CAN YOU PREDICT THE FUTURE BY LOOKING AT THE PAST?

Since accurate forecasting requires more than just inserting historical data into a model, *Forecasting: Methods and Applications, 3/e*, adopts a managerial, business orientation. Integrated throughout this text is the innovative idea that explaining the past is not adequate for predicting the future.

Inside, you will find the latest techniques used by managers in business today, discover the importance of forecasting and learn how it's accomplished. And you'll develop the necessary skills to meet the increased demand for thoughtful and realistic forecasts.

New features in the third edition include:

- An emphasis placed on the practical uses of forecasting.
- All data sets used in this book are available on the Internet.
- Comprehensive coverage provided on both quantitative and qualitative forecasting techniques.
- Includes many new developments in forecasting methodology and practice.

John Wiley & Sons, Inc.
New York • Chichester • Weinheim
Brisbane • Singapore • Toronto

http://www.wiley.com/college
http://www.maths.monash.edu.au/~hyndman/forecasting/

ISBN 0-471-53233-9
90000>
9 780471 532330

Makridakis
Wheelwright
Hyndman

FORECASTING
METHODS AND APPLICATIONS

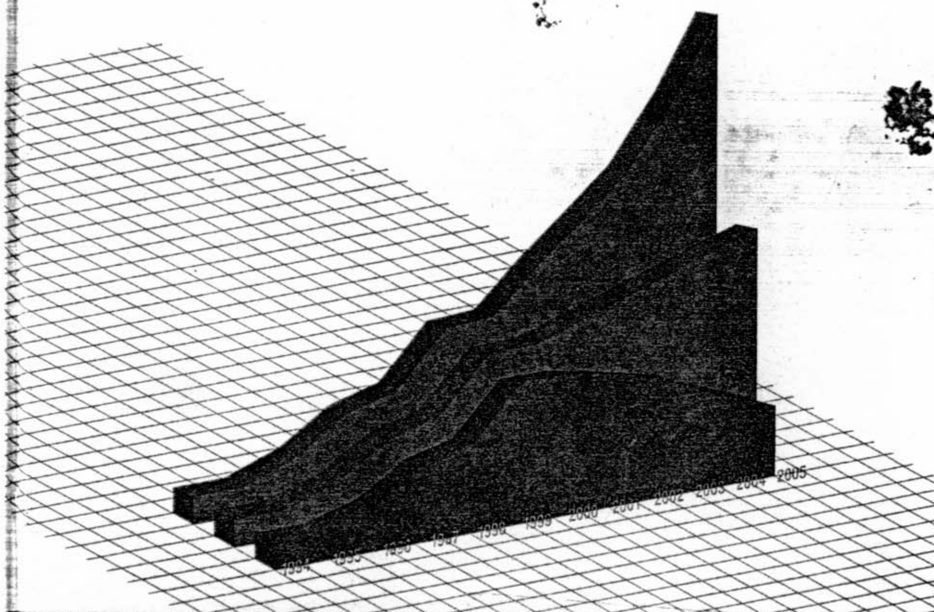
THIRD EDITION

WILEY

FORECASTING

METHODS AND APPLICATIONS

THIRD EDITION



Spyros Makridakis

Steven C. Wheelwright

Rob J. Hyndman

FORECASTING

Methods and Applications

Third Edition

Spyros Makridakis
European Institute of Business Administration (INSEAD)

Steven C. Wheelwright
Harvard University, Graduate School of Business Administration

Rob J. Hyndman
Monash University, Department of Mathematics and Statistics

John Wiley & Sons, Inc.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

Acquisitions Editor Beth L Golub
Marketing Manager Leslie Hines
Production Editor Kelly Tavares
Senior Designer Laura Boucher

This book was set in Computer Modern Roman by one of the authors and printed and bound by Malloy Lithographing. The cover was printed by Lehigh Press.

This book is printed on acid-free paper. ∞

This work is dedicated to Nikolas, Yiorgus, Petros and Ari
Margaret
Leanne, Naomi, Timothy, David and Abigail, whose patience
and support make such a project possible and worthwhile.

Makridakis, Spyros G.
Forecasting: methods and applications. —3rd ed. / Spyros Makridakis, Steven C. Wheelwright, Rob J. Hyndman.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-53233-9 (cloth: alk. paper)
1. Forecasting—Statistical methods. I. Wheelwright, Steven C., 1943- . II. Hyndman, Rob J. III. Title.
HD30.27.M34 1998
338.5'442—dc21 97-44416
CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2

PREFACE

The field of organizational forecasting, born in the 1950s, is reaching maturity. Significant theoretical developments in estimation and prediction, powerful and inexpensive computers coupled with appropriate software, several large scale empirical studies investigating the accuracy of the major forecasting methods, and, most importantly, the considerable experience gained through the actual application of such methods (in business and non-profit organizations) have contributed toward achieving this maturity. Today, the field of (organizational) forecasting rests on solid theoretical foundations while also having a realistic, practical base that increases its relevance and usefulness to practicing managers.

The preparation of this third edition, like the previous two, is based on the authors' view that the book should: (1) cover the full range of major forecasting methods, (2) provide a complete description of their essential characteristics, (3) present the steps needed for their practical application, (4) avoid getting bogged down in the theoretical details that are not essential to understanding how the various methods work, (5) provide systematic comparison of the advantages and drawbacks of various methods so that the most appropriate method can be selected for each forecasting situation, and (6) cover a comprehensive set of forecasting horizons (from the immediate to the long-term) and approaches (time series, explanatory, mixed) to forecasting.

New in this edition

While meeting the above criteria, this third edition includes major revisions of all the chapters and the addition of several completely new chapters. Our purpose has not been to merely revise the second edition, but rewrite it to include the contributions of the latest theoretical developments, and practical concerns, while presenting the most recent empirical findings and thinking. We have tried to make this edition both complete and fully updated, as well as theoretically correct and relevant, for those who want to apply forecasting in practice.

Some of the new material covered includes

- the X-12-ARIMA and the STL methods of time series decomposition
- local regression smoothing, best subsets regression and regression with time series errors.
- the use of Akaike's Information Criterion (AIC) for model selection
- neural networks and non-linear forecasting
- state space modeling and vector autoregression
- a modern approach to forecasting the long-term based on mega trends, analogies and scenarios
- new ideas for combining statistical and judgmental forecasts.
- experience gained from forecasting competitions including the latest M3-IJF Competition.
- recent research on forecast accuracy.
- the features of the major forecasting packages
- forecasting resources on the internet.

Unique features

This book is distinctive for its attention to practical forecasting issues, its comprehensive coverage of both statistical models and how to implement them in practice within a modern business environment, and the inclusion of many recent developments in forecasting research. In particular:

- There are dozens of real data examples and a number of examples from the authors' consulting experience. All data sets in the book are available on the internet (see below).
- We emphasise graphical methods and using graphs to help understand the analyses.

- Our perspective is that forecasting is much more than fitting models to historical data. While explaining the past is important, it is not adequate for accurately predicting the future.
- Much of the modern research on forecasting accuracy, based on surveys of forecast users, is summarized.
- Many recent developments in forecasting methodology and implementation are included.

Chapter outline

The book is divided into twelve chapters and three appendices.

Chapter 1: The forecasting perspective. This chapter provides a conceptual framework for understanding existing methodologies of forecasting and the tasks for which they can be used. It also provides an overview of the forecasting task and gives a useful structure for studying various forecasting methods.

Chapter 2: Basic forecasting tools. Chapter 2 outlines the basic quantitative foundations for the remainder of the book. It gives an overview of the notations and computations required to apply quantitative methods. Furthermore, it summarizes the various measures commonly used in evaluating and comparing the forecasts of different methods.

Chapter 3: Time series decomposition. Chapter 3 looks at the smoothing and decomposition of time series. These are not strictly forecasting methods, but they are useful tools for understanding time series better, and therefore they assist in forecasting.

Chapter 4: Exponential smoothing methods. Chapters 4 to 8 focus on quantitative (time series and explanatory) forecasting methods. In Chapter 4 exponential smoothing methods are examined, particularly single exponential smoothing, Holt's method and Holt-Winters' method.

Chapter 5: Simple regression. The explanatory regression methods (simple regression, multiple regression and econometric models) are discussed in Chapters 5 and 6. Chapter 5 deals with the case where there is only one explanatory variable.

Chapter 6: Multiple regression. Regression methods involving more than one explanatory variable are discussed in Chapter 6.

Chapter 7: The Box-Jenkins methodology for ARIMA models. In Chapter 7, the modeling approach of Box and Jenkins is described and extended using more recent methods such as the AIC.

Chapter 8: Advanced forecasting models. A collection of more advanced time series methods is discussed in Chapter 8 including regression with ARIMA errors, dynamic regression (or transfer function) models, intervention analysis, state space models and neural networks.

Chapter 9: Forecasting the long-term. This chapter discusses the problems and challenges when making long-term forecasts. It also provides three approaches (mega trends, analogies and scenarios) for arriving at such forecasts.

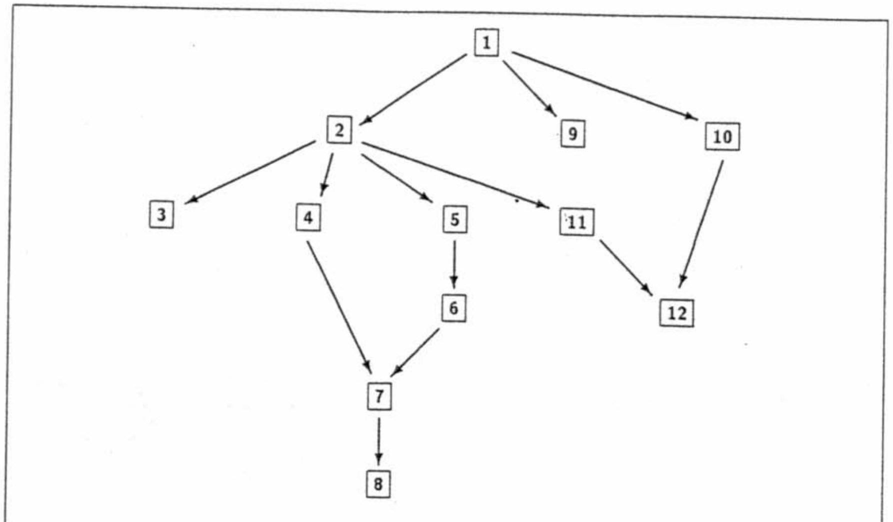
Chapter 10: Judgmental forecasting and adjustments. Chapter 10 describes the various biases and limitations that affect our judgment, as they relate to forecasting, and proposes ways to avoid or minimize them.

Chapter 11: The use of forecasting methods in practice. This chapter presents information about the usage of the various forecasting methods in business organizations as well as empirical findings about the accuracy of such methods. In addition it discusses the value of combining forecasts and the ability to improve the accuracy of the resulting forecasts through such combining.

Chapter 12: Implementing forecasting: its uses, advantages and limitations. The final chapter is concerned with practical implementation issues while also dealing with the usages, advantages and limitations of forecasting.

Appendix I: Forecasting resources. Several resources available to assist in the pursuit of forecast are discussed in Appendix I. These include software, journals, associations and useful Internet sites.

Appendix II: Glossary of forecasting terms. Appendix II is a glossary of forecasting terms covering the techniques, concepts,



Relationships between the chapters.

and tools that are the essential components of forecasting. This glossary can serve as a dictionary and reference for terms that may be new or not clearly understood by readers

Appendix III: Statistical tables. Appendix III contains statistical tables for the various tests of significance used in evaluating quantitative forecasts and methods.

These chapters need not be covered in the order given. The diagram above shows the relationships between the chapters and the necessary prerequisites for studying the material in each chapter.

Supplementary resources

There is a web page for the book at

www.maths.monash.edu.au/~hyndman/forecasting/

This contains a number of resources including all the data sets used throughout the book. Over 500 other data sets may also be obtained from this site as well as the 3003 series used in the latest M3-IJF Competition.

An instructor's manual is also available from the publisher. This includes course outlines, teaching suggestions, solutions to all exercises, and some suggestions for case studies and projects.

Acknowledgments

We would like to thank a number of reviewers for carefully reading our manuscript and providing many helpful suggestions. Leonard E. Ross (California State Polytechnic University) and Elaine L. Tatham (University of Kansas) reviewed the entire book, Gary Grunwald (University of Colorado) and Mindi Nath (Monash University) reviewed Chapters 1 to 8, and Brian Monsell (U.S. Bureau of the Census) reviewed Chapter 3, particularly the X-12-ARIMA methodology. Gary Grunwald also provided some ideas for exercises at the ends of the chapters. Victoria Briscoe, Betsy Brink and Linda Mayer provided excellent administrative assistance for which we are also grateful.

Spyros Makridakis
Fontainebleau, France

Steven Wheelwright
Boston, Massachusetts

Rob Hyndman
Melbourne, Australia

CONTENTS

1

THE FORECASTING PERSPECTIVE

1/1 Why forecast?

Frequently there is a time lag between awareness of an impending event or need and occurrence of that event. This lead time is the main reason for planning and forecasting. If the lead time is zero or very small, there is no need for planning. If the lead time is long, and the outcome of the final event is conditional on identifiable factors, planning can perform an important role. In such situations, forecasting is needed to determine when an event will occur or a need arise, so that appropriate actions can be taken.

In management and administrative situations the need for planning is great because the lead time for decision making ranges from several years (for the case of capital investments) to a few days or hours (for transportation or production schedules) to a few seconds (for telecommunication routing or electrical utility loading). Forecasting is an important aid in effective and efficient planning.

Opinions on forecasting are probably as diverse as views on any set of scientific methods used by decision makers. The layperson may question the validity and efficacy of a discipline aimed at predicting an uncertain future. However, it should be recognized that substantial progress has been made in forecasting over the past several centuries. There are a large number of phenomena whose outcomes can now be predicted easily. The sunrise can be predicted, as can the speed of a falling object, the trajectory of a satellite, rainy weather, and a myriad of other events. However, that was not always the case.

The evolution of science has increased the understanding of various aspects of the environment and consequently the predictability of many events. For example when the Ptolemaic system of astronomy was developed almost 1900 years ago, it could predict the movement of any star with an accuracy unheard of before that time. Even then, however, systematic errors were common. Then came the emergence of Copernican astronomy, which was much more accurate than its Ptolemaic predecessor and could predict the movement of the stars to within hundredths of a second. Today, modern astronomy is far more accurate than Copernican astronomy. The same increase in accuracy is shown in the theory of motion, which Aristotle, Galileo, Newton, and Einstein each improved.

The trend to be able to more accurately predict a wider variety

of events, particularly those in the economic/business environment, will continue to provide a better base from which to plan. Formal forecasting methods are the means by which this improvement is occurring.

Regardless of these improvements, two important comments must be kept in view. The first is that successful forecasting is not always directly useful to managers and others. More than 100 years ago, Jules Verne correctly predicted such developments as submarines, nuclear energy, and travel to the moon. Similarly, in the mid-1800s, Charles Babbage not only predicted the need for computers, but also proposed the design and did the actual construction for one. In spite of the accuracy of these forecasts, they were of little value in helping organizations to profit from such forecasts or achieve greater success.

A second important point is the distinction between uncontrollable external events (originating with the national economy, governments, customers, and competitors) and controllable internal events (such as marketing or manufacturing decisions within the firm). The success of a company depends on both types of events, but forecasting applies directly to the former, while decision making applies directly to the latter. Planning is the link that integrates both. decision making planning

For the important areas of sales forecasting, planning, and decision making, these relationships are shown in Figure 1-1. Recognizing the role of forecasting in its organizational and managerial context is usually as important as selecting the forecasting method itself, and thus it will be addressed throughout this book.

A wide variety of forecasting methods are available to management (see, for example, Makridakis and Wheelwright, 1989). These range from the most naïve methods, such as use of the most recent observation as a forecast, to highly complex approaches such as neural nets and econometric systems of simultaneous equations. In addition, the widespread introduction of computers has led to readily available software for applying forecasting techniques. Complementing such software and hardware has been the availability of data describing the state of economic events (GNP, consumption, etc.) and natural phenomena (temperature, rainfall, etc.). These data in conjunction with organizational statistics (sales, prices, advertising, etc.) and technological know-how provide the base of past information needed for the various forecasting methods.

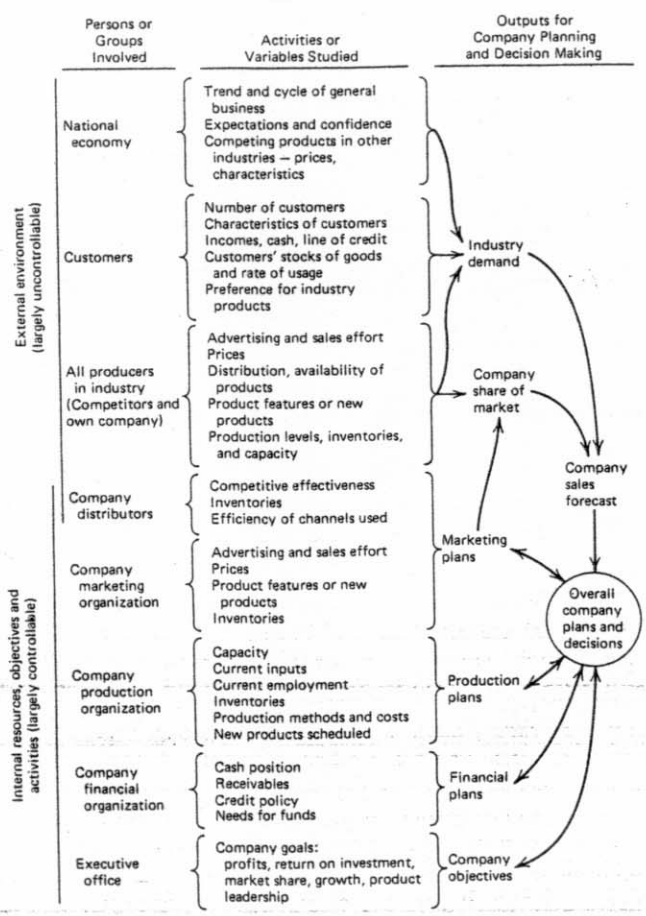


Figure 1-1: *Information flows in sales forecasting and business planning. (Adapted from Lippitt, 1969. Used by permission.)*

As suggested above, forecasting is an integral part of the decision making activities of management. An organization establishes goals and objectives, seeks to predict environmental factors, then selects actions that it hopes will result in attainment of these goals and objectives. The need for forecasting is increasing as management attempts to decrease its dependence on chance and becomes more scientific in dealing with its environment. Since each area of an organization is related to all others, a good or bad forecast can affect the entire organization. Some of the areas in which forecasting currently plays an important role are:

1. *Scheduling:* Efficient use of resources requires the scheduling of production, transportation, cash, personnel, and so on. Forecasts of the level of demand for product, material, labor, financing, or service are an essential input to such scheduling.

2. *Acquiring resources:* The lead time for acquiring raw materials, hiring personnel, or buying machinery and equipment can vary from a few days to several years. Forecasting is required to determine future resource requirements.

3. *Determining resource requirements:* all organizations must determine what resources they want to have in the long-term. Such decisions depend on market opportunities, environmental factors, and the internal development of financial, human, product, and technological resources. These determinations all require good forecasts and managers who can interpret the predictions and make appropriate decisions.

Although there are many different areas requiring forecasts, the preceding three categories are typical of the short-, medium-, and long-term forecasting requirements of today's organizations. This range of needs requires that a company develop multiple approaches to predicting uncertain events and build up a system for forecasting. This, in turn, requires that an organization possess knowledge and skills covering at least four areas: identification and definition of forecasting problems; application of a range of forecasting methods; procedures for selecting the appropriate methods for a specific situation; and organizational support for applying and using formalized forecasting methods.

A forecasting system must establish linkages among forecasts made by different management areas. There is a high degree of interdependence among the forecasts of various divisions or departments, which cannot be ignored if forecasting is to be successful. For example, errors in sales projections can trigger a series of reactions affecting budget forecasts, operating expenses, cash flows, inventory levels, pricing, and so on. Similarly, budgeting errors in projecting the amount of money available to each division will affect product development, modernization of equipment, hiring of personnel, and advertising expenditures. This, in turn, will influence, if not determine, the level of sales, operating costs, and cash flows. Clearly there is a strong interdependence among the different forecasting areas in an organization.

A major aim of this book is not only to examine the techniques available for meeting an organization's forecasting requirements, but also to consider the interdependence of needs in areas such as purchasing, production, marketing, finance, and general management.

1/2 An overview of forecasting techniques

Forecasting situations vary widely in their time horizons, factors determining actual outcomes, types of data patterns, and many other aspects. Figure 1-2 shows graphs of four variables for which forecasts might be required.

Figure 1-2a Monthly Australian electricity production from March 1956 to August 1995. (Source: Australian Bureau of Statistics.) Note the increasing trend, increasing variation each year, and the strong seasonal pattern that is slowly changing in shape. These strong historical patterns make this variable an easy one to forecast. Because of the changing seasonal patterns, some of the early data may not be useful in constructing a model. Forecasts are important for future planning of electricity production facilities and for ensuring existing facilities can meet peak demands.

Figure 1-2b U.S. Treasury Bill contracts on the Chicago market for 100 consecutive trading days in 1981. The downward trend is interesting, but it may only be a short downward movement

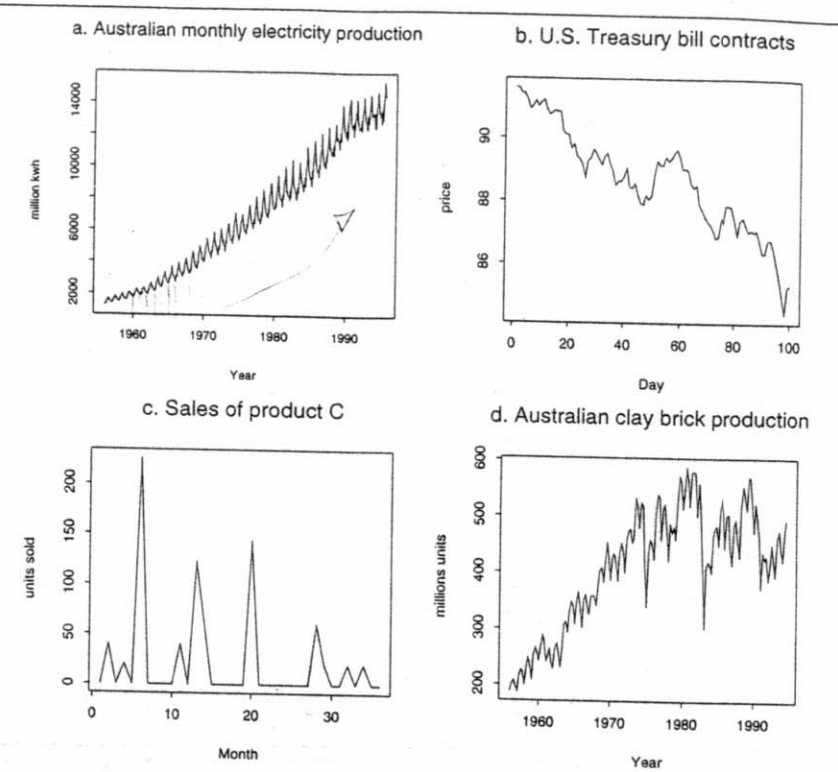


Figure 1-2: *Historical data on four variables for which forecasts might be required.*

in the middle of a highly variable series of observations. The critical question is whether this downward trend is likely to continue.

Figure 1-2c Sales of "product C" from a major oil company. This product was a lubricant sold only in units of large volume. To forecast a variable of this nature, it is necessary to investigate the nature of the product market, who is buying it, and what their future needs are likely to be.

Figure 1-2d Australian monthly clay brick production from March 1956 to September 1994. (Source: Australian Bureau of Statistics.) Clearly, the market is seasonal and quite volatile. Accurate forecasts are very difficult unless the cause of the fluctuations can be identified.

To deal with such diverse applications, several techniques have been developed. These fall into two major categories: quantitative and qualitative methods. Table 1-1 summarizes this categorization scheme and provides examples of situations that might be addressed by forecasting methods in these categories.

QUANTITATIVE: *Sufficient quantitative information is available.*

- **Time series:** Predicting the continuation of historical patterns such as the growth in sales or gross national product.
- **Explanatory:** Understanding how explanatory variables such as prices and advertising affect sales.

QUALITATIVE: *Little or no quantitative information is available, but sufficient qualitative knowledge exists.*

- Predicting the speed of telecommunications around the year 2020.
- Forecasting how a large increase in oil prices will affect the consumption of oil.

UNPREDICTABLE: *Little or no information is available.*

- Predicting the effects of interplanetary travel.
- Predicting the discovery of a new, very cheap form of energy that produces no pollution.

Table 1-1: *Categories of forecasting methods and examples of their application.*

Quantitative forecasting can be applied when three conditions exist: quantitative forecasting

1. Information about the past is available.
2. This information can be quantified in the form of numerical data.
3. It can be assumed that some aspects of the past pattern will continue into the future.

This last condition is known as the *assumption of continuity*; it is an underlying premise of all quantitative and many qualitative forecasting methods, no matter how sophisticated they may be.

Quantitative forecasting techniques vary considerably, having been developed by diverse disciplines for different purposes. Each has its own properties, accuracies, and costs that must be considered in choosing a specific method. Quantitative forecasting procedures fall on a continuum between two extremes: intuitive or ad hoc methods, and formal quantitative methods based on statistical principles. The first type is based on empirical experience that varies widely from business to business, product to product, and forecaster to forecaster. Intuitive methods are simple and easy to use but not always as accurate as formal quantitative methods. Also, they usually give little or no information about the accuracy of the forecast. Because of these limitations, their use has declined as formal methods have gained in popularity. Many businesses still use these methods, either because they do not know about simple formal methods or because they prefer a judgmental approach to forecasting instead of more objective approaches.

Formal statistical methods can also involve extrapolation, but it is done in a standard way using a systematic approach that attempts to minimize the forecasting errors. There are several formal methods, often requiring limited historical data, that are inexpensive and easy to use and that can be applied in a mechanical manner (see Chapter 4). These methods are useful when forecasts are needed for a large number of items and when forecasting errors on a single item will not be extremely costly.

Persons unfamiliar with quantitative forecasting methods often think that the past cannot describe the future accurately because everything is constantly changing. After some familiarity with data and forecasting techniques, however, it becomes clear that although nothing remains exactly the same, some aspects of history do repeat themselves in a sense. Application of the right method can often identify the relationship between the variable to be forecasted and time itself (or several other variables), making improved forecasting possible.

1/2/1 Explanatory versus time series forecasting

An additional dimension for classifying quantitative forecasting methods is to consider the underlying model involved. There are two major types of forecasting models: time series and explanatory models.

explanatory models

Explanatory models assume that the variable to be forecasted exhibits an explanatory relationship with one or more independent variables. For example,

$$\begin{aligned}\text{GNP} &= f(\text{monetary and fiscal policies, inflation,} \\ &\qquad \text{capital spending, imports, exports, error}).\end{aligned} \tag{1.1}$$

Notice that the relationship is not exact. There will always be changes in GNP that can not be accounted for by the variables in the model, and thus some part of GNP changes will remain unpredictable. Therefore, we include the "error" term on the right which represents random effects, beyond the variables in the model, that affect the GNP figures.

Explanatory models can be applied to many systems—a national economy, a company's market, or a household. The purpose of the explanatory model is to discover the form of the relationship and use it to forecast future values of the forecast variable. According to explanatory forecasting, any change in inputs will affect the output of the system in a predictable way, assuming the explanatory relationship will not change (assumption of continuity).

The procedure for selecting an appropriate functional form of equation (1.1) and estimating its parameters will be discussed in detail later on. At this point it should be emphasized that according to (1.1), GNP depends upon, or is explained by, the factors on the right-hand side of the equation. As these factors change, GNP will vary in the manner specified by (1.1).

time series models

Unlike explanatory forecasting, time series forecasting treats the

system as a black box and makes no attempt to discover the factors affecting its behavior. Therefore, prediction of the future is based on past values of a variable and/or past errors, but not on explanatory variables which may affect the system. The objective of such time series forecasting methods is to discover the pattern in the historical data series and extrapolate that pattern into the future.

There are two main reasons for wanting to treat a system as a black box. First, the system may not be understood, and even if it were understood it may be extremely difficult to measure the relationships assumed to govern its behavior. Second, the main concern may be only to predict what will happen and not to know why it happens. During the eighteenth, nineteenth, and first part of the twentieth centuries, for example, there were several people concerned with the magnitude of sunspots. There was little known at that time as to the reasons for the sunspots or the sources of energy of the sun. This lack of knowledge, however, did not hinder many investigators who collected and analyzed the frequency of sunspots. Schuster (1906) found that there was a regular pattern in the magnitude of sunspots, and he and several others were able to predict their continuation through time series analysis.

If the only purpose is to forecast future values of GNP without concern as to why a certain level of GNP will be realized, a time series approach would be appropriate. It is known that the magnitude of GNP does not change drastically from one month to another, or even from one year to another. Thus the GNP of next month will depend upon the GNP of the previous month and possibly that of the months before. Based on this observation, GNP might be expressed as follows:

$$\text{GNP}_{t+1} = f(\text{GNP}_t, \text{GNP}_{t-1}, \text{GNP}_{t-2}, \text{GNP}_{t-3}, \ldots, \text{error}), \tag{1.2}$$

where t is the present month, $t+1$ is the next month, $t-1$ is the last month, $t-2$ is two months ago, and so on.

Equation (1.2) is similar to (1.1) except that the factors on the right-hand side are previous values of the left-hand side. This makes the job of forecasting easier once (1.2) is known, since it requires no special input values as (1.1) does. However, a requirement with both equations (1.1) and (1.2) is that the relationship between the left- and right-hand sides of the equations must be discovered and measured.

Both time series and explanatory models have advantages in certain situations. Time series models can often be used more easily to forecast, whereas explanatory models can be used with greater success for policy and decision making. Whenever the necessary data are available, a forecasting relationship can be hypothesized either as a function of time or as a function of explanatory variables, and tested. As demonstrated by the GNP example, quite often it is possible to forecast by using either explanatory or time series approaches. It is also possible to combine the two approaches. Models which involve both time series and explanatory features are discussed in Chapter 8.

1/2/2 Qualitative forecasting

qualitative forecasting

Qualitative forecasting methods, on the other hand, do not require data in the same manner as quantitative forecasting methods. The inputs required depend on the specific method and are mainly the product of judgment and accumulated knowledge. (See Table 1-1.) Qualitative approaches often require inputs from a number of specially trained people.

As with their quantitative counterparts, qualitative techniques vary widely in cost, complexity, and value. They can be used separately but are more often used in combination with each other or in conjunction with quantitative methods.

It is more difficult to measure the usefulness of qualitative forecasts. They are used mainly to provide hints, to aid the planner, and to supplement quantitative forecasts, rather than to provide a specific numerical forecast. Because of their nature and cost, they are used almost exclusively for medium- and long-range situations such as formulating strategy, developing new products and technologies, and developing long-range plans. Although doubts are often expressed about the value of qualitative forecasting, it frequently provides useful information for managers. It is a premise of the authors that qualitative methods can be used successfully in conjunction with quantitative methods in such areas as product development, capital expenditures, goal and strategy formulation, and mergers, by even medium and small organizations. Whatever the shortcomings of qualitative methods, frequently the only alternative is no forecast at all.

The forecaster has a wide range of methods available that vary in accuracy, scope, time horizon, and cost. Key tasks are deciding which method to apply in each situation, how much reliance to place on the method itself, and how much modification is required to incorporate personal judgment before predictions are used as a basis for planning future actions. These issues will be addressed throughout this book.

1/3 The basic steps in a forecasting task

There are five basic steps in any forecasting task for which quantitative data are available.

Step 1: Problem definition

problem definition

The definition of the problem is sometimes the most difficult aspect of the forecaster's task. It involves developing a deep understanding of how the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organization. It is worth spending time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.

Consider the following statement by the manager of a paper products manufacturing company:

> We have a computerized inventory control system and we can get daily, weekly, and monthly reports at the drop of a hat. But our inventory situation is bad. We have far too much inventory at the factories, in the warehouses, and in the pipeline. Can we get better forecasts of future production and demand so we can reduce our inventory and save storage costs?

A forecaster has a great deal of work to do to properly define the forecasting problem, before any answers can be provided. For example, we need to know exactly what products are stored, who uses them, how long it takes to produce each item, what level of unsatisfied demand the company is prepared to bear, and so on.

Step 2: Gathering information

gathering information

There are always at least two kinds of information available: (a) statistical (usually numerical) data, and (b) the accumulated judgment and expertise of key personnel. Both kinds of information must be tapped.

It is necessary to collect historical data of the items of interest. We use the historical data to construct a model which can be used for forecasting. In the case of the paper products inventory, the data collected may consist of monthly demand and production for each item of interest over the previous three years. Other relevant data such as the timing and length of any significant production downtime due to equipment failure or industrial disputes may also need to be collected.

Step 3: Preliminary (exploratory) analysis

preliminary analysis

What do the data tell us? We start by graphing the data for visual inspection. Then we compute some simple descriptive statistics (e.g., mean, standard deviation, minimum, maximum, percentiles) associated with each set of data. Where more than one series of historical data is available and relevant, we can produce scatter plots of each pair of series and related descriptive statistics (e.g., correlations). These graphical and numerical summaries are discussed in Chapter 2. Another useful tool is decomposition analysis (Chapter 3) to check the relative strengths of trend, seasonality, cycles, and to identify unusual data points.

The purpose in all cases at this stage is to get a feel for the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers (extreme points) in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis?

Such preliminary analyses will help suggest a class of quantitative models that might be useful in the forecasting assignment.

Step 4: Choosing and fitting models

forecasting models

This step involves choosing and fitting several quantitative forecasting models. In this book we will be discussing many types of quantitative forecasting models and will explain the

technical details with completely worked-out examples. For now, we merely mention that the preliminary analysis (Step 3 above) serves to limit the search for an appropriate forecasting model and we would pursue one or two leading contenders for subsequent analysis.

Each model is itself an artificial construct. It is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be "fitted" using the known historical data. We will discuss exponential smoothing methods (Chapter 4), regression models (Chapters 5 and 6), Box-Jenkins ARIMA models (Chapter 7), and a variety of other topics including non-linear models, regression with ARIMA errors, intervention models, transfer function models, multivariate ARMA models, and state space models (Chapter 8).

When forecasting the long-term, a less formal approach is often better. This can involve identifying and extrapolating mega trends going back in time, using analogies, and constructing scenarios to consider future possibilities. These issues are discussed in Chapter 9.

Step 5: Using and evaluating a forecasting model

Once a model has been selected judiciously and its parameters estimated appropriately, the model is to be used to make forecasts, and the users of the forecasts will be evaluating the pros and cons of the model as time progresses. A forecasting assignment is not complete when the model has been fitted to the known data. The performance of the model can only be properly evaluated after the data for the forecast period have become available.

In this book, we have made a clear distinction between "fitting errors" and "forecasting errors." We will examine a variety of accuracy measures for both fitting and forecasting (in Chapter 2) and we will emphasize that, in practice, the model's forecasts are seldom used without modification. Expert judgment is invariably brought to bear on the use of the forecasts. The incorporation of expert judgment is addressed in Chapter 10.

errors

It is important to be aware of how each forecasting method has performed in practice in other forecasting contexts. There has now been quite a lot of research on this issue looking at users' preferences and experiences with a range of forecasting methods. This research is summarized in Chapter 11.

In addition, the accuracy of future forecasts is not the only criterion for assessing the success of a forecasting assignment. A successful forecasting assignment will usually also be a stimulus to action within the organization. If the forecasts suggest a gloomy picture ahead, then management will do its best to try to change the scenario so that the gloomy forecast will not come true. If the forecasts suggest a positive future, then the management will work hard to make that come true. In general, forecasts act as new information and management must incorporate such information into its basic objective to enhance the likelihood of a favorable outcome. Implementing forecasting is often at least as important as the forecasts themselves. Chapter 12 addresses this important subject.

References and selected bibliography

ARMSTRONG, J.S. (1978) *Long-range forecasting: from crystal ball to computer*, New York: John Wiley & Sons.

ASHER, W. (1978) *Forecasting: an appraisal for policy makers and planners*, Baltimore: Johns Hopkins University.

BAILS, D.G. and L.C. PEPPER (1993) *Business fluctuations: forecasting techniques and applications*, 2nd ed., Englewood Cliffs, N.J.: Prentice-Hall.

BRIGHT, J.R. (1978) *Practical technological forecasting*, Austin, Texas: Industrial Management Center.

CHAMBERS, J.C., S.K. MULLICK, and D.D. SMITH (1971) How to choose the right forecasting technique, *Harvard Business Review*, July–August, 45–57.

———— (1974) *An executive's guide to forecasting*, New York: John Wiley & Sons.

CLARK, A.C. (1971) *Profiles of the future*, New York: Bantam Books.

CLEARY, J.P. and H. LEVENBACH (1982) *The professional forecaster: the forecasting process through data analysis*, Belmont, Cal.: Lifetime Learning Publications.

CLIFTON, P., H. NGUYEN, and S. NUTT (1991) *Market research using forecasting in business*, Stoneham, Mass.: Butterworth and Heinemann.

KWONG, K.K., C. LI, V. SIMUNEK, and C.L. JAIN (1995) *Bibliography on forecasting and planning*, Flushing, N.Y.: Graceway Publishing Company.

LIPPITT, V.G. (1969) *Statistical sales forecasting*, New York: Financial Executives Institute.

MAKRIDAKIS, S. and S.C. WHEELWRIGHT (1979) "Forecasting," *TIMS studies in forecasting*, Vol. 12, Amsterdam: North-Holland.

———— (1981) "Forecasting an organization's futures," in *Handbook of organizational design*, Vol. 1, eds. P.C. Nystrom and W.H. Starbuck, Oxford: Oxford University Press, 122–138.

———— (1989) *Forecasting methods for management*, 5th ed., New York: John Wiley & Sons.

SCHUSTER, R. (1906) On the periodicity of sunspots, *Philosophical Transactions*, Series A, **206**, 69–100.

STEINER, G.A. (1979) *Strategic planning*, New York: The Free Press.

THOMOPOULOS, N.T. (1980) *Applied forecasting methods*, Englewood Cliffs, N.J.: Prentice-Hall.

Exercises

1.1 Several approaches have been suggested by those attempting to predict stock market movements. Three of them are described briefly below. How does each relate to the different approaches to forecasting described in this chapter?

(a) Dow Theory: There tend to be support levels (lower bounds) and resistance levels (upper bounds) for stock prices both for the overall market and for individual stocks. These levels can be found by plotting prices of the market or stock over time.

(b) Random Walk Theory: There is no way to predict future movements in the stock market or individual stocks, since all available information is quickly assimilated by the investors and moves market prices in the appropriate direction.

(c) The prices of individual stocks or of the market in general are largely determined by earnings.

1.2 You are asked to provide sales forecasts of several products for a large biscuit manufacturing company. Define the five steps of forecasting in the context of this project.

2

BASIC FORECASTING TOOLS

To develop an understanding of the field of quantitative forecasting requires some basic notation and terminology. This chapter presents such fundamentals. In Appendix 2-A the notation used throughout the book is presented, and in the body of this chapter the following topics are discussed: graphical methods for visualizing data to be used for forecasting (Section 2/2), the most important summary statistics (Section 2/3), and the various measures of forecasting accuracy that are used to help judge the appropriateness of a model (Section 2/4), calculation of prediction intervals (Section 2/5), the least squares procedure for estimating parameters of a model (Section 2/6), and the use of transformations to simplify data patterns (Section 2/7).

2/1 Time series and cross-sectional data

Throughout this chapter, we will use two data sets to illustrate ideas.

- Price (\$US), mileage (mpg), and country of origin for 45 automobiles from *Consumer Reports*, April 1990, pp. 235–255.
- Monthly Australian beer production (megaliters, Ml) from January 1991–August 1995.

These data are given in Tables 2-1 and 2-2.

Often our historical data will consist of a sequence of observations over time. We call such a sequence a *time series*. For example, monthly sales figures, daily stock prices, weekly interest rates, yearly profits, daily maximum temperatures, annual crop production, and electrocardiograph measurements are all time series.

time series

In forecasting, we are trying to estimate how the sequence of observations will continue into the future. To make things simple, we will assume that the times of observation are equally spaced. This is not a great restriction because most business series are measured daily, monthly, quarterly, or yearly and so will be equally spaced.

Of the two examples above, the beer data form a time series as they are monthly figures over a period of time. However, the automobile data do not form a time series. They are *cross-sectional* data; all observations are from the same time.

cross-sectional data

Make	Country	Mileage (mpg)	Price ($)
Chevrolet Caprice V8	USA	18	14525
Chevrolet Lumina APV V6	USA	18	13995
Dodge Grand Caravan V6	USA	18	15395
Ford Aerostar V6	USA	18	12267
Ford Mustang V8	USA	19	12164
Mazda MPV V6	Japan	19	14944
Nissan Van 4	Japan	19	14799
Chevrolet Camaro V8	USA	20	11545
Acura Legend V6	Japan	20	24760
Ford LTD Crown Victoria V8	USA	20	17257
Mitsubishi Wagon 4	Japan	20	14929
Nissan Axxess 4	Japan	20	13949
Mitsubishi Sigma V6	Japan	21	17879
Nissan Stanza 4	Japan	21	11650
Buick Century 4	USA	21	13150
Mazda 929 V6	Japan	21	23300
Oldsmobile Cutlass Ciera 4	USA	21	13150
Oldsmobile Cutlass Supreme V6	USA	21	14495
Chrysler Le Baron Coupe	USA	22	12495
Chrysler New Yorker V6	USA	22	16342
Eagle Premier V6	USA	22	15350
Ford Taurus V6	USA	22	13195
Nissan Maxima V6	Japan	22	17899
Buick Skylark 4	USA	23	10565
Oldsmobile Calais 4	USA	23	9995
Ford Thunderbird V6	USA	23	14980
Toyota Cressida 6	Japan	23	21498
Buick Le Sabre V6	USA	23	16145
Nissan 240SX 4	Japan	24	13249
Ford Tempo 4	USA	24	9483
Subaru Loyale 4	Japan	25	9599
Chrysler Le Baron V6	USA	25	10945
Mitsubishi Galant 4	Japan	25	10989
Plymouth Laser	USA	26	10855
Chevrolet Beretta 4	USA	26	10320
Dodge Daytona	USA	27	9745
Honda Prelude Si 4WS 4	Japan	27	13945
Subaru XT 4	Japan	28	13071
Ford Probe	USA	30	11470
Mazda Protege 4	Japan	32	6599
Eagle Summit 4	USA	33	8895
Ford Escort 4	USA	33	7402
Honda Civic CRX Si 4	Japan	33	9410
Subaru Justy 3	Japan	34	5866
Toyota Tercel 4	Japan	35	6488

Table 2-1: *Price, mileage, and country of origin for 45 automobiles from* Consumer Reports, *April 1990, pp. 235–255.*

Month	1991	1992	1993	1994	1995
January	164	147	139	151	138
February	148	133	143	134	136
March	152	163	150	164	152
April	144	150	154	126	127
May	155	129	137	131	151
June	125	131	129	125	130
July	153	145	128	127	119
August	146	137	140	143	153
September	138	138	143	143	
October	190	168	151	160	
November	192	176	177	190	
December	192	188	184	182	

Table 2-2: *Monthly Australian beer production: January 1991–August 1995.*

2/2 Graphical summaries

The single most important thing to do when first exploring the data is to visualize the data through graphs. The basic features of the data including patterns and unusual observations are most easily seen through graphs. Sometimes graphs also suggest possible explanations for some of the variation in the data.

data visualization

For example, industrial disputes will often affect time series of production; changes in government will affect economic time series; changes in definitions may result in identifiable changes in time series patterns. Graphs are the most effective way of identifying the effect of such events in the data. Where possible, these events should be adjusted for or included in the eventual model.

The type of data will determine which type of graph is most appropriate. Figures 2-1, 2-2, and 2-3 show three plots that provide useful information for forecasting. These graphical forms should be routinely used in forecasting projects and will be utilized throughout the rest of the book.

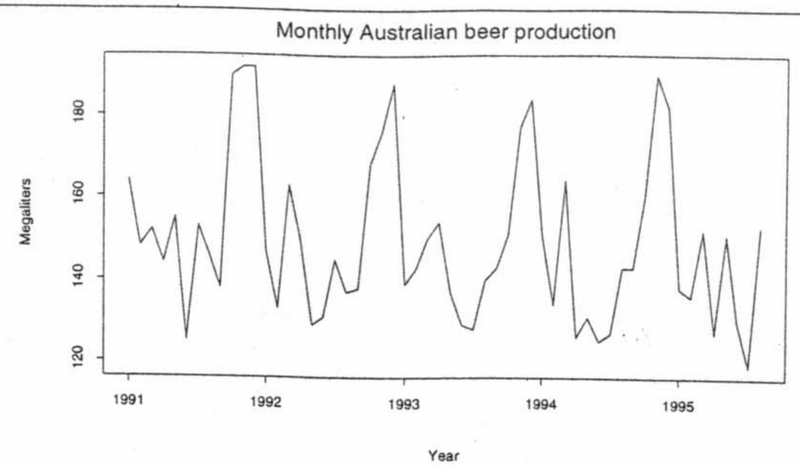


Figure 2-1: *Time plot of monthly Australian beer production (megaliters, Ml) from January 1991–August 1995.*

2/2/1 Time plots and time series patterns

time plot

For time series, the most obvious graphical form is a *time plot* in which the data are plotted over time. Figure 1-2 (p. 7) shows some examples. A time plot immediately reveals any trends over time, any regular seasonal behavior, and other systematic features of the data. These need to be identified so they can be incorporated into the statistical model.

Figure 2-1 shows a time plot of the beer data. This reveals the range of the data and the time at which peaks occur. It also shows the relative size of the peaks compared with the rest of the series and the randomness in the series since the data pattern is not perfect.

time series patterns

An important step in selecting an appropriate forecasting method is to consider the types of data patterns, so that the methods most appropriate to those patterns can be utilized. Four types of time series data patterns can be distinguished: horizontal, seasonal, cyclical, and trend.

horizontal

stationary

1. A *horizontal* (H) pattern exists when the data values fluctuate around a constant mean. (Such a series is called "stationary" in its mean.) A product whose sales do not increase or decrease over time would be of this type. Similarly, a quality control situation involving sampling from a continuous production process that theoretically does not change would also show a horizontal pattern.

seasonal

periodic

2. A *seasonal* (S) pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Sales of products such as soft drinks, ice creams, and household electricity consumption all exhibit this type of pattern. The beer data show seasonality with a peak in production in November and December (in preparation for Christmas) each year. Seasonal series are sometimes also called "periodic" although they do not exactly repeat themselves over each period.

cyclical

3. A *cyclical* (C) pattern exists when the data exhibit rises and falls that are *not of a fixed period*. For economic series, these are usually due to economic fluctuations such as those associated with the business cycle. The sales of products such as automobiles, steel, and major appliances exhibit this type of pattern. The clay brick production shown in Figure 1-2d (p. 7) shows cycles of several years in addition to the quarterly seasonal pattern. The major distinction between a seasonal and a cyclical pattern is that the former is of a constant length and recurs on a regular periodic basis, while the latter varies in length. Moreover, the average length of a cycle is usually longer than that of seasonality and the magnitude of a cycle is usually more variable than that of seasonality.

trend

4. A *trend* (T) pattern exists when there is a long-term increase or decrease in the data. The sales of many companies, the gross national product (GNP), and many other business or economic indicators follow a trend pattern in their movement over time. The electricity production data shown in Figure 1-2a (p. 7) exhibit a strong trend in addition to the monthly seasonality. The beer data in Figure 2-1 show no trend.

Many data series include combinations of the preceding patterns. For example, Figure 1-2d (p. 7) shows trend, seasonality, and cyclical

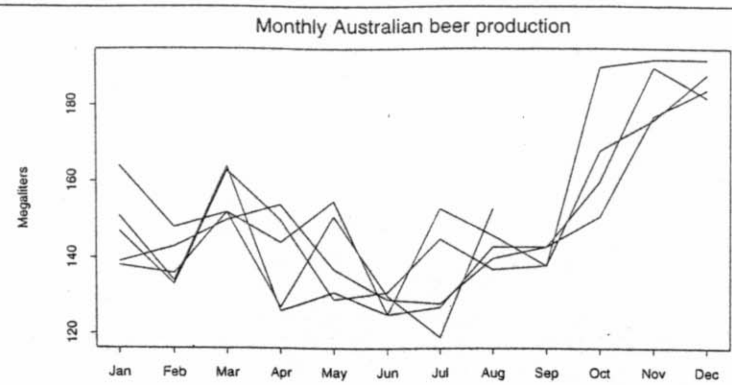


Figure 2-2: *A seasonal plot of the Australian beer production data. Note that production peaks in November and December in preparation for the southern hemisphere summer and is least in winter.*

behavior. One of the things that makes forecasting interesting and challenging is the huge variety of patterns that commonly occur in real time series data. Forecasting methods that are capable of distinguishing each of the patterns must be employed if a separation of the component patterns is needed. Similarly, alternative methods of forecasting can be used to identify the pattern and to best fit the data so that future values can be forecasted.

2/2/2 Seasonal plots

seasonal plot

For time series data that are seasonal, it is often useful to also produce a seasonal plot. Figure 2-2 shows a seasonal plot of the beer data. This graph consists of the data plotted against the individual "seasons" in which the data were observed. (In this case a "season" is a month.) This is something like a time plot except that the data from each season are overlapped. A seasonal plot enables the underlying seasonal pattern to be seen more clearly, and also allows any substantial departures from the seasonal pattern to be easily

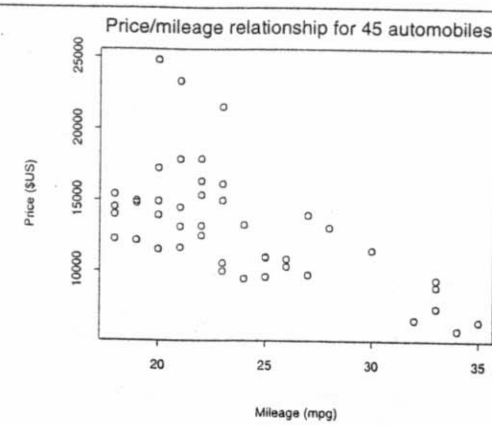


Figure 2-3: *A scatterplot of price versus mileage for the automobile data.*

identified. For example, there is one year (1991) in which the October beer production was higher than the pattern evident in the other years.

2/2/3 Scatterplots

The automobile data of Table 2-1 are not a time series making time or seasonal plots inappropriate for these data. However, these data are well suited to a scatterplot (see Figure 2-3) such as that of price against mileage. In Figure 2-3 we have plotted the variable we wish to forecast (price) against one of the explanatory variables (mileage). Each point on the graph represents one type of vehicle. The plot shows the relationship between price and mileage: vehicles with high mileage per gallon are generally cheaper than less fuel-efficient vehicles. (Both price and fuel-efficiency are related to the vehicle and engine size.) Vehicles with low mileage per gallon are generally priced over a range from around \$12,000 to \$18,000, with three vehicles much more expensive than other vehicles of comparable efficiency. The scatterplot helps us visualize the relationship and suggests that a forecasting model must include mileage as an explanatory variable.

scatterplot

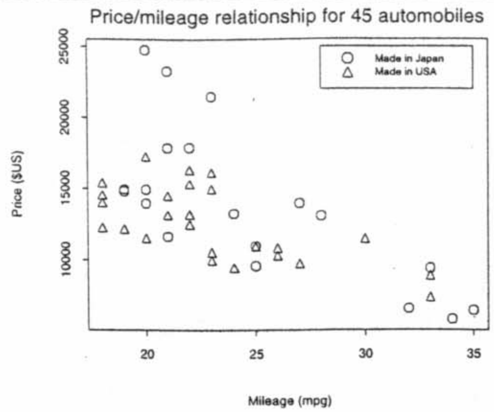


Figure 2-4: *A scatterplot showing price, mileage, and the country of origin for the automobile data.*

With the automobile data there is an additional explanatory variable, country of origin, which is a categorical (qualitative) variable rather than a numerical variable (its values are categories rather than numbers). Thus we cannot plot price against country in the same way. However, we can augment our scatterplot of price against mileage to also show the country of origin information. This is displayed in Figure 2-4. Here we have a scatterplot showing three variables. It shows that the two most efficient automobiles and the three expensive automobiles are all Japanese. It also shows that overall U.S.A. automobiles may be more efficient than the Japanese ones. A forecasting model might also include the country of origin as an explanatory variable, but that it is probably going to be less effective than mileage in giving accurate predictions.

2/3 Numerical summaries

statistic

In addition to graphics, it is also helpful to provide numerical summaries. A summary number for a data set is called a *statistic*.

univariate data

For a single data set (univariate data) or a single time series, the most common descriptive statistics are the mean, the standard deviation, and the variance. Section 2/3/1 deals with these univariate statistics.

bivariate data

For a pair of random variables (bivariate data) it is of interest to describe how the two data sets relate to each other. The most widely used summary numbers (statistics) for this purpose are the covariance and the correlation, and these will be defined in Section 2/3/2.

Then for a single time series, it is very useful to compare the observation at one time period with the observation at another time period. The two most common statistics here are the autocovariance and the autocorrelation, which are defined in Section 2/3/3. These measures will be used extensively in Chapter 7.

2/3/1 Univariate statistics

Consider the mileage of the 19 Japanese automobiles given in Table 2-1, reproduced in Table 2-3. The vehicles have been numbered from 1 to 19 for easy reference.

mean

Using the letter M to denote mileage and a subscript i ($i = 1, 2, 3, \ldots, 19$) to denote the ith vehicle, the mean mileage can be written[1]

$$\begin{aligned} \bar{M} &= (M_1 + M_2 + M_3 + \cdots + M_{19})/19 \\ &= \frac{1}{19}\sum_{i=1}^{19} M_i \\ &= 469/19 = 24.68 \text{ mpg.} \end{aligned}$$

median

The mean should not be confused with the *median*, which is the middle observation. So for the preceding 19 vehicles, the median is the mileage of the tenth vehicle when they are arranged in increasing order as shown in Table 2-3. That is, the median of the mileage data is 23. Both the mean and the median are designed to provide a numerical measure of the center of the data set.

[1]The summation notation, Σ, used here is explained in Appendix 2-A.

Make	Vehicle	Mileage (mpg)	
Mazda MPV V6	1	19	
Nissan Van 4	2	19	
Acura Legend V6	3	20	
Mitsubishi Wagon 4	4	20	
Nissan Axxess 4	5	20	
Mitsubishi Sigma V6	6	21	
Nissan Stanza 4	7	21	
Mazda 929 V6	8	21	
Nissan Maxima V6	9	22	
Toyota Cressida 6	10	23	← median
Nissan 240SX 4	11	24	
Subaru Loyale 4	12	25	
Mitsubishi Galant 4	13	25	
Honda Prelude Si 4WS 4	14	27	
Subaru XT 4	15	28	
Mazda Protege 4	16	32	
Honda Civic CRX Si 4	17	33	
Subaru Justy 3	18	34	
Toyota Tercel 4	19	35	

Table 2-3: *Mileage of Japanese automobiles listed in Table 2-1.*

As well as measuring the center of a data set, it is also valuable to measure the spread of the data. That is, we want a numerical measure indicating if the data are tightly bunched together or spread across a wide range.

To develop a measure of spread, we first need to calculate for each vehicle how far its mileage is from the mean mileage. The mean $\bar{M}$ is subtracted from each M_i to give the ith deviation from the mean, $(M_i - \bar{M})$.

deviation from mean

The sum of the deviations will always equal zero (as shown under column 3 of Table 2-4). Therefore, to develop a useful descriptive statistic from these deviations, they are either squared (as in column 5 of Table 2-4), or, occasionally, the absolute value is taken (as in column 4). The mean of the absolute deviations is denoted MAD,

mean absolute deviation

and for the mileage data

$$\text{MAD} = \frac{1}{19}\sum_{i=1}^{19}|M_i - \bar{M}| = 83.1/19 = 4.37 \text{ mpg}.$$

The mean of the squared deviations is designated MSD:

mean squared deviation

$$\text{MSD} = \frac{1}{19}\sum_{i=1}^{19}(M_i - \bar{M})^2 = 514.11/19 = 27.1.$$

Closely related to the mean squared deviation (MSD) is the *variance*, which is defined as the sum of squared deviations divided by one less than the total number of observations.[2] For the mileage data the variance of age is

variance

$$S^2 = \frac{1}{18}\sum_{i=1}^{19}(M_i - \bar{M})^2 = 514.11/18 = 28.6.$$

Note that since the variance formula uses 18 in the denominator and MSD formula uses 19, S^2 is larger than MSD. The variance S^2 is less intuitive than MSD but it has some desirable mathematical properties.[3]

The deviations $(M_i - \bar{M})$ are defined in units of miles per gallon. Therefore, the squared deviations are in units of squared mpg and so the mean squared deviation, MSD, and the variance, S^2, are also defined in units of squared mpg. By taking the square root of these two summary numbers, we get summary statistics in the same units as the data. In particular, we will define the *standard deviation* as

standard deviation

$$S = \sqrt{S^2} = \sqrt{28.6} = 5.34.$$

Both the MAD and the standard deviation S provide measures of spread. They (roughly) measure the average deviation of the

measures of spread

[2]The sum of squared deviations is divided by the "degrees of freedom" which can be defined as the number of data points minus the number of parameters estimated. When calculating a variance, we have to estimate the mean using the data, so the degrees of freedom is one less than the total number of observations.

[3]In statistics, a distinction is made between a biased estimator and an unbiased estimator. For sample data, the MSD is a biased estimator of population variance and the variance S^2 is an unbiased estimator of the population variance. See Rice (1995), p. 192, for a definition of unbiasedness.

observations from their mean. If the observations are spread out, they will tend to be far from the mean, both above and below. Some deviations will be large positive numbers, and some will be large negative numbers. But the squared deviations (or the absolute deviations) will be all positive. So both MAD and S will be large when the data are spread out, and small when the data are close together. Both MAD and S have the same units as the observations.

For many data sets the following useful rules of thumb hold:

- approximately two-thirds of the observations lie within 1 standard deviation of the mean; and
- approximately 95% of the observations lie within 2 standard deviations of the mean.

To summarize, suppose there are n observations and the individual observations are denoted by Y_i for $i = 1, \ldots, n$. Then the univariate statistics (summary numbers) that will be used in this text are defined (generally) as follows:

mean

$$\bar{Y} = \frac{1}{n}\sum Y_i \tag{2.1}$$

median

$$\text{Median} = \begin{cases} \text{middle observation if } n \text{ odd;} \\ \text{average of middle two} \\ \quad \text{observations if } n \text{ even.} \end{cases} \tag{2.2}$$

mean absolute deviation

$$\text{MAD} = \frac{1}{n}\sum |Y_i - \bar{Y}| \tag{2.3}$$

mean squared deviation

$$\text{MSD} = \frac{1}{n}\sum (Y_i - \bar{Y})^2 \tag{2.4}$$

variance

$$S^2 = \frac{1}{n-1}\sum (Y_i - \bar{Y})^2 \tag{2.5}$$

standard deviation

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum (Y_i - \bar{Y})^2}. \tag{2.6}$$

All summations are over the index i from 1 through n.

Of course, with much statistical software readily accessible, there is usually no need to compute these statistics by hand. However, it is important to understand the formulae behind them in order to understand what they mean.

(1) Vehicle i	(2) Mileage M_i	(3) Deviation $(M_i - \bar{M})$	(4) $\lvert M_i - \bar{M}\rvert$	(5) $(M_i - \bar{M})^2$
1	19	−5.7	5.7	32.31
2	19	−5.7	5.7	32.31
3	20	−4.7	4.7	21.94
4	20	−4.7	4.7	21.94
5	20	−4.7	4.7	21.94
6	21	−3.7	3.7	13.57
7	21	−3.7	3.7	13.57
8	21	−3.7	3.7	13.57
9	22	−2.7	2.7	7.20
10	23	−1.7	1.7	2.84
11	24	−0.7	0.7	0.47
12	25	0.3	0.3	0.10
13	25	0.3	0.3	0.10
14	27	2.3	2.3	5.36
15	28	3.3	3.3	10.99
16	32	7.3	7.3	53.52
17	33	8.3	8.3	69.15
18	34	9.3	9.3	86.78
19	35	10.3	10.3	106.42
Sums	469	0.0	83.1	514.11

Mean	$\bar{M}$ = (col 2 sum)/19 = 24.68	using (2.1)
Median	Median = (col 2 middle observation) = 23	using (2.2)
Mean Absolute Deviation	MAD = (col 4 sum)/19 = 4.37	using (2.3)
Mean Squared Deviation	MSD = (col 5 sum)/19 = 27.1	using (2.4)
Variance	S^2 = (col 5 sum)/18 = 28.6	using (2.5)
Standard Deviation	$S = \sqrt{S^2} = 5.34$	using (2.6)

Table 2-4: *Computation of the univariate statistics for the mileage of Japanese automobiles.*

Make	Vehicle	Mileage (mpg)	Price ($'000)
Mazda MPV V6	1	19	14.944
Nissan Van 4	2	19	14.799
Acura Legend V6	3	20	24.760
Mitsubishi Wagon 4	4	20	14.929
Nissan Axxess 4	5	20	13.949
Mitsubishi Sigma V6	6	21	17.879
Nissan Stanza 4	7	21	11.650
Mazda 929 V6	8	21	23.300
Nissan Maxima V6	9	22	17.899
Toyota Cressida 6	10	23	21.498
Nissan 240SX 4	11	24	13.249
Subaru Loyale 4	12	25	9.599
Mitsubishi Galant 4	13	25	10.989
Honda Prelude Si 4WS 4	14	27	13.945
Subaru XT 4	15	28	13.071
Mazda Protege 4	16	32	6.599
Honda Civic CRX Si 4	17	33	9.410
Subaru Justy 3	18	34	5.866
Toyota Tercel 4	19	35	6.488

Table 2-5: *Price and mileage for the Japanese automobiles listed in Table 2-1.*

2/3/2 Bivariate statistics

Table 2-5 shows the price and mileage for the Japanese automobiles given in Table 2-1. To prevent the computations from becoming cumbersome, we will deal with the price variable in units of thousands of dollars.

negative/positive relationships

When these data are plotted, as in Figure 2-5, it can be seen that a negative relationship exists between these two variables. By negative relationship we mean that as mileage increases, price tends to decrease. (A positive relationship would be similar to the height versus weight relationship—as height increases weight increases too.) Whenever we are dealing with two paired observations (e.g., price and mileage, height and weight, price and demand), it is of interest to examine and measure the extent of the relationship between the two variables.

Suppose we denote the two variables by X and Y. A statistic

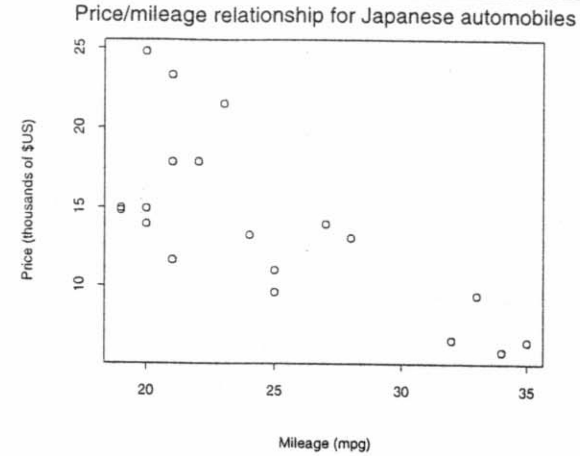


Figure 2-5: *A scatterplot of price versus mileage for the Japanese automobile data in Table 2-5.*

which indicates how two variables "co-vary" is called the *covariance* and is defined as follows:

covariance

$$\text{Cov}_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{2.7}$$

where $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively, and n is the number of observations on each variable.

For the price and mileage data in Table 2-5, the computations necessary for determining the covariance (Cov_{PM}) between price (P) and mileage (M) are shown in Table 2-6. First, the mean price ($\bar{P}$) and the mean mileage ($\bar{M}$) are computed using columns 2 and 3, respectively. Then deviations from the mean are calculated in columns 4 and 5, and column 8 gives the product of these two deviations. Summing the deviation products (column 8) and dividing by the degrees of freedom, $n - 1 = 18$, yields the desired covariance,

$$\text{Cov}_{PM} = -378.03/18 = -21.00.$$

(1) i	(2) M_i	(3) P_i	(4) $M_i - \bar{M}$	(5) $P_i - \bar{P}$	(6) $(M_i - \bar{M})^2$	(7) $(P_i - \bar{P})^2$	(8) $(M_i - \bar{M})(P_i - \bar{P})$
1	19	14.944	−5.68	1.01	32.31	1.01	−5.72
2	19	14.799	−5.68	0.86	32.31	0.74	−4.89
3	20	24.760	−4.68	10.82	21.94	117.12	−50.69
4	20	14.929	−4.68	0.99	21.94	0.98	−4.64
5	20	13.949	−4.68	0.01	21.94	0.00	−0.05
6	21	17.879	−3.68	3.94	13.57	15.53	−14.52
7	21	11.650	−3.68	−2.29	13.57	5.24	8.43
8	21	23.300	−3.68	9.36	13.57	87.65	−34.49
9	22	17.899	−2.68	3.96	7.21	15.69	−10.63
10	23	21.498	−1.68	7.56	2.84	57.15	−12.73
11	24	13.249	−0.68	−0.67	0.47	0.48	0.47
12	25	9.599	0.32	−4.34	0.10	18.83	−1.37
13	25	10.989	0.32	−2.95	0.10	8.70	−0.93
14	27	13.945	2.32	0.01	5.36	0.00	0.02
15	28	13.071	3.32	−0.87	10.99	0.75	−2.88
16	32	6.599	7.32	−7.34	53.52	53.86	−53.69
17	33	9.410	8.32	−4.53	69.15	20.50	−37.65
18	34	5.866	9.32	−8.07	86.78	65.16	−75.20
19	35	6.488	10.32	−7.45	106.42	55.50	−76.85
Sums	469	264.823	0.00	0.00	514.11	524.88	−378.03

Mean mileage $\bar{M} = 469/19 = 24.68$ mpg

Mean price $\bar{P} = 264.823/19 = 13.938$ thousands of dollars

Standard deviation of M $S_M = \sqrt{514.11/18} = 5.34$

Standard deviation of P $S_P = \sqrt{524.88/18} = 5.40$

Covariance between P and M $\text{Cov}_{PM} = -378.03/18 = -21.00$

Correlation between P and M $r_{PM} = \dfrac{\text{Cov}_{PM}}{S_P S_M} = \dfrac{-21.00}{(5.40)(5.34)} = -0.73$

Table 2-6: *Computations for determining the covariance and the correlation of the price and mileage data of Table 2-5.*

Note that the units of covariance are problematical. It is difficult to interpret thousands of dollar-miles per gallon. Hence the value of computing the correlation coefficient, described below. Note that the covariance between price and mileage is negative, but the magnitude of Cov_{PM} clearly depends on the units involved. If the mileage figures were converted to km per liter and the prices to dollars, the plot (Figure 2-5) would look the same but the covariance would be quite different.

correlation

The *correlation coefficient*, designated r, is a special covariance measure that takes care of the scale problem just mentioned. If the covariance (Cov_{XY}) is divided by the two standard deviations (S_X and S_Y), then the units in the numerator and the denominator cancel out, leaving a dimensionless number, which is the correlation coefficient between X and Y. This is written as follows:

$$r_{XY} = \frac{\text{Cov}_{XY}}{S_X S_Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}} \quad (2.8)$$

The effect of this scaling (dividing Cov_{XY} by S_X and S_Y) is to restrict the range of r_{XY} to the interval −1 to +1. No matter what the units of measurement for X and Y the correlation coefficient, r_{XY}, is always restricted to lie within that interval. More information about correlation is given in Sections 5/2/2 and 5/2/3.

For the data in Table 2-5 the computations involved in getting to the correlation coefficient are included in Table 2-6. Columns 6 and 7 are the squared deviations for height and weight, respectively, and can be used to determine the standard deviations S_P and S_M, using equation (2.6). Then the covariance between P and M can be divided by S_P and S_M to yield the correlation between price and mileage,

$$r_{PM} = \frac{-21.00}{(5.34)(5.40)} = -0.73. \quad (2.9)$$

This summary number is readily interpretable. There is a correlation of −0.73 between price and mileage, which is negative and substantial. There is a strong negative association between price and mileage.

Covariance and especially correlation are the basic statistics for bivariate data sets, and for more extensive multivariate data sets. Care should be taken, however, to remember that these are measures of linear association between two variables, so that it is not appropriate (meaningful) to apply the correlation measure when there is a pronounced curvilinear relationship between the two variables. This point is amplified in Chapter 5.

In summary, the two "vital" statistics for bivariate data sets are

covariance

correlation

$$\text{Cov}_{XY} = \frac{1}{n-1}\sum(X_i - \bar{X})(Y_i - \bar{Y})$$

$$r_{XY} = \frac{\text{Cov}_{XY}}{S_X S_Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

In practice these calculations are done by computer.

2/3/3 Autocorrelation

The covariance and correlation coefficient are statistics (summary measures) that measure the extent of the linear relationship between two variables. As such, they can be used to identify *explanatory relationships*. Autocovariance and autocorrelation are comparable measures that serve the same purpose for a single time series.

autocovariance
autocorrelation

For example, if we compare Y_t (the observation at time t) with Y_{t-1} (the observation at time $t-1$), then we see how consecutive observations are related. The observation Y_{t-1} is described as "lagged" by one period. Similarly, it is possible to compare observations lagged by two periods, three periods, and so on.

lagged variable

Table 2-7 shows the beer series, which is a single time series over 56 months from January 1991 to August 1995. The observations Y_1, Y_2, ..., Y_{56} are observed at time periods 1, 2, ..., 56, respectively. If we lag the series by one period, as shown in column 3, then there will be 55 pairs of observations to compare. For these 55 overlapping observations we can compute the covariance and the correlation as if they were two separate series. However, since they are one and the same series (with a lag of one period) the summary measures are called autocovariance and autocorrelation.

Because the two series are almost the same, rather than use

the equations (2.7) and (2.8) to compute the autocovariance and autocorrelation, we normally use simpler formulas which give almost the same answers. We denote the autocovariance at lag k by c_k and the autocorrelation at lag k by r_k. Then define

$$c_k = \frac{1}{n}\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y}) \qquad (2.10)$$

autocovariance

and

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \qquad (2.11)$$

autocorrelation

By way of illustration, consider the beer data and the calculations of autocovariance and autocorrelation in Table 2-7. The mean of *all* the data points in column 2 is $\bar{Y} = 149.30$, and the deviations in columns 4 and 5 are deviations from $\bar{Y}$. Column 6 is the squared deviations from column 4 and the sum of these squares is the denominator in equation (2.11). Column 7 is the column of deviation products (column 4 times column 5). The calculations, using (2.10) and (2.11), are given at the bottom of Table 2-7.

Using exactly similar procedures, the autocorrelations for lags two, three, and beyond can be obtained. The results for the beer data are as follows:

$r_1 = 0.421$	$r_8 = -0.156$
$r_2 = 0.057$	$r_9 = -0.008$
$r_3 = -0.059$	$r_{10} = 0.051$
$r_4 = -0.188$	$r_{11} = 0.374$
$r_5 = -0.287$	$r_{12} = 0.596$
$r_6 = -0.424$	$r_{13} = 0.303$
$r_7 = -0.343$	$r_{14} = 0.082$

Notice that the autocorrelation at lag 12 is higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be 12 months apart and the troughs tend to be 12 months apart. Similarly, the autocorrelation at lag 6 is more negative than for the other lags because troughs tend to be 6 months behind peaks.

(1) t	(2) Y_t	(3) Y_{t-1}	(4) $(Y_t - \bar{Y})$	(5) $(Y_{t-1} - \bar{Y})$	(6) $(Y_t - \bar{Y})^2$	(7) $(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$
1	164	—	14.70	—	215.99	—
2	148	164	−1.30	14.70	1.70	−19.16
3	152	148	2.70	−1.30	7.27	−3.51
4	144	152	−5.30	2.70	28.13	−14.30
5	155	144	5.70	−5.30	32.45	−30.21
6	125	155	−24.30	5.70	590.66	−138.44
⋮	⋮	⋮	⋮	⋮	⋮	⋮
53	151	127	1.70	−22.30	2.88	−37.84
54	130	151	−19.30	1.70	372.63	−32.75
55	119	130	−30.30	−19.30	918.31	584.97
56	153	119	3.70	−30.30	13.66	−112.01
Sums	8361				21135.84	8893.51

$$\text{Mean} \quad \bar{Y} = \frac{8361}{56} = 149.30$$

$$\text{Autocovariance lag 1} \quad c_1 = \frac{8893.51}{56} = 158.8$$

$$\text{Autocorrelation lag 1} \quad r_1 = \frac{8893.51}{21135.84} = 0.421$$

Table 2-7: *Computing the autocovariance and the autocorrelation using equations (2.10) and (2.11), and a lag of one period.*

autocorrelation function

correlogram

Together, the autocorrelations at lags 1, 2, ..., make up the *autocorrelation function* or ACF. Rather than scanning a list of numbers, it is much easier to plot the autocorrelations against the lag. Such a plot is known as a *correlogram* and helps us visualize the ACF quickly and easily. Figure 2-6 shows the ACF for the beer data. Here the seasonal pattern is seen very clearly.

A plot of the ACF is a standard tool in exploring a time series before forecasting. It provides a useful check for seasonality, cycles, and other time series patterns. In the exercises at the end of this chapter are some data sets that display various kinds of pattern (trend, seasonality, and cycles) and the autocorrelations for these

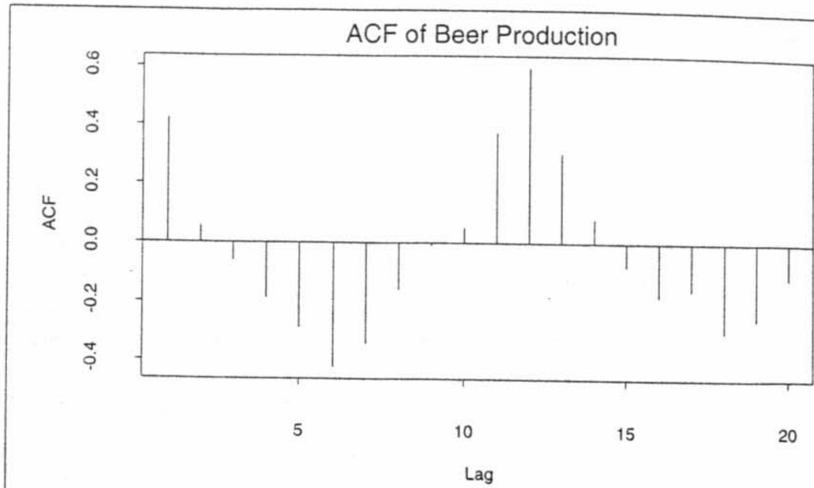


Figure 2-6: *The correlogram (or ACF plot) for the beer production data.*

series will be very helpful in verifying the pattern.

The ACF also helps us identify if previous values of the series contain much information about the next value, or whether there is little relationship between one observation and the next.

To sum up, much is to be learned about a single time series by examining the autocorrelations of the series with itself, lagged one period, two periods, and so on. The ACF plays a very important role in time series forecasting.

2/4 Measuring forecast accuracy

We now turn to another fundamental concern—how to measure the suitability of a particular forecasting method for a given data set. In most forecasting situations, accuracy is treated as the overriding criterion for selecting a forecasting method. In many instances, the word "accuracy" refers to "goodness of fit," which in turn refers to how well the forecasting model is able to reproduce the data that are

goodness of fit

already known. To the consumer of forecasts, it is the accuracy of the *future* forecast that is most important.

In this section, a variety of measures of forecasting (or modeling) accuracy will be defined and in subsequent chapters these measures will be used in the context of worked examples.

To illustrate the computations involved, we will refer to the Australian monthly beer production. Table 2-8 shows the last 8 months of observations (January–August 1995). The second column shows forecasts for these values, obtained using a very simple method, by taking the average of each month over the past four years. So, for example, the forecast for January 1995 is the average production for January 1991, January 1992, January 1993, and January 1994.

Period t	Observation Y_t	Forecast F_t
1	138	150.25
2	136	139.50
3	152	157.25
4	127	143.50
5	151	138.00
6	130	127.50
7	119	138.25
8	153	141.50
9	—	140.50
10	—	167.25

Table 2-8: *The last eight beer production figures and forecasts obtained by taking the average of each month over the past four years.*

2/4/1 Standard statistical measures

If Y_t is the actual observation for time period t and F_t is the forecast for the same period, then the error is defined as

$$e_t = Y_t - F_t. \tag{2.12}$$

one-step forecast

Usually, F_t is calculated using data $Y_1, \dots, Y_{t-1}$. It is a *one-step* forecast because it is forecasting one period ahead of the last observation used in the calculation. Therefore, we describe e_t as a *one-step forecast error*. It is the difference between the observation Y_t and the forecast made using all the observations up to but not including Y_t.

If there are observations and forecasts for n time periods, then there will be n error terms, and the following standard statistical measures can be defined:

mean error

$$\text{ME} = \frac{1}{n}\sum_{t=1}^{n} e_t \tag{2.13}$$

mean absolute error

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{2.14}$$

mean squared error

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2. \tag{2.15}$$

Table 2-9 illustrates the computation of these standard statistical measures.

Equation (2.12) can be used to compute the error for each period. These can then be averaged as in equation (2.13) to give the mean error. However, the ME is likely to be small since positive and negative errors tend to offset one another. In fact, the ME will only tell you if there is systematic under- or over-forecasting, called the forecast bias. It does not give much indication as to the size of the typical errors.

Therefore, the MAE is defined by first making each error positive by taking its absolute value, and then averaging the results. A similar idea is behind the definition of MSE. Here the errors are made positive by squaring each one, then the squared errors are averaged. The MAE has the advantage of being more interpretable and is easier to explain to non-specialists. The MSE has the advantage of being easier to handle mathematically (and so it is often used in statistical optimization).

Each of these statistics deals with measures of accuracy whose size depends on the scale of the data. Therefore, they do not facilitate comparison across different time series and for different time intervals. An error of 10 Ml when forecasting monthly beer production is quite different from an error of 10 Ml when forecasting

Period t	Observation Y_t	Forecast F_t	Error $Y_t - F_t$	Absolute Error $\lvert Y_t - F_t\rvert$	Squared Error $(Y_t - F_t)^2$
1	138	150.25	−12.25	12.25	150.06
2	136	139.50	−3.50	3.50	12.25
3	152	157.25	−5.25	5.25	27.56
4	127	143.50	−16.50	16.50	272.25
5	151	138.00	13.00	13.00	169.00
6	130	127.50	2.50	2.50	6.25
7	119	138.25	−19.25	19.25	370.56
8	153	141.50	11.50	11.50	132.25
Total			−29.75	83.75	1140.20

$$\text{ME} = -29.75/8 = -3.72 = \quad \text{using equation (2.13)}$$
$$\text{MAE} = 83.75/8 = 10.47 = \quad \text{using equation (2.14)}$$
$$\text{MSE} = 1140.20/8 = 142.52 = \quad \text{using equation (2.15)}$$

Table 2-9: *Computations of the standard measures for the beer data.*

annual beer production or an error of 10 Ml when forecasting the water consumption of a city. To make comparisons like these, we need to work with relative or percentage error measures.

relative or percentage error

First we need to define a relative or percentage error as

$$\text{PE}_t = \left(\frac{Y_t - F_t}{Y_t}\right) \times 100. \qquad (2.16)$$

Then the following two relative measures are frequently used:

mean percentage error

mean absolute percentage error

$$\text{MPE} = \frac{1}{n}\sum_{t=1}^{n} \text{PE}_t \qquad (2.17)$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |\text{PE}_t| \qquad (2.18)$$

Equation (2.16) can be used to compute the percentage error for any time period. These can then be averaged as in equation (2.17) to give the mean percentage error. However, as with the ME, the MPE is likely to be small since positive and negative PEs tend to offset one

Period t	Observation Y_t	Forecast F_t	Error $Y_t - F_t$	Percent Error $\left(\frac{Y_t-F_t}{Y_t}\right)100$	Absolute Percent Error $\left\lvert\frac{Y_t-F_t}{Y_t}\right\rvert 100$
1	138	150.25	−12.25	−8.9	8.9
2	136	139.50	−3.50	−2.6	2.6
3	152	157.25	−5.25	−3.5	3.5
4	127	143.50	−16.50	−13.0	13.0
5	151	138.00	13.00	8.6	8.6
6	130	127.50	2.50	1.9	1.9
7	119	138.25	−19.25	−16.2	16.2
8	153	141.50	11.50	7.5	7.5
Total				−26.0	62.1

$$\text{MPE} = -26.0/8 = -3.3\% \quad \text{using equation (2.17)}$$
$$\text{MAPE} = 62.1/8 = 7.8\% \quad \text{using equation (2.18)}$$

Table 2-10: *Computations of the percentage measures for the beer data.*

another. Hence the MAPE is defined using absolute values of PE in equation (2.18). Table 2-10 shows how to compute the PE, MPE, and MAPE measures.

From the point of view of the ultimate user of forecasting, knowing that the MAPE of a method is 5% means a great deal more than simply knowing that the MSE is 183. However, the MAPE is only meaningful if the scale has a meaningful origin. For example, one would not use MAPE for assessing the accuracy of temperature forecasting since the common temperature scales (Fahrenheit and Celsius) have fairly arbitrary zero points. Difficulties also arise when the time series contains zeros, since the percentage error (2.16) cannot then be computed. (When the time series values are very close to zero, the computations involving PE can be meaningless.)

2/4/2 Out-of-sample accuracy measurement

The summary statistics described so far measure the goodness of fit of the model to *historical* data. Such fitting does not necessarily imply good forecasting. An MSE or MAPE of zero can always be obtained over-fitting

in the fitting phase by using a polynomial of sufficiently high order. Over-fitting a model to a data series, which is equivalent to including randomness as part of the generating process, is as bad as failing to identify the systematic pattern in the data.

A second drawback of these measures of accuracy is that that different methods use different procedures in the fitting phase. For example, smoothing methods (Chapter 4) are highly dependent upon initial forecasting estimates; decomposition methods (Chapter 3) include the trend-cycle in the fitting phase as though it were known; regression methods (Chapter 5–6) minimize the MSE by giving equal weight to all observations; and Box-Jenkins methods (Chapter 7) minimize the MSE by a non-linear optimization procedure. Thus, comparison of such methods on a single criterion is of limited value.

test set
holdout set

These problems can be overcome by measuring true *out-of-sample* forecast accuracy. That is, the total data are divided into an "initialization" set and a "test" set or "holdout" set. Then, the initialization set is used to estimate any parameters and to initialize the method. Forecasts are made for the test set. Since the test set was not used in the model fitting, these forecasts are genuine forecasts made without using the values of the observations for these times. The accuracy measures are computed for the errors in the test set only.

2/4/3 Comparing forecast methods

comparing forecast methods

None of these measures give a good basis of comparison as to the gains in accuracy made by applying a specific forecasting method. Does a MSE of 5 or a MAPE of 3.2% indicate a good or bad forecasting performance? One basis for making such a comparison is to define some very simple naïve methods against which the performance of more sophisticated methods can be compared.

Naïve Forecast 1

We have found it useful to define two different naïve methods of forecasting for use as a basis in evaluating other methods in a given situation. The first is referred to as Naïve Forecast 1 or NF1. This method uses the most recent observation available as a forecast. Table 2-11 shows NF1 used to forecast the monthly beer production. Each forecast is produced by taking the value of the previous month's production. So the forecast for January 1995 is the production figure from December 1994, the forecast for February 1995 is the production figure from January 1995, and so on.

Period	Observation	NF1 Forecast	Absolute Error	Absolute Percent Error
t	Y_t	F_t	$\lvert Y_t - F_t \rvert$	$\left\lvert \frac{Y_t - F_t}{Y_t} \right\rvert 100$
1	138	182	44	31.9
2	136	138	2	1.5
3	152	136	16	10.5
4	127	152	25	19.7
5	151	127	24	15.9
6	130	151	21	16.2
7	119	130	11	9.2
8	153	119	34	22.2
Total			177	127.1

$$\text{MAE} = 177/8 = 22.1 \qquad \text{using equation (2.17)}$$
$$\text{MAPE} = 127.1/8 = 15.9\% \qquad \text{using equation (2.18)}$$

Table 2-11: *Computations of the percentage measures for the NF1 forecasts of the beer data.*

The difference between the MAE or MAPE obtained from a more sophisticated method of forecasting and that obtained using NF1 provides a measure of the improvement attainable through use of that more sophisticated forecasting method. This type of comparison is much more useful than simply computing the MAPE or MAE of the first method, since it provides a basis for evaluating the relative accuracy of those results. In this case, the first forecasting method achieved a MAPE of 7.8% compared to about twice that for NF1 and a MAE of 10.5 Ml compared to 22.1 Ml for NF1. Clearly the first method provides much better forecasts.

Naïve Forecast 2

A second naïve method of forecasting has also been found to be extremely useful as a basis for evaluating more formal forecasting methods. This method is referred to as Naïve Forecast 2 or NF2 and goes beyond NF1 in that it considers the possibility of seasonality in the series. Since seasonality often accounts for a substantial percentage of the fluctuation in a series, this method can frequently do much better than NF1 and yet is still a very simple straightforward approach. The procedure is to remove seasonality from the original data in order to obtain seasonally adjusted data. Once the seasonality has been removed, NF2 is comparable to NF1 in that it uses the most recent seasonally adjusted value as a forecast for the next seasonally adjusted value. In practice, NF2 allows one to decide whether or not the improvement obtained from going beyond a simple seasonal adjustment of the data is worth the time and cost involved.

2/4/4 Theil's U-statistic

The relative measures in the previous section all give equal weight to all errors in contrast to the MSE, which squares the errors and thereby emphasizes large errors. It would be helpful to have a measure that considers both the disproportionate cost of large errors and provides a relative basis for comparison with naïve methods. One measure that has these characteristics is the U-statistic developed by Theil (1966).

This statistic allows a relative comparison of formal forecasting methods with naïve approaches and also squares the errors involved so that large errors are given much more weight than small errors. The positive characteristic that is given up in moving to Theil's U-statistic as a measure of accuracy is that of intuitive interpretation. This difficulty will become more apparent as the computation of this statistic and its application are examined. Mathematically, Theil's U-statistic is defined as

Theil's U-statistic

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}(\text{FPE}_{t+1} - \text{APE}_{t+1})^2}{\sum_{t=1}^{n-1}(\text{APE}_{t+1})^2}} \qquad (2.19)$$

where $\text{FPE}_{t+1} = \dfrac{F_{t+1} - Y_t}{Y_t}$ (forecast relative change)

and $\text{APE}_{t+1} = \dfrac{Y_{t+1} - Y_t}{Y_t}$ (actual relative change).

Equation (2.19) is actually very straightforward, as can be seen

Period t	Observation Y_t	Forecast F_t	Numerator $\left(\frac{F_{t+1}-Y_{t+1}}{Y_t}\right)^2$	Denominator $\left(\frac{Y_{t+1}-Y_t}{Y_t}\right)^2$
1	138	150.25	0.0006	0.0002
2	136	139.50	0.0015	0.0138
3	152	157.25	0.0118	0.0271
4	127	143.50	0.0105	0.0357
5	151	138.00	0.0003	0.0193
6	130	127.50	0.0219	0.0072
7	119	138.25	0.0093	0.0816
8	153	141.50	—	—
Total			0.0560	0.1849

$$\text{Theil's } U = \sqrt{\frac{0.0560}{0.1849}} = 0.550$$

Table 2-12: *Computations involved in determining Theil's U-statistic for the beer forecasts.*

by simplifying it to the form shown in (2.20). When the values of FPE_{t+1} and APE_{t+1} are substituted into equation (2.19), the result is

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{F_{t+1} - Y_t - Y_{t+1} + Y_t}{Y_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{Y_{t+1} - Y_t}{Y_t}\right)^2}} = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{F_{t+1} - Y_{t+1}}{Y_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{Y_{t+1} - Y_t}{Y_t}\right)^2}}. \qquad (2.20)$$

Comparing the numerator of equation (2.20) with equation (2.18) shows that it is similar to what was defined previously as the MAPE of a given forecasting method. The denominator is equivalent to the numerator with F_{t+1} replaced by Y_t. So it is similar to the MAPE of NF1. Thus, the U-statistic is an accuracy measure that incorporates both concepts.

Table 2-12 shows how to compute Theil's U-statistic for the beer data.

Theil's U-statistic can be better understood by examining its interpretation. The value of the U-statistic given by equation (2.19) will be 0 only if $\text{FPE}_{t+1} = \text{APE}_{t+1}$ for $t = 1, 2, \ldots, n-1$. That in turn occurs only when the forecasts are exact (give 0 error). Alternatively, the U-statistic will have a value of 1 only when FPE_{t+1} is equal to 0. That would be the case only if the errors in the forecasting method were the same as those that would be obtained by forecasting no change at all in the actual values. That is comparable to assuming an NF1 approach. If FPE_{t+1} is in the opposite direction of APE_{t+1}, the U-statistic will be greater than unity since the numerator will be larger than the denominator. The ranges of the U-statistic can thus be summarized as follows:

$U = 1$: the naïve method is as good as the forecasting technique being evaluated.

$U < 1$: the forecasting technique being used is better than the naïve method. The smaller the U-statistic, the better the forecasting technique is relative to the naïve method.

$U > 1$: there is no point in using a formal forecasting method, since using a naïve method will produce better results.

2/4/5 ACF of forecast error

One other tool for analyzing forecast error needs to be mentioned. The autocorrelation function of the one-step forecast errors is very useful in determining if there is any remaining pattern in the errors (or residuals) after a forecasting model has been applied. This is not a measure of accuracy per se, but rather can be used to indicate if the forecasting method could be improved.

For example, suppose the naïve forecast method NF1 was used for the beer data. We analyze the forecast errors by a time plot and an ACF plot as shown in Figure 2-7.

Note that there is a pattern remaining in these errors—the January value is particularly low each year. Clearly, there is not what would be called a random set of errors. The autocorrelation statistics are sensitive to such patterns.

The ACF plot in the lower panel of Figure 2-7 tells a similar story. The autocorrelation at lag 12 is much larger than the other

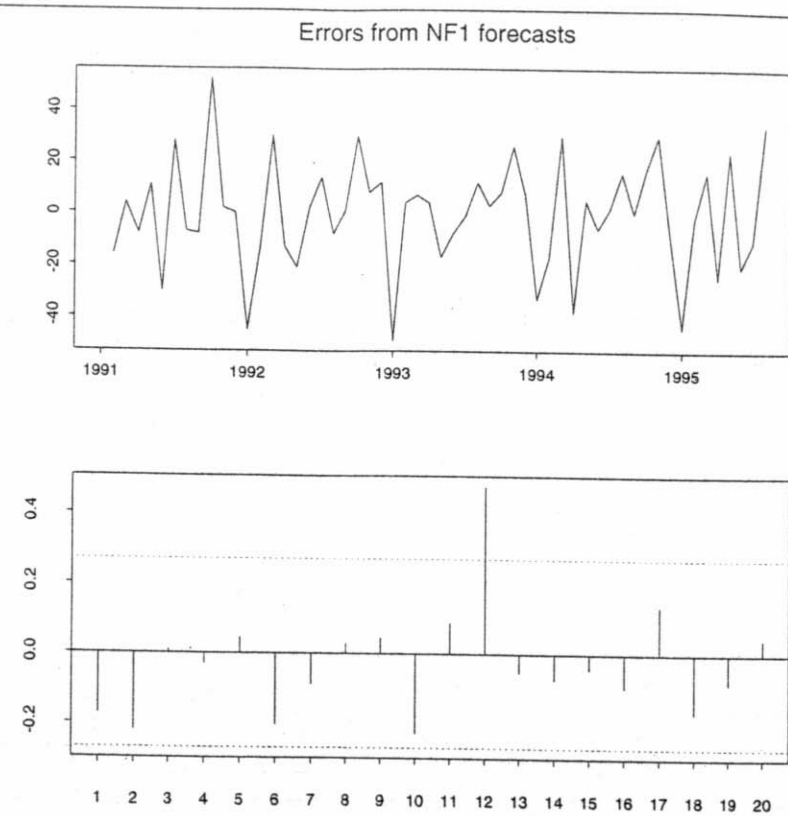


Figure 2-7: *Top: Forecast errors obtained by applying the NF1 method to the beer data. Bottom: The ACF of the forecast errors.*

autocorrelations. This shows there is some seasonal pattern in the error series (the low January values are 12 months apart).

It is important not to read too much into the other autocorrelations shown in Figure 2-7. With random series, no autocorrelations will be exactly zero, even if the series is entirely random. These small fluctuations around zero are quite acceptable and do not indicate there is information in the series which is not being captured by the forecasts.

It is useful to have a benchmark to determine if an autocorrelation is significantly large. A simple rule is to only consider autocorrelations that are larger than the *critical values* of $2/\sqrt{n}$ in magnitude. (Why this works will be discussed in Chapter 7.) For the beer data, $n = 56$, so the critical values are at $2/\sqrt{56} = 0.27$. Figure 2-7 shows the boundary as two horizontal lines at ± 0.27. Autocorrelations that fall within these boundaries can be safely ignored. Autocorrelations that fall outside the boundaries suggest there may be some additional information in the series which is not being captured by the forecast method.

ACF critical values

In this example, only the autocorrelation at lag 12 falls outside the critical values. This is another indication that there is some seasonality in the forecast errors, and that the forecast method could be improved. This is not surprising since the forecast method NF1 does not use the seasonal pattern in producing forecasts.

2/5 Prediction intervals

It is usually desirable to provide not only forecast values but accompanying uncertainty statements, usually in the form of *prediction intervals.* This is useful because it provides the user of the forecasts with "worst" or "best" case estimates and with a sense of how dependable the forecast is, and because it protects the forecaster from the criticism that the forecasts are "wrong." Forecasts cannot be expected to be perfect and intervals emphasize this.

prediction intervals

Prediction intervals are usually based on the MSE because it provides an estimate of the variance of the one-step forecast error. So the square root of the MSE is an estimate of the standard deviation of the forecast error. The usual assumption for constructing prediction intervals is that the forecast errors are normally distributed with zero mean. Under this assumption, an approximate prediction interval for the next observation is

$$\boxed{F_{n+1} \pm z\sqrt{\text{MSE}}\,.}$$

The value of z determines the width and probability of the prediction interval. For example, $z = 1.96$ gives a 95% prediction interval. That is, the interval has probability of 95% of containing the true value, as yet unknown.

For other percentages, different values of z can be used. The table below gives the most common values of z.

z	Probability
0.674	0.50
1.000	0.68
1.150	0.75
1.282	0.80
1.645	0.90
1.960	0.95
2.576	0.99

Other values of z are given in Table D of Appendix III. The rules of thumb given on page 32 are based on this table.

The September 1995 forecast for the beer data can be calculated as the average production for the previous four Septembers. This gives $F_{n+1} = 140.50$. Table 2-9 gives the MSE for the beer forecasts as 142.52. So a 90% prediction interval is

$$140.50 \pm 1.645\sqrt{142.52} = 140.50 \pm 19.64 = [120.86, 160.01].$$

That is, we can be 90% sure that the actual beer production figure for September 1995 will lie between 120.86 and 160.01 megaliters. (In fact, it was 144 megaliters.) A similar calculation gives the October 1995 forecast as 167.25 Ml with a 90% prediction interval of

$$167.25 \pm 1.645\sqrt{142.52} = 167.25 \pm 19.64 = [147.61, 186.89].$$

The actual figure for October was 166 Ml.

Although October is two periods ahead of the last observation, these are both *one-step* forecasts since the forecast method is based only on data from previous years. The forecast method treats the data from each month as separate series. That is why we could use the same MSE value in computing the prediction interval for October. If we had used the NF1 method, then the forecast for October would have been a two-step forecast, and the MSE would not have been valid for calculating a prediction interval.

This procedure only works for one-step forecasts since the MSE is based on one-step forecasts. For multi-step forecasts, we need to h-step MSE

modify the MSE to be based on multi-step forecasts. One approach is to define the h-step MSE as

$$\text{MSE}_h = \frac{1}{n-h}\sum_{t=h+1}^{n}(e_t^{(h)})^2$$

where $e_t^{(h)}$ is the error from making an h-step forecast of the observation at time t. Then, if we assume the h-step forecast error is normally distributed with zero mean, we have the prediction interval

$$\boxed{F_{n+h} \pm z\sqrt{\text{MSE}_h}\,.}$$

Before calculating any prediction intervals in this way, the errors should be checked to ensure the assumptions of zero mean and normal distribution have been met.

2/6 Least squares estimates

In Chapter 1 we introduced two kinds of quantitative forecasting models: explanatory models and time series models. An explanatory model for GNP is of the form

$$\begin{aligned}\text{GNP} \;&=\; f(\text{monetary and fiscal policies, inflation,}\\ &\qquad \text{capital spending, imports, exports, error})\end{aligned}$$

whereas a time series model is of the form

$$\text{GNP}_{t+1} = f(\text{GNP}_t, \text{GNP}_{t-1}, \text{GNP}_{t-2}, \text{GNP}_{t-3}, \ldots, \text{error}).$$

random error

Neither model can be exact. That is why the error term is included on the right-hand sides of these equations. The error term represents variations in GNP that are not accounted for by the relationship f.

In both cases, what is observed as the output of the system is dependent on two things: the functional relationship governing the system (or the pattern, as it will be called from now on) and randomness (or error). That is,

$$\text{data} = \text{pattern} + \text{error}. \tag{2.21}$$

The critical task in forecasting is to separate the pattern from the error component so that the former can be used for forecasting.

The general procedure for estimating the pattern of a relationship, whether explanatory or time series, is through fitting some functional form in such a way as to minimize the error component of equation (2.21). One form of this estimation is least squares. This approach is very old (developed first by Gauss in the 1800s) and is the one most widely used in classical statistics.

least squares

The name *least squares* is based on the fact that this estimation procedure seeks to minimize the sum of the squared errors in equation (2.21). The example shown below illustrates the basis of the least squares method. Its application to all types of functional forms (i.e., linear or non-linear) is analogous to that shown here.

Suppose that the manager of a supermarket wants to know how much a typical customer spends in the store. The manager might start by taking a sample of say 12 clients, at random, obtaining the results shown in Table 2-13.

From Table 2-13, it is clear that not all customers spend the same amount. Some of the variation might be explained through factors such as time of the day, day of the week, discounts offered, maximum or minimum amount of checks cashed, and so on, while part of the variation may be random or unexplainable. For purposes of this illustration, it will be assumed that no variation can be explained through explanatory or time series relationships. In such a case, the store manager faced with finding an appropriate estimator to describe

Client	Amount Spent ($)	Client	Amount Spent ($)
1	9	7	11
2	8	8	7
3	9	9	13
4	12	10	9
5	9	11	11
6	12	12	10

Table 2-13: *Sample expenditures for supermarket clients.*

		Estimate of $\hat{Y}=7$		Estimate of $\hat{Y}=10$		Estimate of $\hat{Y}=12$	
Client	Amount spent	Error[a]	Error squared	Error	Error squared	Error	Error squared
1	9	2	4	−1	1	−3	9
2	8	1	1	−2	4	−4	16
3	9	2	4	−1	1	−3	9
4	12	5	25	2	4	0	0
5	9	2	4	−1	1	−3	9
6	12	5	25	2	4	0	0
7	11	4	16	1	1	−1	1
8	7	0	0	−3	9	−5	25
9	13	6	36	3	9	1	1
10	9	2	4	−1	1	−3	9
11	11	4	16	1	1	−1	1
12	10	3	9	0	0	−2	4
SSE (sum of squared errors)			144		36		84
MSE (mean squared error)			12		3		7

[a]Error = amount spent − estimated value.

Table 2-14: *Mean squared errors for estimates of client expenditure.*

the data may take a fixed value as an estimate. Having made this decision, the manager might decide to select an estimate in such a way as to minimize the mean (average) squared error. This could be done by trial and error. Suppose we denote the estimate by the symbol $\hat{Y}$. The manager tries values of $\hat{Y}=7$, $\hat{Y}=10$, and $\hat{Y}=12$. The resulting mean squared errors are shown in Table 2-14.

From Table 2-14 it is clear that the squared error is least when the manager chooses 10 as the estimate. However, there may be a better estimate. Figure 2-8 shows the resulting MSEs for all estimates from 0 through 20, and it can be seen that the MSEs form a parabola. Furthermore, the minimum value on this parabola is indeed at the point where the estimate is 10. Thus, the minimum MSE will be achieved when the value of the estimate is 10, and we say that 10 is the least squares estimate of customer spending.

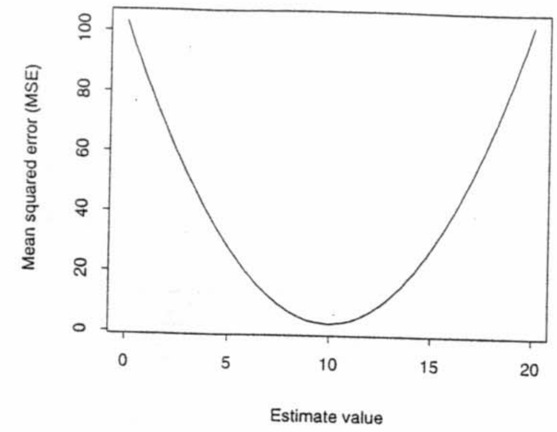


Figure 2-8: *Relating MSE to estimate value.*

Because Figure 2-8 is a mathematical function whose properties can be found exactly, it is not necessary to use trial and error to find the estimator that minimizes the MSE. Rather, this value can be found mathematically with the help of differentiation. The first step is to rewrite equation (2.21) so as to isolate the error on the left-hand side:

$$\text{error} = \text{data} - \text{pattern}. \tag{2.22}$$

As usual, the error will be denoted by e, the data by Y, and the estimate pattern by $\hat{Y}$. In addition, the subscript i ($i = 1,2,3,\ldots,12$) will be added to denote the ith customer. Using this notation, equation (2.22) becomes: $e_i = Y_i - \hat{Y}$. Then the sum of squared errors is

$$\text{SSE} = \sum_{i=1}^{12} e_i^2 = \sum_{i=1}^{12}(Y_i - \hat{Y})^2 \tag{2.23}$$

and the mean squared error is

$$\text{MSE} = \frac{1}{12}\text{SSE} = \frac{1}{12}\sum_{i=1}^{12}(Y_i - \hat{Y})^2.$$

Now initially the value $\hat{Y}$ will not be known, but the store manager wants that value of $\hat{Y}$ which will minimize the sum of the squared errors (or, equivalently, to minimize the MSE).[4]

This can be found by taking the derivative of (2.23), setting it equal to zero, and solving for $\hat{Y}$, as follows:

$$\frac{d\,\mathrm{SSE}}{d\hat{Y}} = -2\sum_{i=1}^{12}(Y_i - \hat{Y}) = 0$$

$$\text{so that} \qquad \sum_{i=1}^{12}(Y_i - \hat{Y}) = 0$$

$$\text{or} \qquad \sum_{i=1}^{12} Y_i - 12\hat{Y} = 0$$

$$\text{which implies} \qquad \hat{Y} = \frac{1}{12}\sum_{i=1}^{12} Y_i = \bar{Y}. \qquad (2.24)$$

Solution (2.24) is easily recognized as the mean of the data, and it gives a value that minimizes the sum of the squared errors. Applying (2.24) to the store manager's data in Table 2-13 gives

$$\hat{Y} = \bar{Y} = \frac{1}{12}\sum_{i=1}^{12} Y_i = \frac{120}{12} = 10.$$

minimum MSE estimate

This value is the minimum point of Figure 2-8. As a single point estimate of the pattern of the data, the mean fits the data as closely as possible, given the criterion of minimizing the MSE. While the mean is a somewhat simple estimate of the data in most situations, the procedure of least squares that was used to determine a MSE estimate can be applied no matter how complex or sophisticated the estimation situation is.

It is of course possible to minimize some other criterion (e.g., MAE) instead of minimizing the MSE. However, minimizing MAE is not as easy mathematically as minimizing the MSE. Also, squaring the

[4] Note that minimizing SSE (the sum of squared errors) is the "least squares" procedure. Dividing by n (which is 12 in the example given) gives the MSE. Thus minimizing the MSE is an exactly equivalent procedure.

errors magnifies (or gives more weight to) extreme values, and this result is attractive because large errors are less desirable than small errors. (Many cost relationships are quadratic in nature, suggesting the appropriateness of squaring.)

2/6/1 Discovering and describing relationships

If the measurable output of a system is viewed as data that include a pattern and some error, a major consideration in forecasting, whether explanatory or time series, is to identify and fit the most appropriate pattern (functional form) so as to minimize the MSE. The basic procedure is illustrated by comparing two different methods for forecasting the price of a Japanese automobile using the data for the 19 vehicles listed in Table 2-5.

The first method is simply to use the mean as a forecast (as was done for the store expenditure data). Table 2-15 gives the price data along with the forecasts and the resulting errors.

The forecasting model underlying Table 2-15 is

$$\hat{Y} = a. \qquad (2.25)$$

That is, the forecasts are the same for all vehicles because we are not using other information about the vehicles. The value of a is estimated from the data to be equal to the mean of the data. Hence, $\hat{Y} = a = \bar{Y}$.

Using the mean as the estimate of the pattern might be acceptable if we had no other information about these vehicles. But we have already seen that price is correlated with mileage. We can use the mileage information to form a more accurate estimate of price than the mean. Figure 2-9 shows a straight line fitted to the price-mileage relationship. Notice that the errors (vertical distances from the points to the line) are smaller in this plot than the errors obtained using the mean as an estimate.

straight line model

Using a straight line forecast model with mileage as the explanatory variable means

$$\hat{Y} = a + b \times \text{mileage} \qquad (2.26)$$

where a and b are suitably chosen values representing the intercept and slope of the line. The values of a and b can be chosen in the same

	Price	Mean value	Error	Squared error
	14.944	13.938	1.01	1.01
	14.799	13.938	0.86	0.74
	24.760	13.938	10.82	117.12
	14.929	13.938	0.99	0.98
	13.949	13.938	0.01	0.00
	17.879	13.938	3.94	15.53
	11.650	13.938	−2.29	5.24
	23.300	13.938	9.36	87.65
	17.899	13.938	3.96	15.69
	21.498	13.938	7.56	57.15
	13.249	13.938	−0.67	0.48
	9.599	13.938	−4.34	18.83
	10.989	13.938	−2.95	8.70
	13.945	13.938	0.01	0.00
	13.071	13.938	−0.87	0.75
	6.599	13.938	−7.34	53.86
	9.410	13.938	−4.53	20.50
	5.866	13.938	−8.07	65.16
	6.488	13.938	−7.45	55.50
Total	264.823	264.823	0.00	524.88

$\bar{Y} = \frac{264.823}{19} = 13.938$ $\qquad$ $\text{MSE} = \frac{524.88}{19} = 27.63$

Table 2-15: *The mean as an estimate of the price of Japanese automobile.*

simple linear regression

way as a was chosen in (2.25): that is, by minimizing the MSE. This procedure is known as *simple linear regression* and will be examined in detail in Chapter 5. The mechanics of how a and b are calculated are not important at this point.

For these data, the values of a and b which minimize the MSE are 32.1 and −0.735 respectively. So the forecast model is

$$\hat{Y} = 32.1 - 0.735 \times \text{mileage}. \qquad (2.27)$$

This line is shown in Figure 2-9.

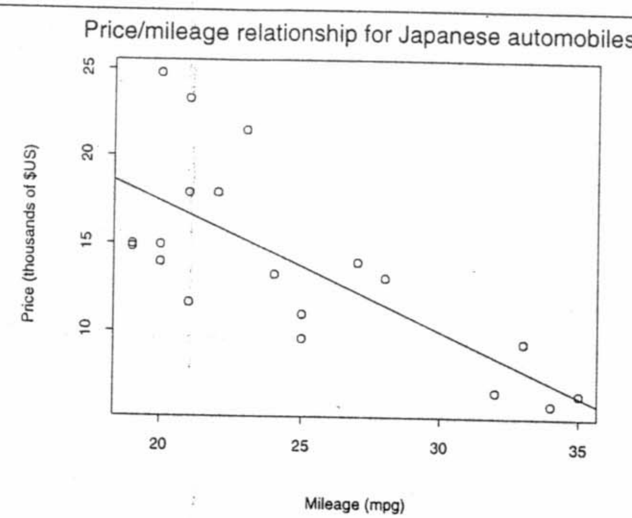


Figure 2-9: *Straight line estimate of vehicle price using mileage as an explanatory variable.*

causal relationship

It is important not to confuse an *explanatory* relationship (such as that between price and mileage) with a *causal* relationship. Of course, higher mileage does not *cause* lower prices—both depend on other variables such as the size of the engine. The straight line forecast model (2.27) is used when we know the mileage of a vehicle and wish to predict its price.

For forecasting purposes, either the mean forecasting model or the straight line forecasting model can be used to predict the price of a Japanese automobile not listed. For example, if a vehicle has a mileage of 23 mpg, the mean gives a forecast of \$13,938 while the straight line model gives a forecast of $[32.1 - 0.736(23)] \times 1,000 =$ \$15,176. From historical data, one would expect the straight line model to be better since it had a smaller MSE value (13.00 compared with 27.63).

A little care must be taken in comparing MSE values (or any other measure of accuracy) from different models. More complicated

Price ($'000)	Mileage (mpg)	Forecast ($'000)	Error	Squared error
14.944	19	18.118	−3.174	10.072
14.799	19	18.118	−3.319	11.014
24.760	20	17.382	7.378	54.429
14.929	20	17.382	−2.453	6.019
13.949	20	17.382	−3.433	11.788
17.879	21	16.647	1.232	1.518
11.650	21	16.647	−4.997	24.971
23.300	21	16.647	6.653	44.261
17.899	22	15.912	1.987	3.949
21.498	23	15.176	6.322	39.962
13.249	24	14.441	−1.192	1.421
9.599	25	13.706	−4.107	16.866
10.989	25	13.706	−2.717	7.381
13.945	27	12.235	1.710	2.923
13.071	28	11.500	1.571	2.468
6.599	32	8.559	−1.960	3.840
9.410	33	7.823	1.587	2.517
5.866	34	7.088	−1.222	1.493
6.488	35	6.353	0.135	0.018
		Sum of squared errors (SSE):		264.913

$$\text{MSE} = 246.913/19 = 13.00$$

Table 2-16: *Straight line estimate of vehicle price using mileage as explanatory variable. Straight line formula:* $\hat{Y} = 32.1 - 0.735(\textit{mileage})$.

models generally have smaller MSE values, even if they do not give more accurate forecasts. This is because they measure the goodness of fit of the model to historical data, rather than true out-of-sample forecasting performance. However, in this case the scatterplot (Figure 2-5) and negative correlation (Equation 2.9) both indicate that mileage should be included in the forecasting model. Also, the straight line model has only half the MSE of the mean model. Such a reduction is unlikely to be due to the additional

complexity of the straight line model.

In general, there is no way that a statistical method can automatically determine the best pattern (functional form) to describe a given set of data. Rather, this decision must be based on judgment. Then a statistical method can be used to fit the specified pattern in such a way as to minimize the MSE.

2/7 Transformations and adjustments

Sometimes, adjusting the historical data will lead to a simpler and more interpretable forecasting model. In this section we deal with three kinds of adjustment: mathematical transformations (such as logarithms and square roots) are discussed in Section 2/7/1; adjustments to remove data variation due to the effects of the calendar are discussed in Section 2/7/2; and adjustments due to population changes and inflation are discussed in Section 2/7/3.

2/7/1 Mathematical transformations

Figure 2-10 shows monthly Australian electricity production data, the same data that were plotted in Figure 1-2a (p. 7). Notice that the size of the annual seasonal variation increases as the level of the series increases. At the start of the series, the total variation throughout the year was only about 300 million kwh, but in the most recent years, when the production is very high, the total variation is over 2,500 million kwh. Clearly any forecasts for these data must take account of the obvious increasing trend, the strong seasonal pattern, and this increasing variation with level. A mathematical transformation is a convenient method for accounting for the increasing variation.

mathematical transformation

One such transformation is the square root function. The top plot in Figure 2-11 shows the square roots of the electricity production data. The new data set was formed simply by taking the square root of each observation in the original data set. This mathematical transformation has helped in reducing the variation in the size of the annual cycles, making it easier to forecast these data than those shown in Figure 2-10.

square root transformation

Square roots are only one kind of transformation that can be

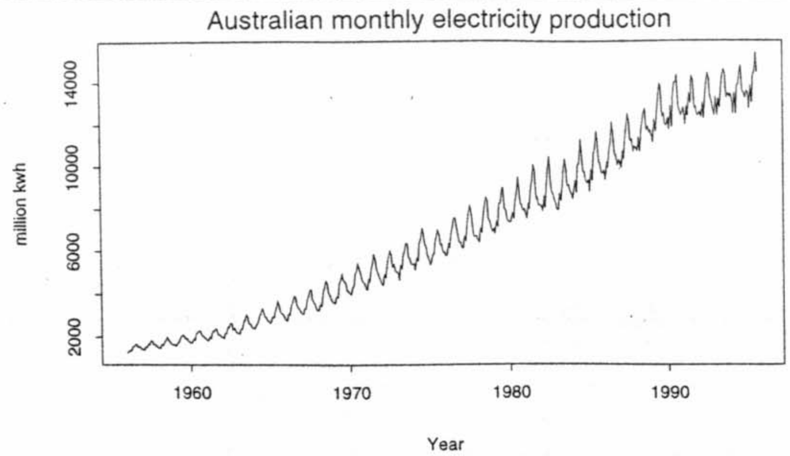


Figure 2-10: *Monthly Australian electricity production from January 1956 to August 1995. Note the increasing variation as the level of the series increases.*

used in this way. Many other transformations are possible, but in practice the square root and logarithm are most useful. Logarithms, in particular, are useful because they are more interpretable: changes in a log value are relative (percent) changes on the original scale.

Other useful transformations are given in Table 2-17. Here we denote the original observations as $Y_1, \dots, Y_n$ and the transformed observations as $W_1, \dots, W_n$. Figure 2-11 displays the electricity data transformed using some of the transformations given in Table 2-17, showing the effect of the increasing strength of the transformations.

power transformation

Each of the transformations in Table 2-17 is a member of the family of power transformations:

$$W_t = \begin{cases} -Y_t^p, & p < 0; \\ \log(Y_t), & p = 0; \\ Y_t^p, & p > 0. \end{cases} \tag{2.28}$$

For $p = 1$ the transformation is simply $W_t = Y_t$, so this leaves the data alone. Choosing $p = \frac{1}{2}$ gives a square root and $p = -1$

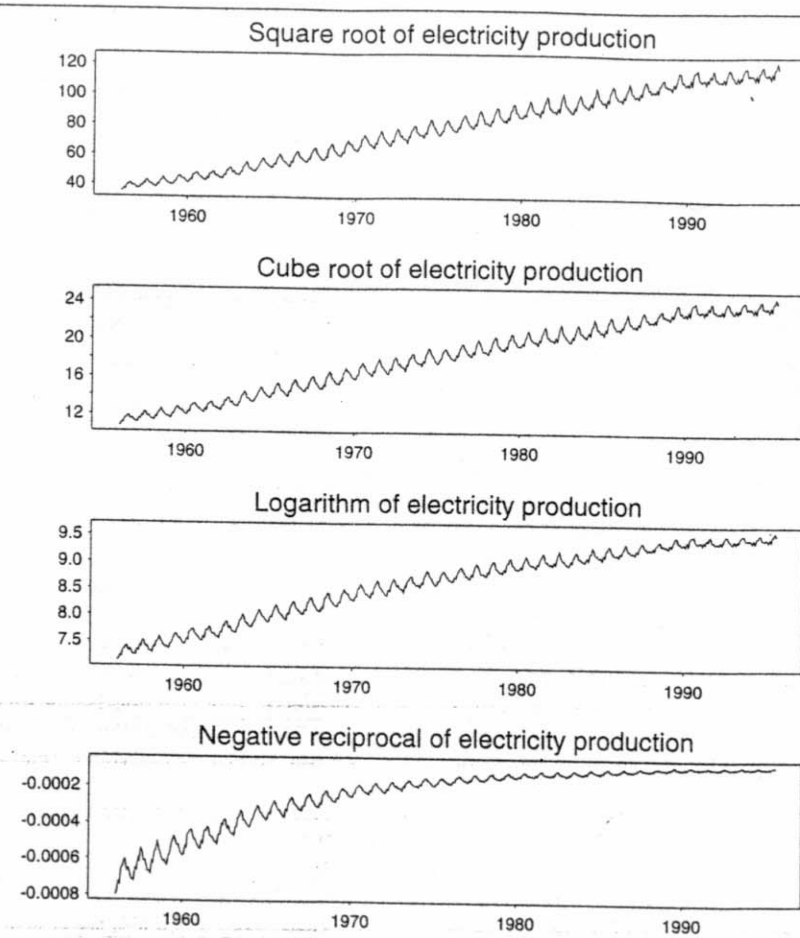


Figure 2-11: *Transformations of the electricity production data. Of these, either a square root, cube root, or a log transformation could be used to stabilize the variation to form a series which has variation approximately constant over the series.*

Square root	$W_t = \sqrt{Y_t}$	↓
Cube root	$W_t = \sqrt[3]{Y_t}$	Increasing
Logarithm	$W_t = \log(Y_t)$	strength
Negative reciprocal	$W_t = -1/Y_t$	↓

Table 2-17: *Mathematical transformations for stabilizing variation.*

gives the negative reciprocal. It might seem artificial to define the power transformation for $p = 0$ to be the logarithm, but it actually belongs there because Y_t^p for p close to zero behaves much like the logarithm. For $p < 0$, the negative of the power transformation is used so that all transformations result in increasing functions (i.e., the transformed variable increases as Y_t increases). The parameter p can be any number if the data are positive, but p must be greater than zero if the data have zeros. If the data have negative values, no power transformation is possible unless they are adjusted first by adding a constant to all values.

back-transforming

Forecasts are calculated on the transformed data rather than the original data. But since we are really interested in forecasts of the original data, not the transformed data, we must reverse the transformation (or *back-transform*) to obtain forecasts on the original scale. For example, the reverse of the square root function is the square function, the reverse of the logarithm function is the exponential function, and so on. Generally, the reverse power transformations are given by

$$Y_t = \begin{cases} (-W_t)^{1/p}, & p < 0; \\ \exp(W_t), & p = 0; \\ (W_t)^{1/p}, & p > 0. \end{cases}$$

For example, if we were to forecast the square root of the electricity data, we could obtain forecasts on the original scale by taking the square of the forecasts of the square root data.

It is preferable to choose a simple value of p to give a transformation such as those given in Table 2-17. Models and forecasts of a time series are relatively insensitive to the value of p chosen—nearby values of p will produce similar results. This is seen in Figure 2-11

where either $p = 1/2$, $p = 1/3$, or $p = 0$ could be used to stabilize the variation. Also, simple values of p such as 0, -1, or $1/2$ make the results much easier to interpret than a number like $p = 0.38463$. Very often it is found that no transformation (i.e., $p = 1$) is needed.

When the data have been transformed, then prediction intervals also need to be transformed back to the original scale. The simplest way to proceed is to apply the inverse transform to the end points of the prediction interval. So, if logarithms have been used, and the forecast on the log scale is F_{n+1} and the prediction interval is (L_{n+1}, U_{n+1}), then the forecast on the original scale is $e^{F_{n+1}}$ with the prediction interval $(e^{L_{n+1}}, e^{U_{n+1}})$. Note that these prediction intervals need not be symmetric around the forecast.

Empirical studies which have considered the merits of mathematical transformations have demonstrated that, for many series, transformation does not often have a major effect on forecast accuracy (Makridakis and Hibon, 1979; Makridakis et al., 1982; Meese and Geweke, 1984). This is because most forecast methods place more weight on the most recent data. Therefore the small annual variation earlier in the electricity series is unlikely to influence the forecasts very much. Only when the series is rapidly changing in variation will mathematical transformations make a large difference to the forecasts.

However, the MSE (and the other measures of accuracy) gives equal weight to all data and so prediction intervals will be affected by transformations. In calculating prediction intervals, it is assumed that the variation is approximately constant over the series.

2/7/2 Calendar adjustments

Some of the variation in a time series may be due to the variation in the number of days (or trading days) each month. It is a good idea to adjust for this known source of variation to allow study of other interesting features.

month length adjustment

Month length can have quite a large effect, since length can differ by about $\frac{31-28}{30} = 10\%$. If this is not removed, it shows up as a seasonal effect, which may not cause problems with forecasts though it does make any seasonal pattern hard to interpret. It is easily adjusted for:

$$W_t = Y_t \times \frac{\text{no. of days in an average month}}{\text{no. of days in month } t} = Y_t \times \frac{365.25/12}{\text{no. of days in month } t}.$$

Figure 2-12 shows monthly milk production per cow over a period of 14 years. Month length will, of course, affect the total monthly milk production of a cow. Figure 2-12 shows the milk production data adjusted for month length. The simpler pattern will lead to better forecasts and easier identification of unusual observations.

trading day adjustment

Trading day adjustment is similar to month length adjustment but is not completely predictable. Trading days adjustments are often necessary because a given month may not have the same number of working, or trading, days in different years. In some industries such as retail sales and banks, this factor becomes very important, since it can have a significant influence on the level of sales. This source of variation occurs in monthly data when there is also a weekly cycle, since the proportions of the various days in a given month vary from year to year. For example, March may have four or five Sundays, and if Sunday is a non-trading day this must be accounted for. While the various proportions are completely predictable from the calendar (like month length adjustment) the effects of the various days are not predictable so this must be estimated.

In the simplest case, days are classified as either trading or non-trading days, and all trading days are assumed to have the same effect. In this case the adjustment is analogous to month length:

$$W_t = Y_t \times \frac{\text{no. of trading days in an average month}}{\text{no. of trading days in month } t}.$$

where Y_t has already been adjusted for month length and transformed if necessary. More complicated cases are discussed in Section 6/2/2.

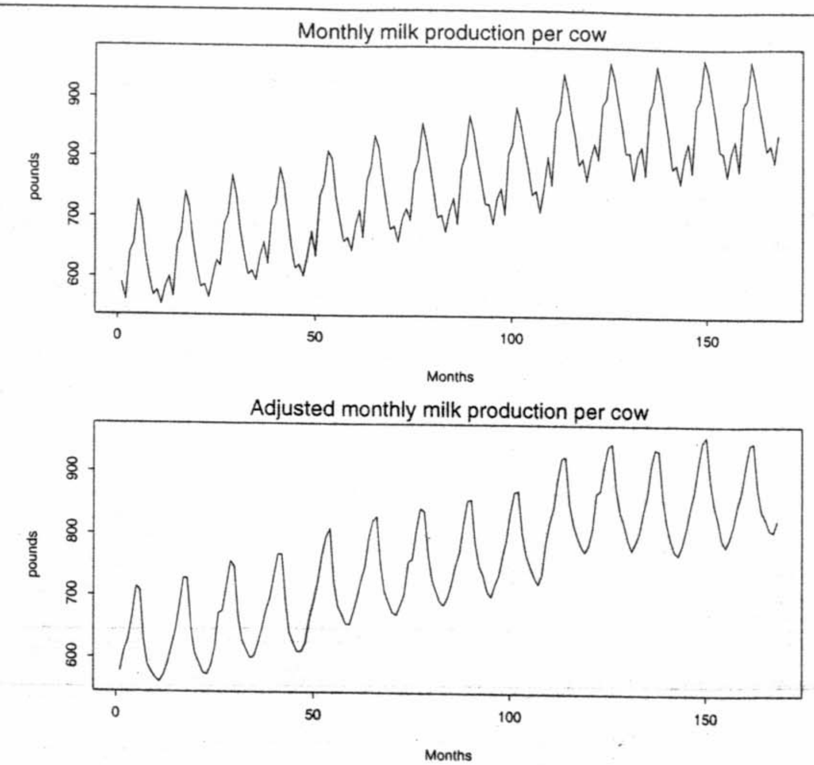


Figure 2-12: *Monthly milk production per cow over 14 years (Source: Cryer, 1986). The second graph shows the data adjusted for the length of month. This yields a simpler pattern enabling better forecasts and easier identification of unusual observations.*

2/7/3 Adjustments for inflation and population changes

adjusting for inflation

One obvious source of variation that afflicts time series is the effect of inflation or changes in population. For example, when forecasting the price of a new motor vehicle, it is essential to take into account the effect of inflation. A $15,000 vehicle this year is not the same as a $15,000 vehicle 10 years ago. The standard approach is to use equivalent value in 1990 dollars (for instance). Then the data are directly comparable and forecasts will not be affected by this additional source of variation.

adjusting for population changes

Adjusting for population changes is similar. For example, when forecasting the number of public transport users in a city, it is preferable to take into account the effect of population changes. In this case, the data could be adjusted by the total *number* of people in the city. Rather than forecasting the total number of public transport users, it will probably be more accurate to forecast the *proportion* of people who are public transport users. Demographic studies are needed to provide forecasts of population and these can then be used to obtain forecasts of the number of public transport users in the future. A more refined approach would be to produce forecasts for different age groups and/or different socioeconomic groups.

Appendix 2-A
Notation for quantitative forecasting

Quantitative forecasts are based on data, or observations, that describe some factor of interest. In this book a single observed value will be represented by Y_t. (See Table 2-18.) This variable can be the actual number of units sold, the cost of production, the advertising budget, price per unit, gross national product, or any other event of interest, as long as it can be quantified. The objective of forecasting is to predict future values of Y. The individual forecasts will be denoted by F_t, or $\hat{Y}_t$, and the error by e_t, where the error is the difference between the actual value and the forecast value for observation i:

$$e_t = Y_t - \hat{Y}_t \qquad \text{or} \qquad e_t = Y_t - F_t.$$

In time series forecasting and in explanatory forecasting, when the data are taken at equal time intervals, n will denote the present time period, $n-1$ last period, $n-2$ two periods ago, and so on. A period can be a day, a week, a month, quarter, year, and so forth. The forecasts usually will be for future time periods such as $n+1$.

	Observed Values						Forecasted Values			
	Y_1	Y_2	Y_3	$\ldots$	Y_{n-1}	Y_n				
Period t	1	2	3	$\ldots$	$n-1$	n	$n+1$	$n+2$	$\ldots$	$n+m$
Estimated values	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\ldots$	$\hat{Y}_{n-1}$	$\hat{Y}_n$	$\hat{Y}_{n+1}$	$\hat{Y}_{n+2}$	$\ldots$	$\hat{Y}_{n+m}$
or	F_1	F_2	F_3	$\ldots$	F_{n-1}	F_n	F_{n+1}	F_{n+2}	$\ldots$	F_{n+m}
Error	e_1	e_2	e_3	$\ldots$	e_{n-1}	e_n				
						↑ Present				

Table 2-18: *Notation used in time series forecasting.*

Appendix 2-B
Summation sign Σ

In order to simplify the manipulation of expressions involving the adding of many numbers, it is convenient to use a summation sign, Σ. The use of this sign and the elements of notation mentioned previously can be demonstrated using the data in Table 2-19.

Based on Table 2-19,

Y_t is the actual sales value,

$\hat{Y}_t$ or F_t is the forecast values for sales, and

e_t is the error or difference $Y_t - \hat{Y}_t$.

If one wants the sum of the errors, it can be obtained from

$$\begin{aligned} e_1 + e_2 + e_3 + \cdots + e_{23} &= \sum_{t=1}^{23} e_t \\ \text{or} \quad 3 - 3 - 2 - \cdots - 7 &= -2. \end{aligned}$$

Below and above the summation sign are the "limits" showing the variable which is indexing the sum (t) and the range of the summation (from 1 to 23). If it is obvious what the limits are, sometimes they are omitted.

The cumulative sales for the years 1985 through 1994 can be obtained from

$$\begin{aligned} \sum_{t=11}^{20} Y_t &= Y_{11} + Y_{12} + Y_{13} + \cdots + Y_{20} \\ &= 175 + 175 + 176 + \cdots + 251 \\ &= 2071. \end{aligned}$$

Year	Period t	No. of Units Sold: Actual	No. of Units Sold: Forecast	Error	Year	Period t	No. of Units Sold: Actual	No. of Units Sold: Forecast	Error
1975	1	123	120	3	1986	12	175	173	2
1976	2	125	128	−3	1987	13	176	177	−1
1977	3	133	135	−2	1988	14	192	188	−4
1978	4	140	138	2	1989	15	199	195	−4
1979	5	144	148	−4	1990	16	210	215	5
1980	6	158	157	3	1991	17	225	230	5
1981	7	161	155	6	1993	18	230	236	−6
1982	8	160	168	−6	1993	19	238	242	−4
1983	9	163	168	−5	1994	20	251	248	3
1984	10	171	171	0	1995	21	259	255	4
1985	11	175	176	−1	1996	22	275	263	12
					1997	23	283	290	−7

Table 2-19: *Use of quantitative forecasting notation.*

The following rules apply to the use of summation signs:

1. $\sum_{t=1}^{n} \bar{Y} Y_t = \bar{Y} \sum_{t=1}^{n} Y_t$, where $\bar{Y}$ is the sample mean (therefore a constant) of the variable Y_t .

2. $\sum_{t=1}^{n} \bar{Y} = n\bar{Y}$.

3. $\sum_{t=1}^{n} (Y_t - \hat{Y}_t) = \sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} \hat{Y}_t$.

4. $\sum_{t=1}^{n} (Y_t - \bar{Y}) = \sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} \bar{Y} = \sum_{t=1}^{n} Y_t - n\bar{Y}$.

5. $$\begin{aligned} \sum_{t=1}^{n} (Y_t - \bar{Y})^2 &= \sum_{t=1}^{n} (Y_t^2 - 2\bar{Y} Y_t + \bar{Y}^2) \\ &= \sum_{t=1}^{n} Y_t^2 - 2\bar{Y} \sum_{t=1}^{n} Y_t + n\bar{Y}^2 \\ &= \sum_{t=1}^{n} Y_t^2 - n\bar{Y}^2 \\ &= \sum_{t=1}^{n} Y_t^2 - (\sum_{t=1}^{n} Y_t)^2/n . \end{aligned}$$

References and selected bibliography

ARMSTRONG, J.S. and F. COLLOPY (1992) Error measures for generalizing about forecasting methods: empirical comparisons (with discussion), *International Journal of Forecasting*, **8**, 69–111.

BIERMAN, H., C.P. BONINI, and W.H. HAUSMAN (1991) *Quantitative analysis for business decisions*, 8th ed., Homewood, Ill.: Richard D. Irwin.

CHATFIELD, C. (1993) Calculating interval forecasts (with discussion), *Journal of Business and Economic Statistics*, **11**, 121–144.

COGGER, K.O. (1979) "Time series analysis and forecasting with an absolute error criterion," in *TIMS Studies, Vol. 12, Forecasting*, S. Makridakis and S.C. Wheelwright, eds., Amsterdam: North-Holland, 89–102.

CONLEY, D.L., G.S. KRAHENBUHL, L.N. BURKETT, and A.L. MILLAR (1981) Physiological correlates of female road racing performance, *Research Quarterly Exercise Sport*, **52**, 441–448.

CONSUMER REPORTS (1990) Vol. 55, No. 4, Consumers Union, Mt Vernon, N.Y..

CRYER, J.D. (1986) *Time series analysis*, Belmont, Cal.: Duxbury Press.

FILDES, R. (1992) The evaluation of extrapolative forecasting methods (with discussion), *International Journal of Forecasting*, **8**, 81–111.

HOLLOWAY, C.A. (1979) *Decision making under uncertainty*, Englewood Cliffs, N.J.: Prentice-Hall.

KOOSIS, D.J. (1972) *Business statistics*, New York: John Wiley & Sons.

LOCKE, F.M. (1972) *Business mathematics*, New York: John Wiley & Sons.

MAKRIDAKIS, S. and M. HIBON. (1979) Accuracy of forecasting: an empirical investigation (with discussion), *Journal of the Royal Statistical Society A*, **142**, 97–145.

MAKRIDAKIS, S., A. ANDERSEN, R. CARBONE, R. FILDES, M. HIBON, R. LEWANDOWSKI, J. NEWTON, E. PARZEN, and R. WINKLER (1982) The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting*, **1**, 111–153.

MCLAUGHLIN, R.L. (1975) The Real Record of the Economic Forecasters, *Business Economics*, **10**, No. 3, 28–36.

MEESE, R. and J. GEWEKE (1984) A comparison of autoregressive univariate forecasting procedures for macroeconomic time series, *Journal of Business and Economic Statistics*, **2**, 191–200.

MONTGOMERY, D.C., L.A. JOHNSON, and J.S. GARDINER (1990) *Forecasting and time series analysis*, 2nd ed., New York: McGraw-Hill.

NEWBOLD, P. and T. BOS (1994) *Introductory business forecasting*, 2nd ed., Cincinnati, Ohio: South-Western Publishing Co..

SPURR, W.A. and C.P. BONINI (1973) *Statistical analysis for business decisions*, revised ed., Homewood, Ill.: Richard D. Irwin.

STEECE, B.M. (1982) "The Evaluation of Forecasts," in *Handbook of Forecasting*, S. Makridakis and S.C. Wheelwright, eds., New York: John Wiley & Sons, 457–468.

THEIL, H. (1966) *Applied economic forecasting*, Amsterdam: North-Holland, 26–32.

RICE, J.A. (1995) *Mathematical statistics and data analysis*, 2nd ed., Belmont, Cal.: Duxbury Press.

WONNACOTT, T.H. and R.J. WONNACOTT (1990) *Introductory statistics for business and economics*, 4th ed., New York: John Wiley & Sons.

Exercises

2.1 Table 2-20 gives average monthly temperatures in Paris.

(a) What is your best estimate of the average temperature in June 1995?

(b) Make a time plot of the data. Is there any time pattern in the temperature readings?

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1994	7.6	7.1	8.3	11.5	13.7	17.2	18.5	19.7	15.1	8.9	8.5	8.5
1995	7.7	6.9	6.1	10.5	12.9							

Table 2-20: *Average monthly temperature in Paris (degrees Celsius).*

2.2 For each of the following series, what sort of time patterns would you expect to see?

(a) Monthly retail sales of computer disks for the past 10 years at your local store.

(b) Hourly pulse rate of a person for one week.

(c) Daily sales at a fast-food store for six months.

(d) Weekly electricity consumption for your local area over the past 10 years.

2.3 For each of the following series on the web page, make a graph of the data (using a computer package), describe the main features and, if transforming seems appropriate, do so and describe the effect.

(a) Monthly total of people on unemployed benefits in Australia (January 1956–July 1992).

(b) Daily morning temperature of a cow for 75 days.

(c) Number of lynx trapped annually in the McKenzie River district of northwest Canada (1821–1934).

(d) Monthly total of accidental deaths in the United States (January 1973–December 1978).

(e) Quarterly production of bricks (in millions of units) at Portland, Australia (March 1956–September 1994).

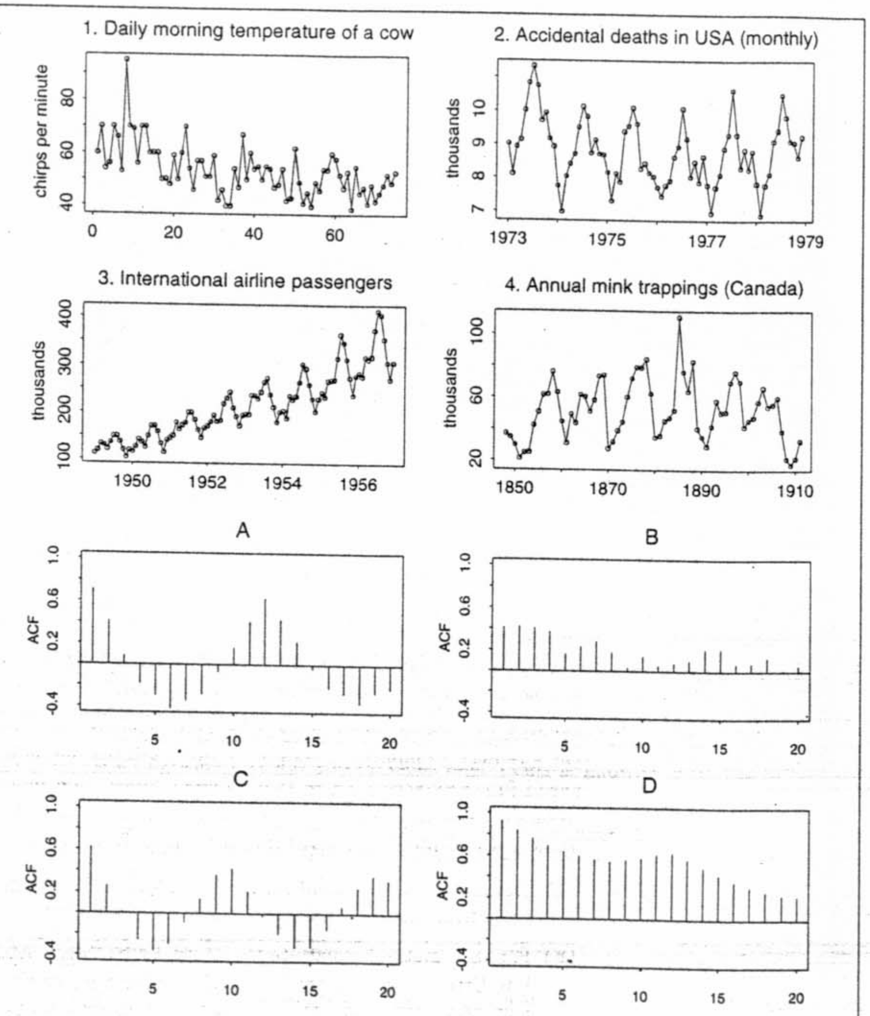


Figure 2-13: *Four time series with their ACFs.*

2.4 In the graphs on the previous page, four time series are plotted along with their ACFs. Which ACF goes with which time series?

2.5 Table 2-21 shows data on the performance of 14 trained female distance runners. The variables measured are the running time (minutes) in a 10 kilometer road race and the maximal aerobic power collected during the week following the run.

(a) Calculate the mean, median, MAD, MSD, and standard deviation for each variable.

(b) Which of these statistics give a measure of the center of data and which give a measure of the spread of data?

(c) Calculate the correlation of the two variables and produce a scatterplot of Y against X.

(d) Why is it inappropriate to calculate the autocorrelation of these data?

X	61.32	55.29	52.83	57.94	53.31	51.32	52.18	52.37	57.91	53.93	47.88	47.41	47.17	51.05
Y	39.37	39.80	40.03	41.32	42.03	42.37	43.93	44.90	44.90	45.12	45.60	46.03	47.83	48.55

Table 2-21: *Running times* (Y) *and maximal aerobic capacity* (X) *for 14 female runners. Source: Conley et al. (1981).*

2.6 Column 1 on the following page is the actual demand for product E15 over 20 months. Columns 2 and 3 are the one-month ahead forecasts according to two different forecasting models to be discussed in Chapter 4. (Method 1 gives forecasts from Table 4-4 and Method 2 gives forecasts from Table 4-6.)

(a) Plot the actual demand on a graph along with the forecasts from the two methods.

(b) For each method, compute the Mean Error, Mean Absolute Error, Mean Squared Error, Mean Percentage Error, and Mean Absolute Percentage Error using equations (2.13) through (2.18).

(c) Repeat Part (b) using columns 1 and 3 below. Which forecasting method appears to be better?

Period	(1) Actual Demand	(2) Method 1 Forecast	(3) Method 2 Forecast
1	139	157	170
2	137	145	162
3	174	140	157
4	142	162	173
5	141	149	164
6	162	144	158
7	180	156	166
8	164	172	179
9	171	167	177
10	206	169	180
11	193	193	199
12	207	193	202
13	218	202	211
14	229	213	221
15	225	223	232
16	204	224	235
17	227	211	225
18	223	221	232
19	242	222	233
20	239	235	243

2.7 Download the Dow Jones index from the web page and produce a time plot of the series using a computer package.

(a) Calculate the change in the index for each day by subtracting the value for the previous day. (This is known as "differencing" the data and is discussed in Chapter 6.)

(b) Forecast the change in the index for each of the next 20 days by taking the average of the historical changes.

(c) From these forecasts, compute forecasts for the original index for each of the 20 days.

(d) Add the forecasts to the graph.

(e) Show that the graphed forecasts are identical to extending the line drawn between the first and last observations.

		1950	32	1960	482	1970	5289	1980	11043
		1951	38	1961	814	1971	5811	1981	11180
		1952	39	1962	991	1972	6294	1982	10732
		1953	50	1963	1284	1973	7083	1983	11112
		1954	70	1964	1702	1974	6552	1984	11465
		1955	69	1965	1876	1975	6942	1985	12271
		1956	111	1966	2286	1976	7842	1986	12260
1947	11	1957	182	1967	3146	1977	8514	1987	12249
1948	20	1958	188	1968	4086	1978	9269	1988	12700
1948	29	1959	263	1969	4675	1979	9636	1989	13026

Table 2-22: *Japanese motor vehicle production (1947–1989) in thousands. Source:* World motor vehicle data, *Motor Vehicle Manufacturers Association of U.S. Inc., Detroit, 1991.*

2.8 Japanese motor vehicle production for 1947–1989 is given in Table 2-22.

(a) Plot the data in a time plot. What features of the data indicate a transformation may be appropriate?

(b) Transform the data using logarithms and do another time plot.

(c) Calculate forecasts for the transformed data for each year from 1948 to 1990 using Naïve Forecast 1.

(d) Compute the forecast errors and calculate the MSE and MAPE from these errors.

(e) Transform your forecast for 1990 back to the original scale by find the exponential of your forecast in (c). Add the forecast to your graph.

(f) From the graphs you have made, can you suggest a better forecasting method?

(g) The world motor vehicle market was greatly affected by the oil crisis in 1973–1974. How did it affect Japanese motor vehicle production? If this information could be included in the forecasts, how would it affect the values of the MSE and MAPE?

3

TIME SERIES DECOMPOSITION