Exercises

6.1 Develop an ARIMA model for the growth rate of the U.S. gross domestic product *(GDP_change_2.xlsx)*. Compare your model with the results obtained in Exercise 3.3. Use the diagnostics discussed in Section 6.5 to guide your choice of model.

6.5 Model Diagnostics

One approach to improving our initial model is to use diagnostic methods based upon the residuals. (The residual plots we use here are the same as those we shall use for regression analysis in Chapter 8.) We plot four figures that directly involve the residuals, along with the *ACF* (and *PACF*) of the residuals. The plots may be used to determine the key features of the residuals, as summarized in Table 6.8.

We first consider the diagnostics for the ARIMA(0,1,1) model for U.S. retail sales, initially fitted without a constant. The plots are given in Figures 6.17 and 6.18.

Table 6.8 Interpretation of Diagnostics for Residuals from ARIMA Models

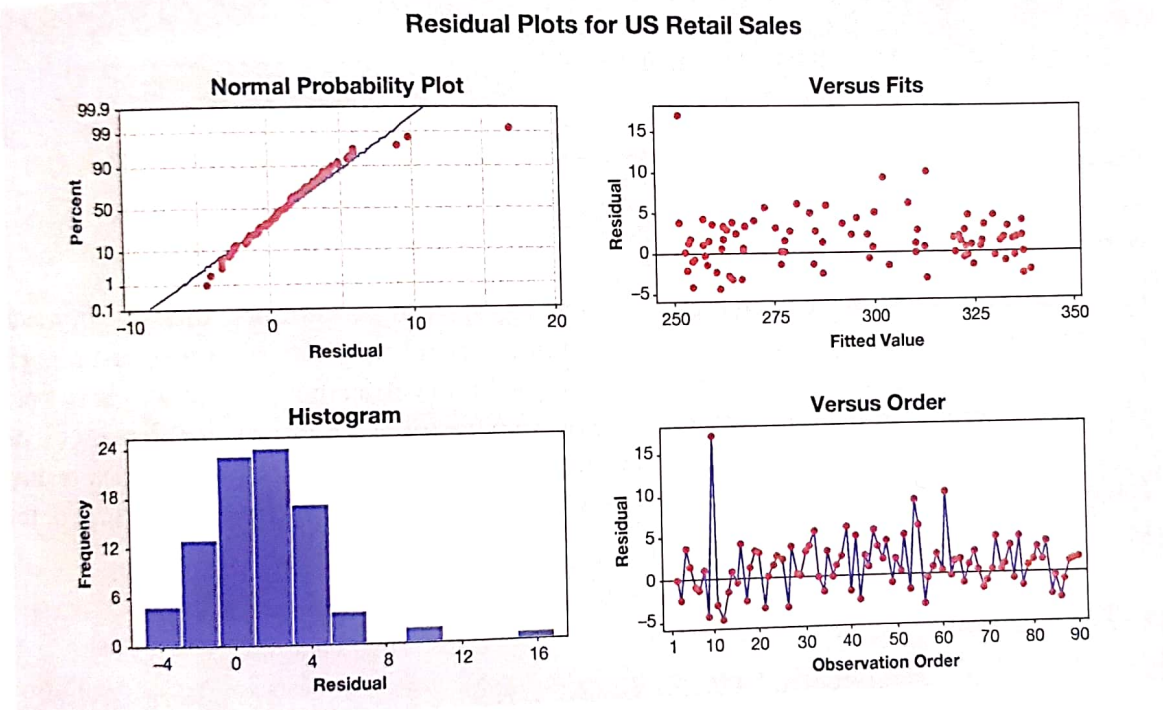
Diagnostic Plots	Issues to be Addressed	Indication of Problems	Remedial Action to Be Taken
Normal probability plot (NPP) and histogram (H)	1. The residuals may not be normally distributed.	1. The residuals do not follow a straight line (NPP), or the histogram (H) is not bell shaped.	1. Consider a transformation.
	2. Are there any outliers?	2. Individual points are far removed from the overall plot (NPP) or other observations (H).	2. Adjust outliers.
Residuals versus fitted values	1. Is there evidence of nonlinearity?	1. The plot shows curvature.	1. Consider a transformation.
	2. Are there any outliers?	2. Individual points are far removed from the overall plot.	2. Adjust outliers.
	3. Does the series show increasing or decreasing variance with the size of the observations (heteroscedasticity)?	3. The plot forms an increasing or decreasing funnel shape.	3. Consider a transformation.
	4. *(Only when the model does not contain a constant term)* Is the model consistent with a zero mean for the errors?	4. The residutals plot shows evidence of a nonzero mean.	4. Add a constant term to the model.
Residuals versus observation order	1. Is there any evidence of residual autocorrelation?	1. The plot shows long runs of residuals with the same sign (positive autocorrelation) or a zigzag pattern (negative autocorrelation).	1. Consider adding terms to the model.
	2. Are there any outliers?	2. Individual points are far removed from the overall plot.	2. Adjust outliers.
	3. Does the series show increasing or decreasing variance over time (changes in volatility)?	3. The plot forms an increasing or decreasing funnel shape.	3. Use a transformation to stabilize the variance.
Autocorrelation function (*ACF*)	1. Are additional differences required?	1. The *ACF* shows a slow linear decay.	1. Try further differencing.
	2. Should more moving average terms be added?	2. The *ACF* shows significant spikes.	2. Include further MA terms.
	3. Is there too much differencing?	3. The lag 1 autocorrelation is large and negative.	3. Reduce the amount of differencing.
Partial autocorrelation function (*PACF*)*	1. Are additional differences required?	1. A "Pollyanna" effect exists.	1. Try further differencing.
	2. Should more AR terms be added?	2. The *PACF* shows significant spikes.	2. Include further AR terms.

*The *PACF* will typically add little to the other diagnostics.

The plot of the *ACF* in Figure 6.18 indicate that the addition of further AR or MA terms is not warranted since there is little pattern to be seen. (This is confirmed by the *PACF* plot which we omit.) Figure 6.17 indicates a clear outlier in the differenced series (October sales minus September sales in 2001), an issue to which we return in Section 6.6. An outlier will often be signaled in the normal probability plot (NPP), as well as in the plots of the residuals. The remaining feature is somewhat unusual: The model under consideration does not include a constant term, and a careful viewing of the two plots of residuals reveals that most of the points lie above the horizontal axis, indicating the need for a constant in the model, as we saw from the results in Table 6.7.

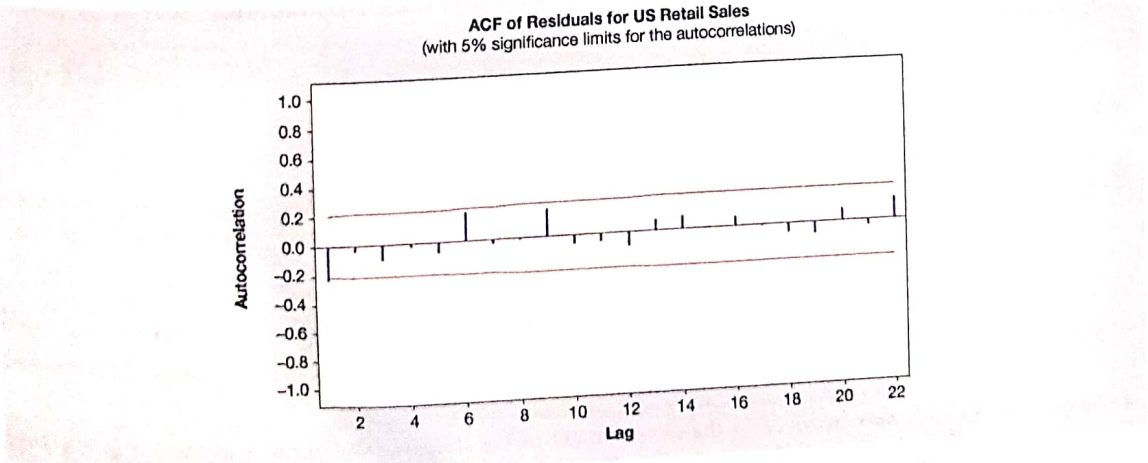
When we include the constant and rerun the model, we obtain the residual plots shown in Figure 6.19. The residuals are now centered about a zero mean, but the other diagnostics are essentially the same as before. In particular, the NPP and the residual plots still signal the large outlier. As we saw in Table 6.7, adding the constant produces a sizeable reduction in the MSE.

Figure 6.17 Residual Plots for ARIMA(0,1,1) Model for U.S. Retail Sales



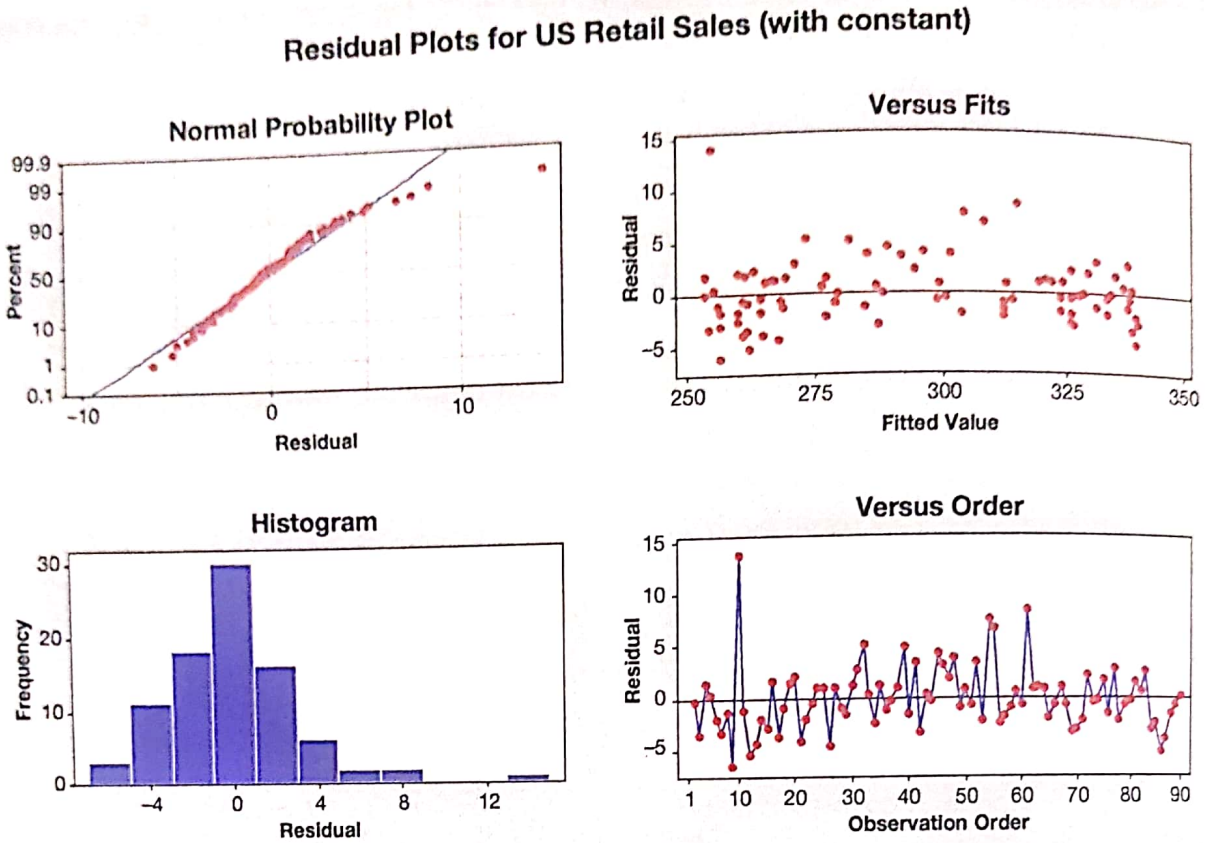
Data: US retail sales_2_SA.xlsx

Figure 6.18 *ACF* for Residuals from the ARIMA(0,1,1) Model for U.S. Retail Sales



Data: US retail sales_2_SA.xlsx

Figure 6.19 Residual Plots for ARIMA(0,1,1)+C Model for U.S. Retail Sales



Data: US retail sales_2_SA.xlsx

7.1 A soft-drink company monitored its television advertising over an eight-week period to evaluate the effect on sales (in millions of dollars) with respect to the number of 30-second spots aired in that week. Estimate the regression equation. The data are as follows:

Week	1	2	3	4	5	6	7	8
No. of spots	8	12	16	10	8	12	16	10
Sales	25	34	39	32	22	30	43	31

Data: Exercise_7_1.xlsx

a. Estimate the regression line for sales on spots.

b. Test whether the slope is significantly different from zero, using $\alpha = 0.10$.

c. Compute S and R^2 and interpret the results.

d. The company has reserved 20 spots for week 9. Forecast the sales and construct a 90 percent prediction interval.

e. Comment on the level of accuracy this analysis provides.

8.1 Exercise 7.1 (in *Exercise_7.1.xlsx*) provided data on television advertising and sales for a soft-drink company over an eight-week period. The purpose was to evaluate the effect on sales (in millions of dollars) with respect to the number of 30-second spots aired during that week. In further analysis, additional price data were collected (also in *Exercise_7_1.xlsx*). Conduct a multiple regression analysis for sales on spot and price.

a. Carry out tests on the overall model and on the individual coefficients. Summarize your conclusions.

b. Compare the performance of the two models. Which model would you recommend?

8.3 Using the unleaded gas prices data (*Gas_prices_1.xlsx*) compare the results of using Personal Disposable Income (rather than Real Retail sales) in a model of *Unleaded price*. Comment on any differences in interpretation and the one-step ahead forecasts from the two alternative models for 2009. Which one would you prefer to have used for 2008?