Chapter 3

Instructor: Pan Guang-Ming

Division of Mathematical Sciences, SPMS
Nanyang Technological University, Singapore

Outline

1. Model selection by AIC
   - *How to select a model based on ACF and PACF*
   - *How to write down the fitted model based on output*
2. diagnostic checking
   - *Using the Ljung-Box test*
   - *P values*
3. Using ACF and PACF of of residuals to improve the model
3. Fit of ARIMA models

Outline

1. Model selection by AIC
   - *How to select a model based on ACF and PACF*
   - *How to write down the fitted model based on output*
2. diagnostic checking
   - *Using the Ljung-Box test*
   - *P values*
3. Using ACF and PACF of of residuals to improve the model
3. Fit of ARIMA models

AIC(1)

- 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.

AIC

Definition

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

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

The log likelihood function of ARMA(p,q) models

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

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

AIC in time series

Definition

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.

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

How it work ?

How does it works ?

Choose the model (choose the values of p, q) with minimum AIC.

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).

SACF of residuals

- 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

$$r_h \approx 0.$$

The guide of determining whether each ACF is zero

Theorem

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.

Ljung-Box statistic

Definition

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.

Counting the number of parameters

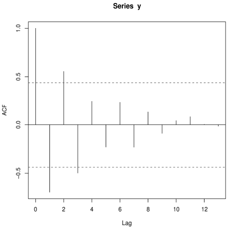
For example, if the model is $X_t = \phi_1 X_{t-1} + Z_t$, then $n_p = 1$.
e.g. if the model is $X_t = \phi_1 X_{t-1} + \phi_3 X_{t-3} + Z_t$, then $n_p = 2$.
e.g. if the model is $X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-3} + \theta_1 Z_{t-1} + Z_t$, then $n_p = 3$ (δ will not be counted).

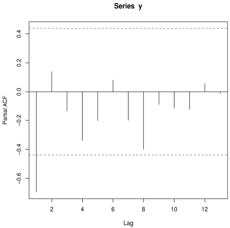
Example 1

Example

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

Its ACF and PACF plots





R codes

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))

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


Coefficients:
          ar1  intercept
      -0.8335    -0.0074
s.e.   0.1311     0.0185

sigma^2 estimated as 0.02201:  log likelihood = 9.19,
 aic = -12.38
```

The fitted model

The fitted model is

$$\hat{X}_t = -0.01357 - 0.8335 X_{t-1},$$

where $-0.01357 = -0.0074(1 + 0.8335)$.

Use arima procedure to plot acf and pacf

Using R code resid(fit) or fit$residuals produces the following residuals $e_t : 2, 3, ..., 20$: 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 fitted model

According to the output the fitted model is

$$\hat{X}_t = -0.00351 - 0.70117 X_{t-1}.$$

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------Autocorrelations--------					
6	5.29	5	0.3819	0.091	-0.012	-0.337	-0.275	-0.028	0.041
12	12.11	11	0.3556	-0.169	-0.045	0.009	0.126	0.300	0.112
18	21.13	17	0.2203	-0.110	-0.220	-0.119	0.091	0.112	0.057

SACF of residuals

The SACF for e_t are

$$r_1 = 0.091, r_2 = -0.012 r_3 = -0.337,$$

$$r_4 = -0.275, r_5 = -0.028, r_6 = 0.041,$$

$$r_7 = -0.169, r_8 = -0.045, r_9 = 0.009, r_{10} = 0.126.$$

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

Using the Ljung-Box test

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

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

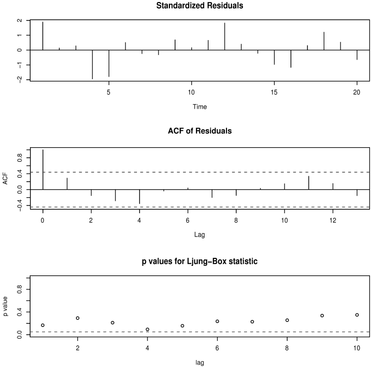
Since $\chi^2_{0.05}(5-1) = 9.49$, $Q(5) < \chi^2_{0.05}(5-1)$. Thus we can not reject the adequacy of the model by setting α equal to 0.05.

Using p values to make a conclusion

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.

Residual plots

Using tsdiag(fit) yields the following plot.



Using ACF and PACF of of residuals to improve the model(I)

Example

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.

Using ACF and PACF of of residuals to improve the model(II)

Example

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.

Types of Nonstationarity

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 1

Example

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 2

Example

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 $U_t = \log(X_t)$. Then $U_t = t + Y_t$, which is not stationary because $EU_t = t + EY_1$. We may apply the first difference operator $\bigtriangledown$ to $\{U_t\}$ to get a stationary process as follows

$$\bigtriangledown U_t = U_t - U_{t-1} = 1 + Y_t - Y_{t-1}.$$

Messages from Example 2

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.

Box-Cox transformation

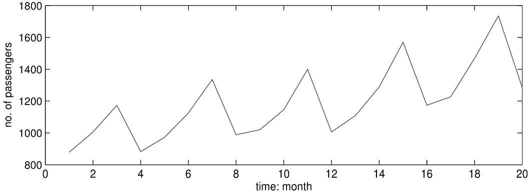
We first consider the power transformation to remove some possible non-stationarity in variance.

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

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)

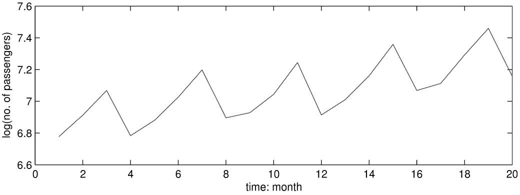
When to perform the power transformation ?

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.



How to perform the power transformation ?

Find the one with minimum sample variance.



Criteria of checking non stationarity

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

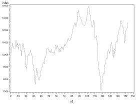
Criteria: The ACFs of non-stationary time series will neither cut off nor die down quickly, but rather will die down extremely slowly, however, those of stationary time series converges to zero **fast**.

Example 3

Example

Figure 2 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 2 and 3 below.

Time plot of data



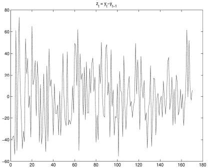


Figure: 2

ACF and PACF of data

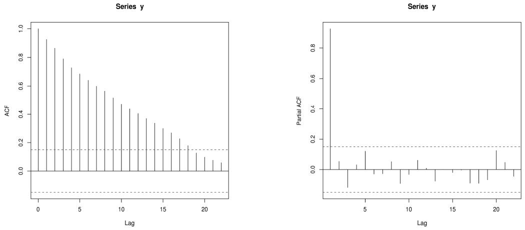


Figure: 3

The differenced data

Let $x_t (= z_t) = Y_t - Y_{t-1}$. Figure 4 suggests that z_t is stationary.

ACF and PACF of the differenced data

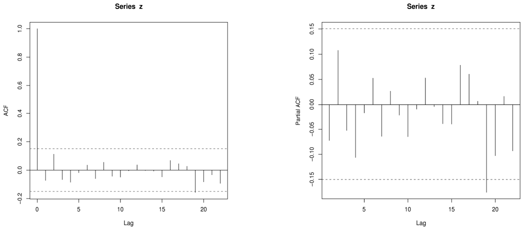


Figure: 4

Fit a model

Moreover, by figure 4 we can fit the following models to the data

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

or

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

which implies that we suggest an ARIMA(0,1,19) for Y_t.

General cases (I)

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.

General cases(II)

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

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

R codes for MA(19)

We may try ARIMA(0,1,19) by looking at SPACF and SACF.
fitma = arima(y, order = c(0, 1, 19))

Output

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

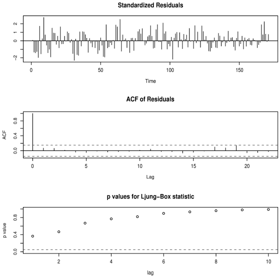
Coefficients:

	ma1	ma2	ma3	ma4	ma5	ma6	ma7	
	-0.2219	0.0249	-0.1093	-0.0702	0.0666	0.0518	0.0191	0.
s.e.	0.0950	0.0832	0.0863	0.0898	0.0908	0.0867	0.0899	0.0
	ma9	ma10	ma11	ma12	ma13	ma14	ma15	m
	-0.0537	-0.0714	-0.0685	0.1185	0.0134	0.0531	-0.0942	0.0
s.e.	0.1016	0.1087	0.0896	0.0893	0.0987	0.1004	0.1145	0.
	ma17	ma18	ma19					
	-0.1438	-0.1234	-0.4681					
s.e.	0.0972	0.1059	0.0971					

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

Diagnostic checking

tsdiag(fitma)



Predictions

predict(fitma, n.ahead= 20)

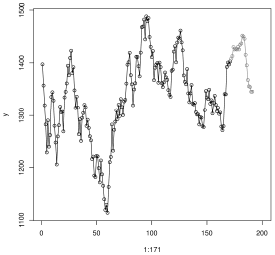


Figure: the black dot is the observation; the red dots are the predictions

R codes

We may try ARIMA(19,1,0) by looking at SPACF and SACF.
fitma1 = arima(x = y, order = c(19, 1, 0))

Output for AR(19)

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

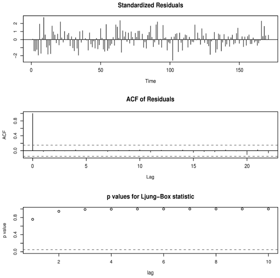
Coefficients:

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

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

Diagnostic checking

tsdiag(fitma)



Predictions

predict(fitma, n.ahead= 20)

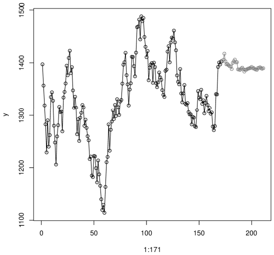


Figure: the black dot is the observation; the red dots are the predictions

The fitted model

The fitted model is MA(19) according to AICs.

Example 4

Example

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

R codes for producing time plot

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

Time plot of Example 4

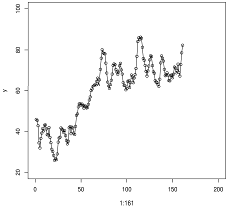


Figure: series y

R codes for plotting ACF and PACF of the data

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

ACF and PACF of the data

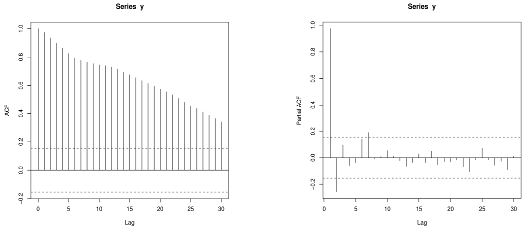


Figure: series y

ACF and PACF of the differenced data

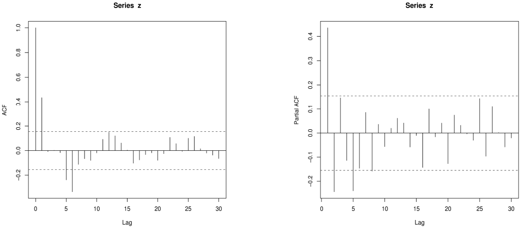


Figure: series x

An appropriate model(I)

The raw data is not stationary according its ACF. We take difference

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

We can use ARIMA(0, 0, 6) for x_t, or ARIMA(0,1,6) for y_t. The R output is follows.

Diagnostic checking

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

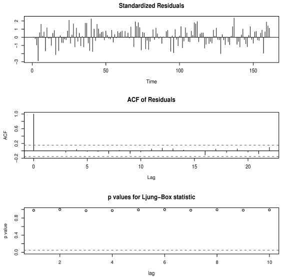


Figure: ACF of Residuals

R output

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	m
	0.6331	-0.0160	0.0361	-0.0264	-0.1490	-0.43
s.e.	0.0771	0.0892	0.0917	0.0879	0.1055	0.07

σ^2 estimated as 4.896: log likelihood = -356.33, aic = 726.65

An appropriate model(II)

The fitted model is

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

Fitted values

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")
```

Forcase

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")
```

How good is the model?

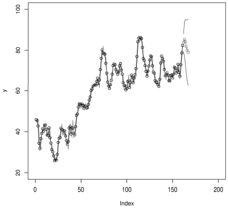


Figure: the black dots are the observation and the red dots are the predictions