CHAPTER 1

INTRODUCTION TO FORECASTING

It is difficult to make predictions, especially about the future

NEILS BOHR, *Danish physicist*

1.1 THE NATURE AND USES OF FORECASTS

A **forecast** is a prediction of some future event or events. As suggested by Neils Bohr, making good predictions is not always easy. Famously "bad" forecasts include the following from the book *Bad Predictions*:

- "The population is constant in size and will remain so right up to the end of mankind." *L'Encyclopedie*, 1756.
- "1930 will be a splendid employment year." U.S. Department of Labor, *New Year's Forecast* in 1929, just before the market crash on October 29.
- "Computers are multiplying at a rapid rate. By the turn of the century there will be 220,000 in the U.S." *Wall Street Journal*, 1966.

Introduction to Time Series Analysis and Forecasting, Second Edition.
Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci.

Forecasting is an important problem that spans many fields including business and industry, government, economics, environmental sciences, medicine, social science, politics, and finance. Forecasting problems are often classified as short-term, medium-term, and long-term. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, and months) into the future. Medium-term forecasts extend from 1 to 2 years into the future, and long-term forecasting problems can extend beyond that by many years. Short- and medium-term forecasts are required for activities that range from operations management to budgeting and selecting new research and development projects. Long-term forecasts impact issues such as strategic planning. Short- and medium-term forecasting is typically based on identifying, modeling, and extrapolating the patterns found in historical data. Because these historical data usually exhibit inertia and do not change dramatically very quickly, statistical methods are very useful for short- and medium-term forecasting. This book is about the use of these statistical methods.

Most forecasting problems involve the use of time series data. A **time series** is a time-oriented or chronological sequence of observations on a variable of interest. For example, Figure 1.1 shows the market yield on US Treasury Securities at 10-year constant maturity from April 1953 through December 2006 (data in Appendix B, Table B.1). This graph is called a **time**

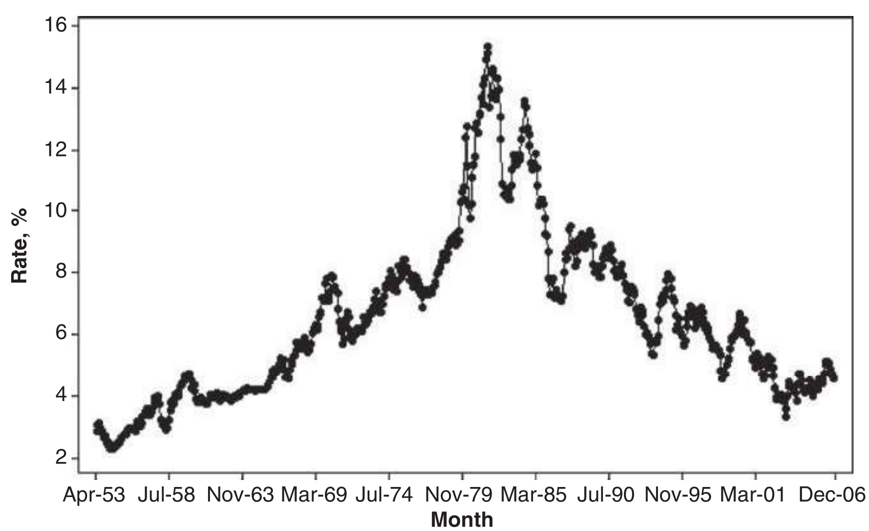


FIGURE 1.1 Time series plot of the market yield on US Treasury Securities at 10-year constant maturity. *Source:* US Treasury.

series plot. The rate variable is collected at equally spaced time periods, as is typical in most time series and forecasting applications. Many business applications of forecasting utilize daily, weekly, monthly, quarterly, or annual data, but any reporting interval may be used. Furthermore, the data may be instantaneous, such as the viscosity of a chemical product at the point in time where it is measured; it may be cumulative, such as the total sales of a product during the month; or it may be a statistic that in some way reflects the activity of the variable during the time period, such as the daily closing price of a specific stock on the New York Stock Exchange.

The reason that forecasting is so important is that prediction of future events is a critical input into many types of planning and decision-making processes, with application to areas such as the following:

1. *Operations Management*. Business organizations routinely use forecasts of product sales or demand for services in order to schedule production, control inventories, manage the supply chain, determine staffing requirements, and plan capacity. Forecasts may also be used to determine the mix of products or services to be offered and the locations at which products are to be produced.
2. *Marketing*. Forecasting is important in many marketing decisions. Forecasts of sales response to advertising expenditures, new promotions, or changes in pricing polices enable businesses to evaluate their effectiveness, determine whether goals are being met, and make adjustments.
3. *Finance and Risk Management*. Investors in financial assets are interested in forecasting the returns from their investments. These assets include but are not limited to stocks, bonds, and commodities; other investment decisions can be made relative to forecasts of interest rates, options, and currency exchange rates. Financial risk management requires forecasts of the volatility of asset returns so that the risks associated with investment portfolios can be evaluated and insured, and so that financial derivatives can be properly priced.
4. *Economics*. Governments, financial institutions, and policy organizations require forecasts of major economic variables, such as gross domestic product, population growth, unemployment, interest rates, inflation, job growth, production, and consumption. These forecasts are an integral part of the guidance behind monetary and fiscal policy, and budgeting plans and decisions made by governments. They are also instrumental in the strategic planning decisions made by business organizations and financial institutions.

5. *Industrial Process Control*. Forecasts of the future values of critical quality characteristics of a production process can help determine when important controllable variables in the process should be changed, or if the process should be shut down and overhauled. Feedback and feedforward control schemes are widely used in monitoring and adjustment of industrial processes, and predictions of the process output are an integral part of these schemes.
6. *Demography*. Forecasts of population by country and regions are made routinely, often stratified by variables such as gender, age, and race. Demographers also forecast births, deaths, and migration patterns of populations. Governments use these forecasts for planning policy and social service actions, such as spending on health care, retirement programs, and antipoverty programs. Many businesses use forecasts of populations by age groups to make strategic plans regarding developing new product lines or the types of services that will be offered.

These are only a few of the many different situations where forecasts are required in order to make good decisions. Despite the wide range of problem situations that require forecasts, there are only two broad types of forecasting techniques—qualitative methods and quantitative methods.

Qualitative forecasting techniques are often subjective in nature and require judgment on the part of experts. Qualitative forecasts are often used in situations where there is little or no historical data on which to base the forecast. An example would be the introduction of a new product, for which there is no relevant history. In this situation, the company might use the expert opinion of sales and marketing personnel to subjectively estimate product sales during the new product introduction phase of its life cycle. Sometimes qualitative forecasting methods make use of marketing tests, surveys of potential customers, and experience with the sales performance of other products (both their own and those of competitors). However, although some data analysis may be performed, the basis of the forecast is subjective judgment.

Perhaps the most formal and widely known qualitative forecasting technique is the **Delphi Method**. This technique was developed by the RAND Corporation (see Dalkey [1967]). It employs a panel of experts who are assumed to be knowledgeable about the problem. The panel members are physically separated to avoid their deliberations being impacted either by social pressures or by a single dominant individual. Each panel member responds to a questionnaire containing a series of questions and returns the information to a coordinator. Following the first questionnaire, subsequent questions are submitted to the panelists along with information about the opinions of the panel as a group. This allows panelists to review their predictions relative to the opinions of the entire group. After several rounds, it is hoped that the opinions of the panelists converge to a consensus, although achieving a consensus is not required and justified differences of opinion can be included in the outcome. Qualitative forecasting methods are not emphasized in this book.

Quantitative forecasting techniques make formal use of historical data and a **forecasting model**. The model formally summarizes patterns in the data and expresses a statistical relationship between previous and current values of the variable. Then the model is used to project the patterns in the data into the future. In other words, the forecasting model is used to extrapolate past and current behavior into the future. There are several types of forecasting models in general use. The three most widely used are regression models, smoothing models, and general time series models. Regression models make use of relationships between the variable of interest and one or more related predictor variables. Sometimes regression models are called **causal forecasting models,** because the predictor variables are assumed to describe the forces that cause or drive the observed values of the variable of interest. An example would be using data on house purchases as a predictor variable to forecast furniture sales. The method of least squares is the formal basis of most regression models. **Smoothing models** typically employ a simple function of previous observations to provide a forecast of the variable of interest. These methods may have a formal statistical basis, but they are often used and justified heuristically on the basis that they are easy to use and produce satisfactory results. General **time series models** employ the statistical properties of the historical data to specify a formal model and then estimate the unknown parameters of this model (usually) by least squares. In subsequent chapters, we will discuss all three types of quantitative forecasting models.

The form of the forecast can be important. We typically think of a forecast as a single number that represents our best estimate of the future value of the variable of interest. Statisticians would call this a **point estimate** or **point forecast.** Now these forecasts are almost always wrong; that is, we experience **forecast error**. Consequently, it is usually a good practice to accompany a forecast with an estimate of how large a forecast error might be experienced. One way to do this is to provide a **prediction interval** (PI) to accompany the point forecast. The PI is a range of values for the future observation, and it is likely to prove far more useful in decision-making than a single number. We will show how to obtain PIs for most of the forecasting methods discussed in the book.

Other important features of the forecasting problem are the **forecast horizon** and the **forecast interval.** The forecast horizon is the number of future periods for which forecasts must be produced. The horizon is often dictated by the nature of the problem. For example, in production planning, forecasts of product demand may be made on a monthly basis. Because of the time required to change or modify a production schedule, ensure that sufficient raw material and component parts are available from the supply chain, and plan the delivery of completed goods to customers or inventory facilities, it would be necessary to forecast up to 3 months ahead. The forecast horizon is also often called the forecast **lead time.** The **forecast interval** is the frequency with which new forecasts are prepared. For example, in production planning, we might forecast demand on a monthly basis, for up to 3 months in the future (the lead time or horizon), and prepare a new forecast each month. Thus the forecast interval is 1 month, the same as the basic period of time for which each forecast is made. If the forecast lead time is always the same length, say, T periods, and the forecast is revised each time period, then we are employing a **rolling** or **moving horizon** forecasting approach. This system updates or revises the forecasts for $T-1$ of the periods in the horizon and computes a forecast for the newest period T. This rolling horizon approach to forecasting is widely used when the lead time is several periods long.

1.2 SOME EXAMPLES OF TIME SERIES

Time series plots can reveal **patterns** such as random, trends, level shifts, periods or cycles, unusual observations, or a combination of patterns. Patterns commonly found in time series data are discussed next with examples of situations that drive the patterns.

The sales of a mature pharmaceutical product may remain relatively flat in the absence of unchanged marketing or manufacturing strategies. Weekly sales of a generic pharmaceutical product shown in Figure 1.2 appear to be constant over time, at about $10{,}400 \times 10^3$ units, in a random sequence with no obvious patterns (data in Appendix B, Table B.2).

To assure conformance with customer requirements and product specifications, the production of chemicals is monitored by many characteristics. These may be input variables such as temperature and flow rate, and output properties such as viscosity and purity.

Due to the continuous nature of chemical manufacturing processes, output properties often are **positively autocorrelated;** that is, a value above the long-run average tends to be followed by other values above the

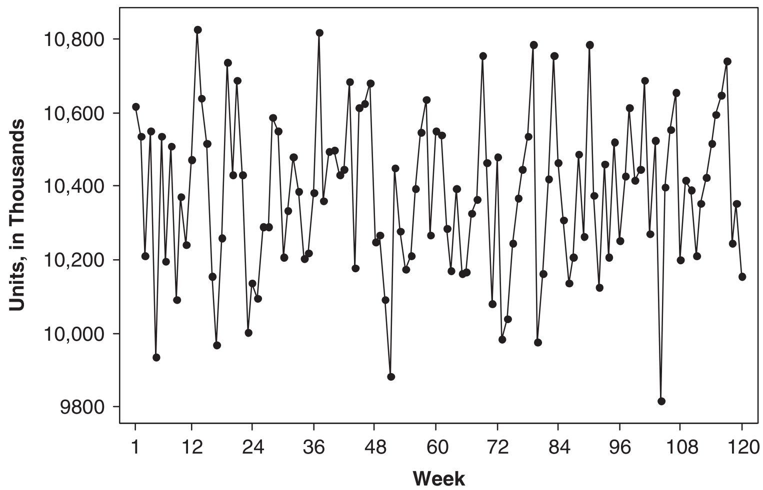


FIGURE 1.2 Pharmaceutical product sales.

average, while a value below the average tends to be followed by other values below the average.

The viscosity readings plotted in Figure 1.3 exhibit autocorrelated behavior, tending to a long-run average of about 85 centipoises (cP), but with a structured, not completely random, appearance (data in Appendix B, Table B.3). Some methods for describing and analyzing autocorrelated data will be described in Chapter 2.

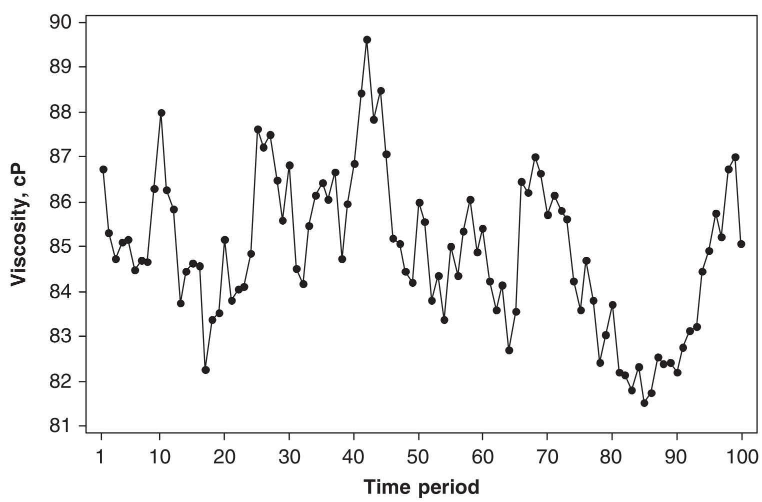


FIGURE 1.3 Chemical process viscosity readings.

The USDA National Agricultural Statistics Service publishes agricultural statistics for many commodities, including the annual production of dairy products such as butter, cheese, ice cream, milk, yogurt, and whey. These statistics are used for market analysis and intelligence, economic indicators, and identification of emerging issues.

Blue and gorgonzola cheese is one of 32 categories of cheese for which data are published. The annual US production of blue and gorgonzola cheeses (in 10^3 lb) is shown in Figure 1.4 (data in Appendix B, Table B.4). Production quadrupled from 1950 to 1997, and the **linear trend** has a constant positive slope with random, year-to-year variation.

The US Census Bureau publishes historic statistics on manufacturers' shipments, inventories, and orders. The statistics are based on North American Industry Classification System (NAICS) code and are utilized for purposes such as measuring productivity and analyzing relationships between employment and manufacturing output.

The manufacture of beverage and tobacco products is reported as part of the nondurable subsector. The plot of monthly beverage product shipments (Figure 1.5) reveals an overall increasing trend, with a distinct **cyclic pattern** that is repeated within each year. January shipments appear to be the lowest, with highs in May and June (data in Appendix B, Table B.5). This monthly, or **seasonal,** variation may be attributable to some cause

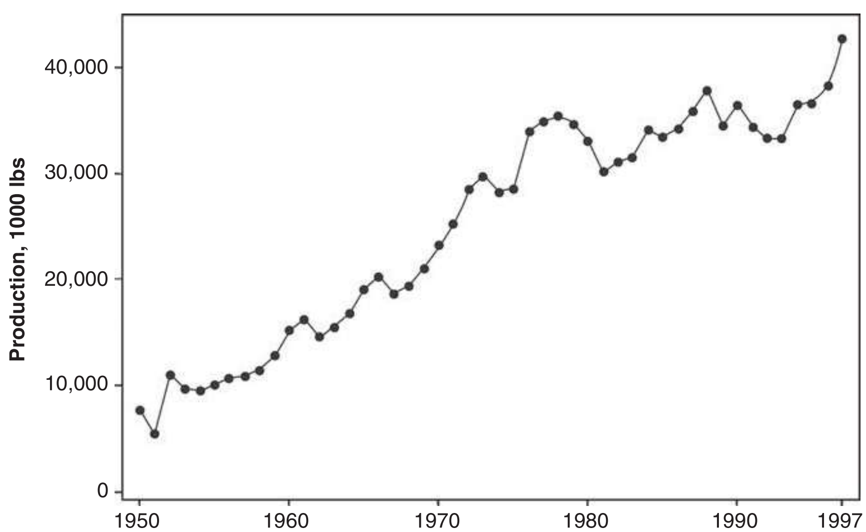


FIGURE 1.4 The US annual production of blue and gorgonzola cheeses. *Source:* USDA–NASS.

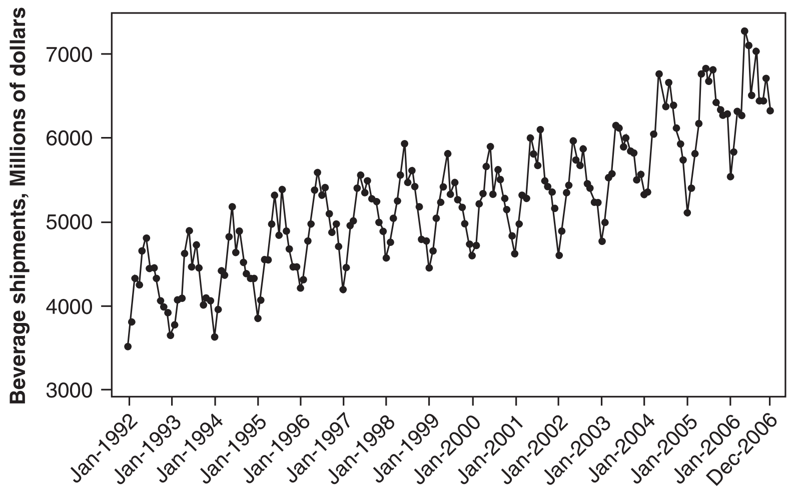


FIGURE 1.5 The US beverage manufacturer monthly product shipments, unadjusted. *Source:* US Census Bureau.

such as the impact of weather on the demand for beverages. Techniques for making seasonal adjustments to data in order to better understand general trends will be discussed in Chapter 2.

To determine whether the Earth is warming or cooling, scientists look at annual mean temperatures. At a single station, the warmest and the coolest temperatures in a day are averaged. Averages are then calculated at stations all over the Earth, over an entire year. The change in global annual mean surface air temperature is calculated from a base established from 1951 to 1980, and the result is reported as an "anomaly."

The plot of the annual mean anomaly in global surface air temperature (Figure 1.6) shows an increasing trend since 1880; however, the slope, or rate of change, varies with time periods (data in Appendix B, Table B.6). While the slope in earlier time periods appears to be constant, slightly increasing, or slightly decreasing, the slope from about 1975 to the present appears much steeper than the rest of the plot.

Business data such as stock prices and interest rates often exhibit **nonstationary** behavior; that is, the time series has no natural mean. The daily closing price adjusted for stock splits of Whole Foods Market (WFMI) stock in 2001 (Figure 1.7) exhibits a combination of patterns for both mean level and slope (data in Appendix B, Table B.7).

While the price is constant in some short time periods, there is no consistent mean level over time. In other time periods, the price changes

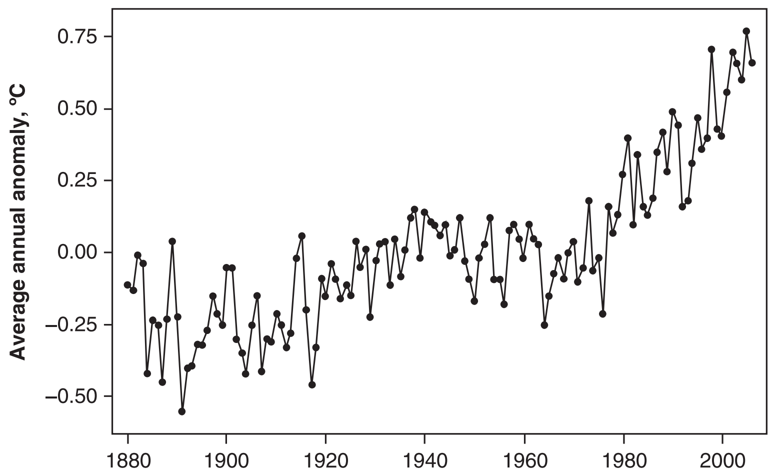


FIGURE 1.6 Global mean surface air temperature annual anomaly. *Source:* NASA-GISS.

at different rates, including occasional abrupt shifts in level. This is an example of nonstationary behavior, which will be discussed in Chapter 2.

The Current Population Survey (CPS) or "household survey" prepared by the US Department of Labor, Bureau of Labor Statistics, contains national data on employment, unemployment, earnings, and other labor market topics by demographic characteristics. The data are used to report

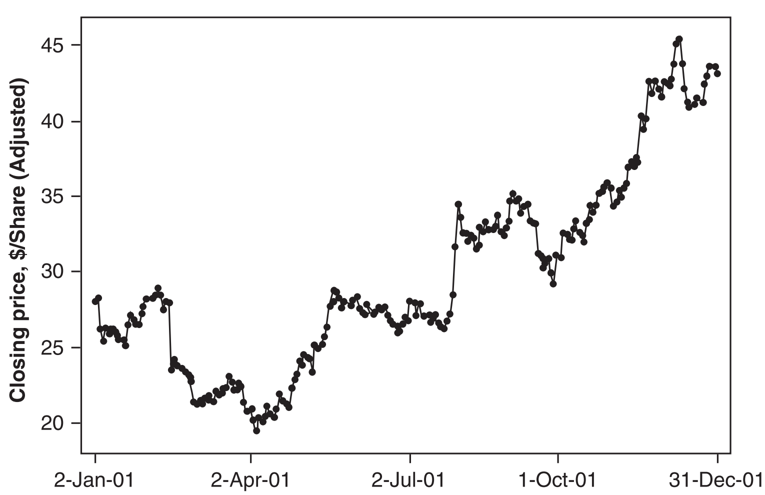


FIGURE 1.7 Whole foods market stock price, daily closing adjusted for splits.

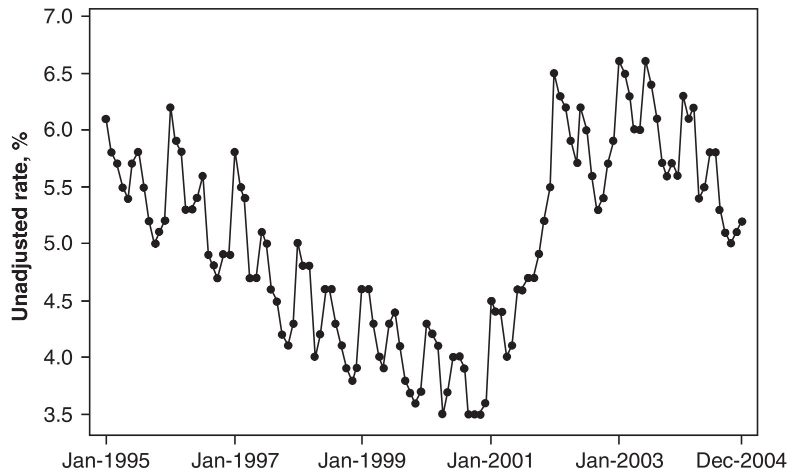


FIGURE 1.8 Monthly unemployment rate—full-time labor force, unadjusted. *Source:* US Department of Labor-BLS.

on the employment situation, for projections with impact on hiring and training, and for a multitude of other business planning activities. The data are reported unadjusted and with seasonal adjustment to remove the effect of regular patterns that occur each year.

The plot of monthly unadjusted unemployment rates (Figure 1.8) exhibits a mixture of patterns, similar to Figure 1.5 (data in Appendix B, Table B.8). There is a distinct cyclic pattern within a year; January, February, and March generally have the highest unemployment rates. The overall level is also changing, from a gradual decrease, to a steep increase, followed by a gradual decrease. The use of seasonal adjustments as described in Chapter 2 makes it easier to observe the nonseasonal movements in time series data.

Solar activity has long been recognized as a significant source of noise impacting consumer and military communications, including satellites, cell phone towers, and electric power grids. The ability to accurately forecast solar activity is critical to a variety of fields. The International Sunspot Number R is the oldest solar activity index. The number incorporates both the number of observed sunspots and the number of observed sunspot groups. In Figure 1.9, the plot of annual sunspot numbers reveals cyclic patterns of varying magnitudes (data in Appendix B, Table B.9).

In addition to assisting in the identification of steady-state patterns, time series plots may also draw attention to the occurrence of **atypical events.** Weekly sales of a generic pharmaceutical product dropped due to limited

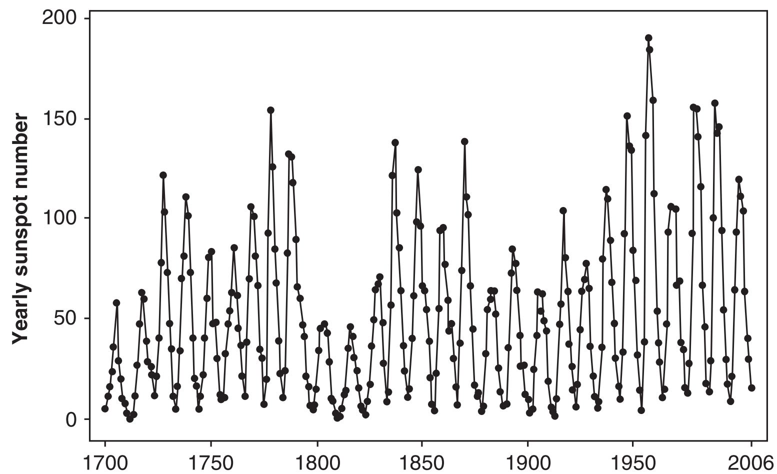


FIGURE 1.9 The international sunspot number. *Source:* SIDC.

availability resulting from a fire at one of the four production facilities. The 5-week reduction is apparent in the time series plot of weekly sales shown in Figure 1.10.

Another type of unusual event may be the failure of the data measurement or collection system. After recording a vastly different viscosity reading at time period 70 (Figure 1.11), the measurement system was

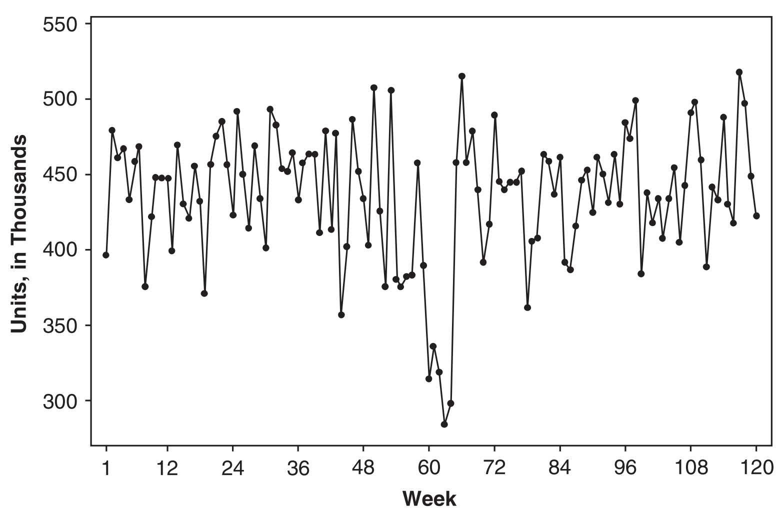


FIGURE 1.10 Pharmaceutical product sales.

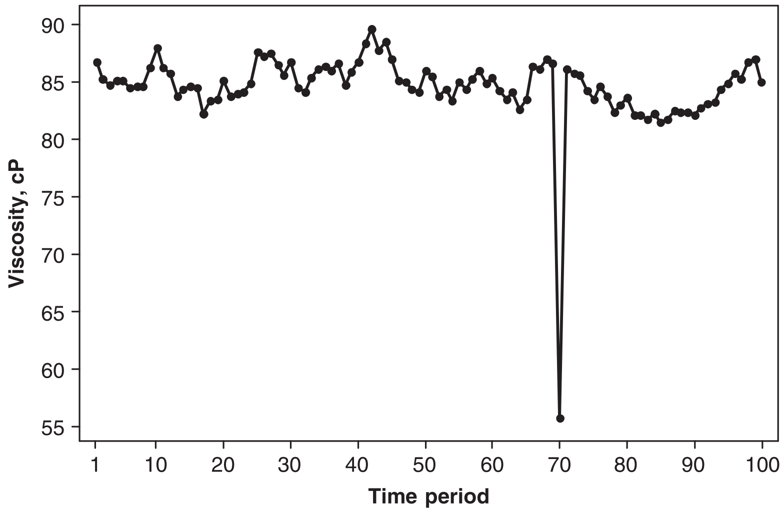


FIGURE 1.11 Chemical process viscosity readings, with sensor malfunction.

checked with a standard and determined to be out of calibration. The cause was determined to be a malfunctioning sensor.

1.3 THE FORECASTING PROCESS

A process is a series of connected activities that transform one or more inputs into one or more outputs. All work activities are performed in processes, and forecasting is no exception. The activities in the forecasting process are:

1. Problem definition
2. Data collection
3. Data analysis
4. Model selection and fitting
5. Model validation
6. Forecasting model deployment
7. Monitoring forecasting model performance

These activities are shown in Figure 1.12.

Problem definition involves developing understanding of how the forecast will be used along with the expectations of the "customer" (the user of

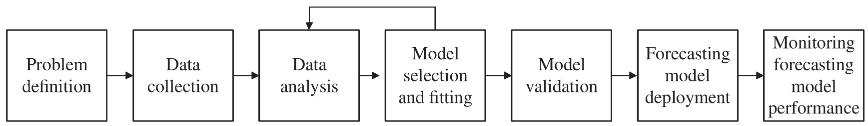


FIGURE 1.12 The forecasting process.

the forecast). Questions that must be addressed during this phase include the desired form of the forecast (e.g., are monthly forecasts required), the forecast horizon or lead time, how often the forecasts need to be revised (the forecast interval), and what level of forecast accuracy is required in order to make good business decisions. This is also an opportunity to introduce the decision makers to the use of prediction intervals as a measure of the risk associated with forecasts, if they are unfamiliar with this approach. Often it is necessary to go deeply into many aspects of the business system that requires the forecast to properly define the forecasting component of the entire problem. For example, in designing a forecasting system for inventory control, information may be required on issues such as product shelf life or other aging considerations, the time required to manufacture or otherwise obtain the products (production lead time), and the economic consequences of having too many or too few units of product available to meet customer demand. When multiple products are involved, the level of aggregation of the forecast (e.g., do we forecast individual products or families consisting of several similar products) can be an important consideration. Much of the ultimate success of the forecasting model in meeting the customer expectations is determined in the problem definition phase.

Data collection consists of obtaining the relevant history for the variable(s) that are to be forecast, including historical information on potential predictor variables.

The key here is "relevant"; often information collection and storage methods and systems change over time and not all historical data are useful for the current problem. Often it is necessary to deal with missing values of some variables, potential outliers, or other data-related problems that have occurred in the past. During this phase, it is also useful to begin planning how the data collection and storage issues in the future will be handled so that the reliability and integrity of the data will be preserved.

Data analysis is an important preliminary step to the selection of the forecasting model to be used. Time series plots of the data should be constructed and visually inspected for recognizable patterns, such as trends and seasonal or other cyclical components. A trend is evolutionary movement, either upward or downward, in the value of the variable. Trends may be long-term or more dynamic and of relatively short duration. Seasonality is the component of time series behavior that repeats on a regular basis, such as each year. Sometimes we will smooth the data to make identification of the patterns more obvious (data smoothing will be discussed in Chapter 2). Numerical summaries of the data, such as the sample mean, standard deviation, percentiles, and autocorrelations, should also be computed and evaluated. Chapter 2 will provide the necessary background to do this. If potential predictor variables are available, scatter plots of each pair of variables should be examined. Unusual data points or potential **outliers** should be identified and flagged for possible further study. The purpose of this preliminary data analysis is to obtain some “feel” for the data, and a sense of how strong the underlying patterns such as trend and seasonality are. This information will usually suggest the initial types of quantitative forecasting methods and models to explore.

Model selection and fitting consists of choosing one or more forecasting models and fitting the model to the data. **By fitting**, we mean estimating the unknown model parameters, usually by the method of least squares. In subsequent chapters, we will present several types of time series models and discuss the procedures of model fitting. We will also discuss methods for evaluating the quality of the model fit, and determining if any of the underlying assumptions have been violated. This will be useful in discriminating between different candidate models.

Model validation consists of an evaluation of the forecasting model to determine how it is likely to perform in the intended application. This must go beyond just evaluating the “fit” of the model to the historical data and must examine what magnitude of forecast errors will be experienced when the model is used to forecast “fresh” or new data. The fitting errors will always be smaller than the forecast errors, and this is an important concept that we will emphasize in this book. A widely used method for validating a forecasting model before it is turned over to the customer is to employ some form of **data splitting,** where the data are divided into two segments—a fitting segment and a forecasting segment. The model is fit to only the fitting data segment, and then forecasts from that model are simulated for the observations in the forecasting segment. This can provide useful guidance on how the forecasting model will perform when exposed to new data and can be a valuable approach for discriminating between competing forecasting models.

Forecasting model deployment involves getting the model and the resulting forecasts in use by the customer. It is important to ensure that the customer understands how to use the model and that generating timely forecasts from the model becomes as routine as possible. Model maintenance, including making sure that data sources and other required information will continue to be available to the customer is also an important issue that impacts the timeliness and ultimate usefulness of forecasts.

Monitoring forecasting model performance should be an ongoing activity after the model has been deployed to ensure that it is still performing satisfactorily. It is the nature of forecasting that conditions change over time, and a model that performed well in the past may deteriorate in performance. Usually performance deterioration will result in larger or more systematic forecast errors. Therefore monitoring of forecast errors is an essential part of good forecasting system design. **Control charts** of forecast errors are a simple but effective way to routinely monitor the performance of a forecasting model. We will illustrate approaches to monitoring forecast errors in subsequent chapters.

1.4 DATA FOR FORECASTING

1.4.1 The Data Warehouse

Developing time series models and using them for forecasting requires data on the variables of interest to decision-makers. The data are the raw materials for the modeling and forecasting process. The terms **data** and **information** are often used interchangeably, but we prefer to use the term data as that seems to reflect a more raw or original form, whereas we think of information as something that is extracted or synthesized from data. The output of a forecasting system could be thought of as information, and that output uses data as an input.

In most modern organizations data regarding sales, transactions, company financial and business performance, supplier performance, and customer activity and relations are stored in a repository known as a **data warehouse**. Sometimes this is a single data storage system; but as the volume of data handled by modern organizations grows rapidly, the data warehouse has become an integrated system comprised of components that are physically and often geographically distributed, such as cloud data storage. The data warehouse must be able to organize, manipulate, and integrate data from multiple sources and different organizational information systems. The basic functionality required includes data extraction, data transformation, and data loading. Data extraction refers to obtaining data from internal sources and from external sources such as third party vendors or government entities and financial service organizations. Once the data are extracted, the transformation stage involves applying rules to prevent duplication of records and dealing with problems such as missing information. Sometimes we refer to the transformation activities as **data**

cleaning. We will discuss some of the important data cleaning operations subsequently. Finally, the data are loaded into the data warehouse where they are available for modeling and analysis.

Data quality has several dimensions. Five important ones that have been described in the literature are accuracy, timeliness, completeness, representativeness, and consistency. Accuracy is probably the oldest dimension of data quality and refers to how close that data conform to its “real” values. Real values are alternative sources that can be used for verification purposes. For example, do sales records match payments to accounts receivable records (although the financial records may occur in later time periods because of payment terms and conditions, discounts, etc.)? Timeliness means that the data are as current as possible. Infrequent updating of data can seriously impact developing a time series model that is going to be used for relatively short-term forecasting. In many time series model applications the time between the occurrence of the real-world event and its entry into the data warehouse must be as short as possible to facilitate model development and use. Completeness means that the data content is complete, with no missing data and no outliers. As an example of representativeness, suppose that the end use of the time series model is to forecast customer demand for a product or service, but the organization only records booked orders and the date of fulfillment. This may not accurately reflect demand, because the orders can be booked before the desired delivery period and the date of fulfillment can take place in a different period than the one required by the customer. Furthermore, orders that are lost because of product unavailability or unsatisfactory delivery performance are not recorded. In these situations demand can differ dramatically from sales. Data cleaning methods can often be used to deal with some problems of completeness. Consistency refers to how closely data records agree over time in format, content, meaning, and structure. In many organizations how data are collected and stored evolves over time; definitions change and even the types of data that are collected change. For example, consider monthly data. Some organizations define “months” that coincide with the traditional calendar definition. But because months have different numbers of days that can induce patterns in monthly data, some organizations prefer to define a year as consisting of 13 “months” each consisting of 4 weeks.

It has been suggested that the output data that reside in the data warehouse are similar to the output of a manufacturing process, where the raw data are the input. Just as in manufacturing and other service processes, the data production process can benefit by the application of quality management and control tools. Jones-Farmer et al. (2014) describe how statistical quality control methods, specifically control charts, can be used to enhance data quality in the data production process.

1.4.2 Data Cleaning

Data cleaning is the process of examining data to detect potential errors, missing data, outliers or unusual values, or other inconsistencies and then correcting the errors or problems that are found. Sometimes errors are the result of recording or transmission problems, and can be corrected by working with the original data source to correct the problem. Effective data cleaning can greatly improve the forecasting process.

Before data are used to develop a time series model, it should be subjected to several different kinds of checks, including but not necessarily limited to the following:

1. Is there missing data?
2. Does the data fall within an expected range?
3. Are there potential outliers or other unusual values?

These types of checks can be automated fairly easily. If this aspect of data cleaning is automated, the rules employed should be periodically evaluated to ensure that they are still appropriate and that changes in the data have not made some of the procedures less effective. However, it is also extremely useful to use graphical displays to assist in identifying unusual data. Techniques such as time series plots, histograms, and scatter diagrams are extremely useful. These and other graphical methods will be described in Chapter 2.

1.4.3 Imputation

Data **imputation** is the process of correcting missing data or replacing outliers with an estimation process. Imputation replaces missing or erroneous values with a "likely" value based on other available information. This enables the analysis to work with statistical techniques which are designed to handle the complete data sets.

Mean value imputation consists of replacing a missing value with the sample average calculated from the nonmissing observations. The big advantage of this method is that it is easy, and if the data does not have any specific trend or seasonal pattern, it leaves the sample mean of the complete data set unchanged. However, one must be careful if there are trends or seasonal patterns, because the sample mean of all of the data may not reflect these patterns. A variation of this is **stochastic mean value imputation**, in which a random variable is added to the mean value to capture some of the noise or variability in the data. The random variable could be assumed to follow a normal distribution with mean zero and standard deviation equal to the standard deviation of the actual observed data. A variation of mean value imputation is to use a subset of the available historical data that reflects any trend or seasonal patterns in the data. For example, consider the time series $y_1, y_2, \dots, y_T$ and suppose that one observation y_j is missing. We can impute the missing value as

$$y_j^* = \frac{1}{2k}\left(\sum_{t=j-k}^{j-1} y_t + \sum_{t-j+1}^{j+k} y_t\right),$$

where k would be based on the seasonal variability in the data. It is usually chosen as some multiple of the smallest seasonal cycle in the data. So, if the data are monthly and exhibit a monthly cycle, k would be a multiple of 12. **Regression imputation** is a variation of mean value imputation where the imputed value is computed from a model used to predict the missing value. The prediction model does not have to be a linear regression model. For example, it could be a time series model.

Hot deck imputation is an old technique that is also known as the last value carried forward method. The term "hot deck" comes from the use of computer punch cards. The deck of cards was "hot" because it was currently in use. **Cold deck imputation** uses information from a deck of cards not currently in use. In hot deck imputation, the missing values are imputed by using values from similar complete observations. If there are several variables, sort the data by the variables that are most related to the missing observation and then, starting at the top, replace the missing values with the value of the immediately preceding variable. There are many variants of this procedure.

1.5 RESOURCES FOR FORECASTING

There are a variety of good resources that can be helpful to technical professionals involved in developing forecasting models and preparing forecasts. There are three professional journals devoted to forecasting:

- *Journal of Forecasting*
- *International Journal of Forecasting*
- *Journal of Business Forecasting Methods and Systems*

These journals publish a mixture of new methodology, studies devoted to the evaluation of current methods for forecasting, and case studies and applications. In addition to these specialized forecasting journals, there are several other mainstream statistics and operations research/management science journals that publish papers on forecasting, including:

- *Journal of Business and Economic Statistics*
- *Management Science*
- *Naval Research Logistics*
- *Operations Research*
- *International Journal of Production Research*
- *Journal of Applied Statistics*

This is by no means a comprehensive list. Research on forecasting tends to be published in a variety of outlets.

There are several books that are good complements to this one. We recommend Box, Jenkins, and Reinsel (1994); Chatfield (1996); Fuller (1995); Abraham and Ledolter (1983); Montgomery, Johnson, and Gardiner (1990); Wei (2006); and Brockwell and Davis (1991, 2002). Some of these books are more specialized than this one, in that they focus on a specific type of forecasting model such as the autoregressive integrated moving average [ARIMA] model, and some also require more background in statistics and mathematics.

Many statistics software packages have very good capability for fitting a variety of forecasting models. Minitab® Statistical Software, JMP®, the Statistical Analysis System (SAS) and R are the packages that we utilize and illustrate in this book. At the end of most chapters we provide R code for working some of the examples in the chapter. Matlab and S-Plus are also two packages that have excellent capability for solving forecasting problems.

EXERCISES

1.1 Why is forecasting an essential part of the operation of any organization or business?

1.2 What is a time series? Explain the meaning of trend effects, seasonal variations, and random error.

1.3 Explain the difference between a point forecast and an interval forecast.

1.4 What do we mean by a causal forecasting technique?

1.5 Everyone makes forecasts in their daily lives. Identify and discuss a situation where you employ forecasts.

a. What decisions are impacted by your forecasts?

b. How do you evaluate the quality of your forecasts?

c. What is the value to you of a good forecast?

d. What is the harm or penalty associated with a bad forecast?

1.6 What is meant by a rolling horizon forecast?

1.7 Explain the difference between forecast horizon and forecast interval.

1.8 Suppose that you are in charge of capacity planning for a large electric utility. A major part of your job is ensuring that the utility has sufficient generating capacity to meet current and future customer needs. If you do not have enough capacity, you run the risks of brownouts and service interruption. If you have too much capacity, it may cost more to generate electricity.

a. What forecasts do you need to do your job effectively?

b. Are these short-range or long-range forecasts?

c. What data do you need to be able to generate these forecasts?

1.9 Your company designs and manufactures apparel for the North American market. Clothing and apparel is a style good, with a relatively limited life. Items not sold at the end of the season are usually sold through off-season outlet and discount retailers. Items not sold through discounting and off-season merchants are often given to charity or sold abroad.

a. What forecasts do you need in this business to be successful?

b. Are these short-range or long-range forecasts?

c. What data do you need to be able to generate these forecasts?

d. What are the implications of forecast errors?

1.10 Suppose that you are in charge of production scheduling at a semiconductor manufacturing plant. The plant manufactures about 20 different types of devices, all on 8-inch silicon wafers. Demand for these products varies randomly. When a lot or batch of wafers is started into production, it can take from 4 to 6 weeks before the batch is finished, depending on the type of product. The routing of each batch of wafers through the production tools can be different depending on the type of product.

a. What forecasts do you need in this business to be successful?

b. Are these short-range or long-range forecasts?

c. What data do you need to be able to generate these forecasts?

d. Discuss the impact that forecast errors can potentially have on the efficiency with which your factory operates, including work-in-process inventory, meeting customer delivery schedules, and the cycle time to manufacture product.

1.11 You are the administrator of a large metropolitan hospital that operates the only 24-hour emergency room in the area. You must schedule attending physicians, resident physicians, nurses, laboratory, and support personnel to operate this facility effectively.

a. What measures of effectiveness do you think patients use to evaluate the services that you provide?

b. How are forecasts useful to you in planning services that will maximize these measures of effectiveness?

c. What planning horizon do you need to use? Does this lead to short-range or long-range forecasts?

1.12 Consider an airline that operates a network of flights that serves 200 cities in the continental United States. What long-range forecasts do the operators of the airline need to be successful? What forecasting problems does this business face on a daily basis? What are the consequences of forecast errors for the airline?

1.13 Discuss the potential difficulties of forecasting the daily closing price of a specific stock on the New York Stock Exchange. Would the problem be different (harder, easier) if you were asked to forecast the closing price of a group of stocks, all in the same industry (say, the pharmaceutical industry)?

1.14 Explain how large forecast errors can lead to high inventory levels at a retailer; at a manufacturing plant.

1.15 Your company manufactures and distributes soft drink beverages, sold in bottles and cans at retail outlets such as grocery stores, restaurants and other eating/drinking establishments, and vending machines in offices, schools, stores, and other outlets. Your product line includes about 25 different products, and many of these are produced in different package sizes.

a. What forecasts do you need in this business to be successful?

b. Is the demand for your product likely to be seasonal? Explain why or why not?

c. Does the shelf life of your product impact the forecasting problem?

d. What data do you think that you would need to be able to produce successful forecasts?

CHAPTER 2

STATISTICS BACKGROUND FOR FORECASTING

The future ain't what it used to be.

YOGI BERRA, *New York Yankees catcher*

2.1 INTRODUCTION

This chapter presents some basic statistical methods essential to modeling, analyzing, and forecasting time series data. Both graphical displays and numerical summaries of the properties of time series data are presented. We also discuss the use of data transformations and adjustments in forecasting and some widely used methods for characterizing and monitoring the performance of a forecasting model. Some aspects of how these performance measures can be used to select between competing forecasting techniques are also presented.

Forecasts are based on data or observations on the variable of interest. These data are usually in the form of a **time series**. Suppose that there are T periods of data available, with period T being the most recent. We will let the observation on this variable at time period t be denoted by y_t, $t = 1, 2, \dots, T$. This variable can represent a cumulative quantity, such as the

Introduction to Time Series Analysis and Forecasting, Second Edition.
Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci.

total demand for a product during period t, or an instantaneous quantity, such as the daily closing price of a specific stock on the New York Stock Exchange.

Generally, we will need to distinguish between a **forecast** or **predicted value** of y_t that was made at some previous time period, say, $t - \tau$, and a **fitted value** of y_t that has resulted from estimating the parameters in a time series model to historical data. Note that τ is the forecast lead time. The forecast made at time period $t - \tau$ is denoted by $\hat{y}_t(t - \tau)$. There is a lot of interest in the **lead − 1** forecast, which is the forecast of the observation in period t, y_t, made one period prior, $\hat{y}_t(t - 1)$. We will denote the fitted value of y_t by $\hat{y}_t$.

We will also be interested in analyzing **forecast errors**. The forecast error that results from a forecast of y_t that was made at time period $t - \tau$ is the **lead − τ forecast error**

$$e_t(\tau) = y_t - \hat{y}_t(t - \tau). \tag{2.1}$$

For example, the lead − 1 forecast error is

$$e_t(1) = y_t - \hat{y}_t(t - 1).$$

The difference between the observation y_t and the value obtained by fitting a time series model to the data, or a fitted value $\hat{y}_t$ defined earlier, is called a **residual**, and is denoted by

$$e_t = y_t - \hat{y}_t. \tag{2.2}$$

The reason for this careful distinction between forecast errors and residuals is that models usually fit historical data better than they forecast. That is, the residuals from a model-fitting process will almost always be smaller than the forecast errors that are experienced when that model is used to forecast future observations.

2.2 GRAPHICAL DISPLAYS

2.2.1 Time Series Plots

Developing a forecasting model should always begin with graphical display and analysis of the available data. Many of the broad general features of a time series can be seen visually. This is not to say that analytical tools are not useful, because they are, but the human eye can be a very sophisticated data analysis tool. To paraphrase the great New York Yankees catcher Yogi Berra, "You can observe a lot just by watching."

The basic graphical display for time series data is the **time series plot**, illustrated in Chapter 1. This is just a graph of y_t versus the time period, t, for $t = 1, 2, \ldots, T$. Features such as trend and seasonality are usually easy to see from the time series plot. It is interesting to observe that some of the classical tools of descriptive statistics, such as the histogram and the stem-and-leaf display, are not particularly useful for time series data because they do not take time order into account.

Example 2.1 Figures 2.1 and 2.2 show time series plots for viscosity readings and beverage production shipments (originally shown in Figures 1.3 and 1.5, respectively). At the right-hand side of each time series plot is a histogram of the data. Note that while the two time series display very different characteristics, the histograms are remarkably similar. Essentially, the histogram summarizes the data across the time dimension, and in so doing, the key time-dependent features of the data are lost. Stem-and-leaf plots and boxplots would have the same issues, losing time-dependent features.

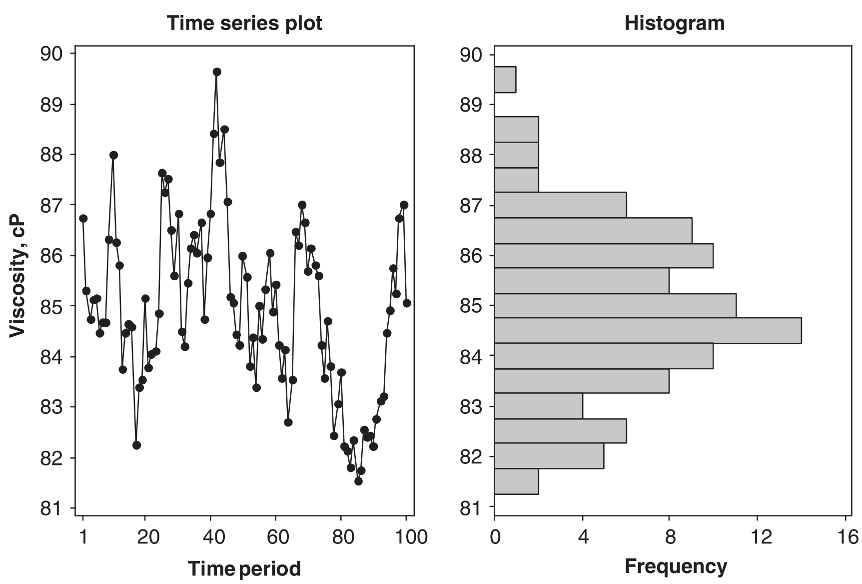


FIGURE 2.1 Time series plot and histogram of chemical process viscosity readings.

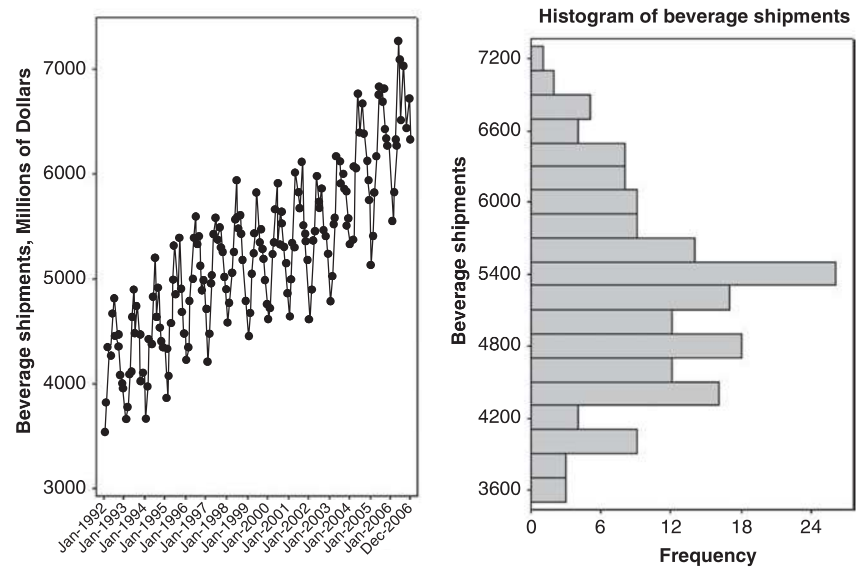


FIGURE 2.2 Time series plot and histogram of beverage production shipments.

When there are two or more variables of interest, **scatter plots** can be useful in displaying the relationship between the variables. For example, Figure 2.3 is a scatter plot of the annual global mean surface air temperature anomaly first shown in Figure 1.6 versus atmospheric CO_2 concentrations. The scatter plot clearly reveals a relationship between the two variables:

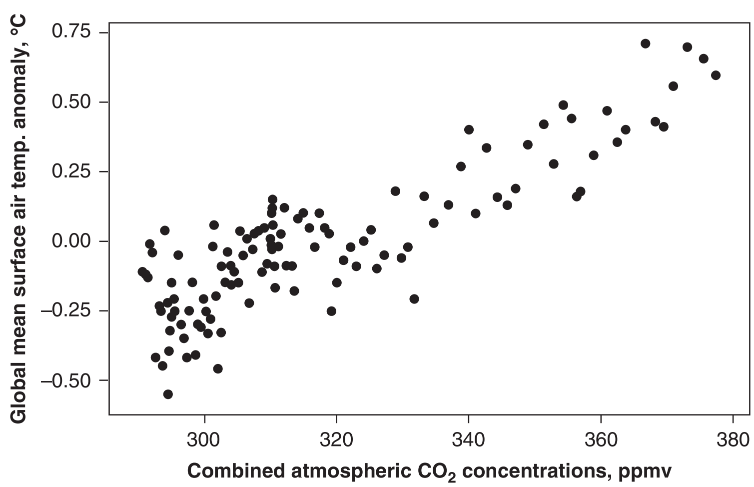


FIGURE 2.3 Scatter plot of temperature anomaly versus CO_2 concentrations. *Sources*: NASA–GISS (anomaly), DOE–DIAC (CO_2).

low concentrations of CO_2 are usually accompanied by negative anomalies, and higher concentrations of CO_2 tend to be accompanied by positive anomalies. Note that this does not imply that higher concentrations of CO_2 actually *cause* higher temperatures. The scatter plot cannot establish a causal relationship between two variables (neither can naive statistical modeling techniques, such as regression), but it is useful in displaying how the variables have varied together in the historical data set.

There are many variations of the time series plot and other graphical displays that can be constructed to show specific features of a time series. For example, Figure 2.4 displays daily price information for Whole Foods Market stock during the first quarter of 2001 (the trading days from January 2, 2001 through March 30, 2001). This chart, created in Excel®, shows the opening, closing, highest, and lowest prices experienced within a trading day for the first quarter. If the opening price was higher than the closing price, the box is filled, whereas if the closing price was higher than the opening price, the box is open. This type of plot is potentially more useful than a time series plot of just the closing (or opening) prices, because it shows the volatility of the stock within a trading day. The volatility of an asset is often of interest to investors because it is a measure of the inherent risk associated with the asset.

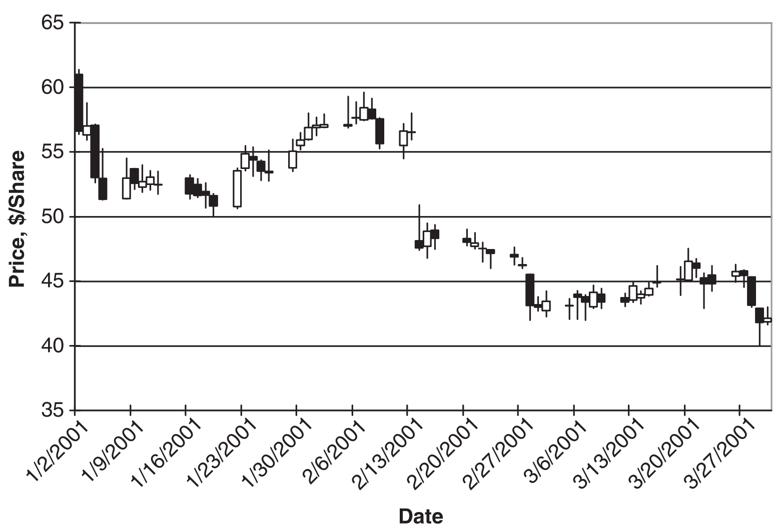


FIGURE 2.4 Open-high/close-low chart of Whole Foods Market stock price. *Source*: `finance.yahoo.com`.

2.2.2 Plotting Smoothed Data

Sometimes it is useful to overlay a **smoothed** version of the original data on the original time series plot to help reveal patterns in the original data. There are several types of data smoothers that can be employed. One of the simplest and most widely used is the ordinary or simple moving average. A simple **moving average** of span N assigns weights $1/N$ to the most recent N observations $y_T, y_{T-1}, \dots, y_{T-N+1}$, and weight zero to all other observations. If we let M_T be the moving average, then the N-span moving average at time period T is

$$M_T = \frac{y_T + y_{T-1} + \cdots + y_{T-N+1}}{N} = \frac{1}{N} \sum_{t=T-N+1}^{T} y_t \tag{2.3}$$

Clearly, as each new observation becomes available it is added into the sum from which the moving average is computed and the oldest observation is discarded. The moving average has less variability than the original observations; in fact, if the variance of an individual observation y_t is σ^2, then assuming that the observations are uncorrelated the variance of the moving average is

$$\mathrm{Var}(M_T) = \mathrm{Var}\left(\frac{1}{N} \sum_{t=T-N+1}^{N} y_t \right) = \frac{1}{N^2} \sum_{t=T-N+1}^{N} \mathrm{Var}(y_t) = \frac{\sigma^2}{N}$$

Sometimes a "centered" version of the moving average is used, such as in

$$M_t = \frac{1}{S+1} \sum_{i=-S}^{S} y_{t-i} \tag{2.4}$$

where the span of the centered moving average is $N = 2S + 1$.

Example 2.2 Figure 2.5 plots the annual global mean surface air temperature anomaly data along with a five-period (a period is 1 year) moving average of the same data. Note that the moving average exhibits less variability than found in the original series. It also makes some features of the data easier to see; for example, it is now more obvious that the global air temperature decreased from about 1940 until about 1975.

Plots of moving averages are also used by analysts to evaluate stock price trends; common MA periods are 5, 10, 20, 50, 100, and 200 days. A time series plot of Whole Foods Market stock price with a 50-day moving

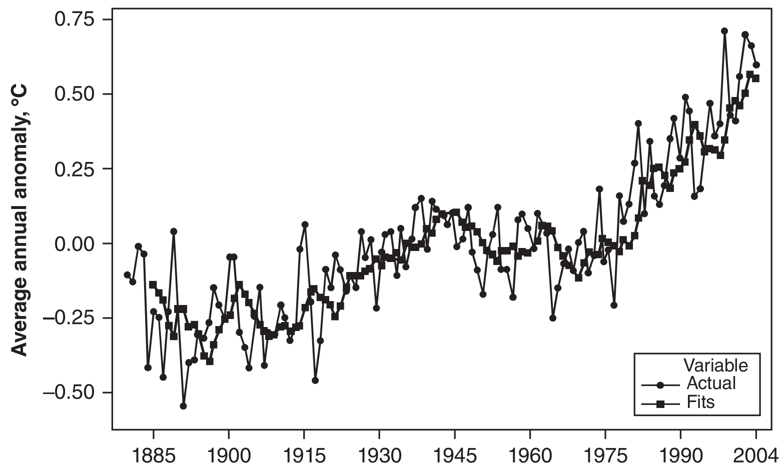


FIGURE 2.5 Time series plot of global mean surface air temperature anomaly, with five-period moving average. *Source*: NASA–GISS.

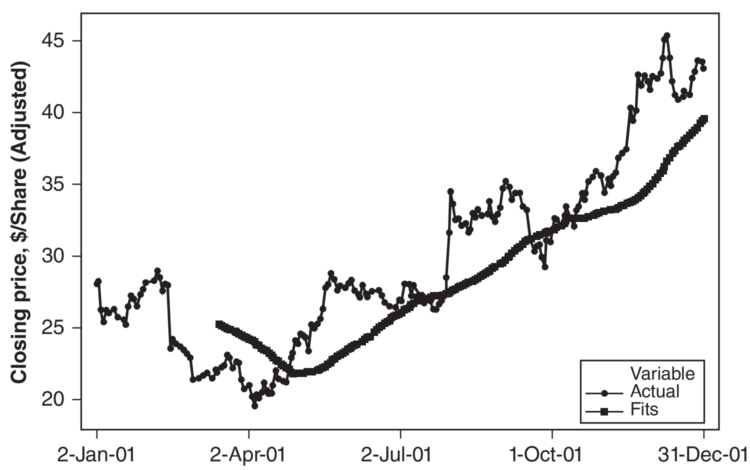


FIGURE 2.6 Time series plot of Whole Foods Market stock price, with 50-day moving average. *Source*: `finance.yahoo.com`.

average is shown in Figure 2.6. The moving average plot smoothes the day-to-day noise and shows a generally increasing trend.

The simple moving average is a **linear data smoother**, or a **linear filter**, because it replaces each observation y_t with a linear combination of the other data points that are near to it in time. The weights in the linear combination are equal, so the linear combination here is an average. Of course, unequal weights could be used. For example, the **Hanning filter** is a weighted, centered moving average

$$M_t^{\mathrm{H}} = 0.25y_{t+1} + 0.5y_t + 0.25y_{t-1}$$

Julius von Hann, a nineteenth century Austrian meteorologist, used this filter to smooth weather data.

An obvious disadvantage of a linear filter such as a moving average is that an unusual or erroneous data point or an outlier will dominate the moving averages that contain that observation, contaminating the moving averages for a length of time equal to the span of the filter. For example, consider the sequence of observations

$$15, 18, 13, 12, 16, 14, 16, 17, 18, 15, 18, 200, 19, 14, 21, 24, 19, 25$$

which increases reasonably steadily from 15 to 25, except for the unusual value 200. Any reasonable smoothed version of the data should also increase steadily from 15 to 25 and not emphasize the value 200. Now even if the value 200 is a legitimate observation, and not the result of a data recording or reporting error (perhaps it should be 20!), it is so unusual that it deserves special attention and should likely not be analyzed along with the rest of the data.

Odd-span **moving medians** (also called **running medians**) are an alternative to moving averages that are effective data smoothers when the time series may be contaminated with unusual values or outliers. The moving median of span N is defined as

$$m_t^{[N]} = med(y_{t-u}, \ldots, y_t, \ldots, y_{t+u}), \tag{2.5}$$

where $N = 2u + 1$. The median is the middle observation in rank order (or order of value). The moving median of span 3 is a very popular and effective data smoother, where

$$m_t^{[3]} = med(y_{t-1}, y_t, y_{t+1}).$$

This smoother would process the data three values at a time, and replace the three original observations by their median. If we apply this smoother to the data above, we obtain

$$\_\_\_, 15, 13, 13, 14, 16, 17, 17, 18, 18, 19, 19, 19, 21, 21, 24, \_\_\_.$$

This smoothed data are a reasonable representation of the original data, but they conveniently ignore the value 200. The end values are lost when using the moving median, and they are represented by "____".

In general, a moving median will pass monotone sequences of data unchanged. It will follow a step function in the data, but it will eliminate a spike or more persistent upset in the data that has duration of at most *u* consecutive observations. Moving medians can be applied more than once if desired to obtain an even smoother series of observations. For example, applying the moving median of span 3 to the smoothed data above results in

$$\_\_\_\_, \_\_\_\_, 13, 13, 14, 16, 17, 17, 18, 18, 19, 19, 19, 21, 21, \_\_\_\_, \_\_\_\_.$$

These data are now as smooth as it can get; that is, repeated application of the moving median will not change the data, apart from the end values.

If there are a lot of observations, the information loss from the missing end values is not serious. However, if it is necessary or desirable to keep the lengths of the original and smoothed data sets the same, a simple way to do this is to "copy on" or add back the end values from the original data. This would result in the smoothed data:

$$15, 18, 13, 13, 14, 16, 17, 17, 18, 18, 19, 19, 19, 21, 21, 19, 25$$

There are also methods for smoothing the end values. Tukey (1979) is a basic reference on this subject and contains many other clever and useful techniques for data analysis.

Example 2.3 The chemical process viscosity readings shown in Figure 1.11 are an example of a time series that benefits from smoothing to evaluate patterns. The selection of a moving median over a moving average, as shown in Figure 2.7, minimizes the impact of the invalid measurements, such as the one at time period 70.

2.3 NUMERICAL DESCRIPTION OF TIME SERIES DATA

2.3.1 Stationary Time Series

A very important type of time series is a **stationary** time series. A time series is said to be **strictly stationary** if its properties are not affected

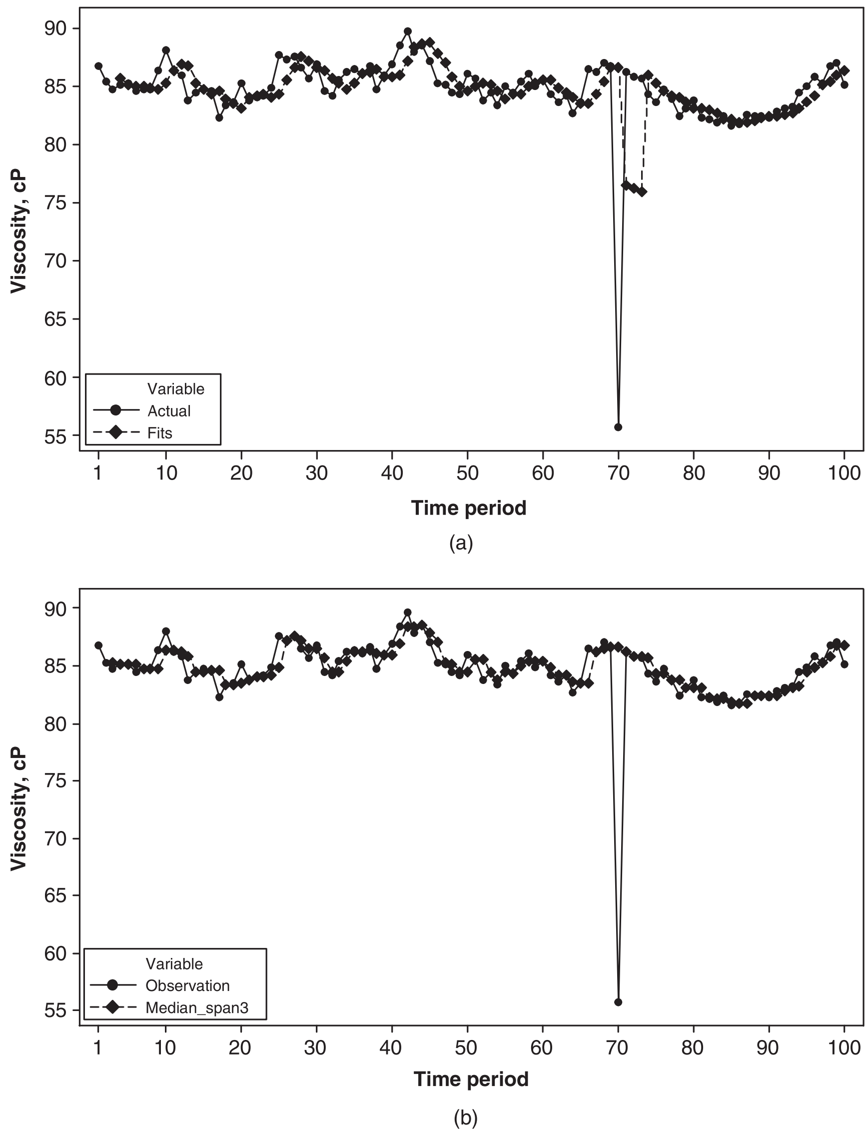


FIGURE 2.7 Viscosity readings with (a) moving average and (b) moving median.

by a change in the time origin. That is, if the joint probability distribution of the observations $y_t, y_{t+1}, \ldots, y_{t+n}$ is exactly the same as the joint probability distribution of the observations $y_{t+k}, y_{t+k+1}, \ldots, y_{t+k+n}$ then the time series is strictly stationary. When $n = 0$ the stationarity assumption means that the probability distribution of y_t is the same for all time periods

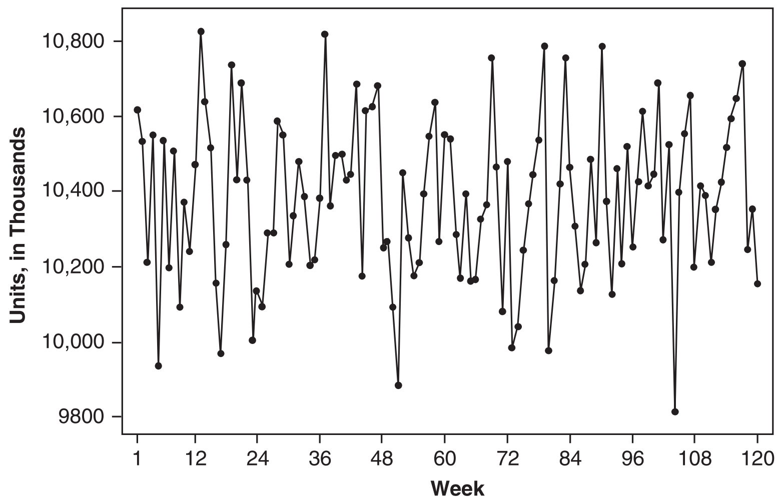


FIGURE 2.8 Pharmaceutical product sales.

and can be written as $f(y)$. The pharmaceutical product sales and chemical viscosity readings time series data originally shown in Figures 1.2 and 1.3, respectively, are examples of stationary time series. The time series plots are repeated in Figures 2.8 and 2.9 for convenience. Note that both time series seem to vary around a fixed level. Based on the earlier definition, this is a characteristic of stationary time series. On the other hand, the Whole

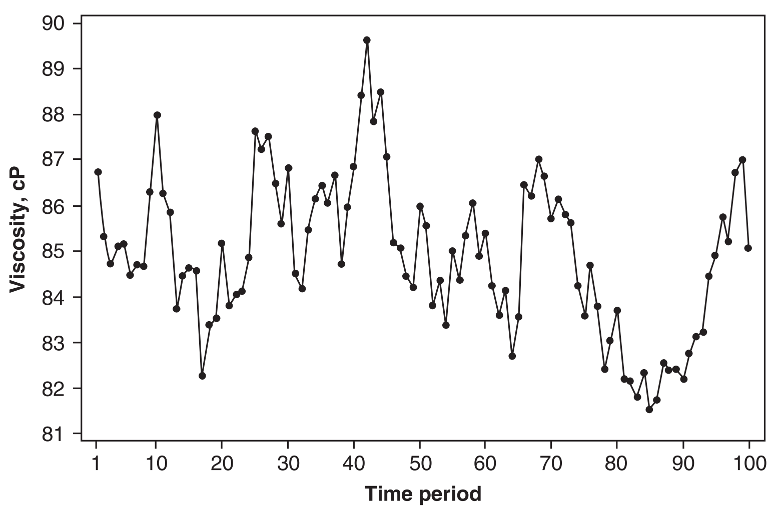


FIGURE 2.9 Chemical process viscosity readings.

Foods Market stock price data in Figure 1.7 tends to wander around or drift, with no obvious fixed level. This is behavior typical of a nonstationary time series.

Stationary implies a type of statistical **equilibrium** or **stability** in the data. Consequently, the time series has a constant mean defined in the usual way as

$$\mu_y = E(y) = \int_{-\infty}^{\infty} y f(y) dy \tag{2.6}$$

and constant variance defined as

$$\sigma_y^2 = \text{Var}(y) = \int_{-\infty}^{\infty} (y - \mu_y)^2 f(y) dy. \tag{2.7}$$

The sample mean and sample variance are used to estimate these parameters. If the observations in the time series are $y_1, y_2, \ldots, y_T$, then the sample mean is

$$\bar{y} = \hat{\mu}_y = \frac{1}{T} \sum_{t=1}^{T} y_t \tag{2.8}$$

and the sample variance is

$$s^2 = \hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})^2. \tag{2.9}$$

Note that the divisor in Eq. (2.9) is T rather than the more familiar $T - 1$. This is the common convention in many time series applications, and because T is usually not small, there will be little difference between using T instead of $T - 1$.

2.3.2 Autocovariance and Autocorrelation Functions

If a time series is stationary this means that the joint probability distribution of any two observations, say, y_t and y_{t+k}, is the same for any two time periods t and $t + k$ that are separated by the same interval k. Useful information about this joint distribution, and hence about the nature of the time series, can be obtained by plotting a scatter diagram of all of the data pairs y_t, y_{t+k} that are separated by the same interval k. The interval k is called the **lag**.

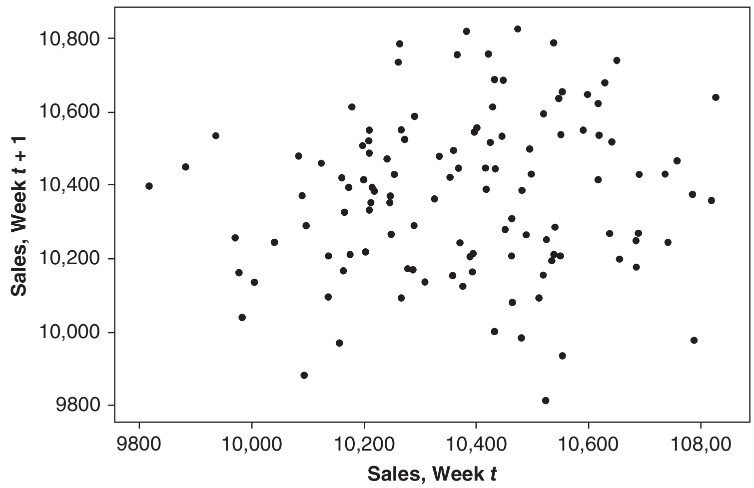


FIGURE 2.10 Scatter diagram of pharmaceutical product sales at lag $k = 1$.

Example 2.4 Figure 2.10 is a scatter diagram for the pharmaceutical product sales for lag $k = 1$ and Figure 2.11 is a scatter diagram for the chemical viscosity readings for lag $k = 1$. Both scatter diagrams were constructed by plotting y_{t+1} versus y_t. Figure 2.10 exhibits little structure; the plotted pairs of adjacent observations y_t, y_{t+1} seem to be **uncorrelated**. That is, the value of y in the current period does not provide any useful information about the value of y that will be observed in the next period. A different story is revealed in Figure 2.11, where we observe that the

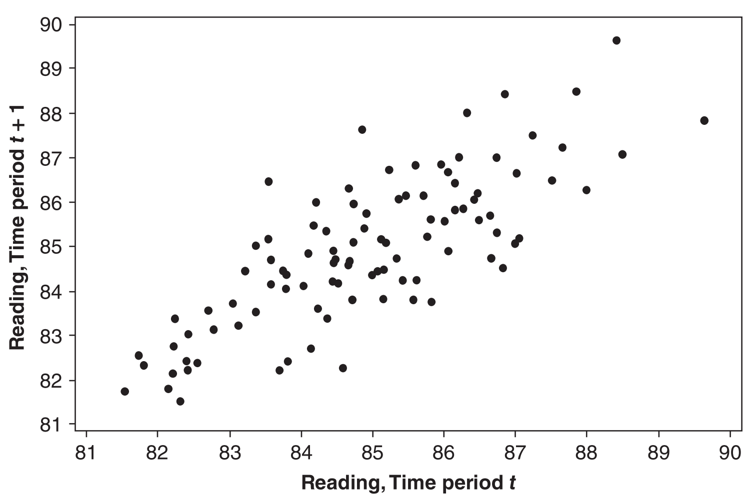


FIGURE 2.11 Scatter diagram of chemical viscosity readings at lag $k = 1$.

pairs of adjacent observations y_{t+1}, y_t are **positively correlated**. That is, a small value of y tends to be followed in the next time period by another small value of y, and a large value of y tends to be followed immediately by another large value of y. Note from inspection of Figures 2.10 and 2.11 that the behavior inferred from inspection of the scatter diagrams is reflected in the observed time series.

The covariance between y_t and its value at another time period, say, y_{t+k} is called the **autocovariance** at lag k, defined by

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]. \tag{2.10}$$

The collection of the values of $\gamma_k, k = 0, 1, 2, \ldots$ is called the **autocovariance function**. Note that the autocovariance at lag $k = 0$ is just the variance of the time series; that is, $\gamma_0 = \sigma_y^2$,which is constant for a stationary time series. The **autocorrelation coefficient** at lag k for a stationary time series is

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}} = \frac{\text{Cov}(y_t, y_{t+k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}. \tag{2.11}$$

The collection of the values of $\rho_k,\ k = 0, 1, 2, \ldots$ is called the **autocorrelation function (ACF)**. Note that by definition $\rho_0 = 1$. Also, the ACF is independent of the scale of measurement of the time series, so it is a dimensionless quantity. Furthermore, $\rho_k = \rho_{-k}$; that is, the ACF is **symmetric** around zero, so it is only necessary to compute the positive (or negative) half.

If a time series has a finite mean and autocovariance function it is said to be second-order stationary (or weakly stationary of order 2). If, in addition, the joint probability distribution of the observations at all times is multivariate normal, then that would be sufficient to result in a time series that is strictly stationary.

It is necessary to estimate the autocovariance and ACFs from a time series of finite length, say, $y_1, y_2, \ldots, y_T$. The usual estimate of the autocovariance function is

$$c_k = \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, 2, \ldots, K \tag{2.12}$$

and the ACF is estimated by the **sample autocorrelation function** (or **sample ACF**)

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, \ldots, K \tag{2.13}$$

A good general rule of thumb is that at least 50 observations are required to give a reliable estimate of the ACF, and the individual sample autocorrelations should be calculated up to lag K, where K is about $T/4$.

Often we will need to determine if the autocorrelation coefficient at a particular lag is zero. This can be done by comparing the sample autocorrelation coefficient at lag k, r_k, to its standard error. If we make the assumption that the observations are uncorrelated, that is, $\rho_k = 0$ for all k, then the variance of the sample autocorrelation coefficient is

$$\mathrm{Var}(r_k) \cong \frac{1}{T} \tag{2.14}$$

and the standard error is

$$se(r_k) \cong \frac{1}{\sqrt{T}} \tag{2.15}$$

Example 2.5 Consider the chemical process viscosity readings plotted in Figure 2.9; the values are listed in Table 2.1.

The sample ACF at lag $k = 1$ is calculated as

$$\begin{aligned}
c_0 &= \frac{1}{100} \sum_{t=1}^{100-0} (y_t - \bar{y})(y_{t+0} - \bar{y}) \\
&= \frac{1}{100}[(86.7418 - 84.9153)(86.7418 - 84.9153) + \cdots \\
&\quad + (85.0572 - 84.9153)(85.0572 - 84.9153)] \\
&= 280.9332 \\
c_1 &= \frac{1}{100} \sum_{t=1}^{100-1} (y_t - \bar{y})(y_{t+1} - \bar{y}) \\
&= \frac{1}{100}[(86.7418 - 84.9153)(85.3195 - 84.9153) + \cdots \\
&\quad + (87.0048 - 84.9153)(85.0572 - 84.9153)] \\
&= 220.3137 \\
r_1 &= \frac{c_1}{c_0} = \frac{220.3137}{280.9332} = 0.7842
\end{aligned}$$

A plot and listing of the sample ACFs generated by Minitab for the first 25 lags are displayed in Figures 2.12 and 2.13, respectively.

TABLE 2.1 Chemical Process Viscosity Readings

Time Period	Reading	Time Period	Reading	Time Period	Reading	Time Period	Reading
1	86.7418	26	87.2397	51	85.5722	76	84.7052
2	85.3195	27	87.5219	52	83.7935	77	83.8168
3	84.7355	28	86.4992	53	84.3706	78	82.4171
4	85.1113	29	85.6050	54	83.3762	79	83.0420
5	85.1487	30	86.8293	55	84.9975	80	83.6993
6	84.4775	31	84.5004	56	84.3495	81	82.2033
7	84.6827	32	84.1844	57	85.3395	82	82.1413
8	84.6757	33	85.4563	58	86.0503	83	81.7961
9	86.3169	34	86.1511	59	84.8839	84	82.3241
10	88.0006	35	86.4142	60	85.4176	85	81.5316
11	86.2597	36	86.0498	61	84.2309	86	81.7280
12	85.8286	37	86.6642	62	83.5761	87	82.5375
13	83.7500	38	84.7289	63	84.1343	88	82.3877
14	84.4628	39	85.9523	64	82.6974	89	82.4159
15	84.6476	40	86.8473	65	83.5454	90	82.2102
16	84.5751	41	88.4250	66	86.4714	91	82.7673
17	82.2473	42	89.6481	67	86.2143	92	83.1234
18	83.3774	43	87.8566	68	87.0215	93	83.2203
19	83.5385	44	88.4997	69	86.6504	94	84.4510
20	85.1620	45	87.0622	70	85.7082	95	84.9145
21	83.7881	46	85.1973	71	86.1504	96	85.7609
22	84.0421	47	85.0767	72	85.8032	97	85.2302
23	84.1023	48	84.4362	73	85.6197	98	86.7312
24	84.8495	49	84.2112	74	84.2339	99	87.0048
25	87.6416	50	85.9952	75	83.5737	100	85.0572

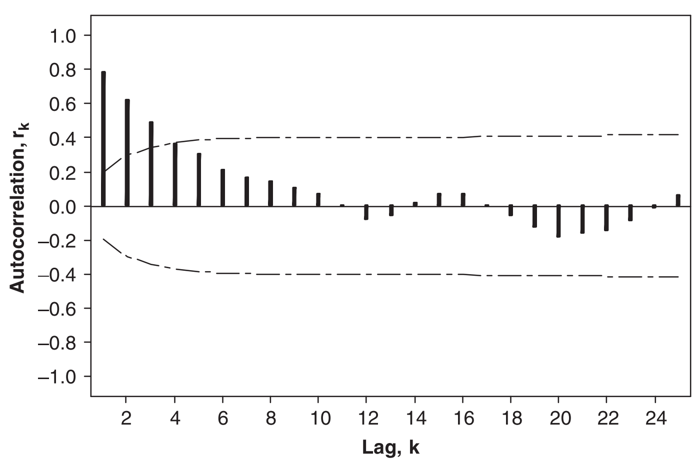


FIGURE 2.12 Sample autocorrelation function for chemical viscosity readings, with 5% significance limits.

Autocorrelation function: reading

Lag	ACF	T	LBQ
1	0.784221	7.84	63.36
2	0.628050	4.21	104.42
3	0.491587	2.83	129.83
4	0.362880	1.94	143.82
5	0.304554	1.57	153.78
6	0.208979	1.05	158.52
7	0.164320	0.82	161.48
8	0.144789	0.72	163.80
9	0.103625	0.51	165.01
10	0.066559	0.33	165.51
11	0.003949	0.02	165.51
12	–0.077226	–0.38	166.20
13	–0.051953	–0.25	166.52
14	0.020525	0.10	166.57
15	0.072784	0.36	167.21
16	0.070753	0.35	167.81
17	0.001334	0.01	167.81
18	–0.057435	–0.28	168.22
19	–0.123122	–0.60	170.13
20	–0.180546	–0.88	174.29
21	–0.162466	–0.78	177.70
22	–0.145979	–0.70	180.48
23	–0.087420	–0.42	181.50
24	–0.011579	–0.06	181.51
25	0.063170	0.30	182.06

FIGURE 2.13 Listing of sample autocorrelation functions for first 25 lags of chemical viscosity readings, Minitab session window output (the definition of T and LBQ will be given later).

Note the rate of decrease or decay in ACF values in Figure 2.12 from 0.78 to 0, followed by a sinusoidal pattern about 0. This ACF pattern is typical of stationary time series. The importance of ACF estimates exceeding the 5% significance limits will be discussed in Chapter 5. In contrast, the plot of sample ACFs for a time series of random values with constant mean has a much different appearance. The sample ACFs for pharmaceutical product sales plotted in Figure 2.14 appear randomly positive or negative, with values near zero.

While the ACF is strictly speaking defined only for a stationary time series, the sample ACF can be computed for *any* time series, so a logical question is: What does the sample ACF of a nonstationary time series look like? Consider the daily closing price for Whole Foods Market stock in Figure 1.7. The sample ACF of this time series is shown in Figure 2.15. Note that this sample ACF plot behaves quite differently than the ACF plots in Figures 2.12 and 2.14. Instead of cutting off or tailing off near zero after a few lags, this sample ACF is very **persistent**; that is, it decays very slowly and exhibits sample autocorrelations that are still rather large even at long lags. This behavior is characteristic of a nonstationary time series. Generally, if the sample ACF does not dampen out within about 15 to 20 lags, the time series is nonstationary.

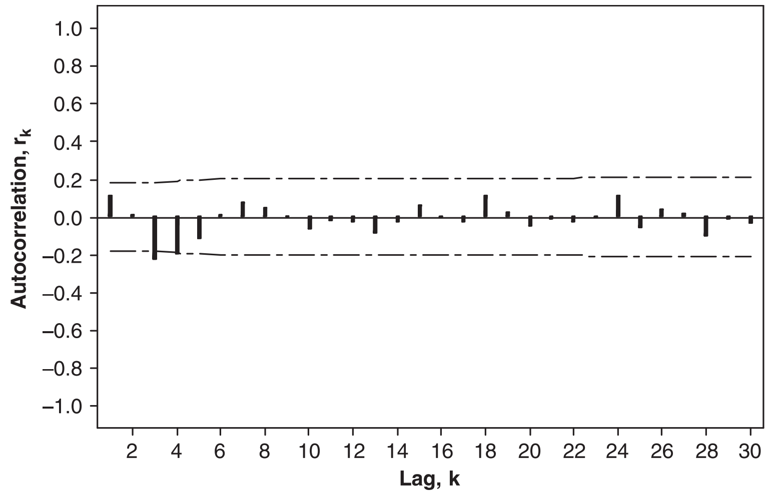


FIGURE 2.14 Autocorrelation function for pharmaceutical product sales, with 5% significance limits.

2.3.3 The Variogram

We have discussed two techniques for determining if a time series is nonstationary, plotting a reasonable long series of the data to see if it drifts or wanders away from its mean for long periods of time, and computing the sample ACF. However, often in practice there is no clear demarcation

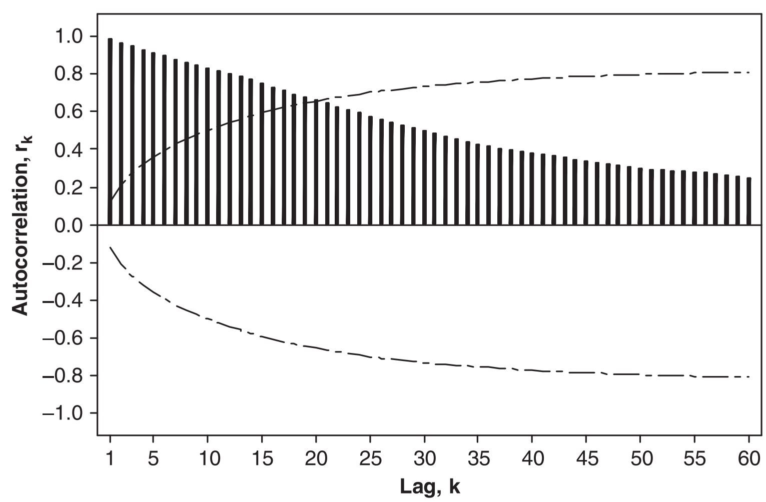


FIGURE 2.15 Autocorrelation function for Whole Foods Market stock price, with 5% significance limits.

between a stationary and a nonstationary process for many real-world time series. An additional diagnostic tool that is very useful is the **variogram**.

Suppose that the time series observations are represented by y_t. The variogram G_k measures variances of the differences between observations that are k lags apart, relative to the variance of the differences that are one time unit apart (or at lag 1). The variogram is defined mathematically as

$$G_k = \frac{\mathrm{Var}\,(y_{t+k} - y_t)}{\mathrm{Var}\,(y_{t+1} - y_t)} \quad k = 1, 2, \ldots \tag{2.16}$$

and the values of G_k are plotted as a function of the lag k. If the time series is stationary, it turns out that

$$G_k = \frac{1 - \rho_k}{1 - \rho_1},$$

but for a stationary time series $\rho_k \to 0$ as k increases, so when the variogram is plotted against lag k, G_k will reach an asymptote $1/(1 - \rho_1)$. However, if the time series is nonstationary, G_k will increase monotonically.

Estimating the variogram is accomplished by simply applying the usual sample variance to the differences, taking care to account for the changing sample sizes when the differences are taken (see Haslett (1997)). Let

$$d_t^k = y_{t+k} - y_t$$
$$\bar{d}^k = \frac{1}{T-k} \sum d_t^k.$$

Then an estimate of $\mathrm{Var}\,(y_{t+k} - y_t)$ is

$$s_k^2 = \frac{\sum_{t=1}^{T-k} \left(d_t^k - \bar{d}^k\right)^2}{T - k - 1}.$$

Therefore the sample variogram is given by

$$\hat{G}_k = \frac{s_k^2}{s_1^2} \quad k = 1, 2, \ldots \tag{2.17}$$

To illustrate the use of the variogram, consider the chemical process viscosity data plotted in Figure 2.9. Both the data plot and the sample ACF in

Lag	Variogram	Plot Variogram
1	1.0000	
2	1.7238	
3	2.3562	
4	2.9527	
5	3.2230	
6	3.6659	
7	3.8729	
8	3.9634	
9	4.1541	
10	4.3259	
11	4.6161	
12	4.9923	
13	4.8752	
14	4.5393	
15	4.2971	
16	4.3065	
17	4.6282	
18	4.9006	
19	5.2050	
20	5.4711	
21	5.3873	
22	5.3109	
23	5.0395	
24	4.6880	
25	4.3416	

FIGURE 2.16 JMP output for the sample variogram of the chemical process viscosity data from Figure 2.19.

Figures 2.12 and 2.13 suggest that the time series is stationary. Figure 2.16 is the variogram. Many software packages do not offer the variogram as a standard pull-down menu selection, but the JMP package does. Without software, it is still fairly easy to compute.

Start by computing the successive differences of the time series for a number of lags and then find their sample variances. The ratios of these sample variances to the sample variance of the first differences will produce the sample variogram. The JMP calculations of the sample variogram are shown in Figure 2.16 and a plot is given in Figure 2.17. Notice that the sample variogram generally converges to a stable level and then fluctuates around it. This is consistent with a stationary time series, and it provides additional evidence that the chemical process viscosity data are stationary.

Now let us see what the sample variogram looks like for a nonstationary time series. The Whole Foods Market stock price data from Appendix Table B.7 originally shown in Figure 1.7 are apparently nonstationary, as it wanders about with no obvious fixed level. The sample ACF in Figure 2.15 decays very slowly and as noted previously, gives the impression that the time series is nonstationary. The calculations for the variogram from JMP are shown in Figure 2.18 and the variogram is plotted in Figure 2.19.

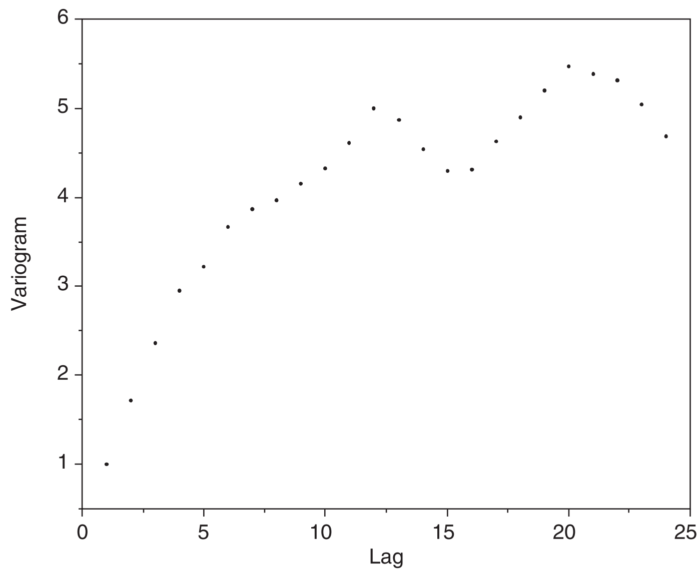


FIGURE 2.17 JMP sample variogram of the chemical process viscosity data from Figure 2.9.

Lag	Variogram	Plot Variogram
1	1.0000	
2	2.0994	
3	3.2106	
4	4.3960	
5	5.4982	
6	6.5810	
7	7.5690	
8	8.5332	
9	9.4704	
10	10.4419	
11	11.4154	
12	12.3452	
13	13.3759	
14	14.4411	
15	15.6184	
16	16.9601	
17	18.2442	
18	19.3782	
19	20.3934	
20	21.3618	
21	22.4010	
22	23.4788	
23	24.5450	
24	25.5906	
25	26.6620	

FIGURE 2.18 JMP output for the sample variogram of the Whole Foods Market stock price data from Figure 1.7 and Appendix Table B.7.

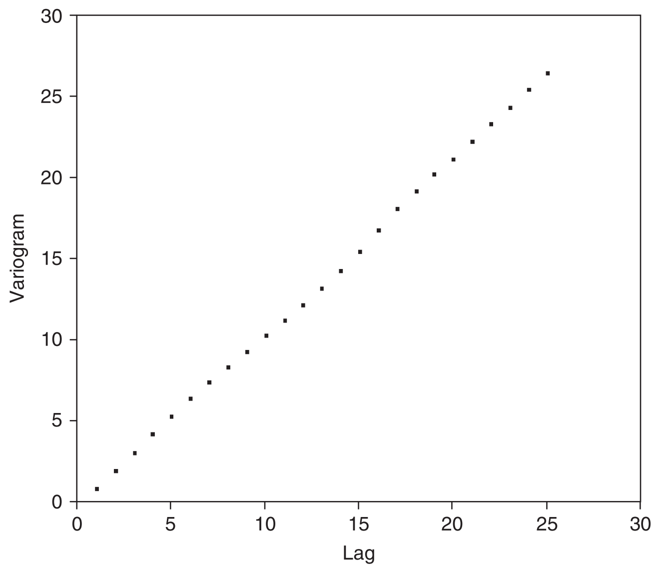


FIGURE 2.19 Sample variogram of the Whole Foods Market stock price data from Figure 1.7 and Appendix Table B.7.

Notice that the sample variogram in Figure 2.19 increases monotonically for all 25 lags. This is a strong indication that the time series is nonstationary.

2.4 USE OF DATA TRANSFORMATIONS AND ADJUSTMENTS

2.4.1 Transformations

Data transformations are useful in many aspects of statistical work, often for stabilizing the variance of the data. Nonconstant variance is quite common in time series data. For example, the International Sunspot Numbers plotted in Figure 2.20a show cyclic patterns of varying magnitudes. The variability from about 1800 to 1830 is smaller than that from about 1830 to 1880; other small periods of constant, but different, variances can also be identified.

A very popular type of data transformation to deal with nonconstant variance is the **power family** of transformations, given by

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y} \ln y, & \lambda = 0 \end{cases}, \tag{2.18}$$

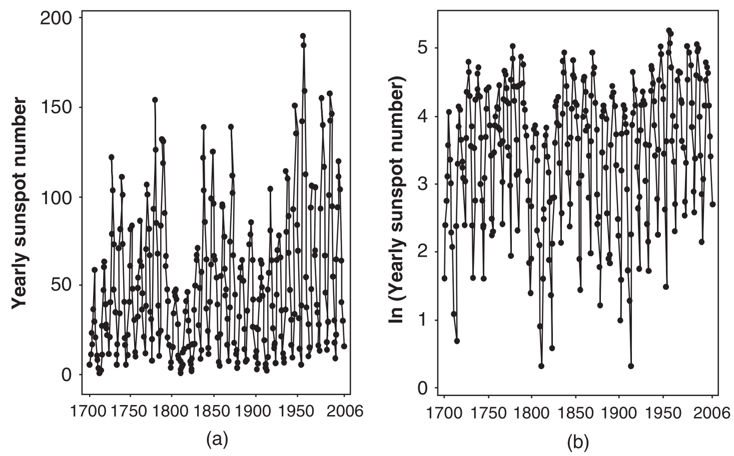


FIGURE 2.20 Yearly International Sunspot Number, (a) untransformed and (b) natural logarithm transformation. *Source*: SIDC.

where $\dot{y} = \exp[(1/T)\sum_{t=1}^{T} \ln y_t]$ is the geometric mean of the observations. If $\lambda = 1$, there is no transformation. Typical values of λ used with time series data are $\lambda = 0.5$ (a square root transformation), $\lambda = 0$ (the log transformation), $\lambda = -0.5$ (reciprocal square root transformation), and $\lambda = -1$ (inverse transformation). The divisor $\dot{y}^{\lambda-1}$ is simply a scale factor that ensures that when different models are fit to investigate the utility of different transformations (values of λ), the residual sum of squares for these models can be meaningfully compared. The reason that $\lambda = 0$ implies a log transformation is that $(y^\lambda - 1)/\lambda$ approaches the log of y as λ approaches zero. Often an appropriate value of λ is chosen empirically by fitting a model to $y^{(\lambda)}$ for various values of λ and then selecting the transformation that produces the minimum residual sum of squares.

The log transformation is used frequently in situations where the variability in the original time series increases with the average level of the series. When the standard deviation of the original series increases linearly with the mean, the log transformation is in fact an optimal variance-stabilizing transformation. The log transformation also has a very nice physical interpretation as percentage change. To illustrate this, let the time series be $y_1, y_2, \dots, y_T$ and suppose that we are interested in the percentage change in y_t, say,

$$x_t = \frac{100(y_t - y_{t-1})}{y_{t-1}},$$

The approximate percentage change in y_t can be calculated from the differences of the log-transformed time series $x_t \cong 100[\ln(y_t) - \ln(y_{t-1})]$ because

$$100[\ln(y_t) - \ln(y_{t-1})] = 100 \ln\left(\frac{y_t}{y_{t-1}}\right) = 100 \ln\left(\frac{y_{t-1} + (y_t - y_{t-1})}{y_{t-1}}\right)$$
$$= 100 \ln\left(1 + \frac{x_t}{100}\right) \cong x_t$$

since $\ln(1 + z) \cong z$ when z is small.

The application of a natural logarithm transformation to the International Sunspot Number, as shown in Figure 2.20b, tends to stabilize the variance and leaves just a few unusual values.

2.4.2 Trend and Seasonal Adjustments

In addition to transformations, there are also several types of adjustments that are useful in time series modeling and forecasting. Two of the most widely used are **trend adjustments** and **seasonal adjustments**. Sometimes these procedures are called trend and seasonal decomposition.

A time series that exhibits a trend is a **nonstationary** time series. Modeling and forecasting of such a time series is greatly simplified if we can eliminate the trend. One way to do this is to fit a **regression model** describing the trend component to the data and then subtracting it out of the original observations, leaving a set of residuals that are free of trend. The trend models that are usually considered are the linear trend, in which the mean of y_t is expected to change linearly with time as in

$$E(y_t) = \beta_0 + \beta_1 t \tag{2.19}$$

or as a quadratic function of time

$$E(y_t) = \beta_0 + \beta_1 t + \beta_2 t^2 \tag{2.20}$$

or even possibly as an exponential function of time such as

$$E(y_t) = \beta_0 e^{\beta_1 t}. \tag{2.21}$$

The models in Eqs. (2.19)–(2.21) are usually fit to the data by using ordinary least squares.

Example 2.6 We will show how least squares can be used to fit regression models in Chapter 3. However, it would be useful at this point to illustrate how trend adjustment works. Minitab can be used to perform trend adjustment. Consider the annual US production of blue and gorgonzola cheeses

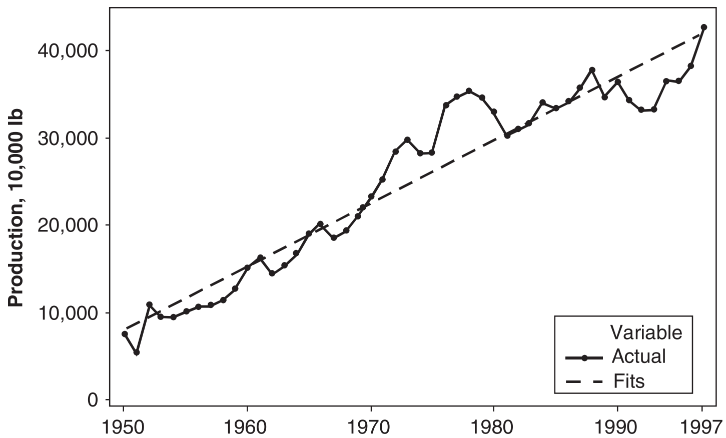


FIGURE 2.21 Blue and gorgonzola cheese production, with fitted regression line. *Source*: USDA–NASS.

shown in Figure 1.4. There is clearly a positive, nearly linear trend. The trend analysis plot in Figure 2.21 shows the original time series with the fitted line.

Plots of the residuals from this model indicate that, in addition to an underlying trend, there is additional structure. The normal probability plot (Figure 2.22a) and histogram (Figure 2.22c) indicate the residuals are

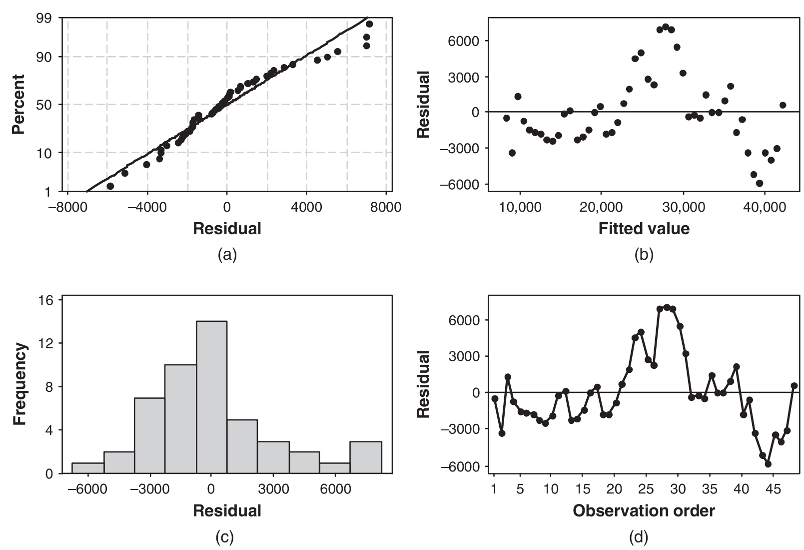


FIGURE 2.22 Residual plots for simple linear regression model of blue and gorgonzola cheese production.

approximately normally distributed. However, the plots of residuals versus fitted values (Figure 2.22b) and versus observation order (Figure 2.22d) indicate nonconstant variance in the last half of the time series. Analysis of model residuals is discussed more fully in Chapter 3.

Another approach to removing trend is by **differencing** the data; that is, applying the difference operator to the original time series to obtain a new time series, say,

$$x_t = y_t - y_{t-1} = \nabla y_t, \tag{2.22}$$

where ∇ is the (backward) difference operator. Another way to write the differencing operation is in terms of a **backshift operator** B, defined as $By_t = y_{t-1}$, so

$$x_t = (1 - B)y_t = \nabla y_t = y_t - y_{t-1} \tag{2.23}$$

with $\nabla = (1 - B)$. Differencing can be performed successively if necessary until the trend is removed; for example, the second difference is

$$x_t = \nabla^2 y_t = \nabla(\nabla y_t) = (1 - B)^2 y_t = (1 - 2B + B^2) = y_t - 2y_{t-1} + y_{t-2} \tag{2.24}$$

In general, powers of the backshift operator and the backward difference operator are defined as

$$\begin{aligned} B^d y_t &= y_{t-d} \\ \nabla^d &= (1 - B)^d \end{aligned} \tag{2.25}$$

Differencing has two advantages relative to fitting a trend model to the data. First, it does not require estimation of any parameters, so it is a more **parsimonious** (i.e., simpler) approach; and second, model fitting assumes that the trend is fixed throughout the time series history and will remain so in the (at least immediate) future. In other words, the trend component, once estimated, is assumed to be **deterministic**. Differencing can allow the trend component to change through time. The first difference accounts for a trend that impacts the change in the mean of the time series, the second difference accounts for changes in the slope of the time series, and so forth. Usually, one or two differences are all that is required in practice to remove an underlying trend in the data.

Example 2.7 Reconsider the blue and gorgonzola cheese production data. A difference of one applied to this time series removes the increasing trend (Figure 2.23) and also improves the appearance of the residuals plotted versus fitted value and observation order when a linear model is fitted to the detrended time series (Figure 2.24). This illustrates that differencing may be a very good alternative to detrending a time series by using a regression model.

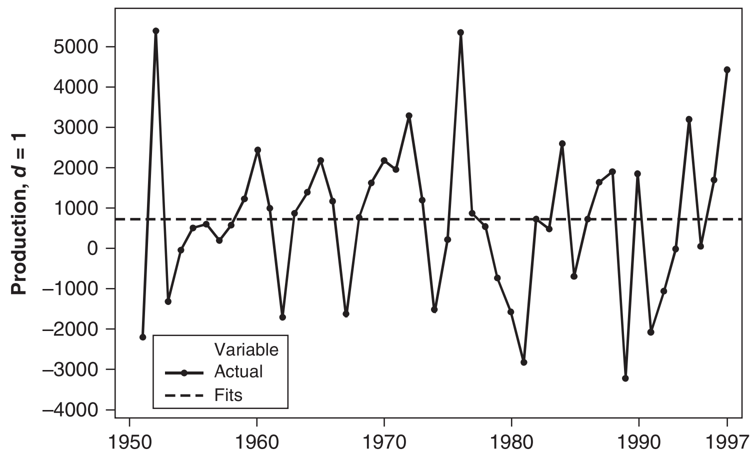


FIGURE 2.23 Blue and gorgonzola cheese production, with one difference. *Source*: USDA–NASS.

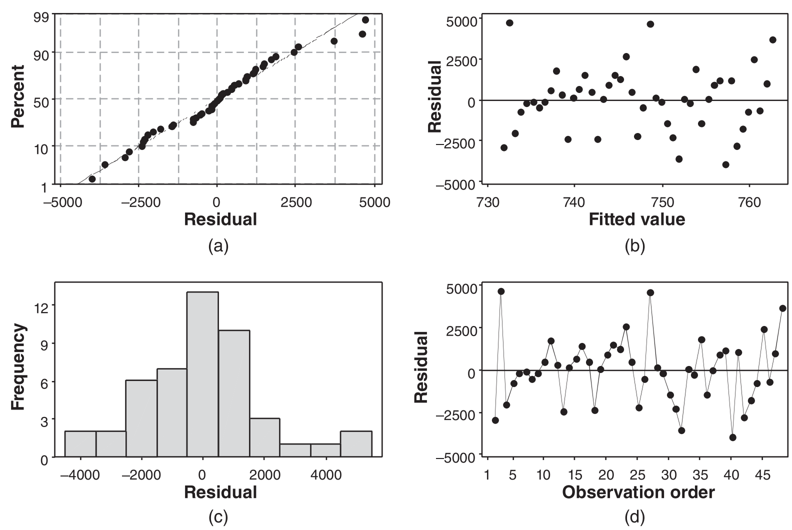


FIGURE 2.24 Residual plots for one difference of blue and gorgonzola cheese production.

Seasonal, or both trend *and* seasonal, components are present in many time series. Differencing can also be used to eliminate seasonality. Define a lag—d **seasonal difference** operator as

$$\nabla_d y_t = (1 - B^d) = y_t - y_{t-d}. \tag{2.26}$$

For example, if we had monthly data with an annual season (a very common situation), we would likely use $d = 12$, so the seasonally differenced data would be

$$\nabla_{12} y_t = (1 - B^{12}) y_t = y_t - y_{t-12}.$$

When both trend *and* seasonal components are simultaneously present, we can sequentially difference to eliminate these effects. That is, first seasonally difference to remove the seasonal component and then difference one or more times using the regular difference operator to remove the trend.

Example 2.8 The beverage shipment data shown in Figure 2.2 appear to have a strong monthly pattern—January consistently has the lowest shipments in a year while the peak shipments are in May and June. There is also an overall increasing trend from year to year that appears to be the same regardless of month.

A seasonal difference of twelve followed by a trend difference of one was applied to the beverage shipments, and the results are shown in Figure 2.25. The seasonal differencing removes the monthly pattern (Figure 2.25a), and the second difference of one removes the overall increasing trend (Figure 2.25b). The fitted linear trend line in Figure 2.25b has a slope of virtually zero. Examination of the residual plots in Figure 2.26 does not reveal any problems with the linear trend model fit to the differenced data.

Regression models can also be used to eliminate seasonal (or trend and seasonal components) from time series data. A simple but useful model is

$$E(y_t) = \beta_0 + \beta_1 \sin \frac{2\pi}{d} t + \beta_2 \cos \frac{2\pi}{d} t, \tag{2.27}$$

where d is the period (or length) of the season and $2\pi/d$ is expressed in radians. For example, if we had monthly data and an annual season, then $d = 12$. This model describes a simple, symmetric seasonal pattern that

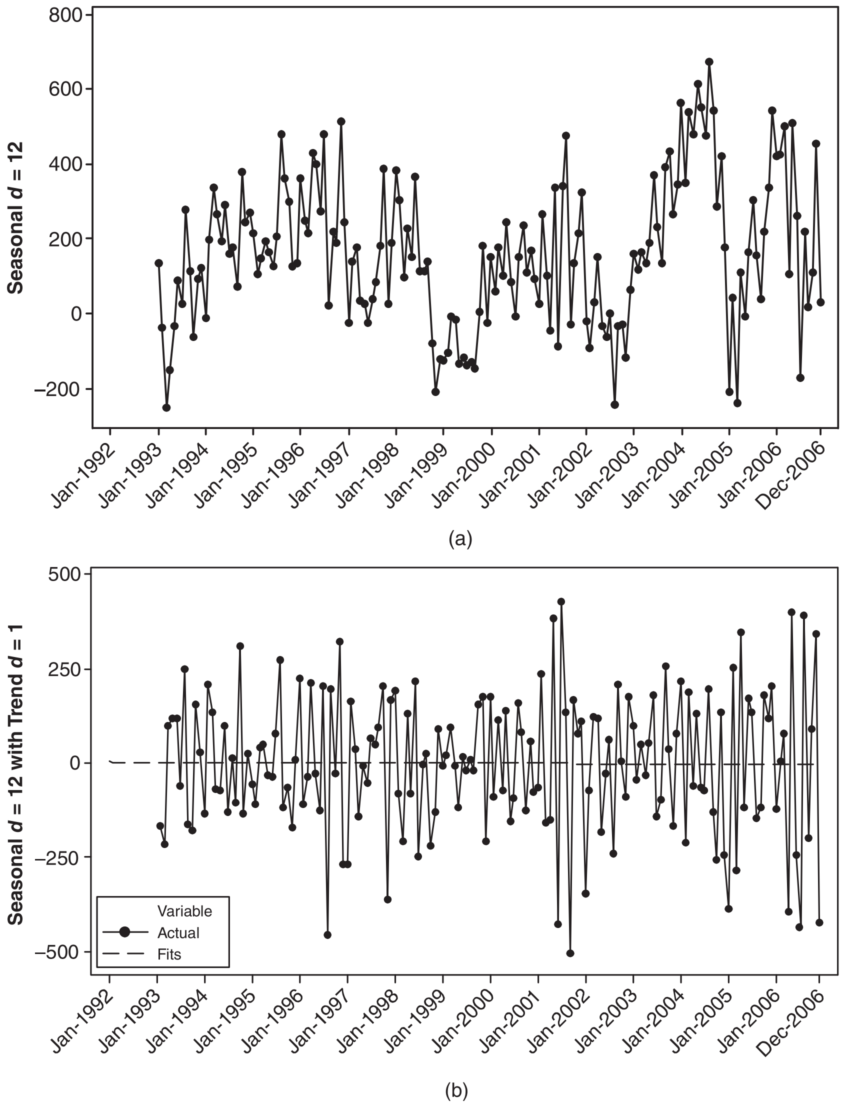


FIGURE 2.25 Time series plots of seasonal- and trend-differenced beverage data.

repeats every 12 periods. The model is actually a sine wave. To see this, recall that a sine wave with amplitude β, phase angle or origin θ, and period or cycle length ω can be written as

$$E(y_t) = \beta \sin \omega(t + \theta). \tag{2.28}$$

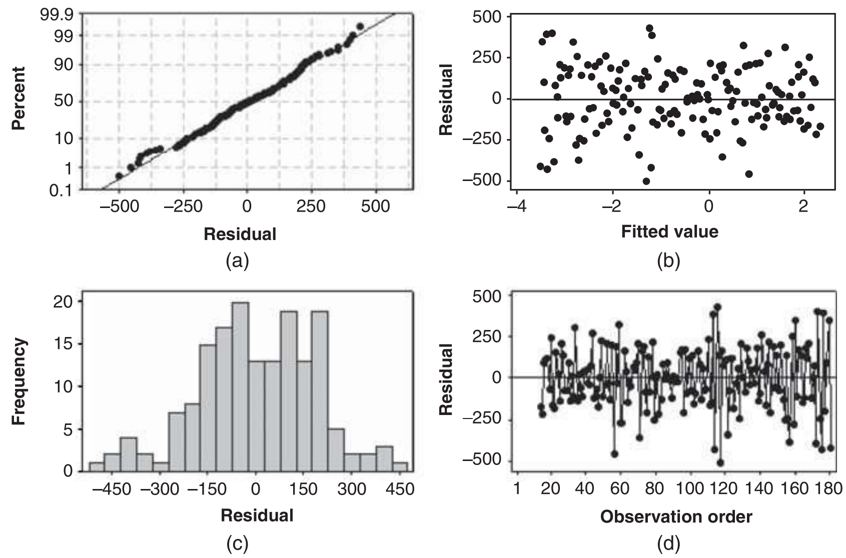


FIGURE 2.26 Residual plots for linear trend model of differenced beverage shipments.

Equation (2.27) was obtained by writing Eq. (2.28) as a sine–cosine pair using the trigonometric identity $\sin(u + v) = \cos u \sin v + \sin u \cos v$ and adding an intercept term β_0:

$$\begin{aligned} E(y_t) &= \beta \sin \omega(t + \theta) \\ &= \beta \cos \omega\theta \sin \omega t + \beta \sin \omega\theta \cos \omega t \\ &= \beta_1 \sin \omega t + \beta_2 \cos \omega t \end{aligned}$$

where $\beta_1 = \beta \cos \omega\theta$ and $\beta_2 = \beta \sin \omega\theta$. Setting $\omega = 2\pi/12$ and adding the intercept term β_0 produces Eq. (2.27). This model is very flexible; for example, if we set $\omega = 2\pi/52$ we can model a yearly seasonal pattern that is observed weekly, if we set $\omega = 2\pi/4$ we can model a yearly seasonal pattern observed quarterly, and if we set $\omega = 2\pi/13$ we can model an annual seasonal pattern observed in 13 four-week periods instead of the usual months.

Equation (2.27) incorporates a single sine wave at the **fundamental frequency** $\omega = 2\pi/12$. In general, we could add **harmonics** of the fundamental frequency to the model in order to model more complex seasonal patterns. For example, a very general model for monthly data and an annual season that uses the fundamental frequency and the first three harmonics is

$$E(y_t) = \beta_0 + \sum_{j=1}^{4} \left(\beta_j \sin \frac{2\pi j}{12} t + \beta_{4+j} \cos \frac{2\pi j}{12} t \right). \tag{2.29}$$

If the data are observed in 13 four-week periods, the model would be

$$E(y_t) = \beta_0 + \sum_{j=1}^{4} \left(\beta_j \sin \frac{2\pi j}{13} t + \beta_{4+j} \cos \frac{2\pi j}{13} t \right). \tag{2.30}$$

There is also a "classical" approach to decomposition of a time series into trend and seasonal components (actually, there are a lot of different decomposition algorithms; here we explain a very simple but useful approach). The general mathematical model for this decomposition is

$$y_t = f(S_t, T_t, \varepsilon_t),$$

where S_t is the seasonal component, T_t is the trend effect (sometimes called the trend-cycle effect), and ε_t is the random error component. There are usually two forms for the function f; an additive model

$$y_t = S_t + T_t + \varepsilon_t$$

and a multiplicative model

$$y_t = S_t T_t \varepsilon_t.$$

The additive model is appropriate if the magnitude (amplitude) of the seasonal variation does not vary with the level of the series, while the multiplicative version is more appropriate if the amplitude of the seasonal fluctuations increases or decreases with the average level of the time series.

Decomposition is useful for breaking a time series down into these component parts. For the additive model, it is relatively easy. First, we would model and remove the trend. A simple linear model could be used to do this, say, $T_t = \beta_0 + \beta_1 t$. Other methods could also be used. Moving averages can be used to isolate a trend and remove it from the original data, as could more sophisticated regression methods. These techniques might be appropriate when the trend is not a straight line over the history of the time series. Differencing could also be used, although it is not typically in the classical decomposition approach.

Once the trend or trend-cycle component is estimated, the series is detrended:

$$y_t - T_t = S_t + \varepsilon_t.$$

Now a seasonal factor can be calculated for each period in the season. For example, if the data are monthly and an annual season is anticipated, we would calculate a season effect for each month in the data set. Then the seasonal indices are computed by taking the average of all of the seasonal factors for each period in the season. In this example, all of the January seasonal factors are averaged to produce a January season index; all of the February seasonal factors are averaged to produce a February season index; and so on. Sometimes medians are used instead of averages. In multiplicative decomposition, ratios are used, so that the data are detrended by

$$\frac{y_t}{T_t} = S_t \varepsilon_t.$$

The seasonal indices are estimated by taking the averages over all of the detrended values for each period in the season.

Example 2.9 The decomposition approach can be applied to the beverage shipment data. Examining the time series plot in Figure 2.2, there is both a strong positive trend as well as month-to-month variation, so the model should include both a trend and a seasonal component. It also appears that the magnitude of the seasonal variation does not vary with the level of the series, so an additive model is appropriate.

Results of a time series decomposition analysis from Minitab of the beverage shipments are in Figure 2.27, showing the original data (labeled "Actual"); along with the fitted trend line ("Trend") and the predicted values ("Fits") from the additive model with both the trend and seasonal components.

Details of the seasonal analysis are shown in Figure 2.28. Estimates of the monthly variation from the trend line for each season (seasonal indices) are in Figure 2.28a with boxplots of the actual differences in Figure 2.28b. The percent of variation by seasonal period is in Figure 2.28c, and model residuals by seasonal period are in Figure 2.28d.

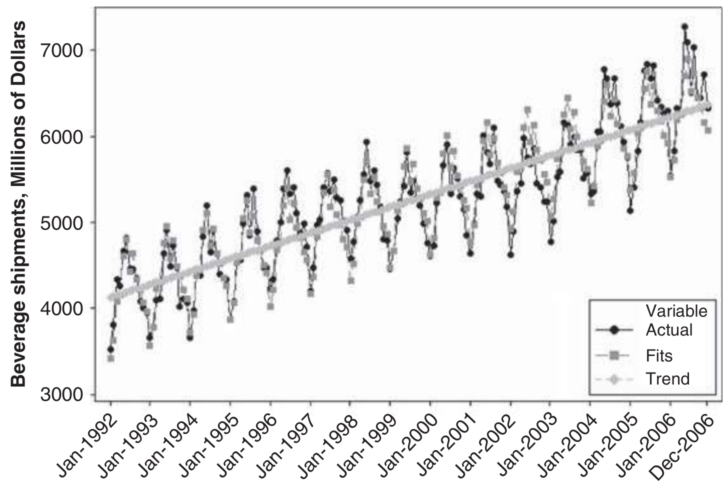


FIGURE 2.27 Time series plot of decomposition model for beverage shipments.

Additional details of the component analysis are shown in Figure 2.29. Figure 2.29a is the original time series, Figure 2.29b is a plot of the time series with the trend removed, Figure 2.29c is a plot of the time series with the seasonality removed, and Figure 2.29d is essentially a residual plot of the detrended and seasonally adjusted data. The wave-like pattern in Figure 2.29d suggests a potential issue with the assumption of constant variance over time.

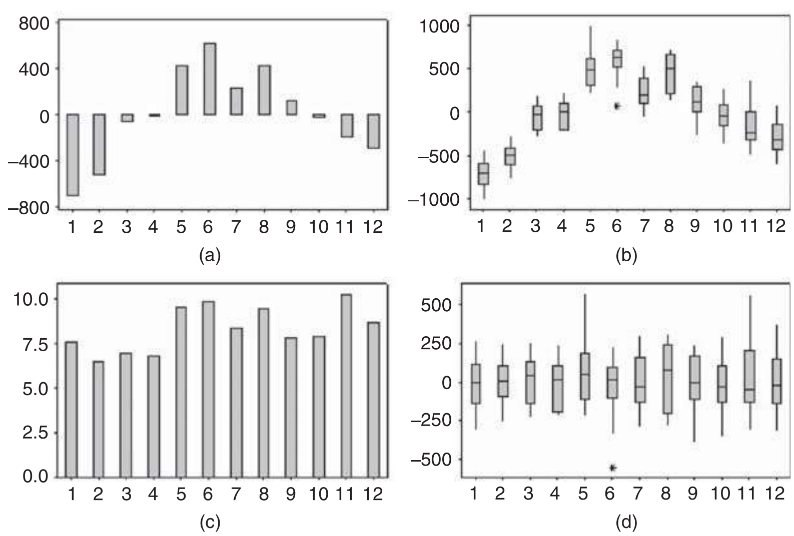


FIGURE 2.28 Seasonal analysis for beverage shipments.

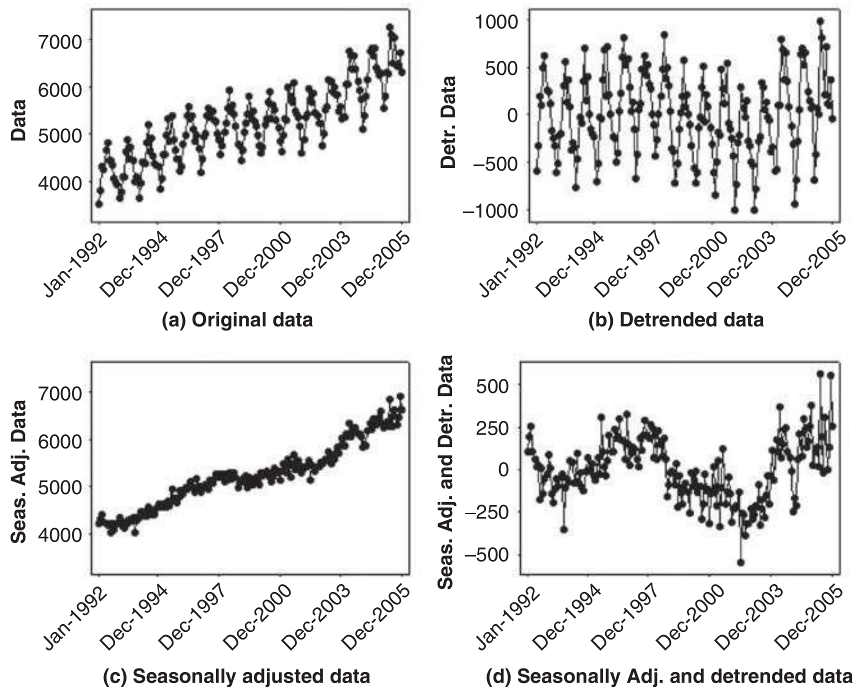


FIGURE 2.29 Component analysis of beverage shipments.

Looking at the normal probability plot and histogram of residuals (Figure 2.30a,c), there does not appear to be an issue with the normality assumption. Figure 2.30d is the same plot as Figure 2.29d. However, variance does seem to increase as the predicted value increases; there is a funnel shape to the residuals plotted in Figure 2.30b. A natural logarithm transformation of the data may stabilize the variance and allow a useful decomposition model to be fit.

Results from the decomposition analysis of the natural log-transformed beverage shipment data are plotted in Figure 2.31, with the transformed data, fitted trend line, and predicted values. Figure 2.32a shows the transformed data, Figure 2.32b the transformed data with the trend removed, Figure 2.32c the transformed data with seasonality removed, and Figure 2.32d the residuals plot of the detrended and seasonally adjusted transformed data. The residual plots in Figure 2.33 indicate that the variance over the range of the predicted values is now stable (Figure 2.33b), and there are no issues with the normality assumption (Figures 2.33a,c). However, there is still a wave-like pattern in the plot of residuals versus time,

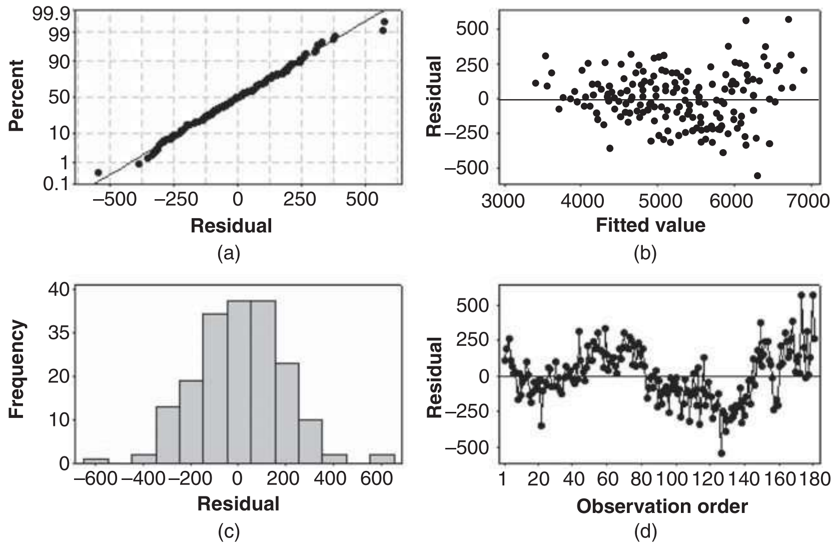


FIGURE 2.30 Residual plots for additive model of beverage shipments.

both Figures 2.32d and 2.33d, indicating that some other structure in the transformed data over time is not captured by the decomposition model. This was not an issue with the model based on seasonal and trend differencing (Figures 2.25 and 2.26), which may be a more appropriate model for monthly beverage shipments.

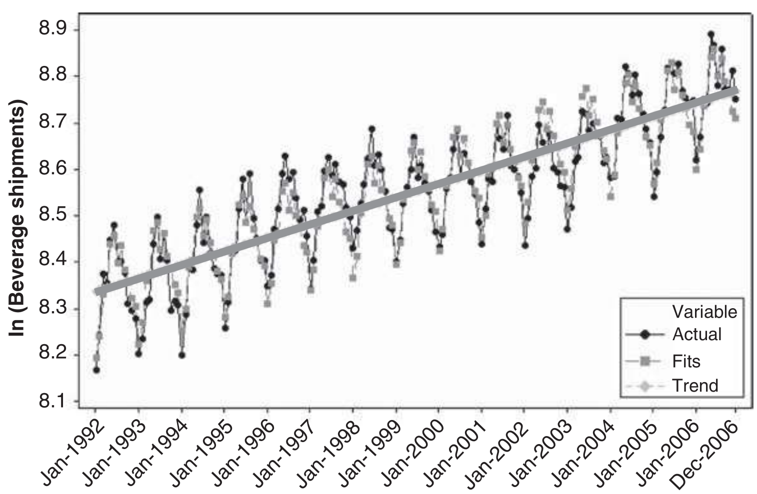


FIGURE 2.31 Time series plot of decomposition model for transformed beverage data.

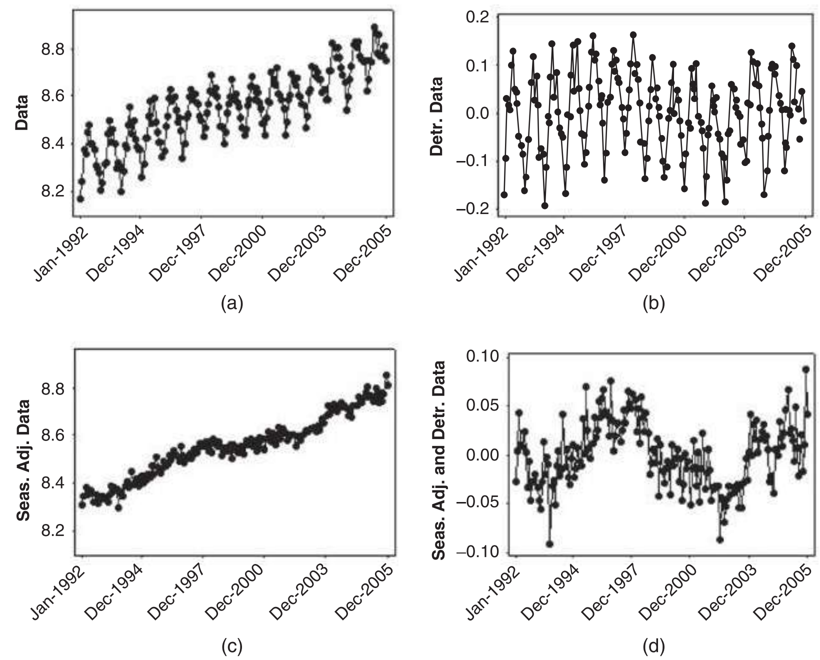


FIGURE 2.32 Component analysis of transformed beverage data.

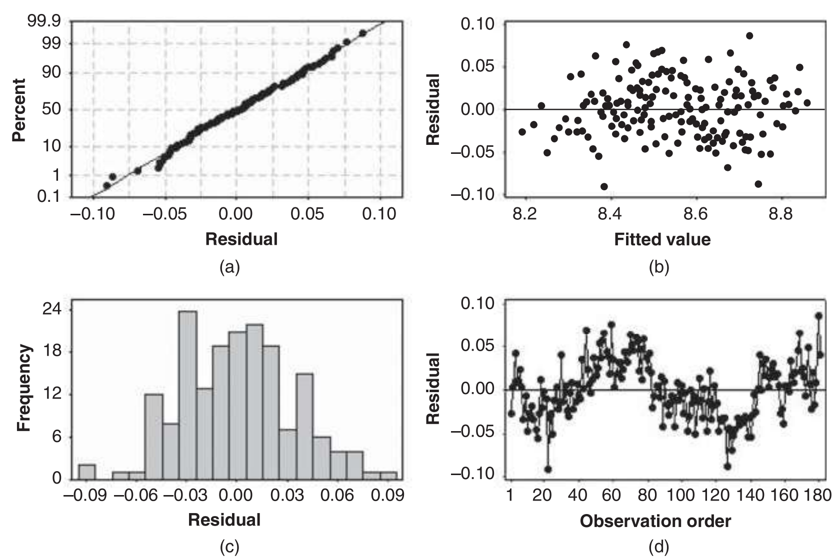


FIGURE 2.33 Residual plots from decomposition model for transformed beverage data.

Another technique for seasonal adjustment that is widely used in modeling and analyzing economic data is the ***X*-11 method**. Much of the development work on this method was done by Julian Shiskin and others at the US Bureau of the Census beginning in the mid-1950s and culminating into the *X*-11 Variant of the Census Method II Seasonal Adjustment Program. References for this work during this period include Shiskin (1958), and Marris (1961). Authoritative documentation for the *X*-11 procedure is in Shiskin, Young, and Musgrave (1967). The *X*-11 method uses symmetric moving averages in an iterative approach to estimating the trend and seasonal components. At the end of the series, however, these symmetric weights cannot be applied. Asymmetric weights have to be used.

JMP (V12 and higher) provides the *X*-11 technique. Figure 2.34 shows the JMP *X*-11 output for the beverage shipment data from Figure 2.2. The upper part of the output contains a plot of the original time series, followed by the sample ACF and PACF. Then Display D10 in the figure shows the final estimates of the seasonal factors by month followed in Display D13 by the irregular or deseasonalized series. The final display is a plot of the original and adjusted time series.

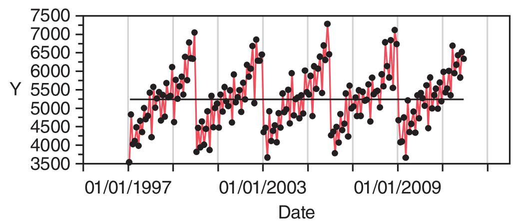
While different variants of the *X*-11 technique have been proposed, the most important method to date has been the *X-11-ARIMA* method developed at Statistics Canada. This method uses Box–Jenkins autoregressive integrated moving average models (which are discussed in Chapter 5) to extend the series. The use of ARIMA models will result in differences in the final component estimates. Details of this method are in Dagum (1980, 1983, 1988).

2.5 GENERAL APPROACH TO TIME SERIES MODELING AND FORECASTING

The techniques that we have been describing form the basis of a general approach to modeling and forecasting time series data. We now give a broad overview of the approach. This should give readers a general understanding of the connections between the ideas we have presented in this chapter and guidance in understanding how the topics in subsequent chapters form a collection of useful techniques for modeling and forecasting time series.

The basic steps in modeling and forecasting a time series are as follows:

1. Plot the time series and determine its basic features, such as whether trends or seasonal behavior or both are present. Look for possible outliers or any indication that the time series has changed with respect



Mean	5238.6611
Std	782.42158
N	180
Zero Mean ADF	–0.793136
Single Mean ADF	–8.334716
Trend ADF	–8.501149

Lag	AutoCorr	Lag	Partial
0	1.0000	0	1.0000
1	0.4398	1	0.4398
2	0.5701	2	0.4670
3	0.3620	3	0.0363
4	0.1353	4	–0.3506
5	0.3132	5	0.2678
6	–0.1007	6	–0.2836
7	0.1533	7	0.1343
8	–0.2213	8	–0.3817
9	–0.1051	9	0.2253
10	–0.0786	10	–0.0379
11	–0.2943	11	–0.0456
12	0.0785	12	0.1256
13	–0.3209	13	–0.2326
14	–0.0537	14	–0.0171
15	–0.1891	15	–0.0413
16	–0.2052	16	0.0353
17	–0.0233	17	–0.1088
18	–0.2957	18	0.0021
19	0.0434	19	0.0036
20	–0.2931	20	–0.1474
21	–0.1244	21	–0.1013
22	–0.1548	22	–0.0729
23	–0.3361	23	–0.0980
24	–0.0158	24	0.0444
25	–0.3572	25	–0.1089

FIGURE 2.34 JMP output for the *X*-11 procedure.

to its basic features (such as trends or seasonality) over the time period history.

2. Eliminate any trend or seasonal components, either by differencing or by fitting an appropriate model to the data. Also consider using data transformations, particularly if the variability in the time series seems to be proportional to the average level of the series. The objective of these operations is to produce a set of stationary residuals.

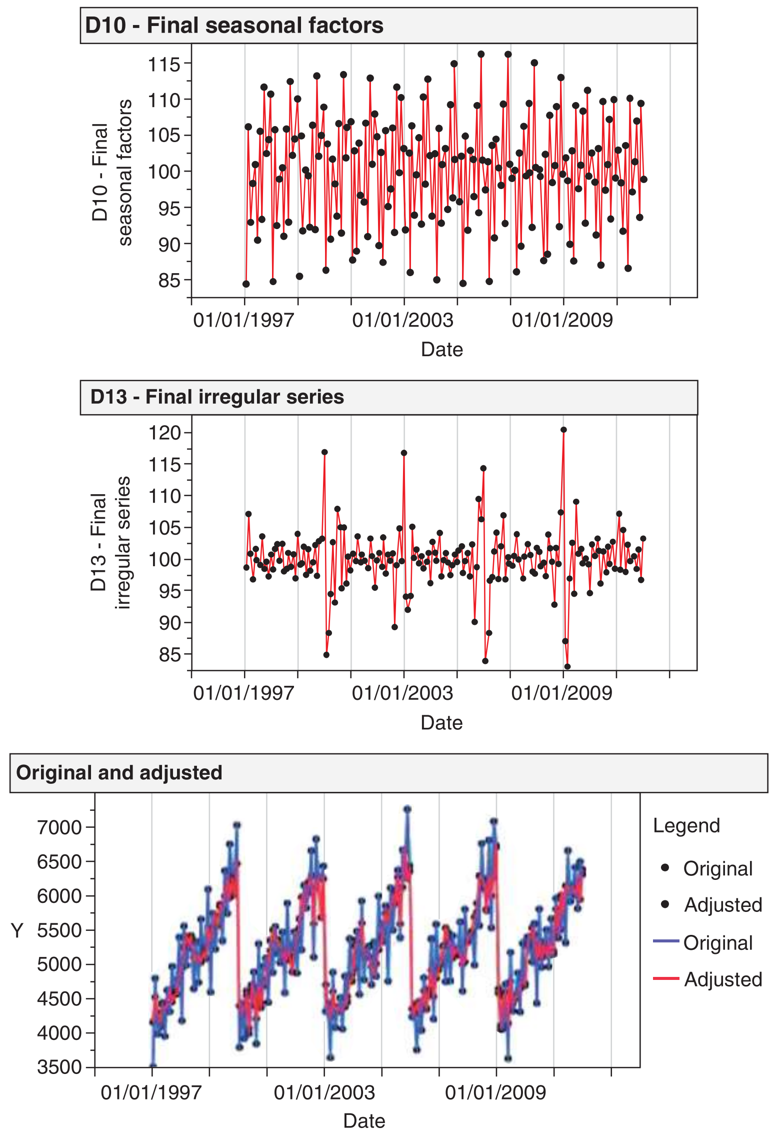


FIGURE 2.34 (*Continued*)

3. Develop a forecasting model for the residuals. It is not unusual to find that there are several plausible models, and additional analysis will have to be performed to determine the best one to deploy. Sometimes potential models can be eliminated on the basis of their fit to the historical data. It is unlikely that a model that fits poorly will produce good forecasts.

4. Validate the performance of the model (or models) from the previous step. This will probably involve some type of split-sample or cross-validation procedure. The objective of this step is to select a model to use in forecasting. We will discuss this more in the next section and illustrate these techniques throughout the book.
5. Also of interest are the differences between the original time series y_t and the values that would be forecast by the model on the original scale. To forecast values on the scale of the original time series y_t, reverse the transformations and any differencing adjustments made to remove trends or seasonal effects.
6. For forecasts of future values in period $T + \tau$ on the original scale, if a transformation was used, say, $x_t = \ln y_t$, then the forecast made at the end of period T for $T + \tau$ would be obtained by reversing the transformation. For the natural log this would be

$$\hat{y}_{T+\tau}(T) = \exp[\hat{x}_{T+\tau}(T)].$$

7. If prediction intervals are desired for the forecast (and we recommend doing this), construct prediction intervals for the residuals and then reverse the transformations made to produce the residuals as described earlier. We will discuss methods for finding prediction intervals for most of the forecasting methods presented in this book.
8. Develop and implement a procedure for monitoring the forecast to ensure that deterioration in performance will be detected reasonably quickly. Forecast monitoring is usually done by evaluating the stream of forecast errors that are experienced. We will present methods for monitoring forecast errors with the objective of detecting changes in performance of the forecasting model.

2.6 EVALUATING AND MONITORING FORECASTING MODEL PERFORMANCE

2.6.1 Forecasting Model Evaluation

We now consider how to evaluate the performance of a forecasting technique for a particular time series or application. It is important to carefully define the meaning of performance. It is tempting to evaluate performance on the basis of the fit of the forecasting or time series model to historical data. There are many statistical measures that describe how well a model fits a given sample of data, and several of these will be described in subsequent chapters. This goodness-of-fit approach often uses the residuals and does not really reflect the capability of the forecasting technique to successfully predict future observations. The user of the forecasts is very concerned about the accuracy of future forecasts, not model goodness of fit, so it is important to evaluate this aspect of any recommended technique. Sometimes forecast accuracy is called "out-of-sample" forecast error, to distinguish it from the residuals that arise from a model-fitting process.

Measure of forecast accuracy should always be evaluated as part of a model validation effort (see step 4 in the general approach to forecasting in the previous section). When more than one forecasting technique seems reasonable for a particular application, these forecast accuracy measures can also be used to discriminate between competing models. We will discuss this more in Section 2.6.2.

It is customary to evaluate forecasting model performance using the one-step-ahead forecast errors

$$e_t(1) = y_t - \hat{y}_t(t-1), \tag{2.31}$$

where $\hat{y}_t(t-1)$ is the forecast of y_t that was made one period prior. Forecast errors at other lags, or at several different lags, could be used if interest focused on those particular forecasts. Suppose that there are n observations for which forecasts have been made and n one-step-ahead forecast errors, $e_t(1), t = 1, 2, \ldots, n$. Standard measures of forecast accuracy are the **average error** or **mean error**

$$\text{ME} = \frac{1}{n} \sum_{t=1}^{n} e_t(1), \tag{2.32}$$

the **mean absolute deviation** (or mean absolute error)

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^{n} |e_t(1)|, \tag{2.33}$$

and the **mean squared error**

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} [e_t(1)]^2. \tag{2.34}$$

The mean forecast error in Eq. (2.32) is an estimate of the expected value of forecast error, which we would hope to be zero; that is, the forecasting technique produces **unbiased** forecasts. If the mean forecast error differs appreciably from zero, bias in the forecast is indicated. If the mean forecast error drifts away from zero when the forecasting technique is in use, this can be an indication that the underlying time series has changed in some fashion, the forecasting technique has not tracked this change, and now biased forecasts are being generated.

Both the mean absolute deviation (MAD) in Eq. (2.33) and the mean squared error (MSE) in Eq. (2.34) measure the **variability** in forecast errors. Obviously, we want the variability in forecast errors to be small. The MSE is a direct estimator of the variance of the one-step-ahead forecast errors:

$$\hat{\sigma}^2_{e(1)} = \text{MSE} = \frac{1}{n}\sum_{t=1}^{n}[e_t(1)]^2. \tag{2.35}$$

If the forecast errors are normally distributed (this is usually not a bad assumption, and one that is easily checked), the MAD is related to the standard deviation of forecast errors by

$$\hat{\sigma}_{e(1)} = \sqrt{\frac{\pi}{2}}\text{MAD} \cong 1.25\,\text{MAD} \tag{2.36}$$

The one-step-ahead forecast error and its summary measures, the ME, MAD, and MSE, are all scale-dependent measures of forecast accuracy; that is, their values are expressed in terms of the original units of measurement (or in the case of MSE, the square of the original units). So, for example, if we were forecasting demand for electricity in Phoenix during the summer, the units would be megawatts (MW). If the MAD for the forecast error during summer months was 5 MW, we might not know whether this was a large forecast error or a relatively small one. Furthermore, accuracy measures that are scale dependent do not facilitate comparisons of a single forecasting technique across different time series, or comparisons across different time periods. To accomplish this, we need a measure of relative forecast error.

Define the **relative forecast error** (in percent) as

$$re_t(1) = \left(\frac{y_t - \hat{y}_t(t-1)}{y_t}\right)100 = \left(\frac{e_t(1)}{y_t}\right)100. \tag{2.37}$$

This is customarily called the **percent forecast error**. The mean percent forecast error (MPE) is

$$\text{MPE} = \frac{1}{n}\sum_{t=1}^{n} re_t(1) \tag{2.38}$$

and the mean absolute percent forecast error (MAPE) is

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |re_t(1)|. \tag{2.39}$$

Knowing that the relative or percent forecast error or the MAPE is 3% (say) can be much more meaningful than knowing that the MAD is 5 MW. Note that the relative or percent forecast error only makes sense if the time series y_t does not contain zero values.

Example 2.10 Table 2.2 illustrates the calculation of the one-step-ahead forecast error, the absolute errors, the squared errors, the relative (percent) error, and the absolute percent error from a forecasting model for 20 time periods. The last row of columns (3) through (7) display the sums required to calculate the ME, MAD, MSE, MPE, and MAPE.

From Eq. (2.32), the mean (or average) forecast error is

$$\text{ME} = \frac{1}{n}\sum_{t=1}^{n} e_t(1) = \frac{1}{20}(-11.6) = -0.58,$$

the MAD is computed from Eq. (2.33) as

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n} |e_t(1)| = \frac{1}{20}(86.6) = 4.33,$$

and the MSE is computed from Eq. (2.34) as

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} [e_t(1)]^2 = \frac{1}{20}(471.8) = 23.59.$$

TABLE 2.2 Calculation of Forecast Accuracy Measures

Time Period	(1) Observed Value y_t	(2) Forecast $\hat{y}_t(t-1)$	(3) Forecast Error $e_t(1)$	(4) Absolute Error $\lvert e_t(1)\rvert$	(5) Squared Error $[e_t(1)]^2$	(6) Relative (%) Error $\left(e_t(1)/y_t\right) 100$	(6) Absolute (%) Error $\lvert \left(e_t(1)/y_t\right) 100\rvert$
1	47	51.1	−4.1	4.1	16.81	−8.7234	8.723404
2	46	52.9	−6.9	6.9	47.61	−15	15
3	51	48.8	2.2	2.2	4.84	4.313725	4.313725
4	44	48.1	−4.1	4.1	16.81	−9.31818	9.318182
5	54	49.7	4.3	4.3	18.49	7.962963	7.962963
6	47	47.5	−0.5	0.5	0.25	−1.06383	1.06383
7	52	51.2	0.8	0.8	0.64	1.538462	1.538462
8	45	53.1	−8.1	8.1	65.61	−18	18
9	50	54.4	−4.4	4.4	19.36	−8.8	8.8
10	51	51.2	−0.2	0.2	0.04	−0.39216	0.392157
11	49	53.3	−4.3	4.3	18.49	−8.77551	8.77551
12	41	46.5	−5.5	5.5	30.25	−13.4146	13.41463
13	48	53.1	−5.1	5.1	26.01	−10.625	10.625
14	50	52.1	−2.1	2.1	4.41	−4.2	4.2
15	51	46.8	4.2	4.2	17.64	8.235294	8.235294
16	55	47.7	7.3	7.3	53.29	13.27273	13.27273
17	52	45.4	6.6	6.6	43.56	12.69231	12.69231
18	53	47.1	5.9	5.9	34.81	11.13208	11.13208
19	48	51.8	−3.8	3.8	14.44	−7.91667	7.916667
20	52	45.8	6.2	6.2	38.44	11.92308	11.92308
		Totals	*−11.6*	*86.6*	*471.8*	*−35.1588*	*177.3*

Because the MSE estimates the variance of the one-step-ahead forecast errors, we have

$$\hat{\sigma}^2_{e(1)} = \text{MSE} = 23.59$$

and an estimate of the standard deviation of forecast errors is the square root of this quantity, or $\hat{\sigma}_{e(1)} = \sqrt{\text{MSE}} = 4.86$. We can also obtain an estimate of the standard deviation of forecasts errors from the MAD using Eq. (2.36)

$$\hat{\sigma}_{e(1)} \cong 1.25\,\text{MAD} = 1.25(4.33) = 5.41.$$

These two estimates are reasonably similar. The mean percent forecast error, MPE, is computed from Eq. (2.38) as

$$\text{MPE} = \frac{1}{n}\sum_{t=1}^{n} re_t(1) = \frac{1}{20}(-35.1588) = -1.76\%$$

and the mean absolute percent error is computed from Eq. (2.39) as

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |re_t(1)| = \frac{1}{20}(177.3) = 8.87\%.$$

There is much empirical evidence (and even some theoretical justification) that the distribution of forecast errors can be well approximated by a **normal** distribution. This can easily be checked by constructing a **normal probability plot** of the forecast errors in Table 2.2, as shown in Figure 2.35. The forecast errors deviate somewhat from the straight line, indicating that the normal distribution is not a perfect model for the distribution of forecast errors, but it is not unreasonable. Minitab calculates the Anderson–Darling statistic, a widely used test statistic for normality. The *P*-value is 0.088, so the hypothesis of normality of the forecast errors would not be rejected at the 0.05 level. This test assumes that the observations (in this case the forecast errors) are uncorrelated. Minitab also reports the standard deviation of the forecast errors to be 4.947, a slightly larger value than we computed from the MSE, because Minitab uses the standard method for calculating sample standard deviations.

Note that Eq. (2.31) could have been written as

$$\text{Error} = \text{Observation} - \text{Forecast}.$$

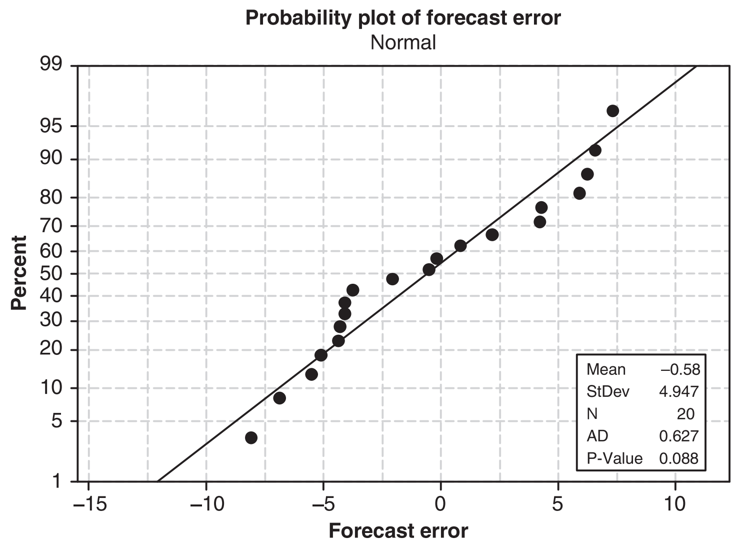


FIGURE 2.35 Normal probability plot of forecast errors from Table 2.2.

Hopefully, the forecasts do a good job of describing the structure in the observations. In an ideal situation, the forecasts would adequately model all of the structure in the data, and the sequence of forecast errors would be structureless. If they are, the sample ACF of the forecast error should look like the ACF of random data; that is, there should not be any large "spikes" on the sample ACF at low lag. Any systematic or nonrandom pattern in the forecast errors will tend to show up as significant spikes on the sample ACF. If the sample ACF suggests that the forecast errors are not random, then this is evidence that the forecasts can be improved by **refining** the forecasting model. Essentially, this would consist of taking the structure out of the forecast errors and putting it into the forecasts, resulting in forecasts that are better prediction of the data.

Example 2.11 Table 2.3 presents a set of 50 one-step-ahead errors from a forecasting model, and Table 2.4 shows the sample ACF of these forecast errors. The sample ACF is plotted in Figure 2.36. This sample ACF was obtained from Minitab. Note that sample autocorrelations for the first 13 lags are computed. This is consistent with our guideline indicating that for T observations only the first $T/4$ autocorrelations should be computed. The sample ACF does not provide any strong evidence to support a claim that there is a pattern in the forecast errors.

TABLE 2.3 One-Step-Ahead Forecast Errors

Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$
1	−0.62	11	−0.49	21	2.90	31	−1.88	41	−3.98
2	−2.99	12	4.13	22	0.86	32	−4.46	42	−4.28
3	0.65	13	−3.39	23	5.80	33	−1.93	43	1.06
4	0.81	14	2.81	24	4.66	34	−2.86	44	0.18
5	−2.25	15	−1.59	25	3.99	35	0.23	45	3.56
6	−2.63	16	−2.69	26	−1.76	36	−1.82	46	−0.24
7	3.57	17	3.41	27	2.31	37	0.64	47	−2.98
8	0.11	18	4.35	28	−2.24	38	−1.55	48	2.47
9	0.59	19	−4.37	29	2.95	39	0.78	49	0.66
10	−0.63	20	2.79	30	6.30	40	2.84	50	0.32

TABLE 2.4 Sample ACF of the One-Step-Ahead Forecast Errors in Table 2.3

Lag	Sample ACF, r_k	Z-Statistic	Ljung–Box Statistic, Q_{LB}
1	0.004656	0.03292	0.0012
2	−0.102647	−0.72581	0.5719
3	0.136810	0.95734	1.6073
4	−0.033988	−0.23359	1.6726
5	0.118876	0.81611	2.4891
6	0.181508	1.22982	4.4358
7	−0.039223	−0.25807	4.5288
8	−0.118989	−0.78185	5.4053
9	0.003400	0.02207	5.4061
10	0.034631	0.22482	5.4840
11	−0.151935	−0.98533	7.0230
12	−0.207710	−1.32163	9.9749
13	0.089387	0.54987	10.5363

If a time series consists of uncorrelated observations and has constant variance, we say that it is **white noise**. If, in addition, the observations in this time series are normally distributed, the time series is **Gaussian white noise**. Ideally, forecast errors are Gaussian white noise. The normal probability plot of the one-step-ahead forecast errors from Table 2.3 are shown in Figure 2.37. This plot does not indicate any serious problem, with the normality assumption, so the forecast errors in Table 2.3 are Gaussian white noise.

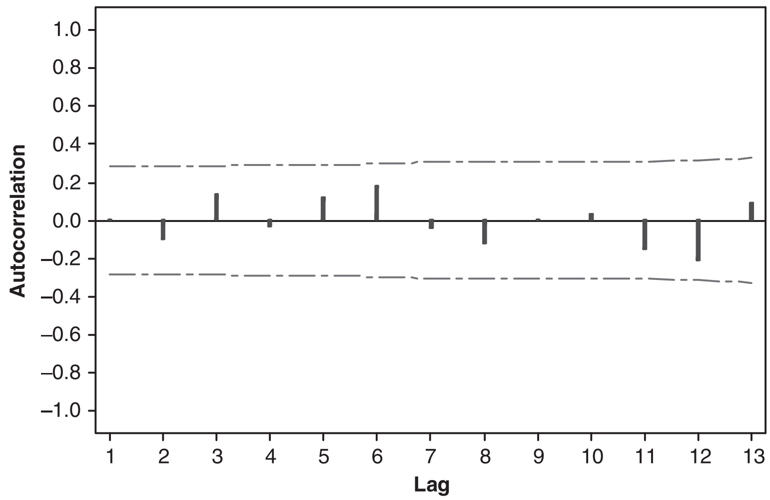


FIGURE 2.36 Sample ACF of forecast errors from Table 2.4.

If a time series is white noise, the distribution of the sample autocorrelation coefficient at lag k in large samples is approximately normal with mean zero and variance $1/T$; that is,

$$r_k \sim N\left(0, \frac{1}{T}\right).$$

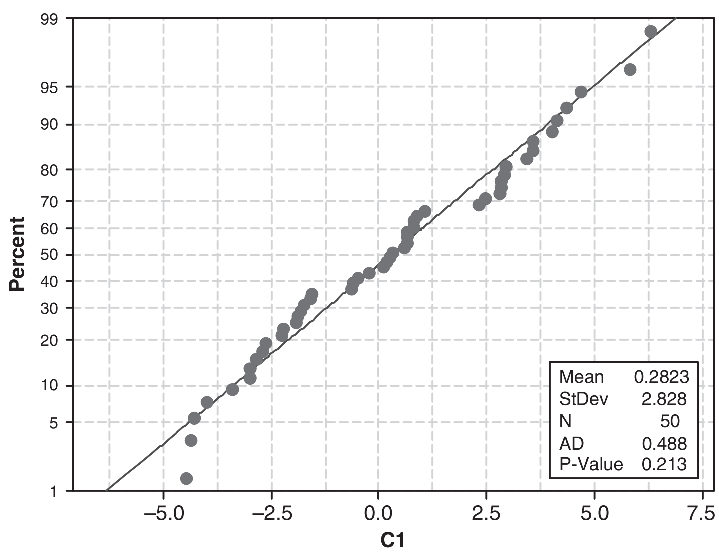


FIGURE 2.37 Normal probability plot of forecast errors from Table 2.3.

Therefore we could test the hypothesis $H_0 : \rho_k = 0$ using the test statistic

$$Z_0 = \frac{r_k}{\sqrt{\frac{1}{T}}} = r_k\sqrt{T}. \tag{2.40}$$

Minitab calculates this Z-statistic (calling it a t-statistic), and it is reported in Table 2.4 for the one-step-ahead forecast errors of Table 2.3 (this is the t-statistic reported in Figure 2.13 for the ACF of the chemical viscosity readings). Large values of this statistic (say, $|Z_0| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution) would indicate that the corresponding autocorrelation coefficient does not equal zero. Alternatively, we could calculate a P-value for this test statistic. Since none of the absolute values of the Z-statistics in Table 2.4 exceeds $Z_{\alpha/2} = Z_{0.025} = 1.96$, we cannot conclude at significance level $\alpha = 0.05$ that any individual autocorrelation coefficient differs from zero.

This procedure is a one-at-a-time test; that is, the significance level applies to the autocorrelations considered individually. We are often interested in evaluating a *set* of autocorrelations jointly to determine if they indicate that the time series is white noise. Box and Pierce (1970) have suggested such a procedure. Consider the square of the test statistic Z_0 in Eq. (2.40). The distribution of $Z_0^2 = r_k^2 T$ is approximately chi-square with one degree of freedom. The Box–Pierce statistic

$$Q_{\text{BP}} = T \sum_{k=1}^{K} r_k^2 \tag{2.41}$$

is distributed approximately as chi-square with Kdegrees of freedom under the null hypothesis that the time series is white noise. Therefore, if $Q_{\text{BP}} > \chi^2_{\alpha,K}$ we would reject the null hypothesis and conclude that the time series is not white noise because some of the autocorrelations are not zero. A P-value approach could also be used. When this test statistic is applied to a set of **residual autocorrelations** the statistic $Q_{\text{BP}} \sim \chi^2_{\alpha,K-p}$, where p is the number of parameters in the model, so the number of degrees of freedom in the chi-square distribution becomes $K - p$. Box and Pierce call this procedure a "Portmanteau" or general **goodness-of-fit statistic** (it is testing the goodness of fit of the ACF to the ACF of white noise). A modification of this test that works better for small samples was devised by Ljung and Box (1978). The Ljung–Box goodness-of-fit statistic is

$$Q_{\text{LB}} = T(T+2) \sum_{k=1}^{K} \left(\frac{1}{T-k}\right) r_k^2. \tag{2.42}$$

Note that the Ljung–Box goodness-of-fit statistic is very similar to the original Box–Pierce statistic, the difference being that the squared sample autocorrelation at lag k is weighted by $(T+2)/(T-k)$. For large values of T, these weights will be approximately unity, and so the Q_{LB} and Q_{BP} statistics will be very similar.

Minitab calculates the Ljung–Box goodness-of-fit statistic Q_{LB}, and the values for the first 13 sample autocorrelations of the one-step-ahead forecast errors of Table 2.3 are shown in the last column of Table 2.4. At lag 13, the value $Q_{\text{LB}} = 10.5363$, and since $\chi^2_{0.05,13} = 22.36$, there is no strong evidence to indicate that the first 13 autocorrelations of the forecast errors considered jointly differ from zero. If we calculate the P-value for this test statistic, we find that $P = 0.65$. This is a good indication that the forecast errors are white noise. Note that Figure 2.13 also gave values for the Ljung–Box statistic.

2.6.2 Choosing Between Competing Models

There are often several competing models that can be used for forecasting a particular time series. For example, there are several ways to model and forecast trends. Consequently, selecting an appropriate forecasting model is of considerable practical importance. In this section we discuss some general principles of model selection. In subsequent chapters, we will illustrate how these principles are applied in specific situations.

Selecting the model that provides the best fit to historical data generally does not result in a forecasting method that produces the best forecasts of new data. Concentrating too much on the model that produces the best historical fit often results in **overfitting**, or including too many parameters or terms in the model just because these additional terms improve the model fit. In general, the best approach is to select the model that results in the smallest standard deviation (or mean squared error) of the one-step-ahead forecast errors when the model is applied to data that were not used in the fitting process. Some authors refer to this as an **out-of-sample** forecast error standard deviation (or mean squared error). A standard way to measure this out-of-sample performance is by utilizing some form of **data splitting**; that is, divide the time series data into two segments—one for model fitting and the other for performance testing. Sometimes data splitting is called **cross-validation**. It is somewhat arbitrary as to how the data splitting is accomplished. However, a good rule of thumb is to have at least 20 or 25 observations in the performance testing data set.

When evaluating the fit of the model to historical data, there are several criteria that may be of value. The **mean squared error** of the residuals is

$$s^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - p} \tag{2.43}$$

where T periods of data have been used to fit a model with p parameters and e_t is the residual from the model-fitting process in period t. The mean squared error s^2 is just the sample variance of the residuals and it is an estimator of the variance of the model errors.

Another criterion is the R-squared statistic

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}. \tag{2.44}$$

The denominator of Eq. (2.44) is just the total sum of squares of the observations, which is constant (not model dependent), and the numerator is just the residual sum of squares. Therefore, selecting the model that maximizes R^2 is equivalent to selecting the model that minimizes the sum of the squared residuals. Large values of R^2 suggest a good fit to the historical data. Because the residual sum of squares always decreases when parameters are added to a model, relying on R^2 to select a forecasting model encourages overfitting or putting in more parameters than are really necessary to obtain good forecasts. A large value of R^2 does not ensure that the out-of-sample one-step-ahead forecast errors will be small.

A better criterion is the "adjusted" R^2 statistic, defined as

$$R_{\text{Adj}}^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2/(T - p)}{\sum_{t=1}^{T} (y_t - \bar{y})^2/(T - 1)} = 1 - \frac{s^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2/(T - 1)}. \tag{2.45}$$

The adjustment is a "size" adjustment—that is, adjust for the number of parameters in the model. Note that a model that maximizes the adjusted R^2 statistic is also the model that minimizes the residual mean square.

Two other important criteria are the **Akaike Information Criterion** (**AIC**) (see Akaike (1974)) and the **Schwarz Bayesian Information Criterion** (**abbreviated as BIC or SIC by various authors**) (see Schwarz (1978)):

$$\mathrm{AIC} = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{2p}{T} \tag{2.46}$$

and

$$\mathrm{BIC} = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{p\ln(T)}{T}. \tag{2.47}$$

These two criteria penalize the sum of squared residuals for including additional parameters in the model. Models that have small values of the AIC or BIC are considered good models.

One way to evaluate model selection criteria is in terms of **consistency**. A model selection criterion is consistent if it selects the true model when the true model is among those considered with probability approaching unity as the sample size becomes large, and if the true model is not among those considered, it selects the best approximation with probability approaching unity as the sample size becomes large. It turns out that s^2, the adjusted R^2, and the AIC are all inconsistent, because they do not penalize for adding parameters heavily enough. Relying on these criteria tends to result in overfitting. The BIC, which caries a heavier "size adjustment" penalty, is consistent.

Consistency, however, does not tell the complete story. It may turn out that the true model and any reasonable approximation to it are very complex. An **asymptotically efficient** model selection criterion chooses a sequence of models as T(the amount of data available) gets large for which the one-step-ahead forecast error variances approach the one-step-ahead forecast error variance for the true model at least as fast as any other criterion. The AIC is asymptotically efficient but the BIC is not.

There are a number of variations and extensions of these criteria. The AIC is a biased estimator of the discrepancy between all candidate models and the true model. This has led to developing a "corrected" version of AIC:

$$\mathrm{AICc} = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{2T(p+1)}{T-p-2}. \tag{2.48}$$

Sometimes we see the first term in the AIC, AICc, or BIC written as $-2\ln L(\boldsymbol{\beta}, \sigma^2)$, where $L(\boldsymbol{\beta}, \sigma^2)$ is the **likelihood function** for the fitted model evaluated at the maximum likelihood estimates of the unknown parameters $\boldsymbol{\beta}$ and σ^2. In this context, AIC, AICc, and SIC are called penalized likelihood criteria.

Many software packages evaluate and print model selection criteria, such as those discussed here. When both AIC and SIC are available, we prefer using SIC. It generally results in smaller, and hence simpler, models, and so its use is consistent with the time-honored model-building principle of **parsimony** (all other things being equal, simple models are preferred to complex ones). We will discuss and illustrate model selection criteria again in subsequent chapters. However, remember that the best way to evaluate a candidate model's potential predictive performance is to use data splitting. This will provide a direct estimate of the one-step-ahead forecast error variance, and this method should always be used, if possible, along with the other criteria that we have discussed here.

2.6.3 Monitoring a Forecasting Model

Developing and implementing procedures to monitor the performance of the forecasting model is an essential component of good forecasting system design. No matter how much effort has been expended in developing the forecasting model, and regardless of how well the model works initially, over time it is likely that its performance will deteriorate. The underlying pattern of the time series may change, either because the internal inertial forces that drive the process may evolve through time, or because of external events such as new customers entering the market. For example, a level change or a slope change could occur in the variable that is being forecasted. It is also possible for the inherent variability in the data to increase. Consequently, performance monitoring is important.

The one-step-ahead forecast errors $e_t(1)$ are typically used for forecast monitoring. The reason for this is that changes in the underlying time series will also typically be reflected in the forecast errors. For example, if a level change occurs in the time series, the sequence of forecast errors will no longer fluctuate around zero; that is, a positive or negative bias will be introduced.

There are several ways to monitor forecasting model performance. The simplest way is to apply **Shewhart control charts** to the forecast errors. A Shewhart control chart is a plot of the forecast errors versus time containing a center line that represents the average (or the target value) of the forecast errors and a set of **control limits** that are designed to provide an indication that the forecasting model performance has changed. The center line is usually taken as either zero (which is the anticipated forecast error for an unbiased forecast) or the average forecast error (ME from Eq. (2.32)), and the control limits are typically placed at three standard deviations of the forecast errors above and below the center line. If the forecast errors plot within the control limits, we assume that the forecasting model performance is satisfactory (or in control), but if one or more forecast errors exceed the control limits, that is a signal that something has happened and the forecast errors are no longer fluctuating around zero. In control chart terminology, we would say that the forecasting process is out of control and some analysis is required to determine what has happened.

The most familiar Shewhart control charts are those applied to data that have been collected in subgroups or samples. The one-step-ahead forecast errors $e_t(1)$ are individual observations. Therefore the Shewhart control chart for individuals would be used for forecast monitoring. On this control chart it is fairly standard practice to estimate the standard deviation of the individual observations using a moving range method. The moving range is defined as the absolute value of the difference between any two successive one-step-ahead forecast errors, say, $|e_t(1) - e_{t-1}(1)|$, and the moving range based on n observations is

$$MR = \sum_{t=2}^{n} |e_t(1) - e_{t-1}(1)|. \tag{2.49}$$

The estimate of the standard deviation of the one-step-ahead forecast errors is based on the average of the moving ranges

$$\hat{\sigma}_{e(1)} = \frac{0.8865 MR}{n-1} = \frac{0.8865 \sum_{t=2}^{n} |e_t(1) - e_{t-1}(1)|}{n-1} = 0.8865\overline{MR}, \tag{2.50}$$

where $\overline{MR}$ is the average of the moving ranges. This estimate of the standard deviation would be used to construct the control limits on the control chart for forecast errors. For more details on constructing and interpreting control charts, see Montgomery (2013).

Example 2.12 Minitab can be used to construct Shewhart control charts for individuals. Figure 2.38 shows the Minitab control charts for the one-step-ahead forecast errors in Table 2.3. Note that both an individuals control chart of the one-step-ahead forecast errors and a control chart of the moving ranges of these forecast errors are provided. On the individuals control chart the center line is taken to be the average of the forecast errors ME defined in Eq. (2.30) (denoted $\overline{X}$ in Figure 2.38) and the upper and lower three-sigma control limits are abbreviated as UCL and LCL, respectively. The center line on the moving average control chart is at the average of the moving ranges $\overline{MR} = MR/(n-1)$, the three-sigma upper control limit UCL is at $3.267MR/(n-1)$, and the lower control limit is at zero (for details on how the control limits are derived, see Montgomery (2013)). All of the one-step-ahead forecast errors plot within the control limits (and the moving range also plot within their control limits). Thus there is no reason to suspect that the forecasting model is performing inadequately, at least from the statistical stability viewpoint. Forecast errors that plot outside the control limits would indicate model inadequacy, or possibly the presence of unusual observations such as outliers in the data. An investigation would be required to determine why these forecast errors exceed the control limits.

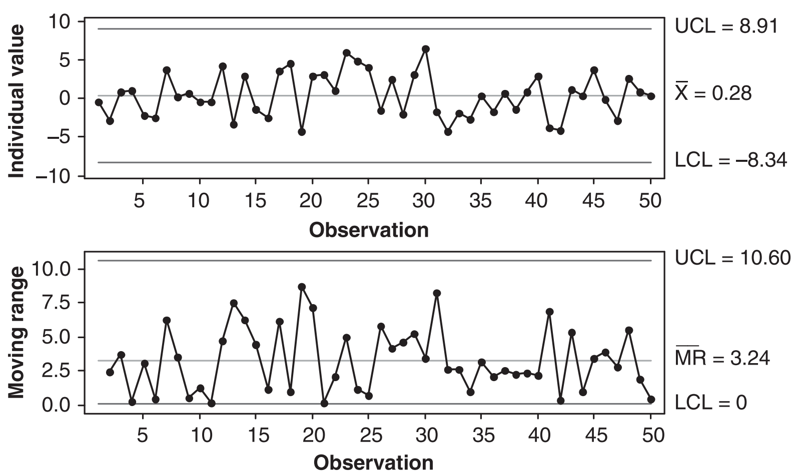


FIGURE 2.38 Individuals and moving range control charts of the one-step-ahead forecast errors in Table 2.3.

Because the control charts in Figure 2.38 exhibit statistical control, we would conclude that there is no strong evidence of statistical inadequacy in the forecasting model. Therefore, these control limits would be retained and used to judge the performance of future forecasts (in other words, we do not recalculate the control limits with each new forecast). However, the stable control chart does not imply that the forecasting performance is satisfactory in the sense that the model results in small forecast errors. In the quality control literature, these two aspects of process performance are referred to as control and capability, respectively. It is possible for the forecasting process to be stable or in statistical control but not capable—that is, produce forecast errors that are unacceptably large.

Two other types of control charts, the cumulative sum (or CUSUM) control chart and the exponentially weighted moving average (or EWMA) control chart, can also be useful for monitoring the performance of a forecasting model. These charts are more effective at detecting smaller changes or disturbances in the forecasting model performance than the individuals control chart. The CUSUM is very effective in detecting level changes in the monitored variable. It works by accumulating deviations of the forecast errors that are above the desired target value T_0 (usually either zero or the average forecast error) with one statistic C^+ and deviations that are below the target with another statistic C^-. The statistics C^+ and C^- are called the upper and lower CUSUMs, respectively. They are computed as follows:

$$\begin{aligned} C_t^+ &= \max[0, e_t(1) - (T_0 + K) + C_{t-1}^+] \\ C_t^- &= \min[0, e_t(1) - (T_0 - K) + C_{t-1}^-] \end{aligned}, \qquad (2.51)$$

where the constant K, usually called the reference value, is usually chosen as $K = 0.5\sigma_{e(1)}$ and $\sigma_{e(1)}$ is the standard deviation of the one-step-ahead forecast errors. The logic is that if the forecast errors begin to systematically fall on one side of the target value (or zero), one of the CUSUMs in Eq. (2.51) will increase in magnitude. When this increase becomes large enough, an out-of-control signal is generated. The decision rule is to signal if the statistic C^+ exceeds a decision interval $H = 5\sigma_{e(1)}$ or if C^- exceeds $-H$. The signal indicates that the forecasting model is not performing satisfactorily (Montgomery (2013) discusses the choice of H and K in detail).

Example 2.13 The CUSUM control chart for the forecast errors shown in Table 2.3 is shown in Figure 2.39. This CUSUM chart was constructed

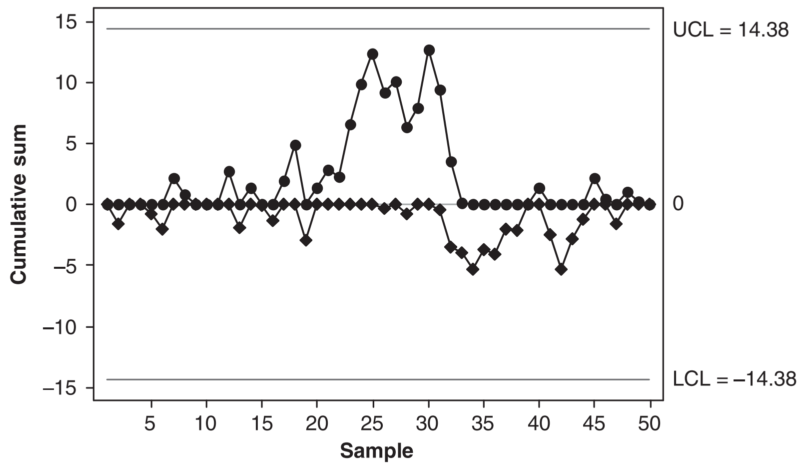


FIGURE 2.39 CUSUM control chart of the one-step-ahead forecast errors in Table 2.3.

using Minitab with a target value of $T = 0$ and $\sigma_{e(1)}$ was estimated using the moving range method described previously, resulting in $H = 5\hat{\sigma}_{e(1)} = 5(0.8865)MR/(T-1) = 5(0.8865)3.24 = 14.36$. Minitab labels H and $-H$ as UCL and LCL, respectively. The CUSUM control chart reveals no obvious forecasting model inadequacies.

A control chart based on the EWMA is also useful for monitoring forecast errors. The EWMA applied to the one-step-ahead forecast errors is

$$\bar{e}_t(1) = \lambda e_t(1) + (1-\lambda)\bar{e}_{t-1}(1), \tag{2.52}$$

where $0 < \lambda < 1$ is a constant (usually called the smoothing constant) and the starting value of the EWMA (required at the first observation) is either $\bar{e}_0(1) = 0$ or the average of the forecast errors. Typical values of the smoothing constant for an EWMA control chart are $0.05 < \lambda < 0.2$.

The EWMA is a weighted average of all current and previous forecast errors, and the weights decrease geometrically with the “age” of the forecast error. To see this, simply substitute recursively for $\bar{e}_{t-1}(1)$, then $\bar{e}_{t-2}(1)$, then $\bar{e}_{t-j}(1)_j$ for $j = 3, 4, \ldots$, until we obtain

$$\bar{e}_n(1) = \lambda \sum_{j=0}^{n-1} (1-\lambda)^j e_{T-j}(1) + (1-\lambda)^n \bar{e}_0(1)$$

and note that the weights sum to unity because

$$\lambda \sum_{j=0}^{n-1} (1-\lambda)^j = 1-(1-\lambda)^n.$$

The standard deviation of the EWMA is

$$\sigma_{\bar{e}_t(1)} = \sigma_{e(1)} \sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]}.$$

So an EWMA control chart for the one-step-ahead forecast errors with a center line of T (the target for the forecast errors) is defined as follows:

$$\begin{aligned} \text{UCL} &= T + 3\sigma_{e(1)}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]} \\ \text{Center line} &= T \\ \text{LCL} &= T - 3\sigma_{e(1)}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2t}]} \end{aligned} \tag{2.53}$$

Example 2.14 Minitab can be used to construct EWMA control charts. Figure 2.40 is the EWMA control chart of the forecast errors in Table 2.3. This chart uses the mean forecast error as the center line, $\sigma_{e(1)}$ was estimated using the moving range method, and we chose $\lambda = 0.1$. None of the forecast

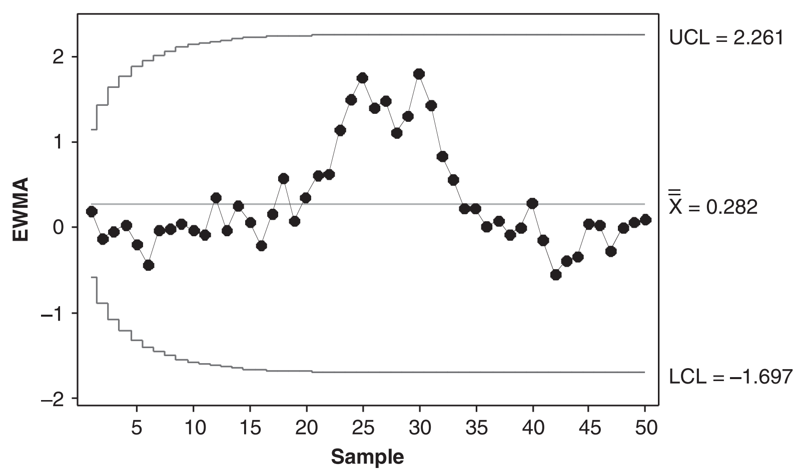


FIGURE 2.40 EWMA control chart of the one-step-ahead forecast errors in Table 2.3.

errors exceeds the control limits so there is no indication of a problem with the forecasting model.

Note from Eq. (2.51) and Figure 2.40 that the control limits on the EWMA control chart increase in width for the first few observations and then stabilize at a constant value because the term $[1 - (1 - \lambda)^{2t}]$ approaches unity as t increases. Therefore steady-state limits for the EWMA control chart are

$$\begin{aligned} \text{UCL} &= T_0 + 3\sigma_{e(1)}\sqrt{\frac{\lambda}{2-\lambda}} \\ \text{Center line} &= T \\ \text{LCL} &= T_0 - 3\sigma_{e(1)}\sqrt{\frac{\lambda}{2-\lambda}}. \end{aligned} \tag{2.54}$$

In addition to control charts, other statistics have been suggested for monitoring the performance of a forecasting model. The most common of these are **tracking signals**. The cumulative error tracking signal (CETS) is based on the cumulative sum of all current and previous forecast errors, say,

$$Y(n) = \sum_{t=1}^{n} e_t(1) = Y(n-1) + e_n(1).$$

If the forecasts are unbiased, we would expect $Y(n)$ to fluctuate around zero. If it differs from zero by very much, it could be an indication that the forecasts are biased. The standard deviation of $Y(n)$, say, $\sigma_{Y(n)}$, will provide a measure of how far $Y(n)$ can deviate from zero due entirely to random variation. Therefore, we would conclude that the forecast is biased if $|Y(n)|$ exceeds some multiple of its standard deviation. To operationalize this, suppose that we have an estimate $\hat{\sigma}_{Y(n)}$ of $\sigma_{Y(n)}$ and form the **cumulative error tracking signal**

$$\text{CETS} = \left|\frac{Y(n)}{\hat{\sigma}_{Y(n)}}\right|. \tag{2.55}$$

If the CETS exceeds a constant, say, K_1, we would conclude that the forecasts are biased and that the forecasting model may be inadequate.

It is also possible to devise a **smoothed error tracking signal** based on the smoothed one-step-ahead forecast errors in Eq. (2.52). This would lead to a ratio

$$\text{SETS} = \left|\frac{\bar{e}_n(1)}{\hat{\sigma}_{\bar{e}_n(1)}}\right|. \tag{2.56}$$

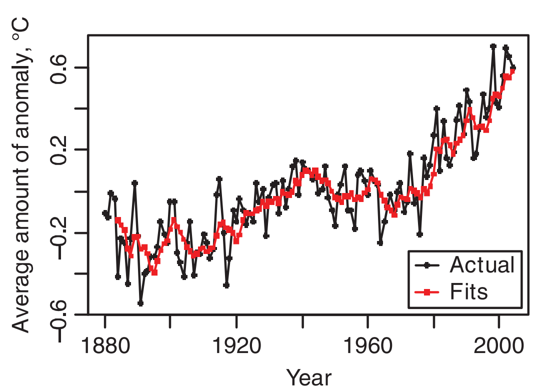
If the SETS exceeds a constant, say, K_2, this is an indication that the forecasts are biased and that there are potentially problems with the forecasting model.

Note that the CETS is very similar to the CUSUM control chart and that the SETS is essentially equivalent to the EWMA control chart. Furthermore, the CUSUM and EWMA are available in standard statistics software (such as Minitab) and the tracking signal procedures are not. So, while tracking signals have been discussed extensively and recommended by some authors, we are not going to encourage their use. Plotting and periodically visually examining a control chart of forecast errors is also very informative, something that is not typically done with tracking signals.

2.7 R COMMANDS FOR CHAPTER 2

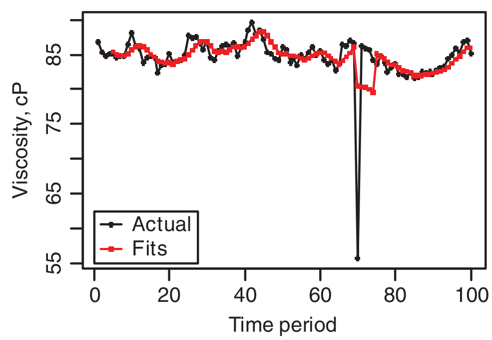
Example 2.15 The data are in the second column of the array called gms.data in which the first column is the year. For moving averages, we use functions from package "zoo."

```
plot(gms.data,type="l",xlab='Year',ylab='Average Amount of
Anomaly, °C')
points(gms.data,pch=16,cex=.5)
lines(gms.data[5:125,1],rollmean(gms.data[,2],5),col="red")
points(gms.data[5:125,1],rollmean(gms.data[,2],5),col="red",pch=15,
cex=.5)
legend(1980,-.3,c("Actual","Fits"), pch=c(16,15),lwd=c(.5,.5),
cex=.55,col=c("black","red"))
```

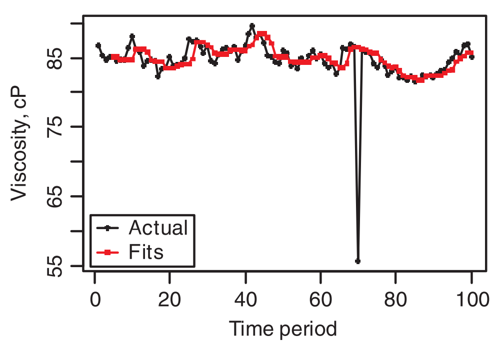


Example 2.16 The data are in the second column of the array called vis.data in which the first column is the time period (or index).

```
# Moving Average
plot(vis.data,type="l",xlab='Time Period',ylab='Viscosity, cP')
points(vis.data,pch=16,cex=.5)
lines(vis.data[5:100,1], rollmean(vis.data[,2],5),col="red")
points(vis.data[5:100,1], rollmean(vis.data[,2],5),col="red",
pch=15,cex=.5)
legend(1,61,c("Actual","Fits"), pch=c(16,15),lwd=c(.5,.5),cex=.55,
col=c("black","red"))
```



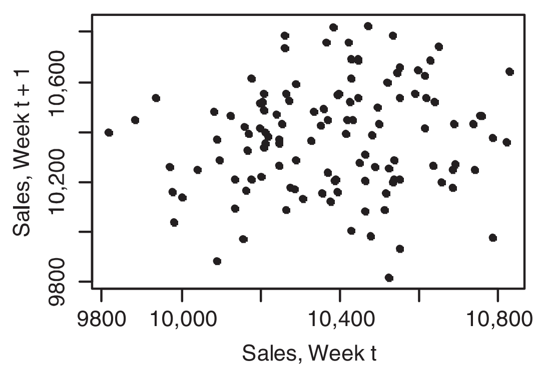

```
# Moving Median
plot(vis.data,type="l",xlab='Time Period',ylab='Viscosity, cP')
points(vis.data,pch=16,cex=.5)
lines(vis.data[5:100,1], rollmedian(vis.data[,2],5),col="red")
points(vis.data[5:100,1], rollmedian(vis.data[,2],5),col="red",
pch=15,cex=.5)
legend(1,61,c("Actual","Fits"), pch=c(16,15),lwd=c(.5,.5),cex=.55,
col=c("black","red"))
```

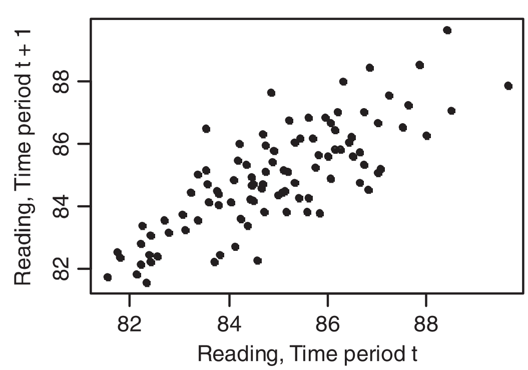


Example 2.17 The pharmaceutical sales data are in the second column of the array called pharma.data in which the first column is the week.

The viscosity data are in the second column of the array called vis.data in which the first column is the year (Note that the 70th observation is corrected).

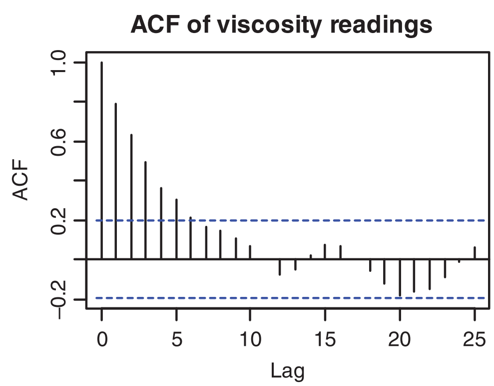
```
nrp<-dim(pharma.data)[1]

nrv<-dim(vis.data)[1]

plot(pharma.data[1:(nrp-1),2], pharma.data[2:nrp,2],type="p",
xlab='Sales, Week t',ylab=' Sales, Week t+1',pch=20,cex=1)

plot(vis.data[1:(nrv-1),2], vis.data[2:nrv,2],type="p", xlab=
'Reading, Time Period t',ylab=' Reading, Time Period t+1',pch=20,
cex=1)
```





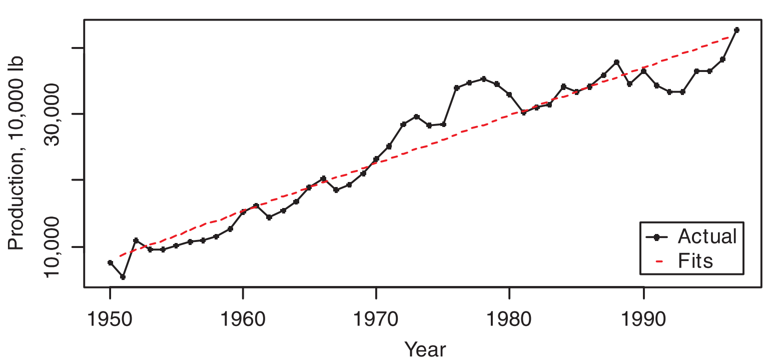
Example 2.18 The viscosity data are in the second column of the array called vis.data in which the first column is the year (Note that the 70th observation is corrected).

```
acf(vis.data[,2], lag.max=25,type="correlation",main="ACF of
viscosity readings")
```

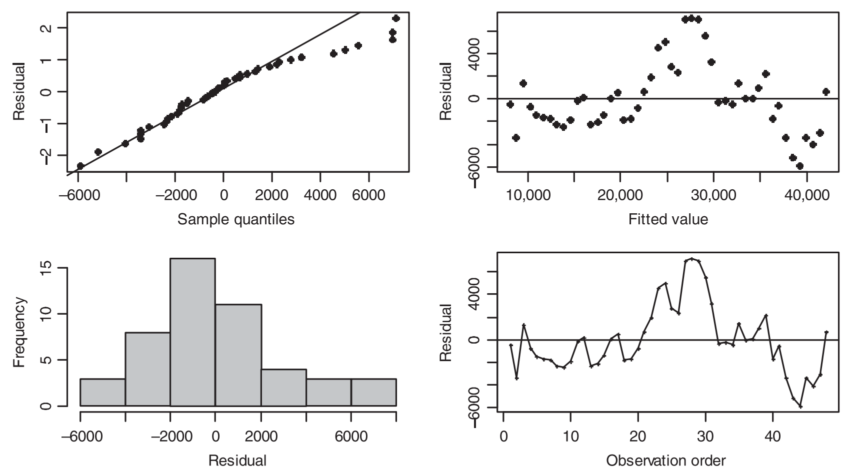


Example 2.19 The cheese production data are in the second column of the array called cheese.data in which the first column is the year.

```
fit.cheese<-lm(cheese.data[,2]~cheese.data[,1])
plot(cheese.data,type="l",xlab='Year',ylab='Production, 10000lb')
points(cheese.data,pch=16,cex=.5)
lines(cheese.data[,1], fit.cheese$fit,col="red",lty=2)
legend(1990,12000,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.55,col=c("black","red"))
```

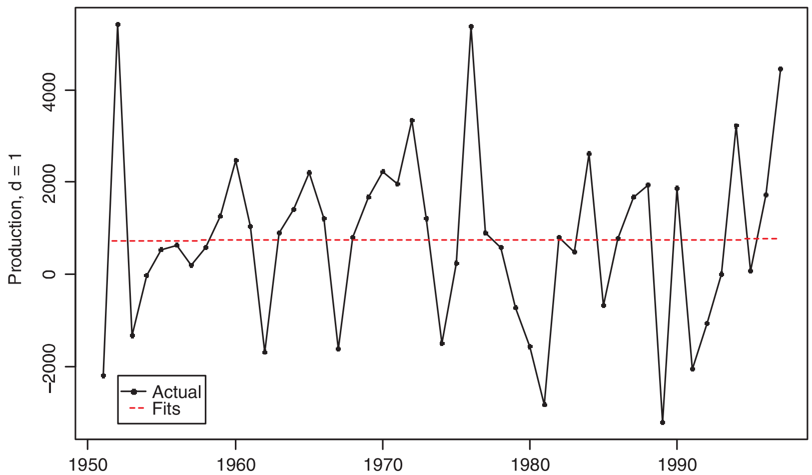


```
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(fit.cheese$res,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(fit.cheese$res,datax=TRUE)
plot(fit.cheese$fit,fit.cheese$res,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(fit.cheese$res,col="gray",xlab='Residual',main='')
plot(fit.cheese$res,type="l",xlab='Observation Order',
ylab='Residual')
points(fit.cheese$res,pch=16,cex=.5)
abline(h=0)
```

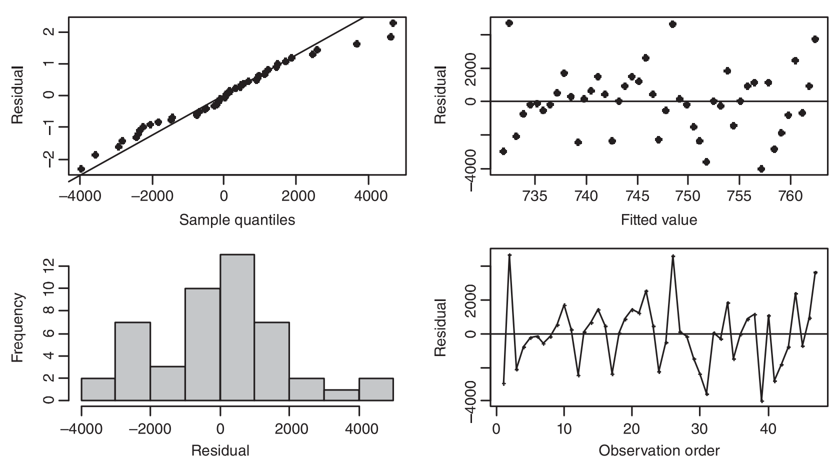


Example 2.20 The cheese production data are in the second column of the array called cheese.data in which the first column is the year.

```
nrc<-dim(cheese.data)[1]
dcheese.data<-cbind(cheese.data[2:nrc,1],diff(cheese.data[,2]))
fit.dcheese<-lm(dcheese.data[,2]~dcheese.data[,1])
plot(dcheese.data,type="l",xlab='',ylab='Production, d=1')
points(dcheese.data,pch=16,cex=.5)
lines(dcheese.data[,1], fit.dcheese$fit,col="red",lty=2)
legend(1952,-2200,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),
cex=.75,col=c("black","red"))
```

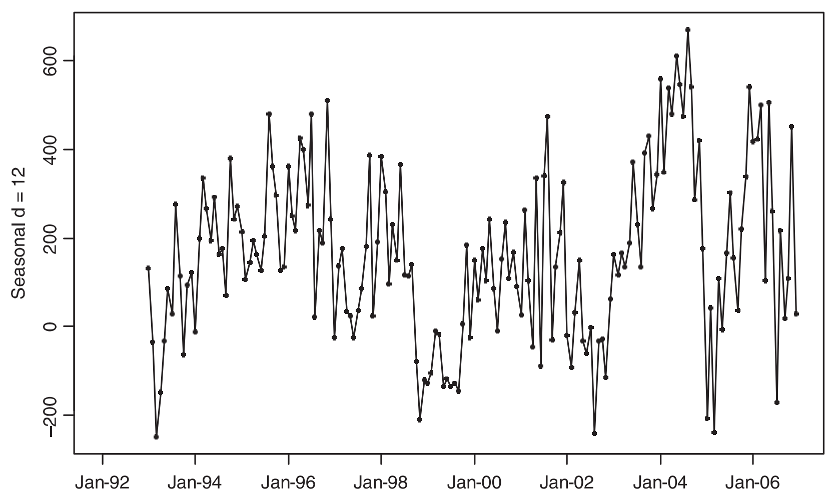



```
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(fit.dcheese$res,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(fit.dcheese$res,datax=TRUE)
plot(fit.dcheese$fit,fit.dcheese$res,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(fit.dcheese$res,col="gray",xlab='Residual',main='')
plot(fit.dcheese$res,type="l",xlab='Observation Order',
ylab='Residual')
points(fit.dcheese$res,pch=16,cex=.5)
abline(h=0)
```

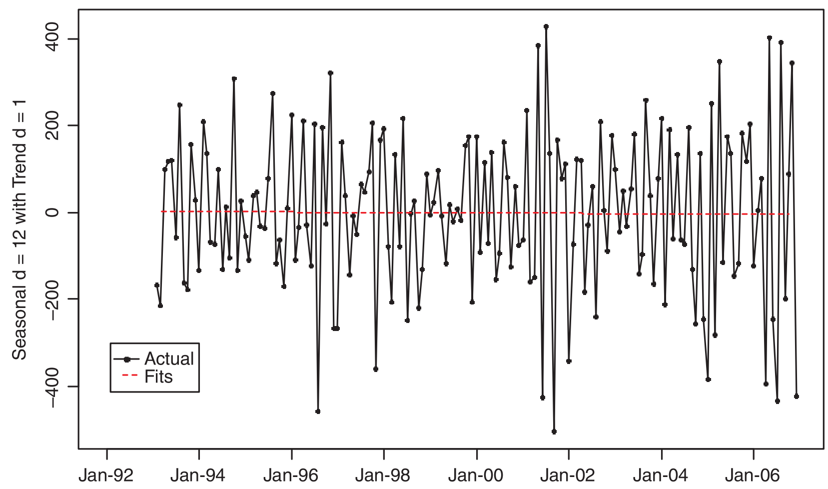


Example 2.21 The beverage sales data are in the second column of the array called bev.data in which the first column is the month of the year.

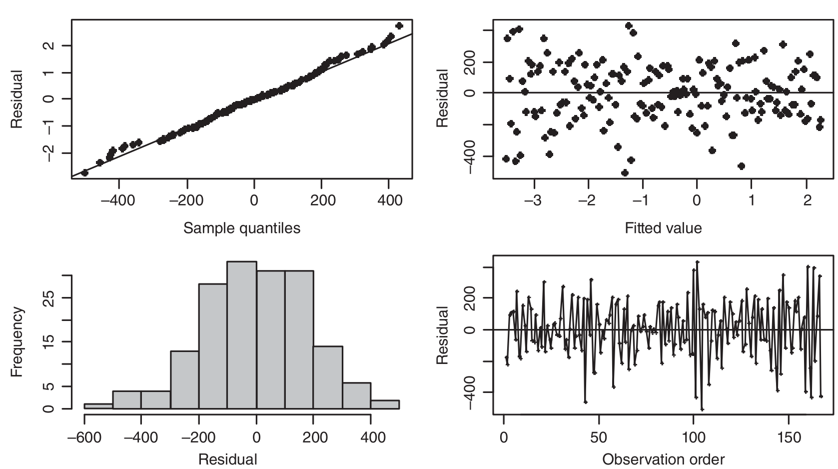
```
nrb<-dim(bev.data)[1]
tt<-1:nrb
dsbev.data<-bev.data
dsbev.data[,2]<- c(array(NA,dim=c(12,1)),diff(bev.data[,2],12))

plot(tt,dsbev.data[,2],type="l",xlab='',ylab='Seasonal d=12',
xaxt='n')axis(1,seq(1,nrb,24),labels=dsbev.data[seq(1,nrb,24),1])
points(tt,dsbev.data[,2],pch=16,cex=.5)
```




```
dstbev.data<-dsbev.data
dstbev.data[,2]<- c(NA,diff(dstbev.data[,2],1))
fit.dstbev<-lm(dstbev.data[,2]~tt)
plot(tt,dstbev.data[,2],type="l",xlab='',ylab='Seasonal d=12 with
Trend d=1',xaxt='n')
axis(1,seq(1,nrb,24),labels=dsbev.data[seq(1,nrb,24),1])
points(tt,dstbev.data[,2],pch=16,cex=.5)
lines(c(array(NA,dim=c(12,1)),fit.dstbev$fit),col="red",lty=2)
legend(2,-300,c("Actual","Fits"),
pch=c(16,NA),lwd=c(.5,.5),lty=c(1,2),cex=.75,col=c("black","red"))
```




```
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(fit.dstbev$res,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(fit.dstbev$res,datax=TRUE)
plot(fit.dstbev$fit,fit.dstbev$res,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(fit.dstbev$res,col="gray",xlab='Residual',main='')
plot(fit.dstbev$res,type="l",xlab='Observation Order',
ylab='Residual')
points(fit.dstbev$res,pch=16,cex=.5)
abline(h=0)
```

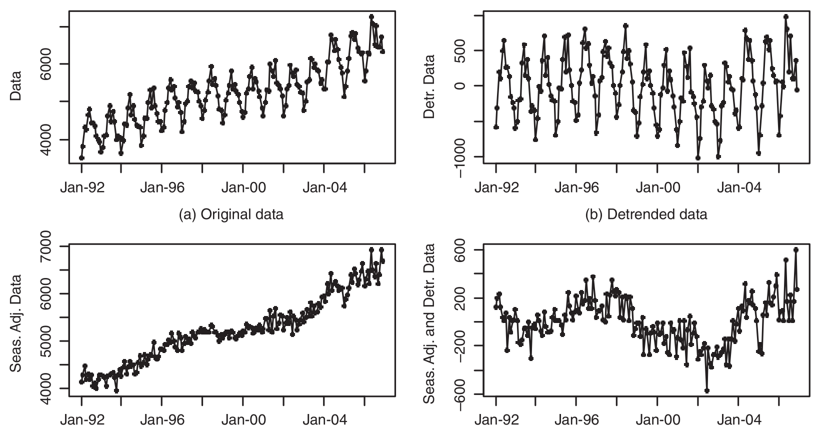


Example 2.22 The beverage sales data are in the second column of the array called bev.data in which the first column is the month of the year.

Software packages use different methods for decomposing a time series. Below we provide the code of doing it in R without using these functions. Note that we use the additive model.

```
nrb<-dim(bev.data)[1]

# De-trend the data
tt<-1:nrb
fit.tbev<-lm(bev.data[,2]~tt)
bev.data.dt<-fit.tbev$res

# Obtain seasonal medians for each month, seasonal period is sp=12
sp<-12
smed<-apply(matrix(bev.data.dt,nrow=sp),1,median)

# Adjust the medians so that their sum is zero
smed<-smed-mean(smed)

# Data without the trend and seasonal components
bev.data.dts<-bev.data.dt-rep(smed,nrb/sp)

# Note that we can also reverse the order, i.e. first take the
 seasonality out
smed2<-apply(matrix(bev.data[,2],nrow=sp),1,median)
smed2<-smed2-mean(smed2)
bev.data.ds<-bev.data[,2]-rep(smed2,nrb/sp)

# To reproduce Figure 2.25

par(mfrow=c(2,2),oma=c(0,0,0,0))
plot(tt,bev.data[,2],type="l",xlab='(a) Original Data',ylab=
'Data',xaxt='n')
axis(1,seq(1,nrb,24),labels=bev.data[seq(1,nrb,24),1])
points(tt,bev.data[,2],pch=16,cex=.75)

plot(tt, bev.data.dt,type="l",xlab='(b) Detrended Data',ylab='Detr.
Data',xaxt='n')
axis(1,seq(1,nrb,24),labels=bev.data[seq(1,nrb,24),1])

points(tt, bev.data.dt,pch=16,cex=.75)
plot(tt, bev.data.ds,type="l",xlab='(c) Seasonally Adjusted Data',
ylab='Seas.
Adj. Data',xaxt='n')
axis(1,seq(1,nrb,24),labels=bev.data[seq(1,nrb,24),1])

points(tt, bev.data.ds,pch=16,cex=.75)
```

```
plot(tt, bev.data.dts,type="l",xlab='(c) Seasonally Adj. and
Detrended Data',ylab='Seas. Adj. and Detr. Data',xaxt='n')
axis(1,seq(1,nrb,24),labels=bev.data[seq(1,nrb,24),1]) points(tt,
bev.data.dts, pch=16,cex=.75)
```



(a) Original data

(b) Detrended data

(c) Seasonally adjusted data

(c) Seasonally Adj. and detrended data

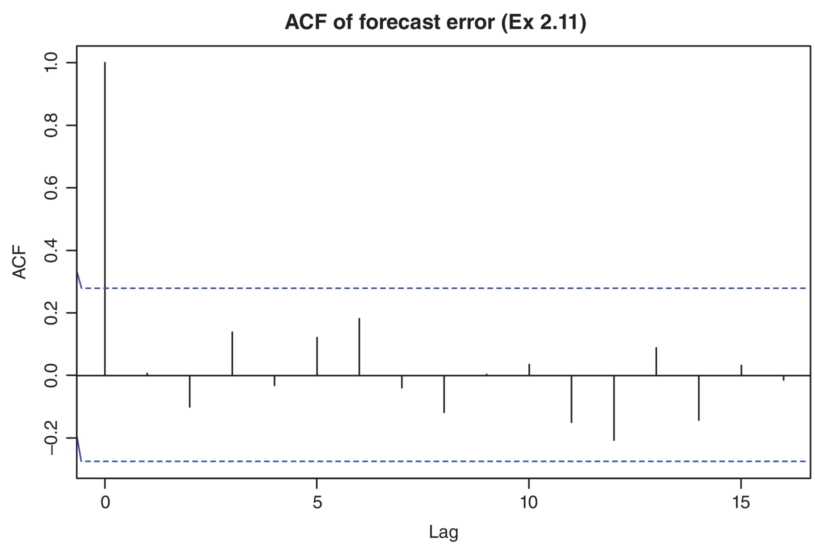
Example 2.23 Functions used to fit a time series model often also provide summary statistics. However, in this example we provide some calculations for a given set of forecast errors as provided in the text.

```
# original data and forecast errors
yt<-c(47,46,51,44,54,47,52,45,50,51,49,41,48,50,51,55,52,53,48,52)
fe<-c(-4.1,-6.9,2.2,-4.1,4.3,-.5,.8,-8.1,-4.4,-.2,-4.3,-5.5,-5.1,
-2.1,4.2,7.3,6.6,5.9,-3.8,6.2)

ME<-mean(fe)
MAD<-mean(abs(fe))
MSE<-mean(fe^2)
ret1<-(fe/yt)*100
MPE<-mean(ret1)
MAPE<-mean(abs(ret1))

> ME
[1] -0.58
> MAD
[1] 4.33
> MSE
[1] 23.59
```

```
> MPE
[1] -1.757938
> MAPE
[1] 8.865001
```

Example 2.24 The forecast error data are in the second column of the array called fe2.data in which the first column is the period.

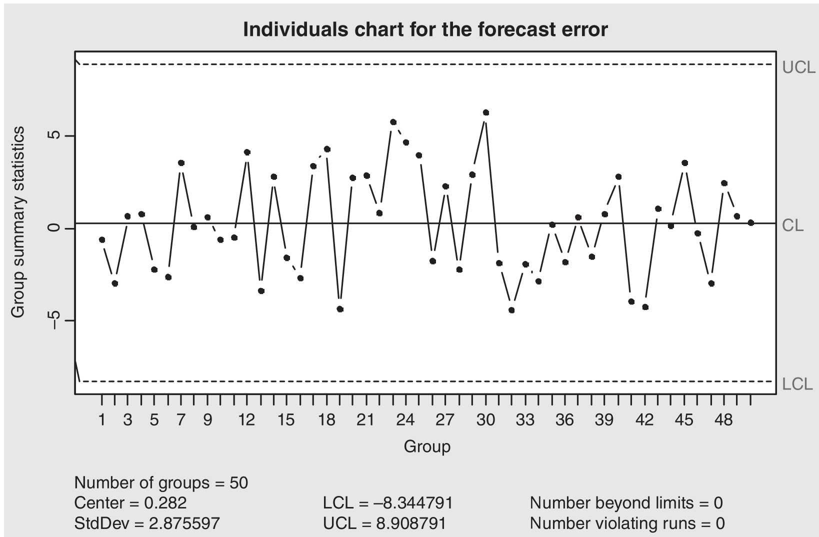
```
acf.fe2<-acf(fe2.data[,2],main='ACF of Forecast Error (Ex 2.11)')
```




```
# To get the Q_LB statistic, we first define the lag K

K<-13
T<-dim(fe2.data)[1]
QLB<-T*(T+2)*sum((1/(T-1:K))*(acf.fe2$acf[2:(K+1)]^2))

# Upper 5% of χ^2 distribution with K degrees of freedom
qchisq(.95,K)
```

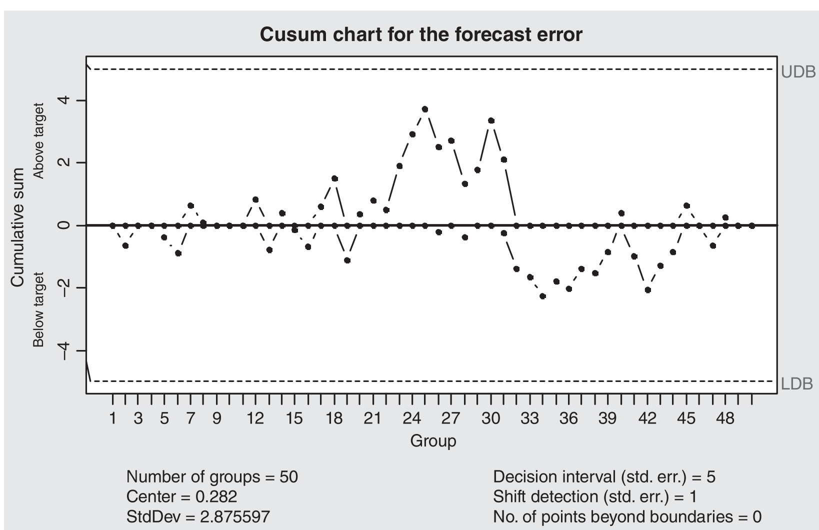
Example 2.25 The forecast error data are in the second column of the array called fe2.data in which the first column is the period.

```
# The following function can be found in qcc package
# Generating the chart for individuals
qcc(fe2.data[,2],type="xbar.one",title="Individuals Chart for the
Forecast Error")
```



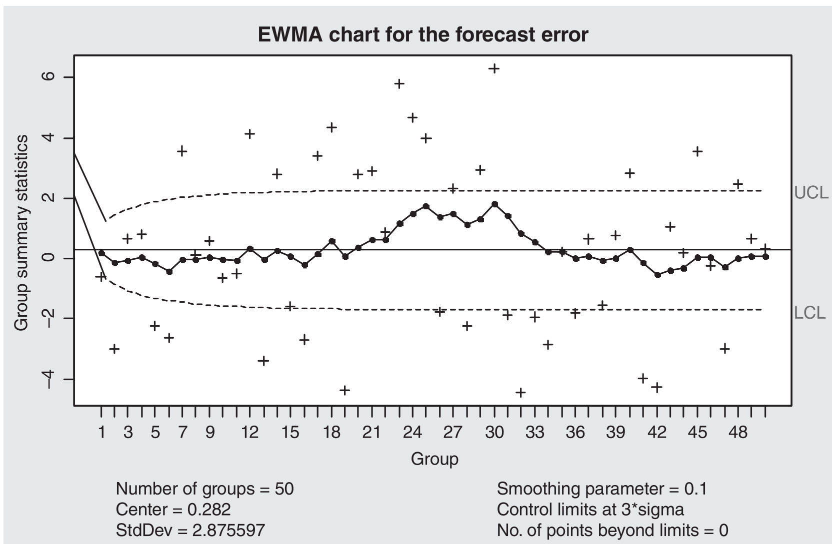
Example 2.26 The forecast error data are in the second column of the array called fe2.data in which the first column is the period.

```
# The following function can be found in qcc package
# Generating the cusum chart

cusum(fe2.data[,2], title='Cusum Chart for the Forecast
Error', sizes=1)
```



Example 2.27 The forecast error data are in the second column of the array called fe2.data in which the first column is the period.

```
# The following function can be found in qcc package
# Generating the EWMA chart
ewma(fe2.data[,2], title='EWMA Chart for the Forecast Error',
lambda=.1,sizes=1)
```



EXERCISES

2.1 Consider the US Treasury Securities rate data in Table B.1 (Appendix B). Find the sample autocorrelation function and the variogram for these data. Is the time series stationary or nonstationary?

2.2 Consider the data on US production of blue and gorgonzola cheeses in Table B.4.

a. Find the sample autocorrelation function and the variogram for these data. Is the time series stationary or nonstationary?

b. Take the first difference of the time series, then find the sample autocorrelation function and the variogram. What conclusions can you draw about the structure and behavior of the time series?

2.3 Table B.5 contains the US beverage product shipments data. Find the sample autocorrelation function and the variogram for these data. Is the time series stationary or nonstationary?

2.4 Table B.6 contains two time series: the global mean surface air temperature anomaly and the global CO_2 concentration. Find the sample autocorrelation function and the variogram for both of these time series. Is either one of the time series stationary?

2.5 Reconsider the global mean surface air temperature anomaly and the global CO_2 concentration time series from Exercise 2.4. Take the first difference of both time series. Find the sample autocorrelation function and variogram of these new time series. Is either one of these differenced time series stationary?

2.6 Find the closing stock price for a stock that interests you for the last 200 trading days. Find the sample autocorrelation function and the variogram for this time series. Is the time series stationary?

2.7 Reconsider the Whole Foods Market stock price data from Exercise 2.6. Take the first difference of the data. Find the sample autocorrelation function and the variogram of this new time series. Is this differenced time series stationary?

2.8 Consider the unemployment rate data in Table B.8. Find the sample autocorrelation function and the variogram for this time series. Is the time series stationary or nonstationary? What conclusions can you draw about the structure and behavior of the time series?

2.9 Table B.9 contains the annual International Sunspot Numbers. Find the sample autocorrelation function and the variogram for this time series. Is the time series stationary or nonstationary?

2.10 Table B.10 contains data on the number of airline miles flown in the United Kingdom. This is strongly seasonal data. Find the sample autocorrelation function for this time series.

a. Is the seasonality apparent in the sample autocorrelation function?

b. Is the time series stationary or nonstationary?

2.11 Reconsider the data on the number of airline miles flown in the United Kingdom from Exercise 2.10. Take the natural logarithm of the data and plot this new time series.

a. What impact has the log transformation had on the time series?

b. Find the autocorrelation function for this time series.

c. Interpret the sample autocorrelation function.

2.12 Reconsider the data on the number of airline miles flown in the United Kingdom from Exercises 2.10 and 2.11. Take the first difference of the natural logarithm of the data and plot this new time series.

a. What impact has the log transformation had on the time series?

b. Find the autocorrelation function for this time series.

c. Interpret the sample autocorrelation function.

2.13 The data on the number of airline miles flown in the United Kingdom in Table B.10 are seasonal. Difference the data at a season lag of 12 months and also apply a first difference to the data. Plot the differenced series. What effect has the differencing had on the time series? Find the sample autocorrelation function and the variogram. What does the sample autocorrelation function tell you about the behavior of the differenced series?

2.14 Table B.11 contains data on the monthly champagne sales in France. This is strongly seasonal data. Find the sample autocorrelation function and variogram for this time series.

a. Is the seasonality apparent in the sample autocorrelation function?

b. Is the time series stationary or nonstationary?

2.15 Reconsider the champagne sales data from Exercise 2.14. Take the natural logarithm of the data and plot this new time series.

a. What impact has the log transformation had on the time series?

b. Find the autocorrelation function and variogram for this time series.

c. Interpret the sample autocorrelation function and variogram.

2.16 Table B.13 contains data on ice cream and frozen yogurt production. Plot the data and calculate both the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the time series and compute the sample autocorrelation function and variogram of the first differences. What impact has differencing had on the time series?

2.17 Table B.14 presents data on CO_2 readings from the Mauna Loa Observatory. Plot the data, then calculate the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the time series and compute the sample autocorrelation function and the variogram of the first differences. What impact has differencing had on the time series?

2.18 Data on violent crime rates are given in Table B.15. Plot the data and calculate the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the time series and compute the sample autocorrelation function and variogram of the first differences. What impact has differencing had on the time series?

2.19 Table B.16 presents data on the US Gross Domestic Product (GDP). Plot the GDP data and calculate the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the GDP time series and compute the sample autocorrelation function and variogram of the first differences. What impact has differencing had on the time series?

2.20 Table B.17 contains information on total annual energy consumption. Plot the energy consumption data and calculate the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the time series and compute the sample autocorrelation function and variogram of the first differences. What impact has differencing had on the time series?

2.21 Data on US coal production are given in Table B.18. Plot the coal production data and calculate the sample autocorrelation function and variogram. Is there an indication of nonstationary behavior in the time series? Now plot the first difference of the time series and compute the sample autocorrelation function and variogram of the first differences. What impact has differencing had on the time series?

2.22 Consider the CO_2 readings from Mauna Loa in Table B.14. Use a six-period moving average to smooth the data. Plot both the smoothed data and the original CO_2 readings on the same axes. What has the moving average done? Repeat the procedure with a three-period moving average. What is the effect of changing the span of the moving average?

2.23 Consider the violent crime rate data in Table B.15. Use a ten-period moving average to smooth the data. Plot both the smoothed data and the original CO_2 readings on the same axes. What has the moving average done? Repeat the procedure with a four-period moving average. What is the effect of changing the span of the moving average?

2.24 Table B.21 contains data from the US Energy Information Administration on monthly average price of electricity for the residential sector in Arizona. Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

2.25 Reconsider the residential electricity price data from Exercise 2.24.

a. Plot the first difference of the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram for the differenced data. Interpret these graphs. What impact did differencing have?

b. Now difference the data again at a seasonal lag of 12. Plot the differenced data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram for the differenced data. Interpret these graphs. What impact did regular differencing combined with seasonal differencing have?

2.26 Table B.22 contains data from the Danish Energy Agency on Danish crude oil production. Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

2.27 Reconsider the Danish crude oil production data from Exercise 2.26. Plot the first difference of the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram for the differenced data. Interpret these graphs. What impact did differencing have?

2.28 Use a six-period moving average to smooth the first difference of the Danish crude oil production data that you computed in Exercise 2.27. Plot both the smoothed data and the original data on the same axes. What has the moving average done? Does the moving average look like a reasonable forecasting technique for the differenced data?

2.29 Weekly data on positive laboratory test results for influenza are shown in Table B.23. Notice that these data have a number of missing values. Construct a time series plot of the data and comment on any relevant features that you observe.

a. What is the impact of the missing observations on your ability to model and analyze these data?

b. Develop and implement a scheme to estimate the missing values

2.30 Climate data collected from Remote Automated Weather Stations (RAWS) are used to monitor the weather and to assist land management agencies with projects such as monitoring air quality, rating fire danger, and other research purposes. Data from the Western Regional Climate Center for the mean daily solar radiation (in Langleys) at the Zion Canyon, Utah, station are shown in Table B.24.

a. Plot the data and comment on any features that you observe.

b. Calculate and plot the sample ACF and variogram. Comment on the plots.

c. Apply seasonal differencing to the data, plot the data, and construct the sample ACF and variogram. What was the impact of seasonal differencing?

2.31 Table B.2 contains annual US motor vehicle traffic fatalities along with other information. Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

2.32 Reconsider the motor vehicle fatality data from Exercise 2.31.

a. Plot the first difference of the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram for the differenced data. Interpret these graphs. What impact did differencing have?

b. Compute a six-period moving average for the differenced data. Plot the moving average and the original data on the same axes. Does it seem that the six-period moving average would be a good forecasting technique for the differenced data?

2.33 Apply the X-11 seasonal decomposition method (or any other seasonal adjustment technique for which you have software) to the mean daily solar radiation in Table B.24.

2.34 Consider the N-span moving average applied to data that are uncorrelated with mean μ and variance σ^2.

a. Show that the variance of the moving average is $\text{Var}(M_t) = \sigma^2/N$.

b. Show that $\mathrm{Cov}(M_t, M_{t+k}) = \sigma^2 \sum_{j=1}^{N-k} (1/N)^2$, for $k < N$.

c. Show that the autocorrelation function is

$$\rho_k = \begin{cases} 1 - \dfrac{|k|}{N}, & k = 1, 2, \ldots, N-1 \\ 0, & k \geq N \end{cases}$$

2.35 Consider an N-span moving average where each observation is weighted by a constant, say, $a_j \geq 0$. Therefore the weighted moving average at the end of period T is

$$M_T^w = \sum_{t=T-N+1}^{T} a_{T+1-t} y_t.$$

a. Why would you consider using a weighted moving average?

b. Show that the variance of the weighted moving average is $\mathrm{Var}(M_T^w) = \sigma^2 \sum_{j=i}^{N} a_j^2$.

c. Show that $\mathrm{Cov}(M_T^w, M_{T+k}^w) = \sigma^2 \sum_{j=1}^{N-k} a_j a_{j+k},\ |k| < N$.

d. Show that the autocorrelation function is

$$\rho_k = \begin{cases} \left(\sum_{j=1}^{N-k} a_j a_{j+k} \right) \Big/ \left(\sum_{j=1}^{N} a_j^2 \right), & k = 1, 2, \ldots, N-1 \\ 0, & k \geq N \end{cases}$$

2.36 Consider the Hanning filter. This is a weighted moving average.

a. Find the variance of the weighted moving average for the Hanning filter. Is this variance smaller than the variance of a simple span-3 moving average with equal weights?

b. Find the autocorrelation function for the Hanning filter. Compare this with the autocorrelation function for a simple span-3 moving average with equal weights.

2.37 Suppose that a simple moving average of span N is used to forecast a time series that varies randomly around a constant, that is, $y_t = \mu + \varepsilon_t$, where the variance of the error term is σ^2. The forecast error at lead one is $e_{T+1}(1) = y_{T+1} - M_T$. What is the variance of this lead-one forecast error?

2.38 Suppose that a simple moving average of span N is used to forecast a time series that varies randomly around a constant, that is,

$y_t = \mu + \varepsilon_t$, where the variance of the error term is σ^2. You are interested in forecasting the cumulative value of y over a lead time of L periods, say, $y_{T+1} + y_{T+2} + \cdots + y_{T+L}$.

a. The forecast of this cumulative demand is LM_T. Why?

b. What is the variance of the cumulative forecast error?

2.39 Suppose that a simple moving average of span N is used to forecast a time series that varies randomly around a constant mean, that is, $y_t = \mu + \varepsilon_t$. At the start of period t_1 the process shifts to a new mean level, say, $\mu + \delta$. Show that the expected value of the moving average is

$$E(M_T) = \begin{cases} \mu, & T \le t_1 - 1 \\ \mu + \dfrac{T - t_1 + 1}{N}\delta, & t_1 \le T \le t_1 + N - 2 \\ \mu + \delta, & T \ge t_1 + N - 1 \end{cases}.$$

2.40 Suppose that a simple moving average of span N is used to forecast a time series that varies randomly around a constant mean, that is, $y_t = \mu + \varepsilon_t$. At the start of period t_1 the process experiences a transient; that is, it shifts to a new mean level, say, $\mu + \delta$, but it reverts to its original level μ at the start of period $t_1 + 1$. Show that the expected value of the moving average is

$$E(M_T) = \begin{cases} \mu, & T \le t_1 - 1 \\ \mu + \dfrac{\delta}{N}, & t_1 \le T \le t_1 + N - 1 \\ \mu, & T \ge t_1 + N \end{cases}.$$

2.41 If a simple $N-$span moving average is applied to a time series that has a linear trend, say, $y_t = \beta_0 + \beta_1 t + \varepsilon_t$, the moving average will lag behind the observations. Assume that the observations are uncorrelated and have constant variance. Show that at time T the expected value of the moving average is

$$E(M_T) = \beta_0 + \beta_1 T - \frac{N-1}{2}\beta_1.$$

2.42 Use a three-period moving average to smooth the champagne sales data in Table B.11. Plot the moving average on the same axes as the original data. What impact has this smoothing procedure had on the data?

TABLE E2.1 One-Step-Ahead Forecast Errors for Exercise 2.44

Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$
1	1.83	11	−2.30	21	3.30	31	−0.07
2	−1.80	12	0.65	22	1.036	32	0.57
3	0.09	13	−0.01	23	2.042	33	2.92
4	−1.53	14	−1.11	24	1.04	34	1.99
5	−0.58	15	0.13	25	−0.87	35	1.74
6	0.21	16	−1.07	26	−0.39	36	−0.76
7	1.25	17	0.80	27	−0.29	37	2.35
8	−1.22	18	−1.98	28	2.08	38	−1.91
9	1.32	19	0.02	29	3.36	39	2.22
10	3.63	20	0.25	30	−0.53	40	2.57

2.43 Use a 12-period moving average to smooth the champagne sales data in Table B.11. Plot the moving average on the same axes as the original data. What impact has this smoothing procedure had on the data?

2.44 Table E2.1 contains 40 one-step-ahead forecast errors from a forecasting model.

a. Find the sample ACF of the forecast errors. Interpret the results.

b. Construct a normal probability plot of the forecast errors. Is there evidence to support a claim that the forecast errors are normally distributed?

c. Find the mean error, the mean squared error, and the mean absolute deviation. Is it likely that the forecasting technique produces unbiased forecasts?

2.45 Table E2.2 contains 40 one-step-ahead forecast errors from a forecasting model.

a. Find the sample ACF of the forecast errors. Interpret the results.

b. Construct a normal probability plot of the forecast errors. Is there evidence to support a claim that the forecast errors are normally distributed?

c. Find the mean error, the mean squared error, and the mean absolute deviation. Is it likely that the forecasting method produces unbiased forecasts?

2.46 Exercises 2.44 and 2.45 present information on forecast errors. Suppose that these two sets of forecast errors come from two different

TABLE E2.2 One-Step-Ahead Forecast Errors for Exercise 2.45

Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$	Period, t	$e_t(1)$
1	−4.26	11	3.62	21	−6.24	31	−6.42
2	−3.12	12	−5.08	22	−0.25	32	−8.94
3	−1.87	13	−1.35	23	−3.64	33	−1.76
4	0.98	14	3.46	24	5.49	34	−0.57
5	−5.17	15	−0.19	25	−2.01	35	−10.32
6	0.13	16	−7.48	26	−4.24	36	−5.64
7	1.85	17	−3.61	27	−4.61	37	−1.45
8	−2.83	18	−4.21	28	3.24	38	−5.67
9	0.95	19	−6.49	29	−8.66	39	−4.45
10	7.56	20	4.03	30	−1.32	40	−10.23

forecasting methods applied to the same time series. Which of these two forecasting methods would you recommend for use? Why?

2.47 Consider the forecast errors in Exercise 2.44. Construct individuals and moving range control charts for these forecast errors. Does the forecasting system exhibit stability over this time period?

2.48 Consider the forecast errors in Exercise 2.44. Construct a cumulative sum control chart for these forecast errors. Does the forecasting system exhibit stability over this time period?

2.49 Consider the forecast errors in Exercise 2.45. Construct individuals and moving range control charts for these forecast errors. Does the forecasting system exhibit stability over this time period?

2.50 Consider the forecast errors in Exercise 2.45. Construct a cumulative sum control chart for these forecast errors. Does the forecasting system exhibit stability over this time period?

2.51 Ten additional forecast errors for the forecasting model in Exercise 2.44 are as follows: 5.5358, –2.6183, 0.0130, 1.3543, 12.6980, 2.9007, 0.8985, 2.9240, 2.6663, and –1.6710. Plot these additional 10 forecast errors on the individuals and moving range control charts constructed in Exercise 2.47. Is the forecasting system still working satisfactorily?

2.52 Plot the additional 10 forecast errors from Exercise 2.51 on the cumulative sum control chart constructed in Exercise 2.38. Is the forecasting system still working satisfactorily?