Time Series (MATH5845)

Dr. Atefeh Zamani
Based on the notes by Prof. William T.M. Dunsmuir

T2 2025

Chapter 7

ARIMA Models

Based on Sections 3.6, 3.7 and 3.9 of Shumway et al. [2000]

Contents

The steps of selecting an appropriate model for $\{X_t,\ t = 1, \ldots, n\}$ are

- If the data

  (a) exhibits no apparent deviations from stationarity

  (b) has a rapidly decreasing autocorrelation function,

  we shall seek a suitable ARMA process to represent the mean-corrected data.

- If not,

  – Look for a transformation of the data which generates a new series with the properties (a) and (b), which can frequently be achieved by differencing (ARIMA (autoregressive-integrated moving average) or SARIMA (Seasonal ARIMA) processes).

  – Once the data has been suitably transformed, the problem becomes one of finding a satisfactory ARMA(p, q) model, and in particular of choosing (or identifying) p and q.

7.1 Integrated Models for Nonstationary Data

In many situations, time series can be thought of as being composed of two components:

1. a nonstationary trend component,

2. zero-mean stationary component.

Differencing such a process will lead to a stationary processes.

Example 7.1 • *Consider the model*

$$X_t = W_t + Y_t \tag{7.1}$$

where $W_t = \beta_0 + \beta_1 t$ *and* Y_t *is stationary. Differencing such a process will lead to a stationary process:*

$$\bigtriangledown X_t = X_t - X_{t-1} = \beta_1 + \bigtriangledown Y_t.$$

- *In* (7.1), *let* W_t *be stochastic and slowly varying according to a random walk. That is,* $W_t = W_{t-1} + V_t$, *where* V_t *is stationary. First differencing makes this process stationary, since*
$$\bigtriangledown X_t = V_t + \bigtriangledown Y_t$$

- *If* W_t *in* (7.1) *is a* k*-th order polynomial,* $W_t = \sum_{j=0}^{k} \beta_j t^j$*, then the differenced series* $\nabla^k X_t$ *is stationary.*

Stochastic trend models can also lead to higher order differencing.

Example 7.2 *In the previous example, suppose* $W_t = W_{t-1} + V_t$ *and* $V_t = V_{t-1} + E_t$*, where* E_t *is stationary. Then,* $\bigtriangledown X_t = V_t + \bigtriangledown Y_t$ *is not stationary, but* $\bigtriangledown^2 X_t = E_t + \bigtriangledown^2 Y_t$ *is stationary.*

In the previous chapters, we talked about how ARMA models are useful for representing stationary series. Now, we introduce ARIMA processes, which are a broader class that can handle non-stationary series. These processes become ARMA models after applying a finite number of differences.

Definition 7.1 *A process X_t is said to be ARIMA(p, d, q) if*

$$\bigtriangledown^d X_t = (1 - B)^d X_t$$

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

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t. \tag{7.2}$$

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

$$\phi(B)(1 - B)^d X_t = \delta + \theta(B) Z_t.$$

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

Note 7.1

- *The process $\{X_t\}$ is stationary if and only if $d = 0$, in which case it reduces to an ARMA(p, q) process.*

- *If $d \geq 1$, we can add an arbitrary polynomial trend of degree $(d - 1)$ to $\{X_t\}$, without violating the difference equation (7.2).*

- *ARIMA models are useful for representing data with trend.*

- *Since for $d \geq 1$, equation (7.2) determines the second order properties of $(1 - B)^d X_t$, but not those of $\{X_t\}$, estimation of parameters will be based on the observed differences $(1 - B)^d X_t$.*

Example 7.3 *Let X_t be an ARIMA(1,1,0) process, for some $\phi \in (-1,1)$,*

$$(1-\phi B)(1-B)X_t = W_t, \quad \{W_t\} \sim WN(0,\sigma_W^2).$$

We can then write

$$X_t = X_0 + \sum_{j=1}^{t} Y_j, \quad t \geq 1,$$

where

$$Y_t = (1-B)X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}.$$

A realization of $\{X_1, \ldots, X_{200}\}$ with $\phi = 0.8$ and $\sigma_W = 1$ is shown in Figure 7.1 together with the sample ACF and PACF.

- *A distinctive feature which suggests the appropriateness of an ARIMA model is the slowly decaying positive sample ACF.*

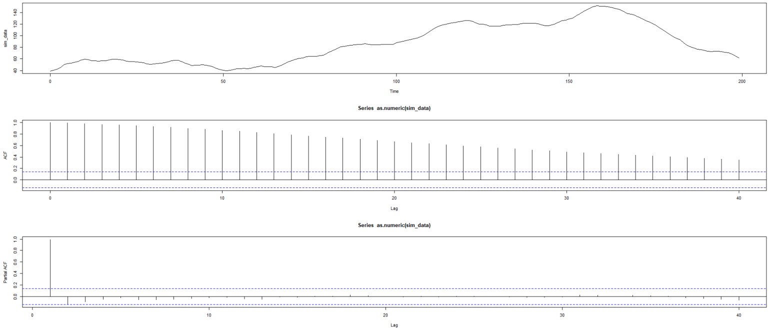

Figure 7.1: A realization of the ARIMA process of Example 7.3 with its sample ACF and PACF.

- *If we were given only the data and wished to find an appropriate model it would be natural to apply the operator* $\bigtriangledown = 1 - B$ *repeatedly in the hope that for some* i, $\{\bigtriangledown^i X_t,\}$ *will have a rapidly decaying sample ACF compatible with that of an ARMA process with no zeroes of the autoregressive polynomial near the unit circle.*

One application of the operator $\bigtriangledown$ *on this time series produces the realization shown in Figure 7.2*

- *The sample ACF and PACF suggest an AR(1) model for* $\{\bigtriangledown^j X_t\}$, *(coefficient estimation: MLE)*

$$(1 - 0.8507B)(1 - B)X_t = W_t, \quad W_t \sim WN(0, 1.035),$$

- *Instead of differencing, we could proceed more directly by fitting an AR(2) process as suggested by the sample PACF (coefficient estimation: OLS),*

$$(1 - 1.8490B + 0.8521B^2)X_t = Z_t, \quad Z_t \sim WN(0, 1.028).$$

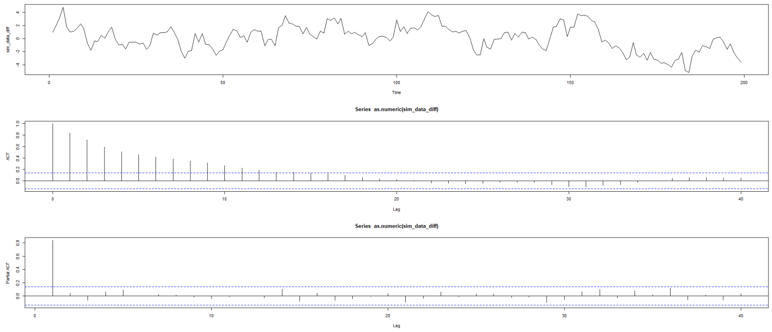

Figure 7.2: The differenced series in Example 7.3 with its sample ACF and PACF.

- Because of the nonstationarity, care must be taken when deriving forecasts.
- Since $Y_t = \bigtriangledown^d X_t$ is ARMA, we can use ARMA forecasting 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_{n+m} for $m = 1, 2, \ldots$, we have $Y^n_{n+m} = X^n_{n+m} - X^n_{n+m-1}$, so that

    $$X^n_{n+m} = Y^n_{n+m} + X^n_{n+m-1}$$

    with initial condition $X^n_{n+1} = Y^n_{n+1} + X_n$ (noting $X^n_n = X_n$).
- It is a little more difficult to obtain the prediction errors, but for large n, the approximation works well:

$$E(X_{n+m} - X^n_{n+m})^2 = \sigma^2_W \sum_{j=0}^{m-1} {\psi_j}^2, \tag{7.3}$$

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

Example 7.4 (RandomWalk with Drift) *Consider the random walk with drift model:*

$$X_t = \delta + X_{t-1} + Z_t, \quad t = 1, 2, \ldots, \quad X_0 = 0.$$

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*

$$X_{n+1}^n = E(X_{n+1}|X_n, \ldots, X_1) = E(\delta + X_n + Z_{n+1}|X_n, \ldots, X_1) = \delta + X_n.$$

- *The two-step-ahead forecast is given by* $X_{n+2}^n = \delta + X_{n+1}^n = 2\delta + X_n$.

- *The m-step-ahead forecast, for* $m = 1, 2, \ldots,$ *is*

$$X_{n+m}^n = m\delta + X_n. \tag{7.4}$$

To obtain the forecast errors, recall that $X_n = n\delta + \sum_{j=1}^{n} Z_j$ *in which case we may write*

$$X_{n+m} = (n+m)\delta + \sum_{j=1}^{n+m} Z_j = m\delta + X_n + \sum_{j=n+1}^{n+m} Z_j$$

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

$$E(X_{n+m} - X_{n+m}^{n})^2 = E\left(\sum_{j=n+1}^{n+m} Z_j\right)^2 = m\sigma_Z^2. \tag{7.5}$$

- *Unlike the stationary case, as the forecast horizon grows, the prediction errors increase without bound and the forecasts follow a straight line with slope* δ *emanating from* X_n.
- *m-step-ahead prediction erro is exact in this case because* $\psi(z) = 1/(1-z) = \sum_{j=0}^{\infty} z^j$, *for* $|z| < 1$, *so that* $\psi = 1$ *for all* j.

- *If Z_t are Gaussian, estimation is straightforward because the differenced data, say $Y_t = \bigtriangledown X_t$, are independent and identically distributed normal variates with mean δ and variance σ_Z^2. Consequently, optimal estimates of δ and σ_Z^2 are the sample mean and variance of the Y_t, respectively.*

Example 7.5 *[IMA(1, 1) and EWMA] The ARIMA(0,1,1), or IMA(1,1) model is of interest because many economic time series can be successfully modeled this way. In addition, the model leads to a frequently used, and abused, forecasting method called* **exponentially weighted moving averages** *(EWMA).*

- *Write the model as*

$$X_t = X_{t-1} + Z_t - \lambda Z_{t-1}, \tag{7.6}$$

  with $|\lambda| < 1$, *for* $t = 1, 2, \ldots$, *and* $x_0 = 0$, *(standard representation for EWMA with no drift).*

- *If we write*

$$Y_t = Z_t - \lambda Z_{t-1},$$

  we may write (7.6) *as* $X_t = X_{t-1} + Y_t$.

- *Because* $|\lambda| < 1$, Y_t *has an invertible representation,* $Y_t = \sum_{j=1}^{\infty} \lambda^j Y_{t-j} + Z_t$, *and substituting* $Y_t = X_t - X_{t-1}$, *for large* t *with* $X_t = 0$ *for* $t \leq 0$, *we may* *write*

$$X_t = \sum_{j=1}^{\infty}(1-\lambda)\lambda^{j-1}X_{t-j} + Z_t. \tag{7.7}$$

- *Using the approximation* (7.7), *we have that the approximate one-step-ahead predictor, is*

$$\begin{aligned}\tilde{X}_{n+1} &= \sum_{j=1}^{\infty}(1-\lambda)\lambda^{j-1}X_{n+1-j} \\ &= (1-\lambda)X_n + \lambda\sum_{j=1}^{\infty}(1-\lambda)\lambda^{j-1}X_{n-j} \\ &= (1-\lambda)X_n + \lambda\tilde{X}_n. \end{aligned} \tag{7.8}$$

- *The new forecast is a linear combination of the old forecast and the new observation.*

- *Based on* (7.8) *and the fact that we only observe* $X_1, \ldots, X_n$, *and consequently* $Y_1, \ldots, Y_n$ *(because* $Y_t = X_t - X_{t-1}$; $X_0 = 0$*), the truncated forecasts are*

$$\tilde{X}_{n+1}^{n} = (1-\lambda)X_n + \lambda \tilde{X}_n^{n-1}, \quad n \geq 1, \tag{7.9}$$

  with $\tilde{X}_1^0 = X_1$ *as an initial value.*

- *The mean-square prediction error can be approximated using* (7.3) *by noting that* $\psi(z) = (1-\lambda z)/(1-z) = 1 + (1-\lambda)\sum_{j=1}^{\infty} z^j$ *for* $|z| < 1.$;

- *Consequently, for large* n, (7.3) *leads to*

$$E(X_{n+m} - X_{n+m}^n)^2 \approx \sigma_W^2[1 + (m-1)(1-\lambda)^2].$$

- *In EWMA, the parameter* $1-\lambda$ *is often called the smoothing parameter and is restricted to be between zero and one. Larger values of* λ *lead to smoother forecasts.*

This method of forecasting is popular because it is easy to use; we need only retain the previous forecast value and the current observation to forecast the next time period. Unfortunately, as previously suggested, the method is often abused because some forecasters do not verify that the observations follow an IMA(1, 1) process, and often arbitrarily pick values of λ.

7.2 Building ARIMA Models

There are a few basic steps to fitting ARIMA models to time series data. These steps involve

1. plotting the data
2. possibly transforming the data,
3. identifying the dependence orders of the model,
4. parameter estimation,
5. diagnostics,
6. model choice.

Plotting the data

First, as with any data analysis, we should construct a time plot of the data, and inspect the graph for any anomalies. If, for example, the variability in the data grows with time, it will be necessary to transform the data to stabilize the variance.

Possibly transforming the data

If the variation in the data is not stable, the Box-Cox class of power transformations could be employed.

Definition 7.2 (Box-Cox Transformations) *The family of Box-Cox transformations are a useful family of transformations, that includes both logarithms and power transformations, which depend on the parameter λ and are defined as follows:*

$$Y_t = \begin{cases} (X_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log(X_t) & \lambda = 0 \end{cases} \tag{7.10}$$

- There exists methods for choosing the power λ (Not discussed here).
- Often, transformations are also used to improve the approximation to normality.

- The particular application might suggest an appropriate transformation.
  - For example, we have seen numerous examples where the data behave as $X_t = (1 + p_t)X_{t-1}$, where p_t is a small percentage change from period $t-1$ to t, which may be negative. If p_t is a relatively stable process, then $\bigtriangledown \log(X_t) \approx p_t$ will be relatively stable.

Identifying the dependence orders of the model

After suitably transforming the data, the next step is to identify preliminary values of the autoregressive order, p, the order of differencing, d, and the moving average order, q.

- A time plot of the data will typically suggest whether any differencing is needed.
  - If differencing is needed, then difference the data once, $d = 1$, and inspect the time plot of $\bigtriangledown(X_t)$.
  - If additional differencing is necessary, then try differencing again and inspect a time plot of $\bigtriangledown^2(X_t)$.
  - Be careful not to overdifference because this may introduce dependence where none exists.

  For example, $X_t = Z_t$ is serially uncorrelated, but $\bigtriangledown(X_t) = Z_t - Z_{t-1}$ is MA(1).

- The sample ACF can help in indicating whether differencing is needed.
  - A slow decay in $\hat{\rho}(h)$, is an indication that differencing may be needed.
- When preliminary values of d have been settled, the next step is to look at the sample ACF and PACF of $\bigtriangledown^d(X_t)$ for whatever values of d have been chosen.
  - Note that it cannot be the case that both the ACF and PACF cut off.
  - Because we are dealing with estimates, it will not always be clear whether the sample ACF or PACF is tailing off or cutting off.
  - so, two models that are seemingly different can actually be very similar.
- We should not worry about being so precise at this stage of the model fitting.
- At this point, a few preliminary values of p, d, and q should be at hand, and we can start estimating the parameters.

Diagnostic Checking

The next step in model fitting is diagnostic checking. This investigation includes

- Analysis of the residuals
- Model comparisons.

1. The first step involves a time plot of the innovations (or residuals), $x_t - \hat{x}_t^{t-1}$, or of the standardized innovations

$$e_t = \left(x_t - \hat{x}_t^{t-1}\right) / \sqrt{\hat{\nu}_{t-1}} \tag{7.11}$$

where $\hat{x}_t^{t-1}$ is the one-step-ahead prediction of x_t based on the fitted model and $\hat{\nu}_{t-1}$ is the estimated one-step-ahead error variance.

- If the model fits well, the standardized residuals should behave as an iid sequence with mean zero and variance one.
- The time plot should be inspected for any obvious departures from this assumption.
- Unless the time series is Gaussian, it is not enough that the residuals are uncorrelated.

For example, it is possible in the non-Gaussian case to have an uncorrelated process for which values contiguous in time are highly dependent.

2. Investigation of marginal normality can be accomplished visually by looking at a histogram of the residuals. In addition to this, a normal probability plot or a Q-Q plot can help in identifying departures from normality.

3. There are several tests of randomness (runs test) that could be applied to the residuals.

4. Inspect the sample ACF of the residuals, say, $\hat{\rho}_e(h)$, for any patterns or large values.

    - For a white noise sequence, the sample autocorrelations are approximately independently and normally distributed with zero means and variances $1/n$.
    - A good check on the correlation structure of the residuals is to plot $\hat{\rho}_e(h)$ versus h along with the error bounds of $\pm 2/\sqrt{n}$.
    - The residuals from a model fit, however, will not quite have the properties of a white noise sequence and the variance of $\hat{\rho}_e(h)$ can be much less than $1/n$.
    - This part of the diagnostics can be viewed as a visual inspection of $\hat{\rho}_e(h)$ with the main concern being the detection of obvious departures from the independence assumption.

5. In addition to plotting $\hat{\rho}_e(h)$, we can perform a general test that takes into consideration the magnitudes of $\hat{\rho}_e(h)$ as a group.

   - The Ljung-Box-Pierce Q-statistic given by

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}_e^2(h)}{n-h}. \tag{7.12}$$

   - The value H in (7.12) is chosen somewhat arbitrarily, typically, $H = 20$.
   - Under the null hypothesis of model adequacy, asymptotically ($n \to \infty$), $Q \sim \chi^2_{H-p-q}$.
   - Reject the null hypothesis at level α if $Q > \chi^2_{(1-\alpha),H-p-q}$.
   - The basic idea is that if W_t is white noise, then $n\hat{\rho}_e^2(h)$, for $h = 1, \ldots, H$, are asymptotically independent χ^2_1 random variables. This means that $n \sum_{h=1}^{H} \hat{\rho}_e^2(h)$ is approximately a χ^2_H random variable.

Model choice

The final step of model fitting is model choice or model selection. That is, we must decide which model we will retain for forecasting. The most popular techniques are AIC, AICc, and BIC. These criteria help us to simply evaluate each model on its own merits instead of a sequential procedure for model selection.

Suppose we consider a normal time series model with k coefficients and denote the maximum likelihood estimator for the variance as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n}. \tag{7.13}$$

The following criteria are based on measuring the goodness of fit for a particular model by balancing the error of the fit against the number of parameters in the model.

Definition 7.3 (Akaike's Information Criterion (AIC))

$$AIC = \log(\hat{\sigma}^2) + \frac{n + 2k}{n} \tag{7.14}$$

where k is the number of parameters in the model.

- The value of k yielding the **minimum AIC** specifies the best model.
- $\hat{\sigma}^2$ decreases monotonically as k increases.
- We ought to penalize the error variance by a term proportional to the number of parameters.
- The choice for the penalty term given by (7.14) is not the only one.

Definition 7.4 (AIC, Bias Corrected (AICc))

$$AICc = \log(\hat{\sigma}^2) + \frac{n+k}{n-k-2}. \tag{7.15}$$

As with the AIC, the AICc should be minimised. Another correction of AIC is based on Bayesian arguments, which leads to the following.

Definition 7.5 (Bayesian Information Criterion (BIC))

$$BIC = \log(\hat{\sigma}^2) + \frac{k\log(n)}{n}. \tag{7.16}$$

- BIC is also called the Schwarz Information Criterion (SIC).
- The penalty term in BIC is much larger than in AIC.
- **BIC tends to choose smaller models**.
- Various simulation studies have tended to verify that **BIC** does well at getting the **correct order in large samples**, whereas **AICc** tends to be superior in **smaller samples** where the **relative number of parameters is large**.

Example 7.6 (Analysis of GNP Data) *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.*

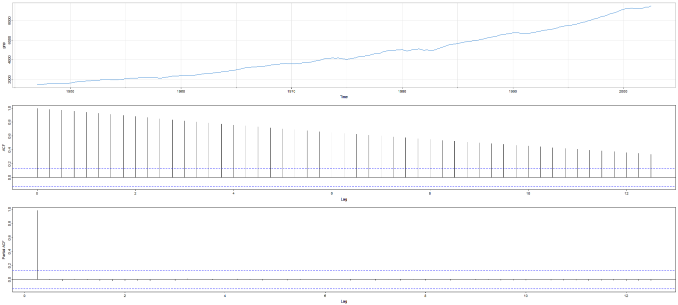

Figure 7.3: Quarterly U.S. GNP from 1947(1) to 2002(3) along with ACF and PACF.

- *Because strong trend, it is difficult to see any other variability in data.*
- *When reports of GNP and similar economic indicators are given, it is often in growth rate* ($x_t = \bigtriangledown \log(y_t)$).

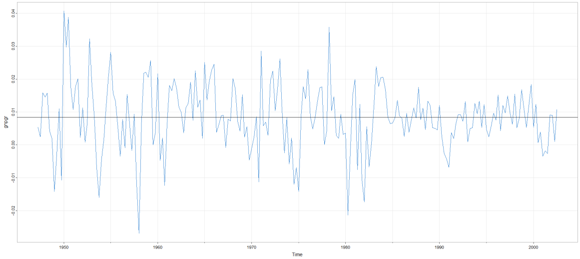

Figure 7.4: U.S. GNP quarterly growth rate.

The sample ACF and PACF of the quarterly growth rate are plotted in Fig. 7.5.

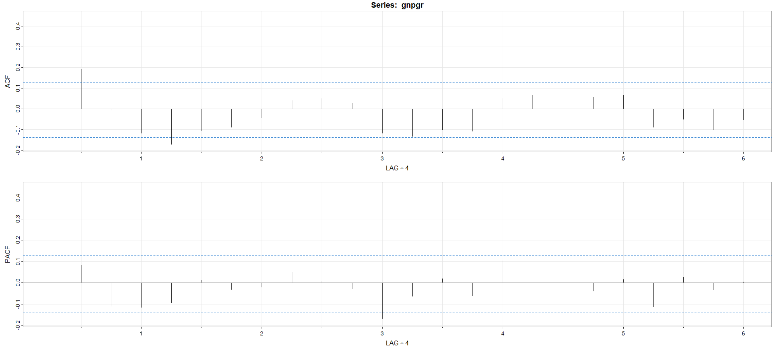

Figure 7.5: Sample ACF and PACF of the GNP quarterly growth rate. Lag is in terms of years.

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

$$\hat{X}_t = .008 + .303\hat{Z}_{t-1} + .204\hat{Z}_{t-2} + \hat{Z}_t, \tag{7.17}$$

where $\hat{\sigma}_Z = .000089$ is based on 219 degrees of freedom.

- *All of the regression coefficients are significant, including the constant.*
- *In this example, not including a constant leads to the wrong conclusions about the nature of the U.S. economy. Not including a constant assumes the average quarterly growth rate is zero, whereas the U.S. GNP average quarterly growth rate is about 1%.*

The estimated AR(1) model is

$$\hat{X}_t = .008(1 - .347) + .347\hat{X}_{t-1} + \hat{Z}_t, \tag{7.18}$$

where $\hat{\sigma}_Z = .000090$ *on 220 degrees of freedom; note that the constant in* (7.18) *is* $.008(1 - .347) = .005$. *We will discuss diagnostics next, but assuming both of these models fit well, how are we to reconcile the apparent differences of the estimated models?*

In fact, the fitted models are nearly the same. To show this, consider an AR(1) model of the form in (7.18) *without a constant term; that is,*

$$X_t = .35X_{t-1} + \hat{Z}_t,$$

and write it in its causal form, $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, *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,*

$$X_t \approx .35Z_{t-1} + .12Z_{t-2} + Z_t,$$

which is similar to the fitted MA(2) model in (7.17).

Diagnostics for MA(2)

- *Inspection of the time plot of the standardized residuals in Fig. 7.6 shows no obvious patterns. Notice that there may be outliers, with a few values exceeding 3 standard deviations in magnitude.*
- *The ACF of the standardized residuals shows no apparent departure from the model assumptions.*
- *The normal Q-Q plot of the residuals shows that the assumption of normality is reasonable, with the exception of the possible outliers.*
- *The Q-statistic is never significant at the lags shown.*

The model appears to fit well.

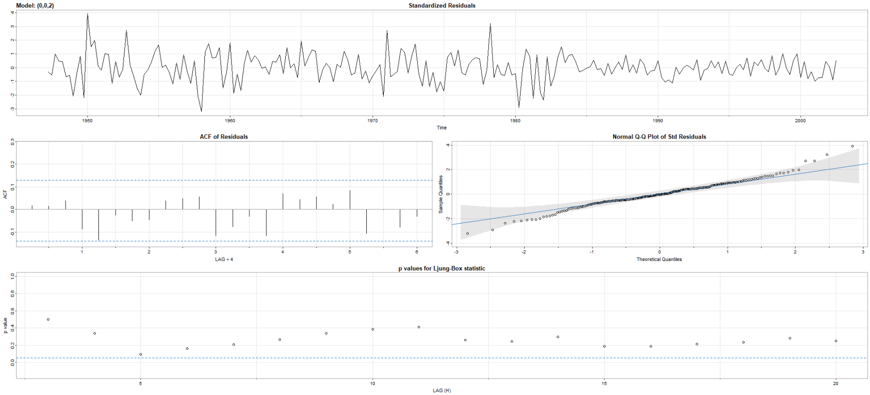

Figure 7.6: Diagnostics of the residuals from MA(2) fit on GNP growth rate

Model Choice

To choose the final model, we compare the AIC, the AICc, and the BIC for both models. 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. In either case, it is not unreasonable to retain the AR(1) because pure autoregressive models are easier to work with.

	AIC	AICc	BIC
AR(1)	-6.44694	-6.446693	-6.400958
MA(2)	-6.450133	-6.449637	-6.388823

Table 7.1: AIC, AICc and BIC of AR(1) and MA(2) models for the U.S. GNP data.

7.3 Multiplicative Seasonal ARIMA Models

- In this section, we introduce several modifications made to the ARIMA model to account for seasonal and nonstationary behavior.
- The idea is that, often, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal lag s.

For example, with monthly economic data, there is a strong yearly component occurring at lags that are multiples of $s = 12$, because of the strong connections of all activity to the calendar year. Data taken quarterly will exhibit the yearly repetitive period at $s = 4$ quarters.

Pure seasonal autoregressive moving average model

The pure seasonal autoregressive moving average model, ARMA$(P, Q)_s$, takes the form

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)Z_t, \tag{7.19}$$

where the operators

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}, \tag{7.20}$$

and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs} \tag{7.21}$$

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 7.7 *[A Seasonal AR Series] A first-order seasonal autoregressive series that might run over months could be written as*

$$(1 - \Phi B^{12})X_t = Z_t,$$

or

$$X_t = \Phi X_{t-12} + Z_t.$$

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 $|\Phi| < 1$. We simulated 3 years of data from the model with $\Phi = .9$, and exhibit the theoretical ACF and PACF of the model, Figure 7.7.

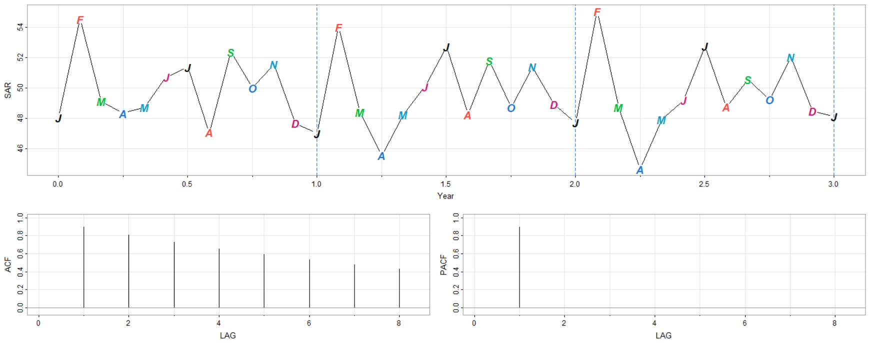

Figure 7.7: Data generated from a seasonal (s = 12) AR(1), and the true ACF and PACF of the model $X_t = .9x_{t-12} + Z_t$

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

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

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

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

In this case, the only non-zero correlations are $\rho(\pm 12k) = \Phi^k$, $k = 0, 1, 2, \ldots$.

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. These results are demonstrated in Figure 7.7.

	$\text{AR}(P)_s$	$\text{MA}(Q)_s$	$\text{ARMA}(P,Q)_s$
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

Table 7.2: Behavior of the ACF and PACF for pure SARMA models (* The values at nonseasonal lags $h \neq ks$, for $k = 1, 2, \dots$, are zero.)

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

$$\Phi_P(B^s)\phi(B)X_t = \Theta_Q(B^s)\theta(B)Z_t, \tag{7.22}$$

as the overall model. Although the diagnostic properties in Table 7.2 are not strictly true for the overall mixed model, the behavior of the ACF and PACF tends to show rough patterns of the indicated form. In fitting such models, focusing on the seasonal autoregressive and moving average components first generally leads to more satisfactory results.

Example 7.8 *[A Mixed Seasonal Model] Consider an* $ARMA(0,1)\times(1,0)_{12}$ *model*

$$X_t = \Phi X_{t-12} + Z_t + \theta Z_{t-1},$$

where $|\Phi| < 1$ *and* $|\theta| < 1$*. Then, because* X_{t-12}*,* Z_t*, and* Z_{t-1} *are uncorrelated, and* X_t *is stationary,* $\gamma(0) = \Phi^2\gamma(0) + \sigma_Z^2 + \theta^2\sigma_Z^2$*, or*

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

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

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

The ACF and PACF for this model, with Φ = .8 and θ = -.5, are shown in Figure 7.8. These type of correlation relationships, although idealized here, are typically seen with seasonal data.

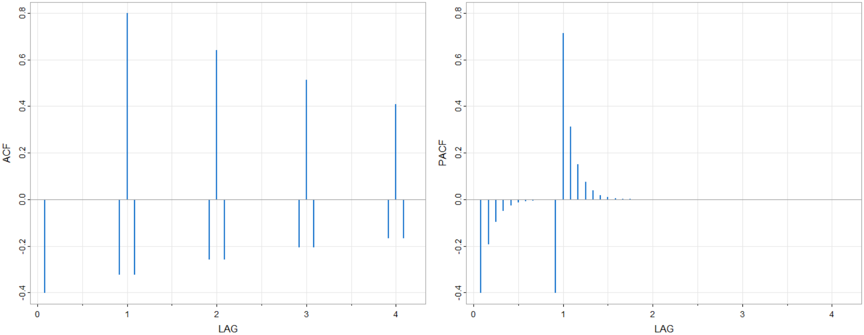

Figure 7.8: ACF and PACF of the mixed seasonal ARMA model $X_t = .8X_{t-12} + Z_t - .5Z_{t-1}$

- Seasonal persistence occurs when the process is nearly periodic in the season.
- For example, with average monthly temperatures over the years, each January would be approximately the same, each February would be approximately the same, and so on.
- In this case, we might think of average monthly temperature X_t as being modeled as
$$X_t = S_t + Z_t,$$
where S_t is a seasonal component that varies a little from one year to the next, according to a random walk,
$$S_t = S_{t-12} + V_t.$$
- In this model, Z_t and V_t are uncorrelated white noise processes. The tendency of data to follow this type of model will be exhibited in a sample ACF that is large and decays very slowly at lags $h = 12k$, for $k = 1, 2, \ldots$.

- If we subtract the effect of successive years from each other, we find that

$$(1 - B^{12})X_t = X_t - X_{t-12} = V_t + Z_t - Z_{t-12}.$$

  This model is a stationary $\mathrm{MA}(1)_{12}$, and its ACF will have a peak only at lag 12.

- In general, seasonal differencing can be indicated when the ACF decays slowly at multiples of some season s, but is negligible between the periods. Then, a seasonal **difference** of order D is defined as

$$\nabla_s^D X_t = (1 - B^s)^D X_t, \tag{7.23}$$

  where $D = 1, 2, \ldots,$ takes positive integer values.

- Typically, $D = 1$ is sufficient to obtain seasonal stationarity.

Definition 7.6 *The* **multiplicative seasonal autoregressive integrated moving average model**, *or SARIMA model (ARIMA$(p, d, q) \times (P, D, Q)_s$.) is given by*

$$\Phi_P(B^s)\phi(B) \bigtriangledown_s^D \bigtriangledown^d X_t = \delta + \Theta_Q(B^s)\theta(B)Z_t, \tag{7.24}$$

where

- Z_t *is the usual Gaussian white noise process.*
- *The ordinary autoregressive and moving average components are the polynomials $\phi(B)$ and $\theta(B)$ of orders p and q.*
- *The seasonal autoregressive and moving average components are polynomials $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q*
- *Ordinary and seasonal difference components are $\bigtriangledown^d = (1 - B)^d$ and $\bigtriangledown_s^D = (1 - B^s)^D$.*

Example 7.9 (An SARIMA Model) *Consider the ARIMA*$(0,1,1) \times (0,1,1)_{12}$ *with* $\delta = 0$*:*

$$\nabla_{12} \nabla X_t = \Theta(B^{12})\theta(B)Z_t \quad \equiv \quad (1-B^{12})(1-B)X_t = (1+\Theta B^{12})(1+\theta B)Z_t. \tag{7.25}$$

Expanding both sides of this model leads to the representation

$$(1 - B - B^{12} + B^{13})X_t = (1 + \theta B + \Theta B^{12} + \Theta\theta B^{13})Z_t,$$

or in difference equation form

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + Z_t + \theta Z_{t-1} + \Theta Z_{t-12} + \Theta\theta Z_{t-13}.$$

- *The multiplicative nature of the model implies that the coefficient of* Z_{t-13} *is the product of the coefficients of* Z_{t-1} *and* Z_{t-12} *rather than a free parameter.*
- *The multiplicative model assumption seems to work well with many seasonal time series data sets while reducing the number of parameters that must be estimated.*

- Selecting the appropriate model for a given set of data from all of those represented by the general form (7.24) is a daunting task.

- We usually think first in terms of finding difference operators that produce a roughly stationary series and then in terms of finding a set of simple autoregressive moving average or multiplicative seasonal ARMA to fit the resulting residual series.

- Differencing operations are applied first, and then the residuals are constructed from a series of reduced length. Next, the ACF and the PACF of these residuals are evaluated. Peaks that appear in these functions can often be eliminated by fitting an autoregressive or moving average components. In considering whether the model is satisfactory, the diagnostic techniques still apply.

Example 7.10 *[Air Passengers] We consider the* `R` *data set AirPassengers, which are the monthly totals of international airline passengers, 1949 to 1960. Various plots of the data and transformed data are shown in Figure 7.9.*

Note that X *is the original series, which shows trend plus increasing variance. The logged data are in log_X, and the transformation stabilizes the variance. The logged data are then differenced to remove trend, and are stored in dlog_X. It is clear the there is still persistence in the seasons (i.e., dlog_X_t $\approx$ dlog_X_{t-12}), so that a twelfth-order difference is applied and stored in ddlog_X. The transformed data appears to be stationary and we are now ready to fit a model.*

The sample ACF and PACF of ddlog_X ($\nabla_{12}\nabla \log X_t$) *are shown in Figure 7.10.*

Seasonsal Component: It appears that at the seasons, the ACF is cutting off a lag $1s$ ($s = 12$)*, whereas the PACF is tailing off at lags* $1s, 2s, 3s, 4s, \ldots$*. These results implies an SMA(1),* $P = 0$, $Q = 1$, *in the season* ($s = 12$).

Non-SeasonsalComponent: Inspecting the sample ACF and PACF at the lower lags, it appears as though both are tailing off. This suggests an ARMA(1, 1) within the seasons, $p = q = 1$.

Thus, we first try an ARIMA$(1, 1, 1) \times (0, 1, 1)_{12}$ *on the logged data, The coeffi-*

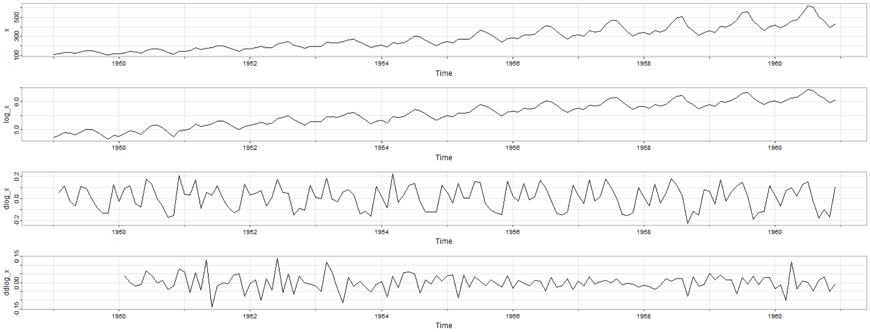

Figure 7.9: `R` data set AirPassengers, which are the monthly totals of international airline passengers x, and the transformed data: $\log X_t$, $\bigtriangledown \log X_t$, and $\bigtriangledown_{12} \bigtriangledown \log X_t$.

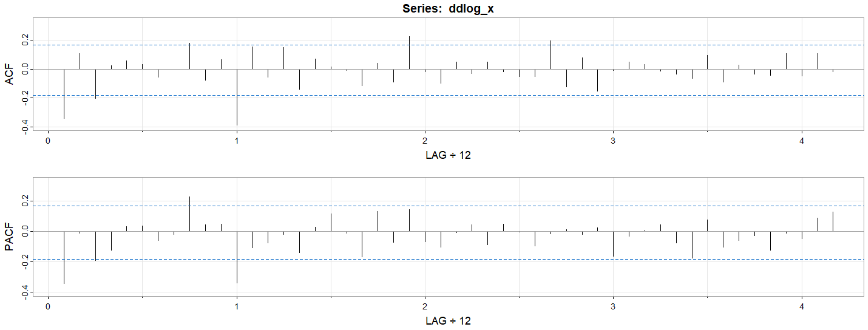

Figure 7.10: Sample ACF and PACF of *ddlog_X* ($\bigtriangledown_{12} \bigtriangledown \log X_t$)

cients of this model are presented in Table 7.3.

However, the AR parameter is not significant, so we should try dropping one

	ar1	ma1	sma1
Coefficients	0.1960	-0.5784	-0.5643
s.e.	0.2475	0.2132	0.0747

Table 7.3: Coefficients of ARIMA$(1,1,1)\times(0,1,1)_{12}$ with σ^2 estimated as 0.001341: log likelihood = 244.95, aic = -481.9

parameter from the within seasons part. In this case,we try an ARIMA$(0,1,1)\times(0,1,1)_{12}$ model and an ARIMA$(1,1,0)\times(0,1,1)_{12}$ model. The coefficients of these two models are displayed in Tables 7.4 and 7.5, respectively.

*All information criteria prefer the ARIMA$(0,1,1)\times(0,1,1)_{12}$ model, which is the model displayed in (**??**). The residual diagnostics are shown in Figure 7.11, and except for one or two outliers, the model seems to fit well.*

Finally, we forecast the logged data out twelve months, and the results are shown in Figure 7.12.

```
x      = AirPassengers
```

	ma1	sma1
Coefficients	-0.4018	-0.5569
s.e.	0.0896	0.0731

Table 7.4: Coefficients of ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ with σ^2 estimated as 0.001348: log likelihood = 244.7, aic = -483.4

	ar1	sma1
Coefficients	-0.3395	-0.5619
s.e.	0.0822	0.0748

Table 7.5: Coefficients of ARIMA$(1, 1, 0) \times (0, 1, 1)_{12}$ with σ^2 estimated as 0.0013678: log likelihood = 243.74, aic = -481.49

```
log_x     = log(x)
dlog_x    = diff(log_x)
ddlog_x   = diff(dlog_x, 12)
```

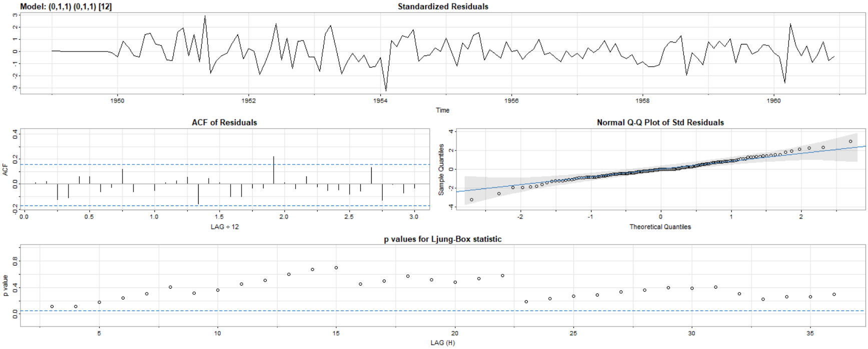

Figure 7.11: Residual analysis for the ARIMA$(0,1,1) \times (0,1,1)_{12}$ fit to the logged air passengers data set

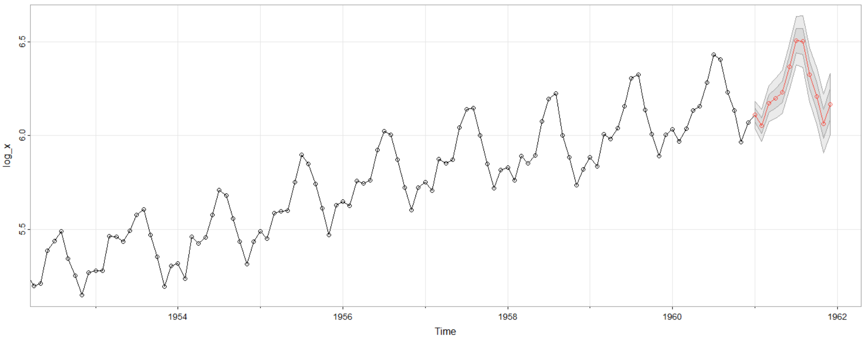

Figure 7.12: Twelve month forecast using the ARIMA$(0,1,1) \times (0,1,1)_{12}$ model on the logged air passenger data set

```
tsplot(cbind(x, log_x, dlog_x, ddlog_x), main="")

acf2(ddlog_x,50)

# below of interest for showing seasonal persistence (not shown here):
par(mfrow=c(2,1))
monthplot(dlog_x)
monthplot(ddlog_x)

sarima(log_x, 1,1,1, 0,1,1, 12)   # model 1
sarima(log_x, 0,1,1, 0,1,1, 12)   # model 2 (the winner)
sarima(log_x, 1,1,0, 0,1,1, 12)   # model 3

dev.new()
sarima.for(log_x, 12, 0,1,1, 0,1,1,12)  # forecasts
```

Listing 7.1: The code used Example 7.10

7.4 Example: Modelling West Virginia Beer Sales

7.4.1 Properties of Beer Sales Series

In this section we develop a simple yet effective model for the monthly sales of beer in West Virginia in the US. We construct the series of litres of ethanol per 100,000 population aged 18 years or over contained in monthly sales of beer. The R code for this analysis if contained in `Chapter6AnalysisWestVABeer.r`

Figure (7.13) shows the original time series and its ACF and PACF. Note that there is general upward trend in the data and variability does not seem to be increasing substantially with trend level so a logarithmic transformation (for variance stabilisation) is not needed here. There is substantial seasonal variation of a reasonably consistent shape over time. The ACF and PACF suggest linear decay at lags 1 to 6 or so and the seasonal peaks in the ACF and PACF are to be expected given the strong seasonal pattern in the series.

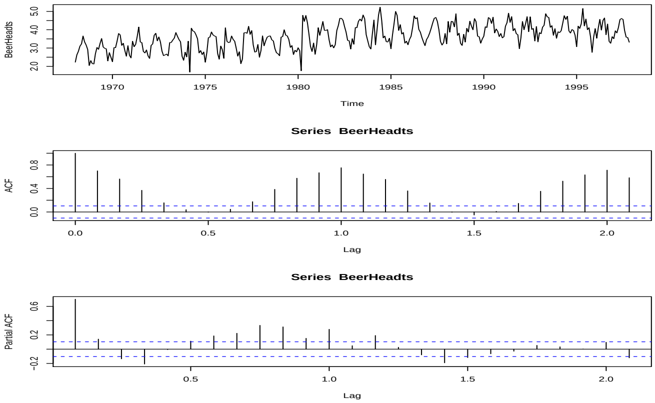

Figure 7.13: Time Series, ACF and PACF of ethanol content of beer sales per head of population aged 18 years and over in West Virginia.

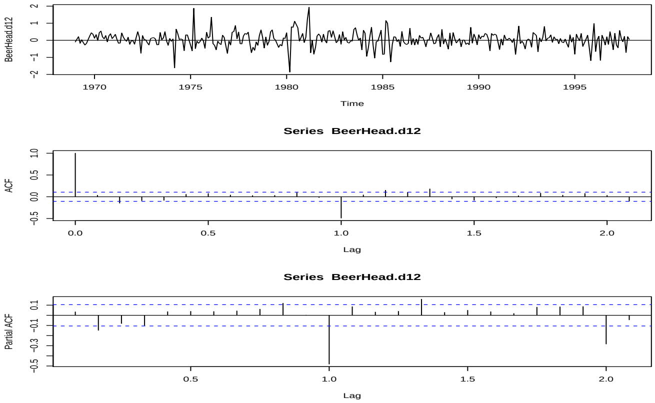

Figure 7.14: Time Series, ACF and PACF of ethanol content of seasonally differenced beer sales per head of population aged 18 years and over in West Virginia.

7.4.2 Seasonal Differenced Series

Because the dominant pattern is the strong seasonal variation we consider seasonally differenced series using the provided R code. Figure (7.14) shows the results for seasonally differenced series.

At this stage there is little to compel us to take lag 1 differences in addition to seasonal differences. We return to this later. The ACF and PACF of the seasonally differenced data shown in Figure 7.14 suggest the need for a moving average term at lag 12 and not much else with the possible exception of some hint of low lag moving average behaviour in the PACF. Based on this we try fitting the ARIMA$(0,0,0) \times (0,1,1)_{12}$ model to the original series. The R-code provided in the `astsa` package of Shumway and Stoffer has a nice way of interfacing to the inbuilt `arima` command in R and also provides graphical diagnostics in one graph. The contents of the `sarima` function is available in `sarima-function-from-atsa-package.R` for your reference.

We apply this function to fit the ARIMA$(0,0,0)\times(0,1,1)_{12}$ model to the original series but note carefully that we include a constant term since there is a slight upward trend in the original series and seasonal differences will conver any such trend into a non-zero mean term in the differenced series. Partial results are as follows:

```
> sarima000011cons
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q),
 seasonal = list(order = c(P, D, Q), period = S),
 xreg = constant,
 optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
         sma1  constant
      -0.7858    0.0036
s.e.   0.0485    0.0004

sigma^2 estimated as 0.1225:
```

```
log likelihood = -123.36,  aic = 252.72
```

In this model the constant term is highly significant with a test statistic $z = 0.0036/0.0004 = 9$. The moving average parameter is also highly significant with $z = -0.7858/0.0485 = -16.2$. The variance of the innovations is estimated to be $\hat{\sigma}^2 = 0.1225$. The graphical display of residuals and their ACF and distributional properties are given in Figure (7.15).

Clearly the residuals are not white noise and substantial and persistent autocorrelation exists for all positive lags. This strongly suggests that an additional lag 1 differencing could be helpful - this was masked in Figure 7.14 and was only when the model was fit that the *residuals* showed this pattern. By the way, you can get all the attributes of the fitted `arima` object by referencing the object `$fit`.

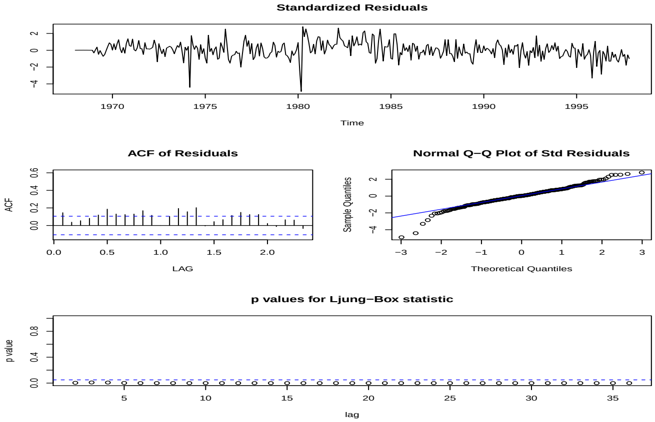

Figure 7.15: Analysis of residuals from ARIMA$(0, 0, 0) \times (0, 1, 1)_{12}$ model with constant term fit to ethanol content of beer sales per head of population aged 18 years and over in West Virginia.

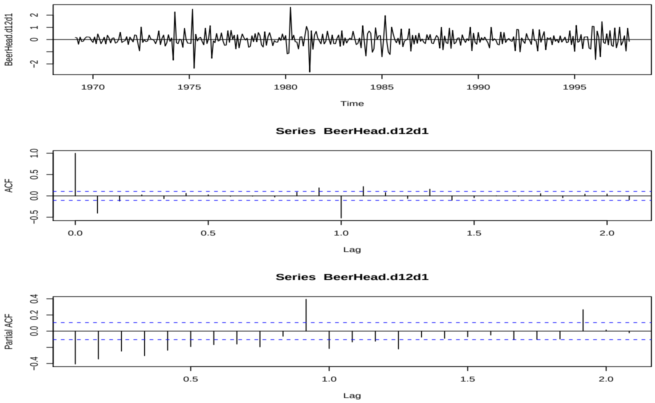

Figure 7.16: Time Series, ACF and PACF of ethanol content of differenced and seasonally differenced beer sales per head of population aged 18 years and over in West Virginia.

7.4.3 Seasonal and lag 1 differenced series

We now consider the seasonal and ordinary diffferenced series in Figure (7.16). The ACF and PACF of this double differenced series strongly suggest that the so-called 'airline' model ARIMA$(0,1,1) \times (0,1,1)_{12}$ would be appropriate but now we will not fit a constant term (since the series are double differenced there would have to be a form of quadratic trend in the original data that would give rise to a non-zero constant in the double differenced series).

The parameter estimates and fit statistics for this model are as follows and the residual diagnostics displayed in Figure (7.17).

```
> sarima011011
$fit

Call:
stats::arima(x = xdata, order = c(p, d, q),
 seasonal = list(order = c(P, D, Q), period = S),
 include.mean = !no.constant,
```

```
 optim.control = list(trace = trc, REPORT = 1, reltol = tol))

Coefficients:
          ma1     sma1
      -0.9237  -0.8935
s.e.   0.0185   0.0357

sigma^2 estimated as 0.1054:
     log likelihood = -112.65,  aic = 231.3
```

Both lag 1 and lag 12 moving average parameters are highly significant. The fit as measured by the AIC criterion is improved over the previous model based on lag 12 differencing only. The estimated innovations variance is $\hat{\sigma}^2 = 0.1054$ which is 89% of that for the previous model.

The ACF of the residuals are improved over the previous model but still shows some evidence of unmodelled autocorrelation. Finding a suitable specification of modifications (i.e. different values of (p, q, P, Q) for the seasonal ARIMA model)is not obvious. Of more immediate concern is the suggestion that the residuals are heavier tailed than normal suggesting some form of volatility in them - we take this up again in the Chapter on GARCH and volatility modelling. There is some evidence of outliers and it would be wise to remodel the series with these removed. Outliers can have large impact on estimation of autocorrelation because the denominator uses sums of squares of values. We will return to these issues later.

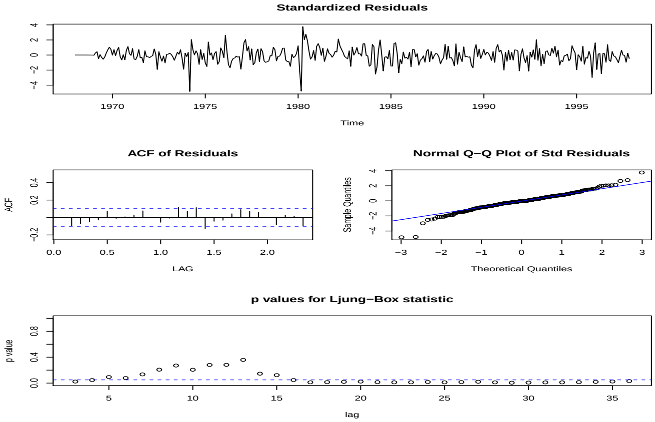

Figure 7.17: Analysis of residuals from ARIMA$(0,1,1) \times (0,1,1)_{12}$ model with constant term fit to ethanol content of beer sales per head of population aged 18 years and over in West Virginia.