Integrated Models for Nonstationary Data

Definition 3.11 *A process x_t is said to be* **ARIMA**(p, d, q) *if*

$$\nabla^d x_t = (1 - B)^d x_t$$

is ARMA(p, q). In general, we will write the model as

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t. \tag{3.143}$$

If $E(\nabla^d x_t) = \mu$, we write the model as

$$\phi(B)(1 - B)^d x_t = \delta + \theta(B) w_t,$$

where $\delta = \mu(1 - \phi_1 - \cdots - \phi_p)$.

It should be clear that, since $y_t = \nabla^d x_t$ is ARMA, we can use §3.5 methods to obtain forecasts of y_t, which in turn lead to forecasts for x_t. For example, if $d = 1$, given forecasts y_{n+m}^n for $m = 1, 2, \ldots$, we have $y_{n+m}^n = x_{n+m}^n - x_{n+m-1}^n$, so that

$$x_{n+m}^n = y_{n+m}^n + x_{n+m-1}^n$$

with initial condition $x_{n+1}^n = y_{n+1}^n + x_n$ (noting $x_n^n = x_n$).

It is a little more difficult to obtain the prediction errors P_{n+m}^n, but for large n, the approximation used in §3.5, equation (3.86), works well. That is, the mean-squared prediction error can be approximated by

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^{*2}, \tag{3.144}$$

where ψ_j^* is the coefficient of z^j in $\psi^*(z) = \theta(z)/\phi(z)(1 - z)^d$.

To better understand integrated models, we examine the properties of some simple cases; Problem 3.29 covers the ARIMA$(1, 1, 0)$ case.

Example 3.36 Random Walk with Drift

To fix ideas, we begin by considering the random walk with drift model first presented in Example 1.11, that is,

$$x_t = \delta + x_{t-1} + w_t,$$

for $t = 1, 2, \ldots$, and $x_0 = 0$. Technically, the model is not ARIMA, but we could include it trivially as an ARIMA$(0, 1, 0)$ model. Given data $x_1, \ldots, x_n$, the one-step-ahead forecast is given by

$$x_{n+1}^n = E(x_{n+1} \mid x_n, \ldots, x_1) = E(\delta + x_n + w_{n+1} \mid x_n, \ldots, x_1) = \delta + x_n.$$

The two-step-ahead forecast is given by $x_{n+2}^n = \delta + x_{n+1}^n = 2\delta + x_n$, and consequently, the m-step-ahead forecast, for $m = 1, 2, \ldots$, is

$$x_{n+m}^n = m\,\delta + x_n, \tag{3.145}$$

To obtain the forecast errors, it is convenient to recall equation (1.4), i.e., $x_n = n\,\delta + \sum_{j=1}^n w_j$, in which case we may write

$$x_{n+m} = (n + m)\,\delta + \sum_{j=1}^{n+m} w_j = m\,\delta + x_n + \sum_{j=n+1}^{n+m} w_j.$$

From this it follows that the m-step-ahead prediction error is given by

$$P_{n+m}^{n} = E(x_{n+m} - x_{n+m}^{n})^2 = E\Big(\sum_{j=n+1}^{n+m} w_j\Big)^2 = m\,\sigma_w^2. \tag{3.146}$$

Building ARIMA Models

Example 3.38 Analysis of GNP Data

In this example, we consider the analysis of quarterly U.S. GNP from 1947(1) to 2002(3), $n = 223$ observations. The data are real U.S. gross

national product in billions of chained 1996 dollars and have been seasonally adjusted. The data were obtained from the Federal Reserve Bank of St. Louis (`http://research.stlouisfed.org/`). Figure 3.12 shows a plot of the data, say, y_t. Because strong trend hides any other effect, it is not clear from Figure 3.12 that the variance is increasing with time. For the purpose of demonstration, the sample ACF of the data is displayed in Figure 3.13. Figure 3.14 shows the first difference of the data, ∇y_t, and now that the trend has been removed we are able to notice that the variability in the second half of the data is larger than in the first half of the data.

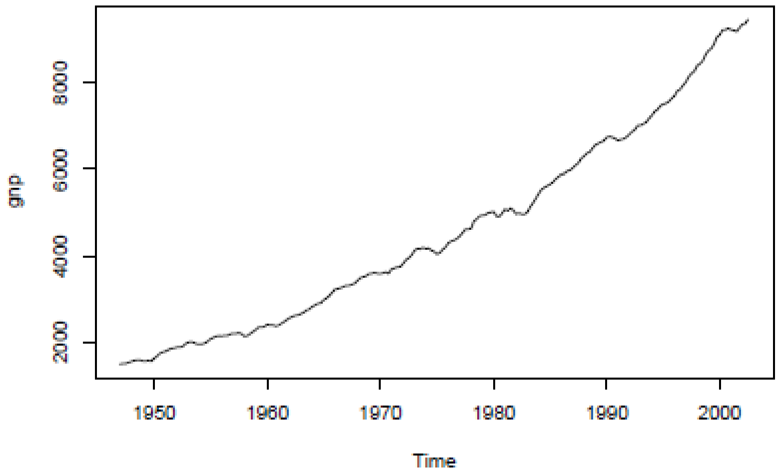

Fig. 3.12. Quarterly U.S. GNP from 1947(1) to 2002(3).

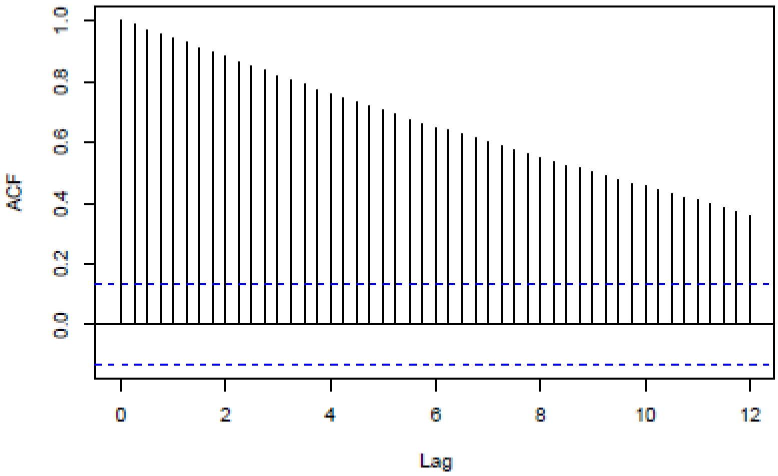

Fig. 3.13. Sample ACF of the GNP data. Lag is in terms of years.

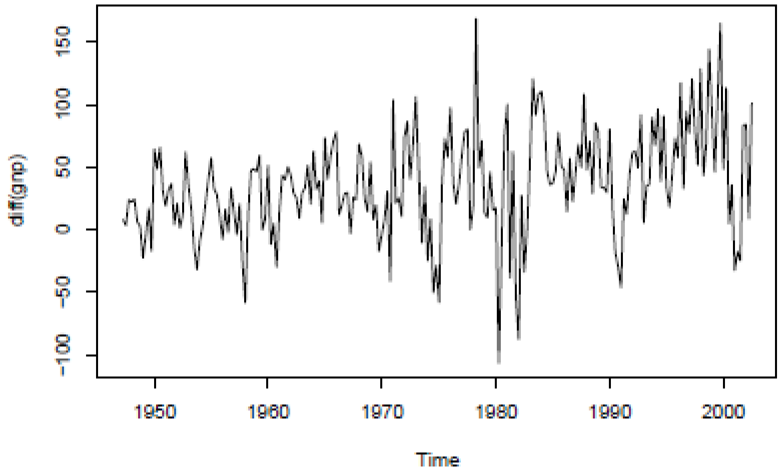

Fig. 3.14. First difference of the U.S. GNP data.

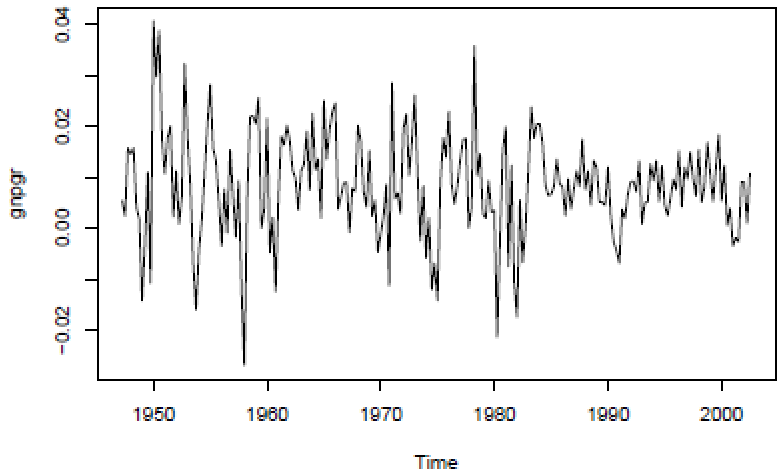

Fig. 3.15. U.S. GNP quarterly growth rate.

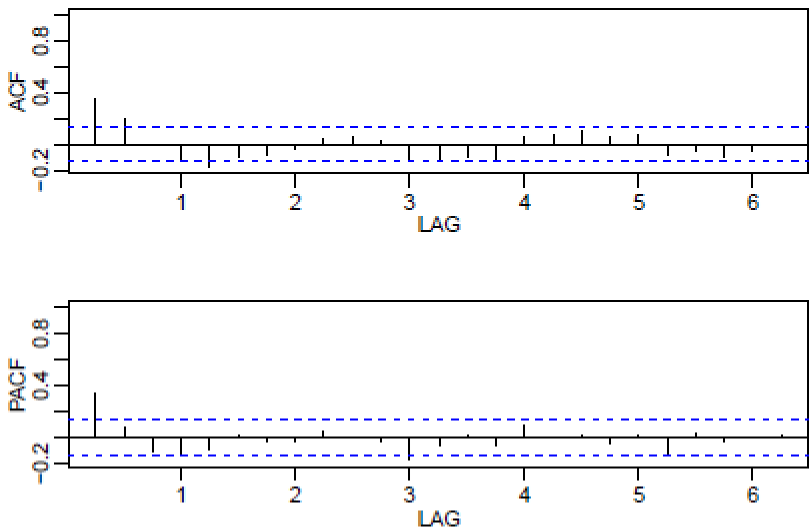

Fig. 3.16. Sample ACF and PACF of the GNP quarterly growth rate. Lag is in terms of years.

Using MLE to fit the MA(2) model for the growth rate, x_t, the estimated model is

$$x_t = .008_{(.001)} + .303_{(.065)}\widehat{w}_{t-1} + .204_{(.064)}\widehat{w}_{t-2} + \widehat{w}_t, \quad (3.151)$$

where $\widehat{\sigma}_w = .0094$ is based on 219 degrees of freedom.

The estimated AR(1) model is

$$x_t = .008_{(.001)}\,(1 - .347) + .347_{(.063)}x_{t-1} + \widehat{w}_t, \quad (3.152)$$

where $\widehat{\sigma}_w = .0095$ on 220 degrees of freedom; note that the constant in (3.152) is $.008\,(1 - .347) = .005$.

$$x_t = .35x_{t-1} + w_t,$$

and write it in its causal form, $x_t = \sum_{j=0}^{\infty}\psi_j w_{t-j}$, where we recall $\psi_j = .35^j$. Thus, $\psi_0 = 1, \psi_1 = .350, \psi_2 = .123, \psi_3 = .043, \psi_4 = .015, \psi_5 = .005, \psi_6 = .002, \psi_7 = .001, \psi_8 = 0, \psi_9 = 0, \psi_{10} = 0$, and so forth. Thus,

$$x_t \approx .35w_{t-1} + .12w_{t-2} + w_t,$$

which is similar to the fitted MA(2) model in (3.152).

Investigation of marginal normality can be accomplished visually by look ing at a histogram of the residuals. In addition to this, a normal probabilit plot or a Q-Q plot can help in identifying departures from normality. Se Johnson and Wichern (1992, Chapter 4) for details of this test as well a additional tests for multivariate normality.

The Ljung–Box–Pierce Q-statistic given by

$$Q = n(n+2)\sum_{h=1}^{H}\frac{\widehat{\rho}_e^2(h)}{n-h} \quad (3.154)$$

can be used to perform such a test. The value H in (3.154) is chosen somewhat arbitrarily, typically, $H = 20$. Under the null hypothesis of model adequacy, asymptotically ($n \to \infty$), $Q \sim \chi^2_{H-p-q}$. Thus, we would reject the null hypothesis at level α if the value of Q exceeds the $(1-\alpha)$-quantile of the χ^2_{H-p-q} distribution. Details can be found in Box and Pierce (1970), Ljung and Box (1978), and Davies et al. (1977). The basic idea is that if w_t is white noise, then by Property 1.1, $n\widehat{\rho}_w^2(h)$, for $h = 1, \ldots, H$, are asymptotically independent χ^2_1 random variables. This means that $n\sum_{h=1}^{H}\widehat{\rho}_w^2(h)$ is approximately a χ^2_H random variable. Because the test involves the ACF of residuals from a model fit, there is a loss of $p+q$ degrees of freedom; the other values in (3.154) are used to adjust the statistic to better match the asymptotic chi-squared distribution.

Example 3.39 Diagnostics for GNP Growth Rate Example

We will focus on the MA(2) fit from Example 3.38; the analysis of the AR(1) residuals is similar. Figure 3.17 displays a plot of the standardized residuals, the ACF of the residuals, a boxplot of the standardized residuals, and the p-values associated with the Q-statistic, (3.154), at lags $H = 3$ through $H = 20$ (with corresponding degrees of freedom $H - 2$).

Inspection of the time plot of the standardized residuals in Figure 3.17 shows no obvious patterns. Notice that there are outliers, however, with a few values exceeding 3 standard deviations in magnitude. The ACF of the standardized residuals shows no apparent departure from the model assumptions, and the Q-statistic is never significant at the lags shown. The normal Q-Q plot of the residuals shows departure from normality at the tails due to the outliers that occurred primarily in the 1950s and the early 1980s.

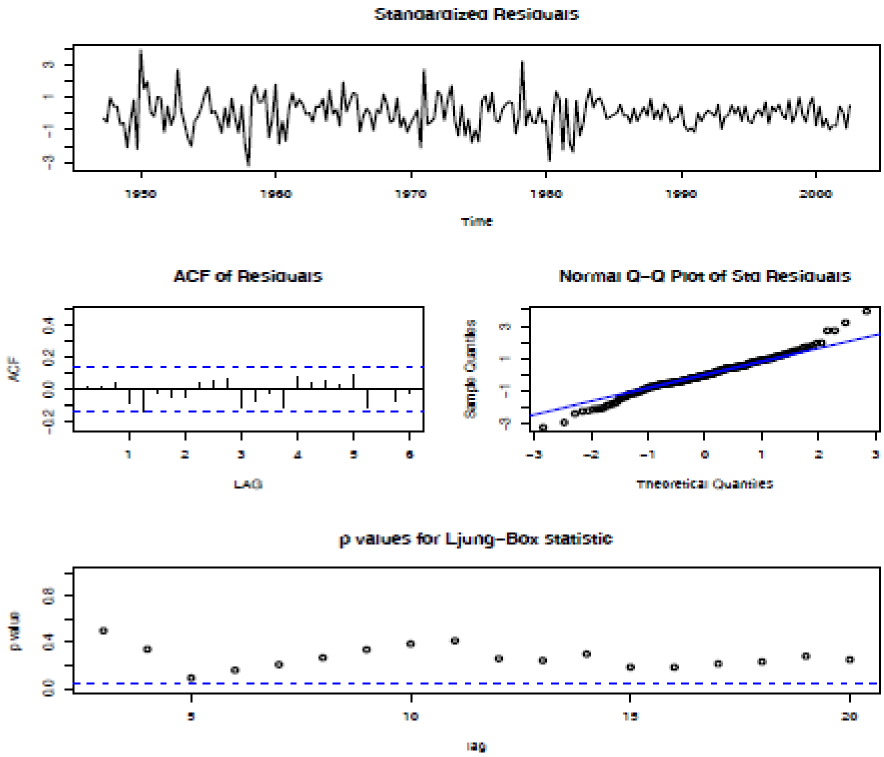

Fig. 3.17. Diagnostics of the residuals from MA(2) fit on GNP growth rate.

Example 3.42 Model Choice for the U.S. GNP Series

Returning to the analysis of the U.S. GNP data presented in Examples 3.38 and 3.39, recall that two models, an AR(1) and an MA(2), fit the GNP growth rate well. To choose the final model, we compare the AIC, the AICc, and the BIC for both models.

```
sarima(gnpgr, 1, 0, 0)  # AR(1)
    $AIC: -8.294403  $AICc: -8.284898   $BIC: -9.263748
sarima(gnpgr, 0, 0, 2)  # MA(2)
    $AIC: -8.297693  $AICc: -8.287854   $BIC: -9.251711
```

The AIC and AICc both prefer the MA(2) fit, whereas the BIC prefers the simpler AR(1) model. It is often the case that the BIC will select a model of smaller order than the AIC or AICc. It would not be unreasonable in this case to retain the AR(1) because pure autoregressive models are easier to work with.

Multiplicative Seasonal ARIMA Models

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t, \tag{3.155}$$

with the following definition.

Definition 3.12 *The operators*

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps} \tag{3.156}$$

and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs} \tag{3.157}$$

are the **seasonal autoregressive operator** *and the* **seasonal moving average operator** *of orders P and Q, respectively, with seasonal period s.*

Analogous to the properties of nonseasonal ARMA models, the pure seasonal ARMA$(P, Q)_s$ is causal only when the roots of $\Phi_P(z^s)$ lie outside the unit circle, and it is invertible only when the roots of $\Theta_Q(z^s)$ lie outside the unit circle.

Example 3.43 A Seasonal ARMA Series

A first-order seasonal autoregressive moving average series that might run over months could be written as

$$(1 - \Phi B^{12})x_t = (1 + \Theta B^{12})w_t$$

or

$$x_t = \Phi x_{t-12} + w_t + \Theta w_{t-12}.$$

This model exhibits the series x_t in terms of past lags at the multiple of the yearly seasonal period $s = 12$ months. It is clear from the above form that estimation and forecasting for such a process involves only straightforward modifications of the unit lag case already treated. In particular, the causal condition requires $|\Phi| < 1$, and the invertible condition requires $|\Theta| < 1$.

Table 3.3. Behavior of the ACF and PACF for Pure SARMA Models

	$AR(P)_s$	$MA(Q)_s$	$ARMA(P,Q)_s$
ACF*	Tails off at lags ks, $k = 1, 2, \dots,$	Cuts off after lag Qs	Tails off at lags ks
PACF*	Cuts off after lag Ps	Tails off at lags ks $k = 1, 2, \dots,$	Tails off at lags ks

*The values at nonseasonal lags $h \neq ks$, for $k = 1, 2, \dots$, are zero.

For the first-order seasonal ($s = 12$) MA model, $x_t = w_t + \Theta w_{t-12}$, it is easy to verify that

$$\begin{aligned} \gamma(0) &= (1 + \Theta^2)\sigma^2 \\ \gamma(\pm 12) &= \Theta\sigma^2 \\ \gamma(h) &= 0, \quad \text{otherwise.} \end{aligned}$$

Thus, the only nonzero correlation, aside from lag zero, is

$$\rho(\pm 12) = \Theta/(1 + \Theta^2).$$

For the first-order seasonal ($s = 12$) AR model, using the techniques of the nonseasonal AR(1), we have

$$\begin{aligned} \gamma(0) &= \sigma^2/(1 - \Phi^2) \\ \gamma(\pm 12k) &= \sigma^2\Phi^k/(1 - \Phi^2) \quad k = 1, 2, \dots \\ \gamma(h) &= 0, \quad \text{otherwise.} \end{aligned}$$

In this case, the only non-zero correlations are

$$\rho(\pm 12k) = \Phi^k, \quad k = 0, 1, 2, \dots .$$

These results can be verified using the general result that $\gamma(h) = \Phi\gamma(h - 12)$, for $h \geq 1$. For example, when $h = 1$, $\gamma(1) = \Phi\gamma(11)$, but when $h = 11$, we have $\gamma(11) = \Phi\gamma(1)$, which implies that $\gamma(1) = \gamma(11) = 0$. In addition to these results, the PACF have the analogous extensions from nonseasonal to seasonal models.

In general, we can combine the seasonal and nonseasonal operators into a multiplicative seasonal autoregressive moving average model, denoted by $ARMA(p, q) \times (P, Q)_s$, and write

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\theta(B)w_t \tag{3.158}$$

Example 3.44 A Mixed Seasonal Model

Consider an ARMA$(0,1) \times (1,0)_{12}$ model

$$x_t = \Phi x_{t-12} + w_t + \theta w_{t-1},$$

where $|\Phi| < 1$ and $|\theta| < 1$. Then, because x_{t-12}, w_t, and w_{t-1} are uncorrelated, and x_t is stationary, $\gamma(0) = \Phi^2\gamma(0) + \sigma_w^2 + \theta^2\sigma_w^2$, or

$$\gamma(0) = \frac{1+\theta^2}{1-\Phi^2}\,\sigma_w^2.$$

In addition, multiplying the model by x_{t-h}, $h > 0$, and taking expectations, we have $\gamma(1) = \Phi\gamma(11) + \theta\sigma_w^2$, and $\gamma(h) = \Phi\gamma(h-12)$, for $h \geq 2$. Thus, the ACF for this model is

$$\begin{aligned} \rho(12h) &= \Phi^h \quad h = 1, 2, \ldots \\ \rho(12h-1) = \rho(12h+1) &= \frac{\theta}{1+\theta^2}\Phi^h \quad h = 0, 1, 2, \ldots, \\ \rho(h) &= 0, \quad \text{otherwise.} \end{aligned}$$

The ACF and PACF for this model, with $\Phi = .8$ and $\theta = -.5$, are shown in Figure 3.20. These type of correlation relationships, although idealized here, are typically seen with seasonal data.

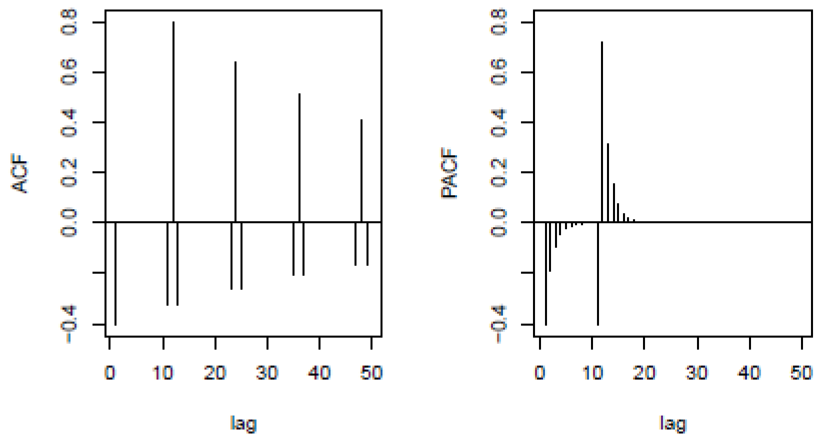

Fig. 3.20. ACF and PACF of the mixed seasonal ARMA model $x_t = .8x_{t-12} + w_t - .5w_{t-1}$.

Definition 3.13 *The multiplicative* **seasonal autoregressive integrated moving average** *model, or* **SARIMA** *model is given by*

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \delta + \Theta_Q(B^s)\theta(B)w_t, \tag{3.160}$$

where w_t is the usual Gaussian white noise process. The general model is denoted as **ARIMA**$(p,d,q) \times (P,D,Q)_s$. *The ordinary autoregressive and moving average components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q, respectively [see (3.5) and (3.18)], and the seasonal autoregressive and moving average components by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ [see (3.156) and (3.157)] of orders P and Q and ordinary and seasonal difference components by $\nabla^d = (1-B)^d$ and $\nabla_s^D = (1-B^s)^D$.*

Example 3.46 The Federal Reserve Board Production Index

A problem of great interest in economics involves first identifying a model within the Box–Jenkins class for a given time series and then producing forecasts based on the model. For example, we might consider applying this methodology to the Federal Reserve Board Production Index shown in Figure 3.21. For demonstration purposes only, the ACF and PACF for this series are shown in Figure 3.22. We note that the trend in the data, the slow decay in the ACF, and the fact that the PACF at the first lag is nearly 1, all indicate nonstationary behavior.

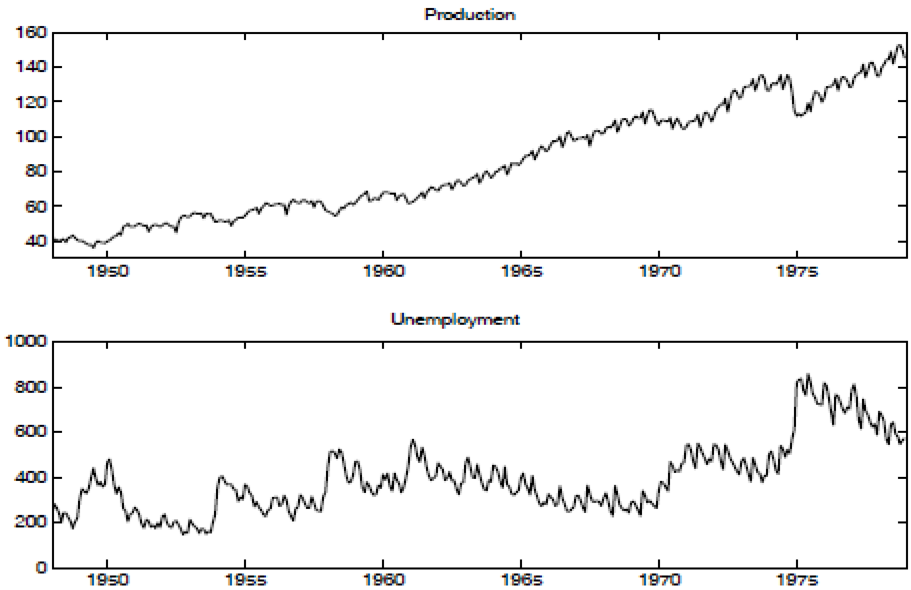

Fig. 3.21. Values of the Monthly Federal Reserve Board Production Index and Unemployment (1948-1978, $n = 372$ months).

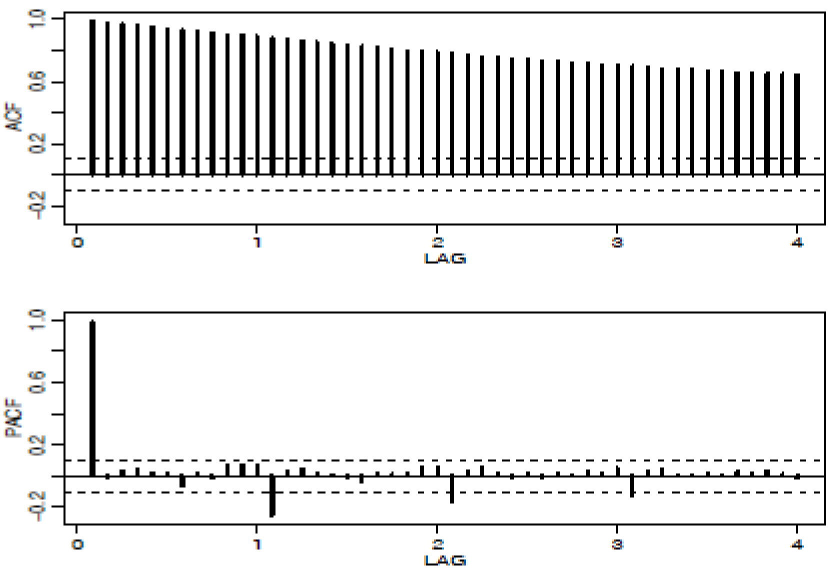

Fig. 3.22. ACF and PACF of the production series.

Following the recommended procedure, a first difference was taken, and the ACF and PACF of the first difference

$$\nabla x_t = x_t - x_{t-1}$$

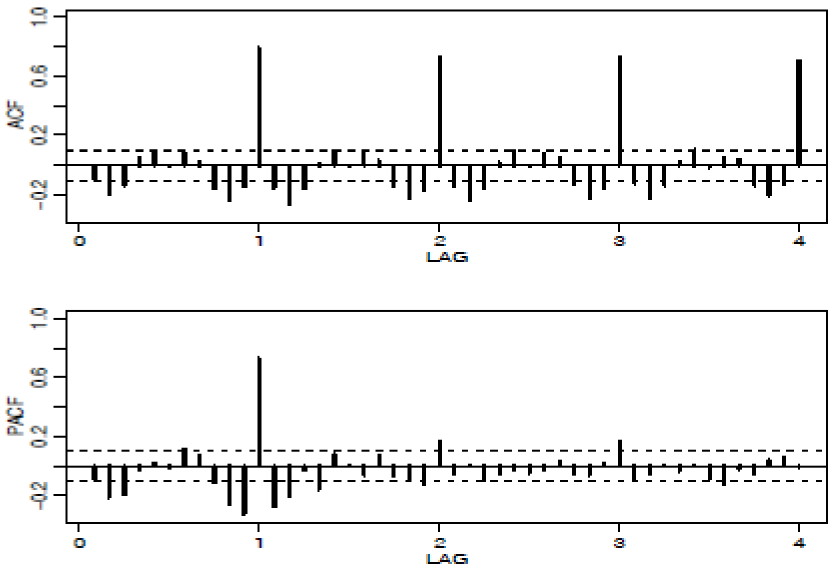

Fig. 3.23. ACF and PACF of differenced production, $(1 - B)x_t$.

$$\nabla_{12}\nabla x_t = (1 - B^{12})(1 - B)x_t.$$

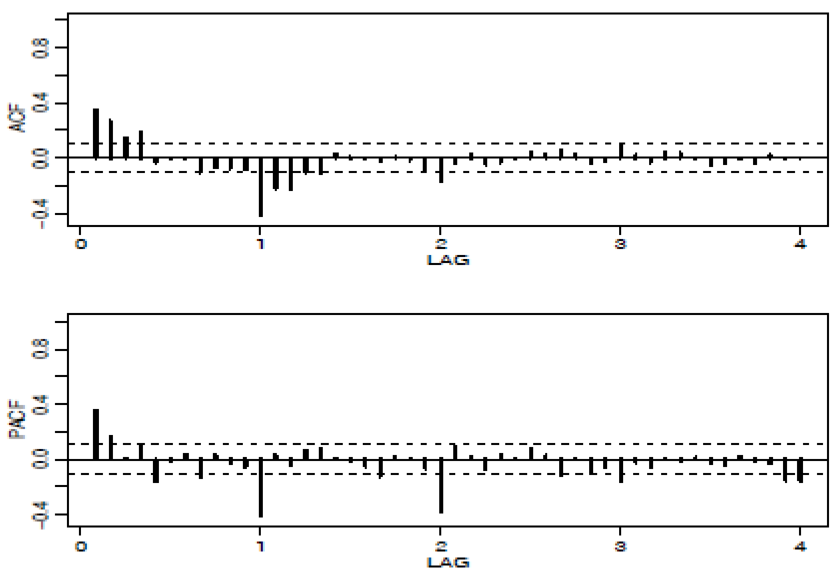

Fig. 3.24. ACF and PACF of first differenced and then seasonally differenced production, $(1 - B)(1 - B^{12})x_t$.

Fitting the three models suggested by these observations we obtain:

(i) ARIMA$(2,1,0)\times(0,1,1)_{12}$:

AIC$=1.372$, AICc$=1.378$, BIC$=.404$

(ii) ARIMA$(2,1,0)\times(0,1,3)_{12}$:

AIC$=1.299$, AICc$=1.305$, BIC$=.351$

(iii) ARIMA$(2,1,0)\times(2,1,1)_{12}$:

AIC$=1.326$, AICc$=1.332$, BIC$=.379$

The ARIMA$(2,1,0)\times(0,1,3)_{12}$ is the preferred model, and the fitted model in this case is

$$(1-.30_{(.05)}B-.11_{(.05)}B^2)\nabla_{12}\nabla\widehat{x}_t$$
$$=(1-.74_{(.05)}B^{12}-.14_{(.06)}B^{24}+.28_{(.05)}B^{36})\widehat{w}_t$$

with $\widehat{\sigma}^2_w=1.312$.

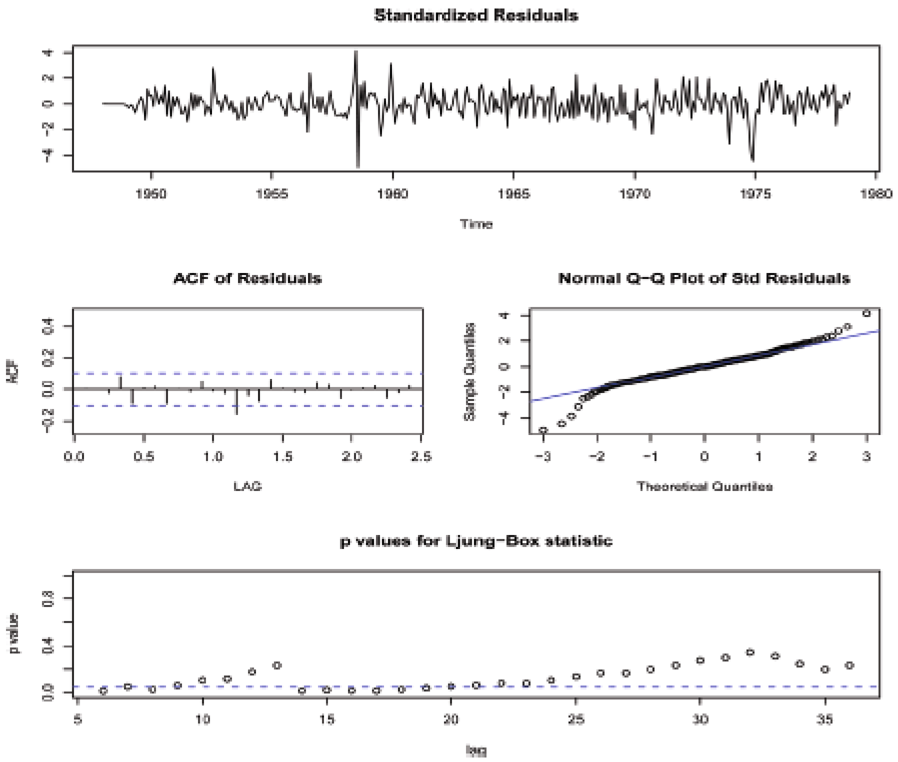

Fig. 3.25. Diagnostics for the ARIMA$(2,1,0)\times(0,1,3)_{12}$ fit on the Production Index.