Chapter 4: Model Specification

Time Series Analysis
WISE,XMU

Content

- Model-building strategy.
- Autocorrelation function (ACF).
- Partial autocorrelation function (PACF).
- Nonstationarity.
- Other specification methods.

§4.1 Model-building strategy

- **Box-Jenkins method for model-building strategy:** Finding appropriate models for time series is a nontrivial task. We will develop a multi-step model-building strategy by Box and Jenkins (1976). There are three main steps in the procedure, each of which may be used several times:

1. **Model specification** (or **identification**)
   - In model specification (or identification), the classes of time series models are selected that may be appropriate for a given observed series.
   - It should be emphasized that the model chosen at this point is *tentative* and subject to revision later in the analysis.
   - In choosing a model, we shall attempt to adhere to the **principle of parsimony**; that is, the model used should require the smallest possible number of parameters that will adequately represent the data.

2. **Model fitting**: Model fitting consists of finding the best possible estimates of those unknown parameters within a given model.
3. **Model diagnostics**
   - Model diagnostics is concerned with analyzing the quality of the model that we have specified and estimated.
   - If no inadequacies are found, the modeling may be assumed to be complete, and the model can be used, for example, to forecast future series values.
   - Otherwise, we choose another model in light of the inadequacies found; that is, we return to model specification. In this way we cycle through the three steps until, ideally, and acceptable model is found.

- In this course, we focus on a large class of parametric models for both stationary and nonstationary time series – the ARIMA(p, d, q) model:

$$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d Z_t = \theta_0+(1-\theta_1 B-\cdots-\theta_q B^q)a_t,$$

where $\{a_t\}$ is a white noise sequence with mean zero and variance σ_a^2.
Our task:

1. determine the orders, i.e. the values of p, d and q.
2. estimate parameters $\phi_1, ..., \phi_p, \theta_0, \theta_1, ..., \theta_q$ and σ_a^2.

- **Definition:** A time series $\{Z_t\}$ is said to follow an **integrated autoregressive-moving average** (ARIMA) process if the dth order difference $W_t = \nabla^d Z_t$ is a stationary ARMA process. If W_t is ARMA(p, q), we say that Z_t is ARIMA(p, d, q).

- According to Box-Jenkins method, we can begin our study of applying the ARIMA(p, d, q) model to the real data by following three steps:
  1. How do we choose appropriate values of the orders p, d and q for a given series? (Identification)
     - To decide **reasonable** – but **tentative** – values of p, d and q.
  2. How do we estimate the parameters of the ARIMA(p, d, q) model with specific orders? (Estimation)
     - To estimate the values of ϕ_is, θ_is and σ_a^2 in **the most efficient way**.
  3. How do we check on the appropriateness of the fitted model with specific orders and parameters? (Diagnostic checking)
     - To look **critically** at the fitted model thus to obtain its adequacy.
     - If the model appears inadequate in some ways, we consider the nature of the inadequacy to help us to select another model.

§4.2 Autocorrelation function (ACF)

- The autocorrelation function (ACF) is defined as

$$\rho_k = \frac{\mathrm{cov}(Z_t, Z_{t-k})}{\sqrt{\mathrm{var}(Z_t)\mathrm{var}(Z_{t-k})}} = \frac{\gamma_k}{\gamma_0}, \qquad k = 1, 2, \cdots .$$

1. For a MA(q) process, it has the cut-off property, i.e. $\rho_q \neq 0$ but $\rho_k = 0$ for $k > q$.
2. For a AR or ARMA process, it decays to zero exponentially, but does not enjoy the cut-off property.

- For a sequence of observations $Z_1, ..., Z_n$, the **sample autocorrelation function** (sample ACF) is defined as follows.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(Z_t - \overline{Z})(Z_{t-k} - \overline{Z})}{\sum_{t=1}^{n}(Z_t - \overline{Z})^2}, \qquad k = 0, 1, 2, \cdots,$$

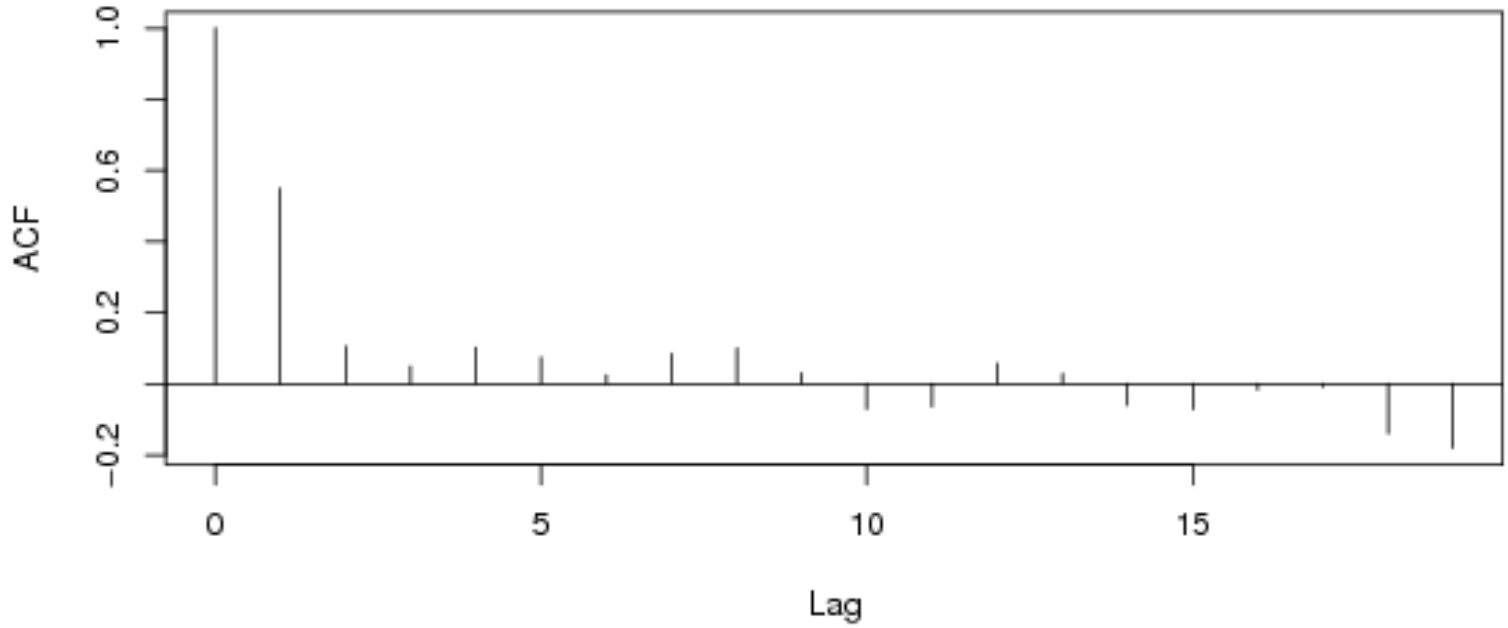
  where $\overline{Z} = n^{-1}\sum_{t=1}^{n} Z_t$.

  Note that we do not assume that $Z_1, ..., Z_n$ are stationary!

- **Theorem:** For a MA(q) process, if $k > q$, then $\rho_k = 0$ and approximately

$$\hat{\rho}_k \sim N\{0, \frac{1}{n}[1 + 2\rho_1^2 + \cdots + 2\rho_q^2]\}.$$

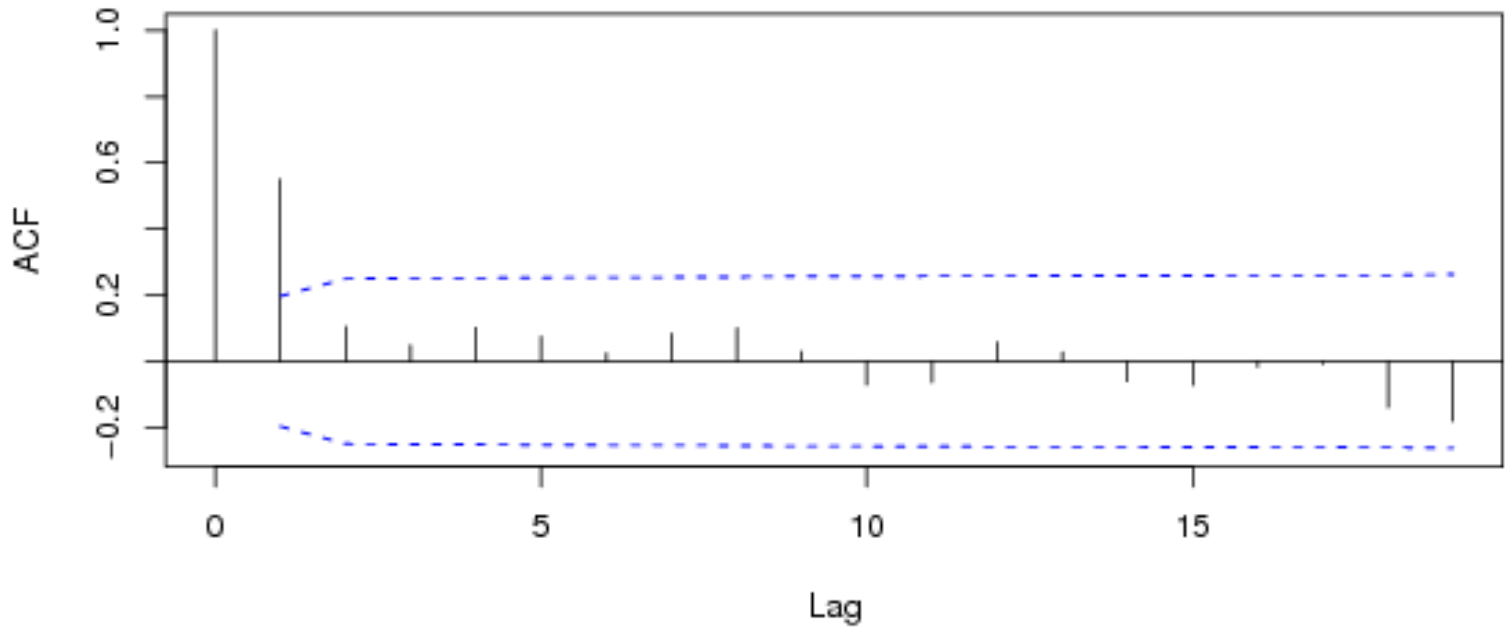
- How to use the sample ACF to identify a MA model?
  1. The following are the sample ACF of a generated MA(1) process, $Z_t = a_t + 0.7a_{t-1}$ with $a_t \sim WN(0, 1)$.



2. **Bartlett's approximation:** The 95% approximate confidence interval of $\hat{\rho}_k$ hence will be

$$\pm \frac{1.96}{\sqrt{n}} \sqrt{1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_{k-1}^2}.$$

   The above CI acts as the accepting region for the test with $H_0 : \rho_k = 0$.
3. Use the Bartlett's approximation to identify the MA model.



4. In real applications, the original condition of the Bartlett's approximation is usually ignored.
   - The upper and lower bond act as the threshold values to measure whether r_ks are significant.
   - The value of 1.96 is the corresponding 95% percentile of the standard normality. The SAS software use the value 1.96, however, the R software use 2.0.

Exercise 1: For a series of length 64, the sample autocorrelations are given as:

Lag	1	2	3	4	5
ACF	0.47	-0.34	0.20	0.02	-0.06

Which models should we consider in this case?
Note: $thre = 1.96/8 * \sqrt{1.00, 1.44, 1.67, 1.75, 1.75} = (0.24, 0.29, 0.32, 0.32, 0.32)$

§4.3 Partial autocorrelation function (PACF)

- **Definition:** The **partial autocorrelation (PACF)** at lag k, denoted by ϕ_{kk}, is defined as the correlation between Z_t and Z_{t-k} *after removing the effect of the intervening variables* $Z_{t-1}, Z_{t-2}, \cdots, Z_{t-k+1}$.
  - The PACF at lag 1 is $\phi_{11} = \text{corr}(Z_t, Z_{t-1}) = \rho_1$.
  - The PACF at lag k is $\phi_{kk} = \text{corr}(Z_t, Z_{t-k} | Z_{t-1}, \cdots, Z_{t-k+1})$

- How to calculate the PACF?
  Suppose $\{Z_t\}$ is a stationary time series with autocorrelation functions $\rho_m, m = 1, 2, 3, ...$, and then it can be shown that ϕ_{kk} satisfies the equations:

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix}$$

  or

$$\begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}.$$

1. Levinson (1947) and Durbin (1960) gave an efficient method to obtain the solutions of $\phi_{k1}, ..., \phi_{kk}$ without calculating the inverse of a series of matrices. They showed that it can be solved recursively as follows:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j},$$

and

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \qquad \text{for } j = 1, 2, \cdots, k-1.$$

2. The above algorithm is also called Durbin-Levinson method. It is commonly used in computer programs.

- Consider the PACF for an AR(2) model, $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$, where $\{a_t\} \sim \text{WN}(0, \sigma_a^2)$.
  1. For PACF at lag 1, $\phi_{11} = \text{corr}(Z_t, Z_{t-1}) = \rho_1$.
  2. For PACF at lag 2,

$$\begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}.$$

Remember that

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \ k = 1 \text{ and } 2.$$

Hence, $\phi_{21} = \phi_1$ and $\phi_{22} = \phi_2$.

3. For PACF at lag 3,

$$\begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix}.$$

Remember that

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \ k = 1, 2, 3.$$

Hence, $\phi_{31} = \phi_1$, $\phi_{32} = \phi_2$ and $\phi_{33} = 0$.

4. For PACF at lag k with $k > 3$,

$$\begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}.$$

Remember that

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2},\ j = 1, 2, ..., k.$$

Hence, $\phi_{k1} = \phi_1$, $\phi_{k2} = \phi_2$ and $\phi_{kj} = 0$ with $j = 3, ..., k$.

- Consider the PACF for general AR models.
  1. For AR(1) processes, $Z_t = \phi Z_{t-1} + a_t$, it is easier to get $\phi_{11} = \phi$, and $\phi_{kk} = 0$ for $k > 1$.
  2. For general AR(p) processes

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t,$$

     it can be shown that $\phi_{pp} = \phi_p$ and $\phi_{kk} = 0$ for $k > p$.
  3. The PACFs for an AR(p) process enjoy the property of cut-off which is the same as that of ACFs for an MA(p) process!

- For an MA(1) series $Z_t = a_t - \theta a_{t-1}$, it can be shown that

$$\phi_{kk} = \frac{-\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \qquad \text{for } k \geq 1.$$

In particular, $\phi_{11} = \rho_1 = -\theta/(1+\theta^2)$,
$\phi_{22} = (\rho_2 - \rho_1^2)/(1-\rho_1^2) = -\theta^2/(1+\theta^2+\theta^4)$.

1. Notice that the partial autocorrelation of an MA(1) model is never zero but essentially decays exponentially to zero, rather like the autocorrelation for an AR(1) series.
2. The behaviors of PACFs for ARMA processes is quite similar to that for MA processes.

- For an observed time series, an obvious method is to estimate the ρ's by r's and then to solve the equations to get estimates of ϕ_{kk}, i.e.,

$$\begin{pmatrix} \widehat{\phi}_{k1} \\ \widehat{\phi}_{k2} \\ \vdots \\ \widehat{\phi}_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_k \end{pmatrix}.$$

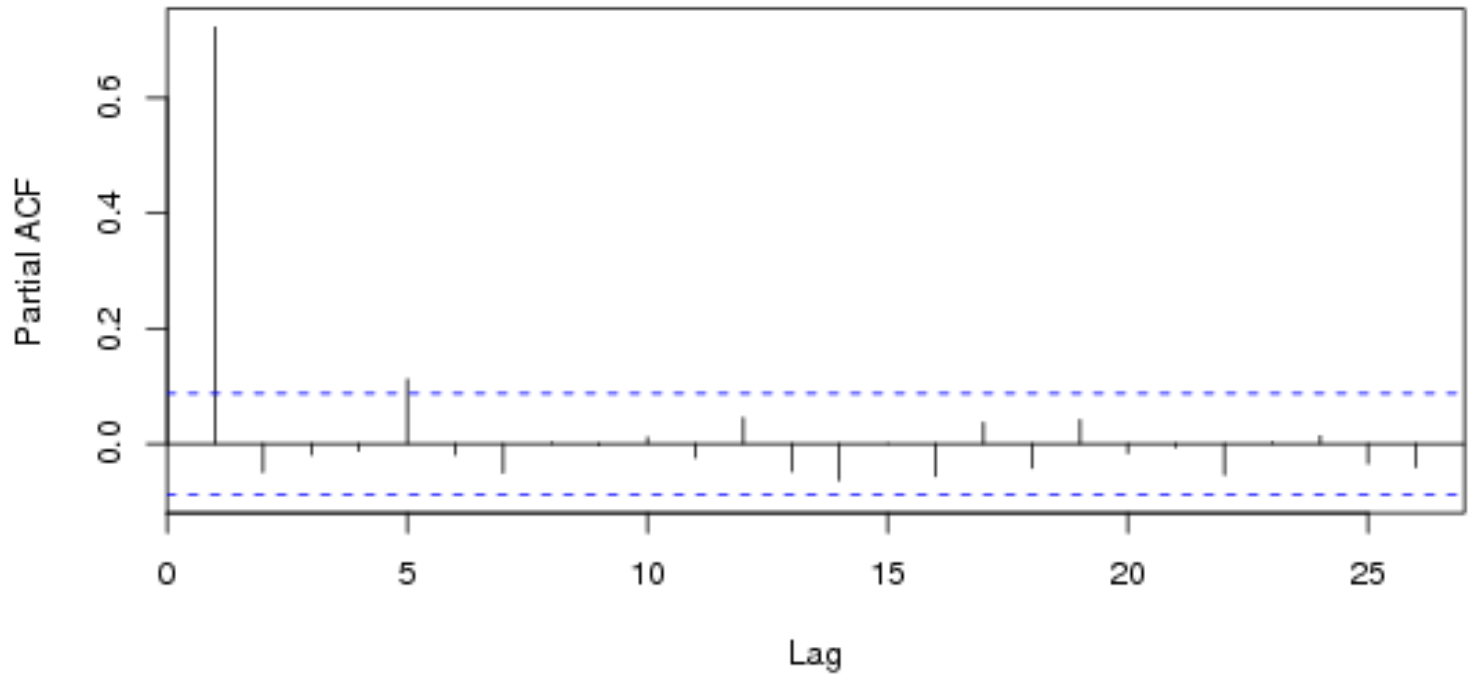
  We call this estimated function the **sample partial autocorrelation function** (sample PACF) and denote it $\widehat{\phi}_{kk}$.

1. The Durbin-Levinson method can also be used to calculate the sample PACF.
2. **Theorem:** For an $AR(p)$ *process*, the estimated values of the partial autocorrelation function at lags greater than p are approximately independently normally distributed with zero means and variance $1/n$, that is,

$$\widehat{\phi}_{kk} \sim N\left(0, \frac{1}{n}\right), \qquad \text{for } k > p.$$

3. Similar to the Bartlett's approximation, in really applications, $\pm 1.96/\sqrt{n}$ can be used as critical limits on $\widehat{\phi}_{kk}$ to measure whether it is significant.
4. The SAS software uses the value of 1.96, however, it is 2.0 in the R software.

5. The following are the sample ACF of a generated AR(1) process, $Z_t = 0.7Z_{t-1} + a_t$ with $a_t \sim WN(0, 1)$.



Example 2: For a series of length 64, the sample partial autocorrelations are given as:

Lag	1	2	3	4	5
PACF	0.47	-0.34	0.20	0.02	-0.06

Which models should we consider in this case?
Note: $1.96/8 = 0.245$

§4.4 Other specification methods

- Some information criteria can also be used to identify the orders of ARMA models by balancing the minimized summation of errors and the complexities of the models.
- **Akaike's information criterion (AIC):** This criterion is to select the model that minimizes

$$\text{AIC}(k) = -2\log(\text{maximum likelihood}) + 2k,$$

where k is the number of parameters. For an ARMA(p, q) model, $k = p + q + 1$ when the model contains an intercept, and $p + q$ otherwise.

Remark: In R, they also account for the variance when calculating k.

- Schwarz **Bayesian information criterion (BIC):** This criterion is to select the model that minimizes

$$\text{BIC}(k) = -2\log(\text{maximum likelihood}) + k\log(n).$$

- There are a lot of information criteria other than AIC and BIC, however, only these two are popular. On the other hand, the AIC is more popular in real applications although the BIC has a well justified theoretical basis.
- How to use these information criteria?
  1. You have 3 models in hands after the complicated modeling procedure. These models are all adequate, and you have no preference for them. You can choose one model with the minimum value of the AIC or BIC.
  2. You want to fit an AR model to a real data, but do not know the order of the model. You can calculate the AIC for an AR(p) model with $p = 0, 1, 2, ..., p_{max}$, say $p_{max} = 10$. You can choose the AR model with the minimum value of the AIC.

Example 3: Consider a special ARMA(1,1) process,
$Z_t = \phi Z_{t-1} + a_t - \phi a_{t-1}$ with $|\phi| < 1$.

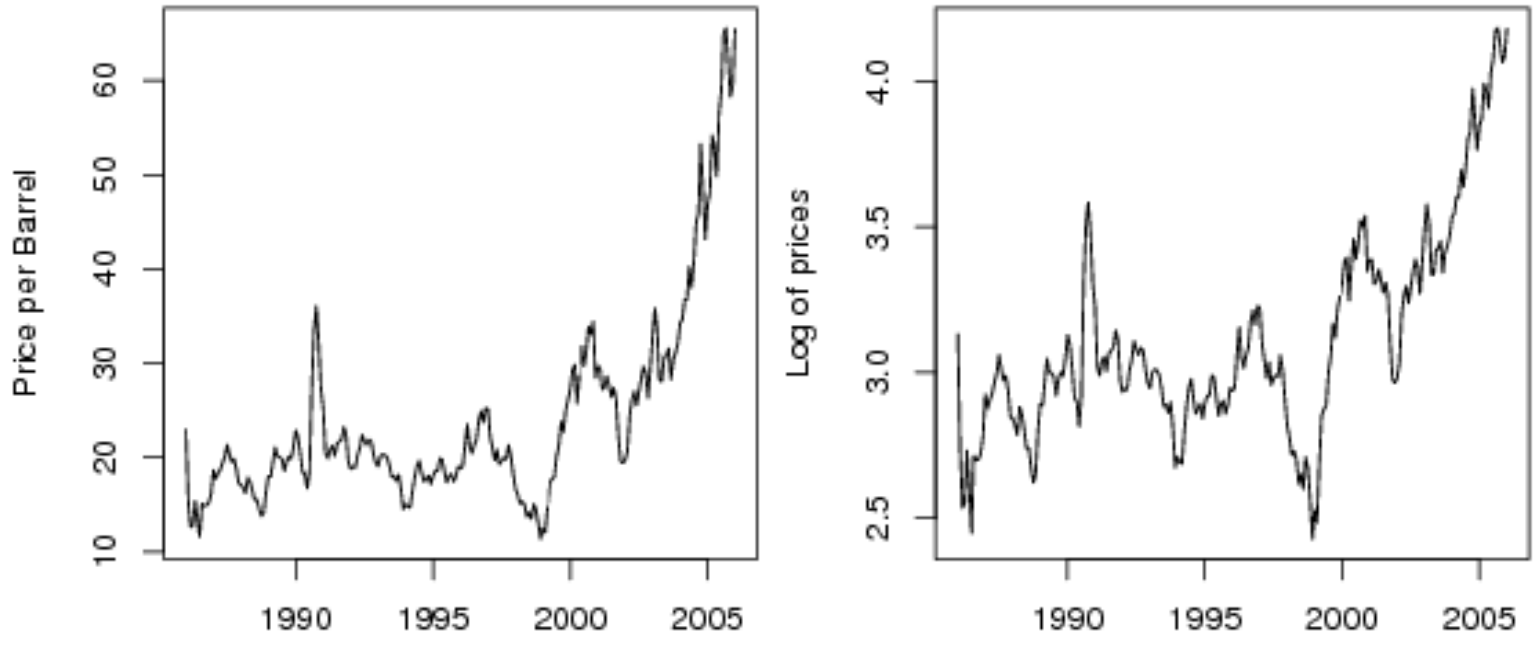
- Calculate the ACFs, PACFs, the MA and AR representations;
- Compare these quantities with those of the white noise $\{a_t\}$.

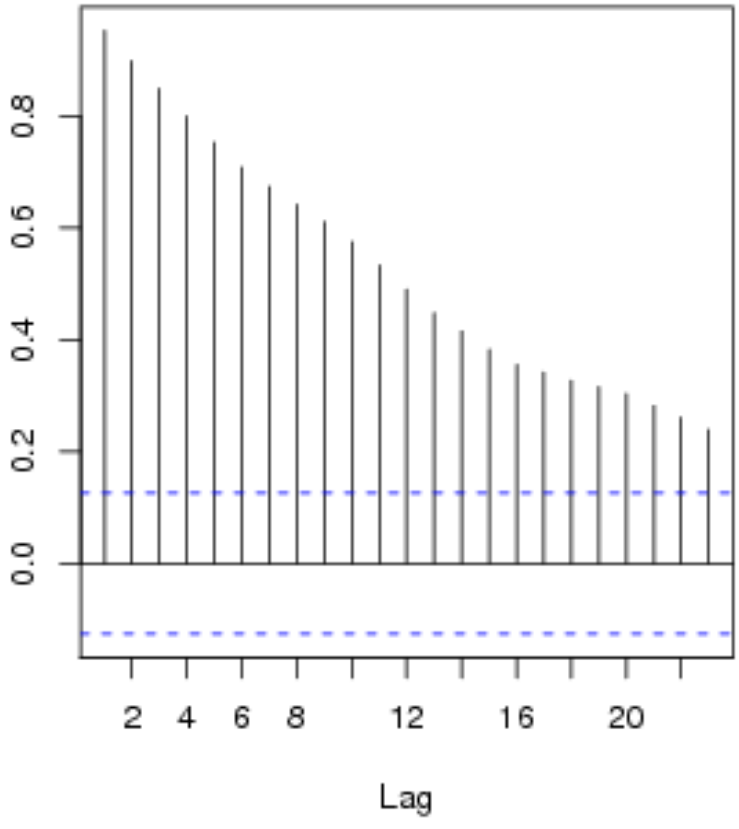
Sol: ACF: $\rho_k = \phi \rho_{k-1}$ for $k > 1$. $\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1-2\phi\theta+\theta^2} = 0$.
PACF: $\phi_{11} = \rho_1 = 0$, $\phi_{22} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2) = 0$,
MA/AR: $(1 - \phi B)Z_t = (1 - \phi B)a_t$. So $Z_t = a_t$.
Suppose $Z_t = \sum_j \psi_j a_{t-j}$. Then
$\psi_0 a_t + \psi_1 a_{t-1} + \cdots = \phi(\psi_0 a_{t-1} + \psi_1 a_{t-2} + \cdots) + a_t - \phi a_{t-1}$.
$\psi_0 = 1$, $\psi_1 = \phi\psi_0 - \phi = 0$...
Suppose $a_t = \sum_j \pi_j Z_{t-j}$. Then
$Z_t = \phi Z_{t-1} + \pi_0 Z_t + \pi_1 Z_{t-1} + \cdots - \phi(\pi_0 Z_{t-1} + \pi_1 Z_{t-2} + \cdots)$.
Hence $1 = \pi_0$, $0 = \phi + \pi_1 - \phi\pi_0$, so $\pi_1 = 0$...

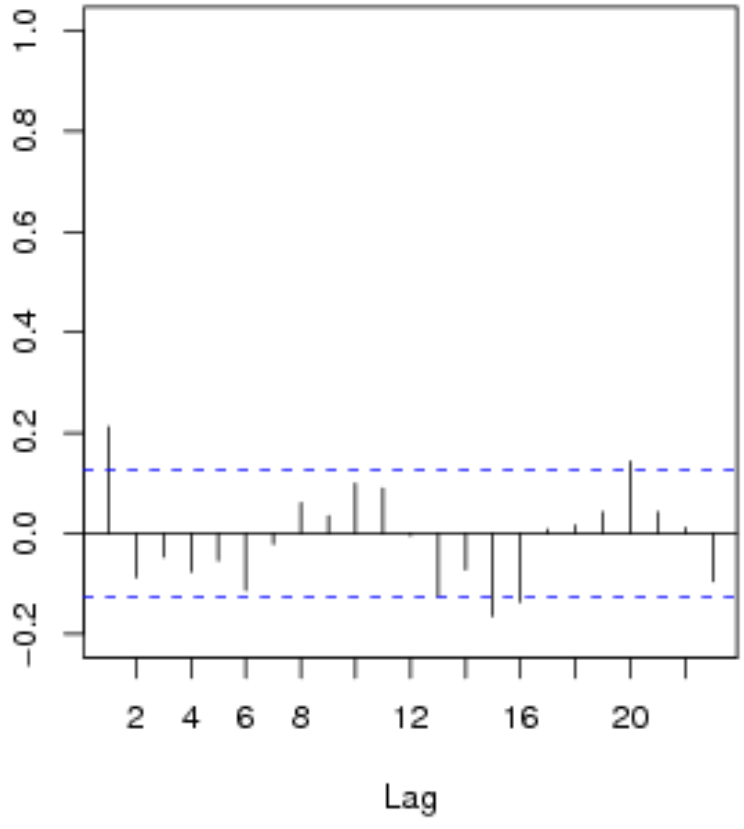
§4.5 Nonstationarity

- How to determine the value of d in ARIMA(p, d, q) models?
  1. To determine the order of d is equivalent to determine whether or not a sequence of values are from a stationary time series.
  2. The definition of the sample ACF implicitly **assumes** stationarity. Thus it is not at all clear what the sample ACF is estimating for a nonstationary process.
  3. However, for a nonstationary series, the sample ACF typically decays slowly as the lags increase. This is due to the tendency for nonstationary series to drift slowly, either up or down, with apparent "trends".
  4. If the sample ACF decays **slowly**, we may say the sequence is nonstationary. Otherwise, it is stationary.

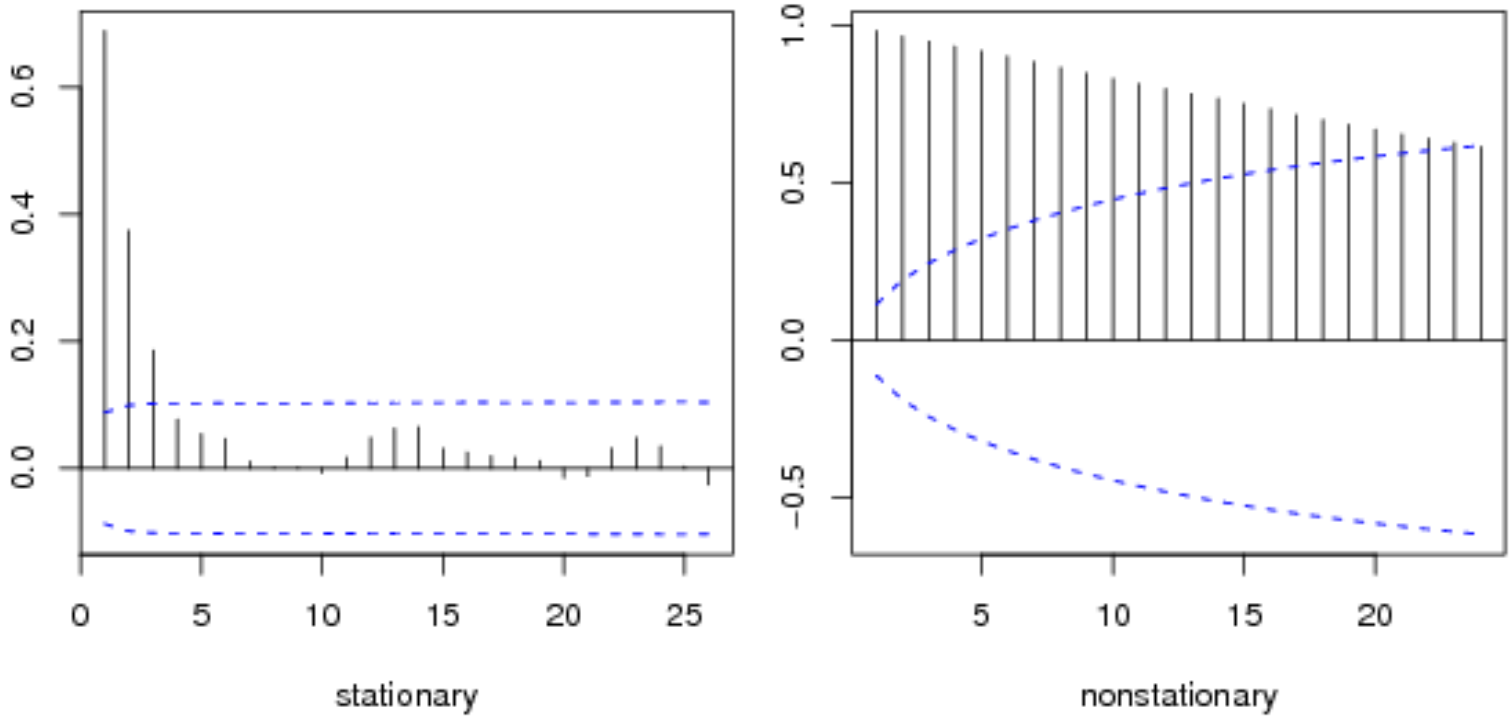
- **Monthly price of a barrel of crude oil from January 1986 to January 2006:** The following are the time plots of original sequence and the log sequence, and the sample ACFs of the log sequence and the differenced log sequence.



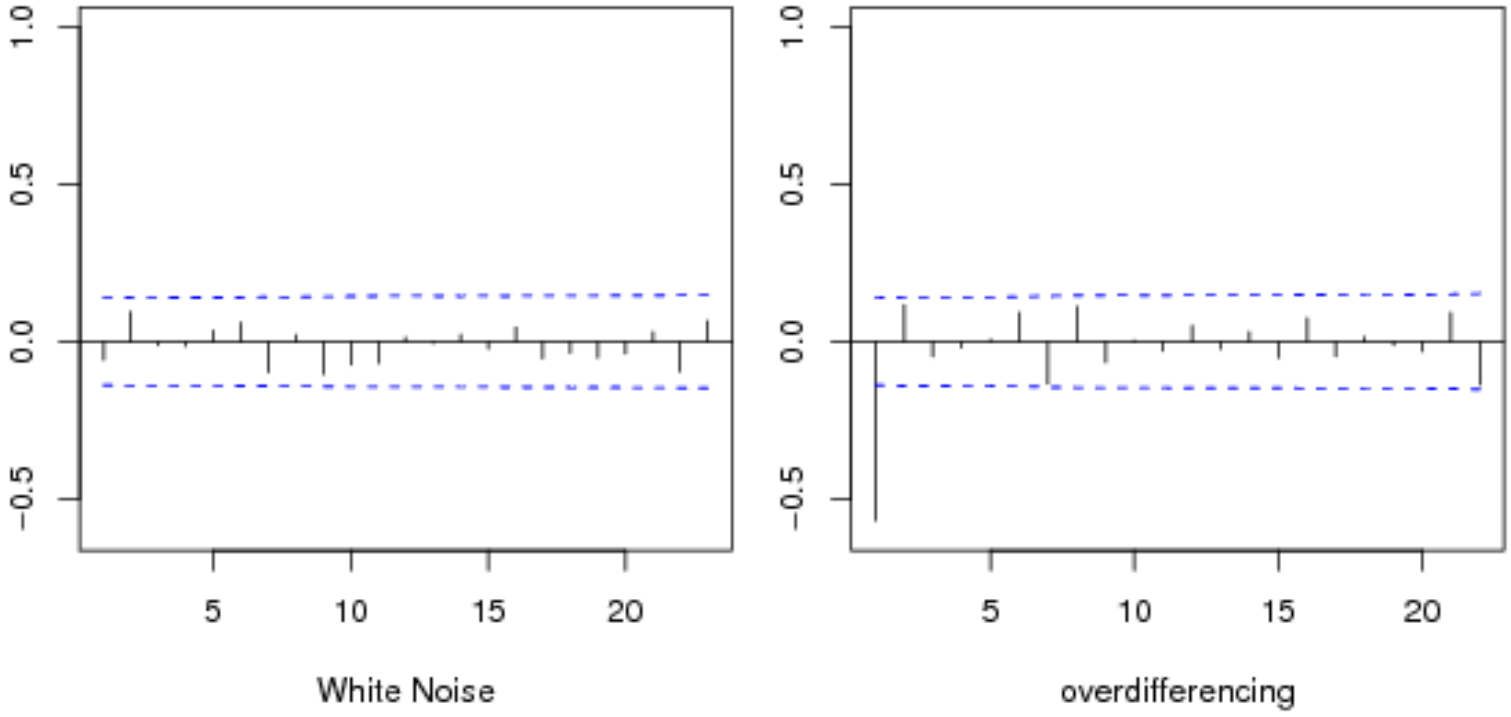

0.0
0.2
0.4
0.6
0.8
2
4
6
8
12
16
20
Lag


−0.2
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
12
16
20
Lag

- The sample ACFs of an ARMA(1,1) process, $Z_t = 0.6Z_{t-1} + a_t + 0.3a_{t-1}$, and the random walk, $Z_t = Z_{t-1} + a_t$.



- **Overdifferencing:**
  1. If a time series $\{Z_t\}$ is stationary, then the differenced sequence $\{\nabla Z_t\}$ is still stationary. Thus, we sometimes tend to overdifference the sequence.
  2. Suppose the original sequence is a white noise, say $Z_t = a_t$. We consider the differenced sequence $\{W_t\}$, where $W_t = a_t - a_{t-1}$. It is a MA(1) process, and not invertible! Comparing $\{Z_t\}$, we need one more parameter to explain $\{W_t\}$. Furthermore, we may meet problems in estimating the parameters and in interpreting the fitted model(s) since it is not invertible.



3. To avoid overdifferencing, we should look carefully at each difference in succession and keep the principle of parsimony always in mind.

- **Augmented Dickey-Fuller unit root tests:** Consider the model

$$Z_t = \alpha Z_{t-1} + X_t,$$

where $\{X_t\}$ is a stationary time series.

1. If $\alpha = 1$, then $\{Z_t\}$ is not stationary, but the differenced sequence $\{\nabla Z_t\}$ is stationary.
   If $-1 < \alpha < 1$, then $\{Z_t\}$ is stationary.
   We want to design a test to check whether or not $\alpha = 1$.
2. Suppose that $\{Z_t\}$ is not stationary, i.e. $\alpha = 1$, and $\{X_t\}$ is an AR(k) process, $X_t = \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + a_t$. Thus, $X_t = Z_t - Z_{t-1} = \nabla Z_t$ and

$$\nabla Z_t = a Z_{t-1} + \phi_1 \nabla Z_{t-1} + \cdots + \phi_k \nabla Z_{t-k} + a_t, \qquad (3.1)$$

   where $a = \alpha - 1 = 0$.
   The augmented Dickey-Fuller unit root tests are

$$H_0 : a = 0 \quad \text{(nonstationary) vs} \quad H_1 : a < 0.$$

3. To handle the possible deterministic trend, instead of equation (3.1), the model for the augmented Dickey-Fuller unit root tests is

$$\nabla Z_t = \mu_t + aZ_{t-1} + \phi_1 \nabla Z_{t-1} + \cdots + \phi_k \nabla Z_{t-k} + a_t.$$

   Three cases are considered: $\mu_t = 0$, $\mu_t = \mu$, $\mu_t = \mu + \beta t$.
4. To choose the value of k is equivalent to fit an AR(k) to the differenced sequence.
   - We may use the sample PACF to select k. For example, we can set $k = 5$ if the sample PACF cuts off at lag 6.
   - We may use some information criteria, which will be introduced later, to select k.
   - We may consider several values of k at the same time, say $k = 1, ..., 8$.

Summary

We want to identify one or more than one possible models by following:

1. Taking transformation or not; (mainly log transformation)
2. Taking possible difference to make stationary the candidate time series; (identify the value of d by plots and ADF test)
3. For stationary sequence, guessing a order p for an AR model by sample PACF, or q for an MA model by sample ACF; If it is an ARMA model, then select p and q by information criteria.

Reference

Please read Chapter 6 of Cryer & Chan (2008).