mu) = \theta(B)e_t &\iff \phi(B)Y_t - \phi(B)\mu = \theta(B)e_t \\ &\iff \phi(B)Y_t = \phi(1)\mu + \theta(B)e_t \\ &\iff \phi(B)Y_t = \theta_0 + \theta(B)e_t,\end{aligned}$$

  where $\phi(1) = (1 - \phi_1 - \cdots - \phi_p)$ and $\theta_0 = \phi(1)\mu$.

- If $\mu = 0$ (this was assumption in Chapter 4), then the constant $\theta_0 = 0$ as well, and we are back at $\phi(B)Y_t = \theta(B)e_t$.
- Adding a constant term μ to the process $\{Y_t\}$ does not affect the stationarity properties of the $\{Y_t\}$.

Non-stationary case: ARIMA(p, d, q)

- The impact of adding a constant term θ_0 to the model when $d > 0$ is profoundly different.
- Consider the ARIMA$(0, 1, 0)$ family, the simplest case, a random walk. Then

$$(1 - B)Y_t = \theta_0 + e_t \Longleftrightarrow Y_t = \theta_0 + Y_{t-1} + e_t.$$

  In Chapter 2, we called this model a random walk with drift.
- We can write it as, by successive substitution,

$$Y_t = \theta_0 + Y_{t-1} + e_t = (t-k)\theta_0 + Y_k + e_t + e_{t-1} + \cdots + e_{t-k+1}.$$

- The $\{Y_t\}$ contains a linear deterministic trend with slope θ_0.
- Figure 2.3 displays four random walk with drift processes.

Remark for non-stationary case: ARIMA(p, d, q)

- The previous finding holds for any ARIMA$(p, 1, q)$ model, that is, adding a constant term θ_0 induces a linear deterministic trend.
- Adding a constant term θ_0 to an ARIMA$(p, 2, q)$ model induces a quadratic deterministic trend,
- Adding a constant term θ_0 to an ARIMA$(p, 3, q)$ model induces a cubic deterministic trend, and so on.
- For very large t, the constant (deterministic trend) term can become very dominating so that it forces the time series to follow a very stringent deterministic pattern.
- Therefore, a constant term should be added to a nonstationary ARIMA model (i.e., $d > 0$) only if it is strongly warranted.

5.4. Transformations

- If we are faced with the task of modelling a nonstationary time series, it can be useful to transform the data first before taking any differences (or "detrending" the data if we use regression methods).
- For example, if there is clear evidence of nonconstant variance over time (e.g., the variance increases as time does, etc.), then a suitable transformation to the data might remove (or lessen the impact of) the nonconstant variance pattern.
- Applying a transformation is regarded as a "first step" to take before using differencing as a means to achieve stationarity.

Example 5.4. Monthly electricity usage in US

Example

- Consider the monthly electricity usage in the United States (usage from coal, natural gas, nuclear, petroleum, and wind) for the time period between January, 1973 and December, 2005.
- Figure 5.8 displays the times series plot of monthly electricity usage in US.

Figure 5.8. Electricity data plot

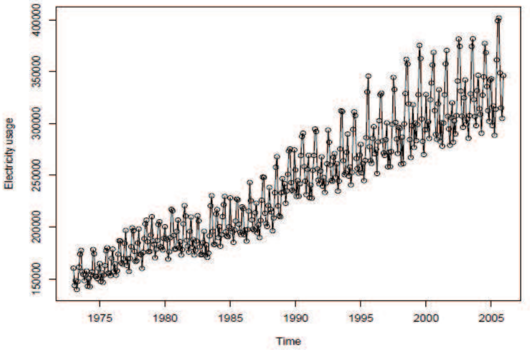

Figure 5.8: *Electricity data. Monthly U.S. electricity generation, measured in millions of kilowatt hours, from 1/1973 to 12/2005.*

Remarks for the time series of monthly electricity data

- From the plot, we can see that there is an increasing variance pattern over time; e.g., the series is much more variable at later years than it is in earlier years.
- Time series that exhibit this "fanning out" shape are not stationary because the variance changes over time.
- Before we model these data, we can first apply a transformation to make the variance constant (that is, we would like to first "stabilize" the variance).

Variance stabilizing transformations

- Suppose that the variance of nonstationary process $\{Y_t\}$ can be written as
$$\text{var}(Y_t) = c_0 f(\mu_t),$$
where $\mu_t = \text{E}(Y_t)$ and c_0 is a positive constant free of μ_t.
- The variance is not constant because it is a function of μ_t, which is changing over time.
- How do we find a function T so that the transformed series $T(Y_t)$ has constant variance? That is, $\text{var}(T(Y_t)) = c_1$.
- Consider approximating T by a first-order Taylor series expansion about the point μ_t, that is,
$$T(Y_t) \approx T(\mu_t) + T'(\mu_t)(Y_t - \mu_t).$$

Variance stabilizing transformations (continued)

- $\text{var}[T(Y_t)] \approx \text{var}[T'(\mu_t)(Y_t - \mu_t)] = c_0[T'(\mu_t)]^2 f(\mu_t)$.
- We want to find the function T such that

$$\text{var}[T(Y_t)] \approx c_0[T'(\mu_t)]^2 f(\mu_t) \stackrel{\text{set}}{=} c_1,$$

  where c_1 is a constant free of μ_t.
- For $T'(\mu_t)$, we get the differential equation

$$T'(\mu_t) = \sqrt{\frac{c_1}{c_0 f(\mu_t)}} = \frac{c_2}{\sqrt{f(\mu_t)}}.$$

  where c_2 is constant, free of μ_t.
- Integrating both sides, we get

$$T(\mu_t) = \int \frac{c_2}{\sqrt{f(\mu_t)}} d\mu_t + c_3,$$

  where c_3 is a constant free of μ_t.

Variance stabilizing transformations (continued)

- If $\mathrm{var}(Y_t) = c_0\mu_t$, so that the variance of the series is proportional to the mean, then

$$T(\mu_t) = \int \frac{c_2}{\sqrt{\mu_t}} d\mu_t = 2c_2\sqrt{\mu_t} + c_3,$$

where c_3 is a constant free of μ_t.

- If we take $c_2 = 1/2$ and $c_3 = 0$, we see that the square root of the series, $T(Y_t) = \sqrt{Y_t}$, will provide a constant variance.

Variance stabilizing transformations (continued)

- If $\text{var}(Y_t) = c_0\mu_t^2$, so that the standard deviation of the series is proportional to the mean, then

$$T(\mu_t) = \int \frac{c_2}{\sqrt{\mu_t^2}} d\mu_t = c_2 \ln(\mu_t) + c_3,$$

  where c_3 is a constant free of μ_t.
- If we take $c_2 = 1$ and $c_3 = 0$, we see that the logarithm of the series, $T(Y_t) = \ln(Y_t)$, will provide a constant variance.

Variance stabilizing transformations (continued)

- If $\text{var}(Y_t) = c_0 \mu_t^4$, so that the standard deviation of the series is proportional to the square of the mean, then

$$T(\mu_t) = \int \frac{c_2}{\sqrt{\mu_t^4}} d\mu_t = c_2 \left(-\frac{1}{\mu_t} \right) + c_3,$$

  where c_3 is a constant free of μ_t.

- If we take $c_2 = -1$ and $c_3 = 0$, we see that the reciprocal of the series, $T(Y_t) = 1/Y_t$, will provide a constant variance.

Box-Cox transformations

- Box and Cox (1964) introduced a power transformation in order to stabilize the variance.
- The transformation is defined by

$$T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(Y_t), & \lambda = 0, \end{cases}$$

  where λ is called the transformation parameter.
- To see why the logarithm transformation $T(Y_t) = \ln(Y_t)$ is used when $\lambda = 0$, by L'Hoptial's Rule,

$$\lim_{\lambda \to 0} \frac{Y_t^\lambda - 1}{\lambda} = \lim_{\lambda \to 0} \frac{Y_t^\lambda \ln(Y_t)}{1} = \ln(Y_t).$$

Box-Cox transformations (continued)

- Some common values of λ, and their implied transformations are listed in the below table.

λ	$T(Y_t)$	Description	Variance style
-2.0	$\frac{1}{Y_t^2}$	Inverse square	$\text{var}(Y_t) = c\mu_t^6$
-1.0	$\frac{1}{Y_t}$	Reciprocal	$\text{var}(Y_t) = c\mu_t^4$
-0.5	$\frac{1}{\sqrt{Y_t}}$	Inverse square root	$\text{var}(Y_t) = c\mu_t^3$
0.0	$\ln(Y_t)$	Logarithm	$\text{var}(Y_t) = c\mu_t^2$
0.5	$\sqrt{Y_t}$	Square root	$\text{var}(Y_t) = c\mu_t$
1.0	Y_t	Identity (no transformation)	$\text{var}(Y_t) = c$
2.0	Y_t^2	Square	$\text{var}(Y_t) = c\mu_t^{-2}$

Remarks for a variance stabilizing transformation

- A variance stabilizing transformation can only be performed on a positive series, that is, when $Y_t > 0$, for all t.
- If some or all of our series is negative, we can simply add a positive constant c to each observation so that the data become positive. Adding c will not affect the (non)stationarity properties of $\{Y_t\}$.
- A variance stabilizing transformation, if needed, should be performed before taking any data differences.
- A transformation performed to stabilize the variance also, frequently, improve the approximation to normality. We will discuss the normality assumption later (Chapters 7-8).

Determining λ in Box-Cox transformations

- We can let the data "suggest" a suitable transformation (i.e., a value of λ) in the Box-Cox power family.
- Treat λ as an unknown parameter, write out the log-likelihood function of the data (under the normality assumption), and find the maximum likelihood estimator (MLE) of λ.
- The full details of this procedure are omitted. The good news is that the authors of CC have written an R function to do it. We provide SAS program figure59.sas for it.
- The function also provides an approximate 95 percent confidence interval for λ, which is constructed using the large sample properties of MLEs.

Figure 5.9. Electricity data: Log-likelihood function vs λ

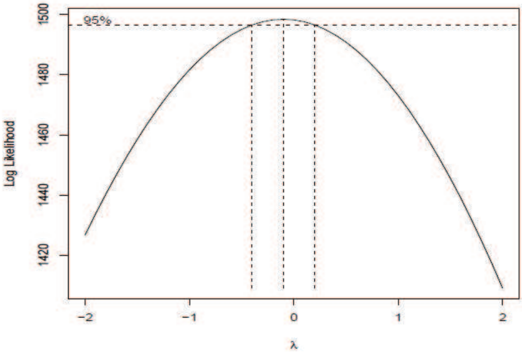

Figure 5.9: *Electricity data. Log-likelihood function versus λ. Note that λ is on the horizontal axis.*

Remarks on λ in Box-Cox transformations to electricity data

- Figure 5.9 displays the (profile) log-likelihood of λ.
- The optimal value of λ is approximately -0.15.
- This value of λ is not that easy to interpret; in other words, the transformation $T(Y_t) = Y_t^{-0.15}$ makes little practical sense.
- An approximate 95 percent confidence interval for λ is about $(-0.37, 0.06)$.
- Because $\lambda = 0$ is in this interval, this suggests that a log transformation $T(Y_t) = \log(Y_t)$ would not be unreasonable.
- The base of the logarithm does not affect the impact of the transformation on the series. Using different logarithm bases (e.g., $10, 2, e$) only changes the scale of the measurement.

Figure 5.10. Transformed electricity data by $\log_{10}$

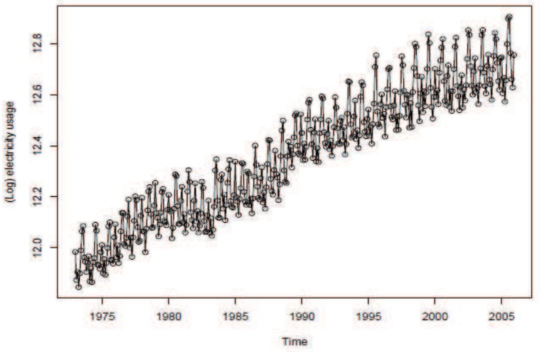

Figure 5.10: *Electricity data. Monthly U.S. electricity generation, measured on the log scale. A* $\log_{10}$ *transformation has been used.*

Analysis to the transformed electricity data

- The transformed process $\{\log Y_t\}$ appears in Figure 5.10.
- Taking logarithms has lessened the nonconstant variance (although there still is a mild increase in the variance over time).
- However, we still have a pronounced linear trend. Therefore, it is necessary to consider the first difference process (on the log scale), given by

$$W_t = \log Y_t - \log Y_{t-1} = \nabla \log Y_t.$$

Figure 5.11. Time plot for difference of transformed electricity data and ACF

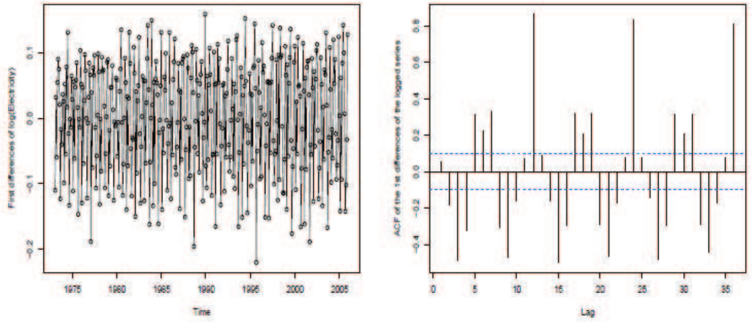

Figure 5.11: *Electricity data. Left:* $W_t = \log Y_t - \log Y_{t-1}$, *the first differences of the log process. Right: The sample autocorrelation function of the* $\{W_t\}$ *process.*

Analysis to the transformed electricity data (continued)

- This process is plotted in Figure 5.11 along with the sample ACF of the difference ∇Y_t process.
- The data now appear to have a constant mean (after differencing).
- However, the sample ACF suggests that there is still a large amount of structure in the data that remains after differencing the transformed series.
- In particular, there looks to be significant autocorrelations that appear according to a seasonal pattern.
- We will consider seasonal models (that model this type of variability) in Chapter 10.

About the differences of a log-transformed series of data

- Taking the differences of a log-transformed series of data arises in financial applications where Y_t tends to have stable percentage changes over time. For example, Y_t denotes stock price, portfolio return, etc..

- See pp 99 (CC).

chapter 5 is over

Thank you for your attention!