TIME SERIES ANALYSIS

Chapter 4
Estimation for nonseasonal Box-Jekins models

Basic Questions:

(a) Use R to implement the estimation.

(b) Estimate ARMA models based on the properties.

1 Estimation of an AR(p) model

Suppose that $X_t : t = 1, 2, \cdots, n$ is a TS (time series), and we want to fit the following regression model

$$X_t = \delta + \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + Z_t$$

time	X_t	intercept	X_{t-1}	X_{t-2}	$\cdots$	X_{t-k}
1	X_1	1	–	–	$\cdots$	–
2	X_2	1	X_1	–	$\cdots$	–
3	X_3	1	X_2	X_1	$\cdots$	–
$\vdots$	1	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
k+1	X_{k+1}	1	X_k	X_{k-1}	$\cdots$	X_1
k+2	X_{k+2}	1	X_{k+1}	X_k	$\cdots$	X_2
$\vdots$	$\vdots$	1	$\vdots$	$\vdots$	$\cdots$	$\vdots$
n	X_n	1	X_{n-1}	X_{n-2}	$\cdots$	X_{n-k}

what is X and Y?

Using LSE (least-squares estimation) we have

$$\begin{pmatrix} \hat{\delta} \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} = (X' * X)^{-1} X' * Y.$$

Similarly, we can calculate

a) fitted value (for $t \leq n$) and prediction (for $t > n$)

$$\hat{X}_t = \hat{\delta} + \hat{\phi}_1 X_{t-1} + \cdots + \hat{\phi}_p X_{t-p}$$

b) Prediction errors

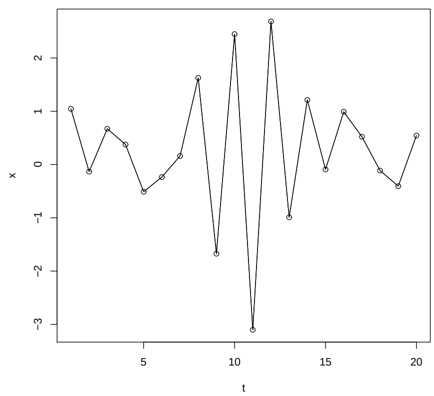
$$\hat{Z}_t = X_t - (\hat{\delta} + \hat{\phi}_1 X_{t-1} + \cdots + \hat{\phi}_p X_{t-p})$$

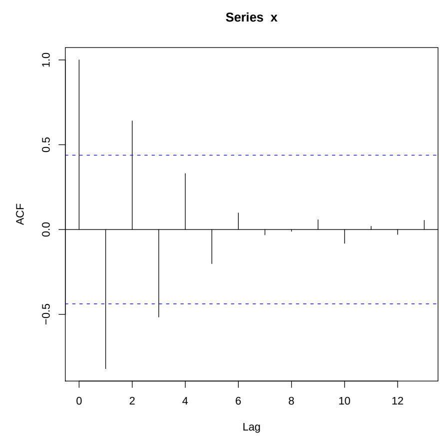
c) The estimator of $\sigma^2 = Var(Z_t)$ is

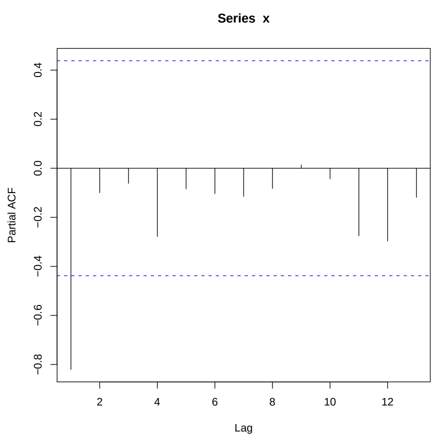
$$\hat{\sigma}^2 = \sum_{t=1}^{n} \hat{Z}_t^2/(n-p-1).$$

d) The standard error $\sqrt{\hat{\sigma}^2 c_{k,k}}$ for δ and ϕ_k, where $c_{k,k}$ is the (k,k) element in $(X' * X)^{-1}$.

EXAMPLE 1 Example (continued) X_t : 1.0445, -0.1338, 0.6706, 0.3755, -0.5110, -0.2352, 0.1595, 1.6258, -1.6739, 2.4478, -3.1019, 2.6860, -0.9905, 1.2113, -0.0929, 0.9905, 0.5213, -0.1139, -0.4062, 0.5438







time	X_t	intercept	X_{t-1}
1	1.0445	1	–
2	-0.1338	1	1.0445
3	0.6706	1	-0.1338
	$\vdots$	1	$\vdots$
20	0.5438	1	-0.4062

we have

$$(X' * X)^{-1} = \begin{pmatrix} 0.054303878 & -0.007102646 \\ -0.007102646 & 0.030166596 \end{pmatrix}$$

and

$$X' * Y = \begin{pmatrix} 3.97280 \\ -26.35337 \end{pmatrix}$$

We have

$$\hat{\delta} = 0.4029171, \hat{\phi}_1 = -0.8232089$$

The fitted values are (time: prediction) 2: -0.4569, 3: 0.5130, 4: -0.1491, 5: 0.0938, 6: 0.8235, 7: 0.5965, 8: 0.2716, 9: -0.9354, 10: 1.7809, 11: -1.6121, 12: 2.9564, 13: -1.8082, 14: 1.2183, 15: -0.5942, 16: 0.47939, 17: -0.4124, 18: -0.0262, 19: 0.4966, 20: 0.73730

The estimate of $\sigma^2 = Var(Z_t)$ is

$$\sum_{i=2}^{20} (X_t - \hat{X}_t)^2/(19-2) = 0.5948365$$

The standard error of δ is

$$\sqrt{(}0.5948365 * 0.054303878) = 0.1797274$$

and that of ϕ_1 is

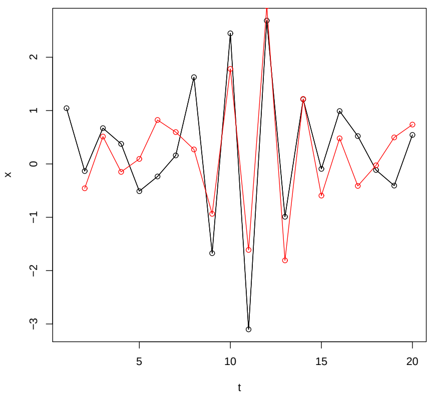
$$\sqrt{(}0.5948365 * 0.030166596) = 0.1339559$$

Therefore, our model is estimated as

$$\begin{aligned} X_t &= 0.4029171 - 0.8232089 X_{t-1} \\ & (0.1797274) \quad (0.1339559) \\ & \hat{\sigma}^2 = 0.5948365. \end{aligned}$$

$\mathbb{R}$ code

```
x = c(1.0445, -0.1338, 0.6706, 0.3755, -0.5110, -0.2352, 0.1595,
1.6258, -1.6739, 2.4478, -3.1019, 2.6860, -0.9905, 1.2113,
-0.0929, 0.9905, 0.5213, -0.1139, -0.4062, 0.5438)
```




```
est = arima(x, order = c(1,0,0))

est

OUTPUT

Call: arima(x = x, order = c(1, 0, 0))

Coefficients:
\[
\begint{rrr}
      &    ar1 & intercept\\
      &-0.8040   &    0.2255\\
s.e.  & 0.1153    &   0.0913\\
\endt
\]

sigma\^{}2 estimated as 0.5181:  log likelihood = -22.32,  aic =
50.65

---------
```

Then the estimated model is

$$(X_t - 0.2255) - 0.8040(X_{t-1} - 0.2255) = 0$$

or

$$X_t = 0.406802 - 0.8040X_{t-1}$$

2 Using R to Estimate of ARMA model (continued)

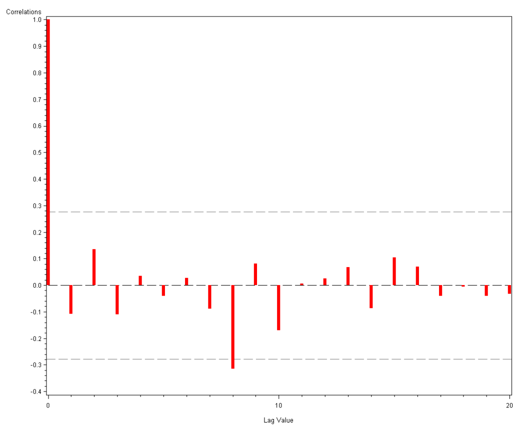
We use R to estimate ARMA(p,q) model

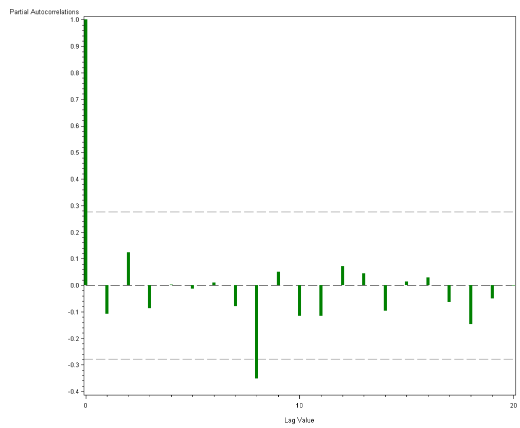
$$X_t = \delta + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}.$$

The formula for the l-step ahead forecast is

$$X_n(l) = E(X_{n+l}|X_n, X_{n-1}, \ldots).$$

Example 2 $y_t, t = 1, \cdots, 50$ are observed as: -1.30, -0.18, 0.94, -0.26, -1.05, -0.78, -0.82, 0.43, 0.57, 1.41, -1.47, 0.49, 0.00, -0.15, -0.64, 0.24, -0.79, 0.82, -0.20, -0.80, -0.22, 0.88, -0.75, 0.55, 0.73, -0.82, 0.70, -1.54, 0.04, -0.70, -0.58, -1.38, -1.28, 0.49, -0.76, 1.08, 0.16, 1.11, -0.06, 0.88, 0.89, 0.31, 0.03,-1.19, -0.38, 0.49, 1.02, -0.98, 0.50, -0.57





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

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

	ar1	ma1	intercept
	-0.7013	0.5768	-0.0946
s.e.	0.3067	0.3377	0.1024

sigma^2 estimated as 0.6086: log likelihood = -58.55, aic = 125.11
predict(fit, n.ahead=10)
$pred
Time Series:
Start = 51
End = 54
Frequency = 1
[1] 0.06499322 -0.20656294 -0.01612905 -0.14967441
$se

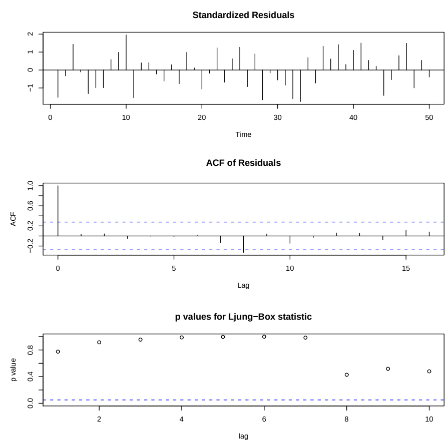
Time Series:
Start = 51
End = 54
Frequency = 1
[1] 0.7801178 0.7861337 0.7890753 0.7905180

The estimted model is

$$\hat{X}_t = \underset{}{-0.1695} - \underset{(0.3067)}{0.7013X_{t-1}} + Z_t + \underset{(0.3377)}{0.5768Z_{t-1}}$$

where $-0.1695 = -0.0946 * (1 + 0.7913)$.

tsdiag(fit)



3 Estimation of ARMA model based on the ACF and PACF: Yule–Walker estimation method

Consider an AR(p) model of the form,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t. \tag{3.1}$$

Our aim is to find estimators of the coefficient vector $\phi = (\phi_1, \ldots, \phi_p)$ and the white noise variance σ^2 based on the observations $X_1, \ldots, X_N$.

Recall **Yule–Walker equations**,

$$\begin{aligned} \gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p) &= \sigma^2, \\ \gamma(p) - \phi_1\gamma(p-1) - \cdots - \phi_p\gamma(0) &= 0; \quad (3.2) \end{aligned}$$

or the **Yule-Walker equations**

$$\begin{cases} \rho(1) = \phi_1 + \phi_2\rho(1) + \cdots + \phi_p\rho(p-1) \\ \rho(2) = \phi_1\rho(1) + \phi_2 + \cdots + \phi_p\rho(p-2) \\ \cdots \\ \rho(p) = \phi_1\rho(p-1) + \phi_2\rho(p-2) + \cdots + \phi_p \end{cases}$$

where $\rho(k)$ is ACF of the time series. We need to calculate sample ACF. We can solve the above equations to estimate $\phi_1, \cdots, \phi_p$ and σ^2.

Suppose that $X_1, X_2, \cdots, X_n$ are observations.

AR(1) model with mean 0: $X_t = \phi_1 X_{t-1} + \varepsilon_t$

Recall that we have

$$\gamma(1) = \phi_1\gamma(0)$$

i.e.

$$\phi_1 = \rho(1)$$

We can use sample ACF r_1 to estimate $\rho(1)$ thus ϕ_1: $\hat{\phi}_1 = r_1$.

Example 3 Fit an AR(1): $X_t = \phi_1 X_{t-1} + Z_t$ to data -0.06, -0.18, 0.06, 0.15, 0.13, -0.02, 0.19, -0.13, -0.26, -0.29, -0.17, -0.10, 0.10, 0.17, 0.04, 0.00, 0.15, 0.11, 0.01, 0.19

Because $r_1 = 0.4755$, the estimated model is

$$\hat{X}_t = 0.4755X_{t-1}.$$

AR(1) model with nonzero mean: $X_t = \delta + \phi_1 X_{t-1} + Z_t$. Let $x_t = X_t - \mu$ where $\mu = EX_t$. By doing so, x_t is now $x_t = \phi_1 x_{t-1} + Z_t$.

Thus, we need to estimate μ first as

$$\hat{\mu} = n^{-1}\sum_{t=1}^{n} X_t$$

Recall that we have

$$\gamma_x(1) = \phi_1\gamma_x(0)$$

i.e.

$$\phi_1 = \rho_x(1).$$

We can use sample ACF r_1 to estimate $\rho(1)$ thus ϕ_1: $\hat{\phi}_1 = r_1$; and $\hat{\delta} = (1-\hat{\phi}_1)\hat{\mu}$.

Example 4 Fit an AR(1): $X_t = \phi_1 X_{t-1} + Z_t$ to data: 5.05, 5.02, 4.78, 4.73, 4.86, 4.81, 4.86, 4.74, 4.89, 5.03, 5.13, 5.16, 5.19, 5.13, 5.16, 5.10, 5.04, 5.07, 4.95, 4.91

We have

$$\hat{\mu} = \bar{X} = n^{-1}\sum_{t=1}^{20} X_t = 4.98$$

and

$$r_1 = \sum_{t=1}^{19}(X_t - \bar{X})(X_{t+1} - \bar{X}) / \sum_{t=1}^{20}(X_t - \bar{X})^2 = 0.7747.$$

Thus

$$\hat{\delta} = (1 - \hat{\phi}_1)\hat{\mu} = 1.1220.$$

Finally the estimated model

$$\hat{X}_t = 1.1220 + 0.7747X_{t-1}.$$

AR(2) model with mean 0: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$
Recall that we have

$$\rho(1) = \phi_1 + \phi_2\rho(1)$$
$$\rho(2) = \phi_1\rho(1) + \phi_2.$$

We can then estimate ϕ_1 and ϕ_2 by solving

$$r_1 = \phi_1 + \phi_2 r_1$$
$$r_2 = \phi_1 r_1 + \phi_2$$

(where r_1, r_2 are sample ACF).

Example 5 Fit an AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$ to data 0.15, -0.06, -0.39, -0.56, -0.52, -0.26, -0.11, 0.32, 0.31, 0.01, 0.00, 0.17, 0.52, 0.32, -0.08, -0.30, -0.16, 0.32, 0.29, 0.07

Because $r_1 = 0.64, r_2 = 0.04$, by

$$0.64 = \phi_1 + 0.64\phi_2$$
$$0.04 = 0.64\phi_1 + \phi_2$$

we have

$$\phi_1 = 1.04, \quad \phi_2 = -0.62.$$

The estimated model is

$$\hat{X}_t = 1.04X_{t-1} - 0.62X_{t-2}.$$

AR(2) model with nonzero mean: $X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$. Let $x_t = X_t - \mu$ where $\mu = EX_t$. By doing so, x_t is now $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + Z_t$.

Thus, we need to estimate μ first as

$$\hat{\mu} = n^{-1}\sum_{t=1}^{n} X_t$$

We estimate model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + Z_t$ first, say

$$\hat{x}_t = \hat{\phi}_1 x_{t-1} + \hat{\phi}_2 x_{t-2}.$$

Then the model for X_t is

$$\hat{X}_t = \hat{\mu}(1 - \hat{\phi}_1 - \hat{\phi}_2) + \hat{\phi}_1 X_{t-1} + \hat{\phi}_2 X_{t-2}.$$

Example 6 Example: X_t: 1.25, 1.64, 1.78, 1.33, 1.21, 1.04, 1.04, 1.55, 1.31, 0.89, 0.78, 1.28, 1.79, 2.42, 2.09, 1.57, 1.05, 0.97, 1.26, 1.70

Because $\bar{X} = 1.40$. Let $x_t = X_t - 1.40$. We have, for x_t

$$r_1 = 0.57, \quad r_2 = -0.11$$

By solving

$$\begin{aligned} 0.57 &= \phi_1 + 0.57\phi_2 \\ -0.11 &= 0.57\phi_1 + \phi_2 \end{aligned}$$

we have $\hat{\phi}_1 = 0.94, \hat{\phi}_2 = -0.65$. The estimated model for X_t is

$$\hat{X}_t = 0.994 + 0.94X_{t-1} - 0.65X_{t-2}$$

where $0.994 = 1.40 * (1 - 0.94 + 0.65)$.

Example 7 Generate an AR(2) model of the form: $X_t = 0.7X_{t-1} - 0.5X_{t-2} + Z_t$ and then estimate $\phi_1 = 0.7, \phi_2 = -0.5$ by $(\widetilde{\phi}_1, \widetilde{\phi}_2)$ using the following codes:

```
> ar.sim<-arima.sim(list(order=c(2,0,0),
+ ar=c(0.7, -0.5)), n=500)
> ts.yw<-ar.yw(ar.sim, order.max=2)
> ts.yw

Call:
ar.yw.default(x = ar.sim, order.max = 2)

Coefficients:
      1        2
 0.7502  -0.4607

Order selected 2 $\sigma^2$ estimated as 1.094.
```

Example 8 We now consider using a real data set `lh` available in `R`: a regular time series giving the luteinizing hormone in blood samples at 10 mins intervals from a human female, 48 samples.

```
> lh
Time Series:
Start = 1
End = 48
```

```
> ts.yw<-ar.yw(lh, order.max=5)
> ts.yw

Call:
ar.yw.default(x = lh, order.max = 5)
Coefficients:
      1        2        3
 0.6534  -0.0636  -0.2269

Order selected 3 $\sigma^2$ estimated
as 0.1959.
```

Conclusion: This suggests using an AR(3) for the `lh` data.

————————————————

MA(1) model with mean 0: $X_t = Z_t + \theta_1 Z_{t-1}$

We have

$$\rho_1 = \frac{\theta_1}{1+\theta_1^2}.$$

Thus

$$\rho_1 \theta_1^2 - \theta_1 + \rho_1 = 0$$

i.e.

$$r_1 \hat{\theta}_1^2 - \hat{\theta}_1 + r_1 = 0.$$

We can estimate θ_1 by solving the above equations. (we discard the root with absolute value greater 1).

MA(1) model with nonzero mean: $X_t = \delta + Z_t + \theta_1 Z_{t-1}$. Because $EX_t = \delta$. define $z_t = X_t - \delta$. We can estimate MA(1) for z_t and then X_t.

Example 9 X_t: -0.89 -0.53 0.54 -0.26 -1.34 -1.97 -0.35 0.46 -0.08 -1.13 0.04 1.64 1.95 0.94 -0.11 0.18 0.72 0.91 -1.09 0.12 1.29 0.79 1.67 -0.60 -1.72 -0.76 -2.60 -1.71 -0.39 -1.18

Fit a MA(1) model

$$\bar{y} = \sum_{t=1}^{30} X_t/30 = -0.182, \quad r_1 = 0.5$$

We have

$$\hat{\theta}_1 = \frac{1-\sqrt{1-4r_1^2}}{2r_1} = 1.00.$$

Thus the estimated model is

$$\hat{X}_t = Z_t + 1.00 Z_{t-1}.$$

MA(q) model with nonzero mean: There is no analytic solution. For some special cases, we still have solutions. For example

$$X_t = \delta + Z_t + \theta_p Z_{t-p}$$

——————-

There are no simple methods for **Estimation of ARMA model ARMA(p,q)** (the details are beyond the scope of the module).