

STAT4005 - Time Series
Tutorial 6 - Model Selection and Diagnostics
Ma Ting Tin, Martin
Week 10

6 Model Selection and Diagnostics

In the previous tutorial, we discuss how to conduct inference on the parameters of the ARMA(p, q) model given the order p and q. In this tutorial, we will discuss how to select a suitable value of p and q.

6.1 Graphical Method (For stationary AR or MA model)

Noticing that for MA(q) model, i.e., $Y_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$, we have

$$\gamma(k) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|k|} \theta_j \theta_{j+|k|} & , \text{ if } |k| \leq q \\ 0 & , \text{ if } |k| > q \end{cases}$$

where $\theta_0 := 1$ for convention. It follows that ACVF $\gamma(\cdot)$ and hence the ACF plot will show a sharp cut-off at lag q. For example, see the following figure

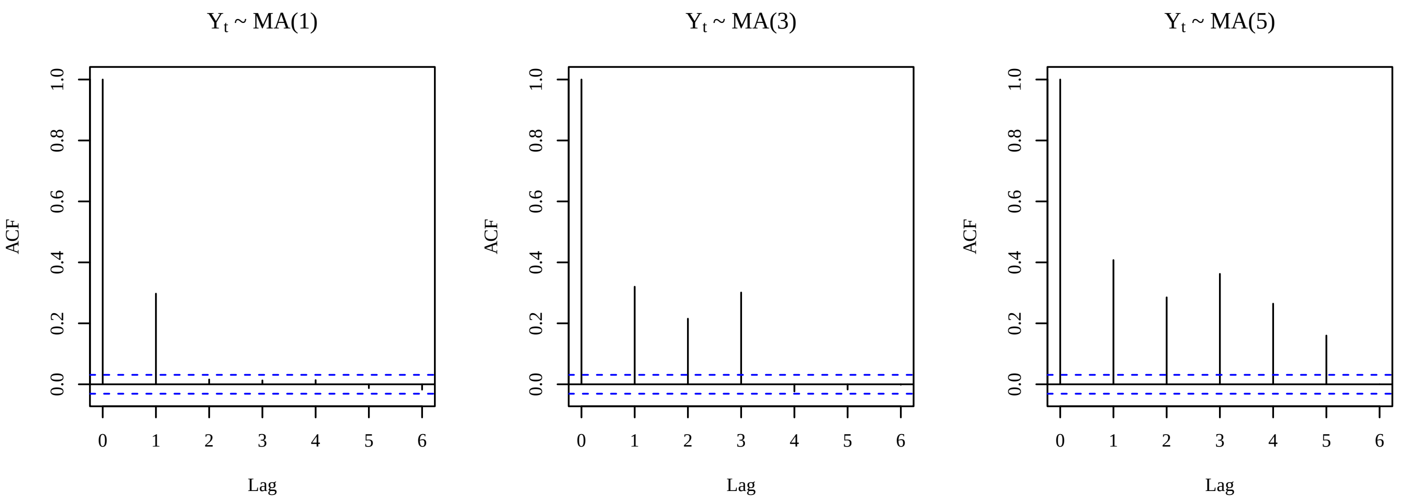


Figure 1: Examples of ACF plot of MA(q) models

Recall from Tutorial 02 that $\sqrt{n} r_k \xrightarrow{\text{d}} \text{N}(0, 1)$ under the assumption that $\gamma(k) = 0$, Hence for large n, the confidence interval for r_k is given by

$$\hat{I} = \left[-\frac{z_{1-\alpha/2}}{\sqrt{n}}, \frac{z_{1-\alpha/2}}{\sqrt{n}} \right],$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)\%$ quantile of standard normal distribution. The blue line in figure 1 refers to the confidence interval $\hat{I}$ with $\alpha = 0.05$. Therefore, we consider the value of ACF to be significantly different from 0 if it lies out of $\hat{I}$.

For the general MA(q) model, it does not show a special pattern in the PACF plot. As a special case, for $\{Y_t\} \sim$ MA(1), we have

$$\phi_{kk} = -\frac{(-\theta)^k (1 - \theta^2)}{1 - \theta^{2(k+1)}},$$

which shows roughly exponential decay (in terms of magnitude). See the following figure

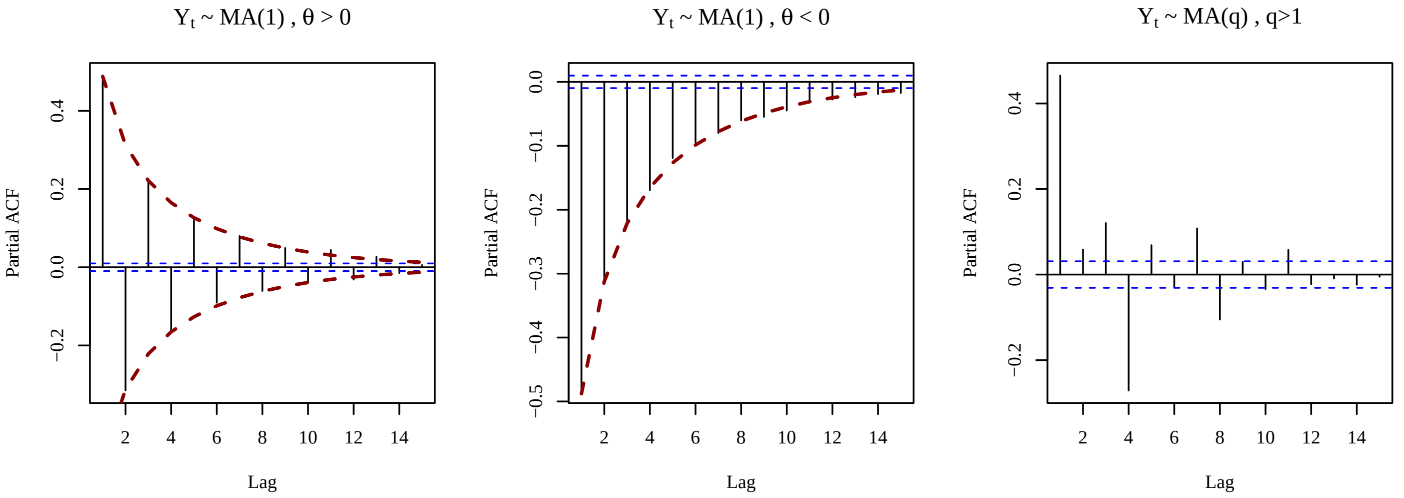


Figure 2: PACF plot of MA(1) and general MA(q) model

For the AR(p) model: $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$, we have $\phi_{k,k} = 0 \quad$ whenever $k > p$, i.e., the PACF shows a sharp cut-off at lag p. For example, see

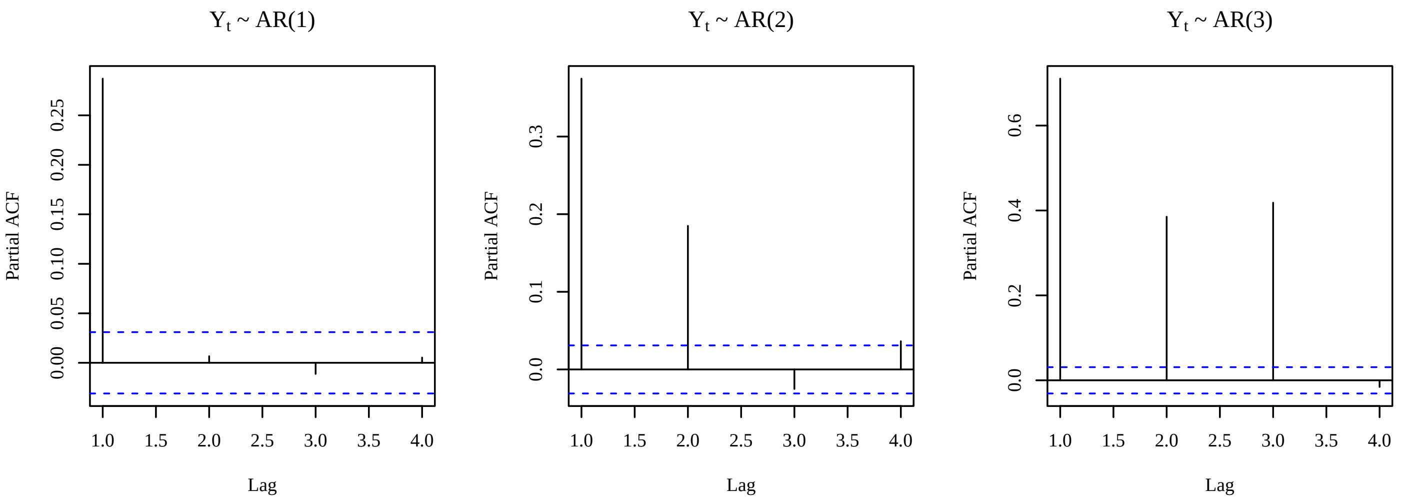


Figure 3: Examples of PACF plot of AR(p) models

For the general AR(p) model, it does not show a special pattern in the ACF plot. As a special case, for $\{Y_t\} \sim$ AR(1), we have

$$\rho(k) = \phi^{|k|}$$

and hence the ACF plot for AR(1) model shows an exponential decay in magnitude. See figure 4.

Remark 1. MA(q) *must be stationary. However, the* AR(p) *model is not stationary in general. The graphical method can only help to specify a stationary AR or MA model.*

The following is a short summary of the graphical model

Plot of	MA(q)	AR(p)	MA(1)	AR(1)
ACF $\rho(k)$	**Cut-off at lag q**	NA	**Cut-off at lag 1**	**Exponential decay**
PACF ϕ_{kk}	NA	**Cut-off at lag p**	**Exponential Decay**	**Cut-off at lag 1**

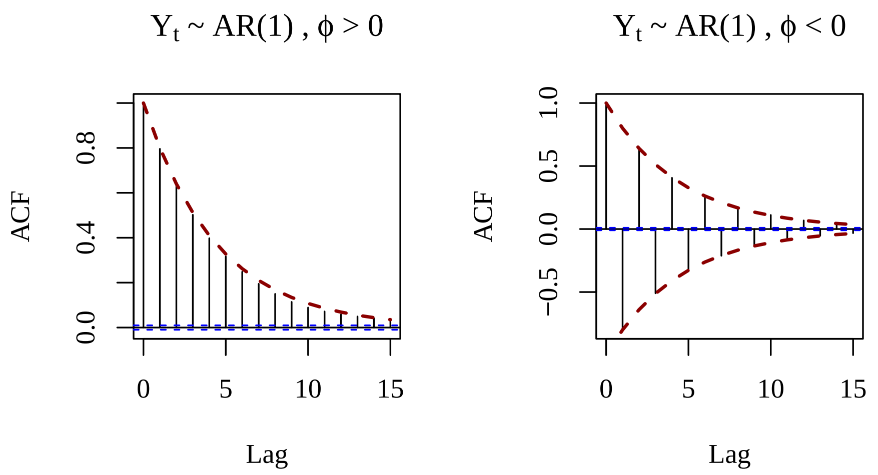
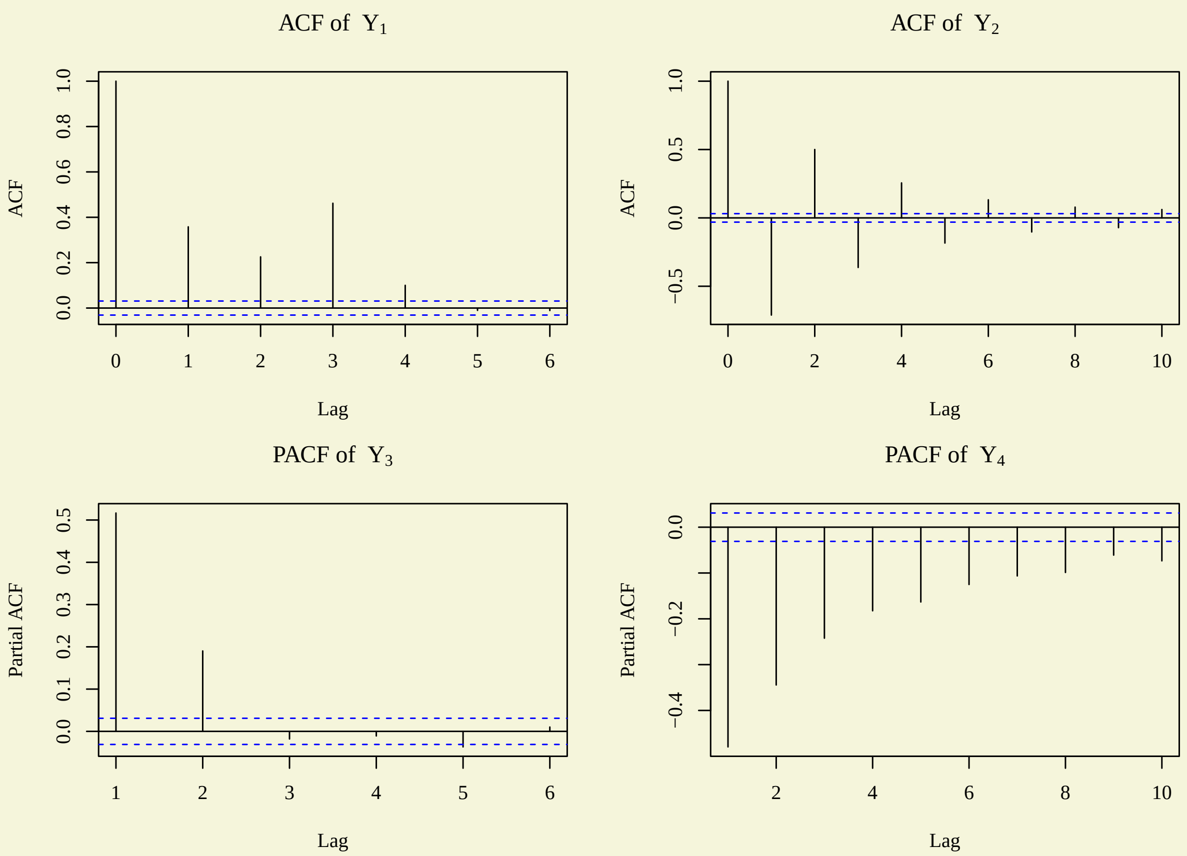


Figure 4: ACF plot of AR(1) models

(★☆☆) Identification of Order (Graphical Method)

✎ **Exercise 1.** *Identify the following AR or MA model and specify the order of the model.*



Attempt

Email contact: martinmtt@link.cuhk.edu.hk ; Personal website: https://sites.google.com/view/martinmtt/home

6.2 Information Criteria (General Approach)

Notice that the graphical method is hard to determine the order of the general ARMA(p, q) model and is a little bit ambiguous. Instead, we can decide on some criteria function for proceeding with rigorous model selection. In this subsection, we consider ARMA(p, q) model with

- $\hat{\beta} = (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$ and $\hat{\sigma}^2$ as the MLE of the model given $Y_1, \ldots, Y_n$.
- $S_Y(\hat{\beta}) = \sum_{t=1}^n \hat{Z}_t^2$, where $\hat{Z}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \cdots - \hat{\phi}_p Y_{t-p} - \hat{\theta}_1 Z_{t-1} - \cdots - \hat{\theta}_q Z_{t-q}$.
- $L(\hat{\beta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} \exp\{-S_Y(\hat{\beta})/(2\hat{\sigma}^2)\}$, as the likelihood function of ARMA(p, q) model.

Information Criteria

Definition 1. *The following are definitions of some common criteria function*

1. ***AIC (Akaike's Information Criterion)***

$$-2 \log L\left(\hat{\beta}, \frac{S_Y(\hat{\beta})}{n}\right) + 2(p+q+1),$$

2. ***AICC (AIC corrected)***

$$-2 \log L\left(\hat{\beta}, \frac{S_Y(\hat{\beta})}{n}\right) + \frac{2(p+q+1)n}{n-p-q-2},$$

3. ***BIC (Bayesian Information Criterion)***

$$(n-p-q)\log\left(\frac{n\hat{\sigma}^2}{n-p-1}\right) + n(1+\log\sqrt{2\pi}) + (p+q)\log\left(\frac{\sum_{i=1}^n Y_i^2 - n\hat{\sigma}^2}{p+q}\right).$$

4. ***FPE (Final Prediction Error)*** *[For AR models only]*

$$\left(\frac{n+p}{n-p}\right)\hat{\sigma}^2.$$

Remark 2. *Given a criteria function f, we should always choose the model which gives a lower value, i.e., given models $\mathscr{F}_1$ and $\mathscr{F}_2$, if $f(\mathscr{F}_1) < f(\mathscr{F}_2)$, we prefer $\mathscr{F}_1$ more.*

Remark 3. *Different criteria functions are developed to achieve different goals.*

1. ***AIC*** *estimates the Expected Predictive Log-likelihood* $\mathrm{E}\{\log f(y_{n+1}|\hat{\beta})|y_1, \ldots, y_n\}$, *i.e. AIC is useful if one wishes to find a model which gives an accurate prediction.*
2. *For a small sample size n, it is likely that AIC will select a model with too many parameters, i.e., overfitted model.* ***AICC*** *is a corrected version of AIC so that it works well for small n.*
3. ***BIC*** *estimates the Marginal Log-likelihood* $\log f(y_1, \ldots, y_n|\hat{\beta})$, *i.e. the intrinsic of the model. As $n \to \infty$, BIC always chooses the correct model. (Unsatisfactory performance for small n)*
4. ***FPE*** *estimates* $\mathrm{E}\{(\hat{\phi} - \phi)^T(\hat{\phi} - \phi)\} = \sum_{k=1}^p \mathrm{MSE}(\hat{\phi}_k)$, *i.e., sum of MSE of the parameter estimate. Notice that it is a valid measure of goodness of fit only for the AR model.*

There are some useful `R`-commands in model fitting

1. `arima.sim(n, model=list( ar=c(phi1,..., phiq), ma=c(theta1,..., thetap) ))` is used for generation of $(Y_1, \dots, Y_n) \sim \text{ARMA}(p, q)$.
2. `arima(Y,order=c(p,d,q))` fit the vector Y to the $\text{ARIMA}(p, d, q)$ model.

(★☆☆) Order Selection by Information Criteria

✎ **Exercise 2.** *Read the attached `R`-code and answer the following:*

(a) Write down the order of the true model and the models fitted. (refer to lines 1 and 2)

(b) Calculate the AIC and BIC of the models. Which model should you choose based on different criteria?

```
1  Y = arima.sim(405,model=list(ar=c(0.3,0.4,0.2),ma=c(0.2,0.1,0.6)))
2  Model_1 = arima(Y,order=c(2,0,3)) ; Model_2 = arima(Y,order=c(3,0,3))
3  sum(Y^2)
4  [1] 3684.08
5
6  Model_1
7  Call: arima(x = Y, order = c(2, 0, 3))
8  Coefficients:
9           ar1     ar2     ma1     ma2     ma3  intercept
10       0.4036  0.4609  0.1826  0.0271  0.6176    -0.4641
11 s.e.  0.0636  0.0631  0.0513  0.0477  0.0404     0.6786
12 sigma^2 estimated as 1.081:  log likelihood = -592.26,  aic = ?
13
14 Model_2
15 Call: arima(x = Y, order = c(3, 0, 3))
16
17 Coefficients:
18          ar1     ar2     ar3     ma1     ma2     ma3  intercept
19       0.3237  0.4004  0.1455  0.2359  0.0986  0.5868    -0.4609
20 s.e.  0.0797  0.0731  0.0771  0.0652  0.0578  0.0475     0.7346
21
22 sigma^2 estimated as 1.072:  log likelihood = -590.51,  aic = ?
```

Attempt

6.3 Model Diagnostics

Recall if $\{Y_t\} \sim \text{ARIMA}(p,d,q)$, we have $\phi(B)(1-B)^d Y_t = \theta(B)Z_t$, where $\{Z_t\} \sim \text{WN}(0,\sigma^2)$. However, if the fitted model is inappropriate, the fitted residual $\{\hat{Z}_t\}$ may NOT be a white noise sequence. Therefore, it is natural to test whether the fitted model is appropriate by testing whether the estimated residuals are a white noise sequence. Recall in tutorial 02,

$$\text{Under } H_0 : \rho(k) = 0, \quad \sqrt{n}r_k \xrightarrow{\text{d}} \text{N}(0,1) \quad \text{as} \quad n \to \infty.$$

Those quantities can help to test the existence of significant dependence. However, those tests only focused on a single lag size k. It motivates us to study more general hypotheses and associated tests.

Ljung-Box Test (Portmanteau Statistics)

Theorem 1. *Let $r_Z(j)$ be the sample ACF of $\{\hat{Z}_t\}$ and ρ_Z be the ACF of the true noise sequence.*

$$H_0 : \rho_Z(k) = 0 \textit{ whenever } |k| \le h \quad \textit{against} \quad H_1 : \rho_Z(k) \neq 0 \textit{ for some } |k| \le h$$

for some prespecified h, The ***Ljung-Box Test*** *is defined by*

$$Q(h) = n(n+2)\sum_{j=1}^{h} \frac{\hat{r}_Z^2(j)}{n-j}$$

and $Q(h) \xrightarrow{\text{d}} \chi^2(h-p-q)$ under H_0 as $n \to \infty$.

Remark 4. *There are several remarks about the Ljung-Box test*

1. *A common choice of h lies between 10 and 30.*
2. *If $Q(h) \ge \chi^2_{h-p-q,0.95}$, H_0 is rejected.*
3. *If H_0 is not rejected, then the model is not a bad fit to the data.*

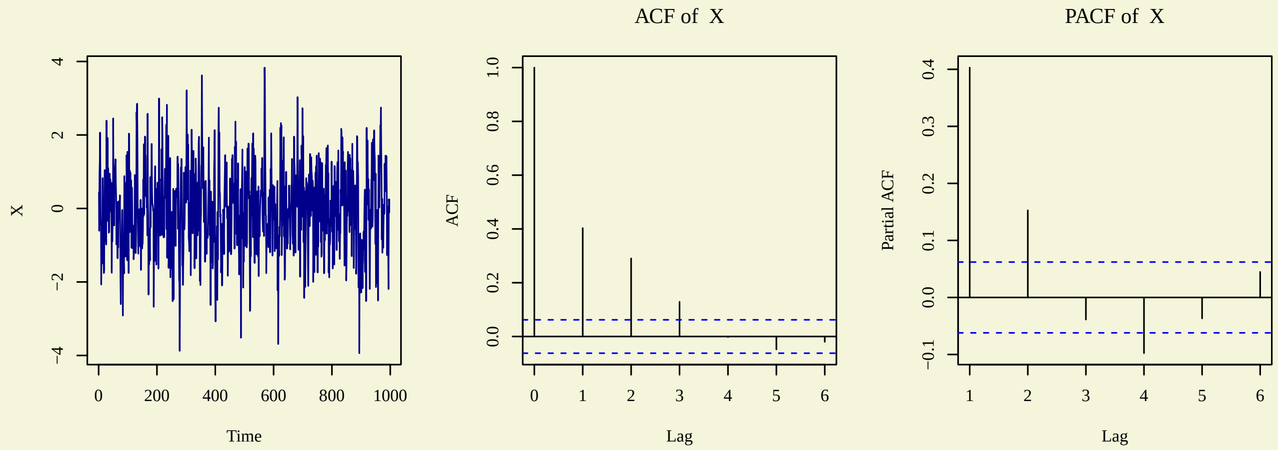
(★★☆) Model Diagnostics

✎ **Exercise 3.** *After we have fitted $Y_1, \ldots, Y_{100}$ to the $\text{ARMA}(1,2)$ model, we have the following information regarding the ACF of the estimated residual: $\hat{r}_Z(1) = -0.1$, $\hat{r}_Z(2) = 0.2$ and $|\hat{r}_Z(k)| < 0.05$ whenever $k \ge 3$. Perform the Ljung-box test with $h = 10$ to test whether the fitted model is a good fit.*

Solution

(★★☆) Comprehensive Exercise

✎ **Exercise 4.** *Given the time series $Y_1, \ldots, Y_n$ ($n = 999$) and $X_t = Y_t - 0.6Y_{t-1} - 0.3Y_{t-2}$, the time-series plot, ACF and PACF plot of $\{X_t\}$ is given in the following figure*



(a) *By observing the above graphs, suggest the most suitable AR or MA model for $\{X_t\}$. Hence, suggest a suitable* $\mathrm{ARMA}(p, q)$ *model for modeling* $\{Y_t\}$.

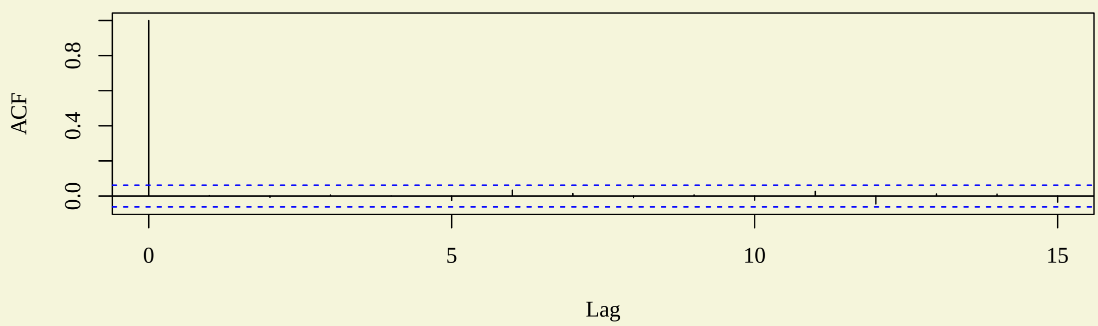
(b) *Kevin suggests that* $\mathrm{ARMA}(2, 2)$ *is a more suitable model for prediction based on AIC. Do you agree with him according to the following* `R`*-output? (p, q is the order of the ARMA model in (a))*

```
Call: arima(x = Y, order = c(p, 0, q)) # Model 1
Coefficients:
          ar1       ar2      ma1     ma2     ma3  intercept
       1.0007   -0.0943  -0.0621  0.3203  0.0726    -0.2451
s.e.   0.2867    0.2629   0.2857  0.0316  0.0953     0.4565

sigma^2 estimated as 1.05:  log likelihood = -1442.93
Call: arima(x = Y, order = c(2, 0, 2)) # Model 2
Coefficients:
          ar1       ar2      ma1     ma2  intercept
       1.1996   -0.2752  -0.2627  0.3206    -0.2421
s.e.   0.0887    0.0846   0.0839  0.0330     0.4501
sigma^2 estimated as 1.05:  log likelihood = -1443.17
```

(c) *Jensen claimed that as $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$ and those two models have the same order in the AR-polynomial, hence they are indifferent if we choose the model based on FPE. Do you agree with him?*

(d) *Check the goodness of fit of the chosen model in part (b) by finding an upper bound for the Portmanteau test with $h = 15$ and refer to the following ACF plot of the estimated residual. Notice that the blue line is the 95% CI for ACF. (Remark: $\chi^2_{15,0.975} = 27.49$ and $\chi^2_{11,0.975} = 21.92$)*



Email contact: martinmtt@link.cuhk.edu.hk ; Personal website: https://sites.google.com/view/martinmtt/home

Attempt

Email contact: martinmtt@link.cuhk.edu.hk ; Personal website: https://sites.google.com/view/martinmtt/home

6.4 R-programming

This section provides some basic techniques and common R-command used in time-series analysis.

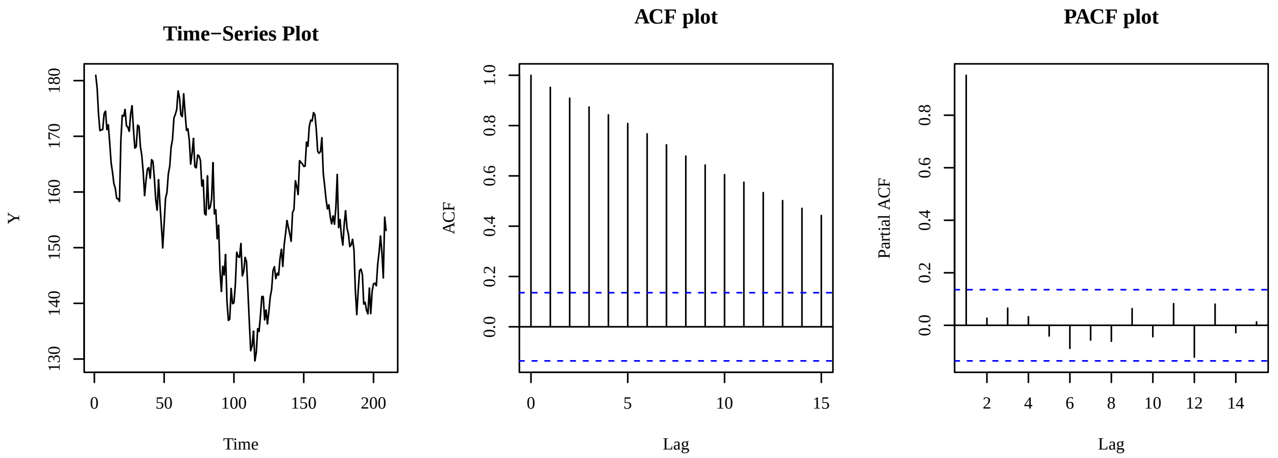
1. Load the time-series data. Let us take the stock price of Apple as an example.

```
install.packages("tseries")
library("tseries")
X0 = get.hist.quote(instrument="AAPL", start="2022-01-01", end="2022-11-01",
   quote="Adjusted", provider=c("yahoo"), compression="d", retclass=c("ts"))
X = X0[-which(is.na(X0)==1)] # Exclude data of non-trading days
```

   Remark 5. *Indeed, we should do the modeling for the log-return instead of the stock price. But let us do it in a simple way here for illustration purpose.*

2. We first visualize the data through the time-series plot, ACF plot, and PACF plot.

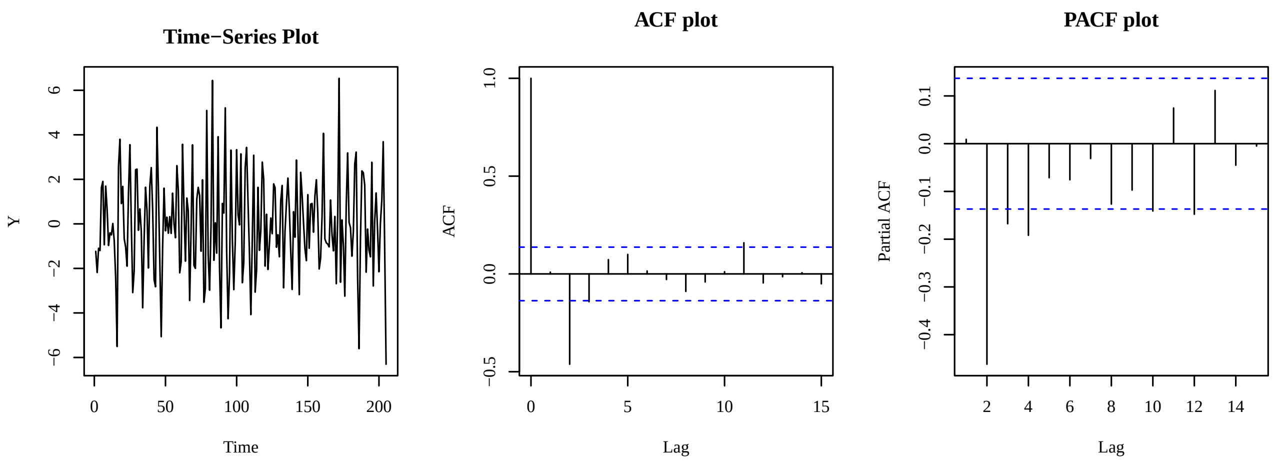
```
par(mfrow=c(1,3))
ts.plot(X,main="Time-Series Plot")
acf(X,lag.max=15,main="ACF plot")
pacf(X,lag.max=15,main="PACF plot")
```



   From the plot, it is obvious that the data is highly non-stationary.

3. Apply the technique in Tutorial 01 to get a sequence of estimated noise. Suppose that there is no seasonal effect. Consider filter $(0.2, 0.2, 0.2, 0.2, 0.2)$, define $\hat{T}_t = \sum_{t=-2}^{2} a_r X_{t+r}$ and let $Y_t = X_t - \hat{T}_t$.

```
n = length(X) ; That = rep(NA,n-4)
for(i in 1:(n-4)){That[i] = mean(X[i:(i+4)])}
Y = X[3:(n-2)] - That ; n = length(Y)
ts.plot(Y,main="Time-Series Plot")
acf(Y,lag.max=15,main="ACF plot") ; pacf(Y,lag.max=15,main="PACF plot")
```



From the plot, the transformed series $\{Y_t\}$ is more likely to be stationary than $\{X_t\}$.

4. We then conduct the model fitting (By Method of Moment or Leas-Square method or MLE).

Method of Moment Estimator for MA model (With R)

✎ **Exercise 5.** *Suppose* $\{Y_t\} \sim \mathrm{MA}(2)$*, i.e.,* $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$ *with* $\{Z_t\} \sim \mathrm{WN}(0,1)$*. Derive a possible MM estimator for* θ_1 *and* θ_2*. Evaluate them through* **R***.*

Solution

$\{Y_t\} \sim \mathrm{MA}(2)$ is causal and therefore stationary. We have

$$\gamma(k) = (1+\theta_1^2+\theta_2^2)\mathbb{1}(k=0) + \theta_1(1+\theta_2)\mathbb{1}(|k=1|) + \theta_2\mathbb{1}(|k|=2)$$

- $\gamma(2) = \theta_2 \quad \Rightarrow \quad \hat{\theta}_2 = C_2.$
- $\gamma(1) = \theta_1(1+\theta_2) \quad \Rightarrow \quad \hat{\theta}_1 = C_1/(1+C_2)$

By the R-code below, we have $\hat{\theta}_1 = -0.03585743$ and $\hat{\theta}_2 = -2.22720951$.

```
# Extract value of sample ACVF
C = as.numeric(acf(Y,type="covariance",plot=FALSE)$acf)
(thetaMM = c(C[2]/(1+C[3]), C[3])) # Value of thetahat1 and thetahat2
-0.03585743 -2.22720951
```

Least-Square Estimator for AR model (With R)

✎ **Exercise 6.** *Suppose* $\{Y_t\} \sim \mathrm{AR}(3)$*, i.e.,* $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + Z_t$ *with* $\{Z_t\} \sim \mathrm{WN}(0,\sigma^2)$*. Evaluate the Least-square estimator of* ϕ_1*,* ϕ_2*,* ϕ_3 *and* σ^2*. Find 95% CI for* ϕ_1*,* ϕ_2 *and* $\phi_2+\phi_3$*.*

Solution

It is equivalent to regressing Y_t against covariates Y_{t-1}, Y_{t-2} and Y_{t-3} (without intercept term). Also, recall from Tutorial 5 that $\sqrt{n}(\hat{\phi}-\phi) \xrightarrow{\mathrm{d}} \mathbb{N}_3(0,\sigma^2\Gamma_3^{-1})$, where $\Gamma_3(i,j) = \gamma(|i-j|)$. Notice that for $\Sigma := \sigma^2\Gamma_3^{-1}$, we have

$$\lim_{n\to\infty} n\mathrm{Var}(\hat{\phi}_2+\hat{\phi}_3) = \lim_{n\to\infty} n\mathrm{Var}(\hat{\phi}_2) + \lim_{n\to\infty} n\mathrm{Var}(\hat{\phi}_3) + 2\lim_{n\to\infty} n\mathrm{Cov}(\hat{\phi}_2,\hat{\phi}_3) = \Sigma(2,2)+\Sigma(3,3)+2\Sigma(2,3).$$

By the R-code below, we have $\hat{\phi}_1 = -0.08$, $\hat{\phi}_2 = -0.482$, $\hat{\phi}_3 = -0.184$.

```
LSE_fit = lm(Y[4:n] ~ Y[3:(n-1)] + Y[2:(n-2)] + Y[1:(n-3)]-1)
(phi = as.numeric(LSE_fit$coefficients))
[1] -0.08034209 -0.48242400 -0.18428814
(sigma2 = sum((LSE_fit$residuals)^2)/(n-2*3))
[1] 3.693695
C = as.numeric(acf(Y, type="covariance", plot=FALSE)$acf)
Gamma_p = array(NA,dim=c(3,3))
for(i in 1:3){for(j in 1:3) Gamma_p[i,j] = C[abs(i-j)+1]}
Sigma_hat = sigma2*solve(Gamma_p) # Estimated Asymptotic Covariance Matrix
V1 = qnorm(0.975)*sqrt(Sigma_hat[1,1]/n)
V2 = qnorm(0.975)*sqrt(Sigma_hat[2,2]/n)
V23 = qnorm(0.975)*sqrt((Sigma_hat[2,2]+Sigma_hat[3,3]+2*Sigma_hat[2,3])/n)
(CI1 = c(phi[1]-V1,phi[1]+V1)) # CI for phi1
[1] -0.21279810  0.05211397
(CI2 = c(phi[2]-V2,phi[2]+V2)) # CI for phi2
[1] -0.5999146 -0.3649336
(CI23 = c(phi[2]+phi[3]-V23,phi[2]+phi[3]+V23)) # CI for phi1+phi2
[1] -0.8424410 -0.4909837
```

Yule-Walker Estimator for AR model (With R)

Exercise 7. *Suppose* $\{Y_t\} \sim \mathrm{AR}(4)$, *i.e.,* $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + Z_t$ *with* $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. *Evaluate the Yule-Walker estimator of* ϕ_1, ϕ_2, ϕ_3 *and* ϕ_4.

Solution

Recall from Tutorial 5, $\hat{\phi} = M^{-1}(r_1, \ldots, r_4)^T$, where $M(i,j) = r_{|i-j|}$. From the R-code below, we have $\hat{\phi} = (-0.096, -0.548, -0.180, -0.192)$.

```r
r = as.numeric(acf(Y, type="correlation", plot=FALSE)$acf)[1:5] #Compute r0,...,r4
M = array(NA,dim=c(4,4))
for(i in 1:4){for(j in 1:4) M[i,j] = r[abs(i-j)+1]}
(phi_YW = solve(M)%*%r[2:5]) # Yule-Walker Estimator
           [,1]
[1,] -0.0962834
[2,] -0.5479180
[3,] -0.1799613
[4,] -0.1919126
```

Conditional Least Square Estimator for ARMA model (With R)

Exercise 8. *Assume that* $\{Y_t\} \sim \mathrm{ARMA}(2,3)$, *i.e.,* $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \theta_3 Z_{t-3}$. *Find the CLS estimator for* $\phi_1, \phi_2, \phi_3, \theta_1, \theta_2$ *and* σ^2.

Solution

Write $Z_t = Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3}$ and define $Z_s = Y_s = 0$ for $s \leq 0$. After we obtain the CLS estimate, we can define $\hat{Z}_t = Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} + \hat{\theta}_1 \hat{Z}_{t-1} + \hat{\theta}_2 Z_{t-2} + \hat{\theta}_3 \hat{Z}_{t-3}$ and hence $\hat{\sigma}^2 = \sum_{t=1}^{n} \hat{Z}_t^2 / n$.

```r
S = function(beta){
  phi1 = beta[1] ; phi2 = beta[2] ; theta1 = beta[3] ; theta2 = beta[4] ; theta3 = beta[5]
  Z = rep(NA,n)
  Z[1] = Y[1] ; Z[2] = Y[2] - phi1*Y[1] + theta1*Z[1]
  Z[3] = Y[3] - phi1*Y[2] - phi2*Y[1] + theta1*Z[2] + theta2*Z[1]
  for(k in 4:n){
    Z[k] = Y[k] - phi1*Y[(k-1)] - phi2*Y[(k-2)] + theta1*Z[(k-1)] + theta2*Z[(k-2)] + theta3*Z[(k-3)]
  }
  return(sum(Z^2))
}
CLS_fit = optim(c(0.1,0.1,0.1,0.1,0.1),S) # Optimize the function S wrt beta
(phihat = CLS_fit$par[1:2])
[1]  0.2207534 -0.2051981
(thetahat = CLS_fit$par[3:5])
[1] 0.39629000 0.39611940 0.03076929
(sig2hat = CLS_fit$value/n)
[1] 3.317136
```

Maximum Likelihood Estimator for ARMA model (With R)

✎ **Exercise 9.** *Assume that* $\{Y_t\} \sim \text{ARMA}(2,3)$*, i.e.,* $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \theta_3 Z_{t-3}$*. Find the MLE for* $\phi_1, \phi_2, \phi_3, \theta_1, \theta_2$ *and* σ^2 *and also the value of the maximized log-likelihood.*

Solution

$\hat{\phi}_1 = 0.130$, $\hat{\phi}_2 = -0.138$, $\hat{\theta}_1 = 0.378$, $\hat{\theta}_2 = 0.529$, $\hat{\theta}_3 = -0.093$ and $\hat{\sigma}^2 = 3.11$.

```
MLE_fit = arima(Y,order=c(2,0,3),include.mean = FALSE)
MLE_fit$coef
        ar1         ar2         ma1         ma2         ma3
 0.12987568 -0.13766270 -0.37820460 -0.52862527 -0.09316591
MLE_fit$loglik
[1] -409.6758
MLE_fit$sigma2
[1] 3.110787
```

❗ **Remark 6.** *In* R*, the ARMA model is of the form* $Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}$ *instead of* $Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + Z_t - \sum_{k=1}^{q} \theta_k Z_{t-k}$*, i.e., we have to be careful with the sign of the MA coefficient.*

5. After fitting the model, say the ARMA(p_1, q_1) model. We can compare its performance with the ARMA(p_2, q_2) model with different criteria.

Model Selection

✎ **Exercise 10.** *Consider the class of* $\text{ARMA}(p,q)$ *model with* $1 \le p, q \le 5$*. Which model is the best in terms of AIC, AICC, and BIC?*

Solution

According to the R-code below, both AIC and AICC suggest ARMA$(5,4)$ while BIC suggest ARMA$(4,5)$.

```
IC=function(x,order.input){
  fit = arima(x,order=c(order.input[1],0,order.input[2]));
  n = length(x) ; p = order.input[1] ; q = order.input[2] ; sig2 = fit$sigma2
  AIC = fit$aic
  AICC = AIC - 2*(p+q+1) + 2*(p+q+1)*n/(n-p-q-2)
  BIC=(n-p-q)*log(n*sig2/(n-p-1))+n*(1+log(sqrt(2*pi)))+
    (p+q)*log((sum(x^2)-n*sig2)/(p+q))
  FPE = sig2*(n+p)/(n-p) # Only valid for AR model
  out =t(as.matrix(c(AIC,AICC,BIC,FPE)));colnames(out)=c("AIC","AICC","BIC","FPE")
  return(out)
}

out = array(NA,dim=c(5,5,4))
for(i in 1:5){for(j in 1:5) out[i,j,] = IC(Y,c(i,j)) }
(order_AIC = arrayInd(which.min(out[,,1]),c(5,5)))
     [,1] [,2]
[1,]    5    4
(order_AICC = arrayInd(which.min(out[,,2]),c(5,5)))
     [,1] [,2]
[1,]    5    4
(order_BIC= arrayInd(which.min(out[,,3]),c(5,5)))
     [,1] [,2]
[1,]    4    5
```


6. After we have chosen a model through model selection, we should conduct residual analysis, i.e.,, consider the time-series plot, ACF, and PACF plot, and carry out the Portmanteau test.

Residual Analysis

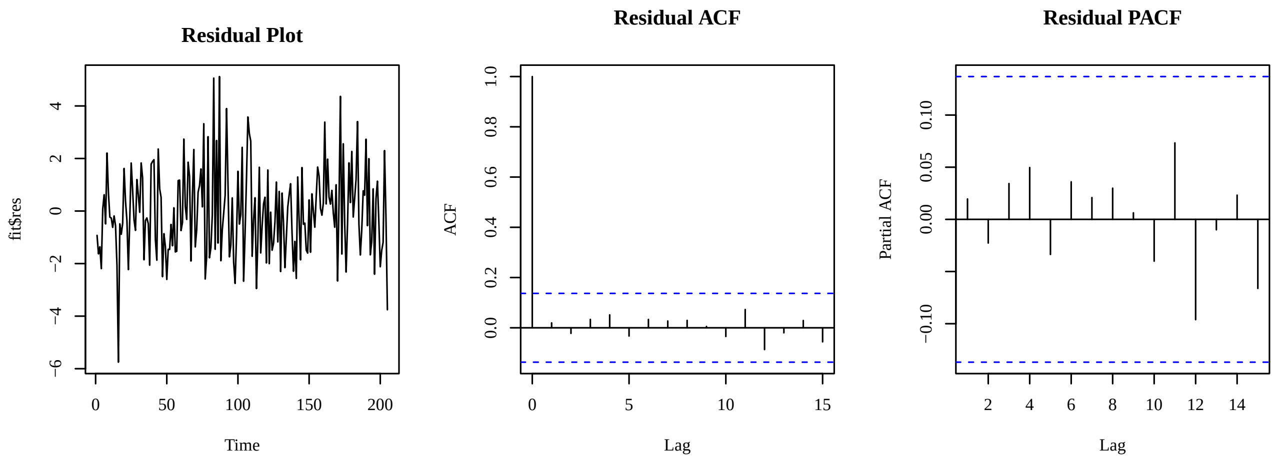
✎ **Exercise 11.** *Suppose we choose the* ARMA$(5,4)$ *model for* $\{Y_t\}$*. We want to check its goodness of fit through the time-series plot, ACF, and PACF plot. Also, perform the portmanteau test with* $h = 15$*. Clearly state the hypothesis and your conclusion.*

Solution

Let Z be the noise under this ARMA$(5,4)$ model and $\rho_Z(\cdot)$ as ACF of Z, we are testing

$$H_0 : \rho_Z(k) = 0 \text{ whenever } |k| \leq 15 \quad \text{against} \quad H_1 : \rho_Z(k) \neq 0 \text{ for some } |k| \leq 15$$

According to the result in the R-code below, we do NOT reject H_0, i.e. at least ARMA$(5,4)$ is NOT a bad fit for $\{Y_t\}$. Also, from the graph below, we can see that both ACF and PACF of residuals are not significant. Suggesting that the residual might be a white-noise sequence.




```
p = 5 ; q = 4 ; h = 15
fit = arima(Y,order=c(p,0,q))
par(mfrow=c(1,3))
ts.plot(fit$res,main="Residual Plot")
r.z = as.numeric(acf(fit$res,lag.max=h,main="Residual ACF")$ acf)
pacf(fit$res,lag.max=h,main="Residual PACF")
Qh = n*(n+2)*sum((r.z[-1]^2)/(n-(1:h)))
Qh > qchisq(0.95,h-p-q)
[1] FALSE
```