TIME SERIES Modeling

Chapter 3: diagnostic checking and model selection

1 Model selection using Information criteria

Evaluation of the graphs of sample ACF and PACF allows to preliminarily choose order p and q of ARMA model to be fit. A final decision is done using AIC criterion which allows us to compare the fit of different models.

Definition 1 Assume that a statistical model of M parameters is fitted to data. The Akaike's Information Criterion (AIC) statistic is defined as

$$AIC = -2ln[\text{maximum likelihood of data}] + 2M.$$

Suppose that the white noise Z_t is Gaussian distribution with variance σ^2. Let $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \hat{X}_j)^2/r_{j-1}$ with r_{j-1} being some constants, independent of σ^2. It turns out that the log likelihood function of ARMA(p,q) models is

$$\ln L = -\frac{n}{2} ln\hat{\sigma}^2 + Const.$$

Definition 2 The Akaike's Information Criterion (AIC) statistic for ARMA(p,q) models is defined as

$$AIC = nln\hat{\sigma}^2 + 2(p+q),$$

where $\hat{\sigma}^2$ stands for the estimated error variance and n is the number of observations.

How does it works ? Choose the model (choose the values of p, q) with minimum AIC.

Intuitively, one can think of $2(p+q)$ as a penalty term to discourage over-parameterization.

2 Model diagnostic checking

OK means that the fitted model can describe the dependence structure of a time series adequately.

If ARMA model

$$X_t = \delta + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

is adequate, then $\{Z_t\}$ should be white noise (WN).

If the ARMA model is adequate, the residuals should be WN (approximately at least).

Recall what is white noise? If $\{Z_t\}$ is WN, then ρ_k (or $\rho(k)$) is zero. In practice we may use its sample ACF. For example, if an AR(1) model is considered, the residuals are

$$\hat{Z}_t = X_t - \hat{\delta} - \hat{\phi}_1 X_{t-1}.$$

To check the dependence structure, we calculate

$$r_k = \frac{\sum_{t=1}^{n-k}(\hat{Z}_t - \bar{a})(\hat{Z}_{t+k} - \bar{a})}{\sum_{t=1}^{n}(\hat{Z}_t - \bar{a})^2}, \qquad k \geq 1$$

where $\bar{a} = \sum \hat{Z}_t / n$. Here r_k is called the residual autocorrelation at lag k. Thus, if a model is adequate we expect

$$r_h \approx 0.$$

Theorem 1 If $H_0 : \rho(k) = 0$ is true, then

$$\hat{\rho}(k) = r_k \sim N(0, \frac{1}{n}).$$

We can use the above as a rough guide on whether each $\rho(k)$ is zero.

Definition 3 Overall test, define

$$Q(m) = n(n+2)\sum_{k=1}^{m} r_k^2/(n-k)$$

where $0 << m << n$ (usually, $m \approx n/5$). $Q(m)$ is called the Ljung-Box statistic (or Portmanteau statistic).

If the fitted model is <u>OK</u> (adequate), then

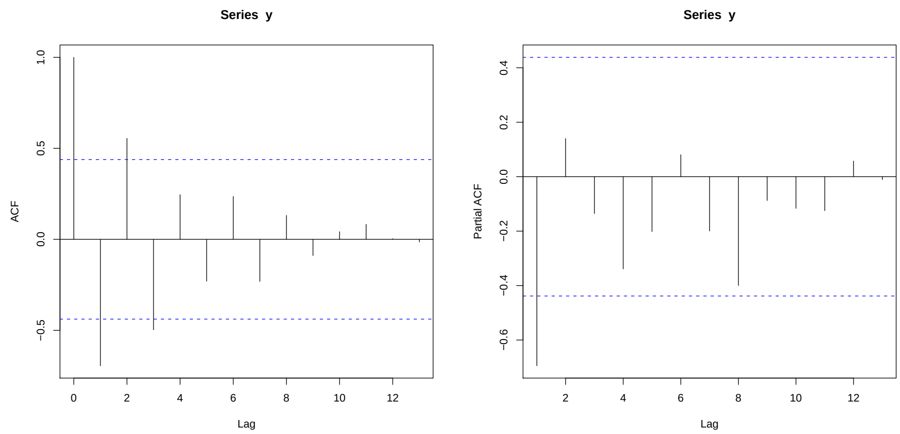
$$Q(m) \sim \chi^2_{m-n_p}$$

where n_p is the number of parameters (exclusive of δ) in the ARMA model. For example, if the model is $X_t = \phi_1 X_{t-1} + Z_t$, then $n_p = 1$.

e.g. if the model is $X_t = \phi_1 X_{t-1} + \phi_3 X_{t-3} + Z_t$, then $n_p = 2$.

e.g. if the model is $X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-3} + \theta_1 Z_{t-1} + Z_t$, then $n_p = 3$ (δ will not be counted).

EXAMPLE 1 The data $X_t, t = 1, \cdots, 20$ are observed as: 0.50, -0.41, 0.37, -0.61, 0.23, -0.13, 0.06, -0.11, 0.18, -0.14, 0.20, 0.09, -0.03, -0.02, -0.14, -0.07, 0.09, 0.09, -0.01, -0.10 ■



We may try model $X_t = \delta + \phi_1 X_{t-1} + Z_t$ by looking at SPACF and SACF.

fit = arima(y, order = c(1,0,0))

The fitted model

$$\hat{X}_t = 0.0075 - 0.832 X_{t-1}$$

The residuals are $e_t : 2, 3, \ldots, 20$: 0.00, 0.02, -0.31, -0.29, 0.05, -0.06, -0.07, 0.08, 0.00, 0.08, 0.25, 0.04, -0.05, -0.16, -0.19, 0.02, 0.16, 0.06, -0.12

The SACF for e_t are

$$r_1 = 0.34, r_2 = -0.21 r_3 = -0.12,$$

$$r_4 = -0.22, r_5 = -0.09, r_6 = 0.09,$$

$$r_7 = -0.18, r_8 = -0.24, r_9 = 0.02, r_{10} = 0.10$$

(each $H_0 : \rho_e(k) = 0$, $k = 1, \ldots, 10$ can be accepted separately, why?)

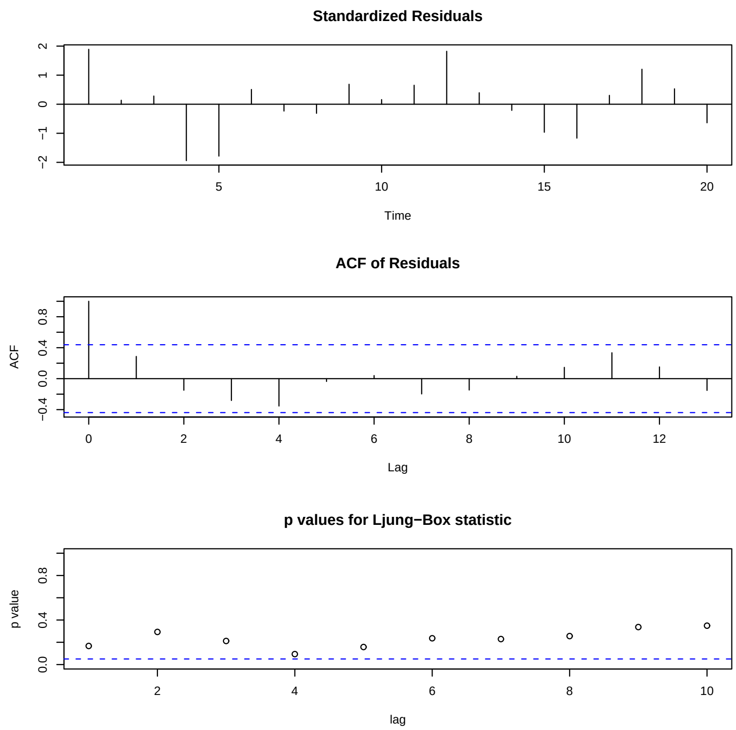
Consider the Ljung-Box test. If we let m = 5, then

$$\begin{aligned} Q(5) &= n*(n+2)*(r_1^2/(n-1) + r_2^2/(n-2) \\ &\quad + r_3^2/(n-3) + r_4^2/(n-4) + r_5^2/(n-5)) \\ &= 5.6964 \end{aligned}$$

Since $\chi^2_{0.05}(5-1) = 9.49$, $Q(5) < \chi^2_{0.05}(5-1)$. Thus we can not reject the adequacy of the model by setting α equal to 0.05.

From the table of the residual we can also see that the p value is $0.3819 > 0.05 = \alpha$ when taking m=6. Again we can not reject the adequacy of the model by setting α equal to 0.05. This is consistent with that obtained by comparing the critical value with the observed statistic.

Using tsdiag(fit) yields the following plot.



3 Using ACF and PACF of of residuals to improve the model

Example 2 : Suppose that we fit AR(1) model

$$X_t = \phi_1 X_{t-1} + Z_t.$$

IF the SACF of $\hat{Z}_t$ has a cut-off after lag 1, then it suggests

$$\hat{Z}_t \sim MA(1)$$

i.e

$$\hat{Z}_t = e_t + \theta e_{t-1}$$

Thus

$$X_t \sim ARMA(1,1).$$

Hopefully, $\hat{e}_t$ is now closer to white noise.

If the SPACF of $\hat{Z}_t$ has a cut-off after lag 1, it suggests

$$\hat{Z}_t \sim AR(1)$$

i.e

$$Z_t = \psi_1 Z_{t-1} + e_{t-1}.$$

Thus

$$(X_t - \phi_1 X_{t-1}) = \psi_1(X_{t-1} - \phi_1 X_{t-2}) + e_{t-1}$$

i.e.

$$X_t \sim AR(2).$$

Hopefully, $\hat{e}_t$ is now closer to white noise.

4 How to change a non-stationary time series into stationary one

According to the definition of stationarity there are 3 types of non-stationarity:

- Non-stationarity in mean: EX_t depends on t;
- Non-stationarity in variance: $\text{var}(X_t)$ depends on t;
- Non-stationarity in covariance: $\text{cov}(X_t, X_{t+k})$ depends on t for some k.

EXAMPLE 3 Suppose that $\{Y_t\}$ is a stationary time series. Let $X_t = a + bt + Y_t$. Apparently, X_t is not a stationary process. However applying the first difference operator $\bigtriangledown$ to $\{X_t\}$ yields a stationary process. Therefore applying the difference operator is one way of obtaining a stationary process. ■

EXAMPLE 4 Suppose $\{Y_t\}$ is an $i.i.d.$ sequence with $\gamma(k) = Cov(Y_t)$ independent of time t. Let $X_t = e^{t+Y_t}$. Can we find a suitable d such that $\{(1-B)^d X_t\}$ is stationary? (No).

Let $U_t = \log(X_t)$. What happen to $\{U_t\}$? ■

Not all non-stationary series can be transformed to stationary ones by differencing. Many time series are stationary in the mean but are not stationary in the variance such as Example 2. To overcome this problem, we need to stabilize the variance of the time series by using pre-differencing transformation.

We first consider the power transformation to remove some possible non-stationarity in variance.

$$T(X_t) = \frac{X_t^\lambda - 1}{\lambda}$$

When and how to perform the power transformation?

Values of lambda	Transformation
-1.0	$\frac{1}{X_t}$
-0.5	$\frac{1}{\sqrt{X_t}}$
0.0	$\ln(X_t)$
0.5	$\sqrt{X_t}$
1.0	X_t (no transformation)

(a) If the variability of a time series increases as time advances it then implies that the time series is non-stationary with respect to its variance; See figure 1 below.

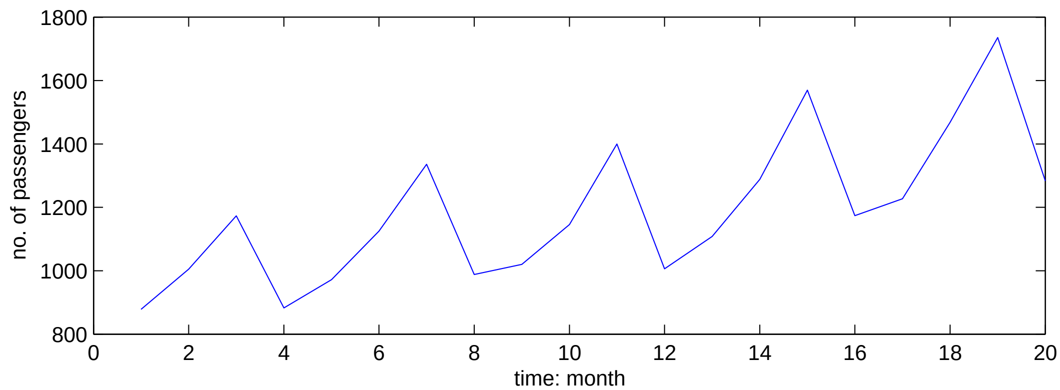


Figure 1:

(b) Find the one with minimum sample variance.

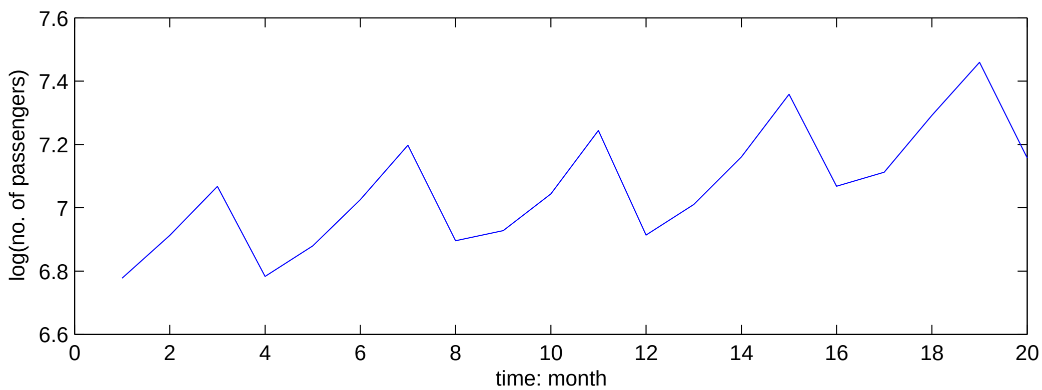


Figure 2: log transformation of data in Fig 1

After transformation, we then consider the possible differecing to make the time series stationary.

Criteria: The ACFs of non-stationary time series converges to zero **slowly**, however, those of stationary time series converges to zero **fast**.

5 ARIMA model

EXAMPLE 5 Figure 3 shows US Dow Jones Industrial Average Market Index $\{Y_t\}$ from 17-Jul-02 to 20-Mar-03.

$\{Y_t\}$ is not stationary. See figures 3 and 4.

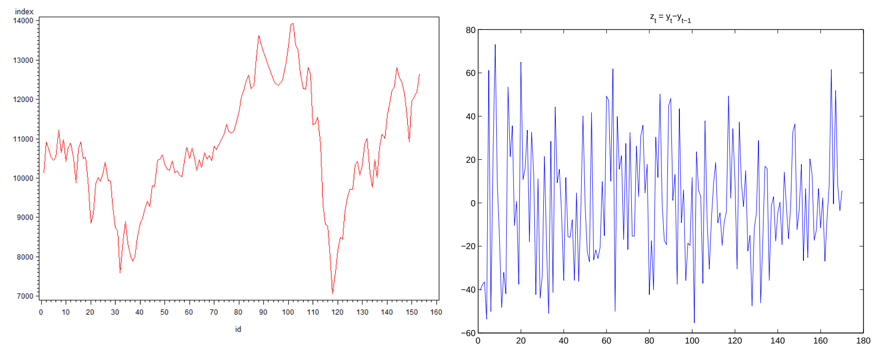


Figure 3:

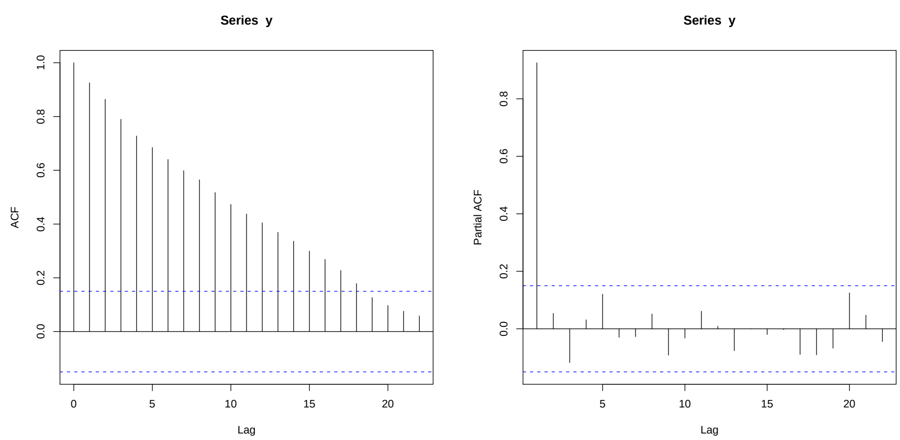


Figure 4: series Y

Moreover, by figure 5 we can fit the following models to the data

$$x_t = Z_t + \theta_{19} Z_{t-19}$$

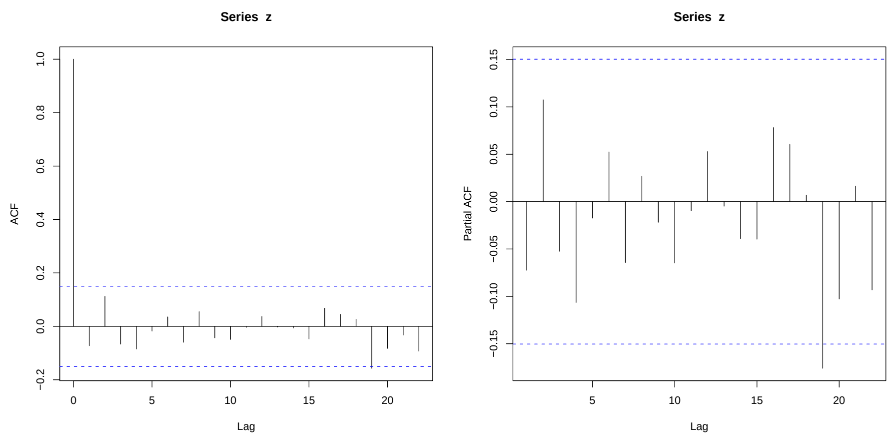


Figure 5: series z or x

or

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_{19} x_{t-19} = Z_t.$$

Generally, we can fit the difference $x_t = X_t - X_{t-1}$ of a time series by a ARMA(p,q) model,

$$\phi_p(B)(X_t - X_{t-1}) = \theta_q(B) Z_t$$

or

$$\phi_p(B)(1 - B) X_t = \theta_q(B) Z_t.$$

This is an ARIMA(p, 1, q) model.

We can also consider higher order difference,

$$\begin{aligned} w_t &= x_t - x_{t-1} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= X_t - 2X_{t-1} + X_{t-2} = (1 - 2B + B^2) X_t \\ &= (1 - B)^2 X_t. \end{aligned}$$

If we fit w_t by

$$\phi_p(B) w_t = \theta_q(B) Z_t$$

or

$$\phi_p(B)(1 - B)^2 X_t = \theta_q(B) Z_t.$$

This is an ARIMA(p,2,q) model. More generally, we define ARIMA(p,d,q) as

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B) Z_t.$$

For the example, we can fit ARIMA(0,1,19) to Y_t in the example

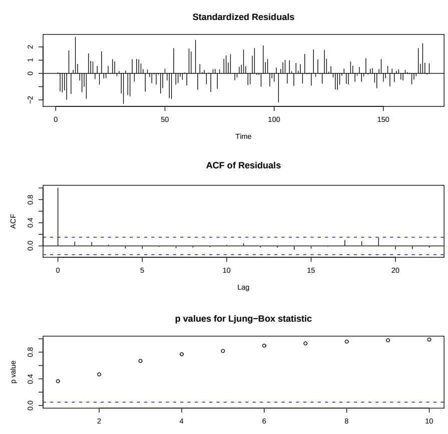
For the example, we can fit ARIMA(0,1,19) to y_t in the example

fitma = arima(y, order = c(0, 1, 19))
Call: arima(x = y, order = c(0, 1, 19))
Coefficients:

	ma1	ma2	ma3	ma4	ma5	ma6	ma7	ma8
	-0.2219	0.0249	-0.1093	-0.0702	0.0666	0.0518	0.0191	0.1042
s.e.	0.0950	0.0832	0.0863	0.0898	0.0908	0.0867	0.0899	0.0948
	ma9	ma10	ma11	ma12	ma13	ma14	ma15	ma16
	-0.0537	-0.0714	-0.0685	0.1185	0.0134	0.0531	-0.0942	0.0521
s.e.	0.1016	0.1087	0.0896	0.0893	0.0987	0.1004	0.1145	0.1075
	ma17	ma18	ma19					
	-0.1438	-0.1234	-0.4681					
s.e.	0.0972	0.1059	0.0971					

```
sigma^2 estimated as 665.5:  log likelihood = -799.39,  aic =
1638.78
```

tsdiag(fitma)



predict(fitma, n.ahead= 20)
Call: arima(x = y, order = c(19, 1, 0))
Coefficients:

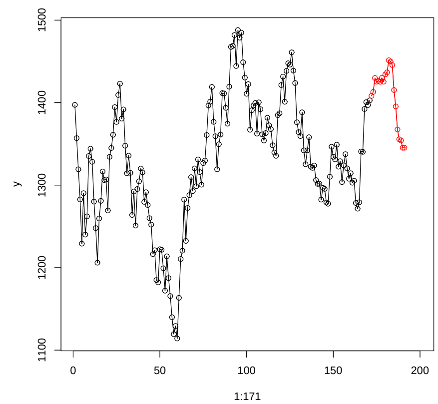


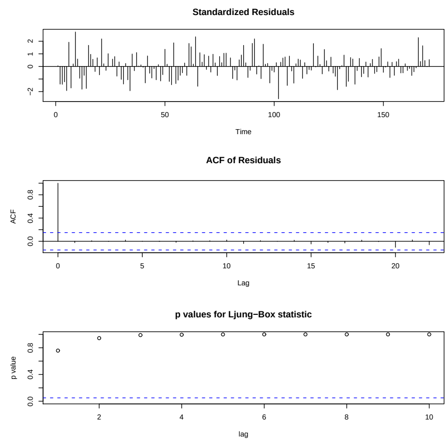
Figure 6: the black dot is the observation; the red dots are the predictions

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	ar8
	-0.0665	0.1378	-0.0443	-0.1369	-0.0163	0.0462	-0.0396	0.0475
s.e.	0.0749	0.0757	0.0764	0.0770	0.0796	0.0808	0.0808	0.0818
	ar9	ar10	ar11	ar12	ar13	ar14	ar15	ar16
	-0.0416	-0.1021	-0.0085	0.038	0.0039	-0.0407	-0.0506	0.0904
s.e.	0.0823	0.0817	0.0821	0.082	0.0826	0.0836	0.0837	0.0836
	ar17	ar18	ar19					
	0.1094	-0.0049	-0.2208					
s.e.	0.0840	0.0837	0.0832					

```
 sigma^2 estimated as 726.2:  log likelihood = -801.95,
aic = 1643.9
```

tsdiag(fitma)
predict(fitma, n.ahead= 20)
The fitted model is

Coefficients:
Factor 1: 1 + 0.05076 B**(1) - 0.03281 B**(2) + 0.0297 B**(3) + 0.00626 B**(4) - 0.01162 B**(5) - 0.01904 B**(6) - 0.06261 B**(7) - 0.09867 B**(8) - 0.14025 B**(9) - 0.11528 B**(10) - 0.02889 B**(11) - 0.00768 B**(12) - 0.2065 B**(13) - 0.18588 B**(14) + 0.03721 B**(15) + 0.00595 B**(16) - 0.22064 B**(17)-0.1234B**(18) -0.4681B**(19)



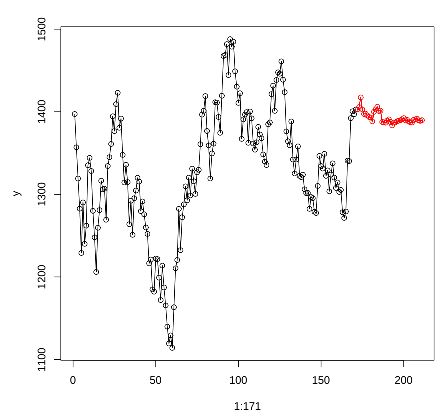


Figure 7: the black dot is the observation; the red dots are the predictions

Example 6 Weekly sales of Super Tech Videocassette Tape [the data can be found at the website].

```
> plot(1:161, y, xlim = c(0, 200), ylim=c(20, 100) )
> lines(1:161, y, type="l" )          # l for L
```

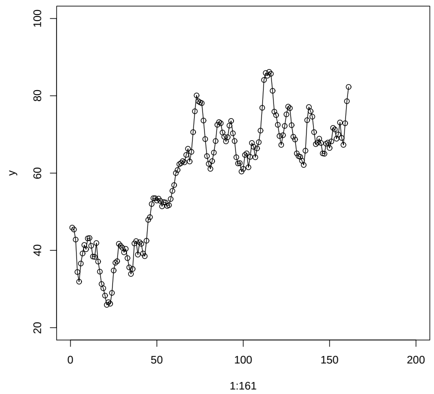


Figure 8: series y

```
> acf(y, lag.max=30)
> pacf(y, lag.max=30)
```

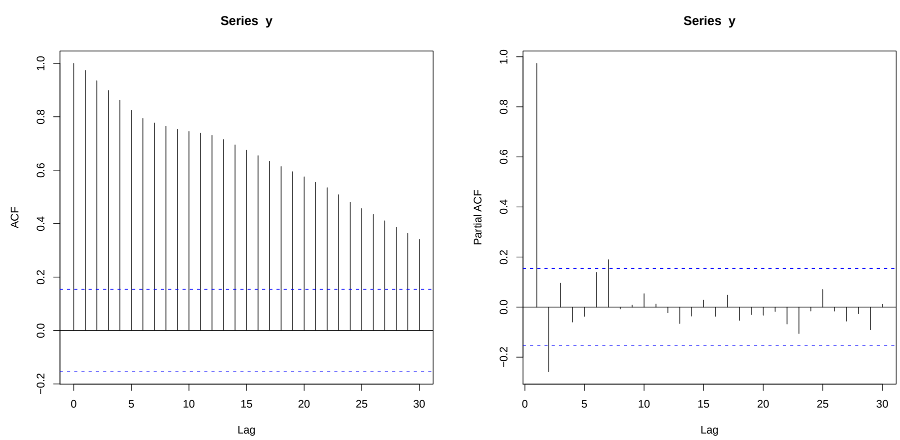


Figure 9: series y

The the raw data is not stationary. We take difference

$$z_t = y_t - y_{t-1} = (1 - B)y_t.$$

```
> acf(z, lag.max=30)
> pacf(z, lag.max=30)
```

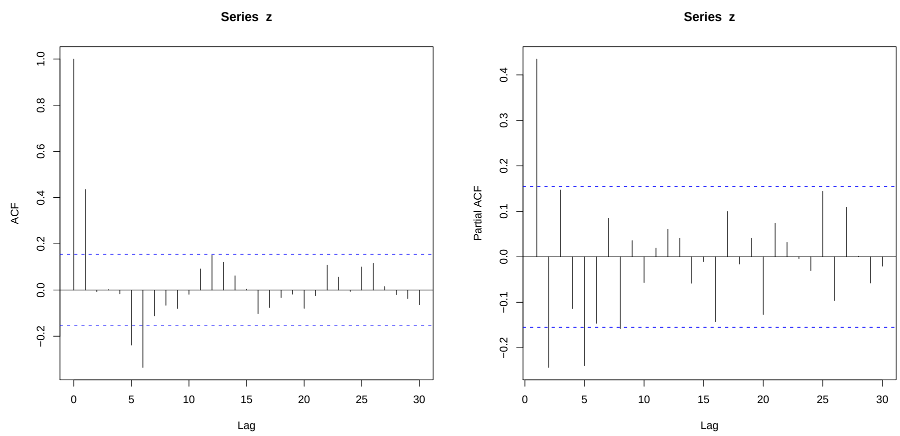


Figure 10: series x

We can use ARIMA(0, 0, 6) for z_t, or ARIMA(0,1,6) for y_t.

```
> fit = arima(z, order = c(0,1,6))
> tsdiag(fit)
```

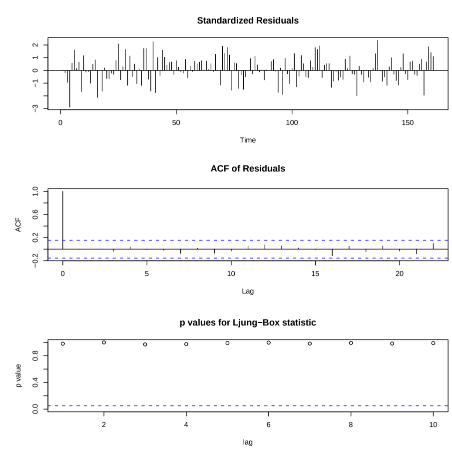


Figure 11: ACF of Residuals

Thus, the model is adequate (OK), i.e. there is no autocorrelation in the residuals.

1. Write down the estiamted model

```
> fit
Call: arima(x = y, order = c(0, 1, 6))
Coefficients:
```

	ma1	ma2	ma3	ma4	ma5	ma6
	0.6331	-0.0160	0.0361	-0.0264	-0.1490	-0.4374
s.e.	0.0771	0.0892	0.0917	0.0879	0.1055	0.0783

sigma^2 estimated as 4.896: log likelihood = -356.33, aic = 726.65
The fitted model is

$$\begin{aligned}(1-B)y_t &= Z_t + 0.6331Z_{t-1} - 0.0160Z_{t-2} + 0.0361Z_{t-3} \\ &\quad -0.0264Z_{t-4} - 0.1490Z_{t-5} - 0.4374Z_{t-6}.\end{aligned}$$

2. Fitted values are as follows.

```
> plot(1:161, y, xlim = c(0, 200), ylim=c(20, 100) )
> lines(1:161, y, type="l" )
> lines(1:161, y-fit$residuals, type="l", col="red")
```

3. Forecast for 6 steps ahead

```
> forecast = predict(fit, n.ahead=6)
> lines(162:167, forecast$pred, type="o", col="red")
> lines(162:167, forecast$pred-1.96*forecast$se, col="blue")
> lines(162:167, forecast$pred+1.96*forecast$se, col="blue")
```

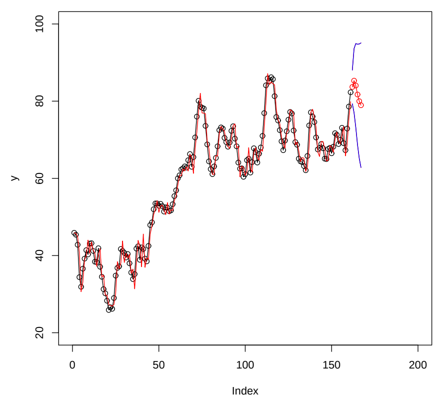


Figure 12: the black dots are the observation and the red dots are the predictions