Statistika, Vol. 6 No. 1, 25 – 35
Mei 2006

Identification of Time Series Model: An Application Part

WAN MUHAMAD AMIR BIN W AHMAD[1], NORHAYATI ROSLI[2], NORIZAN MOHAMED[3], ZALILA BINTI ALI[4]

[1,2,3]Jabatan Matematik, Fakulti Sains dan Teknologi, Malaysia, Kolej Universiti Sains dan Teknologi Malaysia (KUSTEM), 21030 Kuala Terengganu, Terenggani Malaysia,
[4]Pusat Pengajian Sains Matematik, Universiti Sains Malaysia, (USM), 11800 Pulau Pinang.

ABSTRACT

Time series analysis generally referred to any analysis which involved to a time series data. In this analysis, any of the continuous observation is commonly dependent. If the continuous observation is dependable, then the values that will come are able to be forecasted from the previous observation (Weir 2006). If the behaviour of coming time series are able to be exactly forecasted based on previous times series, so it's called deterministic time series. The objective of times series can be summarized as to find the statistical model to describe the behaviour of the time series data and afterwards made use of skilled statistical techniques for estimation, forecasting but also the controlling. The use of time series analysis very much spread in various fields like biology, medical and many more that had a purpose for forecasting. In this paper the recognition of concerning the Autoregressive Process model AR (p), Moving Average Process MA (q), Autoregressive Moving Average ARMA (p,q), Autoregressive Integrated Moving Average ARIMA (p,d,q) was given attention through the approach to the Autocorrelation Function ACF and Partial Autocorrelation Function (PACF) theory plot.
Keywords: Autoregressive Process model AR (p), Moving Average Process MA (q), Autoregressive Moving Average ARMA (p,q), Autoregressive Integrated Moving Average ARIMA (p,d,q).

1. Introduction

A time series is an ordered sequence of observations. Although the ordering is usually through time, particularly in terms of some equally spaced time intervals, the ordering may also be taken through other dimensions, such as space. Time series occur in a variety of fields as such in agriculture, biology, medical, business, economics and much more. There are various objectives for studying time series. They include the understanding and description of the generating mechanism, the forecasting of future values, and optimal control of a system. The intrinsic nature of a time series is that its observations are dependent or correlated, and the order of the observations is therefore important. Hence, statistical procedures and techniques that rely on independence assumption are no longer applicable, and different methods are needed. The body of statistical methodology available for analyzing time series analysis. In time series the most crucial step are to identify and built a model based on the available data. These steps require a good understanding of the process. Thus, in model identification, our goal is to match patterns in the sample ACF and, $\hat{\rho}_k$ and sample PACF, $\hat{\phi}_{kk}$ with the known patterns of the ACF, ρ_k and PACF, ϕ_{kk}, for the ARMA models. In this paper we will illustrated the identification ARIMA model through the plot of *ACF* and *PACF* theory.

2. Moving Average And Autoregressive Representations Of Time Series Process

In time series analysis, there are two useful representations to express a time series process. They are Moving Average *(MA)* and Autoregressive *(AR)*. Autoregressive process model of order *p* is denoted as *AR (p)* and it is given by

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t$$

or

$$\phi_p(B)\dot{Z}_t = a_t$$

where $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and $\dot{Z}_t = Z_t - \mu$.

And Moving Average process or model of order q and is denoted as *MA (q)* and it is given by

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

or

$$\dot{Z}_t = \theta(B) a_t$$

where $\theta(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$.

A natural extension of the pure autoregressive and the pure moving average process is the mixed autoregressive moving average process, which includes the autoregressive and moving average as special cases. The general mixed *ARMA (p,q)* process is given by

$$\phi_p(B)\dot{Z}_t = \theta_q(B) a_t$$

where

$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \text{ and } \theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

A homogenous nonstationary time series can be reduced to a stationary times series by taking a proper degree of differencing. The homogenous nonstationary model below has been referred to as the autoregressive integrated moving average model of order *(p,d,q)* and denoted as the ARIMA *(p,d,q).*

3. The Autocovariance, Autocorrelation Functions and Partial Autocorrelation Function

For a stationary process $\{Z_t\}$ we have the mean $E\{Z_t\} = \mu$ and variance $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$ which are constant, and the covariance $Cov(Z_t, Z_s)$ which are functions only of the time difference $|t - s|$. Hence, in this case, we write the covariance between Z_t and Z_{t+k} as

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu)$$

and the correlation between Z_t and Z_{t+k} as

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}$$

where we note $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$. As function of *k*, γ_k is called the autocovariance function and ρ_k is called autocorrelation function *(ACF)*. In time series analysis because they represent the covariance and correlation between Z_t and Z_{t+k} from the same process, separated only by *k* time lags.

In addition to the autocorrelation between Z_t and Z_{t+k}, we may want to investigate the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the intervening variables $Z_{t+1}, Z_{t+2}, \ldots, Z_{t+k-1}$ has been removed. The conditional correlation

$Corr(Z_t, Z_{t+k} \mid Z_{t+1}, \ldots, Z_{t+k-1})$ is usually referred to as partial autocorrelation in time series analysis.

For the autocorrelation function for the residuals, the confidence limits for the ith autocorrelation are

$$upper\ limit = 2\sqrt{1 + 2\sum_{k=0}^{i-1} r_k^2} / \sqrt{n}$$ and the lower = -(upper limit)

where n = the number of observations in the series and r_k = the kth autocorrelation.

For the partial autocorrelation function for the residuals, the upper limit for all partial autocorrelation is $2/\sqrt{n}$ and the lower limit $-2/\sqrt{n}$.

4. Identification of Time Series Model from Acf and Pacf Plot

A large number of procedures has been suggested for identifying which model in the autoregressive-moving average family of models is appropriate for a given time series data. In times series analysis, there are two useful representations to express the suitable model by using ACF and PACF plot. Below is the guide of the identifications of the model.

1. The First Order Autoregressive AR (1) Process

The first-order autoregressive process AR(1), we write

$$(1-\phi_1 B)\dot{Z}_t = a_t$$

or

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t$$

The identification model for *AR(1)* from the view of theory plot is given by Figure 1.1

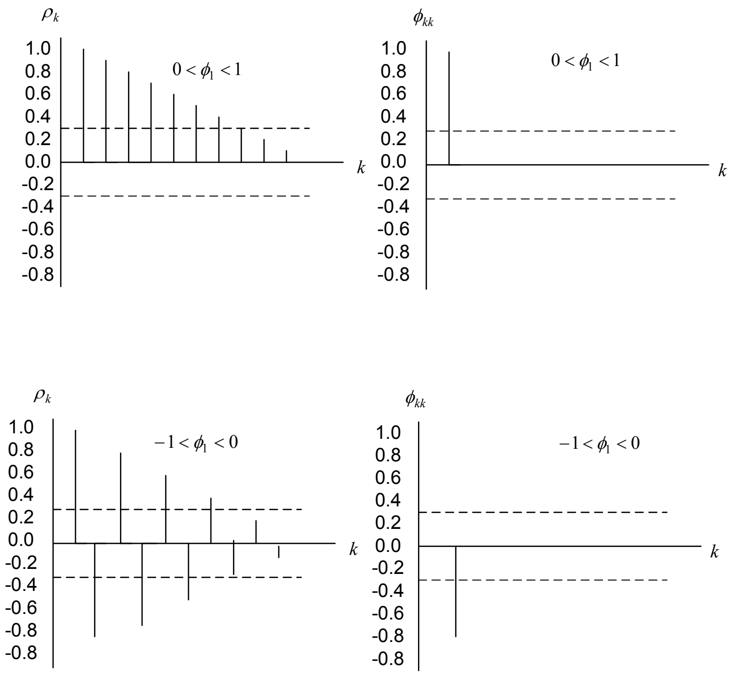

Figure 1.1 ACF and PACF plot for the *AR (1)* process: $(1-\phi_1 B)\dot{Z}_t = a_t$

2. The Second-Order Autoregressive AR (2) Process

For the second-order *AR(2)* process, we have

$$(1-\phi_1 B-\phi_2 B^2)\dot{Z}_t = a_t \text{ or}$$

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + a_t$$

The identification model for *AR (2)* from the view of theory is given by Figure 1.2.

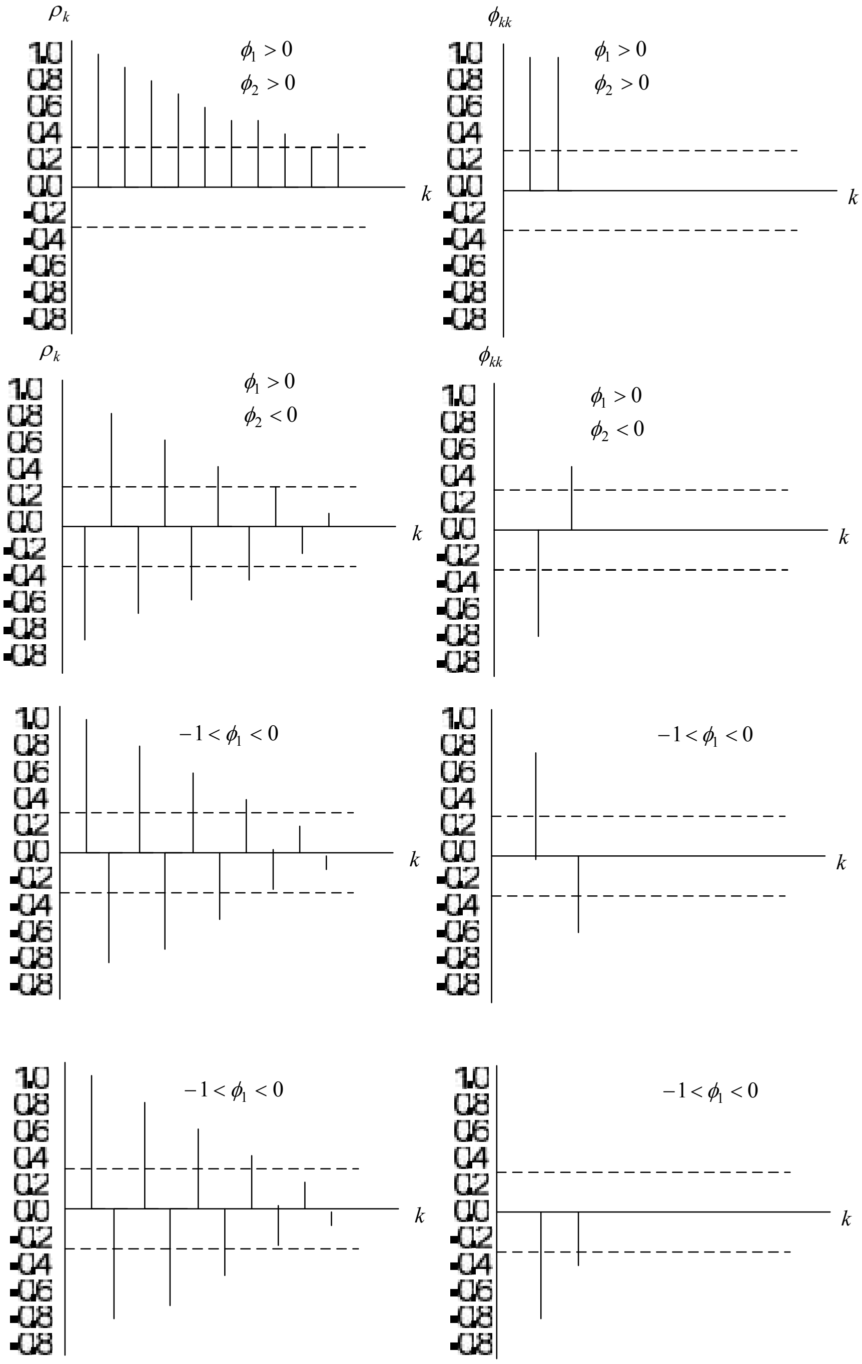

Figure 1.2 ACF and PACF plot for the *AR (2)* process: $(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t$

3. The First Order Moving Average MA (1) Process

When $\theta(B)=(1-\theta_1 B)$, we have the first order moving average *MA (1)* process

$$\dot{Z}=a_t-\theta_1 a_{1-1}$$
$$=(1-\theta_1 B)a_t$$

where $\{a_t\}$ is a zero mean white noise with constant variance σ_a^2. The mean of $\{\dot{Z}_t\}$ is $E(\dot{Z}_t)=0$, and hence $E(Z_t)=\mu$.

The identification model for *MA (1)* from the view of theory plot is given by Figure 1.3.

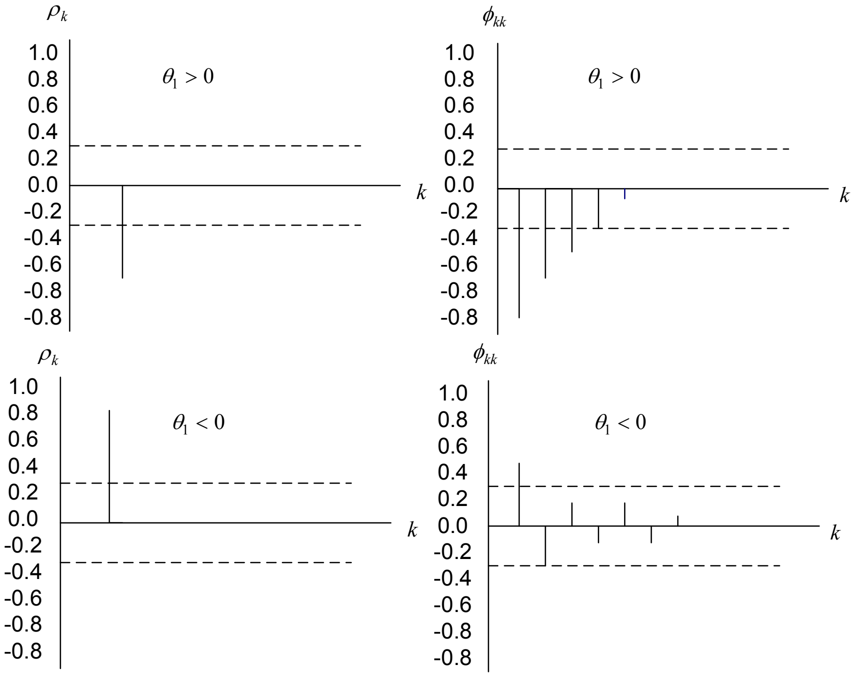

Figure 1.3 ACF and PACF plot for the *MA (1)* process: $\dot{Z}_t=(1-\theta_1 B)a_t$

4. The Second-Order Moving Average MA (2) Process

For the second-order *MA* (2) process, we have

$$Z_t=(1-\theta_1 B-\theta_2 B^2)a_t$$

The identification model for *MA (2)* from the view of theory plot is given by Figure 1.4

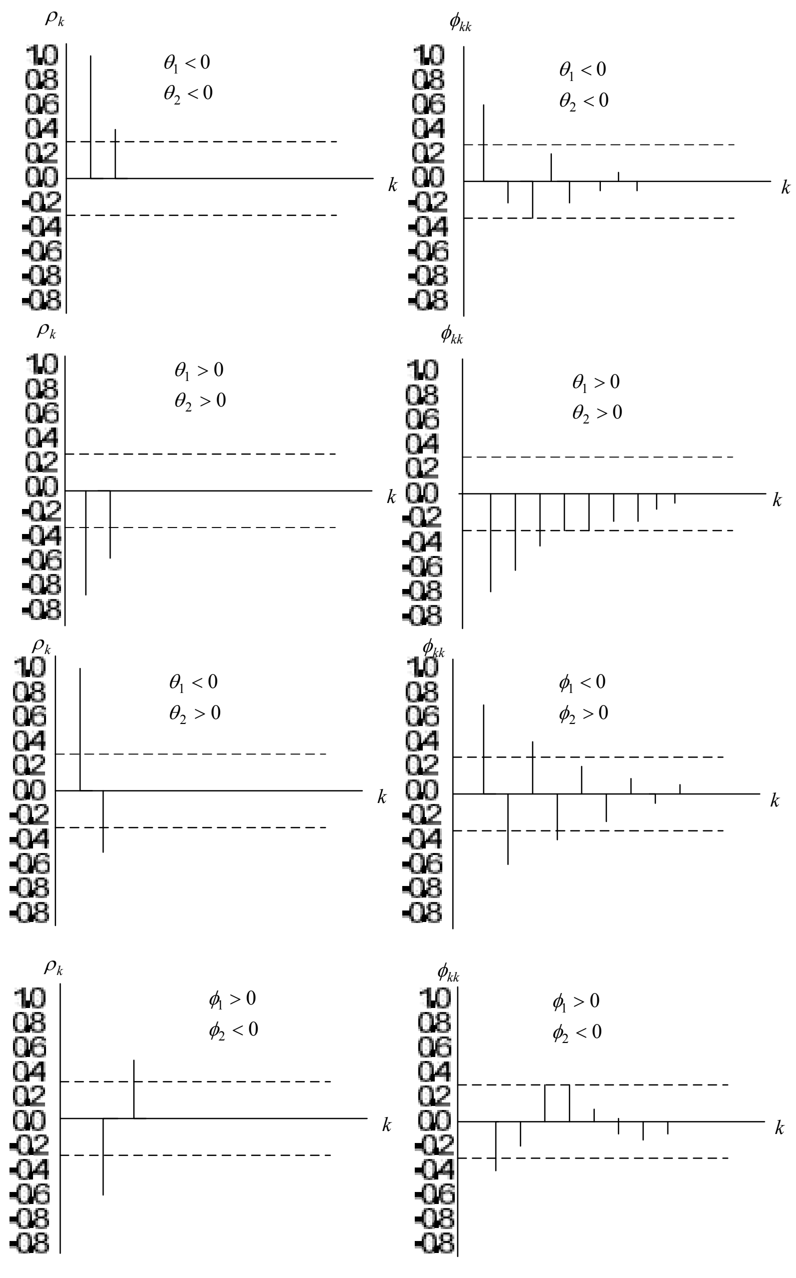

Figure 1.4 ACF and PACF plot for the *MA (2)* process: $Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$

5. Autoregressive Moving Average ARMA (p, q) Processes

A natural extension of the pure autoregressive and the pure moving average processes is the mixed autoregressive moving average process, which includes the autoregressive and moving average processes as a special case. The process contains a large of parsimonious time series model that are useful in describing a wide variety of time encountered in practice. The mixed autoregressive moving average (ARMA) process:

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t$$

where

$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \text{and}\, \theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

The identification plot for the *ARMA (1, 1)* from the view of theory plot is given by Figure 1.5

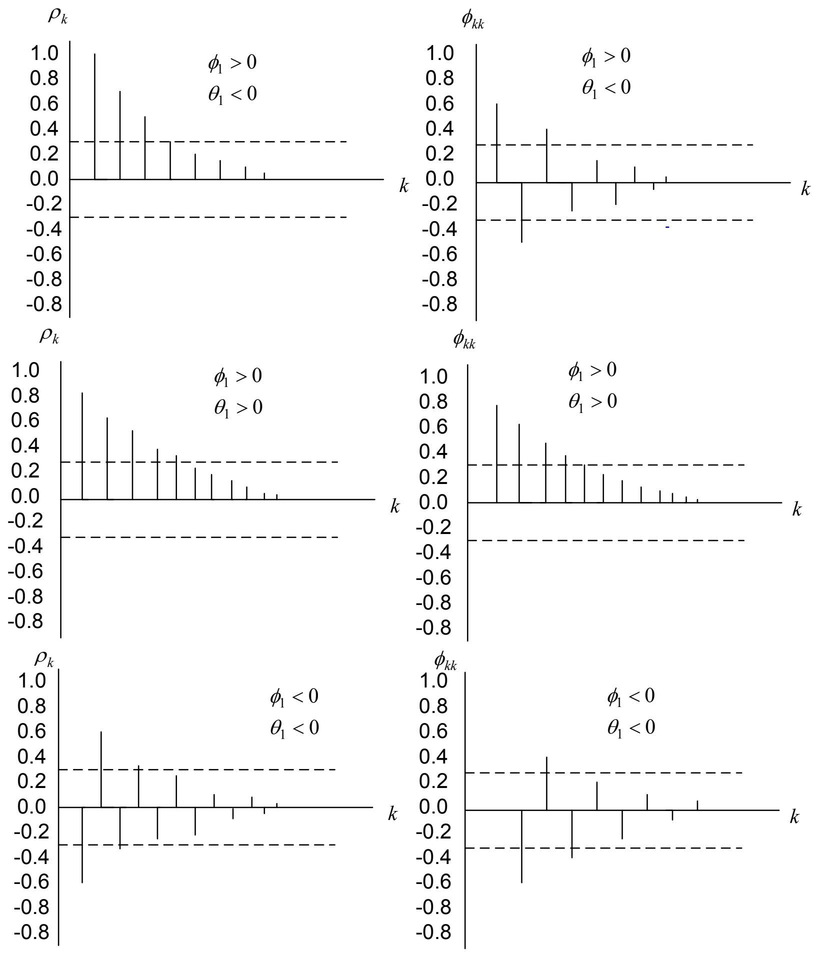

Figure 1.5 ACF and PACF plot for the *ARMA (1, 1)* process: $(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t$

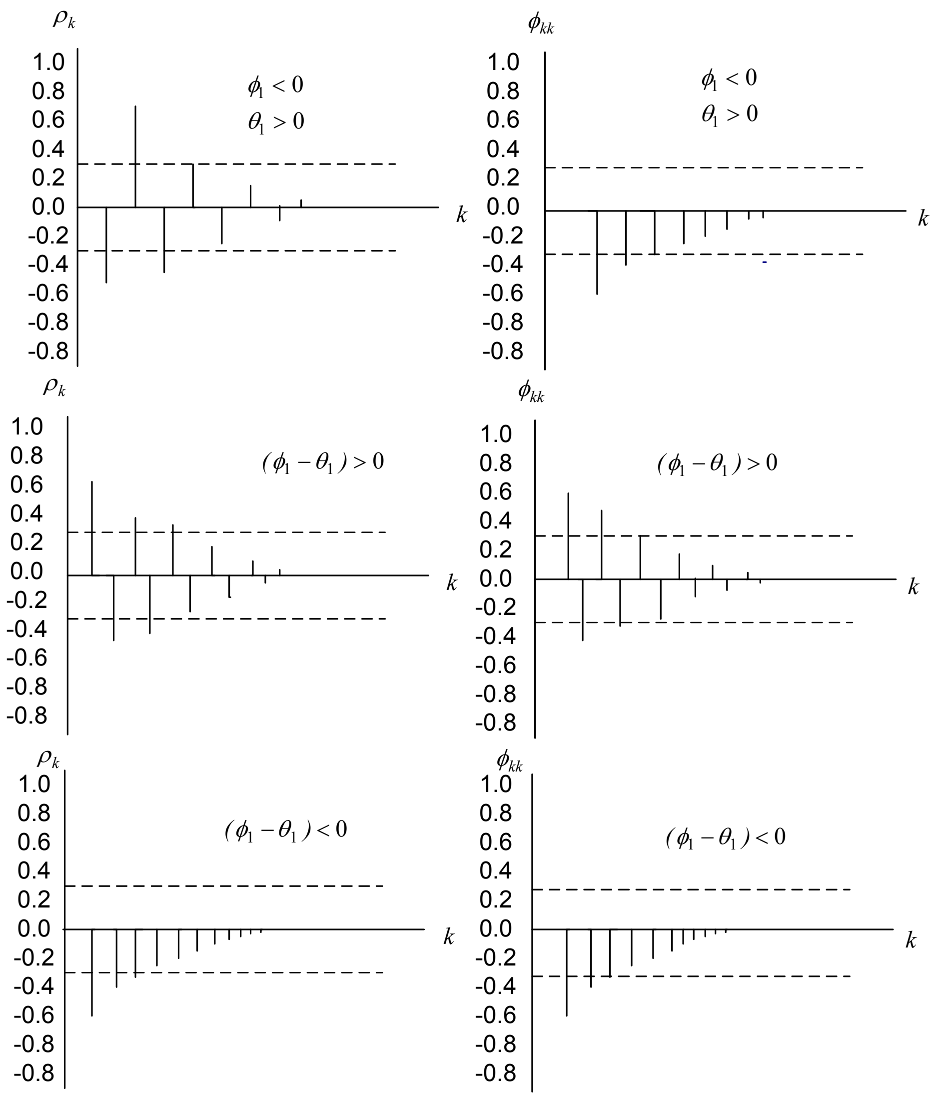

Figure 1.5 (Continued)

6. Autoregressive Integrated Moving Average ARIMA (p, d, q)Processes

A homogeneous nonstationary time series can be reduced to a stationary time series by taking a proper degree of differencing. The autoregressive moving average models are useful in describing stationary times series, so in this section, we discuss the use of differencing to built a large class of time series models, autoregressive integrated moving average models, which are useful in describing various homogeneous nonstationary time series.

Obviously, the stationary process resulting from a properly differenced homogenous nonstationary series is not necessarily white noise as in $(1-B)^d Z_t = a_t$. More generally, the differenced series $(1-B)^d Z_t$ follows the general stationary *ARMA (p, q).*

To illustrate the model identification, we consider the general *ARIMA (p, d, q)* model**;**

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t$$

where the stationary *AR* operator $\phi_p(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$ and the invertible *MA* operator $\theta_q(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$ share no common factors.

The identification plot for the *ARIMA (1, 1,0)* from the view of theory plot is given by Figure 1.6

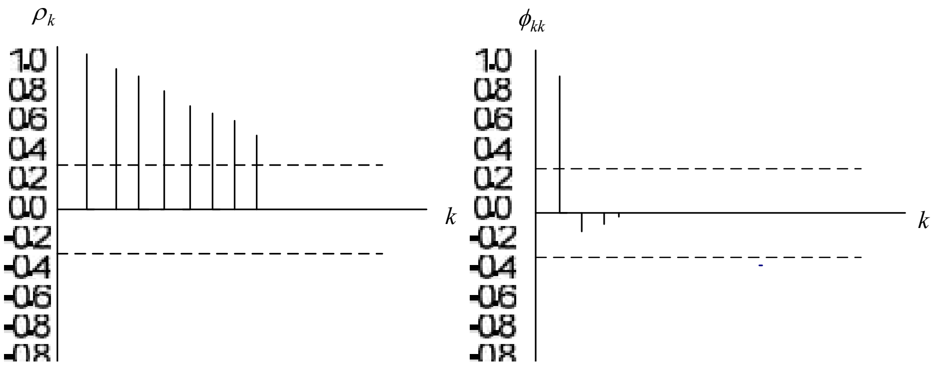

Figure 1.6 ACF and PACF plot for the *ARIMA (1, 1, 0)* process: $(1-\phi_1 B)(1-B)Z_t = a_t$

The identification plot for the *ARIMA (0,1,1)* from the view of theory plot is given by Figure 1.7

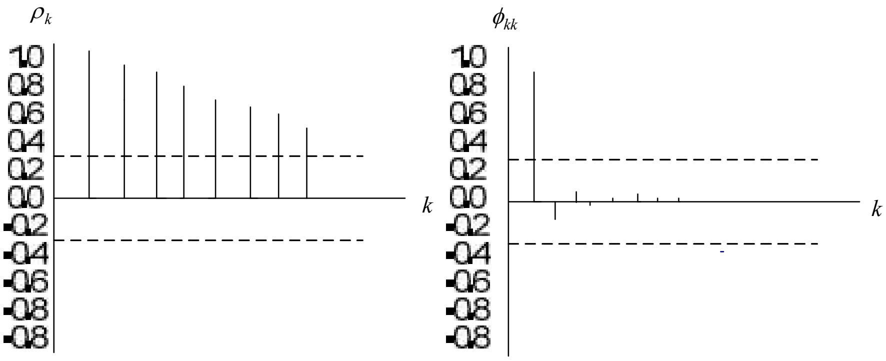

Figure 1.7 ACF and PACF plot for the *ARIMA (0, 1, 1)* process: $(1-B)Z_t = (1-\theta B)a_t$

The identification plot for the *ARIMA (1,1,1)* from the view of theory plot is given by Figure 1.8

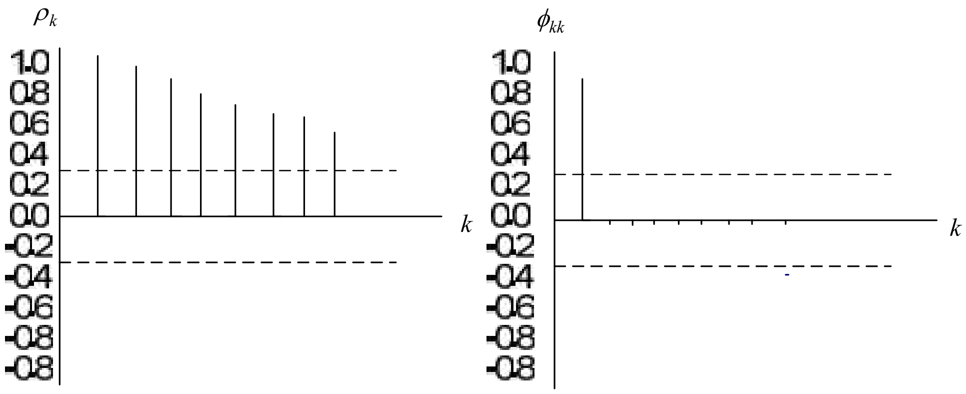

Figure 1.8 ACF and PACF plot for the *ARIMA (1, 1, 1)* process: $(1-\phi_1 B)(1-B)Z_t = (1-\theta B)a_t$

From all the ACF and PACF plots, here we can summarize the results of the model identification as mention below.

Table 1.1 Characteristics of theoretical ACF and PACF for stationary processes.

Process	**ACF**	**PACF**
AR (p)	Tails off exponential decay or damped sine wave	Cuts off after lag *p*
MA (q)	Cuts off after lag *q*	Tails off exponential decay or damped sine wave
ARMA(p,q)	Tails off after lag *(q-p)*	Tails off after lag *(p-q)*

5. Forecasting and Conclusion

One of the most important objectives in the analysis of a time series is to forecast its future values. Event if the final purpose of times series modeling is for the control of a system, its operation is usually based on forecasting. After selection of the appropriate model that we have discussed, we can do the forecasting part which is more useful in biology and other related field. The accurate of the selection model giving us the good results of the prediction. Most forecasting, results, however are derived from a general theory of linear prediction developed by kolmogorov (1939, 1941). At this point, it is appropriate to say that model identification is both a science and art. Through careful examination of the ACF and PACF of the times series, model identification becomes the most interesting aspect of times series analysis. Most of the researchers have more interest in recognizing the model identification before making the prediction especially in biostatistics field.

6. Reference

Abraham, B., Ledolfter, J. (1983) *Statistical Method for Forecasting*, John Wiley, New York.

Ali, M. M. (1977). *Analysis of ARMA models- estimation and prediction,* Biometrika, **64**, 1-11.

Box, G. E. P, Jenkins, G. M., and Reinsel, G. C. (1994). Times Series Analysis, Forecasting and Control.

Bowerman, B. L., O'connel R. T., (1992). *Penelahan Siri Masa, Konsep Satuan Dan Pelaksaan Komputer*, Terjemahan Dewan Bahasa dan Pustaka.