5. NONSEASONAL BOX-JENKINS MODELS

In this section, we will discuss a class of models 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. Such a series is said to be stationary in the mean but nonstationary in the variance.

We note that the methods we have discussed in the previous sections for dealing with nonstationarity in the mean and seasonal time series may not be effective in modelling the trend and seasonality in some very complicated time series. In such cases, one may try a technique called *differencing*.

In summary, given a time series data y_t, $t = 1, 2, \ldots, n$, we proceed as follows.

1. Plot the time series data.
2. If the plot shows that the time series is nonstationary by means of one or a combination of

   (a) nonconstant variance

   (b) seasonal effects

   (c) cyclical fluctuations,

   use one of the methods we have discussed or the method of differencing we shall discuss to remove the components in order to obtain a stationary series, say u_t or z_t.
3. Use one of the nonseasonal Box-Jenkins models to be discussed in this section to model the stationary series u_t or z_t.

5.1. Eliminating Trend By Differencing

For the purpose of introducing the concept of differencing, suppose that the trend in a given series is linear. That is,

$$y_t = TR_t + u_t = \beta_0 + \beta_1 t + u_t, \;\; t = 1, 2, \ldots, n.$$

Notice that if we take the difference of adjacent observations we will be able to remove the linear trend. That is, if we take the *first difference*

$$z_t = y_t - y_{t-1}, \;\; t = 2, 3, \ldots, n,$$

then, the new time series z_t will be free of the linear trend component TR_t. Now,

$$\begin{aligned} z_2 &= y_2 - y_1 = (\beta_0 + \beta_1 \times 2 + u_2) - (\beta_0 + \beta_1 \times 1 + u_1) \\ &= \beta_1 + (u_2 - u_1) = \beta_1 + v_2. \end{aligned}$$

Similarly,

$$\begin{aligned} z_3 &= y_3 - y_2 = (\beta_0 + \beta_1 \times 3 + u_3) - (\beta_0 + \beta_1 \times 2 + u_2) \\ &= \beta_1 + (u_3 - u_2) = \beta_1 + v_3, \end{aligned}$$

and so on. In general,

$$\begin{aligned} z_t &= y_t - y_{t-1} = (\beta_0 + \beta_1 \times t + u_t) - (\beta_0 + \beta_1 \times (t-1) + u_{t-1}) \\ &= \beta_1 + (u_t - u_{t-1}) = \beta_1 + v_t, \end{aligned}$$

where $v_t = u_t - u_{t-1}$. Thus, z_t will be a time series with constant mean β_1. Note that first differencing will result in a series z_t with $n-1$ observations. Using the same approach, it is possible to show that second differencing can be used to eliminate quadratic trend in a time series. By second differencing we mean, taking the first difference of z_t. That is, second difference simply means differencing y_t twice as in,

$$w_t = z_t - z_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}, t = 3, 4, \ldots, n.$$

Continuing in this way, we can eliminate any polynomial trend of order p by diferencing the series y_t, p times.

As an example, consider the beer production data in Figure 1(c) which clearly exhibits a linear trend. The first difference of this series is shown both in the table below and in Figure 16.

First difference of beer production series

	1Q	2Q	3Q	4Q
1975:		8.46	-0.45	-8.43
1976:	0.47	8.44	2.32	-10.05
1977:	2.76	10.06	-5.23	-7.95
1978:	4.90	7.63	-0.09	-9.39
1979:	4.70	5.80	-1.67	-7.03
1980:	4.72	7.33	-0.44	-10.48
1981:	2.09	10.57	-2.94	-10.58
1982:	6.18	6.43	-1.96	-10.28

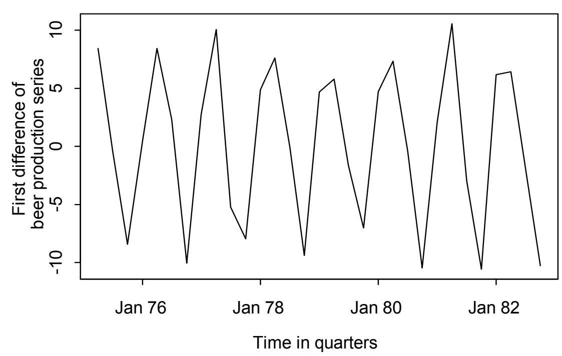

Figure 16: Plot of first difference of quarterly U.S. beer production shown in Figure 1(c).

Observe that we have successfully eliminated the linear trend in the series by taking the first difference. We recall that the polynomial regression methods performed poorly when used to model the trend in the monthly employment series shown in Figures 3(a) and 7(c). Thus, the trend and cyclical pattern in the monthly employment series shown in Figures 3(a) and 7(c) were also eliminated by taking the first difference. The series plotted in Figure 8(c) is the first difference of the monthly employment data.

We are now ready to introduce the backward shift operator $\mathbf{B}$. The backward shift operator is defined by $\mathbf{B}^j y_t = y_{t-j}$. Thus, the first difference can be written in terms of the backward shift operator as

$$z_t = y_t - \mathbf{B}y_t = (1 - \mathbf{B})y_t, \quad t = 2, 3, \ldots, n.$$

and the second difference can be written as

$$w_t = z_t - \mathbf{B}z_t = (1 - \mathbf{B})y_t - \mathbf{B}(1 - \mathbf{B})y_t.$$

By factorization we find that

$$w_t = (1 - \mathbf{B})(1 - \mathbf{B})y_t = (1 - \mathbf{B})^2 y_t = (1 - 2\mathbf{B} - \mathbf{B}^2)y_t, \quad t = 3, \ldots, n.$$

5.2. Eliminating Seasonal Fluctuations By Differencing

We note that the seasonal fluctuation in the beer data were not affected by the first difference as shown in Figure 16. Thus a different type of differencing is required for eliminating seasonal effects in a time series. Let L be the number of seasons in a series. Recall that for constant ($i.e$ not shifting) seasonal fluctuations the seasonal effects are the same from year to year for each season. Thus, it seems reasonable to consider differences between observations in the same season of adjacent years as a means of eliminating the constant seasonal fluctuation. To illustrate this point, consider a seasonal quarterly time series and denote the seasonal component at season t, $t = 1, 2, 3, 4$ by sn_t. Then, we can write

$$y_t = sn_t + u_t, t = 1, 2, \ldots, 12.$$

Then $sn_1 = sn_5 = sn_9$; $sn_2 = sn_6 = sn_{10}$; $sn_3 = sn_7 = sn_{11}$; and $sn_4 = sn_8 = sn_{12}$. Now consider the difference of observations in adjacent quarters, and define

$$w_5 = y_5 - y_1 = y_5 - \mathbf{B}^4 y_5 = (1 - \mathbf{B}^4)y_5 = (sn_5 + u_5) - (sn_1 + u_1).$$

Since $sn_1 = sn_5 = sn_9$ we have that

$$w_5 = (1 - \mathbf{B}^4)y_5 = u_5 - u_1 = v_5.$$

Thus, the new observation w_5 is free of the seasonal component. This process can be continued and defined in general as

$$w_t = (1 - \mathbf{B}^L)y_t = y_t - y_{t-L}, \quad t = L + 1, L + 2, \ldots, n,$$

where $L = 4$ in the above example.

Lag 4 differences of first differences of beer series

	1Q	2Q	3Q	4Q
1976:		-0.02	2.77	-1.62
1977:	2.29	1.62	-7.55	2.10
1978:	2.14	-2.43	5.14	-1.44
1979:	-0.20	-1.83	-1.58	2.36
1980:	0.02	1.53	1.23	-3.45
1981:	-2.63	3.24	-2.50	-0.10
1982:	4.09	-4.14	0.98	0.30

Now, we observe that the first differences of the beer production data contains only the seasonal component. Thus, we can apply this lag L differencing to the first difference to eliminate the seasonal effects. The remainder will then be a stationary series. The lag L differences are shown in the table above and the plot of the deseasonalized series is shown in Figure 17(a). Notice that lag L differencing leads to $L = 4$ missing observations.

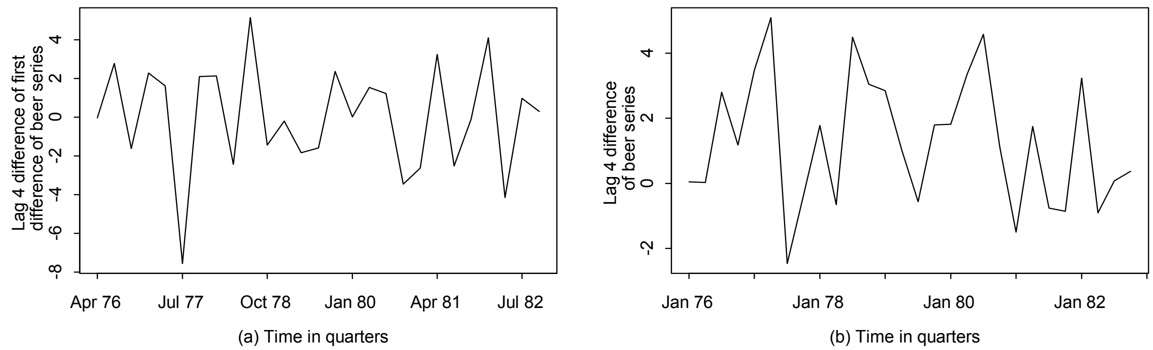

Figure 17: Plot of (a) Lag 4 differences of first difference of quarterly U.S. beer production series (b) Lag 4 differences of quarterly U.S. beer production series shown in Figure 1(c).

Lag 4 differences of beer series

	1Q	2Q	3Q	4Q
1976:	0.05	0.03	2.80	1.18
1977:	3.47	5.09	-2.46	-0.36
1978:	1.78	-0.65	4.49	3.05
1979:	2.85	1.02	-0.56	1.80
1980:	1.82	3.35	4.58	1.13
1981:	-1.50	1.74	-0.76	-0.86
1982:	3.23	-0.91	0.07	0.37

Figure 17(b) provides a good illustration of the fact that applying lag L differencing to the original time series will eliminate both trend and seasonal effects, if the trend in the series is linear. This is easy to see from the fact that if

$$y_t = \beta_0 + \beta_1 t + SN_t + u_t,$$

and the series is a quarterly series, then

$$w_5 = (1 - \mathbf{B}^4)y_5 = y_5 - y_1 = (\beta_0 + 5\beta_1 + SN_5 + u_5) - (\beta_0 + \beta_1 + SN_1 + u_1).$$

Since $SN_5 = SN_1$, we have

$$w_5 = 4\beta_1 + v_5.$$

where $v_5 = u_5 - u_1$. In general, it is easily verified that

$$w_t = (1 - \mathbf{B}^4)y_t = 4\beta_1 + v_t, \quad t = 5, 6, \ldots, n,$$

which is a series with constant mean $4\beta_1$, no trend and no seasonal effects. The lag 4 differences of the actual beer production series is shown in the table above.

5.3. 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 observations k distances apart (Y_t, Y_{t+k}) or (Y_{t-k}, Y_t)) is defined by

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

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. Roughly speaking, if there is no relationship between observations k distances apart, then $c(k) = 0$, for all values of k. Hence, $\rho(k) = 0$, for all k.

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

For a stationary process Y_t, $c(k)$ and $\rho(k)$ have the following properties.

1. Clearly, $c(0) = Var(Y_t)$. It follows that $\rho(0) = 1$.

2. $c(k) \leq c(0)$; $|\rho(k)| \leq 1$.
   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. From the definition of $c(k)$, we have $c(k) = c(-k)$ 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 $k > 0$ only. 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.

4. $c(k)$ 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 $\rho(k)$.

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

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2},$$

where the autocovariance function at lag k is estimated by

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$$

and

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n},$$

is the mean of the time series. It is common to refer to c_k and r_k as the *sample autocovariance function* and the *sample autocorrelation function* (SAC) 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 is always positive definite like $c(k)$, whereas the estimator with divisor $n-k$ is not always positive definite. In addition, 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. This is the reason that, for a given n, it is recommended to compute values of c_k for $k = 1, 2, \ldots, \frac{n}{4}$.

In a paper titled, "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. has shown that for processes in which $\rho(k) = 0$ whenever k is larger than some number, say m,

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

It follows that the large-lag standard error of r_k is

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

If a stationary process Y_t has no correlation structure but simply a random variable with constant mean (usually assumed to be zero) and constant variance σ^2, then $r_0 = 1$ and $r_k = 0$ for all values of $k \geq 1$. Such a process is called a *white noise process*. In this case,

$$S_{r_k} = \sqrt{\frac{1}{n}}. \tag{1.12}$$

Thus, the SAC is a useful tool in testing for lag k autocorrelation in a time series. To test whether the underlying process of a given series is a white noise process (*i.e* no correlation structure) we will use expression (1.12) and the SAC values r_k to construct a confidence interval for $\rho(k)$. Then, the SAC values that fall outside the interval will be considered significant, whereas those that fall within the interval will be considered not significant. For a white noise process, all the SAC values must fall within the interval. Now, under $H_0: \ \rho(k) = 0$, for all k, and the assumption of normality of $\rho(k)$, a 95% large-lag confidence interval for $\rho(k)$ is

$$\left(-1.96\sqrt{\frac{1}{n}}, \ 1.96\sqrt{\frac{1}{n}}\right)$$

where n is the number of observations used in computing r_k.

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

$$\begin{aligned} r_1 &= \frac{(0.05-1.28)(0.03-1.28)+(0.03-1.28)(2.8-1.28)+\ldots+(0.07-1.28)(0.37-1.28)}{(0.05-1.28)^2+(0.03-1.28)^2+\ldots+(0.37-1.28)^2} \\ &= \frac{4.6958}{108.0466} = 0.04346. \\ r_2 &= \frac{(0.05-1.28)(2.80-1.28)+(0.03-1.28)(1.18-1.28)+\ldots+(-0.91-1.28)(0.37-1.28)}{(0.05-1.28)^2+(0.03-1.28)^2+\ldots+(0.37-1.28)^2} \\ &= \frac{14.8561}{108.0466} = -0.1375. \end{aligned}$$

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 $r_7 = 0.19098$. Now, if the underlying process Y_t is white noise, then we expect $r_k = 0$, for all k. Then a large-lag 95% C. I. for $\rho(k)$ is

$$r_k \pm 1.96 \times S_{r_k} = 0 \pm 1.96\sqrt{\frac{1}{28}} \approx \pm 0.37.$$

That is, any value of r_k that falls within the inetrval $(-0.37,\ 0.37)$ will be considered to be zero; whereas any value that falls outside of the interval will be considered significantly different from zero.

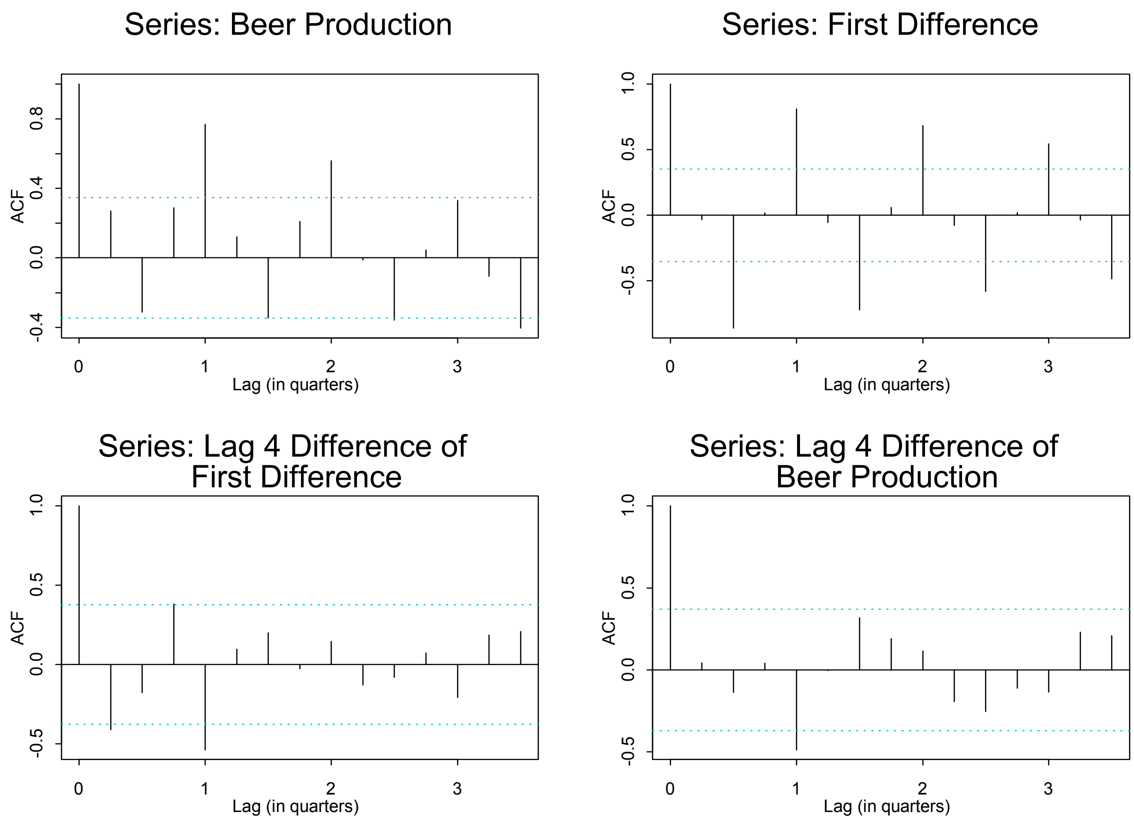

Figure 18: Plot of the sample autocorrelation function for: (a) beer production series (b) first difference of beer series (c) lag 4 difference of the first difference (d) lag 4 difference of the beer series.

2. A graph of the SAC values in a confidence band is shown in Figure 18. The oscillating pattern in the first two plots is an indication of the presence of seasonal effects in beer production series and in the first difference. The fact that the values continue to fall outside the confidence band is an indication that the series is not stationary. For a stationary series, the SAC 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.

5.4. 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, $\phi(k)$, denote the partial autocorrelation function, this conditional correlation is given by

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

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

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

where $\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$, for $j = 1, 2, \ldots, k-1$. Under the null hypothesis that the underlying process is a white noise, the variance of ϕ_{kk} can be approximated by

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

Hence, the critical limits of a 95% confidence interval on $\phi(k)$ to test the hypothesis of a white noise process is approximately $\frac{\pm 1.96}{\sqrt{n}}$.

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

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

In order to compute ϕ_{33} we first compute ϕ_{21}.

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

Then,

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

The careful reader would have noticed that in order to compute ϕ_{44} you first have to compute ϕ_{31} and ϕ_{32}, and so on.

2. A graph of the SPAC values in a confidence band is shown in Figure 19.

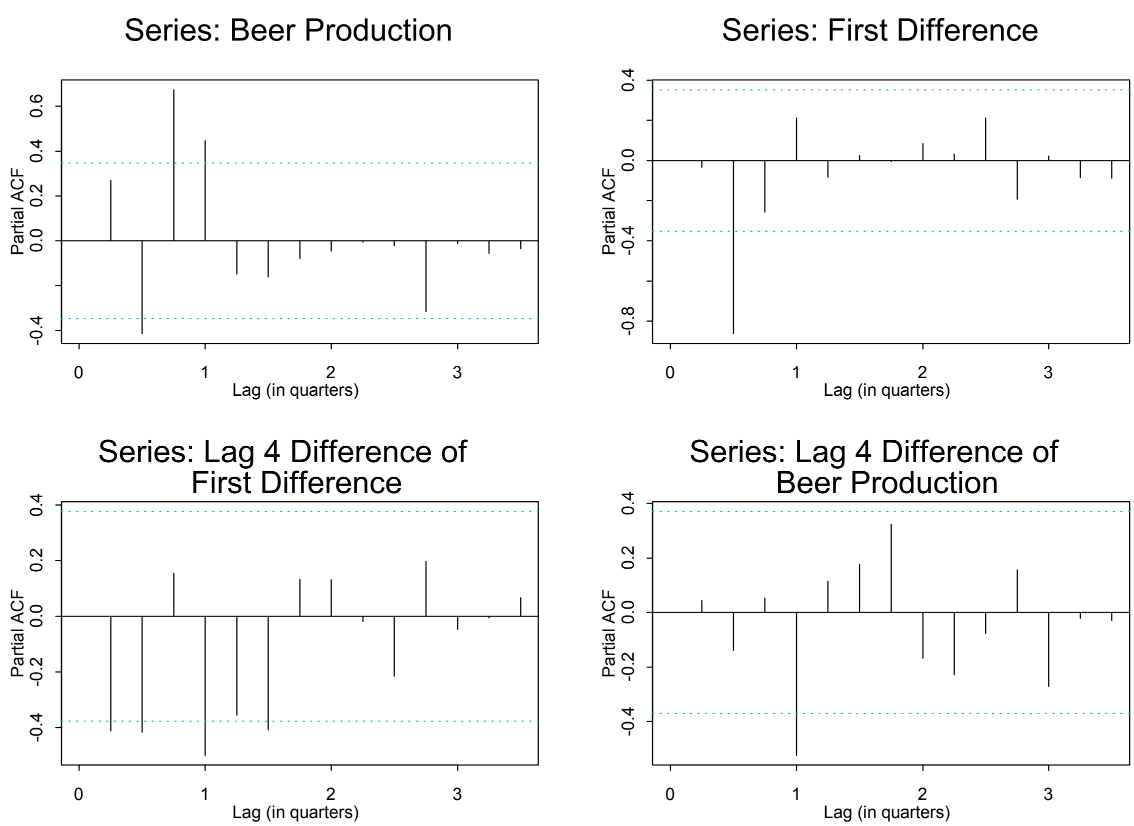

Figure 19: Plot of the sample partial autocorrelation function for: (a) beer production series (b) first difference of beer series (c) lag 4 difference of the first difference (d) lag 4 difference of the beer series.

5.5. Moving Average Representation of Time Series Processes

In the analysis of time series, there are two 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. That is, $E(a_t) = 0$ and $Var(a_t) = \sigma_a^2$ for all t. Then, a moving average representation (MA) of Y_t is,

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \ldots = \mu + a_t - \sum_{j=1}^{\infty} \theta_j a_{t-j}, \tag{1.13}$$

where $1 + \sum_{j=1}^{\infty} \theta_j^2 < \infty$. Note that we can use the backshift operator $\mathbf{B}$ to write the moving average representation as

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

Let $\widetilde{Y}_t = Y_t - \mu$ be the mean deleted process, and

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

Then, we can rewrite the moving average representation as

$$\widetilde{Y}_t = \Theta(\mathbf{B}) a_t. \tag{1.14}$$

Clearly, $E\left(\widetilde{Y}_t\right) = 0$. Recall that infinite sums are only well-defined when their n partial sums converge to zero. Therefore, the MA representation is well-defined only if

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

This implies that the larger the value of n the more accurate the MA representation becomes. Now, it is quite straightforward to show, from (1.13) or (1.14) that,

$$E(Y_t) = \mu, \quad Var(Y_t) = \sigma_a^2 \left(1 + \sum_{j=1}^{\infty} \theta_j^2\right),$$

and for $\theta_0 = -1$,

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

It follows that

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

We can see that the autocovariance function $c(k)$ and autocorrelation function $\rho(k)$ are clearly functions of the lag k only. However, since they involve infinite sums, we need to show that $\rho(k)$ is finite for each k, for a stationary process. Again, we use the Cauchy-Schwarz inequality to write

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

Therefore a condition for the process (1.13) to be stationary is that

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

5.6. 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

$$\widetilde{Y}_t = \phi_1\widetilde{Y}_{t-1} + \phi_2\widetilde{Y}_{t-2} + \phi_3\widetilde{Y}_{t-3} + \cdots + a_t = \sum_{j=1}^{\infty}\phi_j\widetilde{Y}_{t-j} + a_t. \tag{1.15}$$

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

$$\Phi(\mathbf{B})\widetilde{Y}_t = a_t, \tag{1.16}$$

where $1+\sum_{j=1}^{\infty}|\phi_j| < \infty$. It is not too difficult to show that, if the process is stationary,

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

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

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

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

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

This difference equation in $\rho(k)$ is called the Yule-Walker Equation. We see 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. Any process that can be written in the AR form is said to be an invertible process. Thus, for an MA process with the representation (1.13) or (1.14) to be invertible, we should be able to rewrite (1.14) 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. 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 (1.15) or (1.16) 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 equivalent to saying that the roots of the polynomial equation $\Phi(\mathbf{B}) = 0$ lie outside the unit circle. That is the absolute value of the roots should be larger than 1. For complex roots, the condition is that $|a + ib| = \sqrt{a^2 + b^2} > 1$.

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

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} + \ldots \theta_q a_{t-q} = \mu - \sum_{j=0}^{q} \theta_j a_{t-j}, \tag{1.17}$$

which we shall refer to 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

$$\widetilde{Y}_t = \phi_1 \widetilde{Y}_{t-1} + \phi_2 \widetilde{Y}_{t-2} + \cdots + \phi_p \widetilde{Y}_{t-p} + a_t = \sum_{j=1}^{\infty} \phi_j \widetilde{Y}_{t-j} + a_t. \tag{1.18}$$

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).

5.7. The First Order Autoregressive Process

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. We begin with a study of the finite order AR representation of a time series with $p = 1$. This model, called the first order autoregressive model, is denoted by $AR(1)$ and defined by

$$Y_t = \phi_1 Y_{t-1} + a_t. \tag{1.19}$$

It is clear from (1.19) that Y_t is independent of a_s whenever $s > t$. For example, a_t and Y_{t-1} are independent and $E(a_t Y_{t-1}) = E(a_t)E(Y_{t-1}) = 0$ since $t > (t-1)$. In terms of the backshift operator, the $AR(1)$ model can be written as

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

As mentioned earlier, the $AR(1)$ process is always invertible. For the process to be stationary, the root of

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

must 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 $|\phi_1| < 1$.

ACF of the $AR(1)$ Process

By definition, the autocovariance function of a process is

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

since $\mu = 0$ by our assumption. Now, replace Y_t by the expression in (1.19) 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 $k = 0$, $E(a_t Y_{t-k}) = E(a_t Y_t) = \phi_1 E(Y_{t-1}a_t) + E(a_t^2) = \sigma_a^2$. When $k \geq 1$, it can be easily verified that $E(a_t Y_{t-k}) = 0$. Thus,

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

The autocorrelation function then becomes

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

This type of equation is called a difference equation which can be solved recursively. That is,

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

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 slowly. 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.

1. **Example**

   For the purpose of illustration, we generate two series each with $n = 250$ values 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 two series generated from the AR(1) process is shown in Figure 20.

2. Observe that the first 18 values of the SAC of the first series are all positive because $\phi_1 > 0$. The SAC also decay exponentially. On the other hand, the SAC plot of the second series with $\phi_1 = -0.65$ exhibits an alternating pattern beginning with a negative value for r_1 and die out faster than the ACF of the first series. The SAC 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, $r_0 = 1$. The first ten values of the SAC for series 2 are listed in the table below

k	1	2	3	4	5	6	7	8	9	10
r_k	-0.597	0.353	-0.179	0.093	-0.086	0.037	-0.032	-0.0004	0.016	-0.014
r_{kk}	-0.597	-0.006	0.047	0.009	-0.061	-0.054	-0.028	-0.034	0.0069	0.005

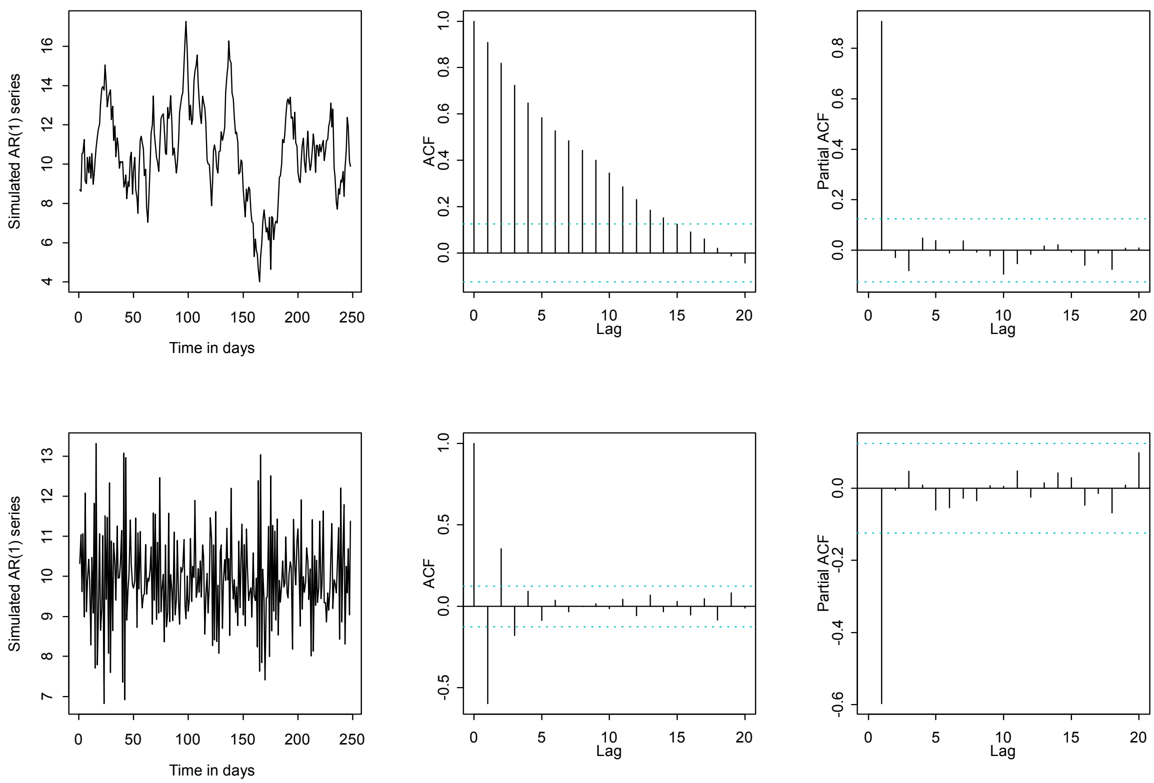

Figure 20: Plot of $AR(1)$ time series, SAC and SPAC functions generated from the model $(1 - \phi_1\mathbf{B})(Y_t - 10) = a_t$ with parameters $\phi_1 = 0.9$ and $\phi_1 = -0.65$ respectively, where $a_t \sim N(0, 1)$ is a white noise process.

PACF of the $AR(1)$ Process

Figure 20 shows that the PSAC of the two $AR(1)$ series cuts off after lag 1 because none of the sample PACF values are significant beyond that lag and more important, these insignificant r_{kk} values do not exhibit any pattern. It can be shown that the PACF of any $AR(1)$ process cuts off after lag 1. In fact, for an $AR(1)$ process, the theoretical PACF is given by

$$p(k) = \begin{cases} \rho(k) = \phi_1, & k = 1 \\ 0, & \text{for } k \geq 2. \end{cases}$$

Hence, 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 cuts off as shown in Figure 20. Note that the sample PACF values for $k \geq 2$ will not be exactly zero but will not be significant after lag 1. The sample PACF values for the second series shown in Figure 20 is listed on the table above.

1. **Example**

2. In this example, we discuss Y_t = weekly sales of absorbent paper towels (in units of 10,000 rolls). The time plot of the observed values y_t of Y_t is shown in Figure 21. Since the SAC of the first difference $z_t = y_t - y_{t-1}$ decays exponentially and the SPAC cuts off after lag 1, we identify the $AR(1)$ model as a tentative model for weekly sales of these paper towels. Since, $\bar{z}_t = 0.0054$ and not zero, the model is defined by

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

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. 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.

3. 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	-0.072 9	0.105	0.084	0.023	-0.133	-0.1191	-0.1738	-0.1182
r_{kk}	0.3066	-0.175	0.0061	0.133	-0.0095	0.0198	-0.1397	-0.0408	-0.1775	-0.0527

4. Another example of a real time series that can be modelled by the $AR(1)$ model is Y_t = daily average truck manufacturing defects shown in Figure 1(a). The SAC and SPAC plots are shown in Figure 22. A tentative time series model for representing daily average defects in the trucks 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. An initial guess for ϕ_1 is $\hat{\phi}_1^{(0)} = r_{11} = 0.4288$.

5.8. The Second Order Autoregressive $AR(2)$ Process

We begin with a reminder that we are assuming that Y_t is a mean deleted series. That is, the mean of Y_t is zero. The second order autoregressive $AR(2)$ model is defined by

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

or

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

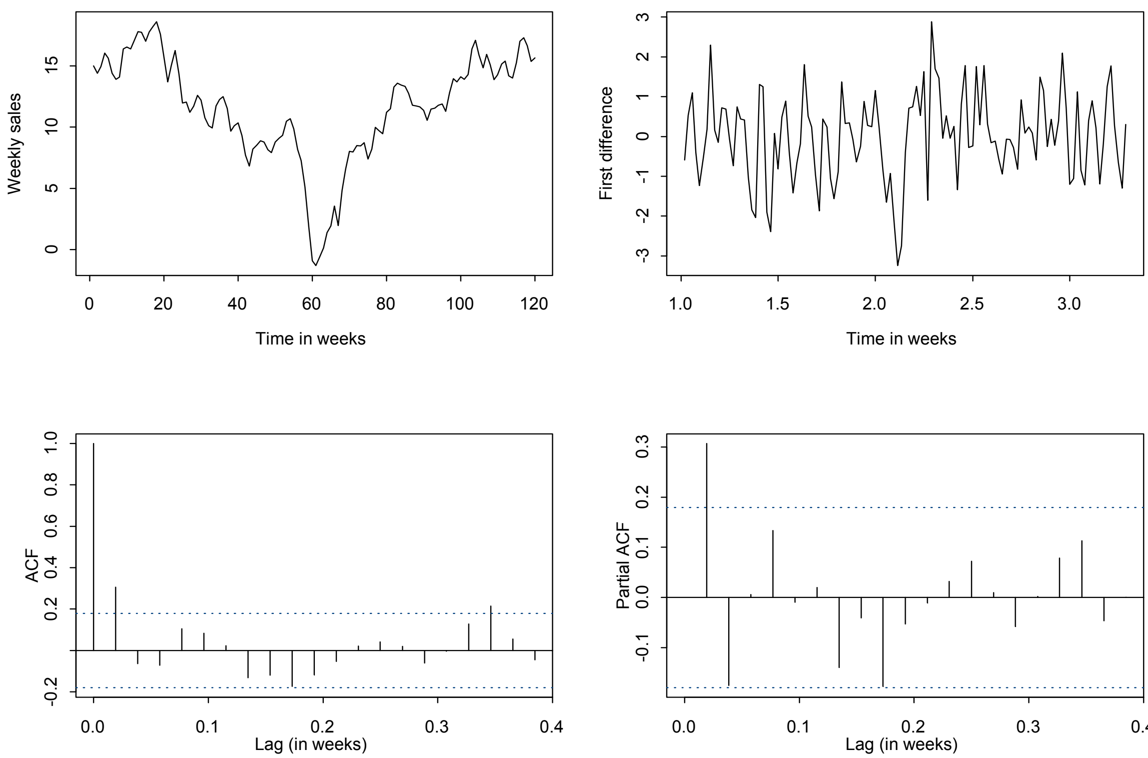

Figure 21: Plot of weekly sales of absorbent paper towels (in units of 10,000 rolls); first difference of weekly sales; SAC and SPAC values.

By virtue of this definition, the $AR(2)$ process is always invertible since $1 + |\phi_1| + |\phi_2| < \infty$, no matter how large these values are. However, the $AR(2)$ will be stationary (*i.e.* can be written as an MA process) if the roots of the quadratic equation

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

lie outside the unit circle. Now the roots of this equation are

$$\mathbf{B}_1 = \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

and

$$\mathbf{B}_2 = \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

Thus, the required condition for stationarity of an $AR(2)$ process is that $|\mathbf{B}_1| > 1$ and $|\mathbf{B}_2| > 1$. These conditions are equivalent to $1/|\mathbf{B}_1| < 1$ and $1/|\mathbf{B}_2| < 1$. It can be shown that from these conditions we have that

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

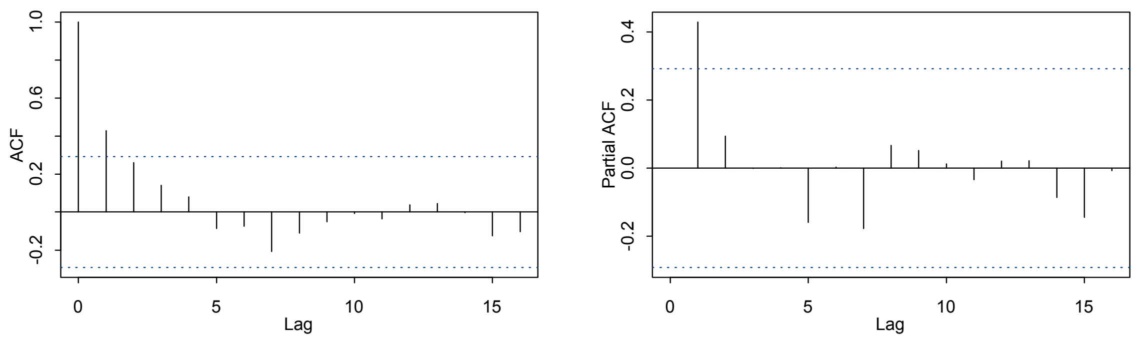

Figure 22: Plot of SAC and SPAC values of daily average number of truck manufacturing defects shown in Figure 1(a).

and

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

It follows that 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 $\mathbf{B} = -1.25$.

ACF of the $AR(2)$ Process

To find the autocorrelation function (ACF) of the $AR(2)$ process, we first obtain the autocovariance function of the process. Now, by definition

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

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

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

We observe that we obtained a similar difference equation for the autocorrelation function of an $AR(1)$ process. In time series, we call these equations Yule-Walker difference equations. We solved the difference equation for $AR(1)$ recursively to obtain a general expression for the autocorrelation function of the $AR(1)$ process. We can also solve this difference equation recursively. For instance, for $k = 1$, we have

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

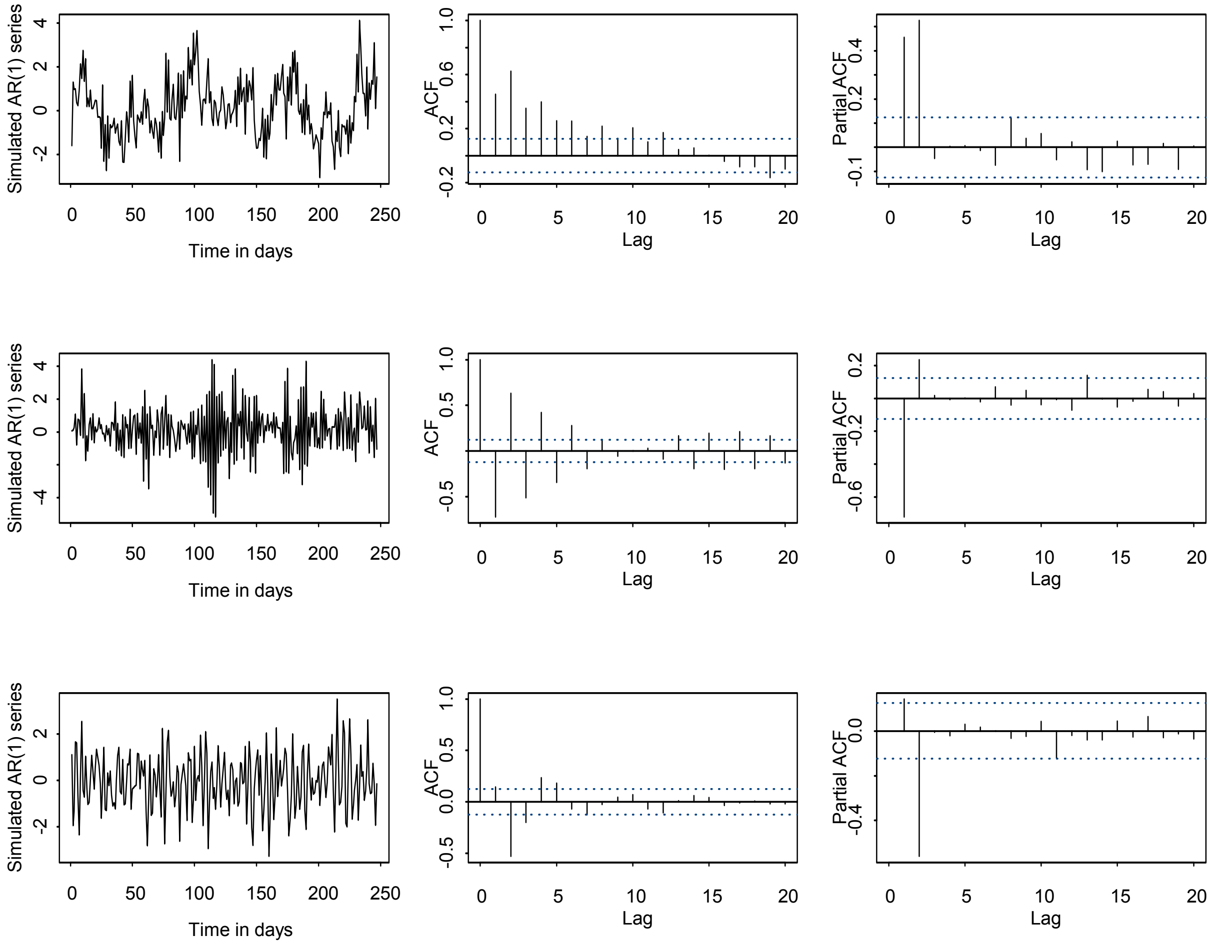

Figure 23: Plot of $AR(2)$ time series, SAC and SPAC functions generated from the model $(1-\phi_1\mathbf{B}-\phi_2\mathbf{B}^2)Y_t = a_t$ with parameters $(\phi_1,\phi_2) = (0.3, 0.5)$, $(\phi_1,\phi_2) = (-0.5, 0.3)$, $(\phi_1,\phi_2) = (0.3, -0.5)$ and $(\phi_1,\phi_2) = (-1.25, -0.65)$ respectively, where $a_t \sim N(0,1)$ is a white noise process.

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

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

When $k = 2$, we have

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

Then, substituting for $\rho(1)$ and for $\rho(0)$, 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 $k \geq 1$. Using more general techniques, we can show that the general solution to the Yule-Walker equation for and $AR(2)$ process is

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

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

One important property we can deduce from this more general result is that the ACF of an $AR(2)$ process will decay exponentially 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.

PACF of the $AR(2)$ Process

For the $AR(2)$ process the PACF cuts off after lag 2. That is, $p(k) = 0$, for $k \geq 3$. In fact it can be shown mathematically that

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

1. **Example**

2. We illustrate the structure of the ACF and PACF of an $AR(2)$ process by generating 250 observations from the model

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

where $a_t \sim N(0, 1)$. We compute and plot the sample ACF and sample PACF values for four different values of (ϕ_1, ϕ_2) as shown in Figure 23.

3. Figure 23 shows that the SAC of the four series decay exponentially and 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. We observe similar patterns between ϕ_1 and r_{11} and between ϕ_2 and r_{22} in the last two rows of Figure 21. This is an indication that one can use the SPAC 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 SPAC 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 SPAC of an $AR(3)$ model will cut off after lag 3; the SPAC of an $AR(4)$ model will cut off after lag 4; and so on.

4. A good example of a real time series that can be modelled 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 SAC values tail off gradually and the SPAC values cut off after lag 2. For the observed series, $\bar{y} = 34.93007$. Thus a tentative model for $Y_t =$ daily readings is

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

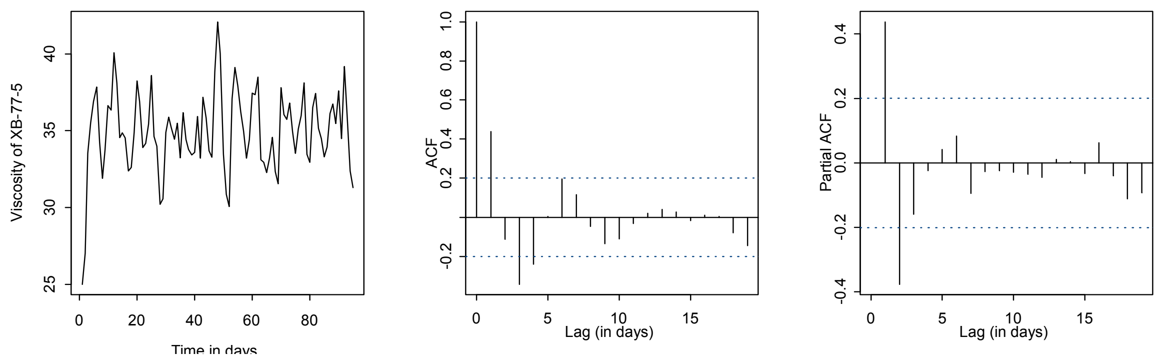

Figure 24: Plot of daily readings of the viscosity of chemical product $XB - 77 - 5$; SAC values and SPAC values .

5.9. 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

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

where a_t is a zero mean white noise process with constant variance σ_a^2. Since $1 + \theta_1^2 < \infty$, always, the $MA(1)$ process is always stationary. However, for the process to be invertible, 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 $|\theta_1| < 1$. It can be shown that 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 of an $MA(1)$ process is given by

$$\rho(k) = \frac{c(k)}{c(0)} = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1}{1+\theta_1^2}, & k = 1 \\ 0, & k > 1. \end{cases} \tag{1.20}$$

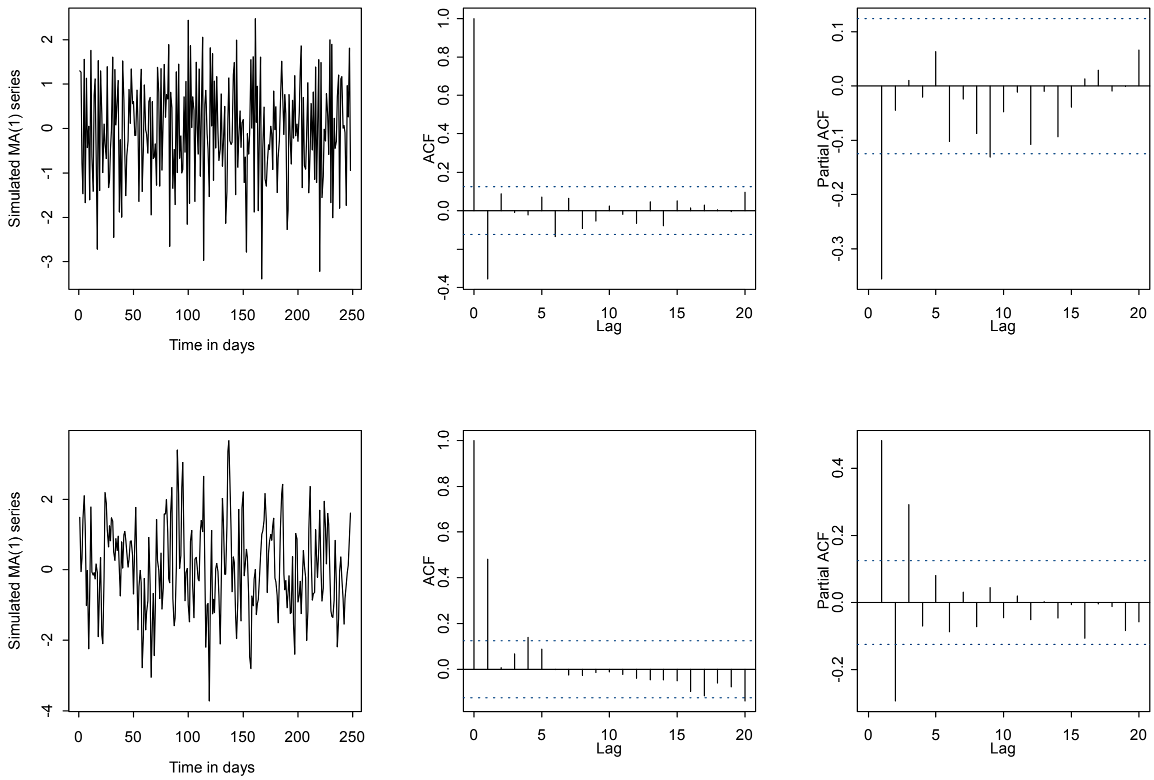

Figure 25: Plot of $MA(1)$ time series, SAC and SPAC functions generated from the model $Y_t = a_t - \theta_1 a_{t-1}$ with parameters $\theta_1 = 0.3$ and $\theta_1 = -0.75$ respectively, where $a_t \sim N(0, 1)$ is a white noise process.

That is, the ACF of an $MA(1)$ process cuts off after lag 1 (see Figure 25). On the contrary, the PACF of an $MA(1)$ process does not cut off but simply decay exponentially in one of two ways, as shown in Figure 25, depending on the sign of θ_1. It will exhibit an alternating pattern when θ_1 is negative. On the other hand, if $\theta > 0$ then all the theoretical PACF values will be negative. We note that for real time series, there may be a few deviations from this pattern.

From the expression (1.20) we note that the following two $MA(1)$ processes

$$Y_t = a_t - 0.4a_{t-1},$$

and

$$Y_t = a_t - 2.5a_{t-1}$$

have the same autocorrelations. However, the process represented by the first model is invertible because $|\theta_1| = 0.4 < 1$, whereas the second process is not invertible since $|\theta_1| = 2.5 > 1$. In other words, amongst these two processes with the same autocorrelations, one and only one is invertible.

Therefore, for uniqueness , we will restrict attention to invertible processes when selecting models for representing a time series. In general, the following two $MA(1)$ processes

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

and

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

have the same autocorrelations.

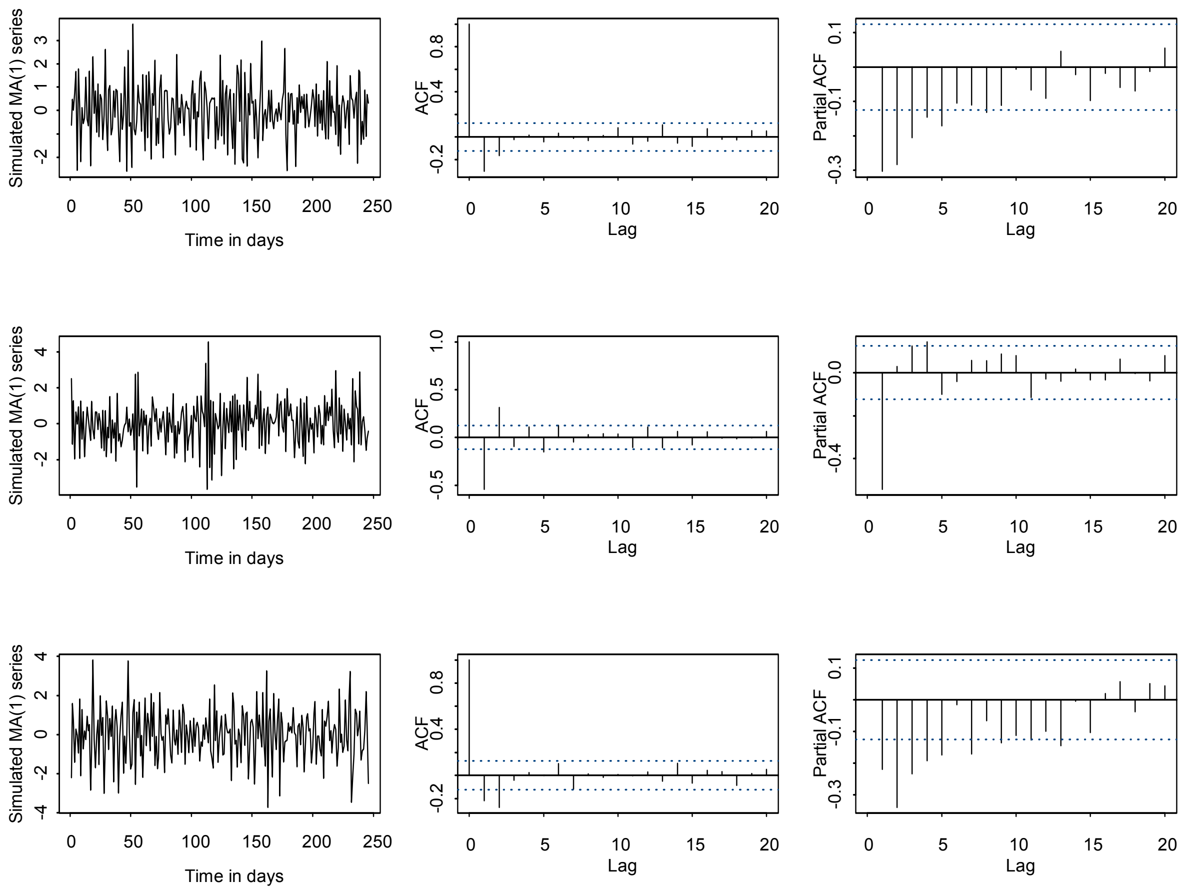

Figure 26: Plot of $MA(2)$ time series, SAC and SPAC functions generated from the model $Y_t = (1 - \theta_1 \mathbf{B} - \theta_2 \mathbf{B}^2)a_t$ with parameters $(\theta_1, \theta_2) = (0.65, 0.3)$ and $(\theta_1, \theta_2) = (0.65, -0.4)$ respectively, where $a_t \sim N(0, 1)$ is a white noise process.

5.10. The Second Order Moving Average $MA(2)$ Process

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

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

Being a finite order MA process, the $MA(2)$ process is always stationary. For invertibility, the roots of $(1 - \theta_1\mathbf{B} - \theta_2\mathbf{B}^2) = 0$ must lie outside the unit circle. This requirement leads to the following restriction on the values of the model parameters:

$$\begin{gathered} \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1. \end{gathered}$$

It can be shown that the autocovariance function of the $MA(2)$ process is given by

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

The autocorrelation function $\rho(k)$ can then be obtained directly from the autocovariance function by using the fact that $\rho(k) = \frac{c(k)}{c(0)}$. We note that $\rho(k) = 0$ for all lags greater than 2. That is, the theoretical ACF of the $MA(2)$ process cuts off after lag 2. On the other hand, the PACF of the $MA(2)$ process tails off exponentially of decay very slowly depending on the sign and magnitudes of θ_1 and θ_2.

1. **Example**

   We use daily viscosity readings for a chemical product $XR - 22$ to illustrate how we can use the patterns in the ACF and PACF of the $MA(2)$ process to identify a $MA(2)$ model as a tentative model for a series. In Figure 27, we see that the SAC cuts off after lag 2 except for two values at time lags 16 and 17. This is to be expected since the level of the confidence band is 95%. We also observed that the SPAC does not cut off. Thus, we tentatively identify the $MA(2)$ model as a suitable model for this data.

2. Based on the argument above our model for the daily viscosity of the chemical product XR-22 is

   $$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

   since the average of the series is not zero but $\bar{y} = 35.2013$.

5.11. The Autoregressive Moving Average Process

The SAC and SPAC values of the lag 12 difference of wages and salaries in Newfoundland from 1951-1969 in Figure 28, show that the SAC and SPAC values of a series may not cut off as quickly as we may desire. In such cases, we may combine both the AR and MA representations to model the series. For

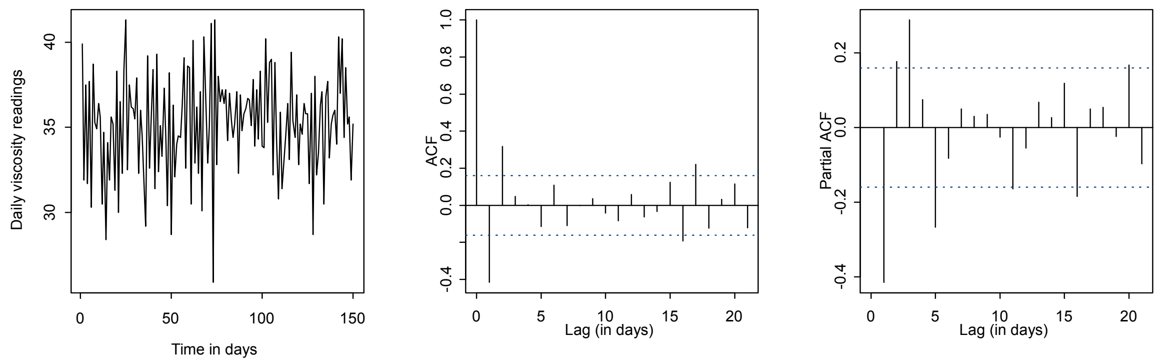

Figure 27: Plot of daily readings of the viscosity of chemical product $XR-22$; SAC values and SPAC values .

example, if we combine the $AR(1)$ and $MA(1)$ representations we obtain the Autoregressive Moving Average Model of order $(1,1)$ denoted by $ARMA(1,1)$. The $ARMA(1,1)$ model is given by

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

or

$$(1-\phi_1 \mathbf{B})\widetilde{Y}_t = (1-\theta_1 \mathbf{B})a_t.$$

The $ARMA(2,1)$ model is

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

and the $ARMA(1,2)$ model is

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

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

We now tabulate some useful patterns in SAC and SPAC values that we can use to identify AR, MA and $ARMA$ models.

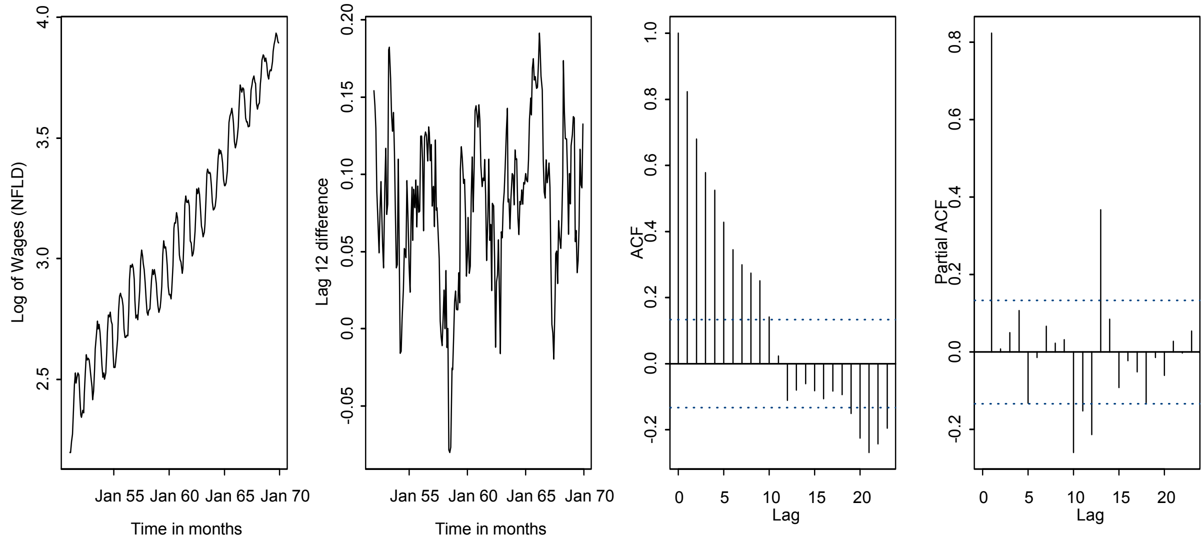

Figure 28: Plot of log of wages and salaries in Newfoundland from 1951-1969, Lag 12 difference of wages, SAC and SPAC values of Lag 12 difference.

Process	Theoretical Autocorrelation (ACF)	Theoretical Partial Autocorrelation (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 $(q-p)$	Decays exponentially after lag $(p-q)$

6. MODEL IDENTIFICATION AND ESTIMATION

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, SAC and sample partial autocorrelation function, SPAC 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, nonconstant 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 (SAC) and the sample PACF (SPAC) of the transformed series from Step 1. Examine the SAC and SPAC 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 (SAC) and the sample PACF (SPAC) of the stationary series from Step 2. If the series from Step 1 was stationary, the sample ACF (SAC) and the sample PACF (SPAC) of the stationary series are the same as in Step 2. Examine the patterns in the SAC and SPAC to identify the orders, p and q in the $ARMA(p, q)$ model for the data.

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 (SAC) and sample PACF (SPAC) 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 SAC and SPAC and not on the fine details. Model improvement can be carried out at the diagnostics stage of the analysis.

6.1. Empirical Examples

1. **Example 1**

   In this example we refer to the weekly sales of absorbent paper towels in Figure 21. After examining the time plot, we transformed the series into a stationary one by differencing. We then computed the SAC and SPAC of the first difference. The patterns in the SAC and SPAC appear to match those of an $AR(1)$ process in that the SAC tails off while the SPAC cuts off after lag 1. In addition, the mean of the first difference of the observed sales is $\hat{\mu} = \bar{z} \approx 0.0054$, which is not zero. Therefore, a tentative model we have identified for the first difference Z_t is

$$\widetilde{Z}_t = \phi_1 \widetilde{Z}_{t-1} + a_t \tag{1.21}$$

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

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

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

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

Theoretically, we know that $p(k) = \phi_1$. Our initial guess for ϕ_1 is therefore $\hat{\phi}_1^{(0)} = r_{11} = 0.3066$.

Estimation

Once a tentative model has been identified, the next step is to estimate the parameters of the model. There are a number of methods in the literature for parameter estimation. It is highly recommended that a mean deleted series $\tilde{z}_t$ or zero mean series be used in estimation. Using the maximum likelihood method of estimation in S-PLUS, we obtain the following parameter estimates

$$\hat{\phi}_1 = 0.30682, \quad \text{and} \quad \hat{\sigma}_a^2 = 1.101269,$$

with $AIC = 378.25214$. The Akaike Information Criterion (AIC) will be discussed in greater details in Example 4. Notice that the final estimate for ϕ_1 is very close to our initial guess. Thus, the fitted model is

$$\hat{\widetilde{Z}}_t = 0.30682 \widetilde{Z}_{t-1}.$$

Model diagnostics

Next, we evaluate the adequacy of our model by constructing plots of the residuals from the fitted model

$$\hat{a}_t = \widetilde{Z}_t - \hat{\widetilde{Z}}_t = \widetilde{Z}_t - 0.30682 \widetilde{Z}_{t-1}.$$

We had assumed that the white noise process a_t is normally distributed with zero mean and constant variance. Model diagnostics involves checking the validity of these assumptions. The SAC and SPAC values of the residuals shown in Figure 29 all fall within the confidence band

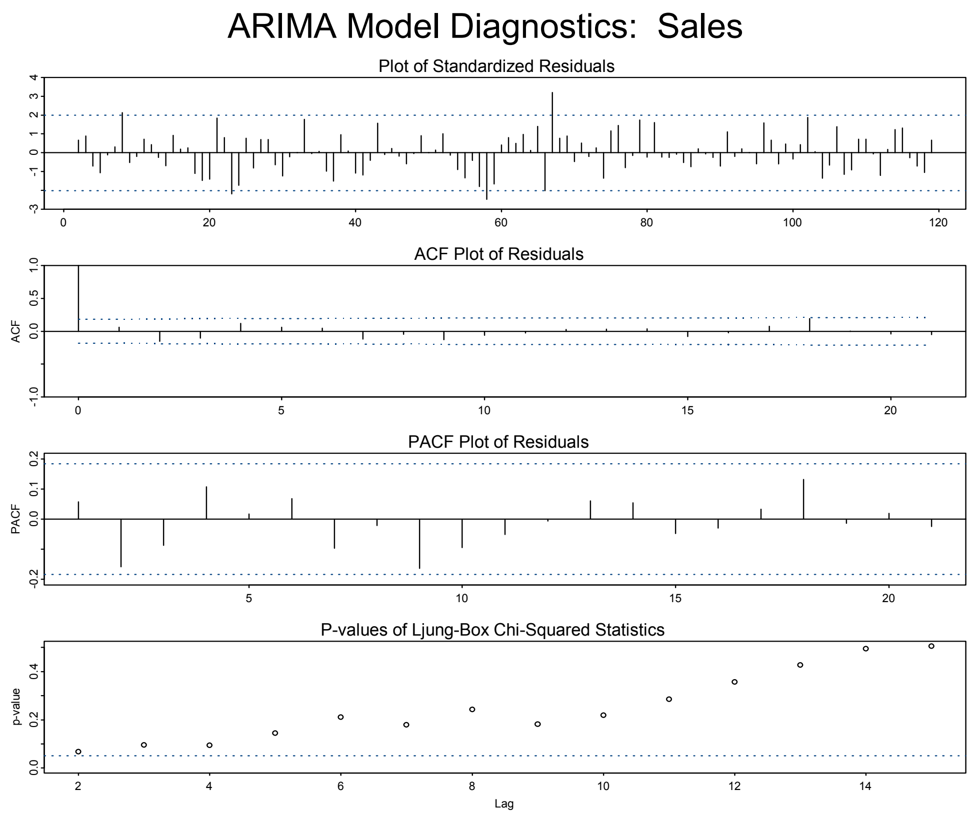

Figure 29: Plot for model diagnostics: $AR(1)$ model in Example 1.

indicating that the residuals are uncorrelated (white noise). Normality can be verified by constructing a histogram of the residuals or a normal probability plot of the residuals.

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

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

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 $K = 2, 3, \ldots, 15$ are shown in Figure 29. Large p-values indicate strong evidence in favour of

H_0. That is, the residuals are uncorrelated. Figure 29 show that the p-values for this test, in Example 1, are all larger than 0.05 and appear to increase for $K = 2, 3, \ldots, 15$ indicating a very strong evidence in favour of H_0. These results show that the $AR(1)$ model (1.21) is adequate for modelling the weekly sales of rolls of paper towels.

An analyst may look at the patterns in the SAC and SPAC and decide to fit an $MA(1)$ model to the weekly sales since the patterns in the SAC and SPAC also match the patterns in the theoretical ACF and PACF of an $MA(1)$ process. In that case, the tentative model becomes

$$\widetilde{Z}_t = a_t - \theta_1 a_{t-1}.$$

Our initial guess for θ_1 is that $\hat{\theta}_1^{(0)} < 0$, since $r_1 > 0$. The maximum likelihood estimates of the parameters for this model are $\hat{\theta}_1 = -0.3518$ and $\hat{\sigma}_a^2 = 1.070804$ with $AIC = 347.98024$. We note that $\hat{\sigma}_a^2$ and the AIC value for this model are slightly lower than those of the $AR(1)$ model. The diagnostic plots for this model also indicate that the model is adequate for the weekly sales series. The $p-value$ at any given lag K, for the portmanteau lack of fit test appear to be larger than that for the $AR(1)$ model. Overall, the $MA(1)$ model seem to offer some improvement over the $AR(1)$ model. However, the improvements are very small and negligible and the $AR(1)$ model will be more useful to us when it comes to forecasting.

2. **Example 2**

The daily viscosity readings of the chemical product $XB-77-5$ shown in Figure 24 appear to be stationary, therefore no transformation is required. The SAC and SPAC also do not indicate any nonstationary phenomena. A careful examination of the patterns in the SAC and SPAC show that the SAC tail off while the SPAC cuts off after lag 2. These patterns matches that of an $AR(2)$ process. We therefore tentatively identify the daily readings as an $AR(2)$ process. The tentative model is then

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

where $\widetilde{Y}_t = Y_t - 34.93007$ and a_t is a normally distributed white noise process. By comparing the pattern in the plot of the SPAC with the theoretical patterns in Figure 23, we infer that $\hat{\phi}_1^{(0)} > 0$ and that $\hat{\phi}_2^{(0)} < 0$.

Estimation

Using the maximum likelihood method in S-PLUS, we find that

$$\hat{\phi}_1 = 0.5706, \quad \hat{\phi}_2 = -0.36499 \text{ and } \hat{\sigma}_a^2 = 3.825336.$$

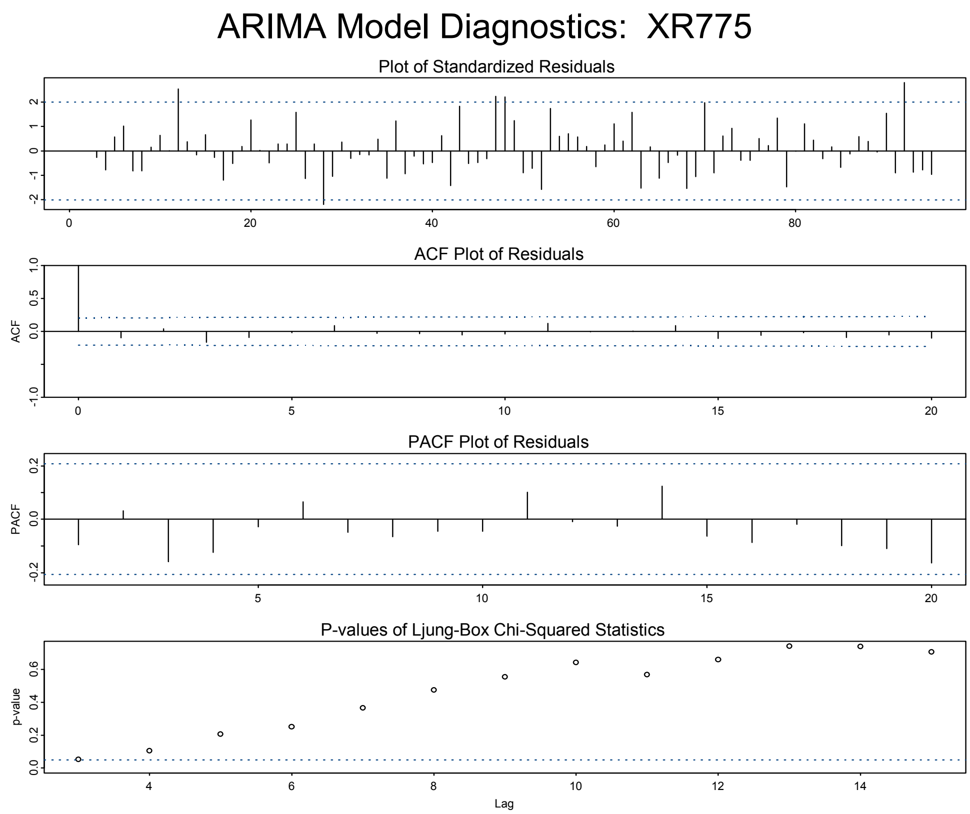

Figure 30: Plot for model diagnostics: $AR(2)$ model in Example 2.

It follows that the fitted $AR(2)$ model is

$$\hat{\tilde{Y}}_t = 0.5706\tilde{Y}_{t-1} - 0.36499\tilde{Y}_{t-2}.$$

The residuals from the fitted model are then computed from

$$\hat{a}_t = \tilde{Y}_t - \hat{\tilde{Y}}_t = \tilde{Y}_t - 0.5706\tilde{Y}_{t-1} + 0.36499\tilde{Y}_{t-2}.$$

These residuals are then used in checking the adequacy of the fitted model. Figure 30 shows that the assumption that a_t is a white noise process is valid. Again, the assumptions of normality and constant variance can be verified by examining appropriate residual plots.

3. **Example 3**

The daily viscosity readings of another chemical product $XR-22$ is shown in Figure 27. For this data, the SAC appear to cut off after lag 2. However, there are two SAC values at lags 16 and 17 that appear to be significant. The SPAC seem to tail off very slowly. Although these patterns do not fit either an AR or a MA process exactly, by concentrating on the broad

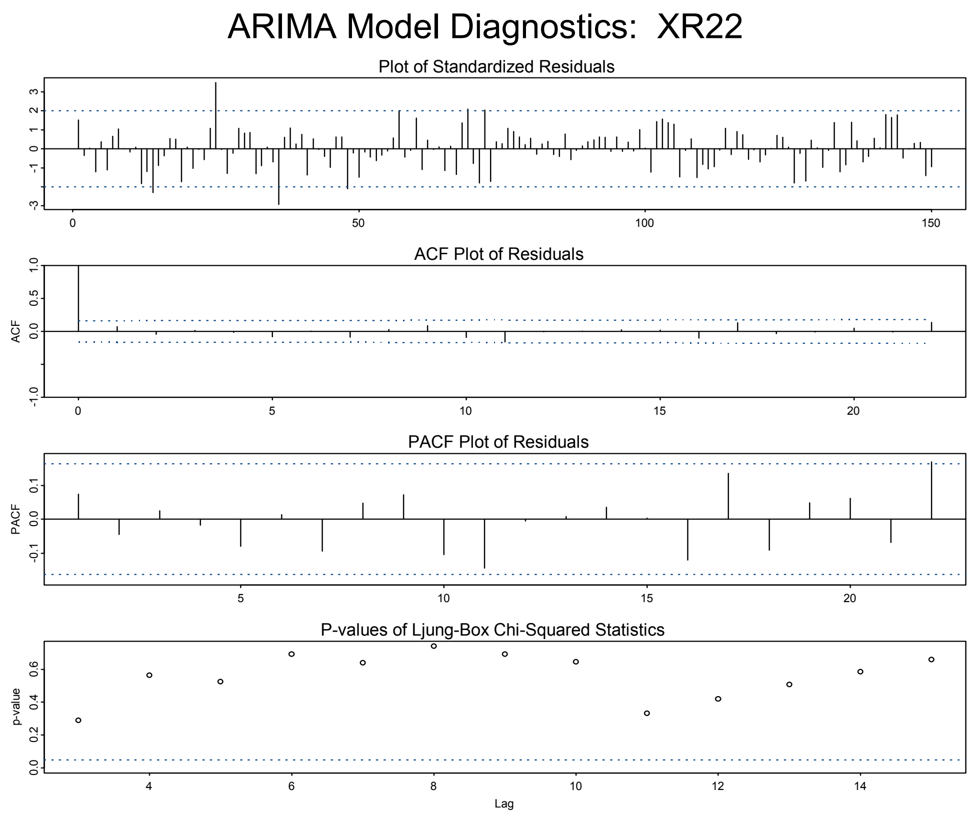

Figure 31: Plot for model diagnostics: $MA(2)$ model in Example 3.

features of the SAC and SPAC, we will tentatively consider an $MA(2)$ model for this series. This initial model may be updated at the stage of model diagnostics, if necessary. Thus our tentative model is

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

where $\hat{\mu} = \bar{y} = 35.20133$. Since $r_1 < 0$ and $r_2 > 0$, we infer that $\hat{\theta}_1^{(0)} > 0$ and $\hat{\theta}_2^{(0)} < 0$.

Estimation

The parameter estimates obtained from S-PLUS were

$$\hat{\theta}_1 = 0.51434, \quad \hat{\theta}_2 = -0.64253 \text{ and } \hat{\sigma}_a^2 = 5.6779.$$

It follows that the fitted $MA(2)$ model is

$$\hat{\widetilde{Y}}_t = -0.51434 a_{t-1} + 0.64253 a_{t-2}.$$

The residuals from the fitted model are then computed from

$$\hat{a}_t = \widetilde{Y}_t - \hat{\widetilde{Y}}_t = \widetilde{Y}_t + 0.51434 a_{t-1} - 0.64253 a_{t-2}.$$

Figure 31 show that the $MA(2)$ model is adequate for the viscosity of the chemical product $XR-22$.

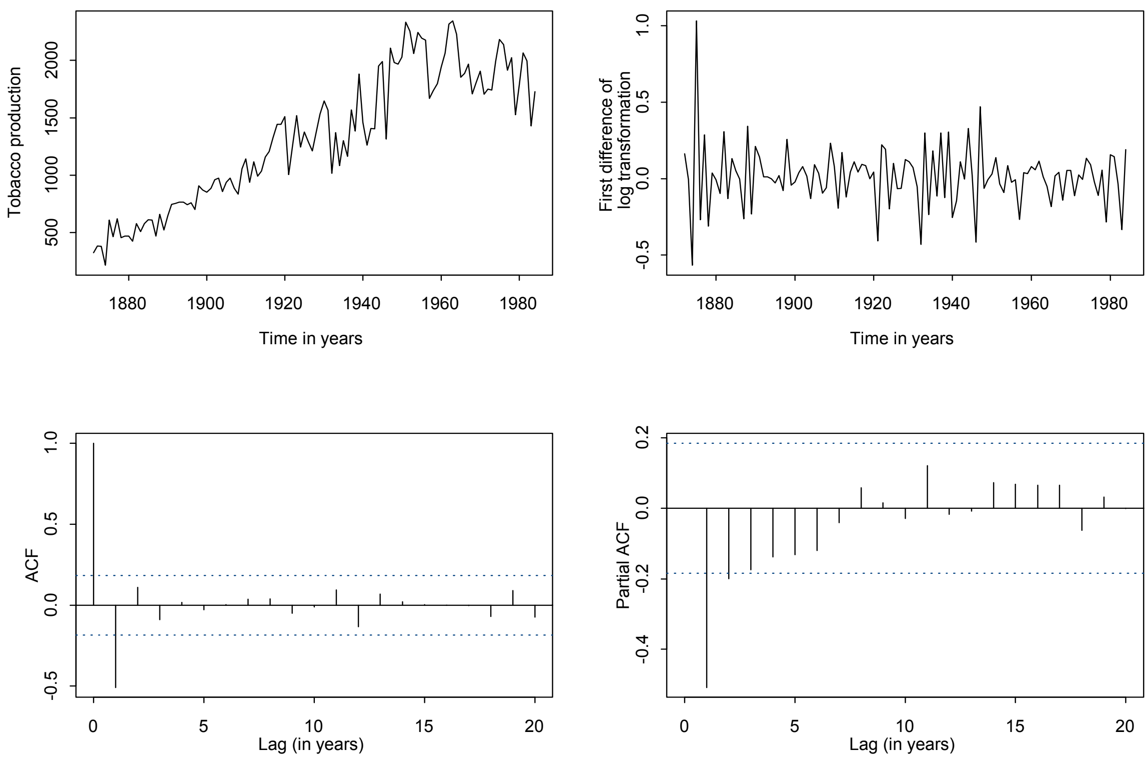

Figure 32: Plot of yearly U.S. tobacco production; First difference of natural logarithm of production figures; SAC and SPAC values of first difference.

4. **Example 4**

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 SAC and SPAC values (see Figure 32) show that the result of applying these two transformations is a stationary series. The patterns in the SAC and SPAC values may lead us to suggest two tentative models for the tobacco series. Let $W_t = lnY_t$ and $U_t = W_t - W_{t-1}$ be the first difference of W_t.

Model I Since the SAC cuts off after lag 1, we may recommend a $MA(1)$ model given by

$$\widetilde{U}_t = a_t - \theta_1 a_{t-1},$$

where $\hat{\mu} = 0.0147$. From the SAC values we observe that $r_1 < 0$. We therefore infer that $\hat{\theta}_1^{(0)} > 0$.

Model II A second model that may be considered for the tobacco series is the $AR(2)$ model

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

We infer that $\hat{\phi}_1$ and $\hat{\phi}_2$ are both negative since r_{11} and r_{22} are both negative (see Figure 32). It is possible to decide on which of these two models is a better model through model diagnostics.

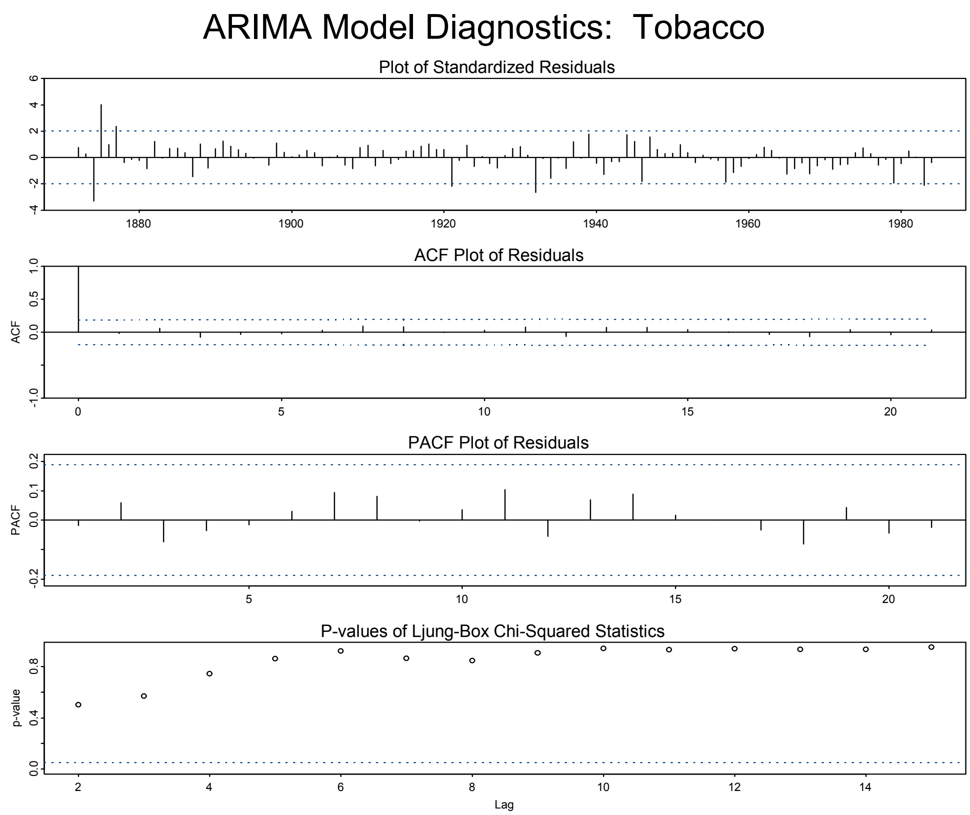

Figure 33: Plot for model diagnostics: $MA(1)$ model in Example 4.

Estimation

Model I The parameter estimates in the $MA(1)$ model are

$$\hat{\theta}_1 = 0.69116, \quad \text{and} \quad \hat{\sigma}_a^2 = 0.02612.$$

That is,

$$\hat{\widetilde{U}}_t = -0.69116 a_{t-1}.$$

Model II The parameter estimates in the $AR(2)$ model are

$$\hat{\phi}_1 = -0.61339, \quad \hat{\phi}_1 = -0.19971 \text{ and } \hat{\sigma}_a^2 = 0.02815.$$

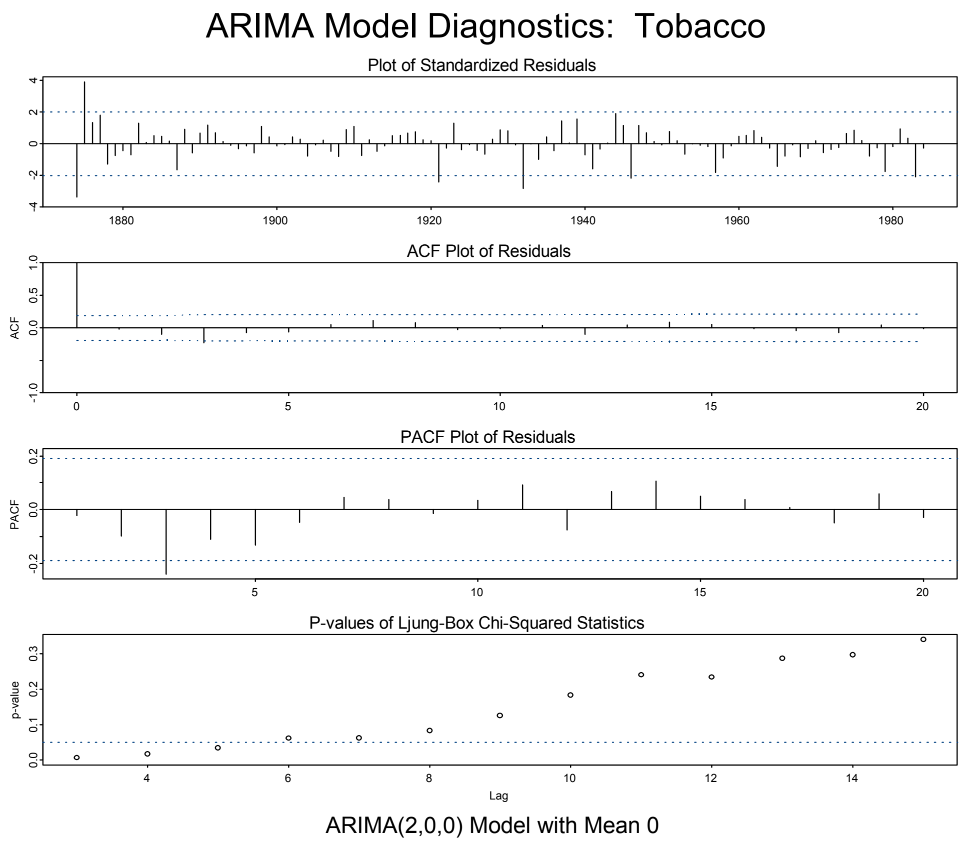

Figure 34: Plot for model diagnostics: $AR(2)$ model in Example 4.

That is,

$$\hat{\widetilde{U}}_t = -0.61339\widetilde{U}_{t-1} - 0.19971\widetilde{U}_{t-2}.$$

A very useful statistic for chosing the best model amongst two or more competing models is the Akaike Information Criterion (AIC) given by

$$AIC(k) = nlog(\hat{\sigma}_a^2) + 2k$$

where k is the number of parameters in a given model. The model with the smallest AIC is considered the best model. For Model I, the AIC is $AIC = -88.57717$ (from S-PLUS) whereas the AIC for Model II is $AIC = -77.28026$. S-PLUS uses a likelihood based approach to compute the AIC. This is why the AIC values from S-PLUS are different from the values computed from the formula given above. Based on the AIC, we conclude that the $MA(1)$ model is a better model since it has the smaller AIC. The diagnostic plots shown in Figures 33 and 34 also show that the residuals from the fitted $MA(1)$ model is a white noise whereas

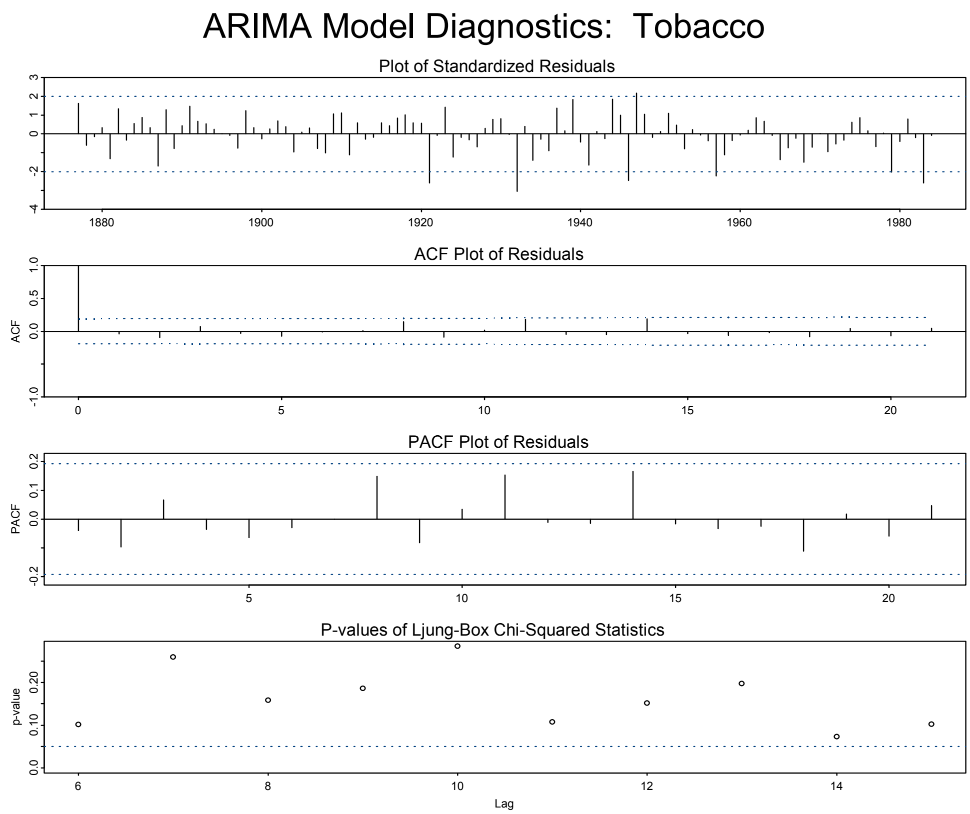

Figure 35: Plot for model diagnostics: $AR(5)$ model in Example 4.

the residuals from the $AR(2)$ are still correlated because the SPAC value at lag 3 is significant. This is further confirmation that the $AR(2)$ model is not adequate whereas the $MA(1)$ model is adequate. Since the SPAC value at lag 3 is significant we may update the $AR(2)$ model by adding three more terms in order to obtain an $AR(5)$ model. The estimated parameters of the fitted $AR(5)$ model are

$$\hat{\phi}_1 = -0.58721, \quad \hat{\phi}_2 = -0.26901, \quad \hat{\phi}_3 = -0.28828, \quad \hat{\phi}_4 = -0.22767, \quad \hat{\phi}_5 = -0.12796,$$

with $\hat{\sigma}_a^2 = 0.0204$ and $AIC = -108.52962$. The model diagnostics plots in Figure 35 show that the $AR(5)$ model is adequate.

Furthermore, given that both the SAC and SPAC decay very fast, one may wish to consider an $ARMA(1,1)$ model for the series. This leads to the fitted $ARMA(1,1)$ model

$$\hat{\widetilde{U}}_t + 0.10695\widetilde{U}_{t-1} = -0.54464a_{t-1}$$

with $AIC = -79.59215$. The diagnostic plots for this model suggests that the model is adequate in the sense that the residuals are a white noise. However, the concept of parsimony (chosing

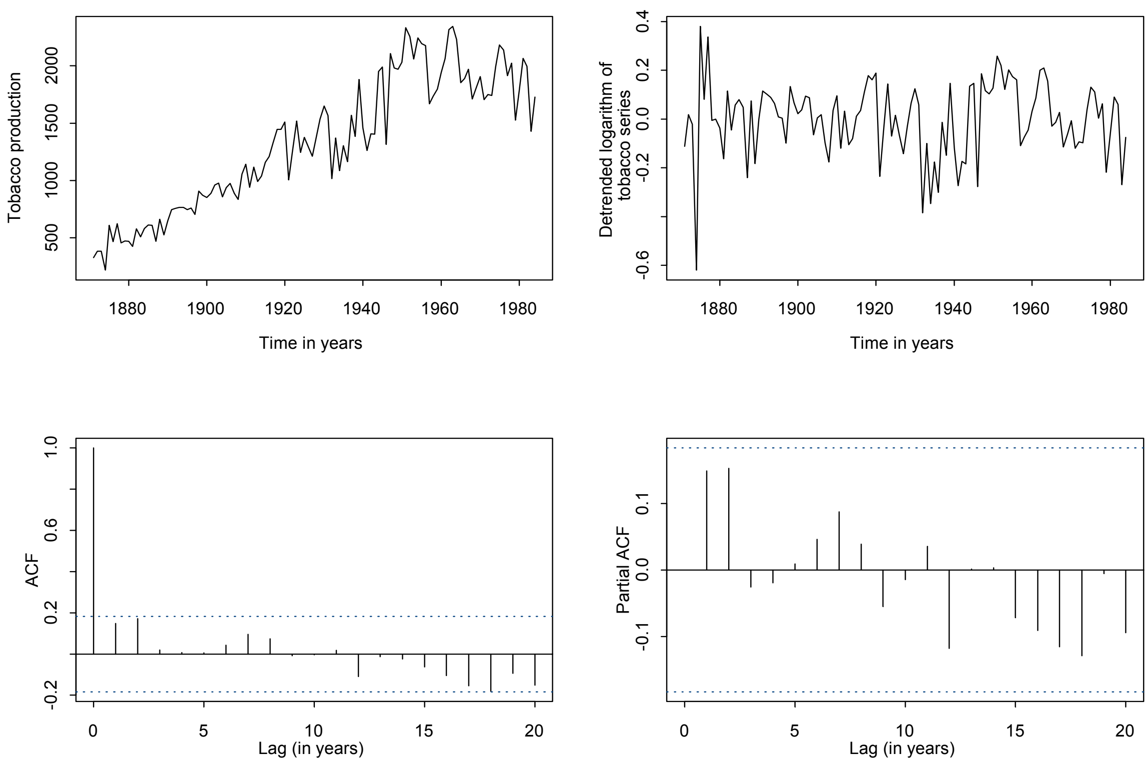

Figure 36: Plot of yearly U.S. tobacco production; Detrended natural logarithm of production figures; SAC and SPAC values of detrended series.

the model with fewer parameters) and the AIC value favour the $MA(1)$ model and the $AR(5)$. Since the difference between the AIC values for the $AR(5)$ model is quite large in magnitude, we would recommend using the $AR(5)$ model to study the tobacco series.

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

$$tr_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 = 5.8676 + 0.0336t - 0.00017t^2,$$

is the estimated trend curve. A plot of the observed values of U_t and the SAC and SPAC values of U_t is shown in Figure 36. 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

$$W_t = tr_t + a_t.$$

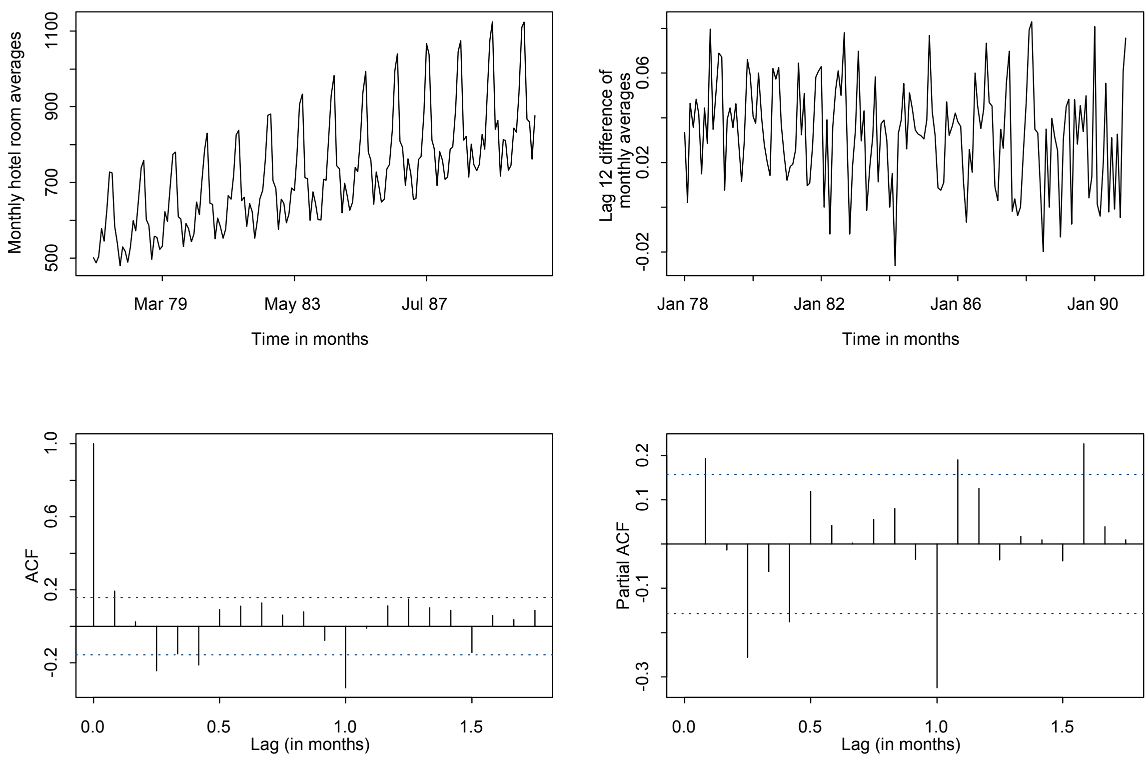

Figure 37: Plot of monthly hotel room averages; Lag 12 difference of natural logarithm of monthly averages; SAC and SPAC values of lag 12 difference.

5. **Example 5**

The monthly number of occupied hotel room averages in Traveller's Rest Inc. is shown in Figure 37. By examining the time plot, we note the presence of trend, seasonal fluctuations and nonconstant variance in the series. We therefore use the logarithmic transformation to stabilize variance before computing the Lag 12 difference of the log transformed series in order to eliminate both trend and seasonal effects. The stationary series that results from this approach is also shown in Figure 37. The next step was to compute the SAC and SPAC values.

This initial approach does not lead to a simple recognizable SAC and SPAC patterns. We therefore tried a second method. We eliminated trend by taking the first difference and then removed seasonal fluctuations through lag 12 differencing. However, this approach also leads to patterns in the SAC and SPAC that cannot be matched with any of the well known theoretical patterns. Thus, we tried a third approach. We applied regression methods to the trend to

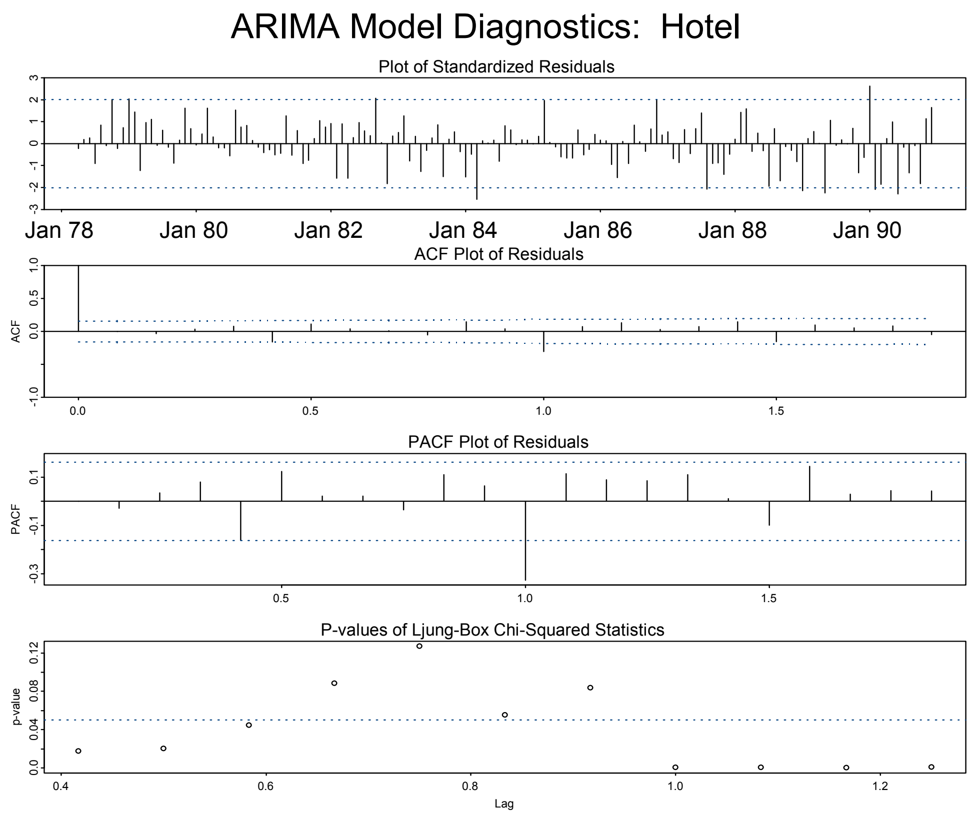

Figure 38: Plot for model diagnostics: $ARMA(3,1)$ model in Example 5.

obtain

$$tr_t = 6.33 + 0.0028t.$$

The lag 12 difference of the detrended series is then computed as well as the SAC and SPAC of the lag 12 difference. However, this method leads to patterns in SAC and SPAC that are similar to those in Figure 37. Other methods, such as polynomial regression with dummy variables/trigonometric functions or multiplicative or additive decomposition methods can then be used until a reasonable pattern and model is found for the hotel series. This particular example illustrates the fact that it may not be easy to find a reasonable model as quickly as we may wish. In order to find a reasonable model for the hotel series, we fit $ARMA(p,q)$ models of various orders to the stationary series from the first approach and then choose the best model based on diagnostic plots and the AIC value. The final model was found to be the $ARMA(3,1)$ model

$$\hat{\widetilde{U}}_t - 0.52792\widetilde{U}_{t-1} + 0.0235\widetilde{U}_{t-2} + 0.26981\widetilde{U}_{t-3} = -0.34855a_{t-1}$$

where $U_t = Y_t - Y_{t-12} = (1 - \mathbf{B}^{12})Y_t$ is the lag 12 difference of the hotel series. The AIC for this model is 729.1524 and the diagnostic plot is shown in Figure 38.

7. FORECASTING

One of the main objectives of analyzing a time series is forecasting. That is, to predict future values $\hat{Y}_{n+l}$ of the series, where n = number of observations and $l = 1, 2, \ldots$. For instance, we had $n = 120$ observed values of weekly sales in Example 1, Section 6 which we used to fit the $AR(1)$ and $MA(1)$ models. We may wish to compute the predicted value $\hat{Y}_{121}$. In this case, $l = 1$. 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. Although the formula SE_{n+l} is not presented here in details, MINITAB, S-PLUS, SAS and other computing software calculate SE_{n+l} and prints the value as part of their output. Using the SE_{n+l} we can then compute a $(1-\alpha)100\%$ prediction interval for the new value Y_{n+l}.

1. **Example:** Weekly sales of absorbent paper towels of Example 1, Section 6

   In order to be able to assess the performance of our model or cross validate our model we will use the first 115 observations in the series to fit the two models and then use the fitted models to predict or forecast the remaining five observed values. The mean of the first difference of the 115 observations is $\bar{z} = 0.00216$.

   Model I: Using the first difference of the 115 observed values, the fitted $AR(1)$ model is

$$\hat{\tilde{Z}}_t = 0.30056\tilde{Z}_{t-1},$$

   with $\hat{\sigma}_a^2 = 1.112459$ and $AIC = 334.72284$. Thus, we can write

$$\hat{\tilde{Z}}_{116} = 0.30056\tilde{Z}_{115},$$

   where $\tilde{Z}_{115} = Y_{115} - Y_{114} - 0.00216 = 1.2453947$. It follows that $\hat{\tilde{Z}}_{116} = (0.30056)(1.2453947) = 0.3743283$. Similarly, we find that $\hat{\tilde{Z}}_{117} = (0.30056)(0.3743283) = (0.30056)^2(1.2453947) = 0.1125119$. That is, in general

$$\hat{\tilde{Z}}_{n+l} = \hat{\phi}_1^l\tilde{Z}_n.$$

   It follows that $\hat{\tilde{Z}}_{118} = (0.30056)^3(1.2453947) = 0.033818$, $\hat{\tilde{Z}}_{119} = (0.30056)^4(1.2453947) = 0.010165$ and $\hat{\tilde{Z}}_{120} = (0.30056)^5(1.2453947) = 0.0030552$. The standard error of the forecasts

$\hat{\tilde{Z}}_{n+l}$, $(l = 1, 2, \ldots, 5)$ are respectively, 1.0547, 1.1013, 1.1055, 1.1058 and 1.1059.

The sales forecast can then be obtained by noting that $\hat{\tilde{Z}}_{n+l} = \hat{Z}_{n+l} - \hat{\mu} = (1 - \mathbf{B})\hat{Y}_{n+l} - \hat{\mu} = \hat{Y}_{n+l} - \hat{Y}_{n+l-1} - \hat{\mu}$. Thus, when $n = 1$, we have

$$0.3743283 = \hat{\tilde{Z}}_{116} = \hat{Y}_{116} - \hat{Y}_{115} - \hat{\mu}.$$

Rearranging the above expression, we find that the sales forecast for $t = 116$ is

$$\begin{aligned} \hat{Y}_{116} &= 0.3743283 + \hat{Y}_{115} + \hat{\mu} \\ &= 0.3743283 + 15.2463 + 0.00216 \approx 15.6228. \end{aligned}$$

In general, we find that the l-th step ahead forecast is given by the expression

$$\hat{Y}_{n+l} = \hat{\tilde{Z}}_{n+l} + \hat{Y}_{n+l-1} + \hat{\mu}.$$

Using this expression, we find that $\hat{Y}_{117} = 15.73711$, $\hat{Y}_{118} = 15.77306$, $\hat{Y}_{119} = 15.78538$ and $\hat{Y}_{120} = 15.79059$.

Alternatively, we can use the fact that $\tilde{Z}_t = Z_t - \mu = (1 - \mathbf{B})Y_t - \mu = Y_t - Y_{t-1} - \mu$ to write

$$\hat{\tilde{Z}}_{n+l} = \hat{Y}_{n+l} - \hat{Y}_{n+l-1} - \hat{\mu} = \hat{\phi}_1^l \tilde{Z}_n.$$

That is, in general

$$\hat{Y}_{n+l} = \hat{\mu}(1 - \hat{\phi}_1^l) + \hat{Y}_{n+l-1} + \hat{\phi}_1^l Y_n - \hat{\phi}_1^l Y_{n-1}.$$

For example, when $l = 1$,

$$\hat{Y}_{116} = \hat{\mu}(1 - \hat{\phi}_1) + \hat{Y}_{115} + \hat{\phi}_1 Y_{115} - \hat{\phi}_1 Y_{114} = 15.62252,$$

where $Y_{114} = 13.9996$ and $\hat{Y}_{115} = Y_{115} = 15.2463$ are known. For $l = 2$, we have

$$\hat{Y}_{117} = \hat{\mu}(1 - \hat{\phi}_1^2) + \hat{Y}_{116} + \hat{\phi}_1^2 Y_{115} - \hat{\phi}_1^2 Y_{114} = 15.73711.$$

Similarly, we find that $\hat{Y}_{118} = 15.77306$, $\hat{Y}_{119} = 15.78537$ and $\hat{Y}_{120} = 15.79059$. Below, we compare the forecast with the observed values in a table.

Weekly Sales Forecasts for Absorbent Paper Towels

l-th step	1	2	3	4	5
Actual	17.0179	17.2929	16.6366	15.3410	15.6453
Forecast	15.62252	15.73711	15.77306	15.78537	15.79059
Error	-1.39538	-1.55579	-0.86354	0.44437	0.14529

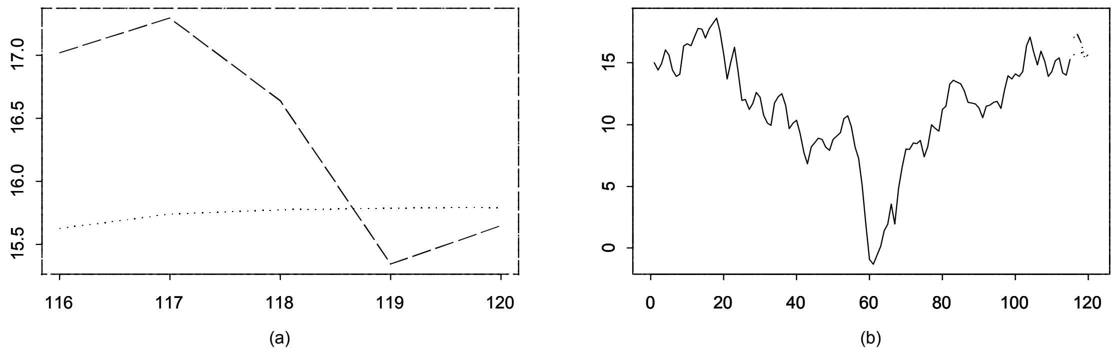

Figure 39: Plot of (a) forecast (dotted lines) and actual (broken lines) sales for week 116 - 120. (b) Weekly sales.

It is clear from Figure 39 that the actual sales from weeks 116 is, in general, decreasing, whereas the forecast is increasing. This points to the danger in forecasting too far into the future.

Model II: The fitted $MA(1)$ model is

$$\hat{\tilde{Z}}_t = 0.34438 a_{t-1},$$

with $\hat{\sigma}_a^2 = 1.083563$ and $AIC = 334.79325$. It follows that

$$\hat{\tilde{Z}}_{n+l} = 0.34438 a_{n+l-1}.$$

When $l = 1$, $a_{115} = \tilde{Z}_{115} - \hat{\tilde{Z}}_{115} \approx 1.168239$ (from S-PLUS). Therefore,

$$\hat{\tilde{Z}}_{116} = 0.34438 a_{115} = (0.34438)(1.168239) = 0.4023183.$$

with standard error 1.040943. Now, since $\tilde{Z}_{116}, \tilde{Z}_{117}, \ldots$ are unknown, it will be impossible to compute $a_{116}, a_{117}, \ldots$ and hence we cannot use the moving average model of order 1 to forecast beyond 1 step. That is, the $MA(1)$ model can only help us compute one-step ahead forecast. This explains why AR models are often preferred when several steps ahead forecasts are desired. In our example, we find that from the $MA(1)$ model

$$0.4023183 = \hat{\tilde{Z}}_{116} = \hat{Y}_{116} - \hat{Y}_{115} - \hat{\mu}.$$

Therefore,

$$\begin{aligned} \hat{Y}_{116} &= 0.4023183 + \hat{Y}_{115} + \hat{\mu} \\ &= 0.4023183 + 15.2463 + 0.00216 \approx 15.6508 \end{aligned}$$