CHAPTER 1

WHY SHOULD WE CARE ABOUT NONLINEARITY?

Linear processes and linear models dominate research and applications of time series analysis. They are often adequate in making statistical inference in practice. Why should we care about nonlinearity then? This is the first question that came to our minds when we thought about writing this book. After all, linear models are easier to use and can provide good approximations in many applications. Empirical time series, on the other hand, are likely to be nonlinear. As such, nonlinear models can certainly make significant contributions, at least in some applications. The goal of this book is to introduce some nonlinear time series models, to discuss situations under which nonlinear models can make contributions, to demonstrate the value and power of nonlinear time series analysis, and to explore the nonlinear world. In many applications, the observed time series are indirect (possibly multidimensional) observations of an unobservable underlying dynamic process that is nonlinear. In this book we also discuss approaches of using nonlinear and non-Gaussian state space models for analyzing such data.

To achieve our objectives, we focus on certain classes of nonlinear time series models that, in our view, are widely applicable and easy to understand. It is not our intention to cover all nonlinear models available in the literature. Readers are referred to Tong (1990), Fan and Yao (2003), Douc et al. (2014), and De Gooijer (2017) for other nonlinear time series models. The book, thus, shows our preference in exploring the nonlinear world. Efforts are made throughout the book to keep applications in mind so that real examples are used whenever possible. We also provide the theory and justifications for the methods and models considered in the book so that readers can have a comprehensive treatment of nonlinear time series analysis. As always, we start with simple models and gradually move toward more complicated ones.

1.1 SOME BASIC CONCEPTS

A scalar process x_t is a discrete-time time series if x_t is a random variable and the time index t is countable. Typically, we assume the time index t is equally spaced and denote the series by $\{x_t\}$. In applications, we consider mainly the case of x_t with $t \geq 1$. An observed series (also denoted by x_t for simplicity) is a realization of the underlying stochastic process.

A time series x_t is *strictly stationary* if its distribution is time invariant. Mathematically speaking, x_t is strictly stationary if for any arbitrary time indices $\{t_1, \ldots, t_m\}$, where $m > 0$, and any fixed integer k such that the joint distribution function of $(x_{t_1}, \ldots, x_{t_m})$ is the same as that of $(x_{t_1+k}, \ldots, x_{t_m+k})$. In other words, the shift of k time units does not affect the joint distribution of the series. A time series x_t is *weakly stationary* if the first two moments of x_t exist and are time invariant. In statistical terms, this means $E(x_t) = \mu$ and $\text{Cov}(x_t, x_{t+\ell}) = \gamma_\ell$, where E is the expectation, Cov denotes covariance, μ is a constant, and γ_ℓ is a function of ℓ. Here both μ and γ_ℓ are independent of the time index t, and γ_ℓ is called the lag-ℓ autocovariance function of x_t. A sequence of independent and identically distributed (iid) random variates is strictly stationary. A martingale difference sequence x_t satisfying $E(x_t \mid x_{t-1}, x_{t-2}, \ldots) = 0$ and $\text{Var}(x_t \mid x_{t-1}, x_{t-2}, \ldots) = \sigma^2 > 0$ is weakly stationary. A weakly stationary sequence is also referred to as a covariance-stationary time series. An iid sequence of Cauchy random variables is strictly stationary, but not weakly stationary, because there exist no moments. Let $x_t = \sigma_t \epsilon_t$, where $\epsilon_t \sim_{iid} N(0, 1)$ and $\sigma_t^2 = 0.1 + 0.2x_{t-1}^2$. Then x_t is weakly stationary, but not strictly stationary.

Time series analysis is used to explore the dynamic dependence of the series. For a weakly stationary series x_t, a widely used measure of serial dependence between x_t and $x_{t-\ell}$ is the lag-ℓ *autocorrelation function* (ACF) defined by

$$\rho_\ell = \frac{\text{Cov}(x_t, x_{t-\ell})}{\text{Var}(x_t)} \equiv \frac{\gamma_\ell}{\gamma_0}, \tag{1.1}$$

where ℓ is an integer. It is easily seen that $\rho_0 = 1$ and $\rho_\ell = \rho_{-\ell}$ so that we focus on ρ_ℓ for $\ell > 0$. The ACF defined in Equation (1.1) is based on the Pearson's correlation coefficient. In some applications we may employ the autocorrelation function using the concept of Spearman's rank correlation coefficient.

1.2 LINEAR TIME SERIES

A scalar process x_t is a linear time series if it can be written as

$$x_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i a_{t-i}, \tag{1.2}$$

where μ and ψ_i are real numbers with $\psi_0 = 1$, $\sum_{i=-\infty}^{\infty} |\psi_i| < \infty$, and $\{a_t\}$ is a sequence of iid random variables with mean zero and a well-defined density function. In practice, we focus on the one-sided linear time series

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \tag{1.3}$$

where $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$. The linear time series in Equation (1.3) is called a *causal* time series. In Equation (1.2), if $\psi_j \neq 0$ for some $j < 0$, then x_t becomes a non-causal time series. The linear time series in Equation (1.3) is weakly stationary if we further assume that $\text{Var}(a_t) = \sigma_a^2 < \infty$. In this case, we have $E(x_t) = \mu$, $\text{Var}(x_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2$, and $\gamma_\ell = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+\ell}$.

The well-known autoregressive moving-average (ARMA) models of Box and Jenkins (see Box et al., 2015) are (causal) linear time series. Any deviation from the linear process in Equation (1.3) results in a nonlinear time series. Therefore, the nonlinear world is huge and certain restrictions are needed in our exploration. Imposing different restrictions leads to different approaches in tackling the nonlinear world which, in turn, results in emphasizing different classes of nonlinear models. This book is no exception. We start with some real examples that exhibit clearly some nonlinear characteristics and employ simple nonlinear models to illustrate the advantages of studying nonlinearity.

1.3 EXAMPLES OF NONLINEAR TIME SERIES

To motivate, we analyze some real-world time series for which nonlinear models can make a contribution.

4   1. WHY SHOULD WE CARE ABOUT NONLINEARITY?

Example 1.1   Consider the US quarterly civilian unemployment rates from 1948.I to 2015.II for 270 observations. The quarterly rate is obtained by averaging the monthly rates, which were obtained from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of St. Louis and were seasonally adjusted. Figure 1.1 shows the time plot of the quarterly unemployment rates. From the plot, it is seen that (a) the unemployment rate seems to be increasing over time, (b) the unemployment rate exhibits a cyclical pattern reflecting the business cycles of the US economy, and (c) more importantly, the rate rises quickly and decays slowly over a business cycle. As usual in time series analysis, the increasing trend can be handled by differencing. Let r_t be the quarterly unemployment rate and $x_t = r_t - r_{t-1}$ be the change series of r_t. Figure 1.2 shows the time plot of x_t. As expected, the mean of x_t appears to be stable over time. However, the asymmetric pattern in rise and decay of the unemployment rates in a business cycle shows that the rate is not time-reversible, which in turn suggests that the unemployment rates are nonlinear. Indeed, several nonlinear tests discussed later confirm that x_t is indeed nonlinear.

If a linear autoregressive (AR) model is used, the Akaike information criterion (AIC) of Akaike (1974) selects an AR(12) model for x_t. Several coefficients of the
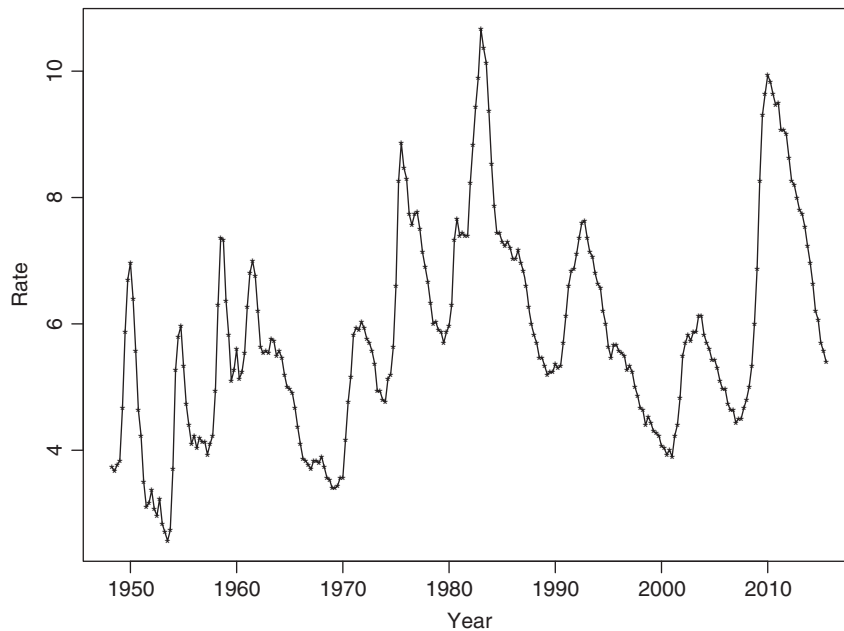
Figure 1.1   Time plot of US quarterly civilian unemployment rates, seasonally adjusted, from 1948.I to 2015.II.
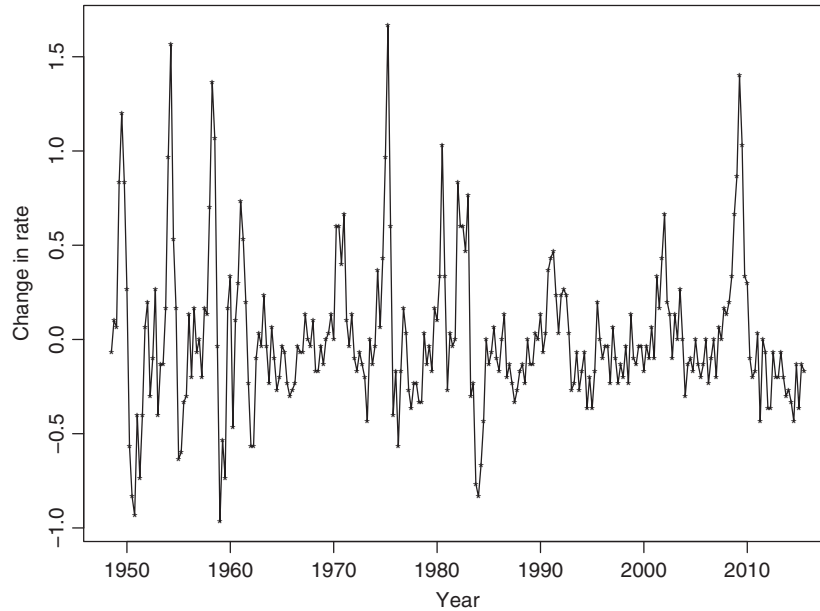
Figure 1.2 Time plot of the changes in US quarterly civilian unemployment rates, seasonally adjusted, from 1948.I to 2015.II.

fitted AR(12) model are not statistically significant so that the model is refined. This leads to a simplified AR(12) model as
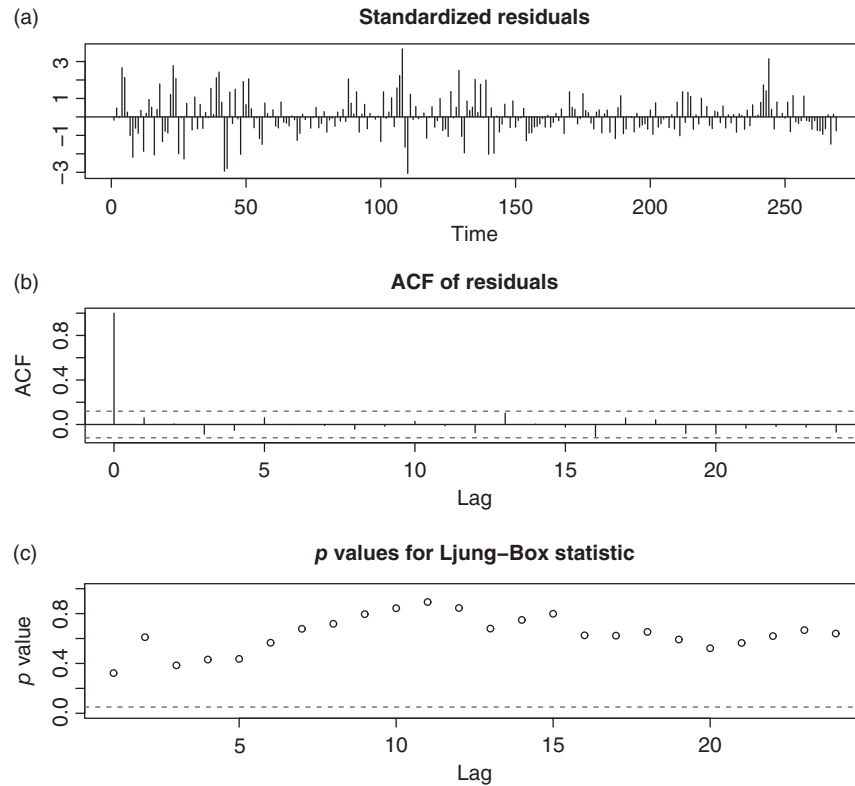
$$x_t = 0.68x_{t-1} - 0.26x_{t-4} + 0.18x_{t-6} - 0.33x_{t-8} + 0.19x_{t-9} - 0.17x_{t-12} + a_t, \quad (1.4)$$

where the variance of a_t is $\sigma_a^2 = 0.073$ and all coefficient estimates are statistically significant at the usual 5% level. Figure 1.3 shows the results of model checking, which consists of the time plot of standardized residuals, sample autocorrelation function (ACF), and the p values of the Ljung–Box statistics $Q(m)$ of the residuals. These p values do not adjust the degrees of freedom for the fitted parameters, but they are sufficient in indicating that the fitted AR(12) model in Equation (1.4) is adequate.

On the other hand, one can use the self-exciting threshold autoregressive (TAR) models of Tong (1978, 1990) and Tong and Lim (1980) to describe the nonlinear characteristics of the data. The TAR model is one of the commonly used nonlinear time series models to be discussed in Chapter 2. Using the TSA package of R (R Development Core Team, 2015), we obtain the following two-regime TAR model

$$x_t = 0.47x_{t-1} + 0.15x_{t-2} - 0.02x_{t-3} - 0.17x_{t-4} + a_{1t}, \quad \text{if } x_{t-1} \leq \delta \qquad (1.5)$$
$$= 0.85x_{t-1} - 0.08x_{t-2} - 0.22x_{t-3} - 0.29x_{t-4} + 0.23x_{t-5} + 0.36x_{t-6} - 0.14x_{t-7}$$
$$- 0.51x_{t-8} + 0.37x_{t-9} + 0.17x_{t-10} - 0.23x_{t-11} - 0.21x_{t-12} + a_{2t}, \text{ if } x_{t-1} > \delta,$$

(a)

Standardized residuals

(b)

ACF of residuals

(c)

***p* values for Ljung–Box statistic**

Figure 1.3 Model checking of the fitted AR(12) model of Equation (1.4) for the change series of US quarterly unemployment rates from 1948.II to 2015.II: (a) standardized residuals, (b) ACF of residuals, and (c) *p* values for Ljung–Box statistics.

where the threshold $\delta = -0.066666$ and the standard errors of a_{1t} and a_{2t} are 0.181 and 0.309, respectively. Some of the coefficient estimates are not statistically significant at the 5% level, but, for simplicity, we do not seek any further refinement of the model. Model checking shows that the fitted TAR model is adequate. Figure 1.4 shows the time plot of standardized residuals of model (1.5) and the sample autocorrelations of the standardized residuals. The Ljung–Box statistics of the standardized residuals show Q(12) = 6.91 (0.86) and Q(24) = 18.28(0.79), where the number in parentheses denotes the asymptotic *p* value. In this particular instance, x_{t-1} is the threshold variable and the fitted model shows that when the quarterly unemployment rate decreased by −0.07% or more, the dynamic dependence of the unemployment rates appears to be simpler than when the rate was increasing or changed mildly. In other words, the model implies that the dynamic dependence
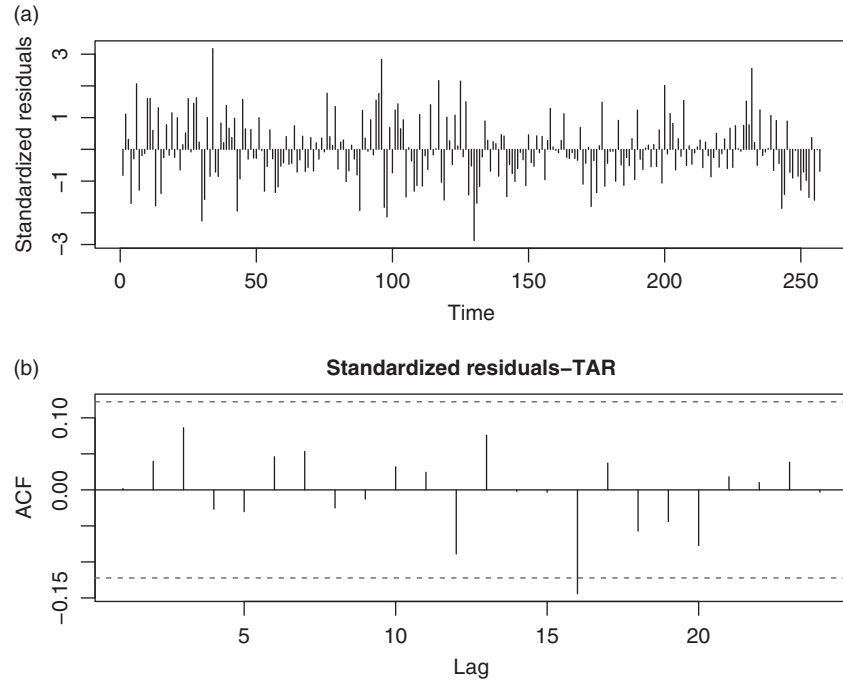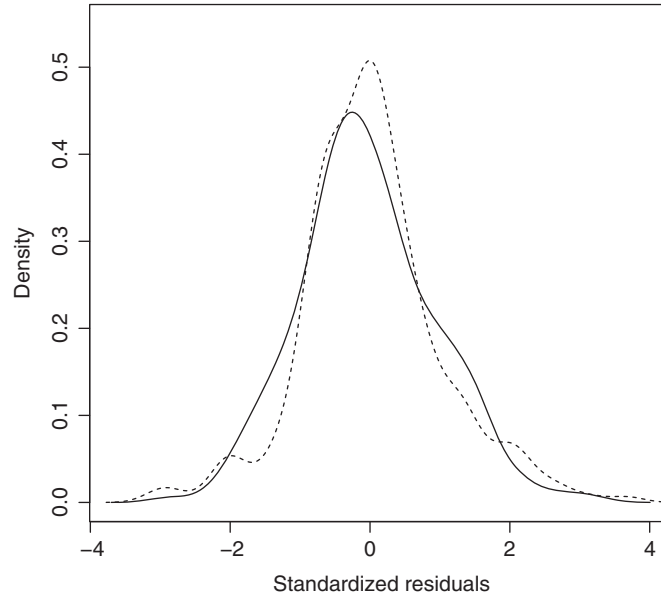
Figure 1.4   Model checking of the fitted TAR(4,12) model of Equation (1.5) for the change series of US quarterly unemployment rates from 1948.II to 2015.II. (a) The time plot of standardized residuals of model (1.5) and (b) the sample autocorrelations of the standardized residuals.

of the US quarterly unemployment rates depends on the status of the US economy. When the economy is improving, i.e. the unemployment rate decreased substantially, the unemployment rate dynamic dependence became relatively simple.

To compare the AR and TAR models in Equations (1.4) and (1.5) for the unemployment rate series, we consider model goodness-of-fit and out-of-sample predictions. For goodness of fit, Figure 1.5 shows the density functions of the standardized residuals of both fitted models. The solid line is for the TAR model whereas the dashed line is for the linear AR(12) model. From the density functions, it is seen that the standardized residuals of the TAR model are closer to the normality assumption. Specifically, the residuals of the TAR model are less skewed and have lower excess kurtosis. In this particular instance, the skewness and excess kurtosis of the standardized residuals of the linear AR(12) model are 0.318 and 1.323, respectively, whereas those of the TAR model are 0.271 and 0.256, respectively. Under the assumption that the standardized residuals are

Figure 1.5   Density functions of standardized residuals of the linear AR(12) model in Equation (1.4) and the TAR(4,12) model of Equation (1.5) for the change series of US quarterly unemployment rates from 1948.II to 2015.II. The solid line is for the TAR model.

independent and identically distributed, the t ratios for excess kurtosis are 4.33 and 0.84, respectively, for the fitted AR(12) and the TAR model. Similarly, the t ratios for the skewness are 2.08 and 1.77, respectively, for AR(12) and the TAR model. If the 5% critical value of 1.96 is used, then one cannot reject the hypotheses that the standardized residuals of the fitted TAR model in Equation (1.5) are symmetric and do not have heavy tails. On the other hand, the standardized residuals of the linear AR(12) model are skewed and have heavy tails.

We also use rolling one-step ahead out-of-sample forecasts, i.e. back-testing, to compare the two fitted models. The starting forecast origin is $t = 200$ so that we have 69 one-step ahead predictions for both models. The forecasting procedure is as follows. Let n be the starting forecast origin. For a given model, we fit the model using the data from $t = 1$ to $t = n$; use the fitted model to produce a prediction for $t = n + 1$ and compute the forecasting error. We then advance the forecast origin by 1 and repeat the estimation–prediction process. We use root mean squared error (RMSE), mean absolute error (MAE) and (average) bias of predictions to quantify the performance of back-testing. In addition, we also classify the predictions based on the regime of the forecast origin. The results are given in Table 1.1. From the table, it is seen that while the nonlinear TAR model shows some improvement in

Criterion	Overall	Origins in Regime 1	Origins in Regime 2
(a) Linear AR(12) model			
RMSE	0.2166	0.1761	0.2536
MAE	0.1633	0.1373	0.1916
Bias	0.0158	0.0088	0.0235
(b) Threshold AR model			
RMSE	0.2147	0.1660	0.2576
MAE	0.1641	0.1397	0.1906
Bias	−0.0060	−0.1003	0.0968

Table 1.1   Performance of out-of-sample forecasts of the linear AR(12) model in Equation (1.4) and the threshold AR model in Equation (1.5) for the US quarterly unemployment rates from 1948.II to 2015.II. The starting forecast origin is 200, and there are 69 one-step ahead predictions.

out-of-sample predictions, the improvement is rather minor and seems to come from those associated with forecast origins in Regime 1.

In this example, a simple nonlinear model can help in prediction and, more importantly, the nonlinear model improves the model goodness of fit as shown by the properties of the residuals.

Example 1.2   As a second example, we consider the weekly crude oil prices from May 12, 2000 to August 28, 2015. The data used are part of the commodity prices available from Federal Reserve Bank of St. Louis and they are the crude oil prices, West Texas Intermediate, Cushing, Oklahoma. Figure 1.6 shows the time plot of the original crude oil prices and the first differenced series of the prices. The differenced series is used in our analysis as the oil prices exhibit strong serial dependence, i.e. unit-root nonstationarity. From the plots, it is seen that the variability of the time series varies over time. Let $x_t = p_t - p_{t-1}$, where p_t denotes the observed weekly crude oil price at week t. If scalar AR models are employed, an AR(8) model is selected by AIC and the fitted model, after removing insignificant parameters, is

$$x_t = 0.197x_{t-1} + 0.153x_{t-8} + a_t, \quad \sigma_a^2 = 6.43, \qquad (1.6)$$

where the standard errors of both AR coefficients are around 0.034. This simple AR model adequately describes the dynamic correlations of x_t, but it fails to address the time-varying variability. Figure 1.7 shows the sample ACF of the
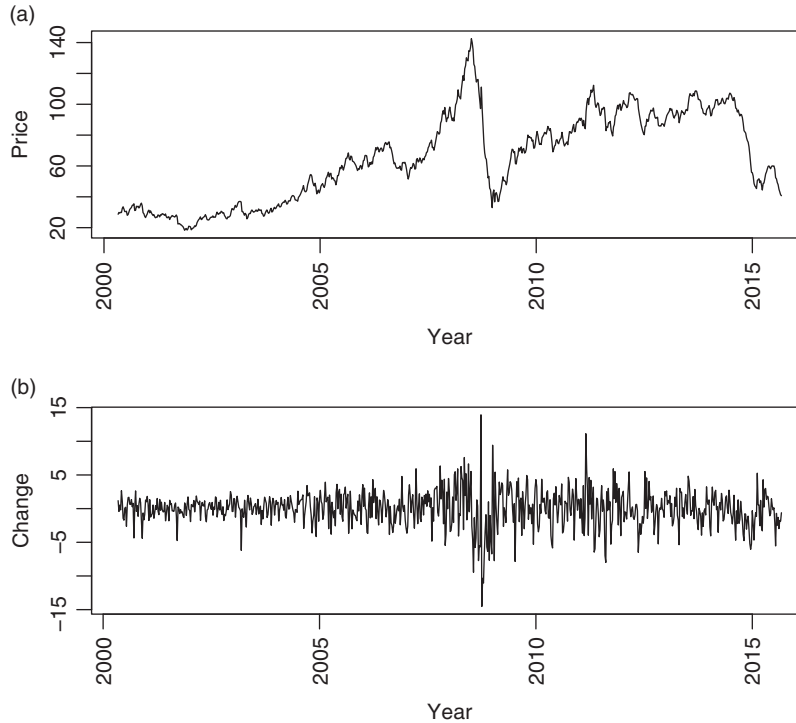
Figure 1.6   Time plot of weekly crude oil prices from May 12, 2000 to August 28, 2015: (a) oil prices and (b) the first differenced series. The prices are West Texas Intermediate, Cushing, Oklahoma, and obtained from FRED of Federal Reserve Bank, St. Louis
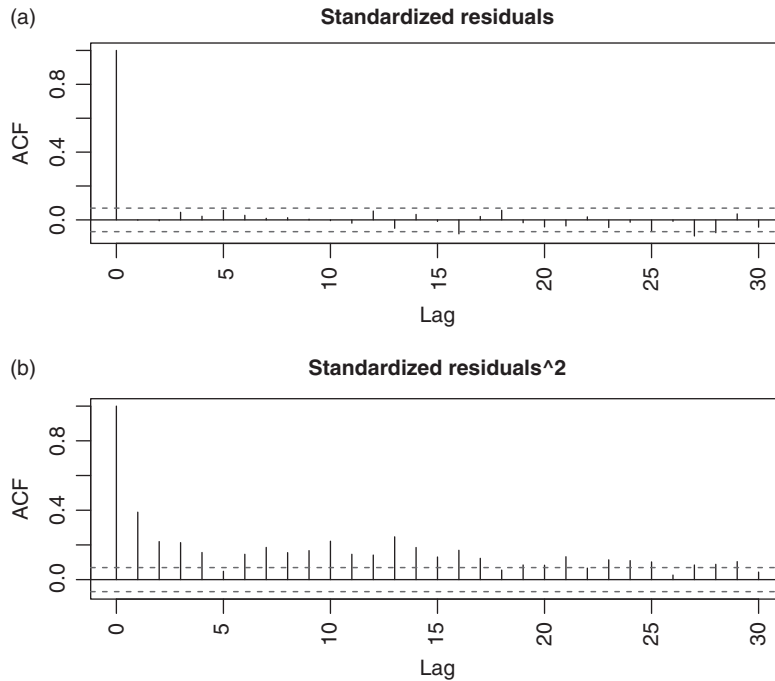
residuals and those of the squared residuals. From the plots, the residuals have no significant serial correlations, but the squared residuals have strong serial dependence. In finance, such a phenomenon is referred to as time-varying volatility or conditional heteroscedasticity.

One approach to improve the linear AR(8) model in Equation (1.6) is to use the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986). GARCH models are nonlinear based on the linearity definition of Equation (1.2). In this particular instance, the fitted AR-GARCH model with Student-t innovations is

$$x_t \approx 0.227x_{t-1} + 0.117x_{t-8} + a_t, \tag{1.7}$$

$$a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} t^*_{7.91},$$

$$\sigma_t^2 = 0.031 + 0.072a_{t-1}^2 + 0.927\sigma_{t-1}^2, \tag{1.8}$$

(a)

Standardized residuals


(b)

Standardized residuals^2

Figure 1.7 Sample autocorrelation functions of (a) the residuals of the AR(8) model in Equation (1.6) for the first differenced series of weekly crude oil prices and (b) the squared residuals.

where t_v^* denotes standardized Student-t distribution with v degrees of freedom. In Equation (1.7) we use the approximation $\approx$, since we omitted the insignificant parameters at the 5% level for simplicity. In Equation (1.8), the standard errors of the parameters are 0.024, 0.016, and 0.016, respectively. This volatility model indicates the high persistence in the volatility. For further details about volatility models, see Chapter 5 and Tsay (2010, Chapter 3). Model checking indicates that the fitted AR-GARCH model is adequate. For instance, we have Q(20) = 14.38 (0.81) and 9.89(0.97) for the standardized residuals and the squared standardized residuals, respectively, of the model, where the number in parentheses denotes the p value. Figure 1.8(a) shows the time plot of x_t along with point-wise two-standard errors limits, which further confirms that the model fits the data well. Figure 1.8(b) shows the quantile-to-quantile (QQ) plot of the standardized residuals versus the Student-t distribution with 7.91 degrees of freedom. The plot shows that it is reasonable to employ the Student-t innovations.

Compared with the linear AR(8) model in Equation (1.6), the fitted AR-GARCH model does not alter the serial dependence in the x_t series

(a)

(b)
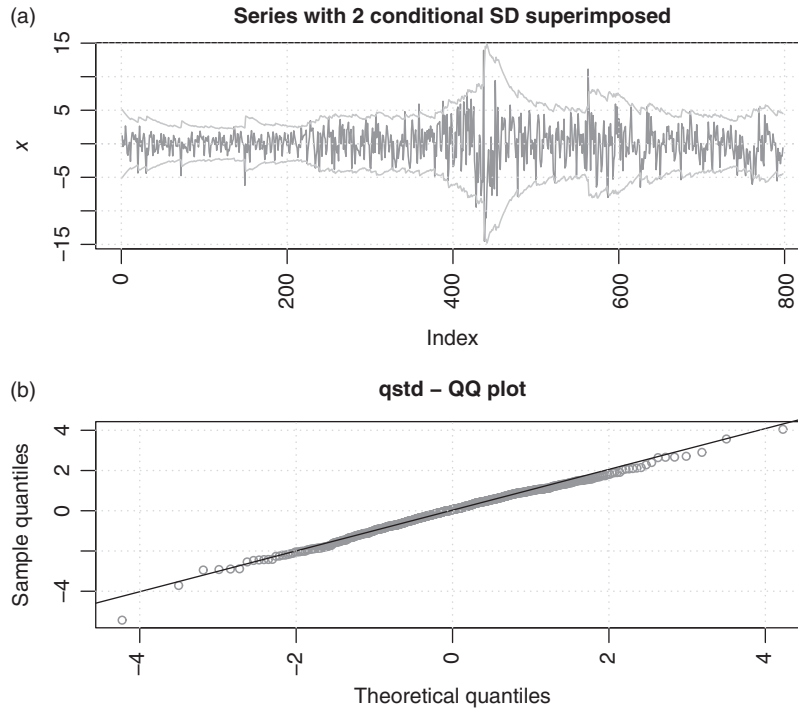
Figure 1.8   (a) The time plot of the first-differenced crude oil prices and point-wise two standard error limits of the model in Equations (1.7) and (1.8). (b) The quantile-to-quantile plot for the Student-t innovations.

because Equation (1.7) is relatively close to Equation (1.6). What the GARCH model does is to handle the time-varying volatility so that proper inference, such as interval predictions, can be made concerning the crude oil prices. In some finance applications volatility plays a key role and the fitted AR-GARCH nonlinear model can be used.

Alternatively, one can use the Markov switching model (MSM) to improve the linear AR(8) model. Details of the MSM are given in Chapters 2 and 7. In this particular application, the model used is

$$
x_t =
\begin{cases}
\phi_{11}x_{t-1} + \phi_{12}x_{t-2} + \cdots + \phi_{18}x_{t-8} + \sigma_1 \epsilon_t, & \text{if } s_t = 1, \\
\phi_{21}x_{t-1} + \phi_{22}x_{t-2} + \cdots + \phi_{28}x_{t-8} + \sigma_2 \epsilon_t, & \text{if } s_t = 2,
\end{cases}
\tag{1.9}
$$

where $\phi_{i,j}$ denotes the coefficient of state i at lag-j, ϵ_t are independent and identically distributed random variables with mean zero and variance 1, σ_i is the innovation standard error of state i, and s_t denotes the status of the state at time t. This is a

Par.	ϕ_{i1}	ϕ_{i2}	ϕ_{i3}	ϕ_{i4}	ϕ_{i5}	ϕ_{i6}	ϕ_{i7}	ϕ_{i8}	σ_i
				State 1					
Est.	−0.60	0.53	−0.43	−0.04	−0.08	−0.14	−0.08	0.12	3.04
s.e.	0.27	0.20	0.24	0.11	0.16	0.13	0.16	0.13	
				State 2					
Est.	0.37	−0.12	0.12	0.03	0.08	0.09	0.05	0.15	1.89
s.e.	0.05	0.04	0.05	0.04	0.04	0.05	0.04	0.04	

Table 1.2   Parameter estimates of the Markov switching model for the first-differenced series of weekly crude oil prices from May 19, 2000 to August 28, 2015. Par., parameter; Est., estimate; s.e., standard error.

simple two-state MSM and the states change over time according to the transition probability matrix

$$P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix},$$

where p_{ij} denotes the probability of switching from state i at time $t-1$ to state j at time t. This notation is used in the MSwM package of R and in Hamilton (1994, Chapter 22). From the definition, the rows of P sum to 1 and p_{ii} denotes the probability of staying in state i from time $t-1$ to time t. Consequently, $1/(1-p_{ii})$ denotes the expected duration for state i.
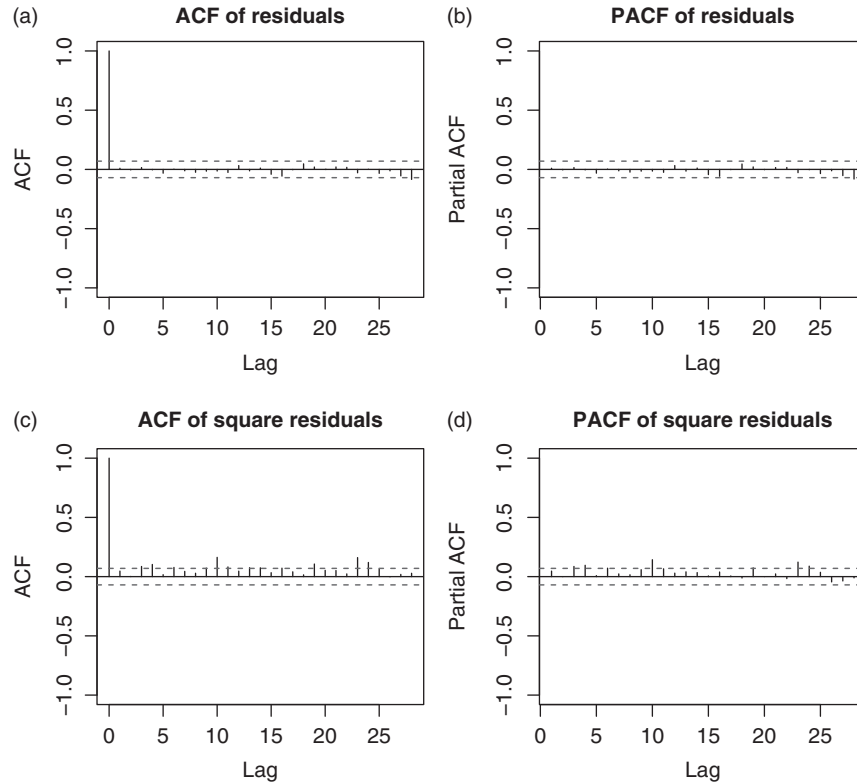
For the first differenced series x_t of the crude oil prices, parameter estimates of the MSM are given in Table 1.2. The fitted transition matrix is

$$\hat{P} = \begin{bmatrix} 0.066 & 0.934 \\ 0.268 & 0.732 \end{bmatrix}.$$

Figure 1.9 shows the sample ACF and PACF of the residuals (a and b) and the squared residuals (c and d) of each state. From the plots, the serial correlations of x_t are well described by the fitted MSM in Equation (1.9). More interestingly, the ACF of squared residuals indicates that the MSM is capable of modeling the time-varying volatility. Figure 1.10 shows the time plots of smoothed and filtered probabilities for each state.

From Table 1.2, it is seen that state $s_t = 1$ corresponds to the volatile state and its coefficient estimates are more uncertain with larger standard errors. This is understandable since if x_{t-1} is in state 1, x_t stays in state 1 with a small probability

Figure 1.9   The sample autocorrelation functions of the residuals (a and b) and squared residuals (c and d), by state, for the Markov switching model of Equation (1.9) fitted to the first-differenced series of weekly crude oil prices from May 19, 2000 to August 28, 2015.

0.066 and it switches to state 2 about 93% of the time. The marginal probability that x_t is in state 1 is 0.223. The table also shows that the serial dependence of x_t depends on the state; the serial dependence can be modeled by an AR(3) model in state 1, but it requires an AR(8) model in state 2. Figure 1.10 confirms that x_t stays in state 2 often, implying that during the data span the crude oil prices encountered some volatile periods, but the high volatility periods are short-lived.

In this example, we show that nonlinear models can provide a deeper understanding of the dynamic dependence of a time series. In addition, various nonlinear models can be used to improve the fitting of a linear model. The Markov switching model allows the dynamic dependence of a time series to change according to the state it belongs to. It is also capable of handling time-varying volatility.

(a)

Filtered probability regime 1

(b)

Smoothed probability regime 1
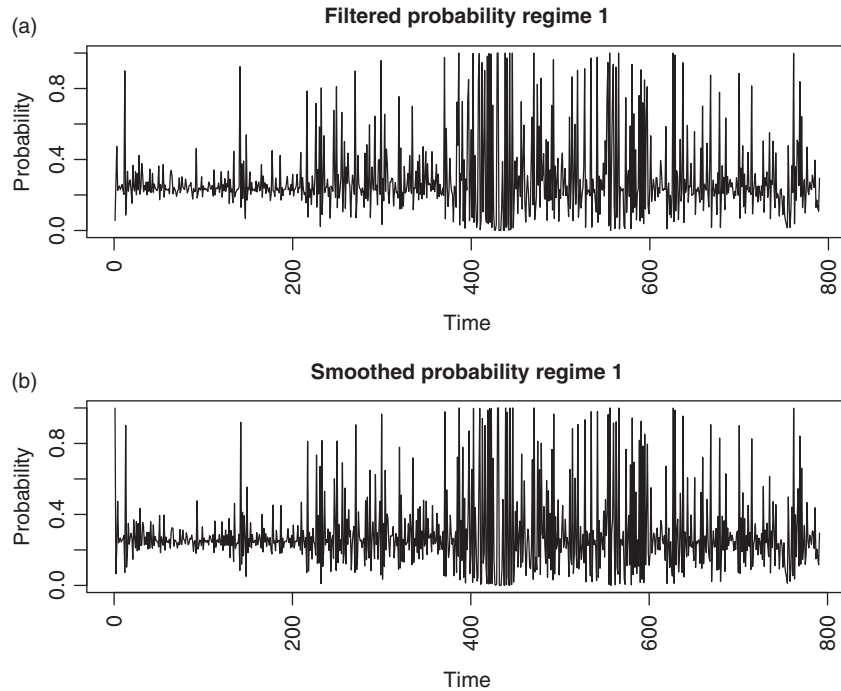
Figure 1.10 (a) Filtered and (b) smoothed probabilities for state 1 of the Markov switching model of Equation (1.9) fitted to the first-differenced series of weekly crude oil prices from May 19, 2000 to August 28, 2015.

Example 1.3 In this example, we revisit the well-known gas furnace series of Box and Jenkins (1976). The data consist of the input gas feed rate, x_t, and corresponding output CO_2 concentration, y_t, from a gas furnace. There are 296 observations taken at time intervals of 9 seconds. The top two plots of Figure 1.11 show the gas rate x_t and the output CO_2 concentration. These two series have been widely used in the literature to demonstrate the analysis of transfer function models. Following Box and Jenkins (1976), one can use linear transfer function models. The model for the gas feed rate is

$$x_t = 1.97x_{t-1} - 1.37x_{t-2} + 0.34x_{t-3} + \epsilon_t, \quad \sigma_\epsilon^2 = 0.035,$$

where the standard errors of the AR coefficients are 0.05, 0.10, and 0.05, respectively. The transfer function model used is

$$y_t = 53.37 - \frac{0.53B^3 + 0.38B^4 + 0.52B^5}{1 - 0.55B}x_t + \frac{1}{1 - 1.53B + 0.63B^2}a_t, \quad (1.10)$$
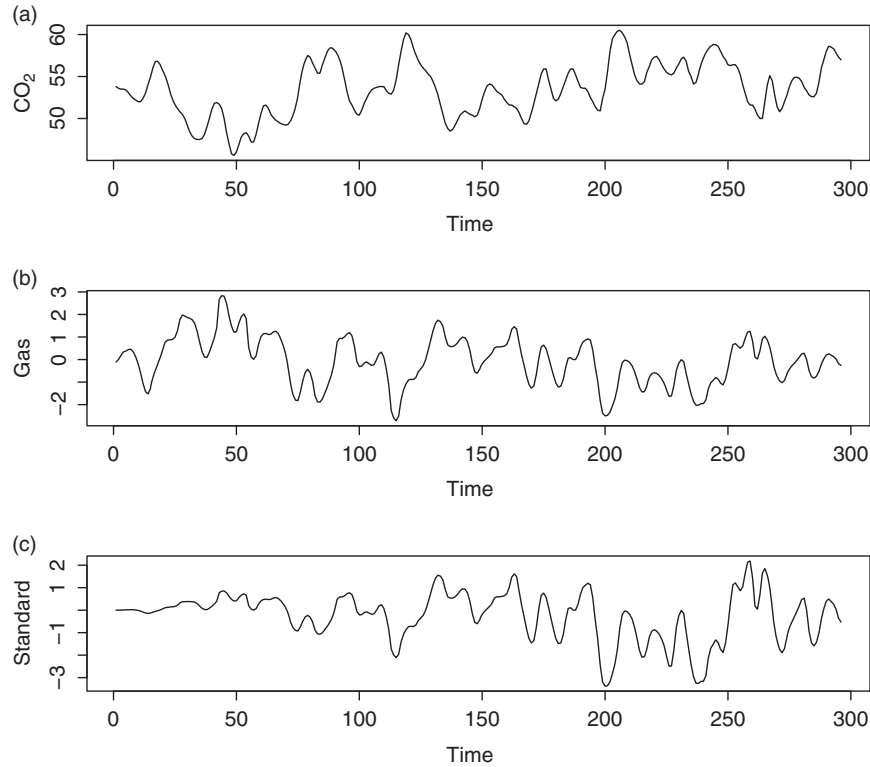
Figure 1.11   Time plots of the gas furnace data: (a) input gas feed rates, (b) output CO_2 concentrations, and (c) gas rates weighted by time trend $s_t = 2t/296$.

where B denotes the back-shift (or lag) operator such that $Bx_t = x_{t-1}$, the variance of a_t is 0.0576, and all parameter estimates are statistically significant at the usual 5% level. Residual analysis shows that the model in Equation (1.10) is adequate. For instance, the Ljung–Box statistics of the residuals gives $Q(12) = 15.51$ with p value 0.21.

On the other hand, nonlinear transfer function models have also been proposed in the literature, e.g. Chen and Tsay (1996). In this example, we follow the approach of Tsay and Wu (2003) by allowing the coefficients of the transfer function model to depend on a state variable s_t. In this particular instance, the state variable used is $s_t = 2t/296$, which simply represents the time sequence of the observations. The corresponding transfer function model would become

$$y_t = c_0 + \frac{\omega_3(s_t)B^3 + \omega_4(s_t)B^4 + \omega_5(s_t)B^5}{1 - \delta(s_t)B}x_t + \frac{1}{1 - \phi_1 B - \phi_2 B^2}a_t,$$

where $\omega_i(s_t)$ and $\delta(s_t)$ are smooth functions of the state variable s_t, c_0 and ϕ_i are constant. This is a special case of the functional-coefficient models of Chen and Tsay (1993). To simplify the model, Tsay and Wu (2003) used the first-order Taylor series approximations of the functional coefficients

$$\omega_i(s_t) \approx \omega_{i0} + \omega_{i1}s_t, \quad \delta(s_t) \approx \delta_0 + \delta_1 s_t,$$

where ω_{ij} and δ_j are constant, and simplified the model further to obtain the transfer function model

$$y_t = 52.65 - \frac{1.22B^3}{1 - 0.61B}x_t + 0.73s_t + \frac{0.99B^3 - 0.99B^4}{1 - 0.65B}(s_t x_t) \quad (1.11)$$
$$+ \frac{1}{1 - 1.42B + 0.49B^2}a_t,$$

where the residual variance is $\sigma_a^2 = 0.0461$ and all coefficient estimates are statistically significant at the 5% level. The Ljung–Box statistics of the results gives $Q(12) = 9.71$ with p value 0.64. This model improves the in-sample fit as the residual variance drops from 0.0576 to 0.0431. Figure 1.11(c) shows the time plot of $s_t x_t$, from which the new variable $s_t x_t$ seems to emphasize the latter part of the x_t series.

Table 1.3 shows the results of out-of-sample forecasts of the two transfer function models in Equations (1.10) and (1.11). These summary statistics are based on 96 one-step ahead forecasts starting with initial forecast origin $t = 200$. As before, the models were re-estimated before prediction once a new observation was available. From the table it is easily seen that the model in Equation (1.11) outperforms the traditional transfer function model in all three measurements of out-of-sample prediction. The improvement in the root mean squared error (RMSE) is $(0.3674 - 0.3311)/0.3311 = 10.96\%$. This is a substantial improvement given that the model in Equation (1.10) has been regarded as a gold standard for the data set.

In this example we show that the functional-coefficient models can be useful in improving the accuracy of forecast. It is true that the model in Equation (1.11)

Model	Bias	RMSE	MAE
Equation (1.10)	0.0772	0.3674	0.2655
Equation (1.11)	0.0300	0.3311	0.2477

Table 1.3 Summary statistics of out-of-sample forecasts for the models in Equations (1.10) and (1.11). The initial forecast origin is 200 and the results are based on 96 one-step ahead predictions.
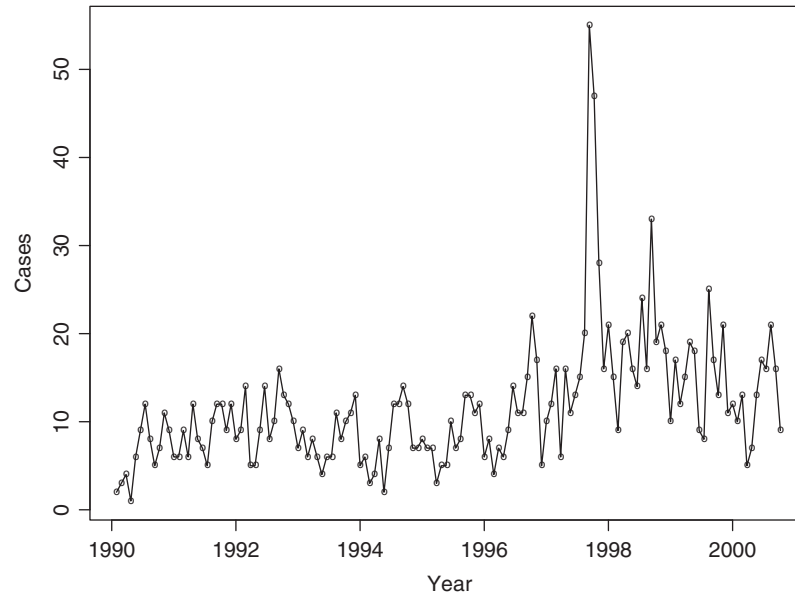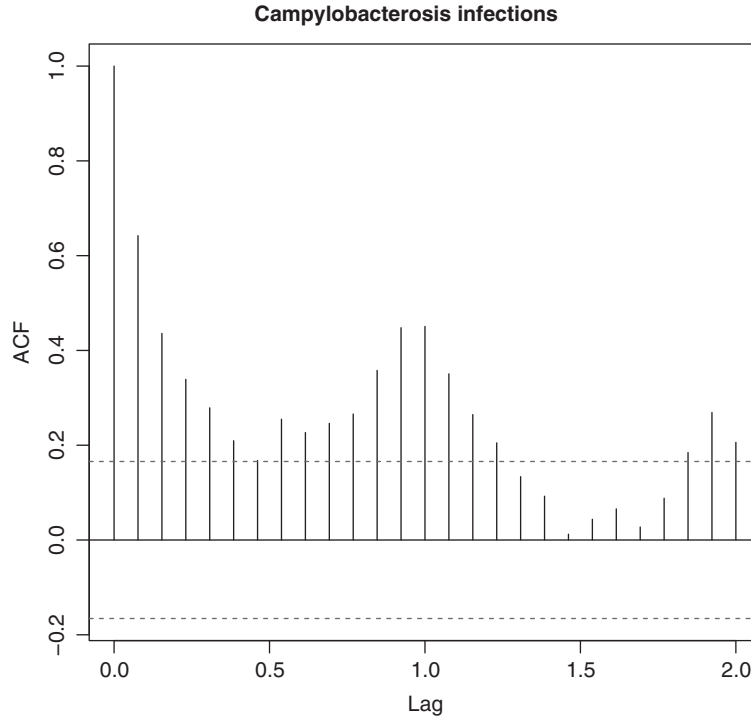
Figure 1.12   Time plot of the number of cases of Campylobacterosis infections in northern Quebec, Canada, in 4-week intervals from January 1990 to October 2000 for 140 observations.

remains linear because the added variable $s_t x_t$ can be treated as a new input variable. However, the model was derived by using the idea of functional-coefficient models, a class of nonlinear models discussed in Chapter 2.

Example 1.4   Another important class of nonlinear models is the generalized linear model. For time series analysis, generalized linear models can be used to analyze count data. Figure 1.12 shows the number of cases of Campylobacterosis infections in the north of the province Quebec, Canada, in 4-week intervals from January 1990 to the end of October 2000. The series has 13 observations per year and 140 observations in total. See Ferland et al. (2006) for more information. The data are available from the R package `tscount` by Liboschik et al. (2015). Campylobacterosis in an acute bacterial infectious disease attacking the digestive system. As expected, the plot of Figure 1.12 shows some seasonal pattern.

This is an example of time series of count data, which occur in many scientific fields, but have received relatively less attention in the time series literature. Similar to the case of independent data, Poisson or negative binomial distribution is often used to model time series of count data. Here one postulates that the time

Campylobacterosis infections

Figure 1.13 Sample autocorrelation function of the cases of Campylobacterosis infections in northern Quebec, Canada, in 4-week intervals from January 1990 to October 2000 for 140 observations.

series of count data is driven by an underlying intensity process λ_t. This latent process governs the time evolution of the conditional expectation of the data, has dynamic dependence, and is related to the data via a link function. In most applications, the link function is either the identify function or the log transformation. Details of modeling time series of count data are given in Chapter 5. Here we use the example to demonstrate yet another application of nonlinear time series analysis. Figure 1.13 shows the sample ACF of the data. Clearly, there is dynamic dependence with seasonality in the data.

Let x_t denote the number of cases of Campylobacterosis infections in week t and $\lambda_t = E(x_t|F_{t-1})$, where F_{t-1} denotes the information available at time $t - 1$. A commonly used distribution for count data is

$$x_t|F_{t-1} \sim \text{Poi}(\lambda_t) \quad \text{or} \quad x_t|F_{t-1} \sim \text{NB}(\lambda_t, \alpha), \tag{1.12}$$

where $\text{Poi}(\lambda_t)$ and $\text{NB}(\lambda_t, \alpha)$ denote, respectively, a Poisson distribution with mean λ_t and a negative binomial distribution with mean λ_t and dispersion parameter $\alpha > 0$. Specifically, for the Poisson distribution we have

$$P(x_t = k|\lambda_t) = \frac{1}{k!}e^{-\lambda_t}\lambda_t^k, \quad k = 0, 1, \dots .$$

For the negative binomial distribution, there are several parameterizations available in the literature. We use the probability mass function

$$P(x_t = k|\lambda_t, \alpha) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k + 1)}\left(\frac{\alpha}{\alpha + k}\right)^\alpha \left(\frac{\lambda_t}{\alpha + \lambda_t}\right)^k, \quad k = 0, 1, \dots .$$

Under this parameterization, we have $E(x_t|F_{t-1}) = \lambda_t$ and $\text{Var}(x_t|F_{t-1}) = \lambda_t + \lambda_t^2/\alpha$. Here $\alpha \in (0, \infty)$ is the dispersion parameter and the negative binomial distribution approaches the Poisson distribution as $\alpha \to \infty$.

For the time series in Figure 1.12, one can introduce the dynamic dependence for the data by using the model

$$\lambda_t = \gamma_0 + \gamma_1 x_{t-1} + \delta_{13}\lambda_{t-13}, \tag{1.13}$$

where λ_{t-13} is used to describe the seasonality of the data. This model is similar to the GARCH model of Bollerslev (1986) for the volatility model and belongs to the class of observation-driven models in generalized linear models. See Ferland et al. (2006), among others. Using Equation (1.13) with negative binomial distribution, the quasi maximum likelihood estimation gives

$$\lambda_t = 2.43 + 0.594x_{t-1} + 0.188\lambda_{t-13},$$

where the dispersion parameter is 0.109 and the standard deviations of the coefficient estimates are 1.085, 0.092, and 0.123, respectively. Figure 1.14 shows various plots of model checking for the fitted model in Equation (1.13). From the time plot of Pearson residuals, it is seen that certain outlying observations exist so that the model can be refined. Comparing the autocorrelations of the data in Figure 1.13 and the residual autocorrelations in Figure 1.14, the simple model in Equation (1.13) does a decent job in describing the dynamic dependence of the count data. Details of model checking and refinement of the generalized linear models for time series of count data will be given in Chapter 5. The analysis can also be handled by the state space model of Chapter 6.

In this example, we demonstrate that the generalized linear models, which are nonlinear, can be used to analyze time series of count data.

1.4 NONLINEARITY TESTS

The flexibility of nonlinear models in data fitting may encounter the problem of finding spurious structure in a given time series. It is, therefore, of importance
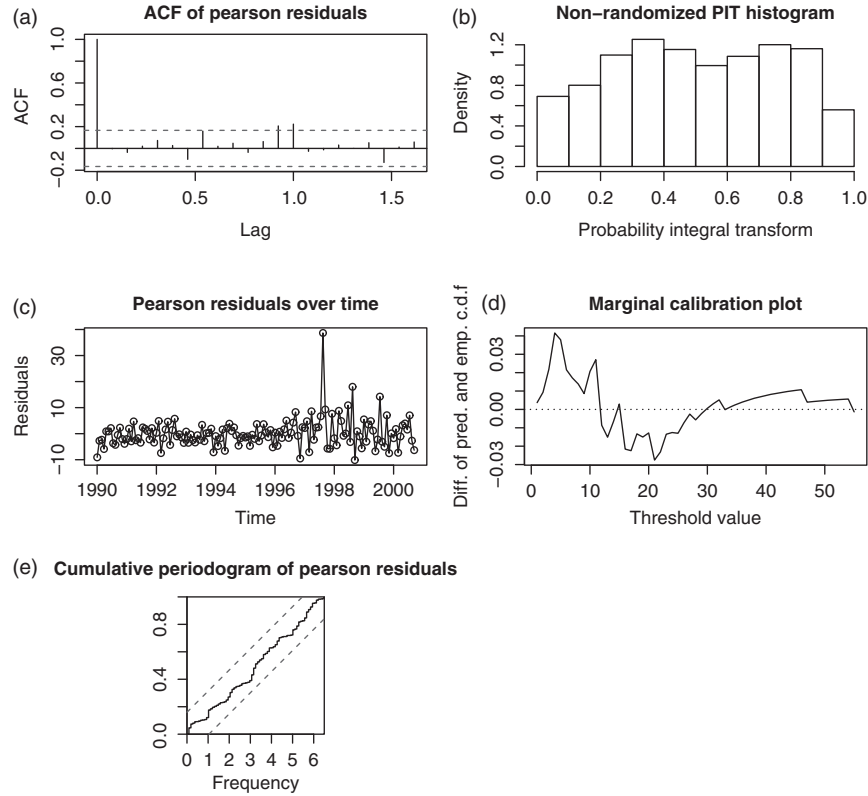
Figure 1.14 Various model diagnostic statistics for the model in Equation (1.13) with negative binomial distribution for the Campylobacterosis infection data. (a) ACF of Pearson residuals, (b) non-randomized PIT histogram, (c) Pearson residuals over time, (d) marginal calibration plot, and (e) cumulative periodogram of Pearson residuals.

to check the need for using nonlinear models. To this end, we introduce some nonlinearity tests for time series data. Both parametric and nonparametric tests are considered.

1.4.1 Nonparametric Tests

Several nonparametric statistics have been proposed in the literature for testing the nonlinearity of a stationary time series. We discuss some of those statistics that are useful and easily available for practical use.

The BDS test: Proposed by Brock et al. (1987) and later published by Brock et al. (1996), this test is widely used to verify the null hypothesis that a given time

series consists of independent and identically distributed (iid) random variables. The test uses the idea of *correlation integral* popular in chaotic time series analysis. Roughly speaking, a correlation integral is a measure of the frequency with which temporal patterns repeat themselves in a data set. Consider a time series $\{x_t | t = 1, \dots, T\}$, where T denotes the sample size. Let m be a given positive integer. Define an m-history of the series as $\boldsymbol{x}_t^m = (x_t, x_{t-1}, \dots, x_{t-m+1})$ for $t = m, \dots, T$. Define the correlation integral at the embedding dimension m as

$$C(m, \epsilon) = \lim_{T_m \to \infty} \frac{2}{T_m(T_m - 1)} \sum \sum_{m \le s < t \le T} I(\boldsymbol{x}_t^m, \boldsymbol{x}_s^m | \epsilon) \qquad (1.14)$$

where $T_m = T - m + 1$ is the number of constructed m-histories, ϵ is a given positive real number, and $I(\boldsymbol{u}, \boldsymbol{v} | \epsilon)$ is an indicator variable that equals to one if $\| \boldsymbol{u} - \boldsymbol{v} \| < \epsilon$ and zero otherwise, where $\| \cdot \|$ denotes the sup-norm of two vectors. For the m-histories, we have $I(\boldsymbol{u}, \boldsymbol{v} | \epsilon) = 1$ if $|u_i - v_i| < \epsilon$ for $i = 1, \dots, m$ and $= 0$ otherwise. For a given ϵ, Equation (1.14) simply measures the probability of m-histories being within a distance ϵ of each other.

For testing purpose, the magnitude of the correlation integral in Equation (1.14) needs to be judged. To this end, the BDS test compares $C(m, \epsilon)$ with $C(1, \epsilon)$ under the null hypothesis. Intuitively, if $\{x_t\}$ are iid, then there exist no patterns in the data so that a probability of the m-history is simply the mth power of the corresponding probability of the 1-history. This is so because under independence $Pr(A \cap B) = Pr(A) \times Pr(B)$. In other words, under the iid assumption, we expect that $C(m, \epsilon) = C(1, \epsilon)^m$. The BDS test is then defined as

$$D(m, \epsilon) = \frac{\sqrt{T}[\hat{C}(m, \epsilon) - \{\hat{C}(1, \epsilon)\}^m]}{s(m, \epsilon)} \qquad (1.15)$$

where $\hat{C}(k, \epsilon)$ is given by

$$\hat{C}(k, \epsilon) = \frac{2}{T_k(T_k - 1)} \sum \sum_{k \le s < t \le T} I(\boldsymbol{x}_t^k, \boldsymbol{x}_s^k | \epsilon), \quad k = 1, m,$$

and $s(m, \epsilon)$ denotes the standard error of $\hat{C}(m, \epsilon) - \{\hat{C}(1, \epsilon)\}^m$, which can be consistently estimated from the data under the null hypothesis. For details, readers may consult Brock et al. (1996) or Tsay (2010, Chapter 4). In practice, one needs to select the embedding dimension m and the distance ϵ.

The BDS test is available in the fNonlinear package of R under the command bdsTest. The user has the option to select the maximum embedding dimension m and the distance ϵ. The default maximum dimension is $m = 3$, which means the embedding dimensions used in the test are 2 and 3. The default choices of ϵ are $(0.5, 1, 1.5, 2)\hat{\sigma}_x$, where $\hat{\sigma}_x$ denotes the sample standard error of x_t. To demonstrate, we consider a simulation of iid random variables from $N(0, 1)$ and a daily log return

series of the stock of International Business Machines (IBM) Corporation. Details are given below with output edited for simplicity.

R demonstration: BDS test using the package `fNonlinear`.

```
> require(fNonlinear)
> set.seed(1)
> x <- rnorm(300)
> bdsTest(x)
Title: BDS Test

Test Results:
  PARAMETER:
    Max Embedding Dimension: 3
    eps[1]: 0.482; eps[2]: 0.964
    eps[3]: 1.446; eps[4]: 1.927
  STATISTIC:
    eps[1] m=2: -1.1256;    eps[1] m=3: -1.4948
    eps[2] m=2: -0.7145;    eps[2] m=3: -1.1214
    eps[3] m=2: -0.6313;    eps[3] m=3: -0.8081
    eps[4] m=2: -0.7923;    eps[4] m=3: -1.2099
  P VALUE:
    eps[1] m=2: 0.2604;    eps[1] m=3: 0.135
    eps[2] m=2: 0.4749;    eps[2] m=3: 0.2621
    eps[3] m=2: 0.5278;    eps[3] m=3: 0.419
    eps[4] m=2: 0.4282;    eps[4] m=3: 0.2263

> require("quantmod")
> getSymbols("IBM",from="2010-01-02",to="2015-09-30")
> head(IBM)
           IBM.Open IBM.High IBM.Low IBM.Close IBM.Volume IBM.Adjusted
2010-01-04   131.18   132.97  130.85    132.45    6155300     117.6580
2010-01-05   131.68   131.85  130.10    130.85    6841400     116.2367
> rtn <- diff(log(as.numeric(IBM[,6])))
> ts.plot(rtn) # not shown
> Box.test(rtn,type="Ljung",lag=10)
        Box-Ljung test
data:  rtn
X-squared = 15.685, df = 10, p-value = 0.109 # No serial correlations
> bdsTest(rtn)
Title: BDS Test

Test Results:
  PARAMETER:
    Max Embedding Dimension: 3
    eps[1]: 0.006;    eps[2]: 0.012
    eps[3]: 0.018;    eps[4]: 0.024
  STATISTIC:
    eps[1] m=2: 4.4058;    eps[1] m=3: 5.3905
```

```
  eps[2] m=2: 4.2309;     eps[2] m=3: 5.429
  eps[3] m=2: 3.9341;     eps[3] m=3: 5.4255
  eps[4] m=2: 3.7848;     eps[4] m=3: 5.3304
 P VALUE:
  eps[1] m=2: 1.054e-05;    eps[1] m=3: 7.026e-08
  eps[2] m=2: 2.327e-05;    eps[2] m=3: 5.666e-08
  eps[3] m=2: 8.349e-05;    eps[3] m=3: 5.78e-08
  eps[4] m=2: 0.0001538;    eps[4] m=3: 9.802e-08
```

From the output, it is clear that in this particular instance the BDS test performs well. It fails to reject the null hypothesis of iid for the random sample from $N(0, 1)$, but successfully rejects the null hypothesis that the daily log returns of IBM are iid.

Discussion: The BDS test is designed to test the null hypothesis of iid, as such one needs to remove any linear dynamic dependence before applying the test to detect nonlinearity in a time series. In other words, care must be exercised in using the BDS test to detect nonlinearity. In practice, the test is typically applied to the residuals of a fitted linear time series model. In addition, a rejection by the BDS test does not provide any specific information to improve the fitted model. Further analysis of the residuals is often needed to seek directions for model refinement after the null hypothesis is rejected by the test. As suggested by the default option, the distance ϵ used in the BDS test should be related to the standard error of x_t.

The McLeod–Li test: McLeod and Li (1983) proposed a general Portmanteau test for nonlinearity under the assumption that the time series x_t is fourth-order stationary, i.e. the x_t^2 process is weakly stationary. Similar to the BDS test, the proposed Portmanteau test is typically applied to the residual $\hat{a}_t$ of a fitted linear time series model. Define the lag-ℓ autocorrelation of the squared residuals as

$$\hat{\rho}_{aa}(\ell) = \frac{\sum_{t=\ell+1}^{T} \left(\hat{a}_t^2 - \hat{\sigma}^2 \right) \left(\hat{a}_{t-\ell}^2 - \hat{\sigma}^2 \right)}{\sum_{t=1}^{T} \left(\hat{a}_t^2 - \hat{\sigma}^2 \right)},$$

where $\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{a}_t^2 / T$ and T is the sample size. McLeod and Li show that, for a fixed positive integer m, the joint distribution of

$$\sqrt{T}[\hat{\rho}_{aa}(1), \hat{\rho}_{aa}(2), \ldots, \hat{\rho}_{aa}(m)]'$$

is asymptotically multivariate normal with mean zero and identity covariance matrix provided that the fitted linear model is adequate for the x_t series. Using this result, McLeod and Li (1983) proposed the Portmanteau statistics

$$Q^*(m) = T(T+2) \sum_{\ell=1}^{m} \frac{\hat{\rho}_{aa}^2(\ell)}{T - \ell} \tag{1.16}$$

to detect nonlinearity in x_t. Under the assumption that x_t is fourth-order stationary and the fitted linear model is adequate, $Q^*(m)$ is asymptotically distributed as χ^2_m. This is essentially a Ljung–Box test on the x^2_t series.

In the literature, the Portmanteau statistics of Equation (1.16) is often used to check for conditional heteroscedasticity. As a matter of fact, the test is asymptotically equivalent to the Lagrange multiplier test of Engle (1982) for the autoregressive conditional heteroscedastic (ARCH) model. To test for ARCH effects, Engle (1982) uses the AR(m) model

$$\hat{a}^2_t = \beta_0 + \beta_1 \hat{a}^2_{t-1} + \cdots + \beta_m \hat{a}^2_{t-m} + \epsilon_t, \qquad (1.17)$$

where ϵ_t denotes the error term, and considers the null hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0$ versus the alternative hypothesis $H_a : \beta_i \neq 0$ for some $i \in \{1, \ldots, m\}$. The F-statistic of the linear regression in Equation (1.17) can be used to perform the test. Alternatively, one can use mF as a test statistic. Under the same conditions as those of McLeod and Li (1983), mF is asymptotically distributed as χ^2_m. The $Q^*(m)$ statistics of Equation (1.16) can be easily computed. We demonstrate this below.

R demonstration: McLeod–Li and Engle tests for nonlinearity.

```
> set.seed(15)
> xt <- rnorm(300)
> Box.test(xt,lag=10,type='Ljung')
        Box-Ljung test
data:  xt
X-squared = 11.095, df = 10, p-value = 0.3502 # No serial correlations
> Box.test(xt^2,lag=10,type='Ljung')
        Box-Ljung test
data:  xt^2
X-squared = 8.4539, df = 10, p-value = 0.5846 # Q-star test

> require(quantmod)
> getSymbols("MSFT",from="2009-01-02",to="2015-10-15",src="google")
> msft <- diff(log(as.numeric(MSFT$MSFT.Close)))  #log returns
> Box.test(msft,lag=10,type='Ljung')
        Box-Ljung test
data:  msft
X-squared = 17.122, df = 10, p-value = 0.0717 # No serial correlations
> Box.test(msft^2,lag=10,type='Ljung')
        Box-Ljung test
data:  msft^2
X-squared = 70.54, df = 10, p-value = 3.487e-11 ## Nonlinearity
### Engle's test with m = 10.
> nT <- length(msft)
> y <- msft[11:nT]^2
```

```
> X <- NULL
> for (i in 1:10)
+ X <- cbind(X,msft[(11-i):(nT-i)]^2)
+ }
> m2 <- lm(y ~X)  ## Linear regression
> anova(m2)
Analysis of Variance Table
Response: y
            Df      Sum Sq     Mean Sq F value     Pr(>F)
X           10 0.00003036 3.0358e-06    4.65 1.447e-06 ***
Residuals 1687 0.00110138 6.5286e-07
> qstar = 4.64*10
> pv = 1 -pchisq(qstar,10)
> print(c(qstar,pv))
[1] 4.65e+01 1.162495e-06  ### ARCH effect exists
```

From the output, the traditional Ljung–Box statistics confirm that both the series of iid $N(0, 1)$ random variables and the daily log returns of Microsoft (MSFT) stock from January 2010 to October 15, 2015 have no serial correlations. On the other hand, the McLeod–Li test statistic of (1.16) cannot reject the null hypothesis of linear time series for the $N(0, 1)$ iid random sample, but it clearly rejects linearity for the daily log returns of MSFT stock. The ARCH effect is also confirmed by the Lagrange multiplier test of Engle (1982).

Rank-based Portmanteau test: The McLeod–Li test requires the existence of the fourth moment of the underlying time series x_t. In some applications, empirical data may exhibit high excess kurtosis. In this situation, the performance of the McLeod–Li test may deteriorate. To overcome the impact of heavy-tails on the McLeod–Li test, one can apply the rank-based Ljung–Box statistics. See, for instance, Dufour and Roy (1986) and Tsay (2014, Chapter 7) and the references therein. Let R_i be the rank of $\hat{a}_i^2$ in the squared residuals $\{\hat{a}_t^2\}$. The lag-ℓ rank-based serial correlation is defined as

$$\tilde{\rho}_\ell = \frac{\sum_{t=\ell+1}^{T}(R_t - \bar{R})(R_{t-\ell} - \bar{R})}{\sum_{t=1}^{T}(R_t - \bar{R})^2}, \quad \ell = 1, 2, \ldots$$

where it can be shown that

$$\bar{R} = \sum_{t=1}^{T} R_t/T = (T + 1)/2$$

$$\sum_{t=1}^{T}(R_t - \bar{R})^2 = T(T^2 - 1)/12.$$

Furthermore, Dufour and Roy (1986) show that

$$E(\tilde{\rho}_\ell) = -(T - \ell)/[T(T - 1)],$$

$$\text{Var}(\tilde{\rho}_\ell) = \frac{5T^4 - (5\ell + 9)T^3 + 9(\ell - 2)T^2 + 2\ell(5\ell + 8)T + 16\ell^2}{5(T - 1)^2 T^2(T + 1)}.$$

The rank-based Portmanteau statistic then becomes

$$Q_R(m) = \sum_{\ell=1}^{m} \frac{[\tilde{\rho}_\ell - E(\tilde{\rho}_\ell)]^2}{\text{Var}(\tilde{\rho}_\ell)}, \tag{1.18}$$

which is asymptotically distributed as χ_m^2 if the $\{\hat{a}_t^2\}$ series has no serial correlations.

To compare the finite-sample performance of the McLeod–Li Portmanteau test in Equation (1.16) and its rank-based counterpart in Equation (1.18), we conduct a simple simulation. We generate iid sequences from $N(0, 1)$ and t_5 distributions with 300 observations and compute the two test statistics for $m = 1$, 5, and 10. This process is repeated for 10,000 iterations. Some selected quantiles of the empirical distributions of $Q^*(1), Q^*(5), Q^*(10)$ and $Q_R(1), Q_R(5), Q_R(10)$ are given in Table 1.4 along with the quantiles of the corresponding χ_m^2 distributions.

Figure 1.15 shows the empirical density functions $Q^*(10)$ and $Q_R(10)$ over the 10,000 iterations. The iid sequences were generated from $N(0, 1)$ for 300 observations. The solid and dashed lines are for $Q^*(10)$ and $Q_R(10)$, respectively. The dotted line of the plot shows the density function of χ_{10}^2. Figure 1.16 shows similar density functions when the iid sequences were generated from a Student-t distribution with 5 degrees of freedom. From Table 1.4 and Figures 1.15 and 1.16, it is seen that both $Q^*(10)$ of McLeod–Li test in Equation (1.16) and the rank-based $Q_R(10)$ in Equation (1.18) follow reasonably well the asymptotic χ_{10}^2 distribution under the normality. On the other hand, the $Q^*(10)$ statistic encounters some size distortion and power loss when the iid sequences are from the t_5 distribution whereas the rank-based Portmanteau statistic $Q_R(10)$ continues to perform well. See the shift of the empirical density of $Q_m(10)$ in Figure 1.16. Therefore, care must be exercised when one applies the McLeod–Li test to detect nonlinearity when the underlying process has heavy tails.

The Peña–Rodriguez test: Peña and Rodriguez (2002) proposed a modified Portmanteau test statistic that can be used for model checking of a fitted linear time series model, including nonlinearity in the residuals. Using simulation studies, the authors show that their modified test statistics perform well in finite samples. For simplicity, let z_t be a function of the residual series $\hat{a}_t$ of a fitted linear

| Distribution	Statistics	0.025	0.05	0.5	0.95	0.975
	χ_1^2	0.00098	0.00393	0.455	3.842	5.024
$N(0,1)$	$Q^*(1)$	0.00096	0.00372	0.451	3.744	4.901
	$Q_R(1)$	0.00088	0.00392	0.463	3.762	5.056
t_5	$Q^*(1)$	0.00074	0.00259	0.288	2.936	4.677
	$Q_R(1)$	0.00098	0.00367	0.441	3.716	4.856
	χ_5^2	0.8312	1.1455	4.351	11.070	12.833
$N(0,1)$	$Q^*(5)$	0.7903	1.0987	4.129	11.045	13.131
	$Q_R(5)$	0.7884	1.1193	4.263	11.042	12.792
t_5	$Q^*(5)$	0.3634	0.5575	2.909	10.992	15.047
	$Q_R(5)$	0.8393	1.1287	4.254	11.123	13.002
	χ_{10}^2	3.2470	3.9403	9.342	18.307	20.483
$N(0,1)$	$Q^*(10)$	3.1077	3.7532	8.870	12.286	21.122
	$Q_R(10)$	3.1737	3.8524	9.184	18.163	20.516
t_5	$Q^*(10)$	1.1347	1.6916	6.685	19.045	23.543
	$Q_R(10)$	3.2115	3.8265	9.216	18.495	20.786

Table 1.4 Empirical quantiles of McLeod–Li Portmanteau statistics and the rank-based Portmanteau statistics for random samples of 300 observations. The realizations are generated from $N(0,1)$ and Student-t distribution with 5 degrees of freedom. The results are based on 10,000 iterations.

model. For instance, $z_t = \hat{a}_t^2$ or $z_t = |\hat{a}_t|$. The lag-ℓ sample autocorrelation of z_t is defined as

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^T (z_{t-\ell} - \bar{z})(z_t - \bar{z})}{\sum_{t=1}^T (z_t - \bar{z})^2}, \tag{1.19}$$

where, as before, T is the sample size and $\bar{z} = \sum_t z_t / T$ is the sample mean of z_t. As usual, $\hat{\rho}_\ell$ is a consistent estimate of the lag-ℓ autocorrelation ρ_ℓ of z_t under some regularity conditions. For a given positive integer m, the Peña and Rodriguez test statistic for testing the null hypothesis of $H_0 : \rho_1 = \cdots = \rho_m = 0$ is

$$\hat{D}_m = T[1 - |\hat{\mathbf{R}}_m|^{1/m}], \tag{1.20}$$
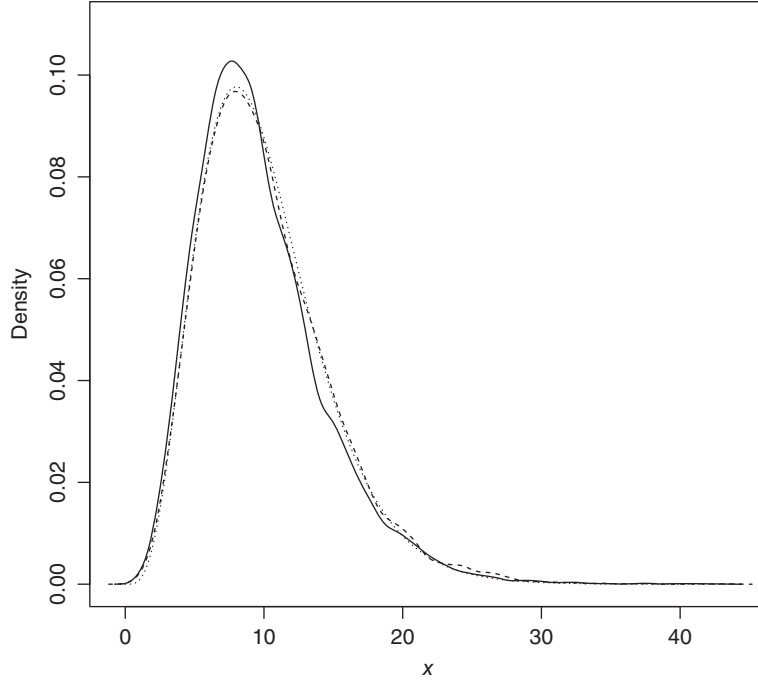
Figure 1.15 Empirical density functions of $Q^*(10)$ in Equation (1.16) (solid line) and $Q_R(10)$ in Equation (1.18) (dashed line) when the iid sequences are generated from $N(0, 1)$. The results are for sample size 300 and 10,000 iterations. The dotted line denotes the density function of χ^2_{10}.

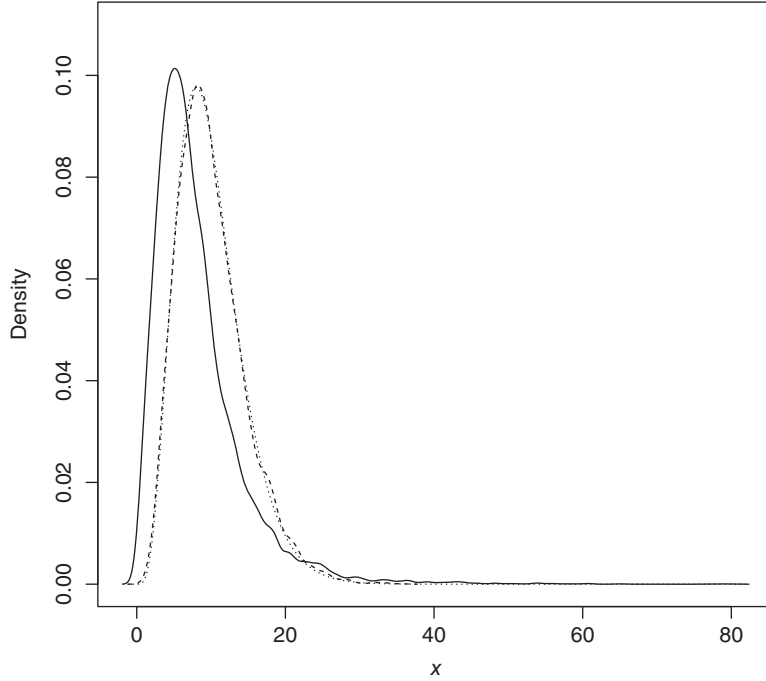where the $(m + 1)$ by $(m + 1)$ matrix $\widehat{\mathbf{R}}_m$ is defined below:

$$
\widehat{\boldsymbol{R}}_m = \begin{bmatrix}
1 & \hat{\rho}_1 & \cdots & \hat{\rho}_m \\
\hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{m-1} \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\rho}_m & \hat{\rho}_{m-1} & \cdots & 1
\end{bmatrix}.
\tag{1.21}
$$

Using the idea of pseudo-likelihood, Peña and Rodriguez (2006) further modified the test statistic in Equation (1.20) to

$$
D_m^* = -\frac{T}{m + 1} \log(|\widehat{\boldsymbol{R}}_m|),
\tag{1.22}
$$

where the denominator $m + 1$ is the dimension of $\widehat{\mathbf{R}}_m$. Under the assumption that the fitted ARMA(p, q) model is correctly specified, Peña and Rodriguez (2006)

Figure 1.16   Empirical density functions of $Q^*(10)$ in Equation (1.16) (solid line) and $Q_R(10)$ in Equation (1.18) (dashed line) when the iid sequences are generated from a Student-t distribution with 5 degrees of freedom. The results are for sample size 300 and 10,000 iterations. The dotted line denotes the density function of χ_{10}^2.

show that the test statistic D_m^* of Equation (1.22) is asymptotically distributed as a mixture of m independent χ_1^2 random variates. The weights of the mixture are rather complicated. However, the authors derived two approximations to simplify the calculation. The first approximation of D_m^* is denoted by GD_m^*, which is approximately distributed as a gamma random variate, $\Gamma(\alpha, \beta)$, where

$$\alpha = \frac{3(m+1)[m-2(p+q)]^2}{2[2m(2m+1)-12(m+1)(p+q)]},$$

$$\beta = \frac{3(m+1)[m-2(p+q)]}{2m(2m+1)-12(m+1)(p+q)}.$$

The second approximation is

$$ND_m^* = (\alpha/\beta)^{-1/\lambda}(\lambda/\sqrt{\alpha})\left[(D_m^*)^{1/\lambda} - (\alpha/\beta)^{1/\lambda}\left(1 - \frac{\lambda-1}{2\alpha\lambda^2}\right)\right], \quad (1.23)$$

where α and β are defined as before, and

$$\lambda = \left[1 - \frac{2(m/2 - (p+q))(m^2/(4(m+1)) - (p+q))}{3(m(2m+1)/(6(m+1)) - (p+q))^2}\right]^{-1}.$$

Asymptotically, ND_n^* is distributed as $N(0, 1)$. For large m, $\lambda \approx 4$. In practice, the authors recommend replacing $\hat{\rho}_\ell$ by $\tilde{\rho}_\ell = (T+2)\hat{\rho}_\ell/(T-\ell)$ in the $\hat{\boldsymbol{R}}_m$ matrix of Equation (1.21) for better performance in finite samples, especially when T is small.

To demonstrate the ND_m^* statistic of Equation (1.23), we consider the daily log returns of Microsoft stock from January 2, 2009 to October 15, 2015. The same series was used before in demonstrating the McLeod–Li test. The sample size is $T = 1708$ so that we select $m = \lfloor \log(T) \rfloor + 1 = 8$. The test, denoted by PRnd, is available in the package NTS.

R demonstration: Peña–Rodriguez test.

```
> require(quantmod); require(NTS)
> getSymbols("MSFT",from="2009-01-02",to="2015-10-15",src="google")
> msft <- diff(log(as.numeric(MSFT$MSFT.Close)))  # log returns
> log(length(msft))
[1] 7.443078
> PRnd(msft,m=8)
ND-stat & p-value  0.1828238 0.8549363
> PRnd(abs(msft),m=8)
ND-stat & p-value  2.302991 0.02127936
```

From the output, the daily log returns have no significant serial correlations, but their absolute values have serial correlations. Again, the results confirm the nonlinearity in the daily MSFT log returns.

1.4.2 Parametric Tests

Most parametric tests for nonlinearity are developed under the assumption of certain nonlinear models as the alternative. Roughly speaking, a well-behaved zero-mean stationary time series x_t can be written as a Volterra series

$$x_t = \sum_{i=1}^{t-1} \psi_i x_{t-i} + \sum_{i=1}^{t-1}\sum_{j=1}^{t-1} \psi_{ij} x_{t-i} x_{t-j} + \sum_i \sum_j \sum_k \phi_{ijk} x_{t-i} x_{t-j} x_{t-k} + \cdots + \epsilon_t, \quad (1.24)$$

where ϵ_t denotes the noise term and the ψs are real numbers. For a linear time series x_t, we have $\psi_{ij} = \psi_{ijk} = \cdots = 0$. Therefore, some of the higher-order coefficients are non-zero if x_t is nonlinear. If the third order and higher coefficients are zero, but $\psi_{ij} \neq 0$ for some i and j, then x_t becomes a bilinear process. Parametric

tests for nonlinearity of x_t are statistics that exploit certain features of the Volterra series in Equation (1.24). We discuss some of the parametric tests in this section.

The RESET test: Ramsey (1969) proposed a specification test for linear least squares regression analysis. The test is referred to as a RESET test and is readily applicable to linear AR models. Consider the linear AR(p) model

$$x_t = X'_{t-1}\phi + a_t, \tag{1.25}$$

where $X_{t-1} = (1, x_{t-1}, \ldots, x_{t-p})'$ and $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$. The first step of the RESET test is to obtain the least squares estimate $\hat{\phi}$ of Equation (1.25) and compute the fitted value $\hat{x}_t = X'_{t-1}\hat{\phi}$, the residual $\hat{a}_t = x_t - \hat{x}_t$, and the sum of squared residuals $SSR_0 = \sum_{t=p+1}^{T}\hat{a}_t^2$, where T is the sample size. In the second step, consider the linear regression

$$\hat{a}_t = X'_{t-1}\alpha_1 + M'_{t-1}\alpha_2 + v_t, \tag{1.26}$$

where $M_{t-1} = (\hat{x}_t^2, \ldots, \hat{x}_t^{s+1})'$ for some $s \geq 1$, and compute the least squares residuals

$$\hat{v}_t = \hat{a}_t - X'_{t-1}\hat{\alpha}_1 - M'_{t-1}\hat{\alpha}_2$$

and the sum of squared residuals $SSR_1 = \sum_{t=p+1}^{T}\hat{v}_t^2$ of the regression. The basic idea of the RESET test is that if the linear AR(p) model in Equation (1.25) is adequate, then α_1 and α_2 of Equation (1.26) should be zero. This can be tested by the usual F statistic of Equation (1.26) given by

$$F = \frac{(SSR_0 - SSR_1)/g}{SSR_1/(T - p - g)} \qquad \text{with} \quad g = s + p + 1, \tag{1.27}$$

which, under the linearity and normality assumption, has an F distribution with degrees of freedom g and $T - p - g$.

Remark: Because the variables $\hat{x}_t^k$ for $k = 2, \ldots, s+1$ tend to be highly correlated with X_{t-1} and among themselves, principal components of M_{t-1} that are not co-linear with X_{t-1} are often used in fitting Equation (1.26).

Keenan (1985) proposed a nonlinearity test for time series that uses $\hat{x}_t^2$ only and modifies the second step of the RESET test to avoid multicollinearity between $\hat{x}_t^2$ and X_{t-1}. Specifically, the linear regression (1.26) is divided into two steps. In step 2(a), one removes linear dependence of $\hat{x}_t^2$ on X_{t-1} by fitting the regression

$$\hat{x}_t^2 = X'_{t-1}\beta + u_t$$

and obtaining the residual $\hat{u}_t = \hat{x}_t^2 - X_{t-1}\widehat{\beta}$. In step 2(b), consider the linear regression

$$\hat{a}_t = \hat{u}_t \alpha + v_t,$$

and obtain the sum of squared residuals $SSR_1 = \sum_{t=p+1}^{T}(\hat{a}_t - \hat{u}_t\hat{\alpha})^2 = \sum_{t=p+1}^{T}\hat{v}_t^2$ to test the null hypothesis $\alpha = 0$.

The F-test: To improve the power of Keenan's and RESET tests, Tsay (1986) used a different choice of the regressor M_{t-1}. Specifically, he suggested using $M_{t-1} = vech(X_{t-1}X'_{t-1})$, where $vech(A)$ denotes the half-stacking vector of the matrix A using elements on and below the diagonal only. For example, if $p = 2$, then $M_{t-1} = (x_{t-1}^2, x_{t-1}x_{t-2}, x_{t-2}^2)'$. The dimension of M_{t-1} is $p(p+1)/2$ for an AR(p) model. In practice, the test is simply the usual partial F statistic for testing $\alpha = 0$ in the linear least squares regression

$$x_t = X'_{t-1}\phi + M'_{t-1}\alpha + e_t,$$

where e_t denotes the error term. Under the assumption that x_t is a linear AR(p) process, the partial F statistic follows an F distribution with degrees of freedom g and $T - p - g - 1$, where $g = p(p+1)/2$. We refer to this F test as the *F-test*. Luukkonen et al. (1988) further extended the test by augmenting M_{t-1} with cubic terms x_{t-i}^3 for $i = 1, \ldots, p$.

To demonstrate nonlinearity tests, consider the US quarterly unemployment rates of Example 1.1. In this particular instance, both Keenan test and F-test confirm that the series is nonlinear. The F-test is referred to as the `Tsay.test` in the `TSA` package.

R demonstration: Parametric tests.

```
> require(TSA)
> da=read.table("q-unrate.txt",header=T)
> rate <- da$rate
> Keenan.test(unrate)
$test.stat
[1] 7.428776
$p.value
[1] 0.006858382
$order
[1] 5
> Tsay.test(unrate)
$test.stat
[1] 2.626
$p.value
```

```
[1] 0.00108
$order
[1] 5
```

Remark: The dimension of M_{t-1} used in the F-test is $p(p+1)/2$, which can be high when the order p is not small compared with the sample size T. In this situation, the F-test might have low power in detecting nonlinearity. To improve the F-test, one can use the idea of thresholding. Specifically, one constructs M_{t-1} using only those lagged variables in X_{t-1} with t-ratio greater than some threshold in modulus. To illustrate, consider the well-known Canadian Lynx data, which is available in the TSA package. The series has 114 observations. With log-transformation, an AR(11) is selected by the Keenan.test and Tsay.test of the TSA package. In this particular case, the Keenan test shows that the log series is nonlinear, but the F-test fails. This is not surprising because the F-test fits $11(12)/2 = 66$ additional parameters whereas the effective sample size is only $114 - 11 = 103$. On the other hand, with threshold 1.645, the F-test also confirms that the log series is nonlinear. Here the threshold 1.645 selects five lagged variables to construct M_{t-1} in the F-test.

R demonstration:

```
> require(TSA)
> data(lynx)
> y <- log10(lynx)
> Keenan.test(y)
$test.stat
[1] 11.66997
$p.value
[1] 0.0009550016
$order
[1] 11
> Tsay.test(y)
$test.stat
[1] 1.316
$p.value
[1] 0.2256
$order
[1] 11
> require(NTS)
> F.test(y,thres=1.645)
$test.stat
[1] 1.971
$p.value
```

```
[1] 0.02858
$order
[1] 11
```

Threshold test: If the alternative model under study is the self-exciting threshold autoregressive (TAR) model shown in Example 1.1, one can derive specific test statistics to increase the power of detecting nonlinearity. One of the specific tests is the likelihood ratio statistic studied by Chan and Tong (1990) and Chan (1990, 1991). The TAR(p) model considered can be written as

$$x_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + I(x_{t-d} > r)\left(\beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i}\right) + a_t, \qquad (1.28)$$

where a_t is a sequence of independent and identically distributed Gaussian random variables with mean zero and variance σ^2, ϕ_i and β_j are real valued parameters, $d > 0$ with x_{t-d} being the threshold variable, r is the threshold, and $I(x_{t-d} > r)$ is an indicator variable such that $I(x_{t-d} > r) = 1$ if $x_{t-d} > r$ and $= 0$ otherwise. The null hypothesis is $H_0 : \beta_i = 0$ for $i \in \{0, \dots, p\}$ and the alternative hypothesis is $H_a : \beta_i \neq 0$ for some i. In Equation (1.28), we assume that the threshold r lies inside a known bounded closed interval $\tilde{R} \subseteq R$. For simplicity, we further assume that $d \leq p$. The likelihood ratio test statistic is then

$$\lambda = \frac{n\left(\hat{\sigma}_0^2 - \hat{\sigma}^2\right)}{\hat{\sigma}^2}, \qquad (1.29)$$

where $n = T - p + 1$ with T being the sample size,

$$\hat{\sigma}^2 = \min_{r \in \tilde{R}, \theta, \phi}\left[\sum_{t=p+1}^{T}\left\{x_t - \phi_0 - \sum_{i=1}^{p}\phi_i x_{t-i} - I(x_{t-d} > r)\left(\beta_0 + \sum_{i=1}^{p}\beta_i x_{t-i}\right)\right\}^2\right],$$

$$\hat{\sigma}_0^2 = \min_{\phi}\left\{\sum_{t=p+1}^{T}\left(x_t - \phi_0 - \sum_{i=1}^{p}\phi_i x_{t-i}\right)^2\right\}.$$

The null hypothesis is rejected if and only if λ of Equation (1.29) is large. The limiting properties of the test statistic λ, however, are non-standard and more involved because the threshold r is undefined under the null hypothesis H_0. This is an issue of *nuisance parameter* under the null hypothesis. Interested readers are referred to Chan and Tong (1990). The percentage points of λ under the null hypothesis have been investigated by Chan (1990). In the literature, it is common to assume that $\tilde{R} = [v, u]$, where v and u are pre-specified real numbers with $v < u$. See, for instance, Davis (1987) and Andrews and Ploberger (1994).

Another specific threshold nonlinearity test is to transform the testing problem into a change point detection problem. The transformation is carried out via the *arranged autoregression*. See Tsay (1989, 1998) and the references therein. Specifically, the arranged autoregression transfers the TAR model under the alternative hypothesis H_a into a model-change problem with the threshold r being the change point. To see this, the TAR model in Equation (1.28) says that x_t follows essentially two linear models depending on whether $x_{t-d} < r$ or $x_{t-d} \geq r$. For a realization of T data points, say $\{x_t\}_{t=1}^T$, the threshold variable x_{t-d} can assume values $\{x_1, \ldots, x_{T-d}\}$. Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(T-d)}$ be the ordered statistics of $\{x_t\}_{t=1}^{T-d}$. The TAR model can then be written as

$$x_{(j)+d} = \theta_0 + \sum_{i=1}^p \theta_i x_{(j)+d-i} + a_{(j)+d}, \quad j = 1, \ldots, T-d, \qquad (1.30)$$

where $\theta_i = \phi_i$ if $x_{(j)} \leq r$, and $\theta_i = \phi_i + \beta_i$ if $x_{(j)} > r$, for $i = 0, \ldots, p$. Consequently, the threshold r is a change point for the linear regression in Equation (1.30), and we refer to Equation (1.30) as an arranged autoregression (in increasing order of the threshold variable x_{t-d}). Note that the arranged autoregression in Equation (1.30) does not alter the dynamic dependence of x_t on x_{t-i} for $i = 1, \ldots, p$ because $x_{(j)+d}$ still depends on $x_{(j)+d-i}$ for $i = 1, \ldots, p$. What is done is simply to present the TAR model in the threshold space instead of in the time space. That is, the equation with a smaller x_{t-d} appears before that with a larger x_{t-d}. The threshold test of Tsay (1989) is obtained as follows.

- Step 1: Fit Equation (1.30) using $j = 1, \ldots, m$, where m is a pre-specified positive integer (e.g., 30). Denote the least squares estimates of θ_i by $\hat{\theta}_{i,m}$, where m denotes the number of data points used in estimation.

- Step 2: Compute the predictive residual

$$\hat{a}_{(m+1)+d} = x_{(m+1)+d} - \hat{\theta}_{0,m} - \sum_{i=1}^p \hat{\theta}_{i,m} x_{(m+1)+d-i}$$

  and its standard error. Let $\hat{e}_{(m+1)+d}$ be the standardized predictive residual.

- Step 3: Use the recursive least squares method to update the least squares estimates to $\hat{\theta}_{i,m+1}$ by incorporating the new data point $x_{(m+1)+d}$.

- Step 4: Repeat Steps 2 and 3 until all data points are processed.

- Step 5: Consider the linear regression of the standardized predictive residual

$$\hat{e}_{(m+j)+d} = \alpha_0 + \sum_{i=1}^p \alpha_i x_{(m+j)+d-i} + v_t, \quad j = 1, \ldots, T-d-m \quad (1.31)$$

and compute the usual F statistic for testing $\alpha_i = 0$ in Equation (1.31) for $i = 0, \ldots, p$. Under the null hypothesis that x_t follows a linear AR(p) model, the F ratio has a limiting F distribution with degrees of freedom $p + 1$ and $T - d - m - p$.

We refer to the resulting F test as a *Tar-F test*. The idea behind the test is that under the null hypothesis there is no model change in the arranged autoregression in Equation (1.30) so that the standardized predictive residuals should be close to iid with mean zero and variance 1. In this case, they should have no correlations with the regressors $x_{(m+j)+d-i}$. For further details, including formulas for a recursive least squares method and some simulation study on performance of the Tar-F test, see Tsay (1989). The Tar-F test avoids the problem of nuisance parameters encountered by the likelihood ratio test. It does not require knowing the threshold r. It simply tests that the predictive residuals have no correlations with regressors if the null hypothesis holds. Therefore, the test does not depend on knowing the number of regimes in the alternative model. Yet the Tar-F test is not as powerful as the likelihood ratio test if the true model is indeed a two-regime TAR model with Gaussian innovations.

The likelihood ratio test for a two-regime Gaussian TAR model is available via the command `tlrt` of the TSA package, whereas the Tar-F test is available via the command `thr.test` of the NTS package. For the US quarterly unemployment rates of Example 1.1, both tests detect nonlinearity with $p = 5$ and $d = 5$ and default options. Details are given below.

R demonstration: Threshold tests.

```
> da <- read.table("q-unrate.txt",header=T)
> unrate <- da$rate
> require(NTS)
> require(TSA)
> thr.test(unrate,p=5,d=5)
SETAR model is entertained
Threshold nonlinearity test for (p,d):  5 5
F-ratio and p-value:  2.849921 0.01081263
> tlrt(unrate,p=5,d=5)
$percentiles
[1] 24.9 75.1
$test.statistic
    33.028
$p.value
    0.0003066953
```

1.5 EXERCISES

1.1 Consider the global land and ocean temperature anomalies from January 1880 to May 2017. The monthly data are available from the National Oceanic and Atmospheric Administration (NOAA) at www.ncdc.noaa.gov and are in the file `globaltemp1880-2017.csv`. The data are in degrees Celsius and the base period is 1901–2000. The temperature anomalies show an upward trend so consider the change series, i.e. the first difference. Is the differenced series nonlinear? Perform some tests and draw a conclusion. Does the series show threshold nonlinearity? Why?

1.2 Consider the daily log returns of the exchange-traded fund for the S&P 500 index with tick symbol `SPY` from January 3, 2007 to June 30, 2017. Daily log returns denote the first difference of log daily closing prices. The first column of the file `SPY0717.txt` contains the log return. Is the return series nonlinear? Perform some tests, including the BDS test to draw a conclusion.

1.3 Asset return series tend to exhibit conditional heteroscedasticity. The GARCH models are often used to handle conditional heteroscedasticity. Consider the SPY daily log returns of previous problem. The second column of the file `SPY0717.txt` shows the standardized residuals of a fitted AR(1)-GARCH(1,1) model with standardized Student-t innovations. Does the standardized residual series consist of independent and identically distributed random variables? Why?

1.4 Consider the grow rate series (first difference of the log index) of the US monthly producer price index from February 1913 to August 2017. The data can be obtained from FRED (PPIACO, 1982 index=100, not seasonally adjusted). The data are also in the file `PPIACO.csv`. Is the growth rate series linear? Perform some nonlinearity tests to draw a conclusion.

1.5 Consider again the growth rate series of the US monthly producer price index of the prior problem. Let x_t be the growth rate series. Use the `fGarch` package to fit the model:

```
require(fGarch)
m1 <- garchFit( arma(3,2)+garch(1,1),data=x_t,trace=F,
            cond.dist="std")
rt <- residuals(m1,standardize=T)
```

The `rt` denotes the standardized residuals. Apply the McLeod–Li test to detect any nonlinearity in the `rt` series and apply the BDS test to check whether the `rt` series consists of iid random variables.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**: 716–723.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**: 1383–1414.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, revised ed. Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*, 5th edition. Wiley, Hoboken, NJ.

Brock, W., Dechert, W.D., and Scheinkman, J. (1987). A test for independence based on the correlation dimension. Working paper, Department of Economics, University of Wisconsin-Madison.

Brock, W., Dechert, W. D., Scheinkman, J., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews* **15**: 197–235.

Chan, K. S. (1990). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B* **53**: 691–696.

Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society Series B* **53**: 691–696.

Chan, K. S. and Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B* **52**: 469–476.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* **88**: 298-308.

Chen, R. and Tsay, R. S. (1996). Nonlinear transfer function. *Journal of Nonparametric Statistics* **6**: 193–204.

Davis, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**: 33–43.

De Gooijer, J. G. (2017). *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer, Cham, Switzerland.

Douc, R., Moulines, E., and Stoffer, D. S. (2014). *Nonlinear Time Series: Theory, Methods, and Applications with R Examples*. CRC Press, Boca Raton, FL.

Dufour, J. M. and Roy, R. (1986). Generalized portmanteau statistics and tests of randomness. *Communications in Statistics – Theory and Methods* **15**: 2953–2972,

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* **50**: 987–1007.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of Time Series Analysis* **27**: 923–942.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

Keenan, D. M. (1985). A Tukey non-additivity-type test for time series nonlinearity. *Biometrika* **72**: 39–44.

Liboschik, T., Fried, R., Fokianos, K., Probst, P., and Rathjens, J. (2015). *Analysis of Count Time Series*, R Package, http://tscount.r-forge.r-project.org.

Luukkonen, R., Saikkonen, P., and Teräsvirta (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* **75**: 491–499.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4**: 269–273.

Peña, D. and Rodriguez, J. (2002). A powerful Portmanteau test of lack of fit for time series. *Journal of the American Statistical Association* **97**: 601–610.

Peña, D. and Rodriguez, J. (2006). The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *Journal of Statistical Planning and Inference* **136**: 2706–2718.

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society Series B* **31**: 350–371.

Tong, H. (1978). On a threshold model. In C.H. Chen (ed.), *Pattern Recognition and Signal Processing*. Sijhoff & Noordhoff, Amsterdam.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Tong, H. and Lim, K. S. (1980). Threshold autoregression,limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* **42**: 245–292.

Tsay, R. S. (1986). Nonlinearity tests for time series. *Biometrika* **73**: 461–466.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* **84**: 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1202.

Tsay, R. S. (2010). *Analysis of Financial Time Series*, 3rd edition. John Wiley, Hoboken, NJ.

Tsay, R. S. (2014). *Multivariate Time Series Analysis with R and Financial Applications*, John Wiley, Hoboken, NJ.

Tsay, R. S. and Wu, C. S. (2003). Forecasting with leading indicators revisited. *Journal of Forecasting* **22**: 603–617.