Linear Time Series Analysis

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Revised on April 1, 2017

Today we are going to learn...

1. Moving average process
2. Autoregressive process
3. Autoregressive moving average process
4. ARIMA model estimation
5. Time series forecasting

Infinite moving average process

- Recall the stable linear process we mentioned in stationary time series
$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$
- We can rewrite it in terms of the **backshift operator** (B) (same as lag operator) ,
$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} = \mu + \sum_{i=0}^{\infty} \psi_i B^i \epsilon_t = \mu + \Psi(B)\epsilon_t$$
- This is called the **infinite moving average**.
- Some properties
  - $\sum_{i=0}^{\infty} \psi_i^2 < \infty$
  - **Wold's decomposition theorem**: any **nondeterministic** weakly stationary time series y_t can be represented as the above form. (i.e. a stationary time series can be seen as the weighted sum of the present and past random "disturbances".)
  - The above means we only need to find ψ.
  - The infinite moving average does not help us much in our modeling and forecasting efforts as it implicitly requires the estimation of the infinitely many weights.

Finite moving average process

- The finite moving average, MA(q) is a special case of infinite moving average when we require
  - $\psi_0 = 0$ (by convention)
  - i is from 0 to q where $q < t$
  - ϵ_t is the white noise.
  - The expression is usually written as

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q}$$

- A MA(q) process is always stationary regardless of values of the weights.

$$\begin{aligned} E(y_t) =& E(\mu + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q}) = \mu \\ Var(y_t) =& Var(\mu + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q}) \\ =& Var(\epsilon_1) + \theta_1^2 Var(\epsilon_{t-1}) + ... + \theta_p^2 Var(\epsilon_{t-p}) \\ =& (1 + \theta_1^2 + ... + \theta_p^2)\sigma^2 \end{aligned}$$

- The interpretation of MA process: At any given time, of the infinitely many past disturbances, only a finite number of those disturbances "contribute" to the current value of the time series.

Finite moving average process

- The autocovariance function for $k < q$

$$
\begin{aligned}
\gamma_y(1) &= Cov(y_t, y_{t+1}) = Cov(\mu + \epsilon_t - \theta_1\epsilon_{t-1} - ... - \theta_q\epsilon_{t-q}, \\
&\qquad\qquad\qquad\qquad \mu + \epsilon_{t+1} - \theta_1\epsilon_t - ... - \theta_q\epsilon_{t+1-q}) \\
&= \sigma^2(-\theta_1 + \theta_1\theta_2 + ... + \theta_{q-1}\theta_q) \\
\gamma_y(2) &= Cov(y_t, y_{t+2}) = Cov(\mu + \epsilon_t - \theta_1\epsilon_{t-1} - ... - \theta_q\epsilon_{t-q}, \\
&\qquad\qquad\qquad\qquad \mu + \epsilon_{t+2} - \theta_1\epsilon_{t+1} - ... - \theta_q\epsilon_{t+2-q}) \\
&= \sigma^2(-\theta_2 + \theta_1\theta_3 + ... + \theta_{q-2}\theta_q) \\
&\dots \\
\gamma_y(k) &= Cov(y_t, y_{t+k}) = Cov(\mu + \epsilon_t - \theta_1\epsilon_{t-1} - ... - \theta_q\epsilon_{t-q}, \\
&\qquad\qquad\qquad\qquad \mu + \epsilon_{t+k} - \theta_{t+k}\epsilon_{t+k-1} - ... - \theta_q\epsilon_{t+k-q}) \\
&= \sigma^2(-\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_q)
\end{aligned}
$$

- The autocovariance function for $k \geqslant q$ is zero. The ACF for MA model "cuts off" (**decays**) after lag q. (**We can use it to identify** q **in application**)
- The autocorrelation function for $k < q$ (**why?**)

$$
\rho_y(k) = \frac{\gamma_y(k)}{\gamma_y(0)} = \frac{-\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_q}{1 + \theta_1^2 + ... + \theta_p^2}
$$

The first-order moving average process

- When $q = 1$, we have the special case called MA(1) model

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$$

- The autocovariance function is

$$\gamma_y(0) = \sigma^2(1+\theta_1^2),\ \gamma_y(1) = -\theta_1\sigma^2,\ \gamma_y(k) = 0$$

- The autocorrelation function is

$$\rho_y(1) = \frac{-\theta_1}{1+\theta_1^2},\ \rho_y(k) = 0 \text{ for } k > 1$$

  which indicates

  - the autocorrelation function cuts off after lag 1, and
  - the first lag autocorrelation in MA(l) is bounded as $|\rho_y(1)| \leqslant 1/2$.

- **R example with MA(1)** $y_t = 40 + \epsilon_t + 0.8\epsilon_{t-1}$

The second-order moving average process

- When $q = 2$, we have the special case called MA(2) model

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

- The autocovariance function is

$$\gamma_y(0) = \sigma^2(1+\theta_1^2+\theta_2^2),\ \gamma_y(1) = (-\theta_1+\theta_1\theta_2)\sigma^2,\ \gamma_y(2) = \sigma^2(-\theta_2),\ \gamma_y(k) = 0$$

- The autocorrelation function is

$$\rho_y(1) = \frac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2},\ \rho_y(1) = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2},\ \rho_y(k) = 0 \text{ for } k > 2$$

  which indicates

  - the autocorrelation function cuts off after lag 2

- **R example with MA(2)** $y_t = 40 + \epsilon_t + 0.7\epsilon_{t-1} - 0.28\epsilon_{t-2}$

First-order autoregressive process

- Let's consider a simple version of MA process

$$y_t = \mu + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \ldots$$

- From the above, we also have

$$y_{t-1} = \mu + \epsilon_{t-1} + \phi \epsilon_{t-2} + \phi^2 \epsilon_{t-3} + \ldots$$

- Then we can combine the two together and the following (**why?**)

$$\begin{aligned} y_t &= \mu - \phi\mu + \phi y_{t-1} + \epsilon_t \\ &= \delta + \phi y_{t-1} + \epsilon_t \end{aligned}$$

  where $\delta = (1 - \phi)\mu$ and $|\phi| < 1$.
- The process is called a **first-order autoregressive process**.
- Because the equation can be seen as a regression of y_t on y_{t-1} and hence the term autoregressive process.
- The assumption of $|\phi| < 1$ that is made to make the weights decay exponentially in time.
- **Question:** Given an AR(1) expression, can you write the infinite MA representation down in terms all ϵ?

First-order autoregressive process

- The AR(1) process is stationary if $|\phi| < 1$.
- The mean of a stationary ar(1) process is (**Why?**)

$$E(y_t) = \mu = \frac{\delta}{1-\phi}$$

- The variance is $\sigma^2/(1-\phi^2)$ due to the fact that

$$\begin{aligned} Var(y_t) &= Var(\mu + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots) \\ &= \sigma^2(1 + \phi^2 + \phi^4 + \phi^6 + \ldots) \\ &= \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

Or under the stationary one may verify this as

$$\begin{aligned} Var(y_t) &= Var(\delta + \phi y_{t-1} + \epsilon_t) \\ &= \phi^2 Var(y_{t-1}) + Var(\epsilon_t) \\ &= \phi^2 Var(y_t) + \sigma^2 \end{aligned}$$

First-order autoregressive process

- The autocovariance function of a **stationary** AR(1)

$$\gamma(k) = \sigma^2 \phi^k \frac{1}{1-\phi^2}, \text{ for } k = 0, 1, \dots$$

- Correspondingly, the autocorrelation function for a stationary AR(I) process is

$$\rho(k) = \phi^k, \text{ for } k = 0, 1, \dots$$

- The ACF for a stationary AR(1) process has an **exponential decay** form.
- **R example**: AR(1) with $y_t = 8 + 0.8 y_{t-1} + \epsilon_t$
- **Take home questions:** Verify the two results (Hint: use the infinite MA representation).

Random Walk

- A special case of AR(1) process is the random Walk

$$y_t = \mu + y_{t-1} + \epsilon_t$$

- The intercept μ is called the **drift**. If $\mu = 0$, it is called random walk without drift. If $\mu =\neq 0$, it is called random walk without drift.
- A random walk is not stationary.
- **Examples**

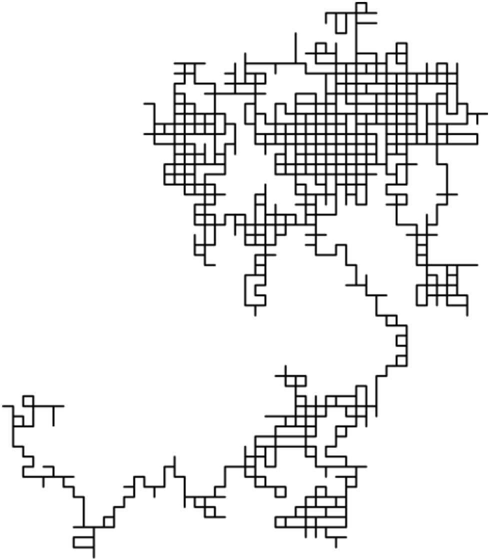

Figure: Random Walk in 2D

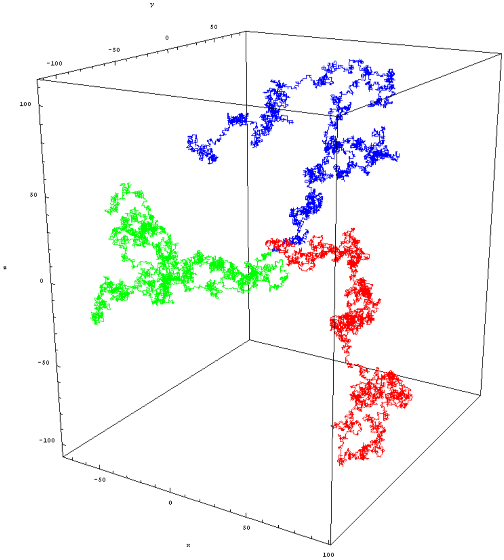

Figure: Random Walk in 3D

Second-order autoregressive process

- We can naturally extend the AR(1) to AR(2)

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

- This can be represented in the backshift operator

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \delta + \epsilon_t$$

- We will see the advantage of this rewriting
  - Let $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$
  - Let $\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i = \Phi(B)^{-1}$. Thus $\Psi(B)\Phi(B) = 1$
  - Now we apply $\Phi(B)^{-1}$ to both sides of $\Phi(B) y_t = \delta + \epsilon_t$

$$\begin{aligned} \Phi(B)\Phi(B)^{-1} y_t &= \delta + \epsilon_t \\ y_t &= \Phi(B)^{-1}\delta + \Phi(B)^{-1}\epsilon_t \\ &= \mu + \Psi(B)\epsilon_t \\ &= \mu + (\sum_{i=0}^{\infty} \psi_i B^i)\epsilon_t \end{aligned}$$

Second-order autoregressive process

- According to Wold's decomposition theorem, we would like to find ϕ
- We start from the equality

$$
\begin{aligned}
\Psi(B)\Phi(B) &= 1 \\
(1-\phi_1 B - \phi_2 B^2)\left(\sum_{i=0}^{\infty}\psi_i B^i\right) &= 1 \\
\psi_0 + (\psi_1 - \phi_1\psi_0)B + (\psi_2 - \phi_1\psi_1 - \phi_2\psi_0)B^2 + \\
\dots + (\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2})B^j + \dots &= 1
\end{aligned}
$$

implies

$$
\begin{aligned}
\psi_0 &= 1 \\
\psi_1 - \phi_1\psi_0 &= 0 \\
&\dots \\
\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} &= 0 \\
&\dots
\end{aligned}
$$

- **Why?**
- **R example:** `ARMAtoMA(ar = numeric(), ma = numeric(), lag.max)`

Second-order autoregressive process

- Notice that the ψ_j satisfy the second-order **linear difference equation**.
- It be expressed as the solution to this equation in terms of the two roots of the associated polynomial

$$m^2 - \phi_1 m - \phi_2 = 0$$

- If the absolute value of all the roots are smaller than one, then the AR(2) model is stationary.
- This is the way to test if an AR process is stationary or not.

Second-order autoregressive process

- Under the conditions for the stationarity of an AR(2) time series
- the mean:

$$\begin{aligned} E(y_t) &= \delta + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + E(\epsilon_t) \\ &= \delta + \phi_1 E(y_t) + \phi_2 E(y_t) \\ &\Rightarrow \mu = E(y_t) = \frac{\delta}{1 - \phi_1 - \phi_2} \end{aligned}$$

- the autocovariance function

$$\begin{aligned} \gamma(k) &= Cov(y_t, y_{t-k}) = Cov(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, y_{t-k}) \\ &= \phi_1 Cov(y_{t-1}, y_{t-k}) + \phi_2 Cov(y_{t-2}, y_{t-k}) + Cov(\epsilon_t, y_{t-k}) \\ &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \begin{cases} \sigma^2, & \text{if } k = 0 \\ 0, & elsewhere \end{cases} \end{aligned}$$

  which is called the **Yule-Walker** equation.
- Similarly, the autocorrelation function

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \text{ for k = 2, 3,...}$$

- **R example** with $y_t = 4 + 0.4 y_{t-1} + 0.5 y_{t-2} + \epsilon_t$

The AR(p) process

- The AR model of order p is of the form

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t$$

- Or we use the backshift operator representation

$$\begin{aligned}(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) y_t &= \delta + \epsilon_t \\ \Phi(B) y_t &= \delta + \epsilon_t\end{aligned}$$

- **The stationary condition:** if the roots of the associated polynomial are less than one in absolute value

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - ... - \phi_p = 0$$

The AR(p) process

- Under the conditions for the stationarity of an AR(2) time series
- the mean:

$$\mu = E(y_t) = \frac{\delta}{1 - \phi_1 - ... - \phi_p}$$

- the autocovariance function

$$\begin{aligned}\gamma(k) &= Cov(y_t, y_{t-k}) \\ &= \sum_{i=1}^{p} \phi_i \gamma(k-i) + \begin{cases} \sigma^2, & \text{if } k = 0 \\ 0, & elsewhere \end{cases}\end{aligned}$$

- Similarly, the autocorrelation function (**Yule-Walker equations** or ***p*th-order linear difference equations**)

$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i), \text{ for } k = 1{,}2{,}...$$

- The ACF for an AR(p) model can be found through the p roots of the associated polynomial

$$\rho(k) = c_1 m_1^k + c_2 m_2^k + ... + c_p m_p^k$$

for $k = 1, 2, ...$

Partial Autocorrelation Function

- We already know that
  - The ACF is expected to "cut off" after lag q for MA(q) model.
  - The ACF for AR(p) process will most likely have a exponential decay but does not tell us what the lag p is. the **partial autocorrelation function** (PACF) will do it for us.
- **The partial correlation**: the correlation between two variables after being adjusted for a common factor that may be affecting them.
- The partial correlation between X and Y after adjusting for Z is defined as

$$\mathrm{Corr}(X - \hat{X}, Y - \hat{Y})$$

where $\hat{X} = a_1 + b_1 Z$ and $\hat{Y} = a_2 + b_2 Z$

Partial Autocorrelation Function

- The **partial autocorrelation function** between y_t and y_{t-k} is the autocorrelation between y_t and y_{t-k} after adjusting for y_{t-1}, y_{t-2},...,y_{t-k+1} and y_{t-k}.
- For an AR(p) model the partial autocorrelation function between y_t and y_{t-k} for $k > p$ should be equal to zero.
- This can be used to detect the p in AR process

Partial Autocorrelation Function

- Recall the Yule-Walker equations for the ACF of AR(p) process

$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i), \text{ for } k = 1,2,...$$

which can be written in terms of matrices

$$\boldsymbol{\rho}_k = \mathbf{P}_k \boldsymbol{\phi}_k$$

where $\boldsymbol{\rho}_k = [\rho(1), ..., \rho(k)]'$, $\boldsymbol{\phi}_k = [\phi_{1k}, ..., \phi_{kk}]'$ and $\mathbf{P}_k$ is the matrix for all the lagged correlations.

- For any given k, $k = 1, 2, ...$ the last coefficient ϕ_{kk} is called the **partial autocorrelation** of the process at lag k.
- The PACF can be used in identifying the order of an AR process similar to how the ACF can be used for an MA process.
- The PACF for MA process exhibits an exponential decay pattern.
- For sample calculations, $\hat{\phi}_{kk}$, the sample estimate of ϕ_{kk}, is obtained by using the sample ACF.
- **R example** PACF for $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \epsilon_t$

Invertibility of MA Models

- The MA(q) process is said to be **invertible** if it has an absolutely summable infinite AR representation.
- Consider the MA(q) process

$$
\begin{aligned}
y_t &= \mu + (1 - \textstyle\sum_{i=1}^{q} \theta_j B^i)\epsilon_t \\
&= \mu + \Phi(B)\epsilon_t
\end{aligned}
$$

- If we multiply both sides with $\Phi(B)^{-1}$, we have

$$
\begin{aligned}
\Phi(B)^{-1} y_t &= \Phi(B)^{-1}\mu + \epsilon_t \\
\Psi(B) y_t &= \delta + \epsilon_t \\
(1 - \textstyle\sum_{i=1}^{\infty} \psi_i B^i) y_t &= \delta + \epsilon_t
\end{aligned}
$$

  by assuming $\Phi(B)\Psi(B) = 1$ where the last equation is the infinite AR representation of an MA(q) process

Invertibility of MA Models

- The ψ_i can be obtained from

$$\Phi(B)\Psi(B) = 1$$
$$(1 - \sum_{i=1}^{q} \theta_j B^i)(1 - \sum_{i=1}^{\infty} \psi_i B^i) = 1$$

- Again this can be checked through the roots of the associated polynomial

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - ... - \theta_q = 0$$

- If all the roots are less than one in absolute value, then MA(q) is said to have an **infinite AR representation**
- **R example (same as AR→ MA, but swap ar and ma parameters. Why?)**:
  `ARMAtoMA(ar = ma.numeric(), ma = ar.numeric(), lag.max)`

ARMA(p,q)

- ARMA model is mixed with autoregressive model and moving average model

$$
\begin{aligned}
y_t &= \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \\
y_t - \sum_{i=1}^{p} \phi_i y_{t-i} &= \delta + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \\
(1 - \sum_{i=1}^{p} \phi_i B^i) y_t &= \delta + (1 - \sum_{i=1}^{q} \theta_i B^i) \epsilon_t \\
\Phi(B) y_t &= \delta + \Theta(B) \epsilon_t
\end{aligned}
$$

  which is called the **ARMA(p,q) model**.
- The **stationary condition** of ARMA(p,q) is same as AR(p) model (**Why?**).

$$
m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - ... - \phi_p = 0
$$

ARMA(p,q)

- Under the stationary condition, ARMA(p,q) has an **infinite MA representation** by multiplying $\Phi(B)^{-1}$ to both sides of

$$\begin{aligned}\Phi(B)y_t &= \delta + \Theta(B)\epsilon_t \\ y_t &= \Phi(B)^{-1}\delta + \Phi(B)^{-1}\Theta(B)\epsilon_t \\ y_t &= \mu + \Phi(B)^{-1}\Theta(B)\epsilon_t\end{aligned}$$

- The **invertibility** of an ARMA process is related to the MA component and can be checked through the roots of the associated polynomial

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - ... - \theta_q = 0$$

ARMA(p,q)

- As in the stationarity and invertibility conditions, the ACF and PACF of an ARMA process are determined by the AR and MA components. respectively
- It can be shown that the ACF and PACF of an ARMA(p, q) both exhibit exponential decay and/or damped sinusoid patterns.
- The identification of the order of the ARMA(p, q) model is relatively more difficult.

TABLE 5.1 Behavior of Theoretical ACF and PACF for Stationary Processes

Model	ACF	PACF
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

ARIMA(p,d,q)

- An **autoregressive integrated moving average** (ARIMA) process of order p, d and q (ARIMA(p,d,q))

$$\Phi(B)(1-B)^d y_t = \delta + \Theta(B)\epsilon_t$$

- It can be viewed a ARMA model with ω_t

$$\Phi(B)\omega_t = \delta + \Theta(B)\epsilon_t$$

  where $\omega_t = (1-B)^d y_t$
- Question: write down the expression of the flowing models
  - ARIMA(1,1,1)
  - ARIMA(0,1,1)
  - ARIMA(1,1,0)
  - ARIMA(0,2,0)

Estimate an ARIMA model

↛ The three steps to build an ARIMA model

- A tentative model of the ARIMA class is identified through analysis of historical data.
- The unknown parameters of the model are estimated. (**what are the unknown parameters in ARIMA models?**)
- Through residual analysis, diagnostic checks are performed to determine the adequacy of the model, or to indicate potential improvements. (same as our usual linear regression model)

Model identification
$\looparrowright$ What model should be used? (AR, MA, ARMA, or ARIMA?)

- It is a piece of experience work.
- Use the sample ACF and PACF.
- **The unit root test** can also be performed to make sure that the differencing is indeed needed.

Model identification I

↬ Is the time series stationary ?

- **The unit root test** tests whether a time series variable is non-stationary using an autoregressive model. A well-known version is the **Dickey-Fuller type test**
- **The null hypothesis** $\beta = 1$ against the alternative hypothesis of $\beta < 1$.
- The regression model

$$x_t = c_t + \beta x_{t-1} + \delta_1 \Delta x_{t-1} + \cdots + \delta_{p-1} \Delta x_{t-p+1} + \varepsilon_t,$$

- **The intuition**: If the series is not integrated, then the lagged level of the series (x_{t-1}) will provide no relevant information in predicting the change in x_t besides the one obtained in the lagged changes.
- If the null hypothesis of $\beta = 1$ is rejected, there is no unit root presented.
- **The augmented Dickey-Fuller (ADF) unit-root test statistic**

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

Model identification II

↛ Is the time series stationary ?

- **Example 2.2** Consider the log series of U.S. quarterly GDP from 1947.I to 2008.IV. The series exhibits an upward trend, showing the growth of the U.S. economy, and has high sample serial correlations; see the lower left panel of Figure 2.11. The first differenced series, representing the growth rate of U.S. GDP and also shown in Figure 2.11, seems to vary around a fixed mean level, even though the variability appears to be smaller in recent years. To confirm the observed phenomenon, we apply the ADF unit-root test to the log series. Based on the sample PACF of the differenced series shown in Figure 2.11, we choose $p = 10$.

Parameter Estimation I

- **Partial likelihood**
  Consider a time series Y_t, $t = 1, ..., N$, with a joint density $f_{theta}(y_1, ..., y_t)$ parametrized by a vector parameter θ. In addition, suppose there is some auxiliary information AI known throughout the period of ob- servation. Then the likelihood is a function of θ defined by the equation

$$f_\theta(y_1, ..., y_t) = f_\theta(y_1|AI) \prod_{t=2}^{N} f_\theta(y_t|y_1, ..., y_{t-1}, AI)$$

- Due to the Markov assumption the joint density can be factored as

$$f_\theta(y_1, ..., y_t) = f_\theta(y_1|AI) \prod_{t=2}^{N} f_\theta(y_t|y_{t-1}, AI)$$

Parameter Estimation II

- Ignoring the first factor $f_\theta(y_1|AI)$, as it is independent of N , inference regarding θ can be based on the product term. This is an example of **conditional likelihood** resulting from dependent observations expressed as a product of conditional densities.
- The factorization (1.3), without $f_\theta(y_1|AI)$, has some desirable properties worth keeping in mind, such as the fact that the dimension of the factors, as well as that of θ , is fixed regardless of N , and that the derivative with respect to θ of the logarithm of the preceding equation is a zero mean square **integrable martingale**.
- Most ARIMA models are nonlinear models.
- Methods such as methods of moments, maximum likelihood, and least squares that can be employed to estimate the parameters.
- Software does this for us.
- **R function:** `arima(y, order = c(p, d, q))`
- **ARIMA estimation example**:

Simulate an ARIMA model

- **Data generating process (DGP)** is a way to sample finite samples from the theoretical model for given assumptions through computer simulation procedures.
- It is an very important tool in statistics to connect the model and data realization.
- DGP is the opposite routine of modeling estimation.
    - **Model estimation**: the data are known, estimated the parameters and obtain the residuals
    - **DGP**: The parameters are known, simulate error term and the data based on model assumptions.
- We already studies how to simulate data from linear models. AR, MA models.
- **R function:**

```
arima.sim(n = 63, list(ar = c(0.8897, -0.4858), ma =
c(-0.2279, 0.2488)))
```

Diagnostic Checking

- Given an ARIMA model

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

  where y_t has be already differenced (the integrated part) .
- The residual is then

$$\hat{\epsilon}_t = y_t - \hat{\delta} - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} + \sum_{i=1}^{q} \hat{\theta}_i \epsilon_{t-i}$$

- Heteroscedasticity and autocorrelations check should then be the same as the usual way in linear models.
- **Durbin-Waston *h* test** should be used for detecting residual autocorrelations where the relation between the Durbin-Waston h statistic and Durbin-Waston d statistic

$$h = \left(1 - \frac{1}{2}d\right)\sqrt{\frac{T}{1 - T \cdot \widehat{\mathrm{Var}}(\hat{\beta}_1)}},$$

ARIMA model forecasting I

- The best forecast in the mean square sense is the conditional expectation of $Y_{T+\tau}$ given current and previous observations

$$\hat{y}_{T+\tau} = E(y_{T+\tau} | y_T, y_{T-1}, \ldots)$$

- So in an ARIMA (p,d,q) process at time $T+\tau$, we have

$$\hat{y}_{T+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{T+\tau-i} + \epsilon_{T+\tau} - \sum_{i=1}^{q} \theta_i \epsilon_{T+\tau-i}$$

- And the predictive variance is

$$\sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2$$

  where ψ_i **are the coefficients of the moving average representation**.

- It should be noted that the variance of the forecast error gets bigger with increasing forecast lead times τ.

ARIMA model forecasting II

- We can then obtain the $100(1-\alpha)$ percent prediction intervals for the future observations from

$$P(\hat{y}_{T+\tau} - z_{\alpha/2}\sigma(\tau) < y_{T+\tau} < \hat{y}_{T+\tau} + z_{\alpha/2}\sigma(\tau)) = 1-\alpha$$

- **The share example** A share follows the model:

$$y_t = 5 + 0.8y_{t-1} + \epsilon_t$$

  where ϵ_t is white noise with known variance $\sigma^2 = 2$. If today's value is CNY 8. determine the probability that the share will exceed CNY 8 after 3 time steps, that is the probability that $y_{t+3} > y_t$ given that $y_t = 8$.

- We can write y_{t+3} in the form of y_t, that is

$$\begin{aligned} y_{t+3} &= 5 + .8y_{t+2} + a_{t+3} \\ &= 5 + .8\left(5 + .8y_{t+1} + a_{t+2}\right) + a_{t+3} \\ &= 5 + .8\left(5 + .8\left(5 + .8y_t + a_{t+1}\right) + a_{t+2}\right) + a_{t+3} \\ &= 12.2 + .512y_t + \left(.64a_{t+1} + .8a_{t+2} + a_{t+3}\right) \end{aligned}$$

  Given $y_t = 8$, $y_{t+3} = 16.296 + \left(.64a_{t+1} + .8a_{t+2} + a_{t+3}\right)$

ARIMA model forecasting III

- Since $a_t \overset{\text{iid}}{\sim} N(0,2)$ (because of the white noise assumption), y_{t+3} is till normal distributed with mean

$$\mu = E\left(16.296 + (.64a_{t+1} + .8a_{t+2} + a_{t+3})\right) = 16.296$$

- and variance

$$\begin{aligned}\sigma^2 &= Var\left(16.296 + (.64a_{t+1} + .8a_{t+2} + a_{t+3})\right) \\ &= .64^2 Var\left(a_{t+1}\right) + .8^2 Var\left(a_{t+2}\right) + Var\left(a_{t+3}\right) \\ &= .64^2 \cdot 2 + .8^2 \cdot 2 + 2 = 4.0992\end{aligned}$$

- We can have the conditional probability

$$\begin{aligned}P\left(y_{t+3} > y_t | y_t = 8\right) = P\left(y_{t+3} > 8\right) &= P\left(\frac{y_{t+3} - \mu}{\sigma} > \frac{8 - \mu}{\sigma}\right) \\ &= P\left(\frac{y_{t+3} - \mu}{\sigma} > \frac{8 - 16.296}{\sqrt{4.0992}}\right) \\ &= P\left(\frac{y_{t+3} - \mu}{\sigma} > -4.097502\right) \approx 1\end{aligned}$$

ARIMA model forecasting IV

ARIMA(p,d,q)

- Why ARIMA(p,d,q)? Differencing can transform a nonstationary time series to stationary.
- In most applications first differencing ($\mathrm{d} = 1$) and occasionally second differencing ($\mathrm{d} = 2$) would be enough to achieve stationarity.
- Differencing is not the only tool to transform to stationary. Taking the logarithm of the original series will be useful in achieving stationarity.

Seasonal Model I

- 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**.
- Log transformation is commonly used in analysis of financial and economic time series.
- In general, for a seasonal time series y_t with periodicity s, **seasonal differencing** means

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

Seasonal Model II

- **Example 2.3** ACF in seasonal time series

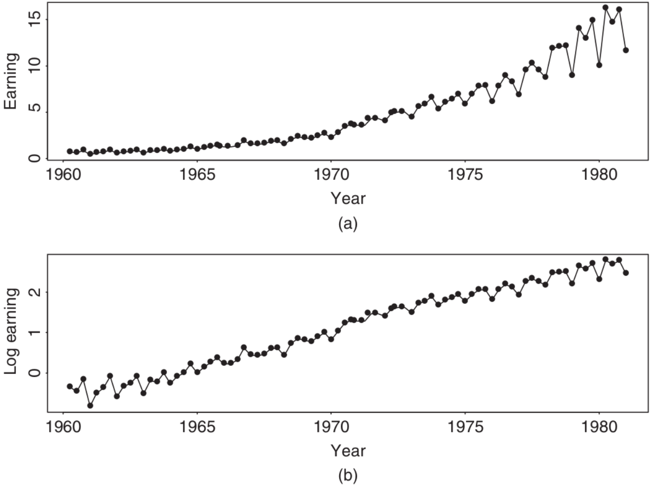

Figure 2.13 Time plots of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) observed earnings and (b) log earnings.

Seasonal Model III

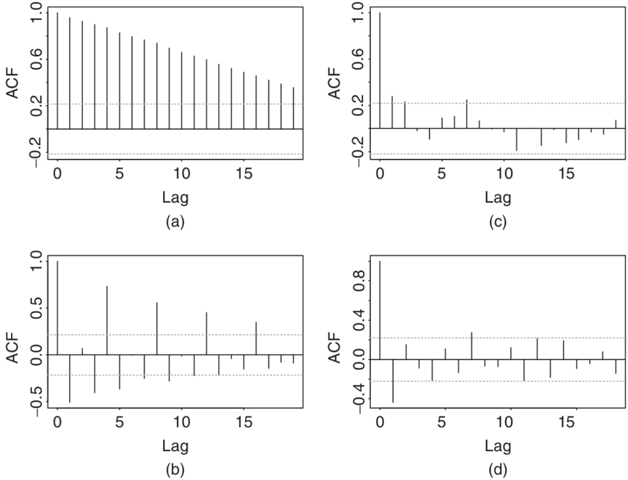

Figure 2.14 Sample ACF of log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980. (a) log earnings, (b) first differenced series, (c) seasonally differenced series, and (d) series with regular and seasonal differencing.

Seasonal Model IV

- **Example 2.4**

Take home questions

- How do you detect if a process y_t is from AR(p) or MA(q) by examining and ACF and PACF?
- What phenomena do AR(p) and MA(q) describe really?

Computer lab

- Data: S&P100 from finance.yahoo.com
- Plot time series for returns (Hint: `ts` package)
- Compute ACF and PACF
- Stationary check using Dickey-Fuller test (Hint: `tseries`)
- Fit an ARIMA(p,d,q) model.
- Explain your model.
- Carry out one-step and two-step ahead forecasting and make comparison.

Suggested Reading

- Tsay (2010) **Chapter 2**