6 NONSEASONAL BOX-JENKINS MODELS

In this section, we will discuss a class of models for describing time series commonly referred to as Box-Jenkins models. There are two types of Box-Jenkins models, seasonal and nonseasonal Box-Jenkins models. Seasonal Box-Jenkins models are used to describe a time series that exhibits seasonal fluctuations but does not contain a trend component. That is, such models are useful in modelling a time series that is nonstationary by reason of seasonal effects only. On the other hand, nonseasonal Box-Jenkins models are used to model stationary time series. This implies that a time series that exhibits either trend, seasonal fluctuations or cyclical fluctuations or a combination of these components, must be transformed into a stationary time series by any one of the methods discussed in previous sections before any of the nonseasonal Box-Jenkins models can be applied to the remainder of the series (detrended and/or deseasonalized series). It is also possible that a series may not exhibit any of these components but the level of the fluctuations about the mean or about a fixed level is not constant, then we must transform the time series to stabilize variance before applying the model.

6.1 Moving Average Representation of Time Series Processes

In the analysis of time series, there are three useful models for representing a time series process Y_t. One representation involves writing the process as a linear combination of a sequence of white noise processes with mean zero. Let $\{a_t\}$ denote a zero mean white noise process with constant variance σ_a^2. Then, a moving average representation (MA) of Y_t is given by

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \ldots = \mu + a_t - \sum_{j=1}^{\infty} \theta_j a_{t-j}, \tag{1}$$

where $1 + \sum_{j=1}^{\infty} \theta_j^2 < \infty$. Using the backwardshift operator $\mathbf{B}$ we can write the moving average representation as

$$Y_t - \mu = \left(1 - \sum_{j=0}^{\infty} \theta_j \mathbf{B}^j\right) a_t.$$

If we let $\widetilde{Y}_t = Y_t - \mu$ be the mean deleted process and

$$\Theta(\mathbf{B}) = 1 - \sum_{j=1}^{\infty} \theta_j \mathbf{B}^j,$$

we can rewrite the moving average representation as

$$\widetilde{Y}_t = \Theta(\mathbf{B}) a_t. \tag{2}$$

Since the sum is infinite, the MA representation is well-defined only if the n partial sums converge to zero. That is,

$$E\left[\left(\widetilde{Y}_t - a_t + \sum_{j=1}^{n}\theta_j a_{t-j}\right)^2\right] \longrightarrow 0, \quad \text{as} \quad n \longrightarrow \infty.$$

This implies that the larger the value of n the more accurate the MA representation becomes. It is quite straightforward to show, from (1) or (2) that the first two moments of Y_t are

$$E(Y_t) = \mu, \quad Var(Y_t) = \sigma_a^2 \sum_{j=0}^{\infty}\theta_j^2,$$

and for $\theta_0 = -1$, the autocovariance function $c(k)$ is

$$\begin{aligned} c(k) &= Cov(Y_t, Y_{t+k}) = E\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\theta_i\theta_j a_{t-i}a_{t+k-j}\right) \\ &= \sigma_a^2\left(\sum_{i=1}^{\infty}\theta_i\theta_{i+k} - \theta_k\right). \end{aligned}$$

It follows that the autocorrelation function is

$$\rho(k) = \frac{c(k)}{c(0)} == \frac{\sigma_a^2\sum_{i=0}^{\infty}\theta_i\theta_{i+k}}{\sigma_a^2\sum_{j=0}^{\infty}\theta_j^2}.$$

We can see that the mean and variance of the MA process Y_t does not depend on time and the autocovariance $c(k)$ and autocorrelation $\rho(k)$ functions are functions of the lag k only. This implies that any MA process is finite provided $c(k)$, $k = 0, 1, \ldots$ is finite. Using the Cauchy-Schwarz inequality we can write

$$\begin{aligned} |c(k)| = |E(Y_t - \mu)(Y_{t+k} - \mu)| &\leq \left[E(Y_t - \mu)^2\right]^{1/2}\left[E(Y_{t+k} - \mu)^2\right]^{1/2} \\ &= [Var(Y_t)]^{1/2}[Var(Y_{t+k})]^{1/2} = Var(Y_t) = \sigma_a^2\sum_{j=0}^{\infty}\theta_j^2. \end{aligned}$$

Therefore a condition for the process (1) to be stationary is that

$$\sum_{j=0}^{\infty}\theta_j^2 < \infty.$$

That is, any process that can be written as a finite or convergent moving average series is a stationary process.

6.2 Autoregressive Representation of Time Series Processes

Another representation that is commonly used is to write the process Y_t as a linear combination of its past values. That is, we write

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \phi_3 \widetilde{Y}_{t-3} + \cdots + a_t = \sum_{j=1}^{\infty} \phi_j \widetilde{Y}_{t-j} + a_t. \tag{3}$$

This representation is called the autoregressive representation (AR) of the process Y_t. If we define

$$\Phi(\mathbf{B}) = 1 - \sum_{j=1}^{\infty} \phi_j \mathbf{B}^j,$$

then we can write the AR representation

$$\widetilde{Y}_t - \sum_{j=1}^{\infty} \phi_j \mathbf{B}^j \widetilde{Y}_t = a_t,$$

as

$$\Phi(\mathbf{B})\widetilde{Y}_t = a_t, \tag{4}$$

where $1 + \sum_{j=1}^{\infty} |\phi_j| < \infty$. Similar to the MA process, any process Y_t that can be written in an AR representation is said to be an invertible process. For the AR process, the condition $\sum_{j=1}^{\infty} |\phi_j| < \infty$ is referred to as the invertibility condition.

It is not too difficult to show that, if Y_t is stationary,

$$E(\widetilde{Y}_t) = 0.$$

That is, $E(Y_t) = \mu$ and

$$c(k) = E(\widetilde{Y}_t \widetilde{Y}_{t+k}) = \begin{cases} \sum_{j=1}^{\infty} \phi_j c(|k-j|) + \sigma_a^2 & \text{if } k = 0 \\ \\ \sum_{j=1}^{\infty} \phi_j c(|k-j|) & \text{if } k \geq 1. \end{cases}$$

It follows that, provided that $\rho(k) < \infty$, the ACF of an AR process is

$$\rho(k) = \frac{c(k)}{c(0)} = \sum_{j=1}^{\infty} \phi_j \rho(|k-j|) \text{ if } k \geq 1.$$

It is clear that the AR representation will be very useful in forecasting future values of a given series since the forecast will depend on past values of the series.

6.3 Duality of AR and MA Processes

We note that an invertible AR process Y_t may not be stationary. To show that it is stationary we have to be able to rewrite the AR process as a convergent MA series. In the same way, a stationary MA process may not be invertible. For it to be invertible, we need to be able to rewrite the MA process as a convergent AR series. The ability to write a MA model as AR and vice versa is referred to as the *duality* of AR and MA processes. The importance of requiring any MA or AR model to be stationary and invertible is the fact that in modelling a real time series, there may be several competing models that may appear suitable for the series. The properties of stationarity and invertibility are then used as one of the criteria for selecting the best model to ensure that the model is well-defined and the forecasts will be finite. To illustrate this point, consider the two moving average models given by

$$\begin{aligned} \widetilde{Y}_t &= a_t - 0.4a_{t-1} = (1 - 0.4\boldsymbol{B})a_t \\ \widetilde{Y}_t &= a_t - 2.5a_{t-1} = (1 - 2.5\boldsymbol{B})a_t. \end{aligned}$$

It is easy to show that $E(Y_t) = \mu$,

$$\text{Var}(Y_t) = \begin{cases} 1.16\sigma_a^2, & \text{if } \widetilde{Y}_t = a_t - 0.4a_{t-1}, \\ 7.25\sigma_a^2, & \text{if } \widetilde{Y}_t = a_t - 2.5a_{t-1}, \end{cases}$$

and

$$\rho(k) = \begin{cases} -\frac{1}{2.9}, & \text{if } k = 1, \\ 0, & \text{if } k > 1. \end{cases}$$

So, the two models used to represent Y_t have the same mean and autocorrelation functions but different variances. The question then is, which model should be used? To answer this question let us consider the properties of stationarity and invertibility for the two models. Clearly, the two models are stationary since for Model 1, $\sum_{j=1}^{\infty} \theta_j^2 = 0.4^2 < \infty$ and for Model 2, $\sum_{j=1}^{\infty} \theta_j^2 = 2.5^2 < \infty$. Now, for invertibility, we have to rewrite both models in an AR representation. For Model 1, we write

$$\widetilde{Y}_t = a_t - 0.4a_{t-1} = (1 - 0.4\boldsymbol{B})a_t.$$

So, to write Model 1 in terms of Y_t's we invert the model in the form

$$(1 - 0.4\boldsymbol{B})^{-1}\widetilde{Y}_t = a_t.$$

Using the formula for Binomial expansion we obtain the AR representation of Model 1 to be

$$(1 + 0.4\boldsymbol{B} + 0.4^2\boldsymbol{B}^2 + 0.4^3\boldsymbol{B}^3 + \cdots)\widetilde{Y}_t = a_t.$$

Clearly, this representation will converge since $0.4 < 1$. Thus Model 1 is invertible. Similarly, we can write Model 2 in the form

$$(1 + 2.5\boldsymbol{B} + 2.5^2\boldsymbol{B}^2 + 2.5^3\boldsymbol{B}^3 + \cdots)\widetilde{Y}_t = a_t.$$

This representation will however diverge since $2.5 > 1$. Therefore Model 2 has no AR representation and hence not invertible.

In general, we observe that if the coefficient of $\boldsymbol{B}$, say $|\theta_1| < 1$ then the MA model will be invertible. Equivalently, if the root of the equation $1 - \theta_1\boldsymbol{B} = 0$ is larger than 1, then the MA model will be invertible. Extending this idea to the general MA representation, for an MA process with the representation (1) or (2) to be invertible, we should be able to rewrite (2) as

$$a_t = \frac{1}{\Theta(\mathbf{B})}\widetilde{Y}_t = \Phi(\mathbf{B})\widetilde{Y}_t.$$

It can be shown that, this is possible if the roots of the polynomial equation $\Theta(\mathbf{B}) = 0$ are all larger than 1 in absolute value or modulus. Thus, if any of the root is complex then the condition for invertibility is that the Euclidean distance from the origin to the root should be larger than 1. That is, if the root is $a+ib$, then the condition is that $|a+ib| = \sqrt{a^2 + b^2} > 1$.

On the other hand, the AR representation (3) or (4) is said to be stationary if it can be written as a linear combination of a white noise process $\{a_t\}$ by inversion

$$\widetilde{Y}_t = \frac{1}{\Phi(\mathbf{B})}\widetilde{Y}_t = \Theta(\mathbf{B})a_t,$$

such that the stationarity condition $\sigma_a^2 \sum_{j=0}^{\infty} \theta_j^2 < \infty$ is satisfied. It can be shown that this condition is also equivalent to saying that the roots of the polynomial equation $\Phi(\mathbf{B}) = 0$ lie outside the unit circle. That is the absolute value or modulus of the roots should be larger than 1. For complex roots, the condition is that $|a + ib| = \sqrt{a^2 + b^2} > 1$.

6.4 Finite Order AR and MA Representations

In the actual representation of a time series process based on a fixed number of observations, we will not be using infinite number of terms in the MA or AR representations because it will be impossible to estimate the infinite number of parameters. Thus, for MA models, we will use, say q terms, as in

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} + \ldots \theta_q a_{t-q} = \mu - \sum_{j=0}^{q} \theta_j a_{t-j}, \quad \theta_0 = -1, \tag{5}$$

and assume that the remainder terms are negligible. We shall refer to this finite order representation as a moving average model of order q, denoted by $MA(q)$. Observe that this process is always stationary, since

$$\sum_{j=0}^{q} \theta_j^2 < \infty.$$

Similarly, for AR models, we will use, say p terms, as in

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \cdots + \phi_p \widetilde{Y}_{t-p} + a_t = \sum_{j=1}^{p} \phi_j \widetilde{Y}_{t-j} + a_t. \tag{6}$$

The AR model with p terms is always invertible because

$$\sum_{j=1}^{p} |\phi_j| < \infty.$$

One issue that we will be discussing later is how to determine the possible values of q or p that is best for a given time series data. Notice that the value of p or q, as the case may be, determines the number of parameters in the model. For a fixed number of observations, it may be possible to represent the underlying process generating the series by several models, each with a different value of p or q. It is often desirable to choose the model with the smallest number of parameters to describe the process because the more the number of parameters, the less efficient is the estimation of the parameters. This is the principle of parsimony in model building recommended by Tukey (1967), ("An introduction to the calculations of numerical spectrum analysis," in *Advanced Seminar on Spectral Analysis* (Ed. B. Harris), 25-46, Wiley, New York) and Box and Jenkins (1976) (*Time Series Analysis Forecasting and Control*, Holden-Day, San Fransisco).

6.5 The Autoregressive Moving Average Representation of Time Series

For some time series data, the AR(p) and MA(q) models may not be suitable for modelling the data. In this case, one may consider a combination of both models commonly refer to as the autoregressive moving average model of order (p, q) and denoted by ARMA(p, q). The ARMA(p, q) model is written as

$$\begin{aligned} \widetilde{Y}_t &= \sum_{j=1}^{\infty} \phi_j \widetilde{Y}_{t-j} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}. \\ &= \sum_{j=1}^{\infty} \phi_j \boldsymbol{B}^j \widetilde{Y}_t + a_t - \sum_{j=1}^{q} \theta_j \boldsymbol{B}^j a_t. \end{aligned} \tag{7}$$

That is, the ARMA(p, q) model can be written as $\Phi_p(\boldsymbol{B})\widetilde{Y}_t = \Theta_q(\boldsymbol{B})a_t$.

6.6 Stationarity

For a process Y_t that can be represented by an ARMA(p, q) model to be stationary we rewrite it as a convergent MA series. That is we invert the model and write

$$\widetilde{Y}_t = \frac{\Theta_q(\boldsymbol{B})}{\Phi_p(\boldsymbol{B})} a_t. \tag{8}$$

Since $\Theta(\boldsymbol{B}$ has a finite number of terms, the condition for stationarity of Y_t is same as that of an AR(p) process.

6.7 Invertibility

Similar to the stationarity condition, any process Y_t that can be represented by an ARMA(p, q) model will be said to be invertible if we can rewrite it as a convergent AR series. So, we invert the model and write

$$\frac{\Phi_p(\boldsymbol{B})}{\Theta_q(\boldsymbol{B})} \widetilde{Y}_t = a_t. \tag{9}$$

Again, we see that since $\Phi(\boldsymbol{B}$ has finite number of terms, the condition for invertibility of Y_t will be the same as that of an MA(q) process.

6.8 Some statistical tools for model identification

Now that we have discussed some of the elementary models we will be using to represent real time series we will discuss some statistics that are useful when choosing suitable models for a given time series. The statistics we shall study are the sample autocorrelation and partial autocorrelation functions.

6.9 The Autocovariance and Autocorrelation Functions

Once we have successfully transformed a nonstationary time series into a stationary time series, we investigate the structure of the relationship between observations in the stationary series k distances apart. This can be done by computing sample versions of the autocorrelation and partial autocorrelation functions of the stationary series which are then used to conduct tests we shall study soon. The Durbin-Watson test we studied earlier is only good enough for detecting first-order autocorrelation.

Let Y_t be a stationary process. Recall that the lag k autocovariance function of Y_t (*i.e.* the covariance between Y_t and Y_{t+k}) is defined by

$$c(k) = Cov(Y_t, Y_{t+k}) = E(Y_t - \mu)(Y_{t+k} - \mu),$$

and the lag k autocorrelation of Y_t is

$$\rho(k) = \frac{Cov(Y_t, Y_{t+k})}{Var(Y_t)} = \frac{c(k)}{c(0)}.$$

We have used the fact that $Var(Y_t) = Var(Y_{t+k}) = c(0)$ due to stationarity of Y_t.

Some properties of $r(k)$ and $\rho(k)$

Earler we noted that for a stationary process Y_t, $c(k)$ and $\rho(k)$ have the following properties.

1. $c(0) = Var(Y_t)$ and $\rho(0) = 1$.

2. $c(k) \leq c(0)$, hence $|\rho(k)| \leq 1$.
   To prove this, we rewrite $c(k)$ as $c(k) = E(Y_t Y_{t+k}) - \mu^2$. The result then follows immediately from the Cauchy-Schwarz inequality $E(Y_t Y_{t+k}) \leq \{E(Y_t^2)\}^{1/2}\{E(Y_{t+k}^2)\}^{1/2} = E(Y_t^2)$ (since Y_t is stationary).

3. $c(k) = c(-k)$ and $\rho(k) = \rho(-k)$.
   That is, $c(k)$ and $\rho(k)$ are even functions. Thus we only need to compute values of $\rho(k)$ for $k > 0$. This result follows from the fact that observations that are k distances apart have the same autocovariance and autocorrelation functions. Here, the time difference between Y_t and Y_{t-k} and Y_t and Y_{t+k} are the same.

In addition, we note that

$c(k)$ and $\rho(k)$ are both positive semidefinite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j c(|t_i - t_j|) \geq 0,$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \rho(|t_i - t_j|) \geq 0,$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$. To prove this, we first define the random variable $X = \sum_{i=1}^{n}\alpha_i Y_{t_i}$ and note that

$$0 \leq Var(X) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j Cov(Y_{t_i}, Y_{t_j}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j c(|t_i - t_j|).$$

We divide this by $c(0)$ to obtain the result for $\rho(k)$.

Now, given a time series $y_1, y_2, \ldots, y_n$ taken from the stationary process Y_t, the autocorrelation function of Y_t at lag k is estimated by

$$r_k = \frac{c_k}{c_0}, \text{ where, } c_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y}), \quad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

It is common to refer to c_k and r_k as the *sample autocovariance function* and the *sample autocorrelation function* (SACF) repectively. One may question why the divisor in the expression for c_k is n instead of $n-k$ since we are summing $n-k$ terms. The point is that c_k with divisor n has some advantages. These advantages include, (i) c_k with divisor n is always positive definite like $c(k)$, whereas the estimator with divisor $n-k$ is not always positive definite; (ii) it can be shown that for certain type of processes, c_k has a smaller mean squared error than the estimator with divisor $n-k$. One disadvantage of using c_k as defined is that it has a larger bias than the estimator with divisor $n-k$, especially when k is large with respect to n. Thus, for a given n, it is recommended to compute values of c_k for $k = 1, 2, \ldots, \frac{n}{4}$.

The autocorrelation structure of some of the models we will be using have the property that their theoretical autocorrelation function $\rho(k) = 0$, after certain lags $k \geq m$. Since r_k values are computed from a sample, r_k may not be exactly zero for $k \geq m$. As a result, we need a procedure that will help us to determine when r_k is not significantly different from zero. It has been shown that for processes in which $\rho(k) = 0$ whenever k is larger than some number, say m, r_k is asymptotically normal, unbiased and have variance

$$Var(r_k) \approx \frac{1}{n}\left\{1 + 2\sum_{j=1}^{m}\rho(j)^2\right\},$$

(see "On the theoretical specification of sampling properties of autocorrelated time series" published 1946 in the *Journal of the Royal Statistical Society*, Series B, Volume 8, pages 27-41, Bartlett, M. S.).
It follows that the large-lag standard error of r_k is

$$S_{r_k} = \sqrt{\frac{1}{n}\left\{1 + 2\sum_{j=1}^{m} r_j^2\right\}}.$$

Thus, to test $H_0: \ \rho(k) = 0,\ k = 1, 2, \ldots, m$, we standardize the sampling distribution of r_k to obtain the test statistic $Z = [r_k - \rho(k)]/S_{r_k}$. When H_0 is true, the statistic becomes $Z = r_k/S_{r_k}$ and S_{r_k} reduces to

$$S_{r_k} = \sqrt{\frac{1}{n}}. \tag{10}$$

Then, we reject H_0 if $|Z| \geq z_{\alpha/2}$ at $(1-\alpha)100\%$ confidence level . After some algebra, we can show that this is equivalent to the condition $|r_k| \geq z_{\alpha/2} S_{r_k}$. In practice, it is common to conduct the test at a 95% confidence level. That means, we reject H_0 if $|r_k| \geq 2S_{r_k}$. Most statistical softwares will automatically produce results at this level of significance.

As an example, we have simulated $n = 100$ observations from MA(1), MA(2) and AR(2) processes and computed and plotted their SACF and the confidence region $\pm 2S_{r_k} \approx 0.258$ (See Figure 18) using the R software. Now, any value of r_k that falls outside the confidence region will be considered significant based on the test procedure. In this example, we see that only r_1 is significant for the MA(1) series and r_1, r_2 are significant for the MA(2) series. The pattern in the SACF for the AR(2) series is similar to that of a damped sine wave. That is, the values of r_k show a decaying pattern while oscillating between positive and negative values. Note that there is a chance that some of the values of r_k may fall outside the interval slightly for some large values of k since we have constructed a 95% region. This does not necessarily mean that these values are significant.

Example: To further illustrate the computation of the SACF, consider the lag 4 difference of the beer production data. The sample mean of the lag 4 difference is $\bar{y} = 1.28$. Thus

$$\begin{aligned} r_1 &= \frac{(0.05-1.28)(0.03-1.28) + (0.03-1.28)(2.8-1.28) + \ldots + (0.07-1.28)(0.37-1.28)}{(0.05-1.28)^2 + (0.03-1.28)^2 + \ldots + (0.37-1.28)^2} \\ &= \frac{4.6958}{108.0466} = 0.04346. \end{aligned}$$

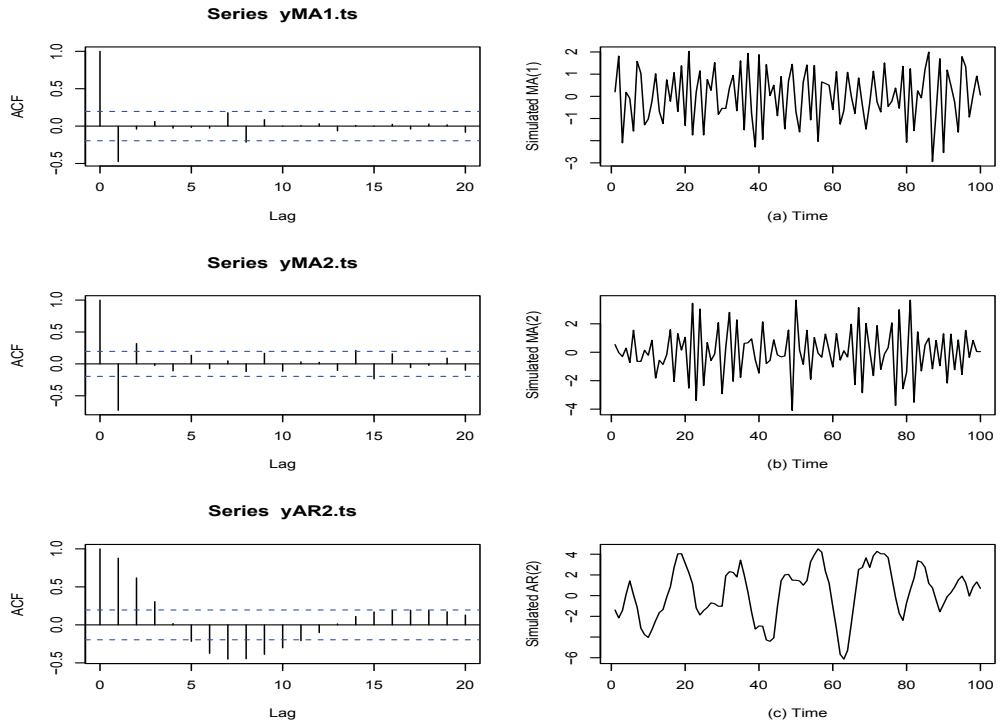

Figure 18: Plot of (a) simulated SACF of MA(1) series with $\theta_1 = 0.9$, (b) simulated SACF of MA(2) series with $\theta_1 = 1$, $\theta_2 = -0.6$ (c) simulated SACF of AR(2) series with $\theta_1 = 1.5$, $\theta_2 = -0.75$.

$$r_2 = \frac{(0.05-1.28)(2.80-1.28)+(0.03-1.28)(1.18-1.28)+\ldots+(-0.91-1.28)(0.37-1.28)}{(0.05-1.28)^2+(0.03-1.28)^2+\ldots+(0.37-1.28)^2}$$
$$= \frac{14.8561}{108.0466} = -0.1375.$$

Similarly, the remaining values of the first $\frac{n}{4} = \frac{28}{4} = 7$ sample autocorrelation function are $r_3 = 0.03963$, $r_4 = -0.4873$, $r_5 = -0.00386$, $r_6 = 0.3161$ and $r_7 = 0.19098$. A graph of the SAC values in a confidence band is shown in Figure 19. We note the following: (i) In Figure 19(a), $r_0 = 1$, r_4, r_8, r_{12} are significant indicating the presence of seasonal effect of period $L = 4$. (ii) In Figure 19(b), $r_0 = 1$, $r_2 < 0, r_4 > 0, r_6 < 0, r_8 > 0, \ldots$. The oscillating pattern is also an indication of the presence of seasonal effects in the first difference of the series. (iii) Only r_4 is significant in the lag 4 differences since it is the only SACF value that fall outside of the band. In general, a good indication that a time series is not stationary is when the values of a SACF continue to fall outside the confidence band and does not decay very quickly. For a stationary series, the SACF values should decay or die down fairly quickly and fall within the confidence band after a few lags. This pattern is shown in the last two

plots representing the lag 4 differences which are stationary series.

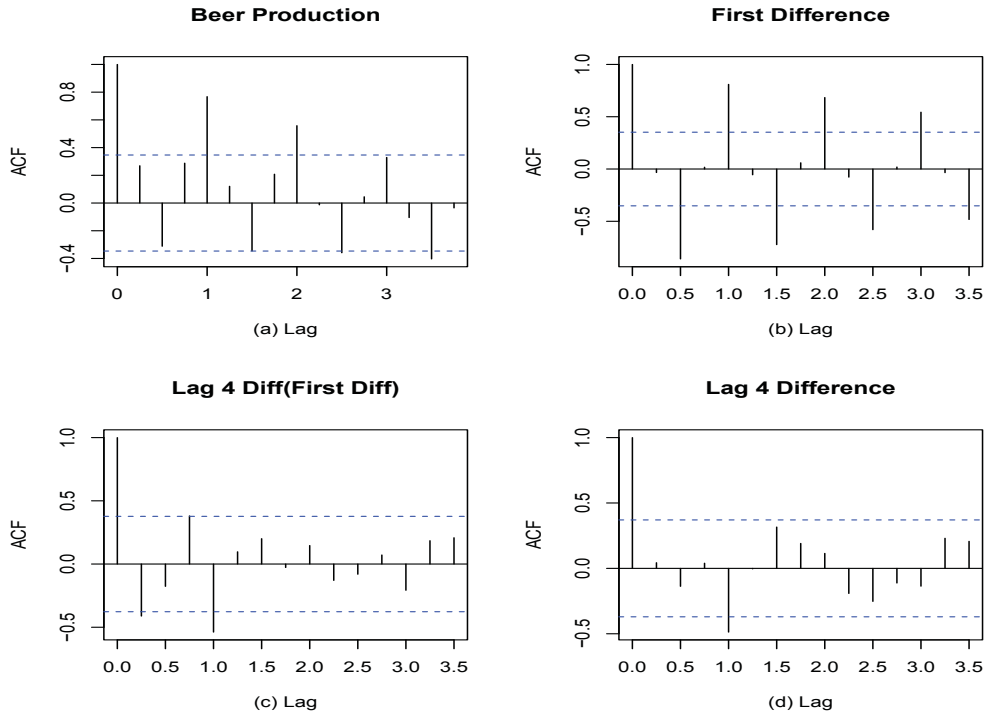

Figure 19: Plot of the sample autocorrelation function for: (a) beer production series (b) first difference of beer series (c) lag 4 difference of the first difference (d) lag 4 difference of the beer series.

To further illustrate some patterns in the SACF of time series, we show the SACF of the male employment series in Figure 20. Once again, we see that the SACF in Figure 20(a), (b) does not die out quickly indicating that the series from which they were computed are non-stationary. We also see that r_{12}, r_{24} and r_{36} are highly significant compared to the others in Figure 20(b) indicating a seasonal pattern of period $L = 12$. Figure 20(c) shows that the series with this SACF is stationary because the SACF die out fairly quickly with the only significant SACF at lag 12 which again indicates a significant 12 month seasonal effect which must be accounted for in any model used to describe the series. The sinusoidal pattern indicated by the SACF in Figure 20(d) simply mirrors the upward and downward swings in the series shown in Figure 17(d).

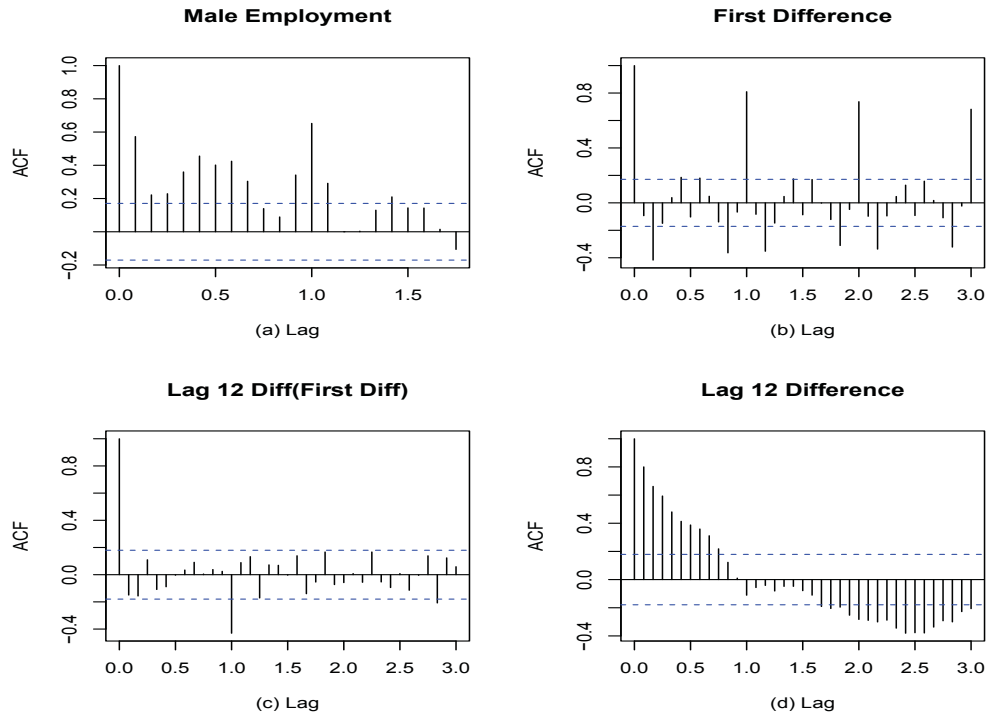

Figure 20: Plot of the sample autocorrelation function for: (a) male employment series (b) first difference of employment series (c) lag 12 difference of the first difference (d) lag 12 difference of the beer series.

6.10 The Partial Autocorrelation Function

Another useful statistic for investigating the correlation structure of a stationary process is called the partial autocorrelation function. The partial autocorrelation function measures the autocorrelation between the variables Y_t and Y_{t+k} after their mutual linear dependency on the intervening variables $Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}$ has been removed. Now letting, ϕ_{kk}, denote the theoretical partial autocorrelation function, this conditional correlation is given by

$$\phi_{kk} = Corr(Y_t, Y_{t+k}|Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k-1}) = \frac{Cov[(Y_t - \hat{Y}_t)(Y_{t+k} - \hat{Y}_{t+k})]}{\sqrt{Var(Y_t - \hat{Y}_t)}\sqrt{Var(Y_{t+k} - \hat{Y}_{t+k})}}.$$

Looking at the expression for $p(k)$, it is clear that it will be very tedious to use this formula to compute the theoretical PACF. An easier approach is by using the expression for the $AR(p)$ model. It turns out that the theoretical PACF at lag k ϕ_{kk} satisfy the $AR(p)$ model given by

$$\widetilde{Y}_t = \phi_{11}\widetilde{Y}_{t-1} + \phi_{22}\widetilde{Y}_{t-2} + \cdots + \phi_{pp}\widetilde{Y}_{t-p} + a_t.$$

Multiply through by $\widetilde{Y}_{t-k}$, take expectation and then divide each term by $c(0)$ to obtain the difference equation in $\rho(k)$

$$\rho(k) = \phi_{11}\rho(k-1) + \phi_{22}\rho(k-2) + \cdots + \phi_{pp}\rho(k-p), \quad k = 1, 2, \ldots.$$

Suppose $p = 1$, then the $AR(1)$ model has only one parameter ϕ_{11} and the difference equation in $\rho(k)$ reduces to

$$\rho(k) = \phi_{11}\rho(k-1), \quad k = 1, 2, \ldots.$$

Then, by using the fact that $\rho(0) = 1$ in the expression for $\rho(k)$, we obtain

$$\phi_{11} = \rho(1). \tag{11}$$

Next, the $AR(2)$ model can be used to obtain ϕ_{22}. When $p = 2$ we have

$$\rho(k) = \phi_{11}\rho(k-1) + \phi_{22}\rho(k-2), \quad k = 1, 2, \ldots.$$

By substituting $k = 1$ and $k = 2$ in the expression for $\rho(k)$ and using the fact the $\rho(k)$ is an even function, we obtain two simultaneous equations given by

$$\begin{aligned} \rho(1) &= \phi_{11}\rho(0) + \phi_{22}\rho(1) \\ \rho(2) &= \phi_{11}\rho(1) + \phi_{22}\rho(0) \end{aligned}$$

which we solve for ϕ_{22} to obtain

$$\phi_{22} = \frac{\begin{vmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(0) \end{vmatrix}} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}. \tag{12}$$

Using the $AR(3)$ model in the same way, we can solve for ϕ_{33} to obtain

$$\phi_{33} = \frac{\begin{vmatrix} \rho(0) & \rho(1) & \rho(1) \\ \rho(1) & \rho(0) & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} \rho(0) & \rho(1) & \rho(2) \\ \rho(1) & \rho(0) & \rho(1) \\ \rho(2) & \rho(1) & \rho(0) \end{vmatrix}}, \tag{13}$$

and so on. It is clear that as the value of k increases, the dimension of the matrices also increases making these computations tedious and difficult. Fortunately, In 1960, Durbin, J. derived a recursive formula for computing the *sample partial autocorrelation function* (SPAC)

denoted by r_{kk} in a paper titled, "The fitting of time series models," in the *Review of the Institute of International Statistics*, Volume 28, pages 233-244. The recursive formula is given by

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2, 3, \ldots, \end{cases}$$

where $r_{kj} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$, for $j = 1, 2, \ldots, k-1$. We can use similar arguments for testing whether the underlying process has zero autocorrelation to show that the large-lag variance of r_{kk} can be approximated by

$$Var(r_{kk}) \approx \frac{1}{n}.$$

Hence, the critical limits of a 95% confidence interval for $\rho(k) = 0$ are approximately $\frac{\pm 2}{\sqrt{n}}$.

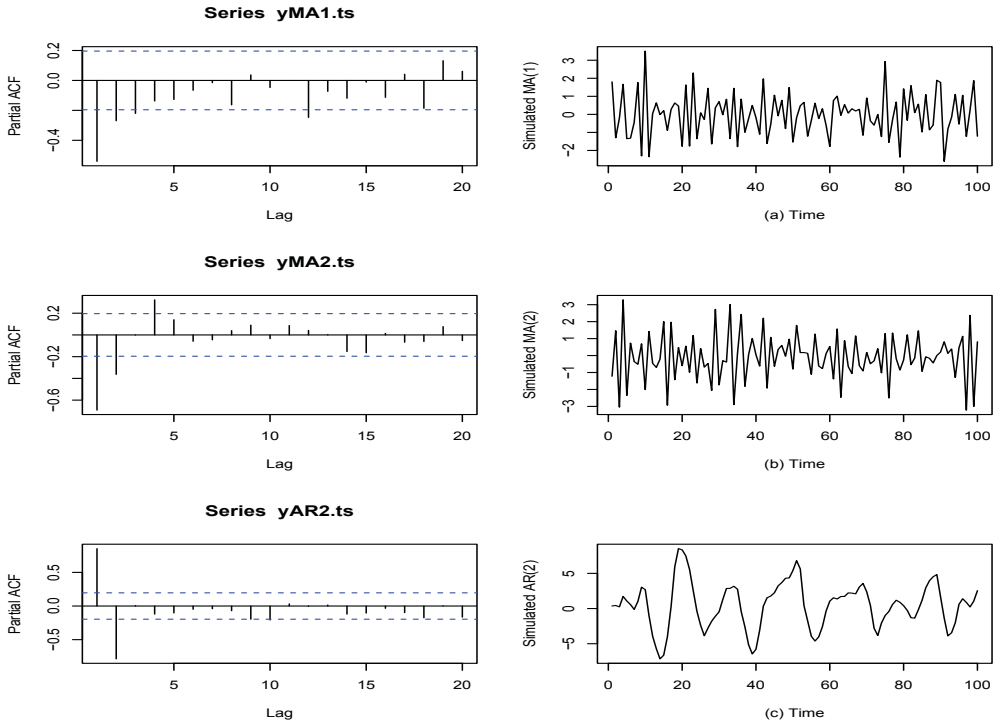

Figure 21: Plot of (a) simulated SPACF of MA(1) series with $\theta_1 = 0.9$, (b) simulated SPACF of MA(2) series with $\theta_1 = 1$, $\theta_2 = -0.6$ (c) simulated SPACF of AR(2) series with $\theta_1 = 1.5$, $\theta_2 = -0.75$.

Again, to illustrate patterns in the SPACF of time series, we have generated 100 observations from the same MA(1), MA(2) and AR(2) models used in illustrating patterns in SACF of

time series. Using the test statistic discussed earlier, we see that most of the SPACF of the MA(1), MA(2) and AR(2) series in Figure 21 are not significant. In particular, for the AR(2) series, only the first 2 values of $\hat{\phi}_{kk}$ are significant.

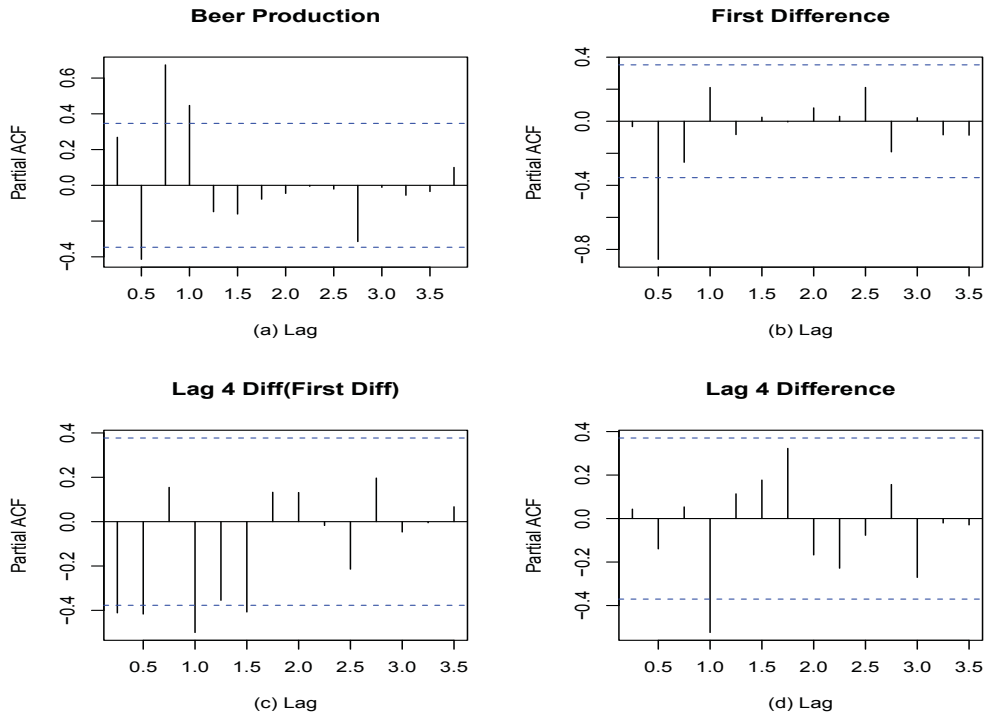

Figure 22: Plot of the sample partial autocorrelation function for: (a) beer production series (b) first difference of beer production series (c) lag 12 difference of the first difference (d) lag 12 difference of the beer series.

Example: To illustrate the computation of the SPACF, consider the lag 4 difference of the beer production data. Now,

$$\begin{aligned} r_{11} &= r_1 = 0.04346. \\ r_{22} &= \frac{r_2 - r_1^2}{1 - r_1^2} \\ &= \frac{-0.1375 - 0.04346^2}{1 - 0.04346^2} = -0.13965 \end{aligned}$$

We observe that the formula for r_{33} requires that we first compute r_{21}. Now,

$$r_{21} = r_{11} - r_{22} \cdot r_{11} = 0.04346 - (-0.13965)(0.04346) = 0.0495293.$$

Then,

$$\begin{aligned} r_{33} &= \frac{r_3 - r_{21}r_2 - r_{22}r_1}{1 - r_{21}r_1 - r_{22}r_2} \\ &= \frac{0.03963 - (0.04952)(-0.1375) - (-0.13965)(0.04346)}{1 - (0.04952)(0.04346) - (-0.13965)(-0.1375)} = 0.05365392. \end{aligned}$$

Other r_{kk} values can be calculated in a similar way. A graph of the SPACF values are shown in Figure 22.

7 Autoregressive Models

Throughout the remainder of this course, we will assume that the time series Y_t is mean deleted. That is, the mean of Y_t is zero. Therefore, we will no longer use the notation $\widetilde{Y}_t$ to denote a mean deleted series but simply Y_t.

Recall that the general $AR(p)$ model given by

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + a_t$$

is always invertible because

$$\sum_{j=1}^{p} |\phi_j| < \infty.$$

Multiplying the model by Y_{t-k} and taking expectation we obtain the autocovariance function of the $AR(p)$ model given by

$$c(k) = \phi_1 c(k-1) + \phi_2 c(k-2) + \cdots + \phi_p c(k-p) + E(Y_{t-k}a_t).$$

Now, it can be shown that

$$E(Y_{t-k}a_t) = \begin{cases} \sigma_a^2, & k = 0, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that when $k > 0$ the autocorrelation function $\rho(k) = c(k)/c(0)$ of the $AR(p)$ model will satisfy the difference equation

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \cdots + \phi_p \rho(k-p),$$

commonly referred to as the Yule-Walker difference equation. We will now consider special cases of the $AR(p)$ model beginning with the simplest case where $p = 1$.

7.1 The First Order Autoregressive Process

The first order autoregressive model, denoted by $AR(1)$ is defined by

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + a_t \\ &= \phi_1(\phi_1 Y_{t-2} + a_{t-1}) + a_t = \cdots. \end{aligned} \tag{14}$$

Clearly, Y_t in (15) is independent of $a_{t+1}, a_{t+2}, \ldots$. That is, Y_t is independent of a_s whenever $s > t$. It follows that $E(a_s Y_t) = E(a_s)E(Y_t) = 0$. Applying this to the $AR(1)$ model (15) we have that $E(a_t Y_{t-1}) = E(a_t)E(Y_{t-1}) = 0$ since $t > (t-1)$. In terms of the backshift operator, the $AR(1)$ model can be written as

$$(1 - \phi_1 \mathbf{B})Y_t = \Phi_1(\mathbf{B})Y_t = a_t.$$

7.1.1 Invertibility

For the $AR(1)$ model

$$\sum_{j=1}^{\infty} |\phi_j| = |\phi_1|.$$

Therefore, any process Y_t that can be written in an AR(1) representation is always invertible since $|\phi_1| < \infty$.

7.1.2 Stationarity

Recall that an AR process is said to be stationary, if the modulus of all the roots of the polynomial equation

$$\phi(\mathbf{B}) = 1 - \phi_1 \mathbf{B} = 0,$$

lie outside the unit circle. Now, the solution to this linear equation in $\mathbf{B}$ is $\mathbf{B} = \frac{1}{\phi_1}$. That is, for the $AR(1)$ process to be stationary, $|1/\phi_1| > 1$ or $|\phi_1| < 1$.

7.1.3 ACF of the $AR(1)$ Process

By definition, the autocovariance function of Y_t is given by

$$c(k) = E(Y_t - \mu)(Y_{t-k} - \mu) = E(Y_t Y_{t-k}) - \mu^2 = E(Y_t Y_{t-k})$$

since $\mu = 0$ by our assumption. Now, replace Y_t by the expression in (15) to obtain

$$c(k) = \phi_1 E(Y_{t-1} Y_{t-k}) + E(a_t Y_{t-k}) = \phi_1 c(k-1) + E(a_t Y_{t-k}).$$

If $k = 0$, $E(a_t Y_{t-k}) = E(a_t Y_t) = \phi_1 E(Y_{t-1} a_t) + E(a_t^2) = \sigma_a^2$. Thus,

$$c(0) = \phi_1 c(1) + \sigma_a^2.$$

When $k \geq 1$, it can be easily verified that $E(a_t Y_{t-k}) = 0$. Thus,

$$c(k) = \phi_1 c(k-1), \quad k \geq 1.$$

This implies that $c(1) = \phi_1 c(0)$. Combining this with the previous result for $c(0)$ we find that

$$c(0) = Var(Y_t) = \frac{\sigma_a^2}{1 - \phi_1^2}.$$

Dividing the difference equation for $c(k)$ by $c(0)$, the autocorrelation function becomes

$$\rho(k) = \phi_1 \rho(k-1), \quad k \geq 1.$$

Difference equations of this type can be solved recursively as follows,

$$\rho(k) = \phi_1 \rho(k-1) = \phi_1 \cdot \phi_1 \rho(k-2) = \cdots = \phi_1^k \rho(k-k) = \phi_1^k \rho(0) = \phi_1^{|k|}, \;\; |k| \geq 1,$$

since $\rho(k)$ is an even function. Here, we have used the fact that $\rho(0) = 1$. This shows that if the $AR(1)$ process is stationary such that $|\phi_1| < 1$, then the ACF will decay exponentially or die out or decrease very quickly. If ϕ_1 is negative, the ACF will exhibit an alternating pattern; whereas if ϕ_1 is positive, the ACF values will be positive for all k.

Example To illustrate the patterns in the ACF of AR(1) processes, we have generated, computed and plotted the SACF of two series of size $n = 250$ each, from the $AR(1)$ process $(1 - \phi_1 \mathbf{B})\widetilde{Y}_t = a_t$ with parameters $\phi_1 = 0.9$ and $\phi_1 = -0.65$ respectively, where $\widetilde{Y}_t = Y_t - 10$ and $a_t \sim N(0, 1)$ white noise. A plot of the SACF of the two series generated from the AR(1) processes is shown in Figure 23. Observe that the values of the SACF of the first series are all positive because $\phi_1 > 0$. The SACF also decays exponentially. On the other hand, the SACF plot of the second series with $\phi_1 = -0.65$ exhibits an alternating pattern beginning with a negative value for r_1 and then dies out faster than the ACF of the first series. The SACF of the first series does not die out very fast because the value of $\phi_1 = 0.9$ is too close to 1. In both cases, $r_0 = 1$.

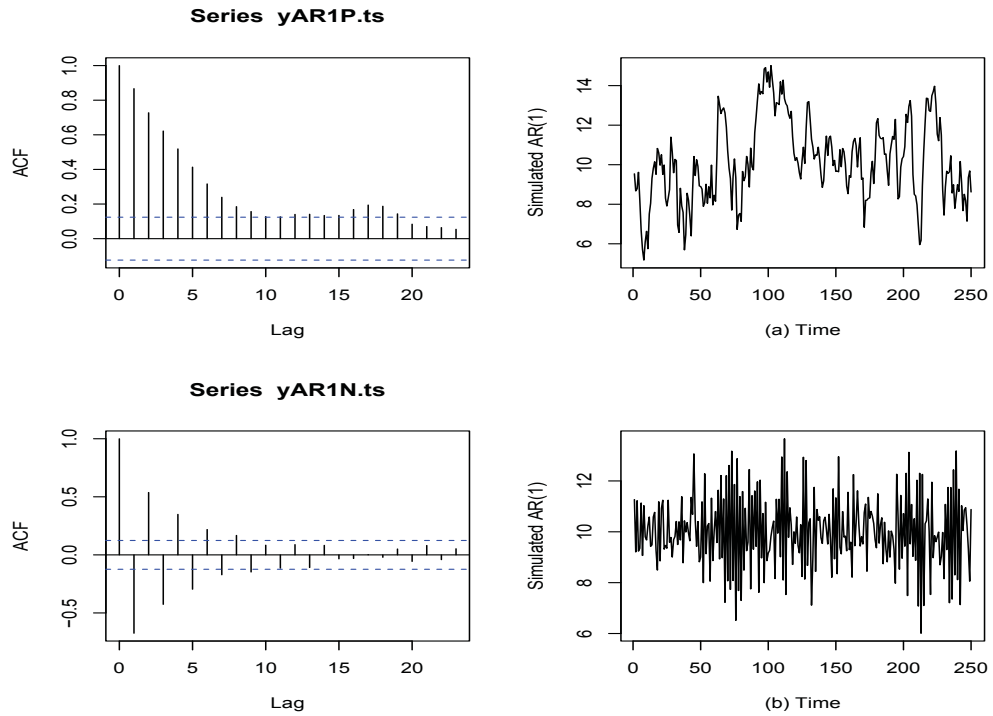

Figure 23: Plot of SACF of $AR(1)$ series generated from the model $(1 - \phi_1 \mathbf{B})(Y_t - 10) = a_t$ with parameters (a) $\phi_1 = 0.9$ and (b) $\phi_1 = -0.65$ respectively, where $a_t \sim N(0, 1)$ is a white noise process.

7.1.4 PACF of the $AR(1)$ Process

In Section 6.8, equation (11) we found that $\phi_{11} = \rho(1)$. It follows that for the $AR(1)$ process the lag 1 PACF is

$$\phi_{11} = \rho(1) = \phi_1.$$

For an $AR(1)$ process, the numerator in the lag 2 PACF given by (12) becomes

$$\begin{vmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix} = \begin{vmatrix} 1 & \phi_1 \\ \phi_1 & \phi_1^2 \end{vmatrix} = \phi_1^2 - \phi_1^2 = 0.$$

Therefore, for any $AR(1)$ process the lag 2 PACF $\phi_{22} = 0$. Furthermore, we see that the numerator in the lag 3 PACF given by (13) for any $AR(1)$ process becomes

$$\begin{vmatrix} \rho(0) & \rho(1) & \rho(1) \\ \rho(1) & \rho(0) & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix} = \begin{vmatrix} 1 & \phi_1 & \phi_1 \\ \phi_1 & 1 & \phi_1^2 \\ \phi_1^2 & \phi_1 & \phi_1^3 \end{vmatrix} = 0.$$

By examining the columns of the matrix, we notice that $\phi \times Col.\ 1 = Col.3$. That is, columns 1 and 3 are linearly dependent. Hence, the determinant in the numerator is zero.

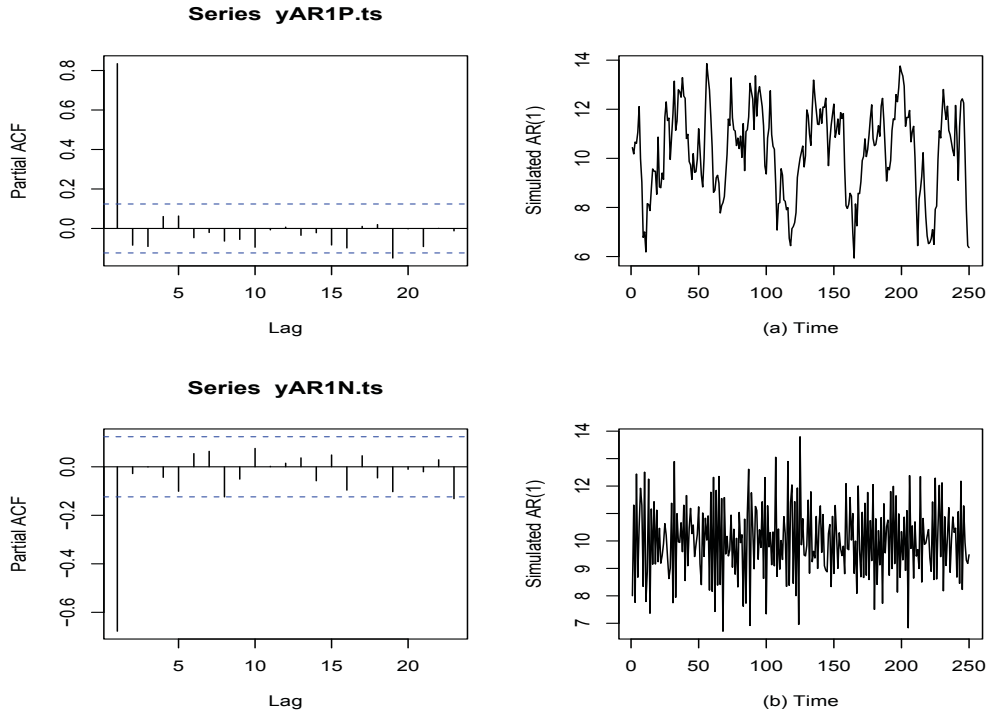

Figure 24: Plot of SPACF of two $AR(1)$ series generated from the model $(1 - \phi_1 \mathbf{B})(Y_t - 10) = a_t$ with parameters (a) $\phi_1 = 0.9$ and (b) $\phi_1 = -0.65$ respectively, where $a_t \sim N(0, 1)$ is a white noise process.

This again implies that the lag 3 PACF $\phi_{33} = 0$. Continuing in this way, we can actually show that the PACF of an $AR(1)$ process becomes zero after lag 1. That is, for any $AR(1)$ process,

$$\phi_{kk} = \begin{cases} \phi_1, & \text{if } k = 1 \\ 0, & \text{if } k > 1 \end{cases}$$

In general, the PACF of an $AR(1)$ process will show a positive or negative spike at lag 1 depending on the sign of ϕ_1, and then cut off as shown in Figure 24. In terms of the sample PACF this means that only the lag 1 SPACF will be significant while the sample PACF values for $k \geq 2$ will lie within the confidence band as noted earlier.

Example As an example, we display the SPACF of two sets of $n = 250$ observations generated from an $AR(1)$ process $(1-\phi_1 \mathbf{B})\widetilde{Y}_t = a_t$ with parameters $\phi_1 = 0.9$ and $\phi_1 = -0.65$ respectively in Figure 24, where $\widetilde{Y}_t = Y_t - 10$ and $a_t \sim N(0, 1)$ white noise. Figure 24 shows that the SPACF of the two $AR(1)$ series cuts off after lag 1 because none of the sample

PACF values are significant beyond that lag. We also note that the insignificant r_{kk} values do not exhibit any pattern.

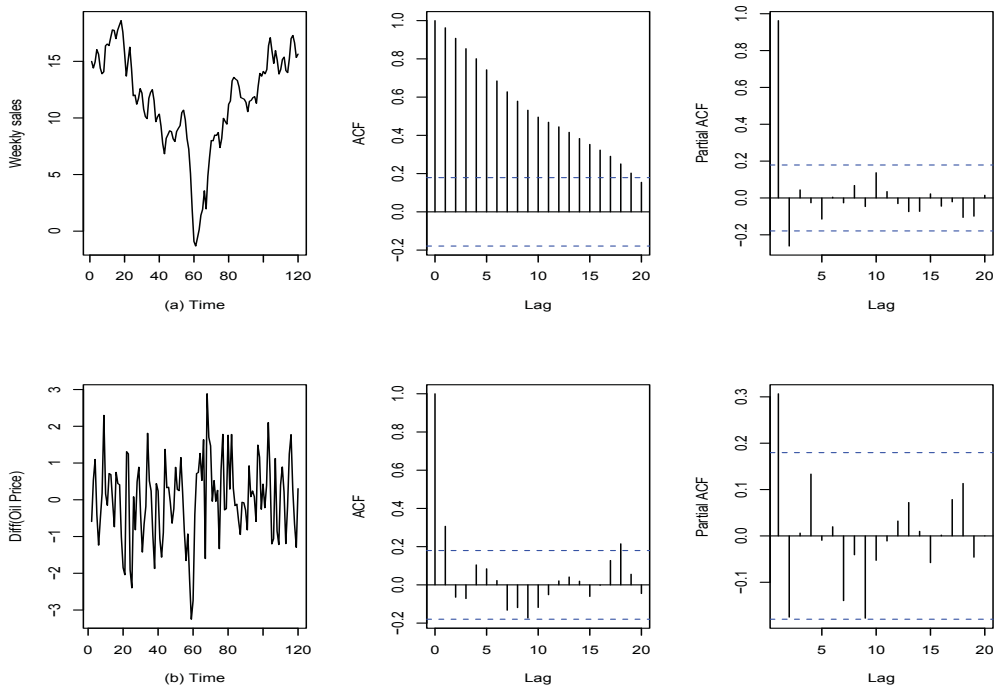

Figure 25: Plot of weekly sales of (a) absorbent paper towels (in units of 10,000 rolls); (b) first difference of weekly sales; and their SACF and SPACF values respectively.

Example: In this example, we discuss $n = 120$ weekly sales of absorbent paper towels (in units of 10,000 rolls). The time plot of the observed values is shown in Figure 25(a). Since the SACF of the first difference $z_t = y_t - y_{t-1}$ decays exponentially and the SPACF cuts off after lag 1, we identify the $AR(1)$ model as one possible model for the first difference of weekly sales of these paper towels. Since $\bar{z}_t = 0.0054$ where $z_t = (1 - \boldsymbol{B})Y_t$, we propose the $AR(1)$ model

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

for the differenced sales data, where $\widetilde{Z}_t = Z_t - \mu$. In terms of Y_t, the model becomes

$$[(1 - \mathbf{B})Y_t - \mu] = \phi_1[(1 - \mathbf{B})Y_{t-1} - \mu] + a_t$$

where a_t is assumed to be a white noise process that is normally distributed. The $AR(1)$ model (15) written in terms of the first difference of the nonstationary Y_t is commonly

referred to as the **autoregressive integrated model of order (1,1)**. Recall that theoretically $\rho(1) = \phi_1$. Thus, our initial guess for ϕ_1 denoted by $\hat{\phi}_1^{(0)}$ is the value $\hat{\phi}_1^{(0)} = r_{11} = 0.3066$. This initial guess has to be updated iteratively to obtain an estimate of ϕ_1. This model can then be used to forecast and study weekly sales of paper towels. The first ten values of r_k and r_{kk} are shown in the table below.

k	1	2	3	4	5	6	7	8	9	10
r_k	0.3066	-0.065	-0.073	0.105	0.084	0.023	-0.133	-0.119	-0.174	-0.118
r_{kk}	0.3066	-0.175	0.006	0.133	-0.0095	0.0198	-0.139	-0.041	-0.177	-0.053

Example: Another example of a real time series with SACF and SPACF patterns that are similar to that of an $AR(1)$ model is the daily average truck manufacturing defects shown in Figure 1(a). Plots of the SACF and SPACF are shown in Figure 26. Thus, one tentative time series model for representing daily average defects in the trucks manufactured by this company is

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + a_t,$$

since $\bar{y} = 1.7887$, where ϕ_1 has to be estimated from the data and a_t is assumed to be a sequence of uncorrelated normal random variables with mean zero and constant variance σ_a^2 which has to be estimated. Based on the SACF values, a good initial guess for ϕ_1 is $\hat{\phi}_1^{(0)} = r_1 = r_{11} = 0.4288$.

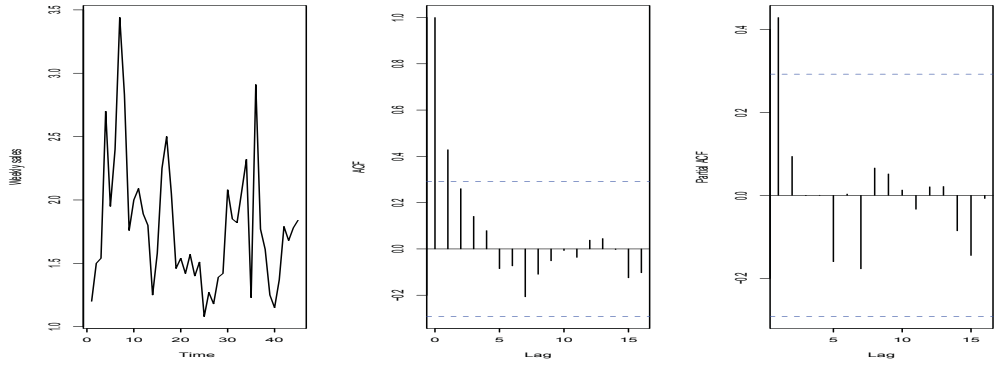

Figure 26: Plot of daily average number of truck manufacturing defects and its SACF and SPACF values.

7.2 The Second Order Autoregressive $AR(2)$ Process

Following our previous notations, the second order autoregressive $AR(2)$ model has two parameters ϕ_1 and ϕ_2 and can be written as

$$(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2) Y_t = a_t$$

or

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t.$$

7.2.1 Invertibility

By virtue of the definition above, any process Y_t that can be represented by an $AR(2)$ model is always invertible since $1 + |\phi_1| + |\phi_2| < \infty$.

7.2.2 Stationarity

Recall that for the $AR(2)$ to be stationary we should be able to rewrite it as a convergent MA series. We noted earlier that this is possible if the modulus of the roots of the quadratic equation in $\boldsymbol{B}$

$$1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2 = 0$$

lies outside the unit circle. Now, the roots of this equation are

$$\mathbf{B}_{1,2} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

Thus, the required condition for stationarity of an $AR(2)$ process is that $|\mathbf{B}_1| > 1$ and $|\mathbf{B}_2| > 1$. Alternatively, these conditions can be written as $1/|\mathbf{B}_1| < 1$ and $1/|\mathbf{B}_2| < 1$. Now,

$$\begin{aligned} \frac{1}{\boldsymbol{B}_{1,2}} &= \frac{2\phi_2}{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}} = \frac{2\phi_2}{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}} \times \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}} \\ &= \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}. \end{aligned}$$

It follows that

$$\left| \frac{1}{\mathbf{B}_1} \cdot \frac{1}{\mathbf{B}_2} \right| = |\phi_2| < 1$$

and

$$|\phi_1| = \left| \frac{1}{\mathbf{B}_1} + \frac{1}{\mathbf{B}_2} \right| \leq \left| \frac{1}{\mathbf{B}_1} \right| + \left| \frac{1}{\mathbf{B}_2} \right| < 2.$$

Thus, a necessary condition for stationarity of an $AR(2)$ process is that $-1 < \phi_2 < 1$ and $-2 < \phi_1 < 2$.

1. **Example**

2. Consider a process Y_t satisfying the model

$$Y_t = 1.5Y_{t-1} - 0.56Y_{t-2} + a_t.$$

   a) Is the process invertible ?

   b) Is the process stationary ?

   Solution

   (a) The process is invertible since $1 + |1.5| + | - 0.56| = 3.06 < \infty$

   (b) To determine whether the process is stationary, we compute the roots of the equation $-0.56\mathbf{B}^2 + 1.5\mathbf{B} - 1 = 0$. Now, the roots are

$$\mathbf{B} = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} = \frac{-1.5 \pm \sqrt{1.5^2 - 4(0.56)}}{2(-0.56)}.$$

   That is,

$$\mathbf{B}_1 = 1.25 \quad \text{and} \quad \mathbf{B}_2 \approx 1.429.$$

   Since the absolute value of these roots are both larger than 1, the process is a stationary process.

3. Is the process satisfying the model

$$Y_t = 0.8Y_{t-1} - 0.52Y_{t-2} + a_t.$$

   stationary ?

   Solution

   It is easy to verify that the roots of the equation $1-0.8\mathbf{B}+0.32\mathbf{B}^2 = 0$ are the complex conjugates

$$\mathbf{B} = \frac{-0.8 \pm i1.2}{-1.04} = 0.769 \pm i1.15.$$

   Now, $|0.769 \pm i1.15| = \sqrt{0.769^2 + 1.15^2} \approx 1.3868 > 1$. Therefore the process Y_t is stationary.

4. The process Y_t satisfying the model

$$Y_t = 0.2Y_{t-1} + 0.8Y_{t-2} + a_t.$$

   is not stationary because one of the roots of $1 - 0.2\mathbf{B} - 0.8\mathbf{B}^2 = 0$ is $\mathbf{B} = 1$, which is not outside the unit circle. The other root is $\mathbf{B} = -1.25$.

7.2.3 ACF of the $AR(2)$ Process

Recall that the autocorrelation function (ACF) of a process Y_t is $\rho(k) = c(k)/c(0)$. Thus, to derive the ACF of an $AR(2)$ process we first obtain the autocovariance function $c(k)$, where

$$\begin{aligned} c(k) &= E(Y_{t-k}Y_t) = \phi_1 E(Y_{t-k}Y_{t-1}) + \phi_2 E(Y_{t-k}Y_{t-2}) + E(a_t Y_{t-k}) \\ &= \phi_1 c(k-1) + \phi_2 c(k-2) + E(a_t Y_{t-k}), \quad \text{for } k \geq 0. \end{aligned}$$

Using the fact that $E(a_t Y_t) = \sigma_a^2$ and $E(a_t Y_{t-k}) = 0$, for $k \geq 1$, we have that

$$c(0) = \phi_1 c(1) + \phi_2 c(2) + \sigma_a^2,$$

and that the ACF of an AR(2) process satisfies the difference equation

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2), \quad k \geq 1.$$

Substituting the expressions for $c(1)$ and $c(2)$ in $c(0)$ and rearranging, we obtain

$$c(0) = Var(Y_t) = \frac{(1-\phi_2)\sigma_a^2}{[(1-\phi_2^2)(1-\phi_2) - \phi_1^2(1+\phi_2)]}.$$

The Yule-Walker difference equation for the $AR(2)$ process can be solved recursively. For instance, for $k = 1$, we have

$$\rho(1) = \phi_1 \rho(0) + \phi_2 \rho(1).$$

We then use the fact that $\rho(0) = 1$ to obtain

$$\rho(1) = \frac{\phi_1}{1-\phi_2}.$$

When $k = 2$, we have

$$\rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0).$$

Then, substituting for $\rho(1)$ and for $\rho(0) = 1$, we find that

$$\rho(2) = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1-\phi_2}.$$

We can continue in this way to obtain expressions for $\rho(k)$ for values of $k \geq 1$. However, we can obtain a general solution by letting $\rho(k) = \alpha^k$. Then by substituting for $\rho(k)$ in the difference equation we obtain

$$\alpha^k = \phi_1 \alpha^{k-1} + \phi_1 \alpha^{k-2}.$$

Now, divide through by α^{k-2} and rearrange to obtain the quadratic equation

$$\alpha^2 - \phi_1 \alpha - \phi_1 = 0,$$

which we solve to obtain

$$\alpha_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

Then, the general solution to the Yule-Walker equation for an $AR(2)$ process becomes

$$\rho(k) = b_1\alpha_1 + b_2\alpha_2 = b_1\left[\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k + b_2\left[\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}\right]^k,$$

where b_1 and b_2 are constants we can obtain by using the initial results for $\rho(1)$ and $\rho(2)$.

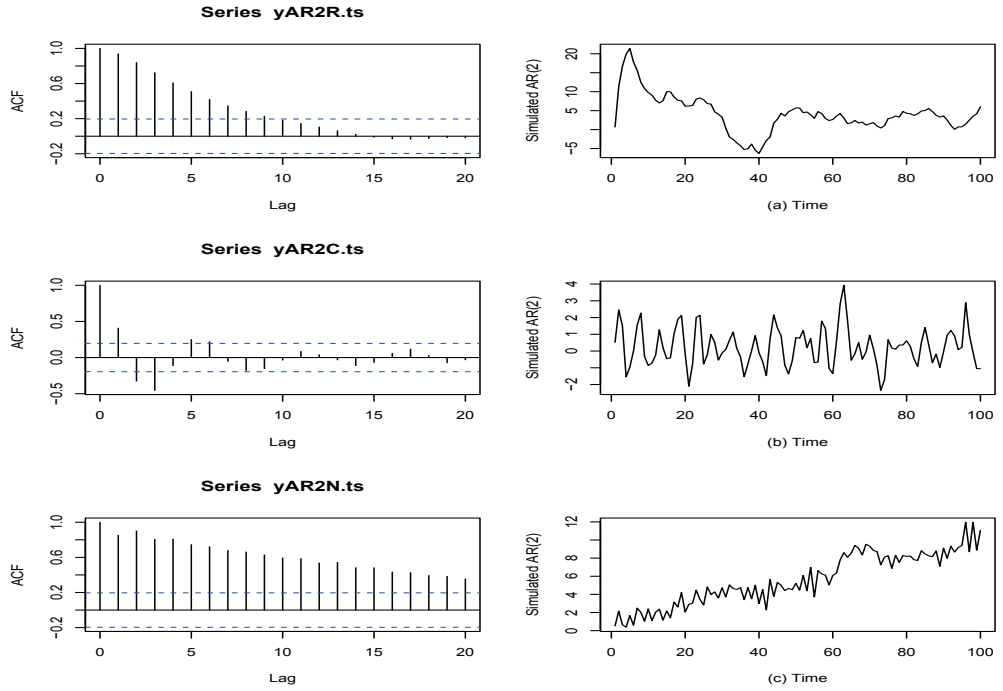

Figure 27: Plot of $AR(2)$ time series generated from the model $(1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2)Y_t = a_t$ with parameters (a) $(\phi_1, \phi_2) = (1.5, -0.56)$, (b) $(\phi_1, \phi_2) = (0.8, -0.52)$, and (c) $(\phi_1, \phi_2) = (0.2, 0.8)$ respectively, and their corresponding SACF, where $a_t \sim N(0, 1)$ is a white noise process.

In Section 7.3, we noted that an $AR(2)$ process is stationary if

$$\left|\frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}\right| < 1.$$

Hence, we can deduce from this more general result that for any stationary $AR(2)$ process, the ACF will decay exponentially as k increases if the roots of $1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 = 0$ are real.

If the roots are complex, a plot of the ACF will exhibit a damped sine wave pattern. The ACF will however not die out if the process is nonstationary. These patterns can be seen in the ACF of the series shown in Figure 26. The observations used in computing the ACF in Figure 27 was generated from an $AR(2)$ model with the parameters ϕ_1 and ϕ_2 carefully chosen to ensure that the data in Figure 27(a) is from a stationary process with two positive roots of the equation $1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 = 0$. The data in Figure 27(b) is from a process with roots that are complex conjugates, whereas the data shown in Figure 27(c) is from a nonstationary process.

7.2.4 PACF of the $AR(2)$ Process

In §6.8 we found that the lag 1 theoretical PACF $\phi_{11} = \rho(1)$. This means that for the $AR(2)$ process

$$\phi_{11} = \rho(1) = \frac{\phi_1}{1 - \phi_2}.$$

We also found that

$$\phi_{22} = (\rho(2) - \rho^2(1))/(1 - \rho^2(1)),$$

where $\rho(2)$ is given in §7.4. To obtain the lag 3 theoretical PACF ϕ_{33}, we consider the determinant in the numerator for the expression for ϕ_{33} given in (13), §6.8 as

$$\begin{vmatrix} \rho(0) & \rho(1) & \rho(1) \\ \rho(1) & \rho(0) & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}.$$

From the Yule-Walker difference equation for the $AR(2)$ process we obtain the following simultaneous equations

$$\begin{aligned} \rho(1) &= \phi_1 + \phi_2\rho(1) \\ \rho(2) &= \phi_1\rho(1) + \phi_2 \\ \rho(3) &= \phi_1\rho(2) + \phi_2\rho(2). \end{aligned}$$

These equations then imply that column three of the determinant can be written as a linear combination of columns one and two, namely, Col.3 = ϕ_1 Col.1 + ϕ_2 Col.2. That is, column 3 is linearly dependent on columns one and two. Therefore, this determinant is zero and hence the theoretical PACF at lag 3, $\phi_{33} = 0$. Similarly, we can show that $\phi_{kk} = 0$, for $k \geq 4$. Thus, the theoretical PACF of an $AR(2)$ process cuts off after lag 2. That is,

$$\phi_{kk} = \begin{cases} \rho(1), & k = 1, \\ \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2}, & k = 2, \\ 0 & k \geq 3. \end{cases}$$

Again, for the purpose of illustration, we generated three sets of observations from an AR(2) process with parameters (a) $(\phi_1, \phi_2) = (1.5, -0.56)$, (b) $(\phi_1, \phi_2) = (0.8, -0.52)$, and (c) $(\phi_1, \phi_2) = (0.2, 0.8)$ respectively, and computed their respective sample PACF. The series and their SPACFs are shown in Figure 28.

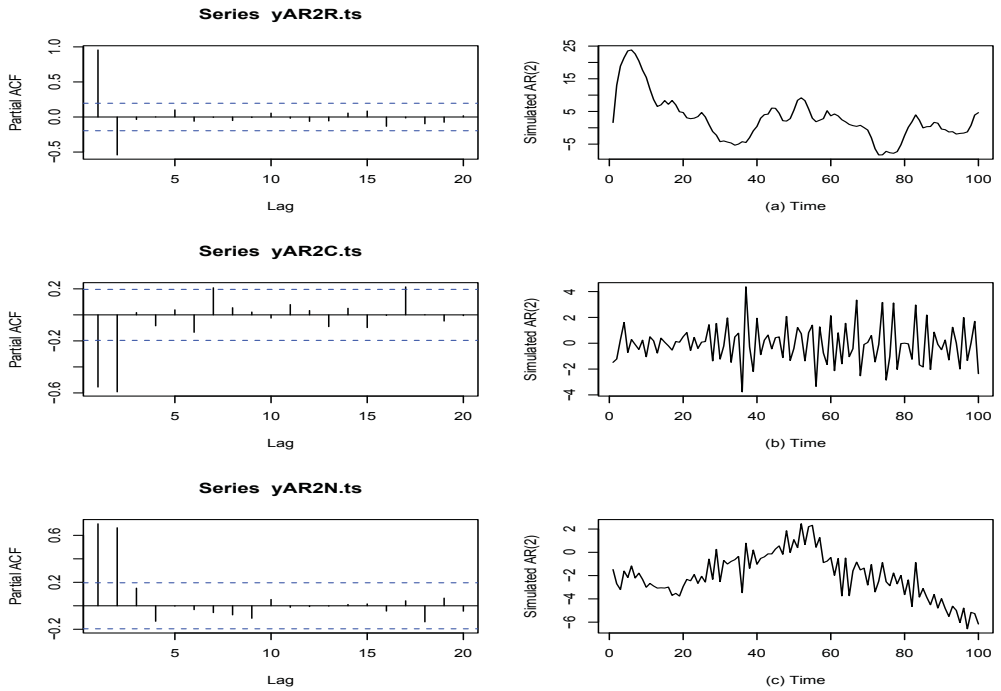

Figure 28: Plot of $AR(2)$ time series generated from the model $(1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2)Y_t = a_t$ with parameters (a) $(\phi_1, \phi_2) = (1.5, -0.56)$, (b) $(\phi_1, \phi_2) = (-0.8, -0.52)$, and (c) $(\phi_1, \phi_2) = (0.2, 0.8)$ respectively, and their corresponding SPACF, where $a_t \sim N(0, 1)$ is a white noise process.

Figure 28 shows that when ϕ_1 and ϕ_2 are both positive, r_{11} and r_{22} are also positive and significant; whereas when $\phi_1 > 0$ and $\phi_2 < 0$ we find that $r_{11} > 0$ and $r_{22} < 0$ and also significant. In addition, when $\phi_1 < 0$ and $\phi_2 < 0$ we find that $r_{11} < 0$ and $r_{22} < 0$ and significant. This is an indication that one can use the SPACF values of an $AR(2)$ process to guess whether the parameters in the $AR(2)$ representation will be positive or negative. Also note that the SPACF values cut off after lag 2. That is, the values after lag 2 where not significant. These properties we have noted will be very useful when modelling a real time series. In fact, a very useful property for identifying an $AR(p)$ model is that the SPACF of an $AR(3)$ model will cut off after lag 3; the SPACF of an $AR(4)$ model will cut off after lag 4; and so on.

Example: A good example of a real time series that can be described by an $AR(2)$ model are the daily readings of the viscosity of a chemical product $XB-77-5$ produced by Chemo, Inc. The time plot for this series is shown in Figure 24. We note that the SACF values tail off gradually and the SPACF values cut off after lag 2. For the observed series, $\bar{y} = 34.93007$. Thus one tentative model for Y_t = daily readings is

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + a_t.$$

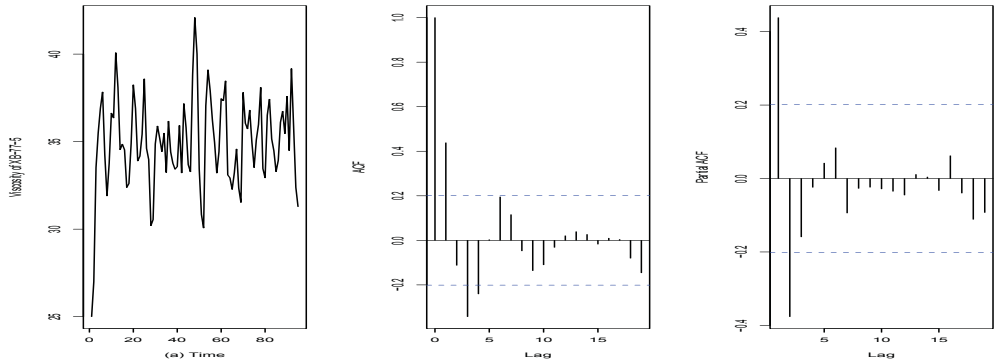

Figure 29: Plot of daily readings of viscosity of chemical product $XB - 77 - 5$; SACF and SPACF values.

In Figure 29, we see that for this series, $r_1 > 0$, $r_2 < 0$, $r_{11} > 0$ and $r_{22} < 0$. this gives us an indication that $\phi_1 > 0$ but ϕ_2 may be negative. So, during parameter estimation, our initial estimates will be chosen to satisfy these conditions.

8 Moving Average Models

We have seen that the moving average model of order q, $MA(q)$ is a linear combination of white noise processes given by (assuming $E(Y_t) = \mu = 0$)

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} + \ldots \theta_q a_{t-q} = \sum_{j=0}^{q} \theta_j \boldsymbol{B}^j a_t, \quad \theta_0 = -1.$$

We also found that the first two moments of Y_t are

$$E(Y_t) = \mu = 0, \quad \text{and} \quad Var(Y_t) = \sigma_a^2 \sum_{j=0}^{q} \theta_j^2,$$

and for $\theta_0 = -1$, the autocovariance function $c(k)$ of an $MA(q)$ process is

$$\begin{aligned} c(k) &= Cov(Y_t, Y_{t+k}) = E\left(\sum_{i=0}^{q}\sum_{j=0}^{q}\theta_i\theta_j a_{t-i}a_{t+k-j}\right) \\ &= \sigma_a^2\left(\sum_{i=1}^{q}\theta_i\theta_{i+k} - \theta_k\right). \end{aligned}$$

Therefore, the autocorrelation function of Y_t is

$$\rho(k) = \frac{c(k)}{c(0)} = \frac{\sigma_a^2\sum_{i=0}^{q}\theta_i\theta_{i+k}}{\sigma_a^2\sum_{j=0}^{q}\theta_j^2}.$$

We can see that the mean and variance of the MA process Y_t does not depend on time and the autocovariance $c(k)$ and autocorrelation $\rho(k)$ functions are functions of the lag k only. This implies that an $MA(q)$ process Y_t is always second order weakly stationary provided $\sum_{j=1}^{q}\theta_j^2 < \infty$. Next, we will now consider special cases of the $MA(q)$ process.

8.1 The First Order Moving Average $MA(1)$ Process

A process Y_t is said to be a moving average process of order 1, $MA(1)$ if it can be represented as

$$Y_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 \boldsymbol{B})a_t,$$

where a_t is a zero mean white noise process with constant variance σ_a^2.

8.1.1 Stationarity

Since $\theta_1^2 < \infty$, the $MA(1)$ process is always stationary.

8.1.2 Invertibility

An invertible MA process is one which can be inverted and written as a convergent AR series. We found earlier that the condition for invertibility of an MA process is that the modulus of the roots of the polynomial equation in the backshift operator $\boldsymbol{B}$ lie outside of the unit circle. For the $MA(1)$ process this means that the modulus of the root of $(1 - \theta_1\mathbf{B}) = 0$ must lie outside the unit circle. That is, $|1/\theta_1|$ must be larger than 1 or $|\theta_1| < 1$.

8.1.3 ACF Of The $MA(1)$ Process

The ACF at lag k is $\rho(k) = c(k)/c(0)$ where $c(k)$ is the autocovariance function. Thus, we start by deriving an expression for $c(k) = E(Y_tY_{t-k})$. Now,

$$c(k) = E(Y_tY_{t-k}) = E[(a_t - \theta_1 a_{t-1})Y_{t-k}] = E(a_tY_{t-k}) - \theta_1 E(a_{t-1}Y_{t-k}).$$

When $k = 0$ we have

$$c(0) = E(a_t Y_t) - \theta_1 E(a_{t-1} Y_t),$$

where

$$\begin{aligned} E(a_t Y_t) &= E[a_t(a_t - \theta_1 a_{t-1})] = E(a_t^2) - \theta_1 E(a_t a_{t-1}) = \sigma_a^2 \\ E(a_{t-1} Y_t) &= E[a_{t-1}(a_t - \theta_1 a_{t-1})] = E(a_t a_{t-1}) - \theta_1 E(a_{t-1}^2) \\ &= -\theta_1 \sigma_a^2. \end{aligned}$$

Therefore,

$$c(0) = (1 + \theta_1^2)\sigma_a^2.$$

Next, when $k = 1$ we have

$$\begin{aligned} c(1) &= E(a_t Y_{t-1}) - \theta_1 E(a_{t-1} Y_{t-1}) \\ &= -\theta_1 \sigma_a^2, \end{aligned}$$

since $E(a_t Y_{t-1}) = 0$. Clearly, for $k \geq 2$ we have that

$$c(k) = E(a_t Y_{t-k}) - \theta_1 E(a_{t-1} Y_{t-k}) = 0.$$

Therefore, the autocovariance function of an $MA(1)$ process is

$$c(k) = \begin{cases} (1 + \theta_1^2)\sigma_a^2, & k = 0 \\ -\theta_1 \sigma_a^2, & k = 1 \\ 0, & k > 1. \end{cases}$$

It follows that the autocorrelation function of an $MA(1)$ process at lag k is

$$\rho(k) = \frac{c(k)}{c(0)} = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1}{1+\theta_1^2}, & k = 1 \\ 0, & k > 1. \end{cases} \tag{15}$$

By examining the expression for the ACF we observe that (a) the ACF of an $MA(1)$ process cuts off after lag 1, (b) if θ_1 is negative, the ACF at lag 1 will be positive, and (c) if θ_1 is positive, the ACF at lag 1 will be negative. This information is quite useful when dealing with a real time series and one is trying to estimate the parameters of the $MA(1)$ model.

Example: To further understand the patterns in the ACF of MA(1) models we have generated $n = 100$ observations from an $MA(1)$ model with (a) $\theta_1 = 0.9$ (positive parameter and stationary and invertible MA(1)); (b) $\theta_1 = -0.65$ (negative parameter and stationary

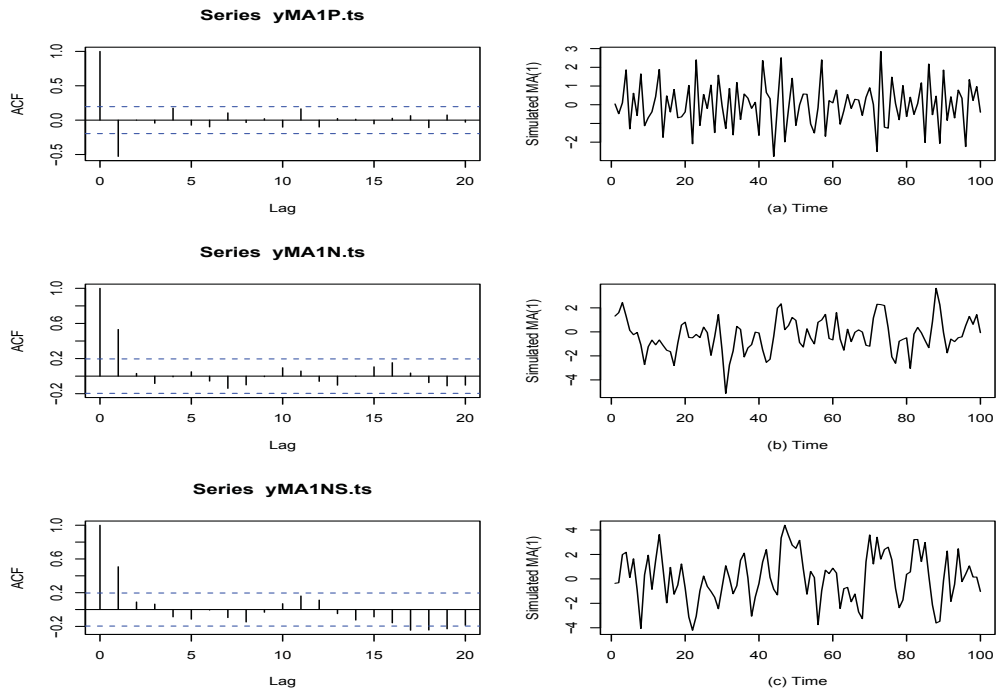

Figure 30: Plot of $MA(1)$ time series generated from the model $Y_t = (1 - \theta_1 \mathbf{B})a_t$ with parameters (a) $\theta_1 = 0.9$, (b) $\theta_1 = -0.65$, and (c) $\theta_1 = -1/0.65$ respectively, and their corresponding SACF, where $a_t \sim N(0, 1)$ is a white noise process.

and invertible MA(1)), and (c) $\theta_1 = -1/0.65$ (negative parameter and stationary but not invertible MA(1)). The patterns in the ACF of the observations are shown in Figure 30. As noted earlier, the SACF in Figure 30(a) is negative since $\theta_1 > 0$ and positive in Figure 30(b), (c) because $\theta_1 < 0$. Observe that the fact that the series in Figure 30(c) is not invertible is not noticeable in the SACF of the series. This indicates that the SACF will not capture whether a series is invertible or not invertible. It is only the partial autocorrelation SPAC function of MA processes that will detect a series that is not invertible. We will see that the SPACF of a process that is not invertible will not decay exponentially, whereas the SACF of a series that is not stationary will not decay exponentially. The SACF in Figure 30(c) decays exponentially because MA processes of finite order are always stationary.

8.1.4 PACF Of The $MA(1)$ Process

Following previous results, we know that the PACF at lag 1 is

$$\phi_{11} = \rho(1) = -\frac{\theta_1}{1 + \theta_1^2}$$

in this case. At lag 2, we have

$$\phi_{22} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} = -\frac{\rho^2(1)}{1 - \rho^2(1)},$$

since $\rho(2) = 0$. After some algebra, we obtain

$$\phi_{22} = -\frac{\theta_1^2}{1 + \theta_1^2 + \theta_1^4} = -\frac{\theta_1^2(1 - \theta_1^2)}{(1 - \theta_1^6)}.$$

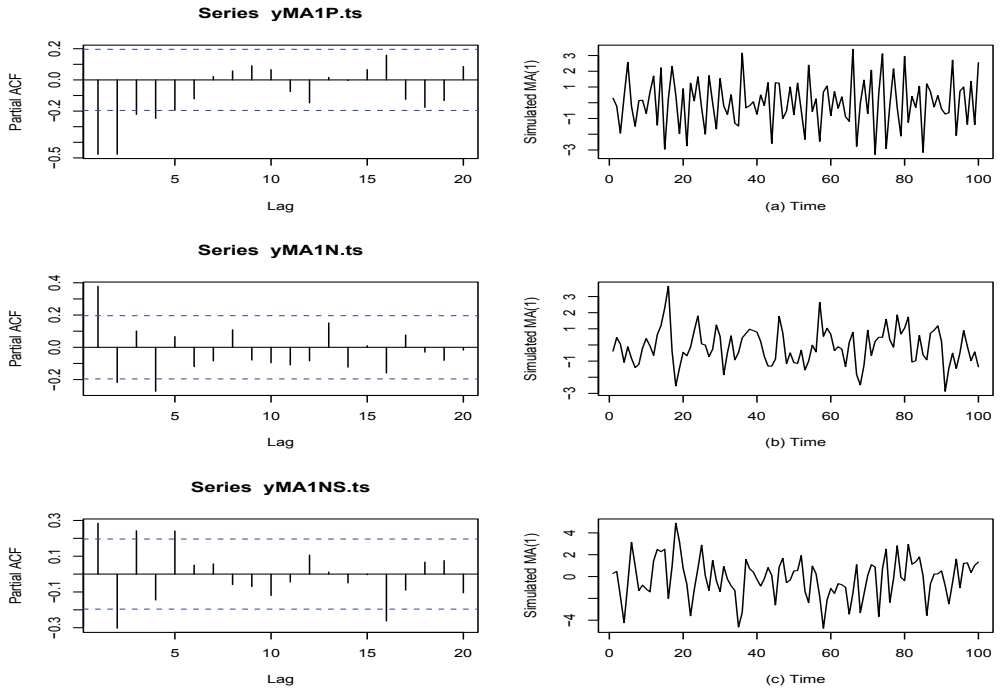

Figure 31: Plot of $MA(1)$ time series generated from the model $Y_t = (1 - \theta_1 \mathbf{B})a_t$ with parameters (a) $\theta_1 = 0.9$, (b) $\theta_1 = -0.65$, and (c) $\theta_1 = -1/0.65$ respectively, and their corresponding SPACF, where $a_t \sim N(0, 1)$ is a white noise process.

Using equation (13) in §6.8, we find that the PACF of the $MA(1)$ process at lag 3 is given by

$$\phi_{33} = \frac{\rho^3(1)}{1 - 2\rho^2(1)} = -\frac{\theta_1^3}{1 + \theta_1^2 + \theta_1^4 + \theta_1^6} = -\frac{\theta_1^3(1 - \theta_1^2)}{(1 - \theta_1^8)}.$$

It can be shown that for the $MA(1)$ model the lag k PACF is given by

$$\phi_{kk} = -\frac{\theta_1^k(1 - \theta_1^2)}{(1 - \theta_1^{2(k+1)})}.$$

This means that if the MA(1) process Y_t is invertible ($|\theta_1| < 1$), then (a) the PACF ϕ_{kk} will decay exponentially as k increases, (b) the PACF values will alternate between positive and

negative values if $|\theta_1| < 0$; whereas (c) all the PACF values will be negative if $|\theta_1| > 0$. We note that for real time series, there may be a few deviations from this pattern.

Example: As an illustration we display the sample PACF of the observations generated from the MA(1) models used in Figure 30. In Figure 31(a) we see that the SPACF values are in most cases negative and does not die out very quickly because the value of $\theta_1 = 0.9$ is positive and close to 1. The alternating pattern which occurs when $\theta_1 < 0$ is seen in Figure 31(b), (c). However, in Figure 31(c) the SPACF at lag 16 is still significant because for this series $|\theta_1| > 1$.

8.2 The Second Order Moving Average $MA(2)$ Process

An $MA(2)$ process Y_t is a process that can be represented as

$$Y_t = (1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2) a_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

where a_t is a zero mean white noise process with constant variance σ_a^2.

8.2.1 Stationarity

Being a finite order MA process, the $MA(2)$ process is always stationary since $\theta_1^2 + \theta_2^2 < \infty$.

8.2.2 Invertibility

For invertibility, the roots of $(1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2) = 0$ must lie outside the unit circle. This condition is similar to that of the $AR(2)$ process we have discussed in §7.2. Following the same arguments as in §7.2 we can show that a necessary condition for invertibility of an $MA(2)$ process is that

$$|\theta_2| < 1 \quad \text{and} \quad |\theta_1| < 2.$$

Suppose that an $MA(2)$ process is invertible, one can express the process as an infinite but convergent AR series. To do this we rewrite the $MA(2)$ model as

$$a_t = (1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2)^{-1} Y_t = \Psi(\boldsymbol{B}) Y_t$$

where $\Psi(\boldsymbol{B}) = 1 - \psi_1 \boldsymbol{B} - \psi_2 \boldsymbol{B}^2 - \psi_3 \boldsymbol{B}^3 - \cdots$. Now, to determine the coefficients ψ_j in the AR representation we note that the above representation implies that

$$1 = (1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2)(1 - \psi_1 \boldsymbol{B} - \psi_2 \boldsymbol{B}^2 - \psi_3 \boldsymbol{B}^3 - \cdots).$$

Now equating coefficients on both sides we have

$$\begin{aligned}
\boldsymbol{B}: &\quad -\psi_1 - \theta_1 = 0 \\
\boldsymbol{B}^2: &\quad -\psi_2 + \theta_1\psi_1 - \theta_2 = 0 \\
\boldsymbol{B}^3: &\quad -\psi_3 + \theta_1\psi_2 + \theta_2\psi_1 = 0 \\
\cdots: &\quad \cdots\cdots\cdots \\
\cdots: &\quad \cdots\cdots\cdots \\
\boldsymbol{B}^j: &\quad -\psi_j + \theta_1\psi_{j-1} + \theta_2\psi_{j-2} = 0, \quad j = 3, 4, \ldots
\end{aligned}$$

From these equations we have that the coefficients in the AR representation of the invertible $MA(2)$ model are

$$\begin{aligned}
\psi_1 &= -\theta_1 \\
\psi_2 &= -\theta_1^2 - \theta_2 \\
\psi_j &= \theta_1\psi_{j-1} + \theta_2\psi_{j-2}, \quad j = 3, 4, \ldots,
\end{aligned}$$

with

$$(1 - \psi_1\boldsymbol{B} - \psi_2\boldsymbol{B}^2 - \psi_3\boldsymbol{B}^3 - \cdots)Y_t = a_t.$$

8.2.3 ACF of an $MA(2)$ Process

As before, we begin by deriving the expression for the autocovariance function $c(k)$ of an $MA(2)$ process Y_t. By definition (assuming $E(Y_t) = 0$),

$$c(k) = E(Y_tY_{t-k}) = E(a_tY_{t-k}) - \theta_1E(a_{t-1}Y_{t-k}) - \theta_2E(a_{t-2}Y_{t-k}).$$

Now, we recall that

$$E(Y_ta_s) = \begin{cases} \sigma_a^2, & \text{if } t = s, \\ 0, & \text{if } t < s. \end{cases} \tag{16}$$

In addition, we have that for the $MA(2)$ process

$$E(a_{t-1}Y_t) = E(a_{t-2}Y_{t-1}) = -\theta_1\sigma_a^2 \quad \text{and } E(a_{t-2}Y_t) = -\theta_2\sigma_a^2.$$

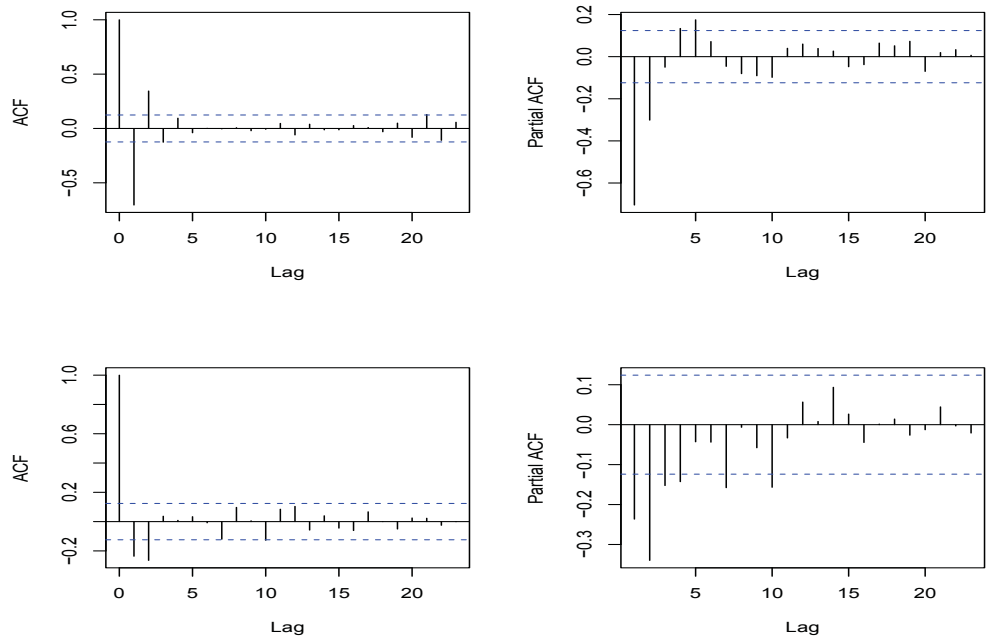

Figure 32: Plot of the SACF and SPACF of $MA(2)$ time series generated from the model $Y_t = (1 - \theta_1\mathbf{B} - \theta_2\mathbf{B}^2)a_t$ with parameters (a) $(\theta_1, \theta_2) = (1, -0.6)$, (b) $(\theta_1, \theta_2) = (0.85, 0.5)$ where $a_t \sim N(0, 1)$ is a white noise process.

Using these results we find that

$$c(k) = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma_a^2, & k = 0 \\ -\theta_1(1 - \theta_2)\sigma_a^2, & k = 1 \\ -\theta_2\sigma_a^2, & k = 2 \\ 0 & k \geq 3 \end{cases}$$

It follows that the autocorrelation function of an $MA(2)$ process Y_t is

$$\rho(k) = \frac{c(k)}{c(0)} = \begin{cases} 1, & k = 0 \\ -\frac{\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, & k = 1 \\ -\frac{\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2 \\ 0 & k \geq 3 \end{cases}$$

That is, the ACF of any $MA(2)$ process will cut-off after lag 2. Furthermore, (a) if $0 < \theta_2 < 1$ and $\theta_1 > 0$, then $\rho(1) < 0$ and $\rho(2) < 0$; (b) if $\theta_2 < 0$ and $\theta_1 > 0$, then $\rho(1) < 0$ and $\rho(2) > 0$; (c) if $\theta_2 > 1$ and $\theta_1 < 0$, then $\rho(1) > 0$ and $\rho(2) < 0$; (d) $\theta_2 > 1$ and $\theta_1 > 0$, then $\rho(1) < 0$ and $\rho(2) < 0$; and so on.

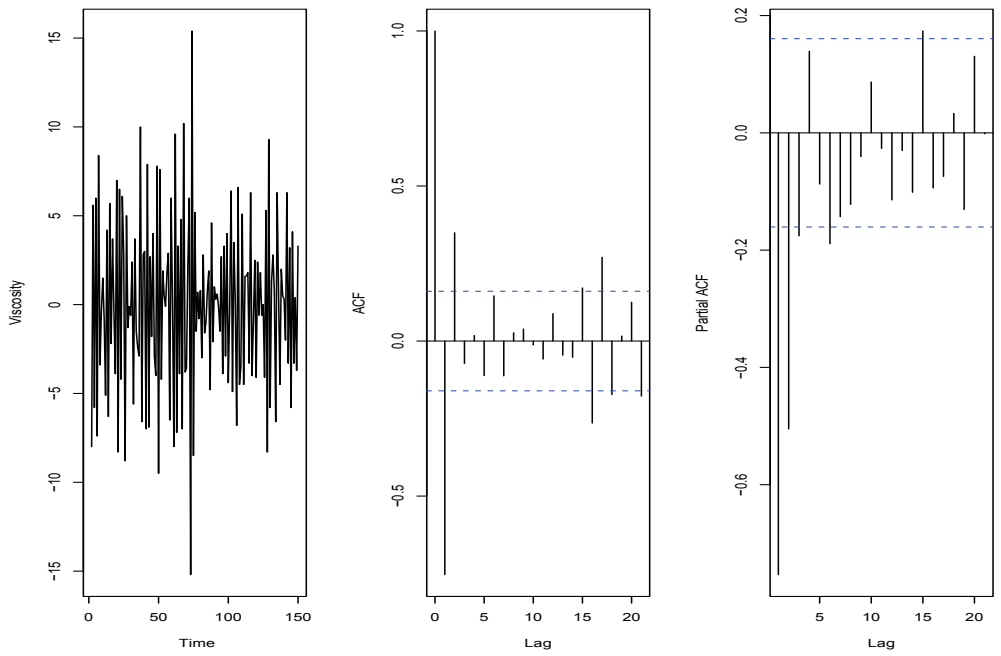

Figure 33: Plot of daily readings of the viscosity of chemical product $XR-22$; SAC values and SPAC values.

8.2.4 PACF of an $MA(2)$ Process

Previously we have shown that

$$\begin{aligned}\phi_{11} &= \rho(1); \quad \phi_{22} = \frac{\rho(2)-\rho^2(1)}{1-\rho^2(1)} \\ \phi_{33} &= \frac{\rho^2(1)-\rho(1)\rho(2)(2-\rho(2))}{1-\rho^2(2)-2\rho^2(1)(1-\rho(2))},\end{aligned}$$

and so on. It is clear that the result of substituting for $\rho(k)$ in the expressions for ϕ_{kk} may be difficult to interpret. However, since the $MA(1)$ model is a special case of the $MA(2)$, the PACF of the $MA(2)$ process will also decay exponentially or as a damped sine wave depending on the sign and magnitude of θ_1 and θ_2. As an example consider the SACF and SPACF in Figure 32 of $n = 250$ observations generated from $MA(2)$ processes with parameters $(\theta_1, \theta_2) = (1, -0.6)$ and $(\theta_1, \theta_2) = (0.65, 0.24)$.

Example: In this example we apply the patterns in the SACF and SPACF of $MA(2)$ series to the daily viscosity readings for a chemical product $XR-22$. Figure 33 shows that the SACF of the data cuts-off after lag 2, except for two values at lags 16 and 17, and the

SPACF decays exponentially. This is to be expected since the level of the confidence band is 95%. Therefore, one possible model for describing the daily viscosity readings is the $MA(2)$ model.

8.3 The Autoregressive Moving Average Process

The SACF and SPACF of the lag 12 difference of wages and salaries in Newfoundland from 1951-1969 in Figure 34, show that the SACF and SPACF values of a series may not cut off as quickly as we may desire or follow any of the patterns discussed earlier. In such cases, we may combine both the AR and MA representations to model the series. Recall that the $ARMA(p,q)$ model is written as

$$\Phi_p(\boldsymbol{B})Y_t = \Theta_q(\boldsymbol{B})a_t,$$

where $\Phi_p = 1 - \sum_{j=1}^{p} \phi_j \boldsymbol{B}^j$ and $\Theta_q = 1 - \sum_{j=1}^{q} \theta_j \boldsymbol{B}^j$. As an example, if we combine the $AR(1)$ and $MA(1)$ representations we obtain the Autoregressive Moving Average Model of order $(1,1)$ denoted by $ARMA(1,1)$ and written as

$$Y_t = \phi_1 Y_{t-1} + a_t - \theta_1 a_{t-1}$$

or

$$(1 - \phi_1 \mathbf{B})Yt = (1 - \theta_1 \mathbf{B})a_t.$$

Other examples are the $ARMA(2,1)$ model given by

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} a_t - \theta_1 a_{t-1},$$

and the $ARMA(1,2)$ model given by

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}.$$

In general, the $ARMA(p,q)$ model will have p autoregressive terms with parameters $\phi_1, \ldots, \phi_p$ and q moving average terms with parameters $\theta_1, \ldots, \theta_q$. Clearly, the $AR(p)$ and the $MA(q)$ models are special cases of the $ARMA(p,q)$ model. In particular, $AR(p) \equiv ARMA(p,0)$ and $MA(q) \equiv ARMA(0,q)$.

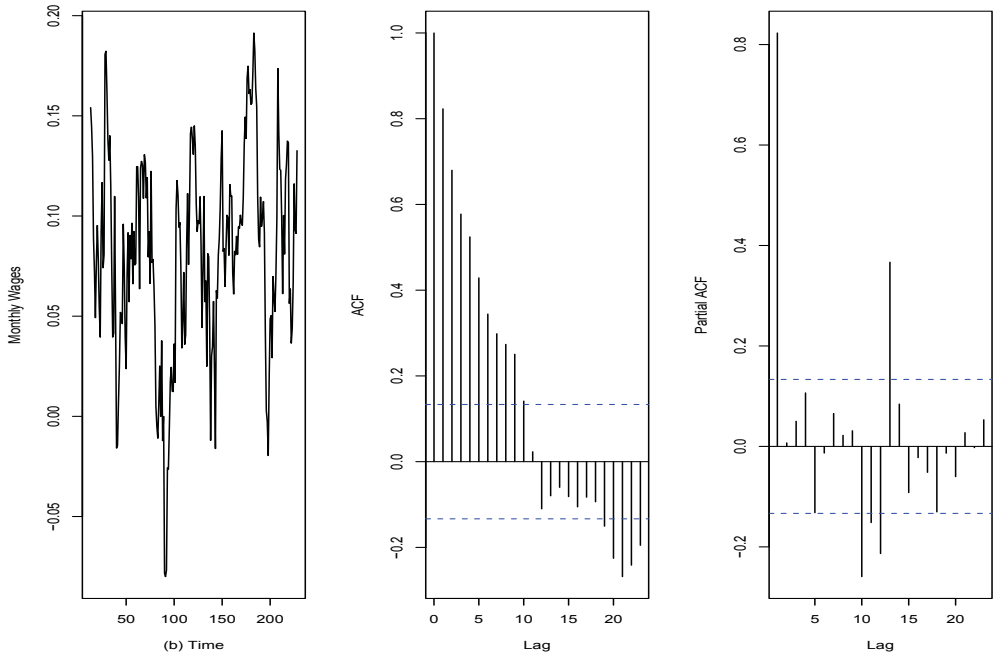

Figure 34: SACF and SPACF of the lag 12 difference of wages and salaries in Newfoundland from 1951-1969.

8.4 The Autoregressive Moving Average Process of order (1,1)

8.4.1 Stationarity and Invertibility

In §6.5 we discussed the stationarity and invertibility conditions for the general $ARMA(p, q)$ model. Applying the conditions in §6.5 to the $ARMA(1, 1)$ model, it is straightforward to see that an $ARMA(1, 1)$ process Y_t will be stationary if $|\phi_1| < 1$ and invertible if $|\theta_1| < 1$.

8.4.2 Autocorrelation and Partial Autocorrelation Functions

Following previous approaches, the theoretical autocovariance function of an $ARMA(1, 1)$ process Y_t can be written as

$$c(k) = \phi_1 c(k-1) + E(a_t Y_{t-k}) - \theta_1 E(a_{t-1} Y_{t-k}).$$

To obtain the variance of Y_t we set $k = 0$. That is,

$$c(0) = Var(Y_t) = \phi_1 c(1) + E(a_t Y_t) - \theta_1 E(a_{t-1} Y_t).$$

By using the result in equation (16) it can be shown that $E(a_{t-1} Y_t) = (\phi_1 - \theta_1)\sigma_a^2$. Therefore,

$$c(0) = Var(Y_t) = \phi_1 c(1) + [1 - \theta_1(\phi_1 - \theta_1)]\sigma_a^2.$$

Similarly, it is easy to show that

$$\begin{aligned} c(1) &= \phi_1 c(0) - \theta_1 \sigma_a^2, \\ c(k) &= \phi_1 c(k-1), \quad k \geq 2. \end{aligned}$$

From these equations we find that the variance of an $ARMA(1,1)$ process Y_t is given by

$$c(0) = \frac{\theta_1^2 - 2\theta_1\phi_1 + 1}{1 - \phi_1^2}\sigma_a^2,$$

and that the ACF

$$\rho(k) = \phi_1 \rho(k-1), \quad k \geq 2.$$

This implies that the ACF of a stationary and invertible $ARMA(1,1)$ process behaves like that of an $AR(1)$ process after lag 1. One can actually show that the ACF of an $ARMA(1,q)$ process will behave like that of an $AR(1)$ process after lag q. In general, the ACF of an $ARMA(p,q)$ process behaves like that of an $AR(p)$ process after lag q. That is, after lag q, the ACF will begin to decay exponentially.

The PACF of an $ARMA(p,q)$ process has a complicated structure. However, since AR and MA processes are special cases of the $ARMA$ process we can imagine that the PACF of an invertible and stationary process will also decay exponentially.

9 MODEL IDENTIFICATION

The most crucial steps in the analysis of time series are to identify and build a model based on available data. This requires a good understanding of the patterns in the SAC and SPAC values and characteristics of the $AR(p)$ and $MA(q)$ processes discussed in Section 5. The goal in model identification is to match patterns in the sample autocorrelation function, SACF and sample partial autocorrelation function, SPACF with known patterns of the theoretical autocorrelation function, ACF and the theoretical partial autocorrelation function, PACF. Given a time series, the following steps have been found to be useful in identifying a tentative model for the data.

Step 1. Plot the time series.

Examine the time plot very carefully for any signs of trend, seasonality, outliers, non-constant variances and other nonstationary phenomena. If necessary, use appropriate transformations to stabilize the variance and also transform the series into a stationary series. These transformations may include one or a combination of the following.

(a) One of the power transformations.

(b) Fitting polynomial regression models with or without dummy variables/trigonometric functions.

(c) Applying multiplicative or additive decomposition methods

(d) Differencing.

Step 2. Compute the sample ACF (SACF) and the sample PACF (SPACF) of the transformed series from Step 1. Examine the SACF and SPACF to determine whether the transformed series is stationary or requires further transformation. If further transformation is required, transform the series from Step 1 until a stationary series is obtained.

Step 3. Compute the sample ACF (SACF) and the sample PACF (SPACF) of the stationary series from Step 2. If the series from Step 1 was stationary, the sample ACF (SACF) and the sample PACF (SPACF) of the stationary series are the same as in Step 2. Examine the patterns in the SACF and SPACF to identify the orders, p and q in the $ARMA(p, q)$ model for the data.

We now tabulate the patterns in the SACF and SPACF we have derived that are helpful in identifying whether an AR, MA or $ARMA$ model is suitable for a time series.

Process	Theoretical Autocorrelation Function (ACF)	Theoretical Partial Autocorrelation Function (PACF)
$AR(p)$	Decays exponentially or as damped sine wave depending on sign and magnitude of model parameters	Cuts off after lag p
$MA(q)$	Cuts off after lag q	Decays exponentially or as damped sine wave depending on sign and magnitude of model parameters
$ARMA(p, q)$	Decays exponentially after lag q	Decays exponentially after lag p

It is important to note that, a large number of observations is needed in order to build a reasonable $ARMA$ model. The sampling variation and the correlation among the sample ACF (SACF) and sample PACF (SPACF) will sometimes disguise the theoretical patterns so that it may not be obvious. Hence, in the initial model identification one is advised to concentrate on the broad features of the SACF and SPACF and not on the fine details. Model improvement can be carried out at the diagnostics stage of the analysis.

10 ESTIMATION OF MODEL PARAMETERS

Once tentative models for a given time series have been identified, the next step is to estimate the parameters of the model using available data. In this section, we will discuss the method of moments, least squares estimation and the maximum likelihood method.

10.1 Method of moments (MM)

The method of moments simply uses the expressions for the ACF of each model by replacing the theoretical ACF $\rho(k)$ by the SACF r_k and solving for the unknown parameter.

10.1.1 AR models

Recall that for the $AR(p)$ model, the ACF satisfies the Yule-Walker difference equation

$$\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p), \quad k \geq 1. \tag{17}$$

When $k = 0$, we have that

$$c(0) = \phi_1 c(k-1) + \cdots + \phi_p c(k-p) + \sigma_a^2.$$

This implies that by replacing $\rho(k)$ by r_k and $c(0)$ by S^2, an estimate of the variance of Y_t, a method of moments estimate of the white noise variance can be written as

$$\hat{\sigma}_a^2 = S^2(1 - \hat{\phi}_1 r_{k-1} - \hat{\phi}_2 r_{k-2} - \cdots - \hat{\phi}_p r_{k-p})$$

where $S^2 = \sum_{t=1}^{n}(Y_t - \bar{Y})^2/(n-1)$.

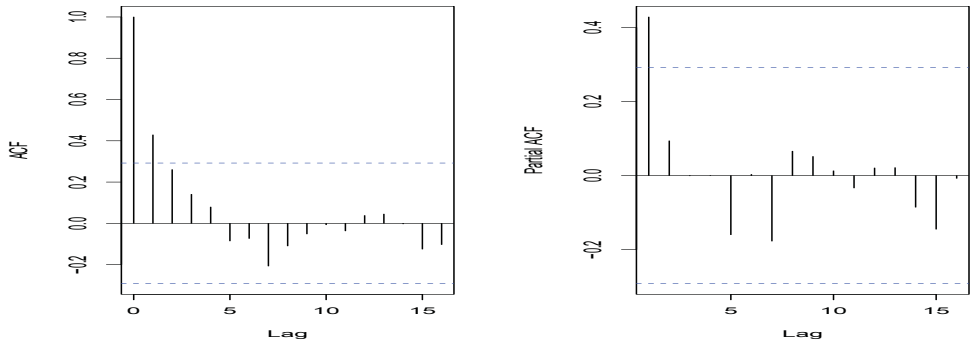

Figure 35: Plots of SACF and SPACF of daily average number of truck manufacturing defects.

Some Examples

Example 1, $AR(1)$: To illustrate the application of the MM we consider the daily average number of truck manufacturing defects shown in Figure 1. Based on the sample ACF and sample PACF shown in Figure 35, we propose that an $AR(1)$ and a $MA(1)$ models are two suitable models for the truck defects series. To estimate the parameters of the AR(1) model, we note that for this series, $r_1 = 0.429$. Since $p = 1$, the Yule-Walker estimating equation in (17) reduces to

$$r_1 = \hat{\phi}_1 r_0.$$

Therefore, the MM estimate is $\hat{\phi}_1 = r_1 = 0.429$. Next, we find that for this data $S^2 = 0.26728$. Thus, an estimate of the white noise variance is

$$\hat{\sigma}_a^2 = S^2(1 - \hat{\phi}_1 r_0) = 0.26728(1 - 0.429) \approx 0.1526.$$

Example 2, $AR(2)$: In order to examine the performance of the MM, we have generated data from an $AR(2)$ process with known parameters $\phi_1 = 1.5$ and $\phi_2 = -0.56$ and used the MM to estimate the parameters. This will enable us compare the estimated parameters with the known parameters in order to assess the performance of the method. Since there are two parameters, we obtain the following two estimating equations from (17)

$$\begin{aligned} r_1 &= \hat{\phi}_1 + \hat{\phi}_2 r_1, \\ r_2 &= \hat{\phi}_1 r_1 + \hat{\phi}_2. \end{aligned}$$

Clearly, solving for $\hat{\phi}_2$ immediately yields $\hat{\phi}_2 = \phi_{22}$, the lag 2 sample partial autocorrelation function. Then, from the first equation we obtain

$$\hat{\phi}_1 = (1 - \hat{\phi}_2) r_1.$$

Computing the SACF and SPACF of the data we have generated, we find that $r_1 = 0.910$, $\phi_{22} = -0.587$. Therefore, we have that

$$\hat{\phi}_2 = \phi_{22} = -0.587, \quad \hat{\phi}_1 = (1 - \hat{\phi}_2) r_1 = 0.910(1 + 0.587) = 1.444.$$

Comparing these estimates to the true values, we see that the MM performed fairly well in estimating the parameters of the $AR(2)$ model.

The results obtained for the $AR(2)$ can be easily extended to higher order AR models. Clearly, for an AR(p) model, $\hat{\phi}_p = \phi_{pp}$. The remaining $p - 1$ parameters can then be obtained by solving the other $p - 1$ equations simultaneously.

10.1.2 MA models

We will use the $MA(1)$ model to illustrate the MM approach for MA models. Recall that the $MA(1)$ model has only one parameter, we therefore need only one equation to obtain the MM estimate. Now, we recall that for MA(1) models

$$\rho(1) = -\frac{\theta_1}{1+\theta_1^2}, \quad \text{and} \quad c(0) = (1+\theta_1^2)\sigma_a^2.$$

Thus, the MM estimating equation for θ_1 is given by

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

and for σ_a^2 is

$$\hat{\sigma}_a^2 = \frac{S^2}{1+\hat{\theta}_1^2}.$$

Solving the equation for θ_1 we obtain

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

It is clear that the MM estimating equations for higher order MA models will be more complicated and nonlinear. Thus, the parameters of MA models are usually estimated by numerical iterative procedures.

As an example, we apply the estimating formula to a data set generated from a MA(1) process with parameter $\theta_1 = -0.65$ and $\sigma_a^2 = 1$. The lag 1 SACF computed from the data was $r_1 = 0.395$ and $S^2 = 0.9859821$. Applying the estimating equation we obtain

$$\hat{\theta}_1^{1,2} = \frac{-1 \pm \sqrt{1-4r_1^2}}{2r_1} = \frac{-1 \pm \sqrt{1-4(0.395^2)}}{2(0.395)}.$$

Simplifying the expression yields $\theta_1^1 = -0.489$ and $\theta_1^2 = -2.0419$. Since $\theta_1^2 = -2.0419$ leads to a MA(1) model that is not invertible we choose the estimate $\theta_1^1 = -0.489$. The estimate for the white noise variance becomes

$$\hat{\sigma}_a^2 = \frac{S^2}{1+\hat{\theta}_1^2} = \frac{0.98598}{1+(-0.489)^2} = 0.7957.$$

When compared to the actual values of the parameters, these results show that the MM estimates are highly biased. In most cases, MM estimates are used as initial values in computing maximum likelihood estimates of model parameters. It is well known that ML estimates are usually more efficient than MM estimates.

10.2 Least Squares Method (LSM)

The least sqaures estimates are obtained by minimizing the sum of squares of the white noise process with respect to the model parameters. We will use the $AR(1)$ and $MA(1)$ models to illustrate the technique.

10.2.1 AR(1)

Suppose that $\mu \neq 0$, then the $AR(1)$ model can be written as

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + a_t.$$

Define

$$Q(\phi_1, \sigma_a^2) = \sum_{t=2}^{n} a_t^2 = \sum_{t=2}^{n}[(Y_t - \mu) - \phi_1(Y_{t-1} - \mu)]^2.$$

First, we differentiate with respect to μ and ϕ_1, equate to zero and solve for $\hat{\mu}$ and $\hat{\phi}_1$. Now, we can show that

$$\begin{aligned}
\frac{\partial Q}{\partial \mu}\Big|_{\hat{\mu},\hat{\phi}_1} &= 2(\hat{\phi}_1 - 1)\sum_{t=2}^{n}[(Y_t - \hat{\mu}) - \hat{\phi}_1(Y_{t-1} - \hat{\mu})] = 0 \\
\frac{\partial Q}{\partial \phi_1}\Big|_{\hat{\mu},\hat{\phi}_1} &= 2\sum_{t=2}^{n}[(Y_t - \hat{\mu}) - \hat{\phi}_1(Y_{t-1} - \hat{\mu})](Y_{t-1} - \hat{\mu}) = 0.
\end{aligned}$$

Solving these equations we obtain

$$\begin{aligned}
\hat{\mu} &= \frac{\sum_{t=2}^{n}(Y_t - \hat{\phi}_1 Y_{t-1})}{(n-1)(1-\hat{\phi}_1)} \\
\hat{\phi}_1 &= \frac{\sum_{t=2}^{n}(Y_t - \hat{\mu})(Y_{t-1} - \hat{\mu})}{\sum_{t=2}^{n}(Y_t - \hat{\mu})^2}.
\end{aligned}$$

The solution indicates that $\hat{\mu}$ can only be computed on the condition that $\hat{\phi}_1$ is known. Similarly, the expression for $\hat{\phi}_1$ is conditional on $\hat{\mu}$ being known. Thus, the approach is sometimes referred to as conditional least squares method. Now, if n is large then

$$\frac{\sum_{t=2}^{n} Y_t}{n-1} \approx \frac{\sum_{t=2}^{n} Y_{t-1}}{n-1} \approx \bar{Y}.$$

Therefore, when n is large $\hat{\mu}$ simplifies to

$$\hat{\mu} \approx \bar{Y}\frac{(1-\hat{\phi}_1)}{(1-\hat{\phi}_1)} = \bar{Y}$$

and will no longer be conditional on known $\hat{\phi}_1$. Also, when n is large, $\hat{\phi}_1$ will be approximately equal to r_1. That is,

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n}(Y_t - \hat{\mu})(Y_{t-1} - \hat{\mu})}{\sum_{t=2}^{n}(Y_t - \hat{\mu})^2} \approx r_1,$$

when n is large (only the first term is missing).

10.2.2 MA(1)

We have noted earlier that an invertible $MA(1)$ model

$$Y_t = a_t - \theta_1 a_{t-1},$$

can be written as a convergent AR series of the form

$$Y_t + \sum_{j=1}^{\infty} \theta_1^j Y_{t-j} = a_t.$$

Thus, the least squares approach of minimizing

$$Q = \sum_t a_t^2 = \sum_t [Y_t + \sum_{j=1}^{\infty} \theta_1^j Y_{t-j}]^2$$

with respect to the parameters will automatically lead to a nonlinear optimization problem in θ_1 which can be solved using numerical techniques such as the Newton-Raphson method. One approach proceeds as follows. We begin by assuming that $a_0 = 0$. Then, given n observations $y_1, y_2, \ldots, y_n$ of a time series Y_t we compute

$$\begin{aligned} a_1 &= y_1 \\ a_2 &= y_2 + \theta_1 a_1 \\ a_3 &= y_3 + \theta_1 a_2 \\ \cdots &= \cdots + \cdots \\ \cdots &= \cdots + \cdots \\ a_n &= y_n + \theta_1 a_{n-1}. \end{aligned}$$

Next, we calculate $Q = \sum_{t=1} a_t^2$ over a grid of the invertible range $(-1, 1)$ and search for the value of θ_1 within this range that gives the smallest sum of squares. A similar approach can be applied to estimate the parameters of more general $MA(q)$ models.

10.3 Maximum Likelihood Estimation (MLE)

In this section, we assume that the white noise process a_t follows the normal distribution with mean 0 and variance σ_a^2. Then, the density function of a_t can be written as

$$f(a_t) = (2\pi\sigma_a^2)^{-1/2} exp\left\{-\frac{a_t^2}{2\sigma_a^2}\right\}, \ |a_t| < \infty.$$

Before discussing the general ML approach, let us consider the simple $AR(1)$ model given by

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + a_t.$$

Since Y_0 is unknown, we can use the model to write

$$\begin{aligned} a_2 &= (Y_2 - \mu) - \phi_1(Y_1 - \mu) \\ a_3 &= (Y_3 - \mu) - \phi_1(Y_2 - \mu) \\ \cdots &= \cdots + \cdots \\ \cdots &= \cdots + \cdots \\ a_n &= (Y_n - \mu) - \phi_1(Y_{n-1} - \mu). \end{aligned}$$

Then, conditional on $a_1 = Y_1$, we can write the joint probability density function (pdf) of $a_2, a_3, \ldots, a_n$ as

$$\begin{aligned} f(a_2, a_3, \ldots, a_n | a_1 = y_1) &= f(y_2, y_3, \ldots, y_n | y_1) \\ &= (2\pi\sigma_a^2)^{-(n-1)/2} exp\left\{ -\frac{1}{2\sigma_a^2} \sum_{t=2}^{n} [(Y_t - \mu) - \phi_1(Y_{t-1} - \mu)]^2 \right\}. \end{aligned}$$

It follows that we can write the joint pdf of $Y_1, \ldots, Y_n$ as

$$L(\phi_1, \mu, \sigma_a^2) = f(y_2, y_3, \ldots, y_n | y_1) f(y_1).$$

Now, we recall that a stationary $AR(1)$ process has mean $E(Y_t) = \mu$ and variance $V(Y_t) = \sigma_a^2/(1 - \phi_1^2)$. This implies that $Y_1 \sim N(\mu, V(Y_1))$ with density

$$f(y_1) = (2\pi\sigma_a^2)^{-1/2}(1 - \phi_1^2)^{1/2} exp\left\{ -\frac{(1 - \phi_1^2)(y - \mu)^2}{2\sigma_a^2} \right\}, \ |a_t| < \infty.$$

Then substituting for $f(y_2, y_3, \ldots, y_n | y_1)$ and $f(y_1)$ and simplifying we obtain the likelihood function for an $AR(1)$ model to be

$$L(\phi_1, \mu, \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2}(1 - \phi_1^2)^{1/2} exp\left\{ -\frac{S(\phi_1, \mu)}{2\sigma_a^2} \right\},$$

where

$$S(\phi_1, \mu) = \sum_{t=2}^{n} [(Y_t - \mu) - \phi_1(Y_{t-1} - \mu)]^2 + (1 - \phi_1^2)(Y_1 - \mu)^2,$$

is the unconditional sum of squares function. The log likelihood

$$logL(\phi_1, \mu, \sigma_a^2) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln\sigma_a^2 + \frac{1}{2} ln(1 - \phi_1^2) - \frac{S(\phi_1, \mu)}{2\sigma_a^2},$$

is then maximized numerically to obtain the maximum likelihood estimator of the parameters. It can be shown that the MLE of σ_a^2 is

$$\hat{\sigma}_a^2 = \frac{S(\hat{\phi}_1, \hat{\mu})}{n},$$

where $\hat{\phi}_1$ and $\hat{\mu}$ are the MLEs of ϕ_1 and μ respectively. Then,

$$logL(\hat{\phi}_1, \hat{\mu}, \hat{\sigma}_a^2) = -\frac{n}{2} ln\hat{\sigma}_a^2 - \frac{n}{2}\{1 + ln(2\pi)\} + \frac{1}{2} ln(1 - \hat{\phi}_1^2). \tag{18}$$

10.4 Empirical Examples

Example 1

In this example we refer to the weekly sales of absorbent paper towels in Figure 25. After examining the time plot, we transformed the series into a stationary one by taking the first difference. We then computed the SACF and SPACF of the first difference. The patterns in the SACF and SPACF appear to match those of an $AR(1)$ or $MA(1)$ process in that the SACF cuts off after lag 1 except for the SACF at lag 18, while the SPAC cuts off after lag 1. Therefore, two tentative model for the first difference Z_t is

$$\begin{aligned} \widetilde{Z}_t &= \phi_1 \widetilde{Z}_{t-1} + a_t \\ \widetilde{Z}_t &= a_t - \theta_1 a_{t-1}. \end{aligned}$$

where $\widetilde{Z}_t = Z_t - \mu$ and a_t is a normally distributed white noise process. The $AR(1)$ model can also be written as

$$Z_t = \mu^* + \phi_1 Z_{t-1} + a_t,$$

where $\mu^* = (1 - \phi_1)\mu$ is the intercept. In terms of Y_t we have

$$(1 - \mathbf{B})Y_t = \mu^* + \phi_1(1 - \mathbf{B})Y_{t-1} + a_t,$$

which is referred to as an autoregressive integrated model of order (1,1) and denoted by $ARI(1, 1)$. Theoretically, we know that $\phi_{kk} = \phi_1$. Our initial guess for ϕ_1 is therefore $\hat{\phi}_1^{(0)} = r_{11} = 0.3066$, whereas our initial guess for θ_1 is that $\hat{\theta}_1^{(0)} < 0$, since $r_1 > 0$.

Estimation

Once tentative models have been identified, the next step is to estimate the parameters of the models. We have discussed three methods for parameter estimation. We will however use the *arima* command in R to obtain maximum likelihood estimates of the parameters.

The format for the *arima* command is *arima(tsdata, order = c(p,d,q), seasonal = list(order = c(p,d,q), period = L), include.mean = TRUE)*. In order to obtain a meaningful estimate for the parameters of an ARMA model we recommend subtracting the mean $\bar{z}$ from each observation and then computing the estimates by setting "include.mean = FALSE". In this example, $\hat{\mu} = \bar{z} \approx 0.0054$. Using the data $\widetilde{z}_t = z_t - \bar{z}$ and the R software, we obtain the following maximum likelihood estimates of the parameters:

$$\begin{aligned} \hat{\phi}_1 &= 0.3050,\ s.e.(\hat{\phi}_1) = 0.0869, \text{ and } \hat{\sigma}_a^2 = 1.095, \\ \hat{\theta}_1 &= -0.3518,\ s.e.(\hat{\theta}_1) = 0.08, \text{ and } \hat{\sigma}_a^2 = 1.071, \end{aligned}$$

with $AIC = 352.58$ for the $AR(1)$ model and $AIC = 349.98$ for the $MA(1)$ model. The Akaike Information Criterion (AIC) will be discussed in greater details in the next section. Notice that the final estimate for ϕ_1 is very close to our initial guess. Thus, the fitted models are

$$\begin{aligned} \hat{\widetilde{Z}}_t &= 0.305\widetilde{Z}_{t-1} + a_t, \\ \hat{\widetilde{Z}}_t &= a_t + 0.3518a_{t-1}, \end{aligned}$$

where $\hat{\widetilde{Z}}_t = \hat{Z}_t - 0.0054$.

11 Model Diagnostics

Next, we evaluate the adequacy of our models by examining the validity of the assumptions of the fitted model. We had assumed that (i) the process a_t is normally distributed with zero mean and constant variance and (ii) that a_t's are uncorrelated. A simple analysis of the normal probability plot of the innovations (residuals0 from the fitted model can be used to assess the validity of the normality assumption. To assess whether the innovations are uncorrelated, we examine a plot of the sample autocorrelation function and SPACF of the residuals. The SACF and SPACF at all lags should not be significant if this assumption is valid. In Figures 36, 37 and 38, we see that the SACF and SPACF of the two models are not significant at all lags.

Another useful correlation test is the portmanteau lack of fit test which uses the Ljung-Box Chi-Squared test statistic

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

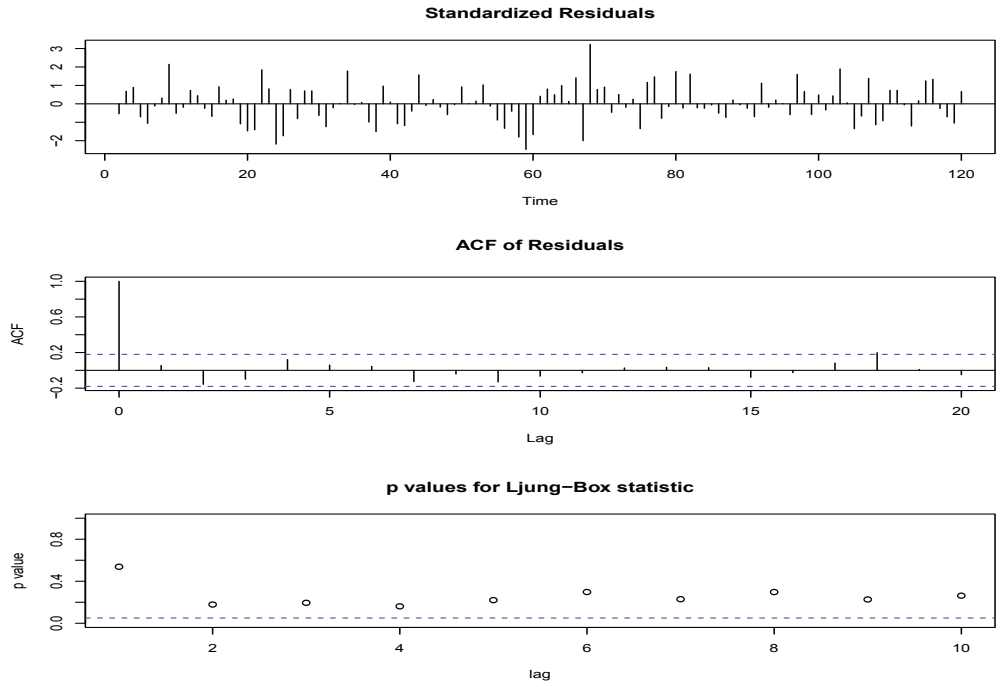

Figure 36: Graphs for model diagnostics: $AR(1)$ model in Example 1.

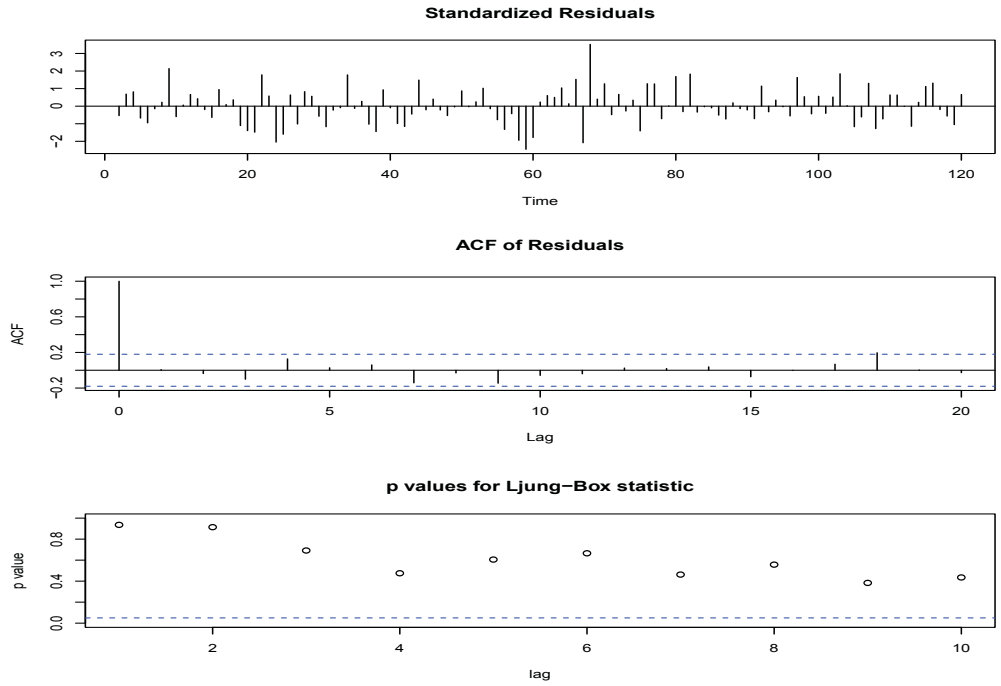

Figure 37: Graphs for model diagnostics: $MA(1)$ model in Example 1.

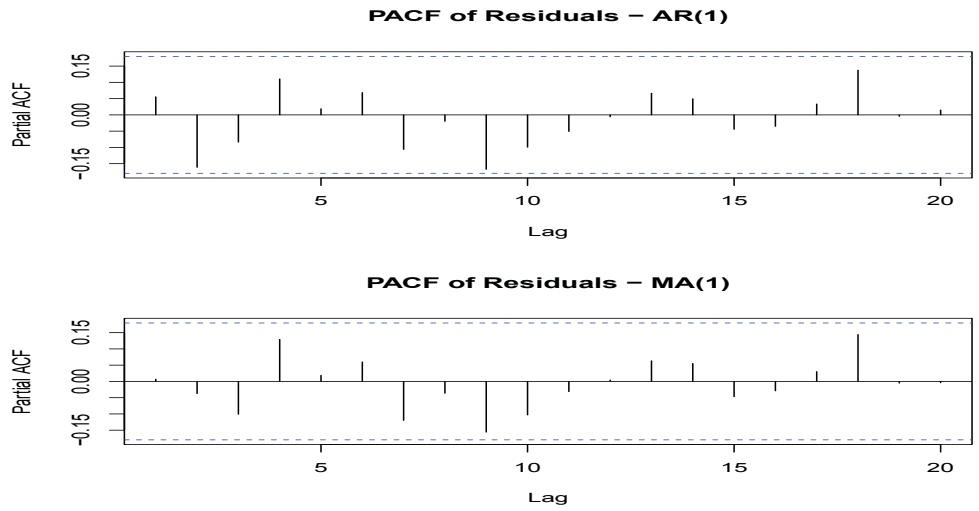

Figure 38: Graphs of SPACF of innovations from fitted $AR(1)$ and $MA(1)$ models in Example 1.

to test the null hypothesis

$$H_0 : \rho(1) = \rho(2) = \cdots = \rho(K) = 0.$$

Under H_0, the test statistic approximately follows the χ^2 distribution with $(K - m)$ degrees of freedom, where m = number of parameters estimated in the model (see *Ansley, C. F. and Newbold, P. (1979). On the finite sample distribution of residual autocorrelations in autoregressive moving average models. Biometrika,* **66***, 547-554*). In our example $m = 1$. The p-values for $K = 2, 3, \ldots, 10$ are shown in Figures 36 and 37. Large p-values indicate strong evidence in favour of H_0. That is, the residuals are uncorrelated. Figures 36 and 37 show that the p-values for this test, in Example 1, are all larger than 0.05 indicating strong evidence in favour of H_0. Based on these results we conclude that the $AR(1)$ and $MA(1)$ models are adequate for modelling the weekly sales of rolls of paper towels.

Example 2

In this example, we demonstrate how model diagnostics can assist us in chosing the parameters to add to a model in order to obtain a better model that satisfies the model assumptions. The daily viscosity readings of the chemical product $XB - 77 - 5$ shown in Figure 29 appear to be stationary, therefore no transformation is required. The SACF and SPACF also do not indicate any nonstationary phenomena. A careful examination of the patterns in the SACF and SPACF show that the SACF tail off and cuts-off after lag 4 while the SPACF

cuts off after lag 2. These patterns matches that of an $AR(2)$ and a $MA(4)$ process. One may also wish to investigate the $ARMA(1,1)$ as a possible alternative. As before, we begin by computing the mean of the observed series then deleting the mean from the observations to obtain $\widetilde{Y}_t = Y_t - 34.93007$. By comparing the pattern in the plot of the SPACF with the theoretical patterns in Figure 28, we infer that $\hat{\phi}_1^{(0)} > 0$ and that $\hat{\phi}_2^{(0)} < 0$.

Estimation

Now, rather than fitting the AR(2) and MA(4) models, suppose an analyst erroneously identifies and fits AR(1) and MA(2) models to obtain

$$\begin{aligned}
\hat{\phi}_1 &= 0.5196,\ s.e.(\hat{\phi}_1) = 0.0975,\ \hat{\sigma}_a^2 = 5.394,\ \text{ and }\ AIC = 434.02,\\
\hat{\theta}_1 &= -0.7009,\ s.e.(\hat{\theta}_1) = 0.0987,\ \hat{\theta}_2 = -0.2634,\ s.e.(\hat{\theta}_2) = 0.1036,\\
\hat{\sigma}_a^2 &= 4.83,\ AIC = 425.73.
\end{aligned}$$

The model diagnostic plots in Figures 39 (for MA(2)), 40 (for AR(1)) and 41 clearly show that these models are not adequate. For the MA(2) model, the PACF plot of the innovations is significant at lag 4 whereas the ACF of the innovations from the AR(1) fit are significant at lags 2, 3, 4 and 6 and the PACF is significant at lags 2 and 3. In addition, the p-values from the Ljung-Box goodness of fit test are below 0.05 for the AR(1) model.

To improve on the MA(2) model we use the fact that the PACF of the innovations is significant at lag 4 to update the model by including an autoregressive component at lag 4. For the AR(1) model we can update the model in two ways. First, based on the fact that the ACF is significant at lags 2, 3, 4 and 6 one may add lags 2,3,4,and 6 moving average components to the AR(1) model or simply update the model to an AR(3) model because lags 2 and 3 of the PACF are significant. Using the concept of parsimony we have chosen the updated model with fewer parameters (i.e. the AR(3)). The updated models become

$$\begin{aligned}
\widetilde{Y}_t &= \phi_1\widetilde{Y}_{t-1} + \phi_2\widetilde{Y}_{t-2} + \phi_3\widetilde{Y}_{t-3} + a_t,\\
\widetilde{Y}_t &= \phi_4\widetilde{Y}_{t-4} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}.
\end{aligned}$$

Using the R software, the parameter estimates are as follows.

$$\begin{aligned}
\hat{\phi}_1 &= 0.5974,\ s.e.(\hat{\phi}_1) = 0.1038,\ \hat{\phi}_2 = -0.3032,\ s.e.(\hat{\phi}_2) = 0.1184,\\
\hat{\phi}_3 &= -0.2286,\ s.e.(\hat{\phi}_3) = 0.1095,\ \hat{\sigma}_a^2 = 4.335,\ \text{ and }\ AIC = 417.84,\\
\hat{\theta}_1 &= -0.7092,\ s.e.(\hat{\theta}_1) = 0.1074,\ \hat{\theta}_2 = -0.3104, s.e.(\hat{\theta}_2) = 0.1102,\\
\hat{\phi}_4 &= -0.2202,\ s.e.(\hat{\phi}_4) = 0.1148,\ \hat{\sigma}_a^2 = 4.646,\ \text{ and }\ AIC = 424.18.
\end{aligned}$$

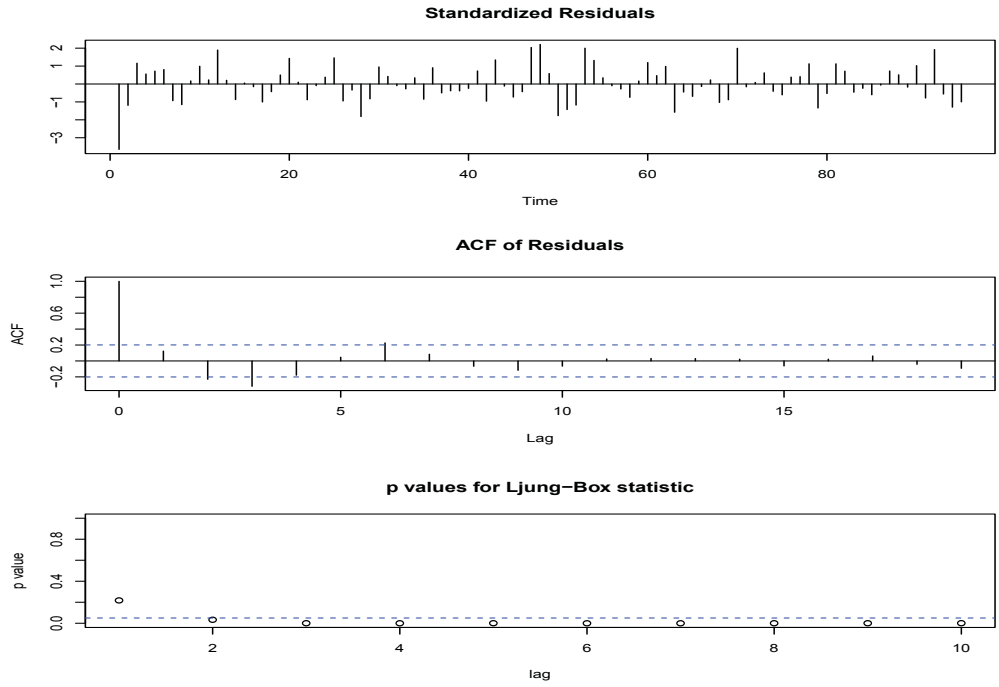

Figure 39: Graphs for model diagnostics: $AR(1)$ model in Example 2.

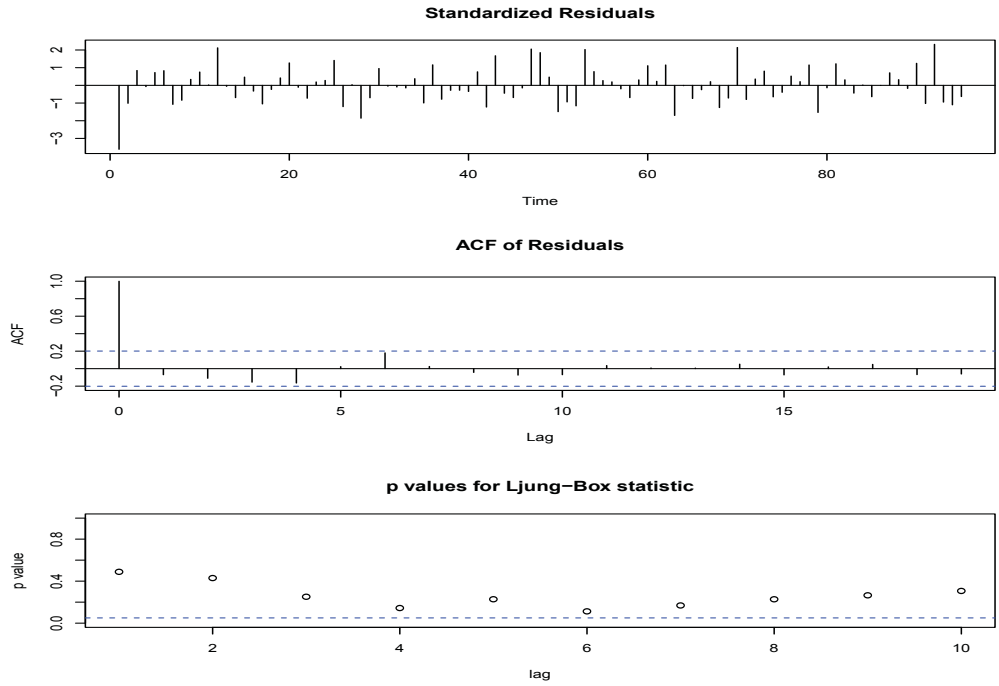

Figure 40: Graphs for model diagnostics: $MA(2)$ model in Example 2.

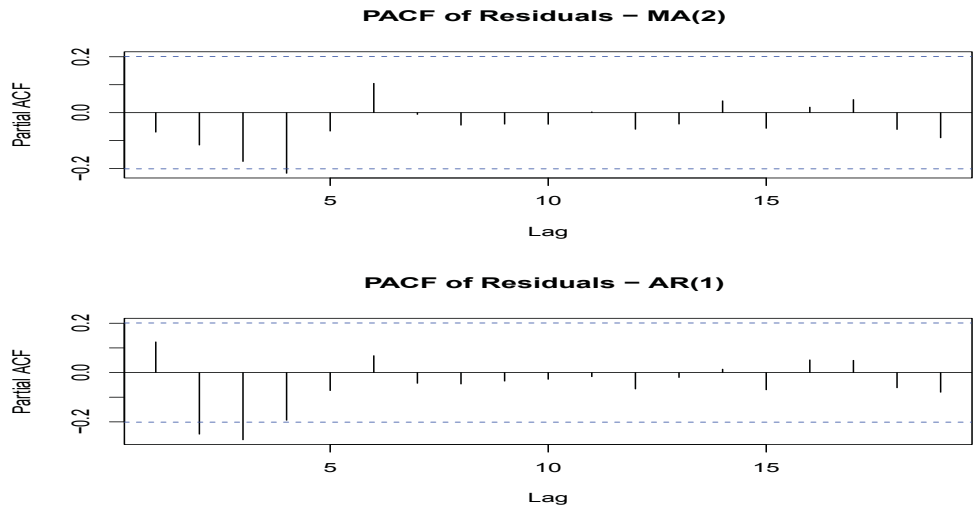

Figure 41: Graphs of SPACF of innovations from fitted $AR(1)$ and $MA(2)$ models in Example 2.

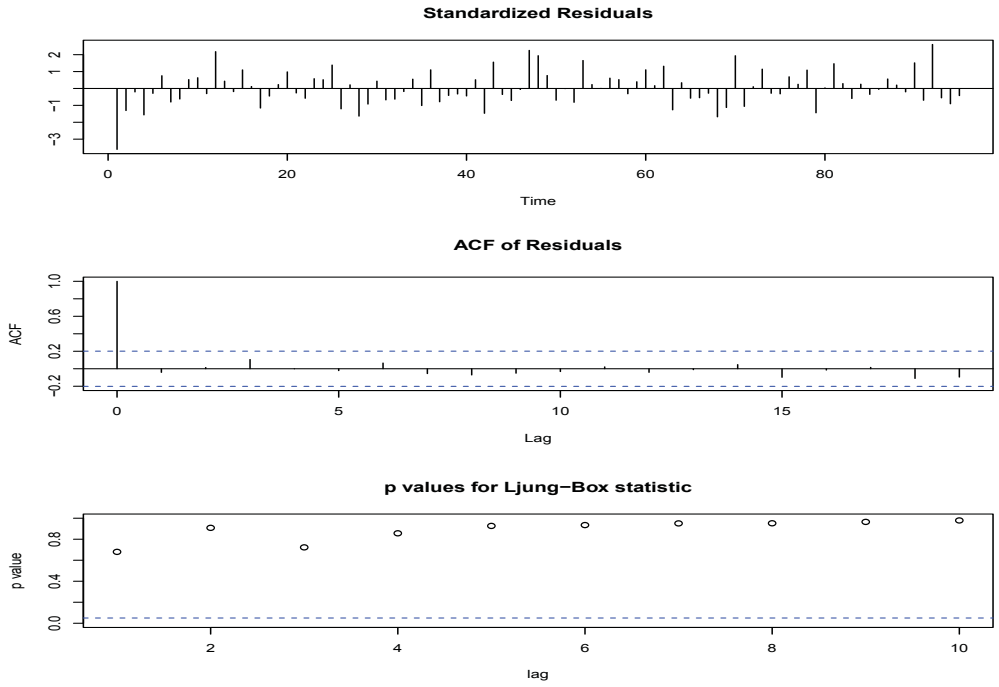

Figure 42: Graphs for model diagnostics: $AR(3)$ model in Example 2.

The model diagnostic plots for the updated models support the conclusion that the new models are adequate. In addition the updated models have smaller estimated innovation variances and AIC values when compared to the original AR(1) and MA(2) models. Since

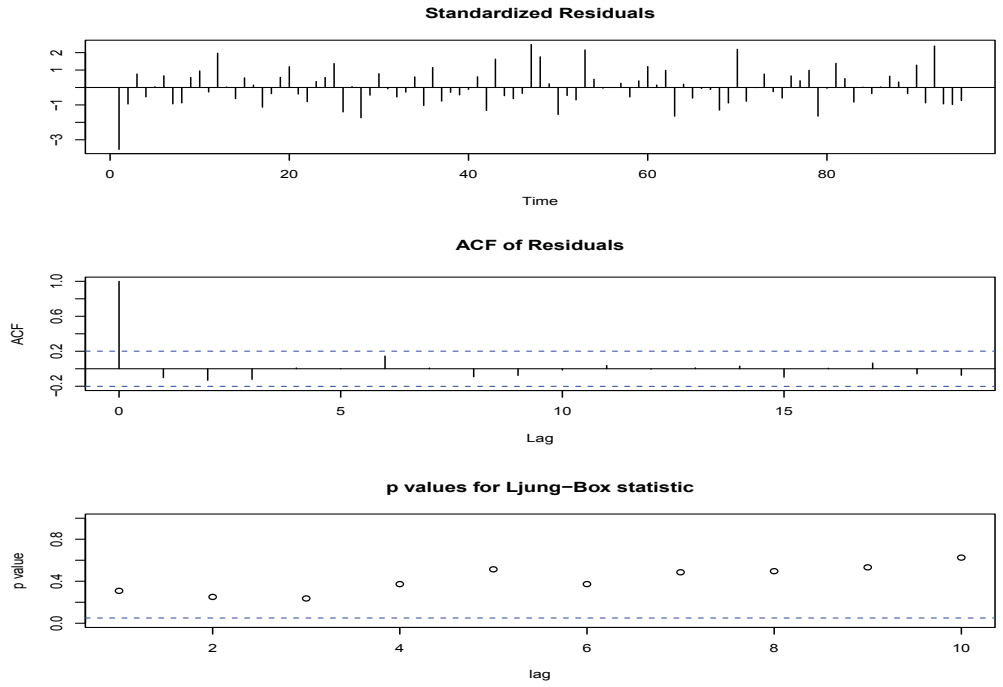

Figure 43: Graphs for model diagnostics: updated $MA(2)$ model in Example 2.

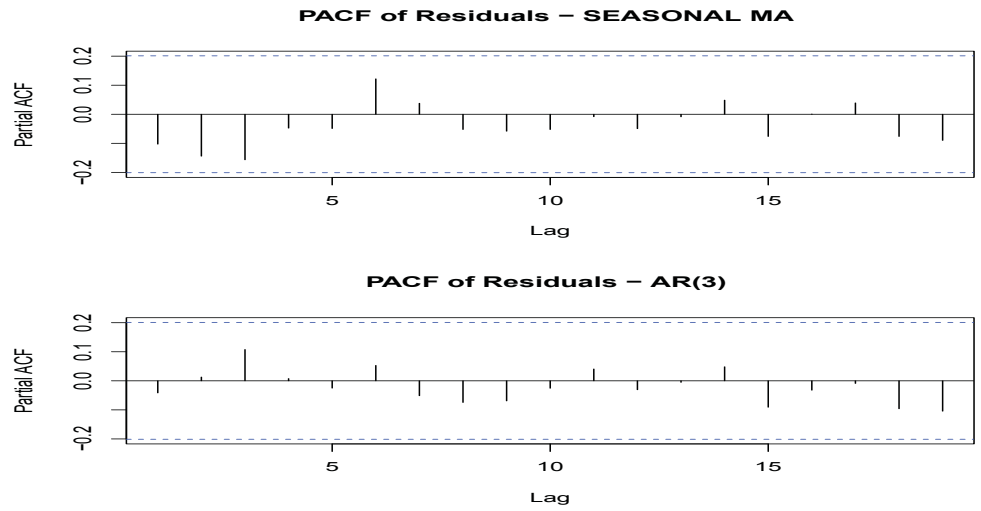

Figure 44: Graphs of SPACF of innovations from fitted $AR(3)$ and updated $MA(2)$ models in Example 2.

there are two models to chose from we now discuss statistical tools for model selection.

12 Model Selection

In the examples we discussed in §10, there were at least two tentative models that were under consideration for each series. Thus, one have to decide on which of the two models is best for each data. A commonly used model selection criterion is the Akaike Information Criterion (AIC) defined as

$$AIC = -2ln(maximumlikelihood) + 2M,$$

where M is the number of estimated parameters in the model. The model with the smaller AIC is considered to be the best model.

In §10.3 we found that the unconditional log likelihood for an $AR(1)$ is given by (18). Now, using only the conditional density obtained in §10.3, we can also show that the conditional log likelihood is given by

$$lnL(\hat{\phi}_1, \hat{\mu}, \hat{\sigma}_a^2) = -\left(\frac{n-1}{2}\right) ln\hat{\sigma}_a^2 - \left(\frac{n-1}{2}\right)\{1 + ln(2\pi)\}. \tag{19}$$

In this case, the AIC becomes

$$AIC = (n-1)ln\hat{\sigma}_a^2 + 2M + K,$$

where $K = \frac{n-1}{2}\{1 + ln(2\pi)\}$ is a constant for any ARMA model. For the unconditional log likelihood the AIC is

$$AIC = nln\hat{\sigma}_a^2 + 2M + K + ln(1 - \hat{\phi}_1^2).$$

Since K is constant for all models it is usually not included in the computation of the AIC.

Shibata (1976) found that the AIC typically overestimates the autoregressive component of an ARIMA model. As a result, Akaike (1978) developed the Bayesian Information criterion (BIC) defined by

$$AIC = nln\hat{\sigma}_a^2 - (n-M)ln\left(1 - \frac{M}{n}\right) + Mlnn + Mln\left[\left(\frac{\hat{\sigma}_y^2}{\hat{\sigma}_a^2} - 1\right)/M\right],$$

where $\hat{\sigma}_y^2$ is the sample variance of the observed series, M is the number of parameters and $\hat{\sigma}_a^2$ is the estimated innovation variance. Again, the model with the minimum BIC is typically selected as the best model.

Example 3

The yearly U.S. tobacco production data in Figure 1(b) is an example of a series that is nonstationary in mean and variance. First, we stabilize the variance with a logarithmic transformation. Then, we take the first difference of the transformed series in order to eliminate the increasing trend. A plot of the series and the SACF and SPACF values (see Figure 32) show that the result of applying these two transformations is a stationary series. The patterns in the SACF and SPACF values lead us to suggest two tentative models, namely, $MA(1)$ and $AR(3)$ for the tobacco series. Let $W_t = lnY_t$ and $U_t = (1 - \boldsymbol{B})W_t$ be the first difference of W_t with sample mean $\bar{u} = 0.014998$.

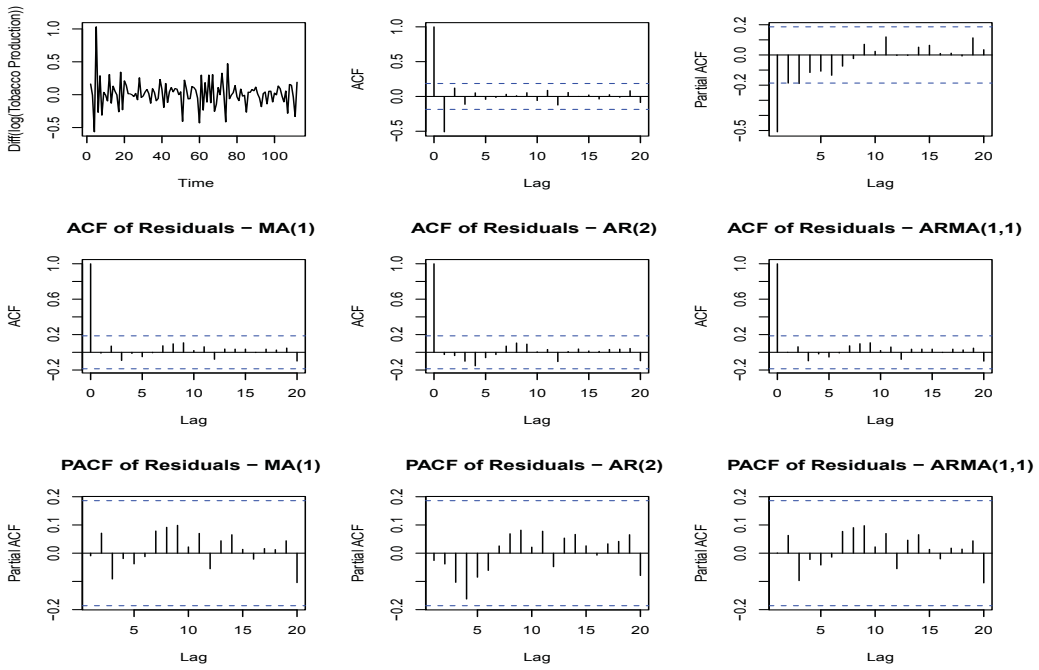

Figure 45: Graphs of SACF, SPACF Tobacco Production Series in Figure 1(b) and diagnostic plots from fitted $MA(1)$, $AR(3)$ and $ARMA(1,1)$ models in Example 3.

Proceeding as before, we use the *arima* command in R to obtain the following estimates of the model parameters. For the MA(1) model the parameter estimates obtained were

$$\hat{\theta}_1 \;=\; 0.6897,\; s.e.(\hat{\theta}_1) = 0.0698, \;\text{ and } \hat{\sigma}_a^2 = 0.02624,$$

with $AIC = -84.42$. For the AR(3) model we obtain

$$\begin{aligned}\hat{\phi}_1 &= -0.6346,\; s.e.(\hat{\phi}_1) = 0.0930,\; \hat{\phi}_2 = -0.3022,\; s.e.(\hat{\phi}_2) = 0.1106,\\ \hat{\phi}_3 &= -0.1943,\; s.e.(\hat{\phi}_1) = 0.0998,\;\; \hat{\sigma}_a^2 = 0.02699,\end{aligned}$$

with $AIC = -77.48$. We also examined whether an ARMA(1,1) model will be suitable for the stationary series. The estimates of the parameters of the ARMA(1,1) model are

$$\begin{aligned}\hat{\phi}_1 &= -0.0196,\ s.e.(\hat{\phi}_1) = 0.1389,\\ \hat{\theta}_1 &= 0.6791,\ s.e.(\hat{\theta}_1) = 0.1056,\ \hat{\sigma}_a^2 = 0.02624,\end{aligned}$$

with $AIC = -82.44$. The ACF and PACF plots of the innovations in Figure 45 clearly shows that all three models are adequate for the series. However, based on the minimum AIC criterion and the concept of parsimony, the MA(1) model is the best model. In terms of the original series Y_t the fitted model becomes

$$(1 - \boldsymbol{B})lnY_t = 0.01499 + a_t - 0.6897a_{t-1}.$$

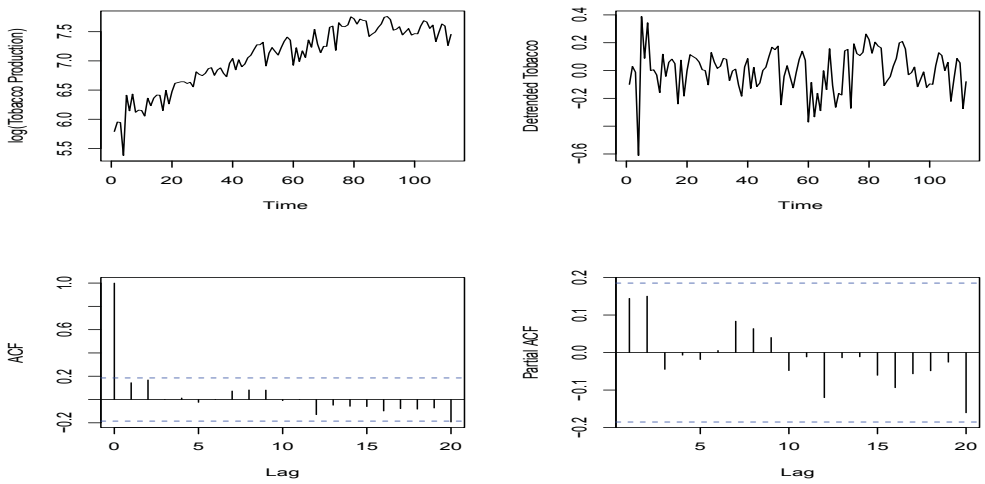

Figure 46: Graph of log(Tobacco series), Detrended log of Tobacco Production Series and SACF and SPACF of detrended series.

Instead of taking first differences of the transformed tobacco series, an analyst may wish to use a quadratic regression model to estimate the trend in the transformed series before computing the SACF and SPACF of the stationary series $U_t = lnY_t - tr_t$, where

$$tr_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 = 5.8557 + 0.0344t - 0.0001735t^2,$$

is the estimated trend curve. A plot of lnY_t, the observed values of U_t and the SAC and SPAC values of U_t is shown in Figure 46. These plots suggests that the detrended series is simply a white noise process since none of the SAC and SPAC values are significant. Based on this approach, the model for the tobacco data is

$$W_t = tr_t + a_t.$$

13 FORECASTING

One of the main objectives of analyzing a time series is for the purpose of forecasting future values Y_{t+l}, $l = 1, 2, \ldots$ of the series. In this section, we shall denote the forecast of Y_{t+l} by $\hat{Y}_t(l)$, referred to as the l-step ahead forecast where t is the forecast origin and l is the lead time. For instance, we had $n = 120$ observed values of weekly sales in Example 1, Section 10.4 which we used to fit the $AR(1)$ and $MA(1)$ models. We may wish to use the fitted models to compute the forecast $\hat{Y}_120(1)$ of Y_{121}. In this case, $l = 1$ is the lead time and $n = 120$ the forecast origin. Predictions are usually made using the "best" fitted model for the time series under consideration. It is also a good idea to compute the standard error SE_{n+l} of each forecasted value in order to assess the accuracy or reliability of our forecast. This is similar to the standard error of the predicted value for a new observation in regression analysis. Using the SE_{n+l} we can then compute a $(1-\alpha)100\%$ prediction interval for the new value Y_{n+l}.

Given $Y_1, Y_2, \ldots, Y_t$, it can be shown that the minimum mean square error (MSE) l-step ahead forecast is

$$\hat{Y}_t(l) = E(Y_{t+l}|Y_1, Y_2, \ldots, Y_t).$$

We shall discuss some of the basic ideas of forecasting by looking at various models we have discussed.

13.1 Deterministic Trend and Seasonal Models

Suppose a time series Y_t has a trend T_t, a seasonal S_t and a random component U_t with $E(U_t) = \mu$ and variance σ_u^2. Then, we can write

$$Y_t = T_t + S_t + U_t.$$

Now, the forecast function of the time series becomes

$$\hat{Y}_t(l) = E(Y_{t+l}|Y_1, Y_2, \ldots, Y_t) = T_{t+l} + S_{t+l} + E(U_{t+l}|Y_1, Y_2, \ldots, Y_t).$$

We note that $U_{t+l} = Y_{t+l} - T_{t+l} - S_{t+l}$ is independent of $Y_1, \ldots, Y_t$ provided $l \geq 1$. Then,

$$\hat{Y}_t(l) = T_{t+l} + S_{t+l} + \mu, \quad l \geq 1.$$

The forecast error $e_t(l)$ is given by

$$e_t(l) = Y_{t+l} - \hat{Y}_t(l) = (T_{t+l} + S_{t+l} + U_t) - (T_{t+l} + S_{t+l} + \mu) = U_t - \mu.$$

So that,

$$E[e_t(l)] = E[Y_{t+l} - \hat{Y}_t(l)] = E(U_t) - \mu = 0.$$

That is, the forecasts are unbiased with constant variance

$$Var[e_t(l)] = Var[Y_{t+l} - \hat{Y}_t(l)] = Var(U_t) = \sigma_u^2,$$

for all lead times l.

13.2 ARMA Forecasting

For the ARMA(p,q) model the future value to be forecasted can be written as

$$Y_{t+l} = \mu + \phi_1(Y_{t+l-1} - \mu) + \phi_2(Y_{t+l-2} - \mu) + \cdots + \phi_p(Y_{t+l-p} - \mu) + a_{t+l} - \theta_1 a_{t+l-1} - \cdots - \theta_q a_{t+l-q}.$$

Then, using the definition of the l-step ahead forecast, the forecasting function for the ARMA(p,q) model can be written as a recursion as follows

$$\begin{aligned} \hat{Y}_t(l) &= \mu + \phi_1(\hat{Y}_t(l-1) - \mu) + \phi_2(\hat{Y}_t(l-2) - \mu) + \cdots + \phi_p(\hat{Y}_t(l-p) - \mu) \\ &+ E(a_{t+l}|Y_1, Y_2, \ldots, Y_t) - \theta_1 E(a_{t+l-1}|Y_1, Y_2, \ldots, Y_t) - \cdots - \theta_q E(a_{t+l-q}|Y_1, Y_2, \ldots, Y_t) \end{aligned}$$

Now, using the fact that $E(Y_1|Y_1, Y_2, \ldots, Y_t) = Y_1$, $E(Y_2|Y_1, Y_2, \ldots, Y_t) = Y_2$ and so on, we have that, in general, $\hat{Y}_t(j) = E(Y_{t+j}|Y_1, Y_2, \ldots, Y_t) = Y_{t+j}$ if $j \leq 0$. Also, using the fact that the residuals from a fitted model are known for $t = 1, 2, \ldots, n$ and the fact that a_t is independent of Y_s if $t > s$ we obtain the approximation $E(a_{t+j}|Y_1, Y_2, \ldots, Y_t) = E(a_{t+j}) = 0$, if $j > 0$ and $E(a_{t+j}|Y_1, Y_2, \ldots, Y_t) = a_{t+j}$, if $j \leq 0$. That is,

$$E(a_{t+j}|Y_1, Y_2, \ldots, Y_t) = \begin{cases} 0, & \text{if } j > 0 \\ a_{t+j}, & \text{if } j \leq 0. \end{cases}$$

As an example, we consider special cases of the general ARMA(p,q) model.

13.2.1 ARMA(1,0) OR AR(1)

Clearly, when $p = 1$ and $q = 0$ we obtain the forecast function for an $AR(1)$ model as

$$\hat{Y}_t(l) = \mu + \phi_1(\hat{Y}_t(l-1) - \mu) + E(a_{t+l}|Y_1, Y_2, \ldots, Y_t).$$

Considering the case $l = 1$ we have that

$$\begin{aligned} \hat{Y}_t(1) &= \mu + \phi_1(\hat{Y}_t(0) - \mu) + E(a_{t+1}|Y_1, Y_2, \ldots, Y_t) \\ &= \mu + \phi_1(Y_t - \mu), \end{aligned}$$

since $\hat{Y}_t(0) = E(Y_t|Y_1, Y_2, \ldots, Y_t) = Y_t$ and by independence $E(a_{t+1}|Y_1, Y_2, \ldots, Y_t) = E(a_{t+1}) = 0$. The forecast error for the one-step ahead forecast is then

$$e_t(1) = Y_{t+1} - \hat{Y}_t(1) = [\mu + \phi_1(Y_t - \mu) + a_{t+1}] - [\mu + \phi_1(Y_t - \mu)] = a_{t+1}.$$

Once again, we see that the one-step ahead forecast is unbiased since

$$E(e_t(1)) = E(Y_{t+1} - \hat{Y}_t(1)) = E(a_{t+1}) = 0,$$

with variance $Var(e_t(1)) = \sigma_a^2$.

In general, we use the recursive expression

$$\begin{aligned} \hat{Y}_t(l) &= \mu + \phi_1(\hat{Y}_t(l-1) - \mu) + E(a_{t+l}|Y_1, Y_2, \ldots, Y_t), \\ &= \mu + \phi_1(\hat{Y}_t(l-1) - \mu), \quad l \geq 1, \end{aligned}$$

to obtain

$$\begin{aligned} \hat{Y}_t(l) &= \mu + \phi_1^2(\hat{Y}_t(l-2) - \mu) = \mu + \phi_1^3(\hat{Y}_t(l-3) - \mu) \\ &= \cdots = \mu + \phi_1^l(Y_t - \mu). \end{aligned}$$

Then, the forecast error becomes

$$\begin{aligned} e_t(l) &= Y_{t+l} - \hat{Y}_t(l) = \mu + \phi_1(Y_{t+l-1} - \mu) + a_{t+l} - \mu - \phi_1(\hat{Y}_t(l-1) - \mu) \\ &= \phi_1(Y_{t+l-1} - \hat{Y}_t(l-1)) + a_{t+l} = \phi_1 e_t(l-1) + a_{t+l}. \end{aligned}$$

Applying the recursion successively and noting that $e_t(0) = Y_t - \hat{Y}_t(0) = 0$ we can write the forecast error as

$$e_t(l) = \phi_1^{l-1} a_{t+1} + \cdots + \phi_1^2 a_{t+l-2} + \phi_1 a_{t+l-1} + a_{t+l}.$$

It is then clear that $E[e_t(l)] = 0$ and hence the forecast is unbiased with variance

$$Var[e_t(l)] = \sigma_a^2 \sum_{j=0}^{l-1} \phi_1^{2j} = \sigma_a^2 \left[\frac{1 - \phi_1^{2l}}{1 - \phi_1^2}\right].$$

We observe that if the AR(1) process is stationary, then as $l \longrightarrow \infty$, $\phi_1^{2l} \longrightarrow 0$ and

$$Var[e_t(l)] \longrightarrow \left[\frac{\sigma_a^2}{1 - \phi_1^2}\right].$$

Example 4

We use the series on the daily average number of truck manufacturing defects in Figure 26 to illustrate the concept of forecasting. For the purpose of validating the model we remove the last 3 observations and use the first 42 observations to fit the model and forecast the last 3 observations. The SACF and SPACF plots in Figure 26 suggests that an $AR(1)$ and a $MA(1)$ models are suitable tentative models for the series. The AIC obtained from fitting the two models are 61.05 and 63.25 respectively. The model diagnostic plots also indicate that both models are adequate. Therefore, we select the $AR(1)$ model since it has the smallest AIC value. The estimated $AR(1)$ model is

$$Y_t = 1.7902 + 0.4318(Y_{t-1} - 1.7902) + a_t,$$

with $\sigma_a^2 = 0.2266$. Now, the forecasting function for an $AR(1)$ model is

$$\hat{Y}_t(l) = \mu + \phi_1(\hat{Y}_t(l-1) - \mu), \quad l \geq 1.$$

Thus, the one step ahead forecast for the truck manufacturing defects series is

$$\hat{Y}_{42}(1) = 1.7902 + 0.4318(\hat{Y}_{42}(0) - 1.7902) = 1.7902 + 0.4318(1.79 - 1.7902) \approx 1.790114.$$

The variance of the 1-step ahead forecast is $\sigma_a^2 = 0.2266$. Next, we compute the forecast for Y_{44}, the 2-step ahead forecast. Following the formula for the forecasting function we have that

$$\hat{Y}_{42}(2) = 1.7902 + 0.4318(\hat{Y}_{42}(1) - 1.7902) = 1.7902 + 0.4318(1.790114 - 1.7902) \approx 1.790163,$$

with forecast variance

$$\hat{Var[e_t(2)]} = \hat{\sigma}_a^2 \left[\frac{1 - \hat{\phi}_1^4}{1 - \hat{\phi}_1^2}\right] = 0.2266 \frac{1 - 0.4318^2}{1 - 0.4318} \approx 0.3244.$$

Similarly, we find that the forecast for Y_{45}, the 3-step ahead forecast is

$$\hat{Y}_{42}(3) = 1.7902 + 0.4318(\hat{Y}_{42}(2) - 1.7902) = 1.7902 + 0.4318(1.790163 - 1.7902) \approx 1.790184,$$

where

$$\hat{Var[e_t(3)]} = \hat{\sigma}_a^2 \left[\frac{1 - \hat{\phi}_1^6}{1 - \hat{\phi}_1^2}\right] = 0.2266 \frac{1 - 0.4318^3}{1 - 0.4318} \approx 0.3667.$$

In R, one can use the command *predict(arima object, n.ahead = l)* to obtain forecasts.

13.2.2 Prediction Limits

Assuming that the white noise process a_t is normally distributed, then the forecast error $e_t(l) = Y_{t+l} - \hat{Y}_t(l)$ will be normally distributed with mean 0 and variance σ_e^2. By standardizing the normal random variable $e_t(l)$ we obtain a standard normal random variable given by

$$Z = \frac{Y_{t+l} - \hat{Y}_t(l)}{\sqrt{\sigma_e^2}}.$$

Given a confidence level $1 - \alpha$ and the standard normal percentile $z_{\alpha/2}$, we have that

$$P\left(-z_{\alpha/2} < \frac{Y_{t+l} - \hat{Y}_t(l)}{\sqrt{\sigma_e^2}} < z_{\alpha/2}\right) = 1 - \alpha.$$

After some algebra, we find that a $(1 - \alpha)100\%$ prediction limits for the future value Y_{t+l} are

$$\hat{Y}_t(l) \pm z_{\alpha/2}\sqrt{\sigma_e^2}.$$

Applying this to the defects in manufactured trucks example we have that a 95% confidence prediction limits for the 2-step ahead forecast of the future value Y_{42+2} are

$$1.790163 \pm 1.96\sqrt{0.3244} \text{ or } (0.6738, 2.9065).$$

13.2.3 ARMA(0,1) OR MA(1)

In this section, we set $p = 0$ and $q = 1$ to obtain the forecast function for a $MA(1)$ process as

$$\hat{Y}_t(l) = \mu + E(a_{t+l}|Y_1, \ldots, Y_t) - \theta_1 E(a_{t+l-1}|Y_1, \ldots, Y_t).$$

When $l = 1$ the forecast function becomes

$$\hat{Y}_t(1) = \mu - \theta_1 E(a_t|Y_1, \ldots, Y_t),$$

since $E(a_{t+1}|Y_1, \ldots, Y_t) = E(a_{t+l}) = 0$. Recall that an invertible $MA(1)$ process can be written as

$$a_t = \widetilde{Y}_t + \theta_1 \widetilde{Y}_{t+1} + \theta_1^2 \widetilde{Y}_{t+2} + \cdots,$$

so that

$$E(a_t|Y_1, \ldots, Y_t) = a_t.$$

Therefore, the 1-step ahead forecast function becomes

$$\hat{Y}_t(1) = \mu - \theta_1 a_t,$$

with forecast error $e_t(1) = Y_{t+1} - \hat{Y}_t(1) = a_{t+1}$. It is then clear that the forecast is unbiased with variance σ_a^2.

For lead times $l \geq 2$, it is straightforward to see that $E(a_{t+j}|Y_1, \ldots, Y_t) = E(a_{t+j}) = 0$, $j \geq 1$. Thus,

$$\hat{Y}_t(l) = \mu, \ l \geq 2.$$

It is easy to show that the l-step ahead forecast is unbiased with forecast variance $(1+\theta_1^2)\sigma_a^2$.

Example 5

In our final example, we return to the data on daily viscosity readings of the chemical compound XB-77-5. In that example, we selected the AR(3) model

$$Y_t = 34.93007 + 0.5974(Y_{t-1} - 34.93007) - 0.3032(Y_{t-2} - 34.93007) - 0.2286(Y_{t-3} - 34.93007) + a_t,$$

based on the minimum AIC criterion and model diagnostics. Here, we remove the last 5 observations and then repeat the process of model identification, estimation and model diagnostics to obtain the fitted $AR(2)$ model given by

$$Y_t = 34.94668 + 0.7239(Y_{t-1} - 34.94668) - 0.4688(Y_{t-2} - 34.93007) + a_t,$$

with $\sigma_a^2 = \hat{4}.277$ and $AIC = 392.97$. The observed values of the last five observations were 34.4895, 39.1692, 35.8242, 32.3875 and 31.2846. The forecasted values obtained based on the $AR(2)$ model were 36.62240 (2.068), 34.91938 (2.5530), 34.14130 (2.556), 34.37646 (2.629) and 34.91148 (2.677) respectively with standard errors of the forecasts in parentheses.