Chapter 1

Introduction

Figure 1.1: This is Not What This Book is About

1.1 Welcome

Forecasting is important — forecasts are constantly made in business, finance, economics, government, and many other fields, and they guide many important decisions. As with anything else, there are good and bad ways to forecast. This book is about the good ways: modern, rigorous, replicable, largely-quantitative statistical/econometric methods – their strengths *and* their limitations. That's why I dislike the above picture of the crystal ball – it bows to the common misconception among the uninitiated that forecasting

is some sort of dubious mystical activity, like fortune telling or astrology. But how could a forecasting book *not* begin with a picture like that? So I decided to lighten up, if only for a moment.

1.2 Who Forecasts, and Why?

Forecasts are made and used in numerous fields. To develop a feel for the tremendous diversity of forecasting applications, let's sketch some of the areas where forecasts feature prominently, and the corresponding diversity of decisions that they support.

One key field is economics, broadly defined. Governments, businesses, policy organizations, central banks, financial services firms, and economic consulting firms around the world routinely forecast major economic variables, such as gross domestic product (GDP), unemployment, consumption, investment, the price level, and interest rates. Governments use such forecasts to guide monetary and fiscal policy, and private firms use them for strategic planning, because economy-wide economic fluctuations typically have industry-level and firm-level effects. In addition to forecasting "standard" variables such as GDP, economists sometimes make more exotic forecasts, such as the stage of the business cycle that we'll be in six months from now (expansion or contraction), the state of future stock market activity (bull or bear), or the state of future foreign exchange market activity (appreciation or depreciation). Again, such forecasts are of obvious use to both governments and firms – if they're accurate!

Another key area is business and all its subfields. These include management strategy of all types including operations management and control (hiring, production, inventory, investment, ...), marketing (pricing distributing, advertising, ...), and accounting (budgeting using revenue and expenditure forecasts), etc. Sales forecasting is a good example. Firms routinely forecast

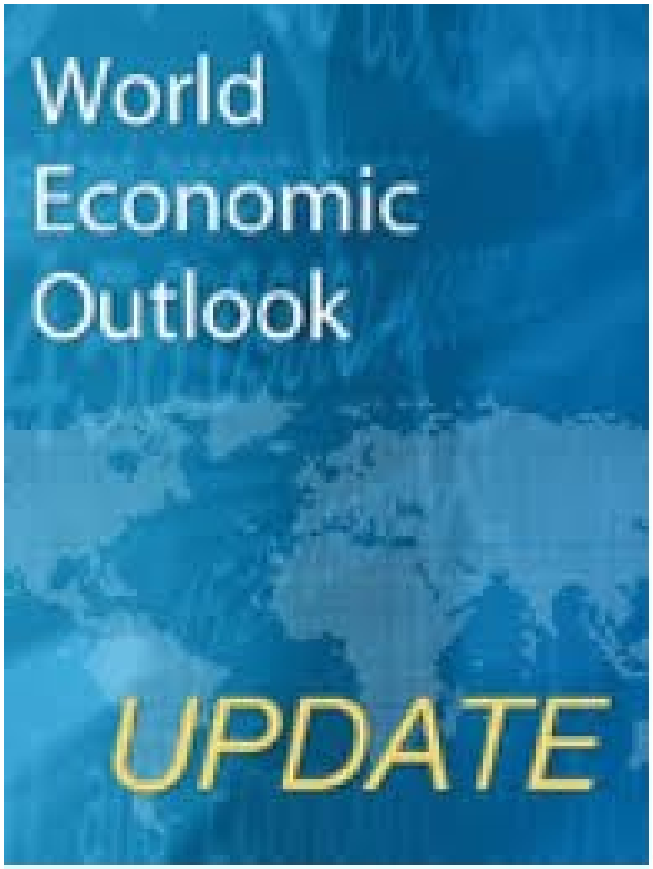

Figure 1.2: Economics: World Economic Outlook

sales to help guide management decisions in inventory management, sales force management, and production planning, as well as strategic planning regarding product lines, new market entry, and so on.

More generally, firms use forecasts to decide what to produce (What product or mix of products should be produced?), when to produce (Should we build up inventories now in anticipation of high future demand? How many shifts should be run?), how much to produce and how much capacity to build (What are the trends in market size and market share? Are there cyclical or seasonal effects? How quickly and with what pattern will a newly-built plant or a newly-installed technology depreciate?), and where to produce (Should we have one plant or many? If many, where should we locate them?). Firms

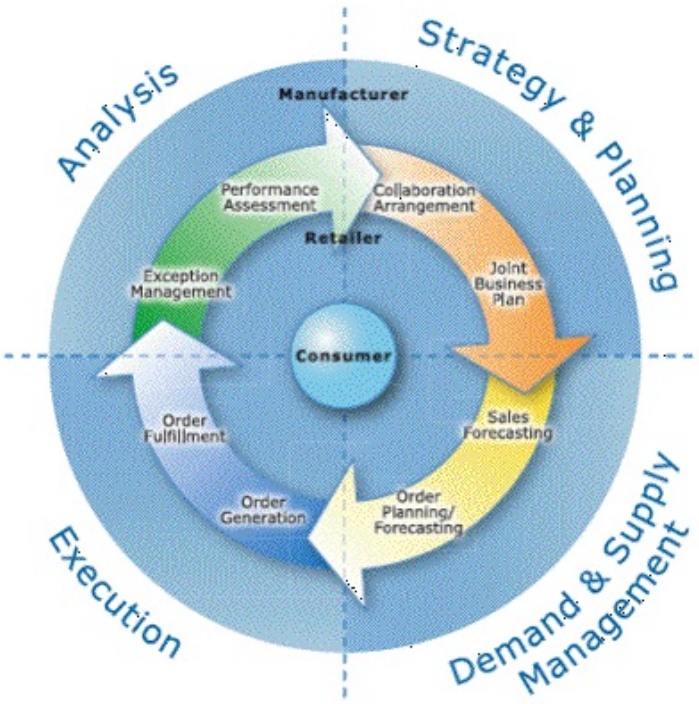

Figure 1.3: Business: Sales Forecasting

also use forecasts of future prices and availability of inputs to guide production decisions.

Forecasting is also crucial in financial services, including asset management, asset pricing, mergers and acquisitions, investment banking, and insurance. Portfolio managers, for example, have keen interest in forecasting asset returns (stock returns, interest rates, exchange rates, and commodity prices) and such forecasts are made routinely. There is endless debate about the success of forecasts of asset returns. On the one hand, asset returns should be very hard to forecast; if they were easy to forecast, you could make a fortune easily, and any such "get rich quick" opportunities should already have been exploited. On the other hand, those who exploited them along the way may well have gotten rich! Thus, we expect that simple, widely-available methods for forecasting should have little success in financial markets, but

Figure 1.4: Finance: A Trading Room

there may well be profits to be made from using new and sophisticated techniques to uncover and exploit previously-unnoticed patterns in financial data (at least for a short time, until other market participants catch on or your own trading moves the market).

Forecasting is similarly central to financial risk management. The forecasting of asset return volatility is related to the forecasting of asset returns. In recent decades, practical methods for volatility forecasting have been developed and widely applied. Volatility forecasts are crucial for evaluating and insuring risks associated with asset portfolios. Volatility forecasts are also crucial for firms and investors who need to price assets such as options and other derivatives.

Finally, forecasting is central to a variety of consulting firms, many of which support the business functions already mentioned. Litigation support is a particularly active area. Forecasting is central to damage assessment (e.g., lost earnings), "but for" analyses and event studies, etc.

The above examples are just the tip of the iceberg. To take another example, demographers routinely forecast the populations of countries and

Figure 1.5: Consulting: Litigation Support

regions all over the world, often in disaggregated form, such as by age, sex, and race. Population forecasts are crucial for planning government expenditure on health care, infrastructure, social insurance, anti-poverty programs, and so forth. Many private-sector decisions, such as strategic product line decisions by businesses, are guided by demographic forecasts of particular targeted population subgroups. Population in turn depends on births, deaths, immigration and emigration, which also are forecasted routinely.

To take just one more example, many events corresponding to crises of various sorts are frequently forecasted. Such forecasts are routinely issued as probabilities. For example, in both consumer and commercial lending, banks generate default probability forecasts and refuse loans if the probability is deemed too high. Similarly, international investors of various sorts are concerned with probabilities of default, currency devaluations, military coups, etc., and use forecasts of such events to inform their portfolio allocation decisions.

The variety of forecasting tasks that we've just sketched was selected to help you begin to get a feel for the depth and breadth of the field. Surely you can think of many more situations in which forecasts are made and used to guide decisions. With so many different forecasting applications, you might fear that a huge variety of forecasting techniques exists, and that you'll have to master all of them. Fortunately, that's not the case. Instead, a rela-

tively small number of tools form the common core of almost all forecasting methods. Needless to say, the details differ if one is forecasting Intel's stock price one day and the population of Scotland the next, but the underlying forecasting principles are identical. We will focus on those underlying core principles.

1.3 Useful Materials

As you begin your study of forecasting, it's important that you begin to develop an awareness of a variety of useful and well-known forecasting textbooks, forecasting journals where original research is published, forecasting software, data sources, professional organizations, etc.

1.3.1 Books

A number of good books exist that complement this one; some are broader, some are more advanced, and some are more specialized. Here we'll discuss a few that are more broad or more advanced. We'll mention more specialized books in subsequent chapters when appropriate.

Wonnacott and Wonnacott (1990) remains a time-honored classic statistics text, which you may wish to consult to refresh your memory on statistical distributions, estimation and hypothesis testing. Anderson et al. (2008) is a well-written and more-recent statistics text, containing a very accessible discussion of linear regression, which we use extensively throughout this book. Pindyck and Rubinfeld (1997) remains one of the all-time great introductory econometrics texts, and it has unusually-strong treatment of time-series and forecasting. It's a useful refresher for basic statistical topics, as well as a good introduction to more advanced **econometric models**.

As a student of forecasting, you'll also want to familiarize yourself with the broader time series analysis literature. Most forecasting methods are

concerned with forecasting time series – data recorded over time. The modeling and forecasting of time series are so important that an entire field called "**time series analysis**" has arisen. Forecasting is intimately related to time series analysis, because quantitative time series forecasting techniques are based on quantitative time series models. Thus, forecasting requires knowledge of time series modeling techniques, and we therefore devote a substantial portion of this book to time series modeling. Chatfield (2006) is a good introductory time series book, which you'll find useful as a background reference. More advanced books, which you may want to consult later, include Granger and Newbold (1986), a classic packed with insight and explicitly oriented toward those areas of time series analysis relevant for forecasting. Finally, Hamilton (1994) and Shumway and Stoffer (2011) are fine advanced texts suitable for Ph.D.-level study.

1.3.2 Online Information and Data

A variety of information of interest to forecasters is available on the web. The best way to learn about what's out there is to spend a few hours searching the web for whatever interests you. Here we mention just a few key "must-know" sites. Resources for Economists, maintained by the American Economic Association, is a fine portal to almost anything of interest to economists. It contains hundreds of links to data sources, journals, professional organizations, and so on. FRED (Federal Reserve Economic Data) at the Federal Reserve Bank of St. Louis is a tremendously convenient source for economic data, as is Quandl. Forecasting Principles has a wealth of data well beyond economics, as well as extensive additional information of interest to forecasters. The National Bureau of Economic Research site has data on U.S. business cycles, and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia has real-time vintage macroeconomic data.

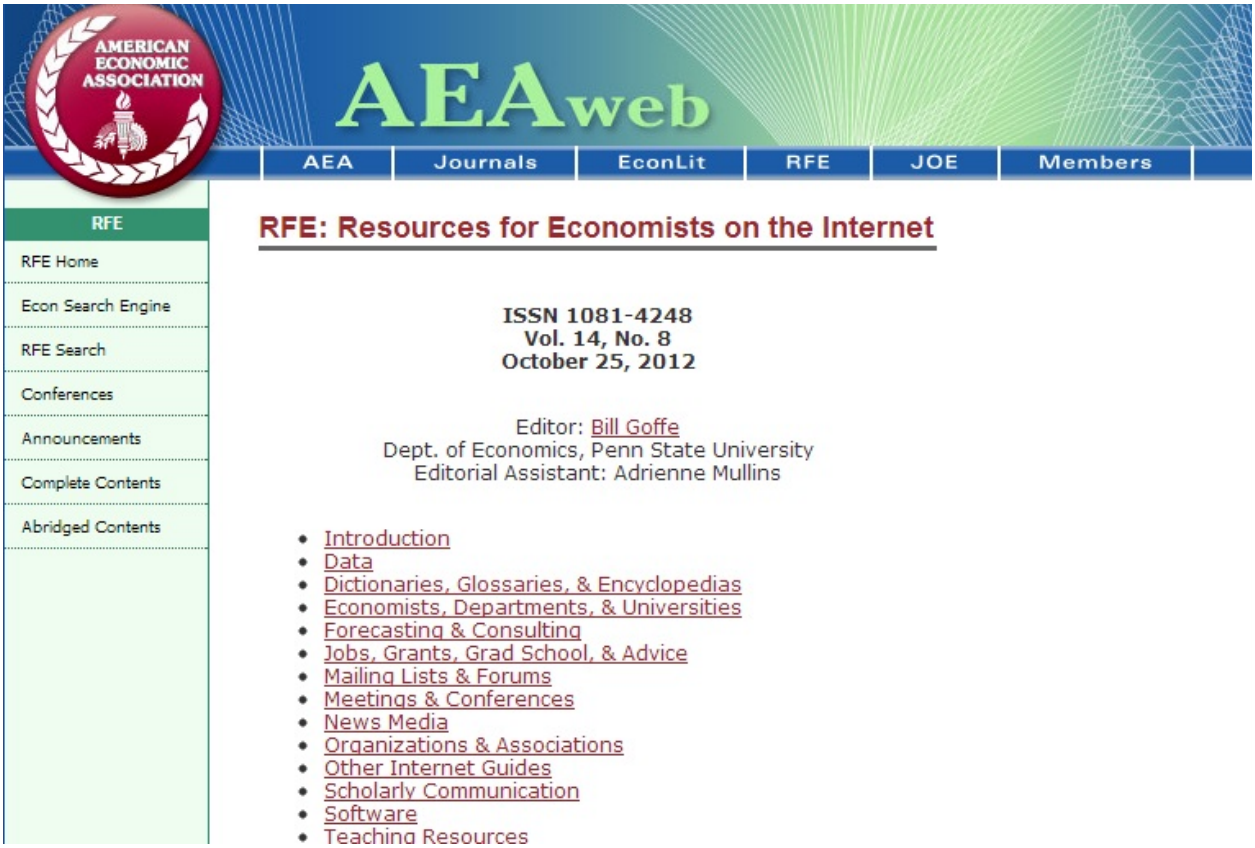

Figure 1.6: Resources for Economists Web Page

1.3.3 Software (and a Tiny bit of Hardware)

Just as some journals specialize exclusively in forecasting, so too do some software packages. But just as the most important forecasting articles often appear in journals much broader than the specialized forecasting journals, so too are forecasting tools scattered throughout econometric / statistical software packages with capabilities much broader than forecasting alone. One of the best such packages is Eviews, a modern object-oriented environment with extensive time series, modeling and forecasting capabilities. It implements almost all of the methods described in this book, and many more. Eviews reflects a balance of generality and specialization that makes it ideal for the sorts of tasks that will concern us, and most of the examples in this book are

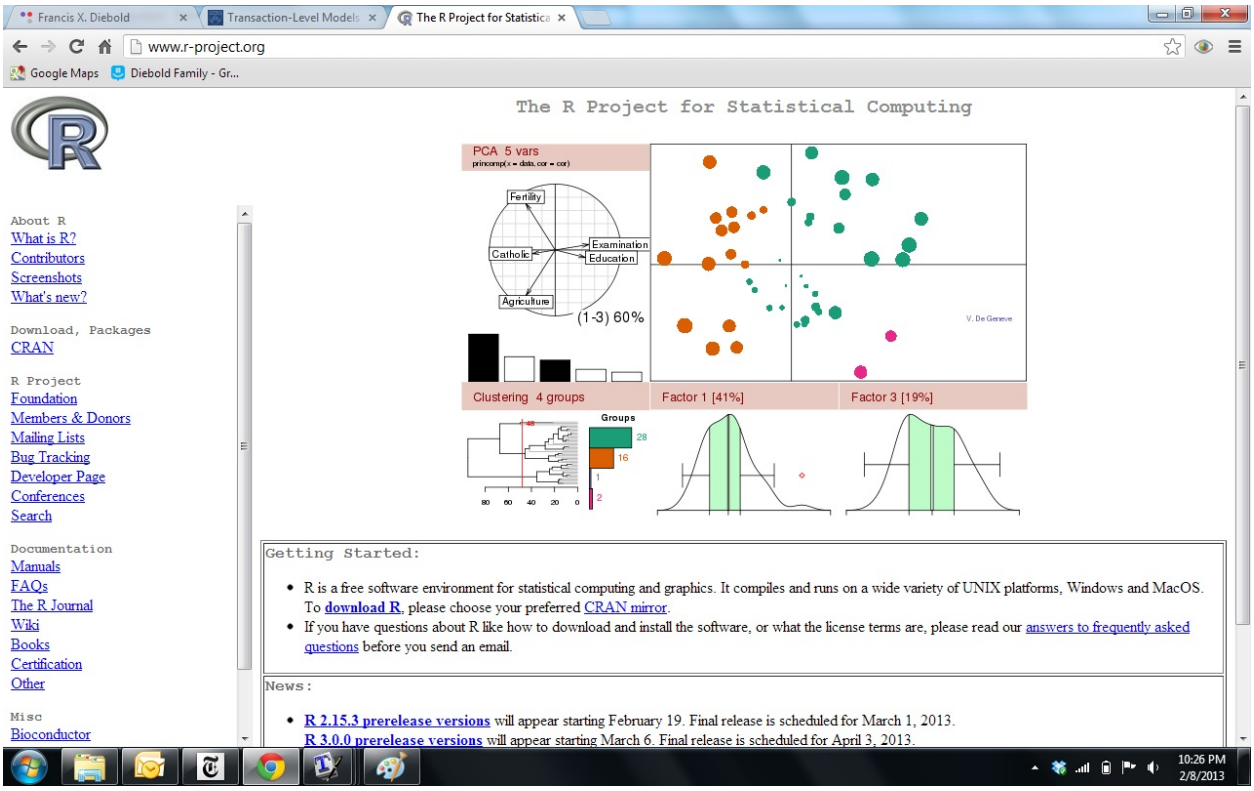

Figure 1.7: The R Homepage

done using it. If you feel more comfortable with another package, however, that's fine – none of our discussion is wed to Eviews in any way, and most of our techniques can be implemented in a variety of packages.

Eviews is an example of a very high-level modeling environment. If you go on to more advanced modeling and forecasting, you'll probably want also to have available slightly lower-level ("mid-level") environments in which you can quickly program, evaluate and apply new tools and techniques. R is one very powerful and popular such environment, with special strengths in modern statistical methods and graphical data analysis. It is available for free as part of a major open-source project. In this author's humble opinion, R is the key mid-level environment for the foreseeable future.[1]

If you need real speed, such as for large simulations, you will likely need

[1] Python and Julia are also powerful mid-level environments.

a low-level environment like Fortran or C++. And in the limit (and on the hardware side), if you need blazing-fast parallel computing for massive simulations etc., graphics cards (graphical processing units, or GPU's) provide stunning gains, as documented for example in Aldrich et al. (2011).

1.3.4 Journals and Professional Organizations

Forecasting cuts across many literatures, including statistics, econometrics, machine learning, and many others.

A number of journals cater to the forecasting community. *International Journal of Forecasting*, for example, is a leading academic forecasting journal, which contains a mixture of newly-proposed methods, evaluation of existing methods, practical applications, and book and software reviews. It is an official journal of the International Institute of Forecasters, which also publishes *Foresight* (a super-applied journal for industry professionals) and *The Oracle* (an online newsletter), and sponsors the Forecasting Principles site. Other organizations with a strong focus on forecasting methods include the Econometric Society and the Society for Financial Econometrics (SoFiE).

Although there are a number of journals devoted to forecasting, its interdisciplinary nature results in a rather ironic outcome: A substantial fraction of the best forecasting research is published not in the forecasting journals, but rather in the broader applied econometrics and statistics journals, such as *Journal of Econometrics*, *Journal of Business and Economic Statistics*, and *Journal of Applied Econometrics*, among many others.

1.4 Final Thoughts

Forecasts guide decisions, and good forecasts help to produce good decisions. In the remainder of this book, we'll motivate, describe, and compare modern forecasting methods. You'll learn how to build and evaluate forecasts and

forecasting models, and you'll be able to use them to improve your decisions.

Forecasting is inextricably linked to the building of **statistical models**. Before we can forecast a variable of interest, we typically build a model for it and estimate the model's parameters using observed historical data. Typically, the estimated model summarizes dynamic patterns in the data; that is, the estimated model provides a statistical characterization of the links between the present and the past. More formally, an estimated **forecasting model** provides a characterization of what we expect in the present, conditional upon the past, from which we infer what to expect in the future, conditional upon the present and past. Quite simply, we use the estimated forecasting model to extrapolate the observed historical data.

In this book we focus on core modeling and forecasting methods that are very widely applicable. We begin by introducing several fundamental issues relevant to any forecasting exercise, and then we treat the construction, use, and evaluation of modern forecasting models. We give special attention to basic methods for forecasting trend, seasonality and cycles, as well as methods for evaluating and combining forecasts. Most chapters contain a detailed application; examples include forecasting retail sales, housing starts, employment, liquor sales, exchange rates, shipping volume, and stock market volatility.

1.5 Tips on How to use this book

As you navigate through the book, keep the following in mind.

- Hyperlinks to internal items (table of contents, index, footnotes, etc.) appear in red.
- Hyperlinks to bibliographical references appear in green.
- Hyperlinks to external items (web pages, video, etc.) appear in cyan.[2]

[2]Obviously web links sometimes go dead. I make every effort to keep them updated in the latest edition

- Hyperlinks to external files appear in blue.
- Many graphics are clickable to reach related material, as are, for example, all pictures in this chapter.
- Key concepts appear in bold. They also appear in the (hyperlinked) index and so can be referenced instantly.
- Additional course-related materials (slides, code, data) appear on the book's website at http://www.ssc.upenn.edu/~fdiebold/Textbooks.html.
- Datasets appear in Appendix C, from which they may be copied and pasted directly.
- The examples that appear throughout should not be taken as definitive or complete treatments – there is no such thing! A good idea is to think of the implicit "Problem 0" in each chapter's Exercises, Problems and Complements (EPC) section as "Critique the modeling and forecasting in this chapter's empirical example, obtain the relevant data, and produce a superior modeling and forecasting analysis."
- All data used in examples are fictitious. Sometimes they are based on real data for various real countries, firms, etc., and sometimes they are artificially constructed. Ultimately, however, any resemblance to particular countries, firms, etc. should be viewed as coincidental and irrelevant.
- The end-of-chapter EPC's are of central importance and should be studied carefully. Exercises are generally straightforward checks of your understanding. Problems, in contrast, are generally significantly more in-

(but no guarantees of course!). If you're encountering an unusual number of dead links, you're probably using an outdated edition.

volved, whether analytically or computationally. Complements generally introduce important auxiliary material not covered in the main text.

1.6 Exercises, Problems and Complements

1. The basic forecasting framework.

   True or false:

   a. The underlying principles of time-series forecasting differ radically depending on the time series being forecast.

   b. Ongoing improvements in forecasting methods will eventually enable perfect prediction.

   c. There is no way to learn from a forecast's historical performance whether and how it could be improved.

2. Data and forecast timing conventions.

   Suppose that, in a particular monthly data set, time $t = 10$ corresponds to September 1960.

   a. Name the month and year of each of the following times: $t+5$, $t+10$, $t+12$, $t+60$.

   b. Suppose that a series of interest follows the simple process $y_t = y_{t-1} + 1$, for $t = 1, 2, 3, ...$, meaning that each successive month's value is one higher than the previous month's. Suppose that $y_0 = 0$, and suppose that at present $t = 10$. Calculate the forecasts $y_{t+5,t}$, $y_{t+10,t}$, $y_{t+12,t}$, $y_{t+60,t}$, where, for example, $y_{t+5,t}$ denotes a forecast made at time t for future time $t + 5$, assuming that $t = 10$ at present.

3. Degrees of forecastability.

   Which of the following can be forecast perfectly? Which can not be forecast at all? Which are somewhere in between? Explain your answers, and be careful!

   a. The direction of change tomorrow in a country's stock market;

b. The eventual lifetime sales of a newly-introduced automobile model;

c. The outcome of a coin flip;

d. The date of the next full moon;

e. The outcome of a (fair) lottery.

4. Forecasting in daily life.

   We all forecast, all the time, implicitly if not explicitly.

   a. Sketch in detail three forecasts that you make routinely, and probably informally, in your daily life. What makes you believe that the things your forecast are in fact forecastable? What does that even *mean*? What factors might introduce error into your forecasts?

   b. What decisions are aided by your three forecasts? How might the degree of predictability of the forecast object affect your decisions?

   c. For each of your forecasts, what is the value to you of a "good" as opposed to a "bad" forecast?

   d. How might you measure the "goodness" of your three forecasts?

5. Forecasting in business, finance, economics, and government.

   What sorts of forecasts would be useful in the following decision-making situations? Why? What sorts of data might you need to produce such forecasts?

   a. Shop-All-The-Time Network (SATTN) needs to schedule operators to receive incoming calls. The volume of calls varies depending on the time of day, the quality of the TV advertisement, and the price of the good being sold. SATTN must schedule staff to minimize the loss of sales (too few operators leads to long hold times, and people hang up if put on hold) while also considering the loss associated with hiring excess employees.

b. You're a U.S. investor holding a portfolio of Japanese, British, French and German stocks and government bonds. You're considering broadening your portfolio to include corporate stocks of Tambia, a developing economy with a risky emerging stock market. You're only willing to do so if the Tambian stocks produce higher portfolio returns sufficient to compensate you for the higher risk. There are rumors of an impending military coup, in which case your Tambian stocks would likely become worthless. There is also a chance of a major Tambian currency depreciation, in which case the dollar value of your Tambian stock returns would be greatly reduced.

c. You are an executive with Grainworld, a huge corporate farming conglomerate with grain sales both domestically and abroad. You have no control over the price of your grain, which is determined in the competitive market, but you must decide what to plant and how much, over the next two years. You are paid in foreign currency for all grain sold abroad, which you subsequently convert to dollars. Until now the government has bought all unsold grain to keep the price you receive stable, but the agricultural lobby is weakening, and you are concerned that the government subsidy may be reduced or eliminated in the next decade. Meanwhile, the price of fertilizer has risen because the government has restricted production of ammonium nitrate, a key ingredient in both fertilizer and terrorist bombs.

d. You run BUCO, a British utility supplying electricity to the London metropolitan area. You need to decide how much capacity to have on line, and two conflicting goals must be resolved in order to make an appropriate decision. You obviously want to have enough capacity to meet average demand, but that's not enough, because demand is uneven throughout the year. In particular, demand skyrockets during summer heat waves – which occur randomly – as more and more peo-

ple run their air conditioners constantly. If you don't have sufficient capacity to meet peak demand, you get bad press. On the other hand, if you have a large amount of excess capacity over most of the year, you also get bad press.

6. Finding and using data on the web.

   Search the web for information on U.S. retail sales, U.K. stock prices, German GDP, and Japanese federal government expenditures. Using graphical methods, compare and contrast the movements of each series.

7. Software differences and bugs: caveat emptor.

   Be warned: no software is perfect. In fact, all software is highly imperfect! The results obtained when modeling or forecasting in different software environments may differ – sometimes a little and sometimes a lot – for a variety of reasons. The details of implementation may differ across packages, for example, and small differences in details can sometimes produce large differences in results. Hence, it is important that you understand *precisely* what your software is doing (insofar as possible, as some software documentation is more complete than others). And of course, quite apart from correctly-implemented differences in details, deficient implementations can and do occur: there is no such thing as bug-free software.

8. Forecasting vs. prediction.

   We will use the terms *prediction* and *forecasting* interchangeably, using either term in all environments (time-series environments), cross-section environments, etc.)