Chapter 6 Model Identification

Suppose we have real data: $y_1, y_2, \cdots, y_n$. How to identify the model?

6.1 Steps for model identification

Step 1. Plot data y_t and choose proper transformations.

some common transformation:

$$Z_t = \ln y_t\,, \quad \text{or} \quad Z_t = \sqrt{y_t}\,, \quad \text{or} \quad Z_t = \frac{y_t^\lambda - 1}{\lambda}.$$

Denote $Z_t = f(y_t)$, where f is the transformation function.

Step 2. Compute and examine the sample ACF of Z_t.

Observe whether or not Z_t is stationary or nonstationary.

If it is stationary, then go to **Step 4**. Otherwise go to **Step 3**.

Step 3. Differenced data.

Let $x_t = Z_t - Z_{t-1}$ [i.e. $d = 1, x_t = (1 - B)Z_t$]. Now, we have data x_t, and go back

Step 2. If x_t is stationary, then go to **Step 4**, otherwise:

Let $x_t^{'} = x_t - x_{t-1}$ [i.e. $d = 2, x_t^{'} = (1 - B)^2 Z_t$]. Now, we have data $x_t^{'}$, and go back **Step 2**. If $x_t^{'}$ is stationary, then go to **Step 4**, otherwise:

Let $x_t^{''} = x_t^{'} - x_{t-1}^{'}$ [i.e. $d = 3, x_t^{''} = (1 - B)^3 Z_t$]. Now, we have data $x_t^{''}$, and go back **Step 2**. If x_t is stationary, then go to **Step 4**, otherwise:

$\cdots$.

Usually, when $d = 1$ or 2, we can obtain a stationary data.

Step 4. Identify p and q in **ARMA**(p, q) model.

Now, we have stationary data x_t. Model is:

$\phi_p(B)x_t = \theta_q(B)a_t\,, \quad$ or $\quad \phi_p(B)(x_t - \mu) = \theta_q(B)a_t.$

The criteria for identifying p and q:

	ACF	PACF
AR(p)	Tails off as exp. decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exp. decay or damped sine wave

ARMA(p,q) ???

For the stationary **ARMA**(p, q) model, there are three important things:

1. Parameter estimation.

2. Diagnostic checking.

3. Model selection.

We will study these in other chapters .

Step 5. Final Model

1. When $x_t = Z_t$ (i.e. $d = 0$), model is

$$\phi_p(B)[f(y_t)] = \theta_q(B)a_t$$

or

$$\phi_p(B)[f(y_t) - \mu] = \theta_q(B)a_t$$

2. When $x_t = (1 - B)^d Z_t$, model is:

$$\phi_p(B)(1 - B)^d[f(y_t)] = \theta_q(B)a_t$$

or

$$\phi_p(B)[(1 - B)^d f(y_t) - \mu] = \theta_q(B)a_t$$

6.2 Empirical Examples

Example 6.1. Series W1.

Bun (1976, p.134).

daily average number of defects per truck found in inspection at the end of the assembly line of a truck manufacturing plant between 04/11-10/01 (45 observations)

Model: $(1-\phi B)(Z_t-\mu)=a_t$.

Example 6.2. Series W2.

Yule (1927), Bartlett (1950), Whittle (1954), Brillinger and Rosenblatt (1967), $\cdots$.

This data set is the classic series of the Wolf yearly sunspot numbers from 1700-1955. Scientists believe that the sunspot numbers affect the weather of Earth and hence human activities such as agriculture, telecommunications and others.

Model:

$$(1-\phi_1 B-\phi_2 B^2)(\sqrt{Z_t}-\mu)=a_t.$$

Box and Jenkins (1976):

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(\sqrt{Z_t}-\mu)=a_t.$$
$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_9 B^9)(\sqrt{Z_t}-\mu)=a_t.$$

Example 6.4. Series W4.

Data: the U.S. monthly series of unemployed females between ages 16 and 19 from Jan. 1961 Dec. 1985.

Model: $(1-B)(Z_t-\mu)=(1-\theta B)a_t$.

Example 6.5. Series W5.

Data: yearly accidental death rate (per 100,000 population) of Pennsylvania (1950-1984).

Model: $(1-B)Z_t=\theta_0+a_t$ or $(1-\phi B)(Z_t-\mu)=a_t$.

Example 6.6. Series W6.

Yearly U.S. tobacco production (1981-1984) published in the 1985 Agricultural Statistics by the United States Department of Revenue.

Model: $(1-B)\ln Z_t=(1-\theta_1 B)a_t$.

Example 6.7. Series W7.

Yearly number of lynx pelts sold by the Hudson's Bay Company in Canada between 1857 and 1911.

Model:

1. More (1953): $(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)(\ln Z_t-\mu) = a_t$.

2. Nicholls and Quin (1982): Random coefficient AR(2) model. (rank 1).

3. Subba-Rao and Gabr (1984): bilinear model. (rank 1)

4. Davis and Blockwell (1986): AR(7) model. (rank 3)

5. Tong (1990): STAR(2; 7, 2) model. (rank 2)