# Elements of Forecasting

*in Business, Finance, Economics and Government*

Francis X. Diebold
Department of Economics
University of Pennsylvania

"Solutions Manual"

# Preface

This is quite a nonstandard "Solutions Manual," but I use the term for lack of something more descriptively accurate. Many of the Problems and Complements don't ask questions, so they certainly don't have solutions; instead, they simply introduce concepts and ideas that, for one reason or another, didn't make it into the main text. Moreover, even for those Problems and Complements that *do* ask questions, the vast majority don't have explicit or unique solutions. Hence the "solutions manual" offers remarks, suggestions, hints, and occasionally, solutions. Most of the Problems and Complements are followed by brief remarks marked with asterisks, and in the (relatively rare) cases where there was nothing to say, I said nothing.


F.X.D.

# Solutions

**Chapter 1 Problems and Complements**

1. (Forecasting in daily life:  we are all forecasting, all the time)

   a. Sketch in detail three forecasts that you make routinely, and probably informally, in your daily life.  What makes you believe that the forecast object is predictable?  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. How might you measure the "goodness" of your three forecasts?

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

* Remarks, suggestions, hints, solutions:  The idea behind all of these questions is to help the students realize that forecasts are of value only in so far as they help with decisions, so that forecasts and decisions are inextricably linked.

2. (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 people 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.

* Remarks, suggestions, hints, solutions: Each of the above scenarios is complex and realistic,

with no clear cut answer. Instead, the idea is to get students thinking about and discussing

relevant issues that run through the questions, such the forecast object, the forecast horizon, the

loss function and whether it might be asymmetric, the fact that some risks can be hedged and

hence need not contribute to forecast uncertainty, etc.

3. (The basic forecasting framework) True or false (explain your answers):

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

time series being forecast.

* Remarks, suggestions, hints, solutions: False - that is the beauty of the situation.

b. Ongoing improvements in forecasting methods will eventually enable perfect

prediction.

* Remarks, suggestions, hints, solutions: False - the systems forecast in the areas that concern us

are intrinsically stochastic and hence can never be perfectly forecast.

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

could be improved.

* Remarks, suggestions, hints, solutions:  False.  Indeed studying series of forecast errors can provide just such information.  The key to forecast evaluation is that good forecasts shouldn't have forecastable forecast errors, so if the errors can be forecast then something is wrong.

4.  (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;

* Remarks, suggestions, hints, solutions:  Some would say imperfectly, some would say not at all.

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

* Remarks, suggestions, hints, solutions:  Imperfectly.

    c.  The outcome of a coin flip;

* Remarks, suggestions, hints, solutions:  Not at all, in the sense of guessing correctly more than fifty percent of the time (assuming a fair coin).

    d.  The date of the next full moon;

* Remarks, suggestions, hints, solutions:  Perfectly.

    e.  The outcome of a (fair) lottery.

5.  (Data on the web)  A huge amount of data of all sorts are available on the web.  Frumkin (2004) and Baumohl (2005) provide useful and concise introductions to the construction, accuracy and interpretation of a variety of economic and financial indicators, many of which are available on the web.  Search the web for information on U.S. retail sales, U.K. stock prices, German GDP, and Japanese federal government expenditures.  (The Resources for Economists page is a fine place to start:  www.rfe.org)  Using graphical methods, compare and contrast the

movements of each series and speculate about the relationships that may be present.

* Remarks, suggestions, hints, solutions: The idea is simply to get students to be aware of what data interests them and whether its available on the web.

6. (Univariate and multivariate forecasting models) In this book we consider both "univariate" and "multivariate" forecasting models. In a univariate model, a single variable is modeled and forecast solely on the basis of its own past. Univariate approaches to forecasting may seem simplistic, and in some situations they are, but they are tremendously important and worth studying for at least two reasons. First, although they are simple, they are not necessarily simplistic, and a large amount of accumulated experience suggests that they often perform admirably. Second, it's necessary to understand univariate forecasting models before tackling more complicated multivariate models.

In a multivariate model, a variable (or each member of a set of variables) is modeled on the basis of its own past, as well as the past of other variables, thereby accounting for and exploiting cross-variable interactions. Multivariate models have the *potential* to produce forecast improvements relative to univariate models, because they exploit more information to produce forecasts.

    a. Determine which of the following are examples of univariate or multivariate forecasting:

        • Using a stock's price history to forecast its price over the next week;

        • Using a stock's price history and volatility history to forecast its price over the next week;

        • Using a stock's price history and volatility history to forecast its price and

volatility over the next week.

b.  Keeping in mind the distinction between univariate and multivariate models, consider a wine merchant seeking to forecast the price per case at which 1990 Chateau Latour, one of the greatest Bordeaux wines ever produced, will sell in the year 2015, at which time it will be fully mature.

• What sorts of univariate forecasting approaches can you imagine that might be relevant?

* Remarks, suggestions, hints, solutions:  Examine the prices from 1990 through the present and extrapolate in some "reasonable" way.  Get the students to try to define "reasonable."

• What sorts of multivariate forecasting approaches can you imagine that might be relevant?  What other variables might be used to predict the Latour price?

* Remarks, suggestions, hints, solutions:  You might also use information in the prices of other similar wines, macroeconomic conditions, etc.

• What are the comparative costs and benefits of the univariate and multivariate approaches to forecasting the Latour price?

* Remarks, suggestions, hints, solutions:  Multivariate approaches bring more information to bear on the forecasting problem, but at the cost of greater complexity.  Get the students to expand on this tradeoff.

• Would you adopt a univariate or multivariate approach to forecasting the Latour price?  Why?

* Remarks, suggestions, hints, solutions:  You decide!

## Chapter 2 Problems and Complements

1. (Interpreting distributions and densities)  The Sharpe Pencil Company has a strict quality control monitoring program.  As part of that program, it has determined that the distribution of the amount of graphite in each batch of one hundred pencil leads produced is continuous and uniform between one and two grams.  That is, $f(y) = 1$ for y in [1, 2], and zero otherwise, where y is the graphite content per batch of one hundred leads.

     a.  Is y a discrete or continuous random variable?

* Remarks, suggestions, hints, solutions: Continuous.

     b.  Is f(y) a probability distribution or a density?

* Remarks, suggestions, hints, solutions: Density.

     c.  What is the probability that y is between 1 and 2?  Between 1 and 1.3?  Exactly equal to 1.67?

* Remarks, suggestions, hints, solutions: 1.00, 0.30, 0.00.

     d.  For high-quality pencils, the desired graphite content per batch is 1.8 grams, with low variation across batches.  With that in mind, discuss the nature of the density f(y).

* Remarks, suggestions, hints, solutions: f(y) is unfortunately centered at 1.5, not 1.8.  Moreover, f(y) unfortunately shows rather high dispersion.

2. (Covariance and correlation)  Suppose that the annual revenues of world's two top oil producers have a covariance of 1,735,492.

     a.  Based on the covariance, the claim is made that the revenues are "very strongly positively related."  Evaluate the claim.

* Remarks, suggestions, hints, solutions: Can't tell - it depends on the units of measurement.  Are they dollars, billions of dollars, or what?

    b.  Suppose instead that, again based on the covariance, the claim is made that the revenues are "positively related."  Evaluate the claim.

* Remarks, suggestions, hints, solutions: True.

    c.  Suppose you learn that the revenues have a *correlation* of 0.93.  In light of that new information, re-evaluate the claims in parts a and b above.

* Remarks, suggestions, hints, solutions: Indeed the revenues are unambiguously "very strongly positively related."

3.  (Conditional expectations vs. linear projections)  It is important to note the distinction between a conditional mean and a linear projection.

    a.  The conditional mean is not necessarily a linear function of the conditioning variable(s). In the Gaussian case, the conditional mean is a linear function of the conditioning variables, so it coincides with the linear projection.  In non-Gaussian cases, however, linear projections are best viewed as approximations to generally non-linear conditional mean functions.

* Remarks, suggestions, hints, solutions: This is one of the amazing and very convenient properties of the normal distribution.

    b.  The U.S. Congressional Budget Office (CBO) is helping the president to set tax policy. In particular, the president has asked for advice on where to set the average tax rate to maximize the tax revenue collected per taxpayer.  For each of 23 countries the CBO has obtained data on the tax revenue collected per taxpayer and the

average tax rate.  Is tax revenue likely related to the tax rate?  Is the relationship

likely linear?  (Hint:  how much revenue would be collected at tax rates of zero or

one hundred percent?)  If not, is a linear regression nevertheless likely to produce a

good approximation to the true relationship?

* Remarks, suggestions, hints, solutions: The relationship is not likely linear.  Revenues would

initially rise with the tax rate, but eventually decline as the rate nears 100 percent and people

simply opt not to work, or to work but not report the income.  (This is the famous "Laffer

curve.")  It appears unlikely that a linear approximation would be accurate.

4.  (Conditional mean and variance)  Given the regression model,

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 z_t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} (0, \sigma^2),$$

find the mean and variance of $y_t$ conditional upon $x_t = x_t^*$ and $z_t = z_t^*$.  Does the conditional mean

adapt to the conditioning information?  Does the conditional variance adapt to the conditioning

information?

* Remarks, suggestions, hints, solutions:  The conditional mean is

$$y_t = \beta_0 + \beta_1 x_t^* + \beta_2 x_t^{*2} + \beta_3 z_t^* + \varepsilon_t.$$

The conditional variance is simply $\sigma^2$.

5.  (Scatter plots and regression lines)  Draw qualitative scatter plots and regression lines for each

of the following two-variable data sets, and state the $R^2$ in each case:

    a.  data set 1:  y and x have correlation 1

    b.  data set 2:  y and x have correlation -1

c. data set 3: y and x have correlation 0.

* Remarks, suggestions, hints, solutions: 1, 1, 0.

6. (Desired values of regression diagnostic statistics)  For each of the diagnostic statistics listed below, indicate whether, other things the same, "bigger is better," "smaller is better," or neither. Explain your reasoning.  (Hint:  Be careful, think before you answer, and be sure to qualify your answers as appropriate.)

a. Coefficient

* Remarks, suggestions, hints, solutions:  neither

b. Standard error

* Remarks, suggestions, hints, solutions:  smaller is better

c. t statistic

* Remarks, suggestions, hints, solutions:  bigger is better

d. Probability value of the t statistic

* Remarks, suggestions, hints, solutions:  smaller is better

e. R-squared

* Remarks, suggestions, hints, solutions:  bigger is better

f. Adjusted R-squared

* Remarks, suggestions, hints, solutions:  bigger is better

g. Standard error of the regression

* Remarks, suggestions, hints, solutions:  smaller is better

h. Sum of squared residuals

* Remarks, suggestions, hints, solutions:  smaller is better

        i.  Log likelihood

* Remarks, suggestions, hints, solutions:  bigger is better

        j.  Durbin-Watson statistic

* Remarks, suggestions, hints, solutions:  neither -- should be near 2

        k.  Mean of the dependent variable

* Remarks, suggestions, hints, solutions:  neither -- could be anything

        l.  Standard deviation of the dependent variable

* Remarks, suggestions, hints, solutions:  neither -- could be anything

        m.  Akaike information criterion

* Remarks, suggestions, hints, solutions:  smaller is better

        n.  Schwarz information criterion

* Remarks, suggestions, hints, solutions:  smaller is better

        o.  F-statistic

* Remarks, suggestions, hints, solutions:  bigger is better

        p.  Probability-value of the F-statistic

* Remarks, suggestions, hints, solutions:  smaller is better

* Additional remarks:  Many of the above answers need qualification.  For example, the fact that,

other things the same, a high $R^2$ is good in so far as it means that the regression has more

explanatory power, does not mean that forecasting models should be selected on the basis of

"high $R^2$."

7.  (Mechanics of fitting a linear regression)  On the book's web page you will find a second set of

data on y, x and z, similar to, but different from, the data that underlie the analysis performed in

this chapter. Using the new data, repeat the analysis and discuss your results.

* Remarks, suggestions, hints, solutions: In my opinion, it's crucially important that students do this exercise, to get comfortable with the computing environment sooner rather than later.

8. (Regression with and without a constant term) Consider Figure 2, in which we showed a scatterplot of y vs. x with a fitted regression line superimposed.

      a. In fitting that regression line, we included a constant term. How can you tell?

* Remarks, suggestions, hints, solutions: The fitted line does not pass through the origin.

      b. Suppose that we had not included a constant term. How would the figure look?

* Remarks, suggestions, hints, solutions: The fitted line would pass through the origin.

      c. We almost always include a constant term when estimating regressions. Why?

* Remarks, suggestions, hints, solutions: Except in very special circumstances, there is no reason to force lines through the origin.

      d. When, if ever, might you explicitly want to exclude the constant term?

* Remarks, suggestions, hints, solutions: If, for example, an economic "production function" were truly linear, then it should pass through the origin. (No inputs, no outputs.)

9. (Interpreting coefficients and variables) Let $y_t = \beta_0 + \beta_1 x_t + \beta_2 z_t + \varepsilon_t$, where $y_t$ is the number of hot dogs sold at an amusement park on a given day, $x_t$ is the number of admission tickets sold that day, $z_t$ is the daily maximum temperature, and $\varepsilon_t$ is a random error.

      a. State whether each of $y_t$, $x_t$, $z_t$, $\beta_0$, $\beta_1$ and $\beta_2$ is a coefficient or a variable.

* Remarks, suggestions, hints, solutions: variable, variable, variable, coefficient, coefficient, coefficient

      b. Determine the units of $\beta_0$, $\beta_1$, and $\beta_2$, and describe the physical meaning of each.

* Remarks, suggestions, hints, solutions:  Units are hot dogs.  The coefficients measure the

responsiveness (formally the partial derivative) of hot dog sales to the various variables.

       c.  What does the sign of a coefficient tell you about its corresponding variable affects the

            number of hot dogs sold?  What are your expectations for the signs of the various

            coefficients (negative, zero, positive or unsure)?

* Remarks, suggestions, hints, solutions: Sign tells whether the relationship is positive or inverse.

Sign on admissions is surely expected to be positive.  I don't have strong feelings about the sign

of the temperature coefficient; that is, I'm not sure whether people eat more or fewer hot dogs

when it's hot.  Maybe the coefficient is zero.

       d.  Is it sensible to entertain the possibility of a non-zero intercept (i.e., $\beta_0 \neq 0$)?  $\beta_0 > 0$?

            $\beta_0 < 0$?

* Remarks, suggestions, hints, solutions: Taken rigidly, it's probably not sensible to allow a non-

zero intercept.  (Presumably hot dog sales must be zero if admissions are zero.)  But more

generally, if we view this linear model a merely a linear approximation to a potentially non-linear

relationship, the intercept may well be non-zero (of either sign).

10.  (Nonlinear least squares) The least squares estimator discussed in this chapter is often called

"ordinary" least squares.  The adjective "ordinary" distinguishes the ordinary least squares

estimator from fancier estimators, such as the nonlinear least squares estimator.  When we

estimate by nonlinear least squares, we use a computer to find the minimum of the sum of squared

residual function directly, using numerical methods.  For the simple regression model discussed in

this chapter, ordinary and nonlinear least squares produce the same result, and ordinary least

squares is simpler to implement, so we prefer ordinary least squares.  As we will see, however,

some intrinsically nonlinear forecasting models can't be estimated using ordinary least squares but can be estimated using nonlinear least squares.  We use nonlinear least squares in such cases.

For each of the models below, determine whether ordinary least squares may be used for estimation (perhaps after transforming the data).

a. $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

b. $y_t = \beta_0 e^{\beta_1 x_t} \varepsilon_t$

c. $y_t = \beta_0 + e^{\beta_1 x_t} + \varepsilon_t$.

* Remarks, suggestions, hints, solutions: OLS is fine for a, fine for b after taking logs, and no good for c.

11.  (Regression semantics)  Regression analysis is so important, and used so often by so many people, that a variety of associated terms have evolved over the years, all of which are the same for our purposes.  You may encounter them in your reading, so it's important to be aware of them.  Some examples:

a.  Ordinary least squares, least squares, OLS, LS.

b.  y, left-hand-side variable, regressand, dependent variable, endogenous variable

c.  x's, right-hand-side variables, regressors, independent variables, exogenous variables, predictors

d.  probability value, prob-value, p-value, marginal significance level

e.  Schwarz criterion, Schwarz information criterion, SIC, Bayes information criterion, BIC

* Remarks, suggestions, hints, solutions:  Students are often confused by statistical/econometric jargon, particularly the many redundant or nearly-redundant terms.  This complement presents

some commonly-used synonyms, which many students don't initially recognize as such.

## Chapter 3 Problems and Complements

1. (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.

* Remarks, suggestions, hints, solutions: t+5 is February 1961, and so on.  $y_{t+5,t} = y_{15,10} = 15$, and

so on.

2. (Properties of loss functions)  State whether the following potential loss functions meet the

criteria introduced in the text, and if so, whether they are symmetric or asymmetric:

    a.  $L(e) = e^2 + e$

    b.  $L(e) = e^4 + 2e^2$

    c.  $L(e) = 3e^2 + 1$

    d.  $L(e) = \begin{cases} \sqrt{e} & \text{if } e>0 \\ |e| & \text{if } e\leq 0. \end{cases}$

* Remarks, suggestions, hints, solutions:  d satisfies the criteria, immediately by inspection.

(L(0)=0, and it is monotonically increasing on each side of the origin.) As for the others, graph

them and see for yourself!

3. (Relationships among point, interval and density forecasts) For each of the following density

forecasts, how might you infer "good" point and ninety percent interval forecasts? Conversely, if

you started with your point and interval forecasts, could you infer "good" density forecasts? Be

sure to defend your definition of "good."

    a. Future y is distributed as N(10,2).

$$P(y) = \begin{cases} \dfrac{y-5}{25} & \text{if } 5<y<10 \\[2mm] -\dfrac{y-15}{25} & \text{if } 10<y<15 \\[2mm] 0 & \text{otherwise.} \end{cases}$$

    b. $P(y) =$

* Remarks, suggestions, hints, solutions: For part a, use E(y)=10 as the point forecast and use

$[p_{.05}, p_{.95}]$ as the interval forecast, where $p_{.05}$ and $p_{.95}$ are the fifth and ninety-fifth percentiles of

a N(10, 2) random variable.

4. (Forecasting at short through long horizons) Consider the claim, "The distant future is harder

to forecast than the near future." Is it sometimes true? Usually true? Always true? Why or why

not? Discuss in detail. Be sure to define "harder."

* Remarks, suggestions, hints, solutions: Usually but not always.

5. (Forecasting as an ongoing process in organizations) We could add another very important

item to this chapter's list of considerations basic to successful forecasting -- forecasting in

organizations is an ongoing process of building, using, evaluating, and improving forecasting

models. Provide a concrete example of a forecasting model used in business, finance, economics or government, and discuss ways in which each of the following questions might be resolved prior to, during, or after its construction.

    a. Are the data "dirty"? For example, are there "ragged edges"? That is, do the starting and ending dates of relevant series differ? Are there missing observations? Are there aberrant observations, called outliers, perhaps due to measurement error? Are the data stored in a format that inhibits computerized analysis?

* Remarks, suggestions, hints, solutions: The idea is to get students to think hard about the myriad of problems one encounters when analyzing real data. The question introduces them to a few such problems; in class discussion the students should be able to think of more.

    b. Has software been written for importing the data in an ongoing forecasting operation?

* Remarks, suggestions, hints, solutions: Try to impress upon the students the fact that reading and manipulating the data is a crucial part of applied forecasting.

    c. Who will build and maintain the model?

* Remarks, suggestions, hints, solutions: All too often, too little attention is given to issues like this.

    d. Are sufficient resources available (time, money, staff) to facilitate model building, use, evaluation, and improvement on a routine and ongoing basis?

* Remarks, suggestions, hints, solutions: Ditto.

    e. How much time remains before the first forecast must be produced?

* Remarks, suggestions, hints, solutions: The model-building time can differ drastically across government and private projects. For example, more than a year may be allocated to a model-

building exercise at the Federal Reserve, whereas just a few months may be allocated at a wall

street investment bank.

f. How many series must be forecast, and how often must ongoing forecasts be produced?

* Remarks, suggestions, hints, solutions: The key is to emphasize that these sorts of questions

impact the choice of procedure, so they should be asked explicitly and early.

g. What level of data aggregation or disaggregation is desirable?

* Remarks, suggestions, hints, solutions: If disaggregated detail is of intrinsic interest, then

obviously a disaggregated analysis will be required. If, on the other hand, only the aggregate is of

interest, then the question arises as to whether one should forecast the aggregate directly, or

model its components and add together their forecasts. It can be shown that there is no one

answer; instead, one simply has to try it both ways and see which works better.

h. To whom does the forecaster or forecasting group report, and how will the forecasts

be communicated?

* Remarks, suggestions, hints, solutions: Communicating forecasts to higher management is a

key and difficult issue. Try to guide a discussion with the students on what formats they think

would work, and in what sorts of environments.

i. How might you conduct a "forecasting audit"?

* Remarks, suggestions, hints, solutions: Again, this sort of open-ended, but nevertheless

important, issue makes for good class discussion.

6. (Assessing forecasting situations) For each of the following scenarios, discuss the decision

environment, the nature of the object to be forecast, the forecast type, the forecast horizon, the

loss function, the information set, and what sorts of simple or complex forecasting approaches

you might entertain.

    a. You work for Airborne Analytics, a highly specialized mutual fund investing exclusively in airline stocks. The stocks held by the fund are chosen based on your recommendations. You learn that a newly rich oil-producing country has requested bids on a huge contract to deliver thirty state-of-the-art fighter planes, but that only two companies submitted bids. The stock of the successful bidder is likely to rise.

    b. You work for the Office of Management and Budget in Washington DC and must forecast tax revenues for the upcoming fiscal year. You work for a president who wants to maintain funding for his pilot social programs, and high revenue forecasts ensure that the programs keep their funding. However, if the forecast is too high, and the president runs a large deficit at the end of the year, he will be seen as fiscally irresponsible, which will lessen his probability of reelection. Furthermore, your forecast will be scrutinized by the more conservative members of Congress; if they find fault with your procedures, they might have fiscal grounds to undermine the President's planned budget.

    c. You work for D&D, a major Los Angeles advertising firm, and you must create an ad for a client's product. The ad must be targeted toward teenagers, because they constitute the primary market for the product. You must (somehow) find out what kids currently think is "cool," incorporate that information into your ad, and make your client's product attractive to the new generation. If your hunch is right, your firm basks in glory, and you can expect multiple future clients from this one

advertisement. If you miss, however, and the kids don't respond to the ad, then your client's sales fall and the client may reduce or even close its account with you.

* Remarks, suggestions, hints, solutions: Again, these questions are realistic and difficult, and they don't have tidy or unique answers. Use them in class discussion to get the students to appreciate the complexity of the forecasting problem.

**Chapter 4 Problems and Complements**

1. (Outliers)  Recall the lower-left panel of the multiple comparison plot of the Anscombe data (Figure 1), which made clear that dataset number three contained a severely anomalous observation.  We call such data points "outliers."

 a. Outliers require special attention because they can have substantial influence on the fitted regression line.  Regression parameter estimates obtained by least squares are particularly susceptible to such distortions.  Why?

* Remarks, suggestions, hints, solutions:  The least squares estimates are obtained by minimizing the sum of *squared* errors.  Large errors (of either sign) often turn into *huge* errors when squared, so least squares goes out of its way to avoid such large errors.

 b. Outliers can arise for a number of reasons.  Perhaps the outlier is simply a mistake due to a clerical recording error, in which case you'd want to replace the incorrect data with the correct data.  We'll call such outliers measurement outliers, because they simply reflect measurement errors.  If a particular value of a recorded series is plagued by a measurement outlier, there's no reason why observations at other times should necessarily be affected.  But they *might* be affected.  Why?

* Remarks, suggestions, hints, solutions:  Measurement errors could be correlated over time.  If, for example, a supermarket scanner is malfunctioning today, it may be likely that it will also malfunction tomorrow, other thinks the same.

 c. Alternatively, outliers in time series may be associated with large unanticipated shocks, the effects of which may linger.  If, for example, an adverse shock hits the U.S.

economy this quarter (e.g., the price of oil on the world market triples) and the U.S. plunges into a severe depression, then it's likely that the depression will persist for some time. Such outliers are called innovation outliers, because they're driven by shocks, or "innovations," whose effects naturally last more than one period due to the dynamics operative in business, economic, and financial series.

d. How to identify and treat outliers is a time-honored problem in data analysis, and there's no easy answer. What factors would you, as a forecaster, examine when deciding what to do with an outlier?

* Remarks, suggestions, hints, solutions: Try to determine whether the outlier is due to a data recording error. If so, the correct data should be obtained if possible. Alternatively, the bad data could be discarded, but in time series environments, doing so creates complications of its own. Robust estimators could also be tried. If the outlier is not due to a recording error or some similar problem, then there may be little reason to discard it; in fact, retaining it may greatly increase the efficiency of estimated parameters, for which variation in the right-hand-side variables is crucial.

2. (Simple vs. partial correlation) The set of pairwise scatterplots that comprises a multiway scatterplot provides useful information about the joint distribution of the N variables, but it's incomplete information and should be interpreted with care. A pairwise scatterplot summarizes information regarding the simple correlation between, say, x and y. But x and y may appear highly related in a pairwise scatterplot even if they are in fact unrelated, if each depends on a third variable, say z. The crux of the problem is that there's no way in a pairwise scatterplot to examine the correlation between x and y *controlling* for z, which we call partial correlation.

When interpreting a scatterplot matrix, keep in mind that the pairwise scatterplots provide

information only on simple correlation.

* Remarks, suggestions, hints, solutions:  Understanding the difference between simple and partial

correlation helps with understanding the fact that correlation does not imply causation, which

should be emphasized.

3.  (Graphical regression diagnostic I:  time series plot of $y_t$, $\hat{y}_t$, and $e_t$)  After estimating a

forecasting model, we often make use of graphical techniques to provide important diagnostic

information regarding the adequacy of the model.  Often the graphical techniques involve the

residuals from the model.  Throughout, let the regression model be

$$y_t = \sum_{i=1}^{k} \beta_i x_{it} + \varepsilon_t,$$

and let the fitted values be

$$\hat{y}_t = \sum_{i=1}^{k} \hat{\beta}_i x_{it}.$$

The difference between the actual and fitted values is the residual,

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

a.  Superimposed time series plots of $y_t$ and $\hat{y}_t$ help us to assess the overall fit of a

forecasting model and to assess variations in its performance at different times

(e.g., performance in tracking peaks vs. troughs in the business cycle).

* Remarks, suggestions, hints, solutions:  We will use such plots throughout the book, so it makes sense to be sure students are comfortable with them from the outset.

      b.  A time series plot of $e_t$ (a so-called residual plot) helps to reveal patterns in the

residuals.  Most importantly, it helps us assess whether the residuals are correlated over time, that is, whether the residuals are serially correlated, as well as whether there are any anomalous residuals.  Note that even though there might be many right-hand side variables in this regression model, the actual values of y, the fitted values of y, and the residuals are simple univariate series which can be plotted easily.  We'll make use of such plots throughout this book.

* Remarks, suggestions, hints, solutions:  Ditto.  Students should appreciate from the outset that inspection of residuals is a crucial part of any forecast model building exercise.

4.  (Graphical regression diagnostic II:  time series plot of $e_t^2$ or $|e_t|$ )  Plots of $e_t^2$ or $|e_t|$ reveal

patterns (most notably serial correlation) in the *squared* or *absolute* residuals, which correspond to non-constant volatility, or heteroskedasticity, in the levels of the residuals.  As with the standard residual plot, the squared or absolute residual plot is always a simple univariate plot, even when there are many right-hand side variables.  Such plots feature prominently, for example, in tracking and forecasting time-varying volatility.

* Remarks, suggestions, hints, solutions:  We make use of such plots in problem 6 below.

5.  (Graphical regression diagnostic III:  scatterplot of $e_t$ vs. $x_t$ )  This plot helps us assess

whether the relationship between y and the set of x's is truly linear, as assumed in linear regression analysis.  If not, the linear regression residuals will depend on x.  In the case where

there is only one right-hand side variable, as above, we can simply make a scatterplot of $e_t$ vs. $x_t$.

When there is more than one right-hand side variable, we can make separate plots for each, although the procedure loses some of its simplicity and transparency.

* Remarks, suggestions, hints, solutions: I emphasize repeatedly to the students that if forecast errors are forecastable, then the forecast can be improved. The suggested plot is one way to help assess whether the forecast errors are likely to be forecastable, on the basis of in-sample residuals. If e appears to be a function of x, then something is probably wrong.

6. (Graphical analysis of foreign exchange rate data) Magyar Select, a marketing firm representing a group of Hungarian wineries, is considering entering into a contract to sell 8,000 cases of premium Hungarian dessert wine to AMI Imports, a worldwide distributor based in New York and London. The contract must be signed now, but payment and delivery is 90 days hence. Payment is to be in U.S. Dollars; Magyar is therefore concerned about U.S. Dollar / Hungarian Forint (USD/HUF) exchange rate volatility over the next 90 days. Magyar has hired you to analyze and forecast the exchange rate, on which it has collected data for the last 620 days. Naturally, you suggest that Magyar begin with a graphical examination of the data. (The USD/HUF exchange rate data are on the book's web page.)

      a. Why might we be interested in examining data on the log rather than the level of the

            USD/HUF exchange rate?

* Remarks, suggestions, hints, solutions: We often work in natural logs, which have the convenient property that the change in the log is approximately the percent change, expressed as a decimal.

      b. Take logs and produce a time series plot of the log of the USD/HUF exchange rate.

Discuss.

* Remarks, suggestions, hints, solutions: The data wander up and down with a great deal of persistence, as is typical for asset prices.

      c. Produce a scatterplot of the log of the USD/HUF exchange rate against the lagged log of the USD/HUF exchange rate. Discuss.

* Remarks, suggestions, hints, solutions: The point cloud is centered on the 45° line, suggesting that the current exchange rate equals the lagged exchange rate, plus a zero-mean error.

      d. Produce a time series plot of the change in the log USD/HUF exchange rate, and also produce a histogram, normality test, and other descriptive statistics. Discuss. (For small changes, the change in the logarithm is approximately equal to the percent change, expressed as a decimal.) Do the log exchange rate changes appear normally distributed? If not, what is the nature of the deviation from normality? Why do you think we computed the histogram, etc., for the differenced log data, rather than for the original series?

* Remarks, suggestions, hints, solutions: The log exchange rate changes look like random noise, in sharp contrast to the level of the exchange rate. The noise is not unconditionally Gaussian, however; the log exchange rate changes are fat-tailed relative to the normal. We analyzed the differenced log data rather than for the original series for a number of reasons. First, the differenced log data is approximately the one-period asset return, a concept of intrinsic interest in finance. Second, the exchange rate itself is so persistent that applying standard statistical procedures directly to it might result in estimates with poor or unconventional properties; moving to differenced log data eliminates that problem.

e. Produce a time series plot of the *square* of the change in the log USD/HUF exchange rate. Discuss and compare to the earlier series of log changes. What do you conclude about the volatility of the exchange rate, as proxied by the squared log changes?

* Remarks, suggestions, hints, solutions: The square of the change in the log USD/HUF exchange rate appears persistent, indicating serial correlation in volatility. That is, large changes tend to be followed by large changes, and small by small, regardless of sign.

7. (Common scales) Redo the multiple comparison of the Anscombe data in Figure 1 using common scales. Do you prefer the original or your newly-created graphic? Why or why not?

* Remarks, suggestions, hints, solutions: The use of common scales facilitates comparison and hence results in a superior graphic.

8. (Graphing real GDP, continued)

a. Consider the final plot at which we arrived when graphing four components of U.S. real GDP. What do you like about the plot? What do you dislike about the plot? How could you make it still better? Do it!

* Remarks, suggestions, hints, solutions: Decide for yourself!

b. In order to help sharpen your eye (or so I claim), some of the graphics in this book fail to adhere strictly to the elements of graphical style that we emphasized. Pick and critique three graphs from anywhere in the book (apart from this chapter), and produce improved versions.

* Remarks, suggestions, hints, solutions: There is plenty to choose from!

9. (Color)

a.  Color can aid graphics both in showing the data and in appealing to the viewer.  How?

* Remarks, suggestions, hints, solutions:  When plotting multiple time series, for example,

different series can be plotted in different colors, resulting in a graphic that is often much easier to

digest than using dash for one series, dot for another, etc.

b.  Color can also confuse.  How?

* Remarks, suggestions, hints, solutions:  One example, too many nearby members of the color

palette used together can be hard to decode.  Another example:  Attention may be drawn to those

series for which "hot" colors are used, which may distort interpretation if care is not taken.

c.  Keeping in mind the principles of graphical style, formulate as many guidelines for

color graphics as you can.

* Remarks, suggestions, hints, solutions:  For example, avoid color chartjunk -- glaring, clashing

colors that repel the viewer.

10.  (Regression, regression diagnostics, and regression graphics in action)  You're a new

financial analyst at a major investment house, tracking and forecasting earnings of the health care

industry.  At the end of each quarter, you forecast industry earnings for the next quarter.

Experience has revealed that your clients care about your forecast accuracy -- that is, they want

small errors -- but that they are not particularly concerned with the sign of your error.  (Your

clients use your forecast to help allocate their portfolios, and if your forecast is way off, they lose

money, regardless of whether you're too optimistic or too pessimistic.)  Your immediate

predecessor has bequeathed to you a forecasting model in which current earnings ($y_t$) are

explained by one variable lagged by one quarter ($x_{t-1}$).  (Both are on the book's web page.)

a.  Suggest and defend some candidate "x" variables?  Why might lagged x, rather than

current x, be included in the model?

b. Graph $y_t$ vs $x_{t-1}$ and discuss.

c. Regress $y_t$ on $x_{t-1}$ and discuss (including related regression diagnostics that you deem relevant).

d. Assess the entire situation in light of the "six considerations basic to successful forecasting" emphasized in Chapter 3: the decision environment and loss function, the forecast object, the forecast statement, the forecast horizon, the information set, and the parsimony principle.

e. Consider as many variations as you deem relevant on the general theme. At a minimum, you will want to consider the following:

-- Does it appear necessary to include an intercept in the regression?

-- Does the functional form appear adequate? Might the relationship be nonlinear?

-- Do the regression residuals seem random, and in particular, do they appear serially correlated or heteroskedastic?

-- Are there any outliers? If so, does the estimated model appear robust to their inclusion/exclusion?

-- Do the regression disturbances appear normally distributed?

-- How might you assess whether the estimated model is structurally stable?

* Remarks, suggestions, hints, solutions: It is necessary that x be lagged for the model to be useful for 1-step-ahead forecasting.

## Chapter 5 Problems and Complements

1.  (Calculating forecasts from trend models) You work for the International Monetary Fund in Washington DC, monitoring Singapore's real consumption expenditures.  Using a sample of real consumption data (measured in billions of 2005 Singapore dollars), $y_t$, t = 1990:Q1, ..., 2006:Q4, you estimate the linear consumption trend model, $y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$, obtaining the estimates $\hat{\beta}_0 = 0.51$, $\hat{\beta}_1 = 2.30$, and $\hat{\sigma}^2 = 16$.  Based upon your estimated trend model, construct feasible point, interval and density forecasts for 2010:Q1.

2.  (Identifying and testing trend models) In 1965, Intel co-founder Gordon Moore predicted that the number of transistors that one could place on a square-inch integrated circuit would double every twelve months.

    a.  What sort of trend is this?

    b.  Given a monthly series containing the number of transistors per square inch for the latest integrated circuit, how would you test Moore's prediction?  How would you test the currently accepted form of "Moore's Law," namely that the number of transistors actually doubles every eighteen months?

* Remarks, suggestions, hints, solutions:  The trend is increasing at an increasing rate.  One could test Moore's law by estimating the model $y_t = \beta y_{t-1} + \varepsilon_t$ and doing a t-test for $\beta = 2$.

3.  (Understanding model selection criteria) You are tracking and forecasting the earnings of a new company developing and applying proprietary nano-technology.  The earnings are trending upward.  You fit linear, quadratic, and exponential trend models, yielding sums of squared residuals of 4352, 2791, and 2749, respectively.  Which trend model would you select, and why?

* Remarks, suggestions, hints, solutions: Assuming that AIC and SIC are used for model selection, exponential trend must be best, because it has the smallest sum of squared residuals, and no other model has fewer parameters.

4. (Mechanics of trend estimation and forecasting)  Obtain from the web an upward-trending monthly series that interests you.  Choose your series such that it spans at least ten years, and such that it ends at the end of a year (i.e., in December).

    a.  What is the series and why does it interest you?  Produce a time series plot of it.

        Discuss.

* Remarks, suggestions, hints, solutions:  Hopefully the plot will reveal a bit about the shape of the trend, as well as the nature of deviations from trend.

    b.  Fit linear, quadratic and exponential trend models to your series.  Discuss the

        associated diagnostic statistics and residual plots.

* Remarks, suggestions, hints, solutions:  Note that if the residuals appear serially correlated, as indicated for example by the Durbin-Watson statistic, then the standard errors are nor necessarily trustworthy, so the results should be interpreted with care.

    c.  Select a trend model using the AIC and using the SIC.  Do the selected models agree?

        If not, which do you prefer?

* Remarks, suggestions, hints, solutions: I would likely prefer the model selected by the SIC, although I would want to dig deeper into the cause of any divergence.

    d.  Use your preferred model to forecast each of the twelve months of the next year.

        Discuss.

    e.  The *residuals* from your fitted model are effectively a *detrended* version of your

original series.  Why?   Plot them and discuss.

* Remarks, suggestions, hints, solutions:  Seasonal and/or cyclical effects (the topics of Chapters 5-9) may be evident.

5.  (Properties of polynomial trends)  Consider a sixth-order deterministic polynomial trend:

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2 + ... + \beta_6 TIME_t^6.$$

a.  How many local maxima or minima may such a trend display?

* Remarks, suggestions, hints, solutions:  A polynomial of degree p can have at most p-1 local optima.  Here p=6, so the answer is 5.

b.  Plot the trend for various values of the parameters to reveal some of the different

possible trend shapes.

* Remarks, suggestions, hints, solutions:  Students will readily see that a huge variety of shapes can emerge, depending on the particular parameter configuration.

c.  Is this an attractive trend model in general?  Why or why not?

* Remarks, suggestions, hints, solutions:  No.  Trends should be smooth; a polynomial of degree six can wiggle too much.

d.  Fit the sixth-order polynomial trend model to the NYSE volume series.  How does it

perform in that particular case?

* Remarks, suggestions, hints, solutions:  The in-sample fit will look very good, although close scrutiny will probably reveal wiggles that would not ordinarily be ascribed to trend.  You can illustrate another source of difficulty with high-order polynomial trends by doing a long extrapolation, with disastrous results.

6. (Specialized nonlinear trends) The logistic trend is

$$T_t = \frac{1}{a + br^t},$$

with $0 < r < 1$.

      a. Display the trend shape for various a and b values. When might such a trend shape be

          useful?

\* Remarks, suggestions, hints, solutions: Qualitatively, the logistic trend is always s-shaped,

although its precise shape of course varies with the parameter values. Logistic trends are

commonly used to model the product life cycle. Sales often build gradually, then briskly, and

finally level off.

      b. Can you think of other specialized situations in which other specialized trend shapes

          might be useful? Produce mathematical formulas for the additional specialized

          trend shapes you suggest.

\* Remarks, suggestions, hints, solutions: One example is a linear trend with a break at a

particular time, say $T^*$. Prior to time $T^*$, the trend is a+bt, but at time $T^*$ and onward the trend is

a'+b't. Trend breaks can occur for many reasons, such as legal or regulatory changes.

7. (Moving average smoothing for trend estimation) The trend regression technique is one way

to estimate and forecast trend. Another way to estimate trend is by *smoothing* techniques, which

we briefly introduce here. We'll focus on three: two-sided moving averages, one-sided moving

averages, and one-sided weighted moving averages. Here we present them as ways to estimate

and examine the trend in a time series; later we'll see how they can actually be used to *forecast*

time series.

Denote the original data by $\{y_t\}_{t=1}^{T}$ and the smoothed data by $\{\bar{y}_t\}$. Then the two-sided

moving average is $\bar{y}_t = (2m+1)^{-1}\sum\limits_{i=-m}^{m} y_{t-i}$, the one-sided moving average is $\bar{y}_t = (m+1)^{-1}\sum\limits_{i=0}^{m} y_{t-i}$,

and the one-sided weighted moving average is $\bar{y}_t = \sum\limits_{i=0}^{m} w_i y_{t-i}$, where the $w_i$ are weights and m is

an integer chosen by the user. The "standard" one-sided moving average corresponds to a one-

sided weighted moving average with all weights equal to $(m+1)^{-1}$.

   a. For each of the smoothing techniques, discuss the role played by m. What happens as

      m gets very large? Very small? In what sense does m play a role similar to p, the

      order of a polynomial trend?

* Remarks, suggestions, hints, solutions: The larger is m, the more smoothing is done, and

conversely. Thus the choice of m governs the amount of smoothing, just as the choice of p

governs the smoothness of a polynomial trend.

   b. If the original data runs from time 1 to time T, over what range can smoothed values

      be produced using each of the three smoothing methods? What are the

      implications for "real-time" or "on-line" smoothing versus "ex post" or "off-line"

      smoothing?

* Remarks, suggestions, hints, solutions: The one-sided smoothers, because they require only

current and past data, are immediately useful for real-time smoothing. Two-sided smoothers, in

contrast, require m observations backward *and* forward, which renders them less useful for real-

time smoothing, because future observations are not known in real time. At any time t, two-sided smoothed values can be computed only through period t-m.

c. You've been hired as a consultant by ICSB, a major international bank, to advise them on trends in North American and European stock markets, and to help them allocate their capital. You have extracted from your database the recent history of EUROStar, an index of eleven major European stock markets. Smooth the EUROStar data using equally-weighted one-sided and two-sided moving averages, for a variety of m values, until you have found values of m that work well. What do we mean by "work well"? Must the chosen value of m be the same for the one- and two-sided smoothers? For your chosen m values, plot the two-sided smoothed series against the actual and plot the one-sided smoothed series against the actual. Do you notice any systematic difference in the relationship of the smoothed to the actual series depending on whether you do a two-sided or one-sided smooth? Explain.

* Remarks, suggestions, hints, solutions: By "work well," we mean a choice of m that delivers a smoothed series that conforms visually with our prior notion of how smooth (or rough) the smoothed series should be. Different values of m can, and typically will, be used for one-sided and two-sided smoothers. A series passed through a one-sided smoother will tend to lag the same series passed through a two-sided smoother with symmetric weights, by virtue of the fact that the one-sided smoother works only from current and past data, whereas the two-sided smoother invokes present observations to balance the past observations.

d. Moving average procedures can also be used to detrend a series -- we simply subtract

the estimated trend from the series.  Sometimes, but not usually, it's appropriate

and desirable to detrend a series before modeling and forecasting it.  Why might it

sometimes be appropriate?  Why is it not usually appropriate?

* Remarks, suggestions, hints, solutions:  Trend is typically responsible for a very large part of the

variation of series we want to forecast, in which case we wouldn't want to remove and discard it;

rather, we'd want to model and forecast it.

8.  (Bias corrections when forecasting from logarithmic models)

a.  In Chapter 3 we introduced squared error loss, $L(e)=e^2$.  A popular measure of

forecast accuracy is out-of-sample mean squared error,  $MSE=E(e^2)$.[1]  The more

accurate the forecast, the smaller is MSE.  Show that MSE is equal to the sum of

the variance of the error and the square of the mean error.

* Remarks, suggestions, hints, solutions:  Recall the elementary decomposition of the variance,

$$var(e) \; = \; E(e^2) \; - \; (E(e))^2 \; = \; MSE \; - \; (E(e))^2.$$

Thus,

$$MSE \; = \; var(e) \; + \; (E(e))^2.$$

b.  A forecast is *unbiased* if the mean forecast error is zero.  Why might unbiased forecasts

be desirable?  Are they *necessarily* desirable?

---

[1] The MSE introduced earlier in the context of model selection is the mean of the *in-sample* residuals, as opposed to out-of-sample prediction errors.  The distinction is crucial.

* Remarks, suggestions, hints, solutions: Think about the question in terms of MSE. Other things constant, a reduction in bias (mean error) is good, as it will clearly reduce MSE. However, other things may not be constant; in particular, an increase in bias may be more than offset by a large reduction in variance, resulting in a decrease in MSE.

    c. Suppose that $(\log y)_{t+h,t}$ is an unbiased forecast of $(\log y)_{t+h}$. Then $\exp((\log y)_{t+h,t})$ is

        a *biased* forecast of $y_{t+h}$. More generally, if $(f(y))_{t+h,t}$ is an unbiased forecast of

        $(f(y))_{t+h}$, then $f^{-1}((f(y))_{t+h,t})$ is a biased forecast of $y_{t+h}$, for the arbitrary nonlinear

        function f. Why? (Hint: Is the expected value of a nonlinear function of the

        random variable the same as the nonlinear function of the expected value?)

* Remarks, suggestions, hints, solutions: The expected value of a nonlinear function of a random variable is *not* the same as the nonlinear function of the expected value, a result usually called "Jensen's inequality." Thus unbiasedness is not preserved under nonlinear transformations.

    d. Various "corrections" for the bias in $\exp((\log y)_{t+h,t})$ have been proposed. In practice,

        however, bias corrections may increase the variance of the forecast error even if

        they succeed in reducing bias. Why? (Hint: In practice the corrections involve

        estimated parameters.)

* Remarks, suggestions, hints, solutions: Because the bias corrections involve estimated parameters, the very act of implementing them introduces a fresh source of error.

    e. In practice will bias corrections necessarily reduce the forecast MSE? Why or why

        not?

* Remarks, suggestions, hints, solutions: Not necessarily, because the harmful effects of variance

inflation may more than offset the beneficial effects of bias reduction.

9. (Model selection for long-horizon forecasting) Suppose that you want to forecast monthly

inventory of Lamborgini autos at an exclusive Manhattan dealership.

a. Using the true data-generating process is best for forecasting at any horizon.

Unfortunately, we never know the true data-generating process! All our models

are approximations to the true but unknown data-generating process, in which case

the best forecasting model may change with the horizon. Why?

* Remarks, suggestions, hints, solutions: The best approximation to the data for one use may be

very different from the best approximation for another use.

b. At what horizon are the forecasts generated by models selected by the AIC and SIC

likely to be most accurate? Why?

* Remarks, suggestions, hints, solutions: The AIC and SIC are designed to select models with

good 1-step-ahead forecasting performance, as can be seen from the fact that they are based on

the mean squared error of the residuals, which are the in-sample analogs of 1-step-ahead

prediction errors.

c. How might you proceed to select a 1-month-ahead forecasting model? 2-month-

ahead? 3-month-ahead? 4-month-ahead?

* Remarks, suggestions, hints, solutions: For 1-month-ahead, use AIC and SIC in the usual way.

For multi-step-ahead forecasting models, an appropriately-modified selection criterion, based on

analogs of multi-step-ahead forecast errors, might be used. Some of the bibliographic references

in the text (e.g., Findley) pursue the idea.

d. What are the implications of your answer for construction of an extrapolation forecast,

at horizons 1-month-ahead through 4-months-ahead?

* Remarks, suggestions, hints, solutions: The extrapolation forecast would be assembled from forecasts from different models, one for each horizon.

      e. In constructing our extrapolation forecasts for retail sales, we used the AIC and SIC to select one model, which we then used to forecast all horizons. Why do you think we didn't adopt a more sophisticated strategy?

* Remarks, suggestions, hints, solutions: Perhaps laziness, or perhaps a hunch that the cost of such an approach (in terms of extra complexity and tedium) might exceed the benefit. At any rate, for better or worse, it remains "standard" practice to use one model for all horizons.

10. (The variety of "information criteria" reported across software packages) Some authors, and software packages, examine and report the logarithms of the AIC and SIC,

$$\ln(\text{AIC}) = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \left(\frac{2k}{T}\right)$$

$$\ln(\text{SIC}) = \ln\left(\frac{\sum_{t=1}^{T} e_t^2}{T}\right) + \frac{k \ln(T)}{T}.$$

The practice is so common that log(AIC) and log(SIC) are often simply called the "AIC" and "SIC." AIC and SIC must be greater than zero, so log(AIC) and log(SIC) are always well-defined and can take on any real value. Other authors and packages use other variants, based for example on the value of the maximized likelihood or log likelihood function. Some software packages have even changed definitions of AIC and SIC across releases! The important insight,

however, is that although these variations will of course change the numerical values of AIC and

SIC produced by your computer, they will not change the *rankings* of models under the various

criteria. Consider, for example, selecting among three models. If $AIC_1 < AIC_2 < AIC_3$, then it must

be true as well that $\ln(AIC_1) < \ln(AIC_2) < \ln(AIC_3)$, so we would select model 1 regardless of the

"definition" of the information criterion used.

Chapter 6 Problems and Complements

1. (Log transformations in seasonal models)  Just as log transformations were useful in trend

models to allow for nonlinearity, so too are they useful in seasonal models, although for a

somewhat different purpose:  stabilization of variance.  Often log transformations stabilize

seasonal patterns whose variance is growing over time.  Explain and illustrate.

* Remarks, suggestions, hints, solutions:  The log is a "squashing" transformation (e.g., it

converts fast exponential growth into slower linear growth).  It is therefore sometimes useful for

converting growing variances into stable variances.

2. (Seasonal adjustment)  Just as we sometimes want to remove the trend from a series,

sometimes we want to seasonally adjust a series before modeling and forecasting it.  Seasonal

adjustment may be done with moving average methods analogous to those used for detrending in

Chapter 5, or with the dummy variable methods discussed in this chapter, or with sophisticated

hybrid methods like the X-11 procedure developed at the U.S. Census Bureau.

     a. Discuss in detail how you'd use dummy variable regression methods to seasonally

        adjust a series.  (Hint:  the seasonally adjusted series is closely related to the

        residual from the seasonal dummy variable regression.)

* Remarks, suggestions, hints, solutions:  The hint gives most of it away.  The only twist is that

the mean value of the series needs to be added back to the residual.

     b. Seasonally adjust the housing starts series using dummy variable regression.  Discuss

        the patterns present and absent from the seasonally adjusted series.

* Remarks, suggestions, hints, solutions:  Obviously the seasonal pattern should be largely absent

for the adjusted series.  It need not be totally absent, however, because the seasonality may be of a

form different from that associated with the deterministic dummy-variable seasonality model.

  c.  Search the Web (or the library) for information on the latest U.S. Census Bureau

    seasonal adjustment procedure, and report what you learned.

* Remarks, suggestions, hints, solutions:  Among other things, we've now moved to X-12.

3.  (Selecting forecasting models involving calendar effects)  You're sure that a series you want to

forecast is trending, and that a linear trend is adequate, but you're not sure whether seasonality is

important.  To be safe, you fit a forecasting model with both trend and seasonal dummies,

$$y_t = \beta_1 \text{TIME}_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t.$$

  a.  The hypothesis of no seasonality, in which case you could drop the seasonal dummies,

    corresponds to equal seasonal coefficients across seasons, which is a set of s-1

    linear restrictions:

$$\gamma_1 = \gamma_2, \ \gamma_2 = \gamma_3, \ \dots, \ \gamma_{s-1} = \gamma_s.$$

    How would you perform an F test of the hypothesis?  What assumptions are you

    implicitly making about the regression's disturbance term?

* Remarks, suggestions, hints, solutions:  This is a standard F test, but the students need to be

reminded that the tests's legitimacy requires that the regression disturbances be white noise,

which may well not hold in a regression on only trend and seasonals.  Otherwise, the F statistic

will not in general have the F distribution.

b. Alternatively, how would you use forecast model selection criteria to decide whether

or not to include the seasonal dummies?

* Remarks, suggestions, hints, solutions:  They can be applied in the usual way to help decide

whether to drop the seasonals.  (Don't forget to include an intercept when the seasonals are

dropped.)

c. What would you do in the event that the results of the  "hypothesis testing" and "model

selection" approaches disagree?

* Remarks, suggestions, hints, solutions:  Typically that wouldn't happen, but if it did, I would

probably go with the SIC results, other things the same.

d. How, if at all, would your answers change if instead of considering whether to include

seasonal dummies you were considering whether to include holiday dummies?

Trading day dummies?

* Remarks, suggestions, hints, solutions:  The answers would not change.

4. (Testing for seasonality) Using the housing starts data:

a. As in the chapter, construct and estimate a model with a full set of seasonal dummies.

b. Test the hypothesis of no seasonal variation.  Discuss your results.

c. Test for the equality of the coefficients on March and November and the coefficients on

all the months in between and construct a model that uses three dummy variables,

one for December, January, and February, one for March and November, and one

for the remaining months.

5. (Seasonal regressions with an intercept and s-1 seasonal dummies)  Reestimate the housing

starts model using an intercept and eleven seasonal dummies, rather than the full set of seasonal

-49-

dummies as in the text. Compare and contrast your results with those reported in the text. What is the interpretation of the intercept? What are the interpretations of the coefficients on the eleven included seasonal dummies? Does it matter which month's dummy you drop?

6. (Applied trend and seasonal modeling) Nile.com, a successful on-line bookseller, monitors and forecasts the number of "hits" per day to its web page. You have daily hits data for 1/1/98 through 9/28/98.

     a. Fit and assess the standard linear, quadratic, and log linear trend models.

* Remarks, suggestions, hints, solutions: An increasing trend is present. Linear trend looks fine.

     b. For a few contiguous days roughly in late April and early May, hits were much higher than usual during a big sale. Do you find evidence of a corresponding group of outliers in the residuals from your trend models? Do they influence your trend estimates much? How should you treat them?

* Remarks, suggestions, hints, solutions: The outliers are certainly there, but they don't influence the trend estimates much. One could drop the offending observations, but again, not much changes.

     c. Model and assess the significance of day-of-week effects in Nile.com web page hits.

* Remarks, suggestions, hints, solutions: There appear to be no such effects operative.

     d. Select a final model, consisting only of trend and seasonal components, to use for forecasting.

* Remarks, suggestions, hints, solutions: AIC, SIC select linear trend, no calendar effects.

     e. Use your model to forecast Nile.com hits through the end of 1998.

* Remarks, suggestions, hints, solutions: The forecast is for continued upward linear trend.

7. (Periodic models) We introduced the seasonal dummy model as a natural and simple method

for generalizing a simple "mean plus noise" model,

$$y_t = \mu + \varepsilon_t,$$

to allow the mean to vary with the seasons,

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t.$$

More generally, we can also allow the coefficients of richer models to vary with the seasons, as

for example when we move from the fixed-coefficient regression model,

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t,$$

to the model,

$$y_t = \left( \sum_{i=1}^{s} \gamma_i^0 D_{it} \right) + \left( \sum_{i=1}^{s} \gamma_i^1 D_{it} \right) x_t + \varepsilon_t.$$

This model, which permits not only a seasonally varying intercept but also a seasonally varying

slope, is an example of a "periodic regression model." The word "periodic" refers to the

coefficients, which vary regularly with a fixed seasonal periodicity.

8. (Interpreting dummy variables) You fit a purely seasonal model with a full set of standard

monthly dummy variables to a monthly series of employee hours worked. Discuss how the

estimated dummy variable coefficients $\hat{\gamma}_1$, $\hat{\gamma}_2$, ... would change if you changed the first dummy

variable $D_1 = (1,0,0,...)$ (with all the other dummy variables remaining the same) to:

    a. $D_1 = (2,0,0,...)$

b. $\mathbf{D_1} = (-10, 0, 0, \ldots)$

c. $\mathbf{D_1} = (1, 1, 0, \ldots)$.

* Remarks, suggestions, hints, solutions: a. $\hat{\gamma}_1$ would change to $\hat{\gamma}_1/2$ and other estimates would be unchanged, b. $\hat{\gamma}_1$ would change to $-\hat{\gamma}_1/10$ and other estimates would be unchanged, c. Hmmm...

9. (Constructing seasonal models) Describe how you would construct a purely seasonal model for the following monthly series. In particular, what dummy variable(s) would you use to capture the relevant effects?

    a. A sporting goods store finds that detrended monthly sales are roughly the same for each month in a given three-month season. For example, sales are similar in the winter months of January, February and March, in the spring months of April, May and June, and so on.

* Remarks, suggestions, hints, solutions: Four dummies, indicating the quarter.

    b. A campus bookstore finds that detrended sales are roughly the same for all first, all second, all third, and all fourth months of each trimester. For example, sales are similar in January, May, and September, the first months of the first, second, and third trimesters, respectively.

* Remarks, suggestions, hints, solutions: Four dummies, indicating the month of the trimester.

    c. A Christmas ornament store is only open in November and December, so sales are zero in all other months.

* Remarks, suggestions, hints, solutions: This is a rather degenerate situation, but one might use two dummies, one for November-December, and one for other (or three, one for November, one for December, and one for other), perhaps even imposing the constraint that the coefficient on the

"other" dummy is zero.

10. (Calendar effects) You run a large catering firm, specializing in Sunday brunches and weddings. You model the firm's monthly income as $y_t = \beta_0 + \delta_s S_t + \delta_w W_t + \varepsilon_t$, where y is monthly income, and S and W are calendar effect variables indicating the number of Sundays and weddings in a month.

    a. What are the units of $\beta_0$, $\delta_s$, and $\delta_w$?

* Remarks, suggestions, hints, solutions: Dollars.

    b. How could you estimate the average income the firm receives per wedding?

* Remarks, suggestions, hints, solutions: $\hat{\delta}_w$.

    c. Over the past thirty years, you have regularly increased your prices to keep pace with inflation. How would you modify the model to account for the effects of such increases?

* Remarks, suggestions, hints, solutions: Include a time trend.

## Chapter 7 Problems and Complements

1. (Lag operator expressions, I)  Rewrite the following expressions without using the lag

operator.

    a.  $(L^\tau)y_t = \varepsilon_t$

    b.  $y_t = \left( \dfrac{2 + 5L + .8L^2}{L - .6L^3} \right) \varepsilon_t$

    c.  $y_t = 2\left( 1 + \dfrac{L^3}{L} \right) \varepsilon_t.$

\* Remarks, suggestions, hints, solutions:  a. $y_{t-\tau} = \varepsilon_t$, b. $y_{t-1} - .6y_{t-3} = 2\varepsilon_t + 5\varepsilon_{t-1} + .8\varepsilon_{t-2}$,

    c. $y_t = 2\varepsilon_t + 2\varepsilon_{t-2}.$

2. (Lag operator expressions, II)  Rewrite the following expressions in lag operator form.

    a. $y_t + y_{t-1} + ... + y_{t-N} = \alpha + \varepsilon_t + \varepsilon_{t-1} + ... + \varepsilon_{t-N}$, where $\alpha$ is a constant

    b. $y_t = \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t.$

\* Remarks, suggestions, hints, solutions: a. $y_t \sum_{i=0}^{N} L^i = \alpha + \varepsilon_t \sum_{i=0}^{N} L^i$, $y_t = \varepsilon_t \sum_{i=0}^{2} L^i.$

3. (Autocorrelation functions of covariance stationary series)  While interviewing at a top

investment bank, your interviewer is impressed by the fact that you have taken a course on time

series forecasting.  She decides to test your knowledge of the autocovariance structure of

covariance stationary series and lists five autocovariance functions:

    a.  $\gamma(t,\tau) = \alpha$

    b.  $\gamma(t,\tau) = e^{-\alpha\tau}$

c. $\gamma(t,\tau) = \alpha\tau$

d. $\gamma(t,\tau) = \dfrac{\alpha}{\tau}$,

where $\alpha$ is a positive constant. Which autocovariance function(s) are consistent with covariance stationarity, and which are not? Why?

* Remarks, suggestions, hints, solutions: Autocovariance functions of covariance stationary processes must eventually decay, as with b and d. The exponential decay of b holds for covariance stationary short-memory (e.g., ARMA) processes, and the hyperbolic decay of d holds for covariance stationary long-memory processes.

4. (Autocorrelation vs. partial autocorrelation) Describe the difference between autocorrelations and partial autocorrelations. How can autocorrelations at certain displacements be positive while the partial autocorrelations at those same displacements are negative?

* Remarks, suggestions, hints, solutions: Correlation measures linear association between two variables, whereas partial autocorrelation measures linear association between two variables *controlling* for the effects of one or more additional variables. Hence the two types of correlation, although related, are nevertheless very different, and they may well be of different signs.

5. (Conditional and unconditional means) As head of sales of the leading technology and innovation magazine publisher TECCIT, your bonus is dependent on the firm's revenue. Revenue changes from season to season, as subscriptions and advertizing deals are entered or renewed. From your experience in the publishing business you know that the revenue in a season is a function of the number of magazines sold in the previous season and can be described as

$y_t = 1000 + .9x_{t-1} + \varepsilon_t$, with uncorrelated residuals $\varepsilon_t \sim N(0, 1000)$, where $y$ is revenue and $x$ is

number of magazines sold.

    a. What is the expected revenue for next season conditional upon total sales of 6,340 this

        season?

\* Remarks, suggestions, hints, solutions: $1000 + .9*6340$

    b. What is unconditionally expected revenue if unconditionally expected sales are 8500?

\* Remarks, suggestions, hints, solutions: $1000 + .9*8500$

    c. A rival publisher offers you a contract identical to your current contract (same base pay

        and bonus). Based upon a confidential interview, you know that the same revenue

        model with identical coefficients is appropriate for your rival. The rival has sold an

        average of 9000 magazines in previous seasons but only 5,650 this season. Will

        you accept the offer? Why or why not?

\* Remarks, suggestions, hints, solutions: Conditional upon sales this season, next year's expected revenue (and hence expected bonus) is higher at the old firm. However, unconditionally expected revenue is higher at the new firm, and that's what's relevant for the longer run. So the decision would presumably depend on the rate of time preference.

6. (White noise residuals) You work for a top five consulting firm and are asked to join a team evaluating a turnaround project at Stardust Cinemas. You are briefed that despite its bad financial condition, the CEO attempted to increase Stardust's market share by renovating every theater to include a bar, an arcade, and a restaurant. Your task on the team is to assess whether the renovations increased box office receipts. To do so, you spend a long night fitting a trend + seasonal model to samples of $T = 100$ observations on box office receipts for each of Stardust's theaters. You find that the residuals (e) from your models approximately follow

$e_t = 0.5 e_{t-1} + v_t$, where $v_t \overset{iid}{\sim} N(0,1)$. You forward your results to your project manager.

   a. You receive an email from your project manager indicating that your residuals do not

      look like white noise. Why? Why care?

* Remarks, suggestions, hints, solutions: Clearly the error at time t is correlated with the error at

time t-1, so the errors are serially correlated. This indicates that something has been neglected

from the model, such that if the model were to be used for forecasting, its errors could be

forecast, meaning that there is room for improvement.

   b. Assuming that the residuals do indeed follow $e_t = 0.5 e_{t-1} + v_t$, what is their

      autocorrelation function? Discuss.

* Remarks, suggestions, hints, solutions: $\rho(\tau) = .5^\tau$.

   c. What type of model might be useful for describing the historical path of box office

      income, and its likely future path in the absence of renovations? How would you

      use it to assess the efficacy of the renovation project?

* Remarks, suggestions, hints, solutions: A trend + cycle model with AR(1) errors, or almost

equivalently, a model with trend, cycle *and* a lagged dependent variable. The efficacy of the

renovation project could be assessed by comparing the forecasted revenue path (based on pre-

renovation historical data) to the actual revenue path subsequent to the renovation.

7. (Selecting an employment forecasting model with the AIC and SIC) Use the AIC and SIC to

assess the necessity and desirability of including trend and seasonal components in a forecasting

model for Canadian employment.

   a. Display the AIC and SIC for a variety of specifications of trend and seasonality. Which

      would you select using the AIC? SIC? Do the AIC and SIC select the same

model?  If not, which do you prefer?

* Remarks, suggestions, hints, solutions:  A variety of answers is possible, depending on the specific models fit, but the upshot is that trend and seasonality are not important parts of the dynamics of Canadian employment.

  b.  Discuss the estimation results and residual plot from your preferred model, and perform a correlogram analysis of the residuals.  Discuss, in particular, the patterns of the sample autocorrelations and partial autocorrelations, and their statistical significance.

* Remarks, suggestions, hints, solutions:  Because trend and seasonality don't contribute much to the variation in Canadian employment, the residuals from trend+seasonal regressions have properties very similar to the original series.

  c.  How, if at all, are your results different from those reported in the text?  Are the differences important?  Why or why not?

* Remarks, suggestions, hints, solutions:  Any differences are likely unimportant.

8.  (Simulating time series processes)  Many cutting-edge estimation and forecasting techniques involve simulation.  Moreover, simulation is often a good way to get a feel for a model and its behavior.  White noise can be simulated on a computer using random number generators, which are available in most statistics, econometrics and forecasting packages.

  a.  Simulate a Gaussian white noise realization of length 200.  Call the white noise $\varepsilon_t$. Compute the correlogram.  Discuss.

* Remarks, suggestions, hints, solutions:  The correlogram should be flat, with most sample autocorrelations inside the Bartlett bands.

b. Form the distributed lag $y_t = \varepsilon_t + .9\varepsilon_{t-1}$, $t = 2, 3, ..., 200$. Compute the sample

autocorrelations and partial autocorrelations. Discuss.

* Remarks, suggestions, hints, solutions: Neither the sample autocorrelation nor the sample

partial autocorrelation function is flat. For now, the precise patterns are unimportant; what's

important is that the distributed lag series is clearly *not* white noise. The illustrates the Slutsky-

Yule effect: distributed lags of white noise are serially correlated.

c. Let $y_1 = 1$ and $y_t = .9y_{t-1} + \varepsilon_t$, $t = 2, 3, ..., 200$. Compute the sample autocorrelations

and partial autocorrelations. Discuss.

* Remarks, suggestions, hints, solutions: Ditto.

9. (Sample autocorrelation functions for trending series) A tell-tale sign of the slowly-evolving

nonstationarity associated with trend is a sample autocorrelation function that damps extremely

slowly.

a. Find three trending series, compute their sample autocorrelation functions, and report

your results. Discuss.

* Remarks, suggestions, hints, solutions: The sample autocorrelation functions will damp very

slowly, if at all.

b. Fit appropriate trend models, obtain the model residuals, compute their sample

autocorrelation functions, and report your results. Discuss.

* Remarks, suggestions, hints, solutions: If the model now contains an appropriate trend, the

sample autocorrelations may damp more quickly. Note, however, that the data can have

nonstationarities other than deterministic trend, such as unit roots, in which case the residuals

from the trend regression will have sample autocorrelations that still fail to damp.

10. (Sample autocorrelation functions for seasonal series)  A tell-tale sign of seasonality is a sample autocorrelation function with sharp peaks at the seasonal displacements (4, 8, 12, etc. for quarterly data, 12, 24, 36, etc. for monthly data, and so on).

      a.  Find a series with both trend and seasonal variation.  Compute its sample

           autocorrelation function.  Discuss.

* Remarks, suggestions, hints, solutions: Presumably it fails to damp, or damps very slowly.

      b.  Detrend the series.  Discuss.

* Remarks, suggestions, hints, solutions:  The detrended series will (hopefully!) not display trend, but it will still be seasonal.

      c.  Compute the sample autocorrelation function of the detrended series.  Discuss.

* Remarks, suggestions, hints, solutions: Presumably it now damps and shows local peaks at the seasonal lag and its multiples.

      d.  Seasonally adjust the detrended series.  Discuss.

* Remarks, suggestions, hints, solutions: Visual appearance of seasonality should no longer be present.

      e.  Compute the sample autocorrelation function of the detrended, seasonally-adjusted

           series.  Discuss.

* Remarks, suggestions, hints, solutions:  Presumably it now damps and has no seasonal peaks.

11. (Volatility dynamics:  correlograms of squares)  In the Chapter 4 Exercises, Problems and Complements, we suggested that a time series plot of a squared residual, $e_t^2$, can reveal serial correlation in squared residuals, which corresponds to non-constant volatility, or

heteroskedasticity, in the levels of the residuals. Financial asset returns often display little

systematic variation, so instead of examining residuals from a model of returns, we often examine

returns directly. In what follows, we will continue to use the notation $e_t$, but you should interpret

$e_t$ it as an observed asset return.

      a. Find a high frequency (e.g., daily) financial asset return series, $e_t$, plot it, and discuss

           your results.

* Remarks, suggestions, hints, solutions: A stock return or the percent change in an exchange

rate would be a good choice. It will probably look like random noise, perhaps with occasional

bursts of volatility.

      b. Perform a correlogram analysis of $e_t$, and discuss your results.

* Remarks, suggestions, hints, solutions: The correlogram will probably be flat, as most asset

returns are very close to (weak) white noise.

      c. Plot $e_t^2$, and discuss your results.

* Remarks, suggestions, hints, solutions: The squared return, in contrast to the return itself, may

well be serially correlated, indicating nonlinear dependence operative through the conditional

variance.

      d. In addition to plotting $e_t^2$, examining the correlogram of $e_t^2$ often proves informative

           for assessing volatility persistence. Why might that be so? Perform a correlogram

           analysis of $e_t^2$ and discuss your results.

* Remarks, suggestions, hints, solutions: Ditto.

## Chapter 8 Problems and Complements

1. (ARMA lag inclusion)  Review Table 1.  Why is the MA(3) term included even though the p-value indicates that it is not significant?  What would be the costs and benefits of dropping the insignificant MA(3) term?

* Remarks, suggestions, hints, solutions:  The convention is to include all lags up to the maximum adopted, but insignificant lags could presumably be dropped without doing damage.  In large samples there would be little gain from doing so, but in small samples, with limited degrees of freedom, the gains would be larger.

2. (Shapes of correlograms)  Given the following ARMA processes, sketch the expected forms of the autocorrelation and partial autocorrelation functions.  (Hint:  examine the roots of the various autoregressive and moving average lag operator polynomials.)

a. $y_t = \left( \dfrac{1}{1 - 1.05L - .09L^2} \right) \varepsilon_t$

* Remarks, suggestions, hints, solutions: Factoring to obtain the roots reveals that the process is not covariance stationary, so the autocorelation function will fail to damp.

b. $y_t = (1 - .4L)\varepsilon_t$

* Remarks, suggestions, hints, solutions: Sharp cutoff beyond lag one, due to MA(1) structure.

c. $y_t = \left( \dfrac{1}{1 - .7L} \right) \varepsilon_t.$

* Remarks, suggestions, hints, solutions: Exponential decay, due to AR(1) structure.

3. (The autocovariance function of the MA(1) process, revisited)  In the text we wrote

$$\gamma(\tau) = E(y_t y_{t-\tau}) = E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-\tau} + \theta\varepsilon_{t-\tau-1})) = \begin{cases} \theta\sigma^2, & \tau=1 \\ 0, & \text{otherwise.} \end{cases}$$

Fill in the missing steps by evaluating explicitly the expectation $E((\varepsilon_t+\theta\varepsilon_{t-1})(\varepsilon_{t-\tau}+\theta\varepsilon_{t-\tau-1}))$.

* Remarks, suggestions, hints, solutions:  When we set $\tau=1$, we get $E((\varepsilon_t+\theta\varepsilon_{t-1})(\varepsilon_{t-1}+\theta\varepsilon_{t-2}))$, but

all expected cross products of $\varepsilon$'s vanish (because $\varepsilon$ is white noise, so the covariance of any two

$\varepsilon$'s at different times is zero) with the exception of $E\theta\varepsilon_{t-1}^2$.  So we get $E\theta\varepsilon_{t-1}^2 = \theta E\varepsilon_{t-1}^2 = \theta\sigma^2$.

When we set $\tau>1$, all expected cross products of $\varepsilon$'s vanish, so we simply get 0.

4. (ARMA algebra)  Derive expressions for the autocovariance function, autocorrelation

function, conditional mean, unconditional mean, conditional variance and unconditional variance

of the following processes:

    a. $y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$

    b. $y_t = \varphi y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$.

* Remarks, suggestions, hints, solutions:  Do the multiplication and move in the expectations

operator.  For $\tau=0$, all cross products vanish in expectation.  For $\tau=1$, all vanish except $E(\theta\varepsilon_{t-1}^2)$,

which is $\theta\sigma^2$.

5. (Diagnostic checking of model residuals)  If a forecasting model has extracted all the

systematic information from the data, then what's left -- the residual -- should be white noise.

More precisely, the true innovations are white noise, and if a model is a good approximation to

the Wold representation then its 1-step-ahead forecast errors should be approximately white

noise.  The model residuals are the in-sample analog of out-of-sample 1-step-ahead forecast

errors.  Hence the usefulness of various tests of the hypothesis that residuals are white noise.

    The Durbin-Watson test is the most popular.  Recall the Durbin-Watson test statistic,

discussed in Chapter 2,

$$DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}.$$

Note that

$$\sum_{t=2}^{T} (e_t - e_{t-1})^2 \approx 2\sum_{t=2}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1}.$$

Thus

$$DW \approx 2(1 - \hat{\rho}(1)),$$

so that the Durbin-Watson test is effectively based only on the first sample autocorrelation and really only tests whether the first autocorrelation is zero. We say therefore that the Durbin-Watson is a test for first-order serial correlation. In addition, the Durbin-Watson test is not valid in the presence of lagged dependent variables.[2]  On both counts, we'd like a more general and flexible framework for diagnosing serial correlation. The residual correlogram, comprised of the residual sample autocorrelations, the sample partial autocorrelations, and the associated Q statistics, delivers the goods.

a.  When we discussed the correlogram in the text, we focused on the case of an observed

---

[2] Following standard, if not strictly appropriate, practice, in this book we often report and examine the Durbin-Watson statistic even when lagged dependent variables are included.  We always supplement the Durbin-Watson statistic, however, with other diagnostics such as the residual correlogram, which remain valid in the presence of lagged dependent variables, and which almost always produce the same inference as the Durbin-Watson statistic.

time series, in which case we showed that the Q statistics are distributed as $\chi^2_m$.

Now, however, we want to assess whether unobserved model disturbances are

white noise. To do so, we use the model residuals, which are estimates of the

unobserved disturbances. Because we fit a model to get the residuals, we need to

account for the degrees of freedom used. The upshot is that the distribution of the

Q statistics under the white noise hypothesis is better approximated by a $\chi^2_{m-k}$

random variable, where k is the number of parameters estimated. That's why, for

example, we don't report (and in fact the software doesn't compute) the p-values

for the Q statistics associated with the residual correlogram of our employment

forecasting model until m>k.

b. Durbin's h test is an alternative to the Durbin-Watson test. As with the Durbin-Watson

test, it's designed to detect first-order serial correlation, but it's valid in the

presence of lagged dependent variables. Do some background reading as well on

Durbin's h test and report what you learned.

c. The Breusch-Godfrey test is another alternative to the Durbin-Watson test. It's

designed to detect $p^{th}$-order serial correlation, where p is selected by the user, and

is also valid in the presence of lagged dependent variables. Do some background

reading on the Breusch-Godfrey procedure and report what you learned.

d. Which do you think is likely to be most useful to you in assessing the properties of

residuals from forecasting models: the residual correlogram, Durbin's h test, or the

Breusch-Godfrey test? Why?

* Remarks, suggestions, hints, solutions: The residual correlogram, supplemented with the

Bartlett bands and perhaps also a Ljung-Box test is the most widely used, and in my view the most useful, because of the visual insights afforded by the correlogram. The other tests nevertheless have their uses, and it's important for students to be introduced to them.

6. (Mechanics of fitting ARMA models) On the book's web page you will find data for daily transfers over BankWire, a financial wire transfer system in a country responsible for much of the world's finance, over a recent span of 200 business days.

    a. Is trend or seasonality operative? Defend your answer.

    b. Using the methods developed in Chapters 7 and 8, find a parsimonious ARMA(p,q) model that fits well, and defend its adequacy.

* Remarks, suggestions, hints, solutions: This exercise is very useful for getting the students comfortable with the mechanics of fitting ARMA models and is intentionally open-ended. The true data-generating process contains no trend and no seasonality, but it does of course have a nonzero mean. The dynamics are ARMA(1,1), with a large autoregressive root and a smaller moving average root.

7. (Modeling cyclical dynamics) As a research analyst at the U.S. Department of Energy, you have been asked to model non-seasonally-adjusted U.S. imports of crude oil.

    a. Find a suitable time series on the web.

    b. Create a model that captures the trend in the series.

    c. Adding to the model from part b, create a model with trend and a full set of seasonal dummy variables.

    d. Observe the residuals of the model from part b and their correlogram. Is there evidence neglected dynamics? If so, what to do?

8. (Aggregation and disaggregation: top-down vs. bottom-up forecasting models) Related to the issue of methods and complexity discussed in Chapter 3 is the question of aggregation. Often we want to forecast an aggregate, such as total sales of a manufacturing firm, but we can take either an aggregated or disaggregated approach.

Suppose, for example, that total sales is composed of sales of three products. The aggregated, or top-down, or macro, approach is simply to model and forecast total sales. The disaggregated, or bottom- up, or micro, approach is to model and forecast separately the sales of the individual products, and then to add them together.

Perhaps surprisingly, it's impossible to know in advance whether the aggregated or disaggregated approach is better. It all depends on the specifics of the situation; the only way to tell is to try both approaches and compare the forecasting results.

However, in real-world situations characterized by likely model misspecification and parameter estimation uncertainty, there are reasons to suspect that the aggregated approach may be preferable. First, standard (e.g., linear) models fit to aggregated series may be less prone to specification error, because aggregation can produce approximately linear relationships even when the underlying disaggregated relationships are not linear. Second, if the disaggregated series depend in part on a common factor (e.g., general business conditions) then it will emerge more clearly in the aggregate data. Finally, modeling and forecasting of one aggregated series, as opposed to many disaggregated series, relies on far fewer parameter estimates.

Of course, if our interest centers on the disaggregated components, then we have no choice but to take a disaggregated approach.

It is possible that an aggregate forecast may be useful in forecasting disaggregated series.

Why? (Hint: See Fildes and Stekler, 2000.)

* Remarks, suggestions, hints, solutions: If factor structure is present, the aggregate forecast may effectively be a (very good) forecast of the factor, which could help with disaggregated forecasting, as each disaggregated series depends in part on the factor.

9. (Nonlinear forecasting models: regime switching) In this chapter we've studied dynamic linear models, which are tremendously important in practice. They're called linear because $y_t$ is a simple linear function of past y's or past $\varepsilon$'s. In some forecasting situations, however, good statistical characterization of dynamics may require some notion of regime switching, as between "good" and "bad" states, which is a type of nonlinear model.

Models incorporating regime switching have a long tradition in business-cycle analysis, in which expansion is the good state, and contraction (recession) is the bad state. This idea is also manifest in the great interest in the popular press, for example, in identifying and forecasting turning points in economic activity. It is only within a regime-switching framework that the concept of a turning point has intrinsic meaning; turning points are naturally and immediately defined as the times separating expansions and contractions.

Threshold models are squarely in line with the regime-switching tradition. The following threshold model, for example, has three regimes, two thresholds, and a d-period delay regulating the switches:

$$
y_t = \begin{cases}
c^{(u)} + \varphi^{(u)} y_{t-1} + \varepsilon_t^{(u)}, & \theta^{(u)} < y_{t-d} \\
c^{(m)} + \varphi^{(m)} y_{t-1} + \varepsilon_t^{(m)}, & \theta^{(l)} < y_{t-d} < \theta^{(u)} \\
c^{(l)} + \varphi^{(l)} y_{t-1} + \varepsilon_t^{(l)}, & \theta^{(l)} > y_{t-d} \, .
\end{cases}
$$

The superscripts indicate "upper," "middle," and "lower" regimes, and the regime operative at any

time t depends on the observable past history of y -- in particular, on the value of $y_{t-d}$.

    Although observable threshold models are of interest, models with *latent* (or unobservable) states as opposed to observed states may be more appropriate in many business, economic and financial contexts. In such a setup, time-series dynamics are governed by a finite-dimensional parameter vector that switches (potentially each period) depending upon which of two unobservable states is realized, with state transitions governed by a first-order Markov process (meaning that the state at any time t depends only on the state at time t-1, not at time t-2, t-3, etc.).

    To make matters concrete, let's take a simple example. Let $\{s_t\}_{t=1}^{T}$ be the (latent) sample path of two-state first-order autoregressive process, taking just the two values 0 or 1, with transition probability matrix given by

$$M = \begin{pmatrix} p_{00} & 1-p_{00} \\ 1-p_{11} & p_{11} \end{pmatrix}.$$

The ij-th element of M gives the probability of moving from state i (at time t-1) to state j (at time t). Note that there are only two free parameters, the staying probabilities, $p_{00}$ and $p_{11}$. Let $\{y_t\}_{t=1}^{T}$ be the sample path of an observed time series that depends on $\{s_t\}_{t=1}^{T}$ such that the density of $y_t$ conditional upon $s_t$ is

$$f(y_t|s_t; \theta) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(\frac{-(y_t-\mu_{s_t})^2}{2\sigma^2}\right).$$

Thus, $y_t$ is Gaussian white noise with a potentially switching mean. The two means around which $y_t$ moves are of particular interest and may, for example, correspond to episodes of differing growth rates ("booms" and "recessions", "bull" and "bear" markets, etc.).

* Remarks, suggestions, hints, solutions: A detailed treatment of nonlinear regime-switching models is largely beyond the scope of the text, but the idea is nevertheless intuitive and worth introducing to the students.

10. (Difficulties with nonlinear optimization) Nonlinear optimization is a tricky business, fraught with problems. Some eye-opening reading includes Newbold, Agiakloglou and Miller (1994) and McCullough and Vinod (1999).

Some problems are generic. It's relatively easy to find a local optimum, for example, but much harder to be confident that the local optimum is global. Simple checks such as trying a variety of startup values and checking the optimum to which convergence occurs are used routinely, but the problem nevertheless remains. Other problems may be software specific. For example, some software may use highly accurate analytic derivatives whereas other software uses approximate numerical derivatives. Even the same software package may change algorithms or details of implementation across versions, leading to different results. Software for ARMA model estimation is unavoidably exposed to all such problems, because estimation of any model involving MA terms requires numerical optimization of a likelihood or sum-of-squares function.

## Chapter 9 Problems and Complements

1. (Forecast accuracy across horizons) You are a consultant to MedTrax, a large pharmaceutical company, which released a new ulcer drug three months ago and is concerned about recovering research and development costs. Accordingly, MedTrax has approached you for drug sales projections at 1- through 12-month-ahead horizons, which it will use to guide potential sales force realignments. In briefing you, MedTrax indicated that it expects your long-horizon forecasts (e.g., 12-month-ahead) to be just as accurate as your short-horizon forecasts (e.g., 1-month-ahead). Explain to MedTrax why that is not likely to be the case, even if you do the best forecasting job possible.

* Remarks, suggestions, hints, solutions: For a variety of reasons, the distant future is likely to be harder to forecast than the near future, even if the forecaster is behaving optimally.

2. (Mechanics of forecasting with ARMA models: BankWire continued) On the book's web page you will find data for daily transfers over BankWire, a wire transfer system in a country responsible for much of the world's finance, over a recent span of 200 business days.

    a. In the Chapter 8 Exercises, Problems and Complements, you were asked to find a parsimonious ARMA(p,q) model that fits the transfer data well, and to defend its adequacy. Repeat the exercise, this time using only the first 175 days for model selection and fitting. Is it necessarily the case that the selected ARMA model will remain the same as when all 200 days are used? Does yours?

    b. Use your estimated model to produce point and interval forecasts for days 176 through 200. Plot them and discuss the forecast pattern.

c. Compare your forecasts to the actual realizations.  Do the forecasts perform well?

Why or why not?

d. Discuss precisely how your software constructs point and interval forecasts.  It should

certainly match our discussion in spirit, but it may differ in some of the details.

Are you uncomfortable with any of the assumptions made?  How, if at all, could

the forecasts be improved?

\* Remarks, suggestions, hints, solutions: the transfer data display quite a lot of persistence, and

the last few observations are rather far below the mean, so the forecast is for gradual mean

reversion.

3.  (Forecasting an AR(1) process with known and unknown parameters)  Use the chain rule to

forecast the AR(1) process,

$$y_t = \varphi y_{t-1} + \varepsilon_t.$$

For now, assume that all parameters are known.

a. Show that the optimal forecasts are

$$y_{T+1,T} = \varphi y_T$$

$$y_{T+2,T} = \varphi^2 y_T$$

$$\ldots$$

$$y_{T+h,T} = \varphi^h y_T.$$

* Remarks, suggestions, hints, solutions: The 1-step-ahead forecast is trivial to construct, as is

the 2-step-ahead forecast given the 1-step-ahead forecast, and so on.

  b. Show that the corresponding forecast errors are

$$e_{T+1,T} = \left(y_{T+1} - y_{T+1,T}\right) = \varepsilon_{T+1}$$

$$e_{T+2,T} = \left(y_{T+2} - y_{T+2,T}\right) = \varphi\, \varepsilon_{T+1} + \varepsilon_{T+2}$$

...

$$e_{T+h,T} = \left(y_{T+h} - y_{T+h,T}\right) = \varepsilon_{T+h} + \varphi\varepsilon_{T+h-1} + ... + \varphi^{h-1}\varepsilon_{T+1}.$$

* Remarks, suggestions, hints, solutions: Simple algebra.

  c. Show that the forecast error variances are

$$\sigma_1^2 = \sigma^2$$

$$\sigma_2^2 = \sigma^2(1 + \varphi^2)$$

...

$$\sigma_h^2 = \sigma^2\sum_{i=0}^{h-1} \varphi^{2i}.$$

* Remarks, suggestions, hints, solutions: Follows immediately by taking variances of the earlier-

-74-

computed errors.

    d. Show that the limiting forecast error variance is

$$\lim_{h \to \infty} \sigma_h^2 = \frac{\sigma^2}{1-\varphi^2},$$

    the unconditional variance of the AR(1) process.

\* Remarks, suggestions, hints, solutions:  As h approaches infinity, the summation approaches a

geometric series with first term 1 and ratio $\varphi^2$, which sums to $1/(1-\varphi^2)$ so long as $|\varphi|<1$.

Now assume that the parameters are unknown and so must be estimated.

    e. Make your expressions for both the forecasts and the forecast error variances

        operational, by inserting least squares estimates where unknown parameters

        appear, and use them to produce an operational point forecast and an operational

        90% interval forecast for $y_{T+2,T}$.

\* Remarks, suggestions, hints, solutions:  The operational forecast is

$$\hat{y}_{T+2,T} = \hat{\varphi}^2 \, y_T$$

and the operational forecast error variance is

$$\hat{\sigma}_2^2 = \hat{\sigma}^2(1 + \hat{\varphi}^2).$$

Hence the operational interval forecast is

$$\hat{y}_{t+2,t} \pm 1.64\hat{\sigma}_2.$$

4. (Forecasting an ARMA(2,2) process)  Consider the ARMA(2,2) process:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

    a. Verify that the optimal 1-step ahead forecast made at time T is

$$y_{T+1,T} = \varphi_1 y_T + \varphi_2 y_{T-1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}.$$

* Remarks, suggestions, hints, solutions:  Proceed in the usual way, projecting on the time-T information set.

    b. Verify that the optimal 2-step ahead forecast made at time T is

$$y_{T+2,T} = \varphi_1 y_{T+1,T} + \varphi_2 y_T + \theta_2 \varepsilon_T,$$

and express it purely in terms of elements of the time-T information set.

* Remarks, suggestions, hints, solutions:  Projection on the time-(T+1) information set gives the expression above.  To express it purely in terms of elements of the time-T information set, insert the earlier expression for $y_{T+1,T}$.

    c. Verify that the optimal 3-step ahead forecast made at time T is

$$y_{T+3,T} = \varphi_1 y_{T+2,T} + \varphi_2 y_{T+1,T},$$

and express it purely in terms of elements of the time-T information set.

* Remarks, suggestions, hints, solutions:  Projection on the time-(T+2) information set gives the expression above.  To express it purely in terms of elements of the time-T information set, insert the earlier expression for $y_{T+2,T}$ in terms of the time-T information set.

     d.  Show that for any forecast horizon h greater than or equal to three,

$$y_{T+h,T} = \varphi_1 y_{T+h-1,T} + \varphi_2 y_{T+h-2,T}.$$

* Remarks, suggestions, hints, solutions:  Beyond h=3, all the moving average terms project to 0, which leaves the above prediction equation.  The upshot is that the moving average terms wash out of the prediction equation once h is greater than the moving average order.

5.  (Optimal forecasting under asymmetric loss)  One of the conditions required for optimality of the conditional mean forecast is symmetric loss.  We make that assumption for a number of reasons.  First, the conditional mean is usually easy to compute.  In contrast, optimal forecasting under asymmetric loss is rather involved, and the tools for doing so are still under development. (See, for example, Christoffersen and Diebold, 1997.)  Second, and more importantly, symmetric loss often provides a good approximation to the loss structure relevant in a particular decision environment.

     Symmetric loss is not *always* appropriate, however.  Here we discuss some aspects of forecasting under asymmetric loss.  Under asymmetric loss, optimal forecasts are biased, whereas the conditional mean forecast is unbiased.[3]  Bias is optimal under asymmetric loss because we can gain on average by pushing the forecasts in the direction such that we make relatively few errors

---

[3]  A forecast is unbiased if its error has zero mean.  The error from the conditional mean forecast has zero mean, by construction.

of the more costly sign.

There are many possible asymmetric loss functions. A few, however, have proved particularly useful, because of their flexibility and tractability. One is the linex loss function,

$$L(e) = b[\exp(ae) - ae - 1], \quad a \neq 0, \, b > 0.$$

It's called linex because when a>0, loss is approximately linear to the left of the origin and approximately exponential to the right, and conversely when a<0. Another is the linlin loss function, given by

$$L(e) = \begin{cases} a|e|, & \text{if } e > 0 \\ b|e|, & \text{if } e \leq 0. \end{cases}$$

Its name comes from the linearity on each side of the origin.

    a. Discuss three practical forecasting situations in which the loss function might be asymmetric. Give detailed reasons for the asymmetry, and discuss how you might produce and evaluate forecasts.

* Remarks, suggestions, hints, solutions: This is a good question for generating discussion among the students. Under asymmetric loss, optimal forecasts are biased. Forecasts should be evaluated using the relevant loss function.

    b. Explore and graph the linex and linlin loss functions for various values of a and b. Discuss the roles played by a and b in each loss function. In particular, which parameter or combination of parameters governs the degree of asymmetry? What happens to the linex loss function as a gets smaller? What happens to the linlin

loss function as a/b approaches one?

* Remarks, suggestions, hints, solutions:  As a gets smaller, linex loss approaches quadratic loss. As a/b approaches 1, linlin loss approaches absolute loss.

6.  (Truncation of infinite distributed lags, state space representations, and the Kalman filter)  This complement concerns practical implementation of formulae that involve innovations ($\varepsilon$'s).  Earlier we noted that as long as a process is invertible we can express the $\varepsilon$'s in terms of the y's.  If the process involves a moving average component, however, the $\varepsilon$'s will depend on the infinite past history of the y's, so we need to truncate to make it operational.  Suppose, for example, that we're forecasting the MA(1) process,

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}.$$

The operational 1-step-ahead forecast is

$$y_{t+1,T} = \hat{\theta}\hat{\varepsilon}_T.$$

But what, precisely, do we insert for the residual, $\hat{\varepsilon}_T$?  Back substitution yields the autoregressive representation,

$$\varepsilon_t = y_t + \theta y_{t-1} - \theta^2 y_{t-2} + \dots$$

Thus,

$$\varepsilon_T = y_T + \theta y_{T-1} - \theta^2 y_{T-2} + ...,$$

which we are forced to truncate at time T=1, when the data begin. This yields the approximation

$$\varepsilon_T \approx y_T + \theta y_{T-1} - \theta^2 y_{T-2} + ... + \theta^T y_1.$$

Unless the sample size is very small, or $\theta$ is very close to 1, the approximation will be very accurate, because $\theta$ is less than one in absolute value (by invertibility), and we're raising it to higher and higher powers. Finally, we make the expression operational by replacing the unknown moving average parameter with an estimate, yielding

$$\hat{\varepsilon}_T \approx y_T + \hat{\theta} y_{T-1} - \hat{\theta}^2 y_{T-2} + ... + \hat{\theta}^T y_1.$$

In the engineering literature of the 1960s, and then in the statistics and econometrics literatures of the 1970s, important tools called state space representations and the Kalman filter were developed. Those tools provide a convenient and powerful framework for estimating a wide variety of forecasting models and constructing optimal forecasts, and they enable us to tailor the forecasts precisely to the sample of data at hand, so that no truncation is necessary.

* Remarks, suggestions, hints, solutions: Apart from this complement, the Kalman filter is an advanced topic that is beyond the scope of the text. It's good, however, to try to give students a feel for the more advanced material.

7. (Point and interval forecasts allowing for serial correlation - Nile.com continued) On the book's website you will find data for the internet retailer Nile.com, giving the number of hits at

the Nile.com website each day from 1/1/1998 through 9/28/1998.  Your marketing firm,

CyberMedia, which specializes in developing quick, intensive marketing strategies based on short

term projections, is hired to develop a forecasting model for hits at the Nile.com website.

    a.  In Chapter 6, Problem 6, you estimated a trend + seasonal model for Nile.com hits,

        ignoring the possible presence of cyclical dynamics.  Now generalize your earlier

        model to allow for cyclical dynamics, if present, via AR(p) disturbances.  Write the

        full specification of your model in general notation (e.g., with p left unspecified).

    b.  Estimate three versions of your full model, corresponding to p = 0, 1, 2, 3, while

        leaving the original trend and seasonal specifications intact, and select the one that

        optimizes SIC.

    c.  Using the model selected in part b, write theoretical expressions for the 1- and 2-day-

        ahead point forecasts and 95% interval forecasts, using estimated parameters.

    d.  Calculate those point and interval forecasts for Nile.com for 9/29 and 9/30.

* Remarks, suggestions, hints, solutions: $y_t = \beta_0 + \sum_{i=1}^{K} \beta_i x_{it} + \varepsilon_t, \quad \varepsilon_t = \varphi \varepsilon_{t-1} + v_t, \quad v_t \sim iid$, where the x

variables are the various included trend and seasonal terms.

7.  (Bootstrap simulation to acknowledge innovation distribution uncertainty and parameter

estimation uncertainty)  A variety of simulation-based methods fall under the general heading of

"bootstrap."  Their common element, and the reason for the name bootstrap, is that they build up

an approximation to an object of interest directly from the data.  Hence they "pull themselves up

by their own bootstrap."  For example, the object of interest might be the distribution of a random

disturbance, which has implications for interval and density forecasts, and about which we might

sometimes feel uncomfortable making a possibly erroneous assumption such as normality.

a. The density and interval forecasts that we've discussed rely crucially on normality. In many situations, normality is a perfectly reasonable and useful assumption; after all, that's why we call it the "normal" distribution. Sometimes, however, such as when forecasting high-frequency financial asset returns, normality may be unrealistic. Using bootstrap methods we can relax the normality assumption. Suppose, for example, that we want a 1-step-ahead interval forecast for an AR(1) process. We know that the future observation of interest is

$$y_{T+1} = \varphi y_T + \varepsilon_{T+1}.$$

We know $y_T$, and we can estimate $\varphi$ and then proceed as if $\varphi$ were known, using the operational point forecast, $\hat{y}_{T+1,T} = \hat{\varphi} y_T$. If we want an operational interval forecast, however, we've thus far relied on a normality assumption, in which case we use $\hat{y}_{T+1,T} \pm z_{\alpha/2} \hat{\sigma}$. To relax the normality assumption, we can proceed as follows. Imagine that we could sample from the distribution of $\varepsilon_{T+1}$ -- whatever that distribution might be. Take R draws, $\{\varepsilon_{T+1}^{(i)}\}_{i=1}^{R}$, where R is a large number, such as 10000. For each such draw, construct the corresponding forecast of $y_{T+1}$ as

$$\hat{y}_{T+1,T}^{(i)} = \hat{\varphi} y_T + \varepsilon_{T+1}^{(i)}.$$

Then form a histogram of the $\hat{y}_{T+1,T}^{(i)}$ values, which is the density forecast. And given the density forecast, we can of course construct interval forecasts at any desired level. If, for example, we want a 90% interval we can sort the $\hat{y}_{T+1,T}^{(i)}$

values from smallest to largest, and find the 5th percentile (call it a) and the 95th

percentile (call it b), and use the 90% interval forecast [a, b].

* Remarks, suggestions, hints, solutions:  Obviously, students at this level will never be experts in

bootstrap theory or application.  Rather, the idea is to introduce them to the simple and powerful

idea of resampling, and more generally, to the uses of simulation in modeling and forecasting.

b.  The only missing link in the strategy above is how to sample from the distribution of

$\varepsilon_{T+1}$.  It turns out that it's easy to do -- we simply assign probability 1/T to each of

the observed residuals  (which are estimates of the unobserved $\varepsilon$'s) and draw from

them R times with replacement.  Describe how you might do so.

* Remarks, suggestions, hints, solutions:  Just split the unit interval into T parts, draw a U(0,1)

variate, determine the cell in which it falls, and use the corresponding residual.

c.  Note that the interval and density forecasts we've constructed thus far -- even the one

above  based on bootstrap techniques -- make no attempt to account for parameter

estimation uncertainty.  Intuitively, we would expect confidence intervals obtained

by ignoring parameter estimation uncertainty to be more narrow than they would

be if parameter uncertainty were accounted for, thereby producing an artificial

appearance of precision.  In spite of this defect, parameter uncertainty is usually

ignored in practice, for a number of reasons.  The uncertainty associated with

estimated parameters vanishes as the sample size grows, and in fact it vanishes

quickly.  Furthermore, the fraction of forecast error attributable to the difference

between estimated and true parameters is likely to be small compared to the

fraction of forecast error coming from other sources, such as using a model that

does a poor job of approximating the dynamics of the variable being forecast.

d.  Quite apart from the reasons given above for ignoring parameter estimation

uncertainty, the biggest reason is probably that, until very recently, mathematical

and computational difficulties made attempts to account for parameter uncertainty

infeasible in many situations of practical interest.  Modern computing speed,

however, lets us use the bootstrap to approximate the effects of parameter

estimation uncertainty.  To continue with the AR(1) example, suppose that we

know that the disturbances are Gaussian, but that we want to attempt to account

for the effects of parameter estimation uncertainty when we produce our 1-step-

ahead density forecast.  How could we use the bootstrap to do so?

* Remarks, suggestions, hints, solutions:  Now we have to account for both innovation and

parameter estimation uncertainty.  First obtain an approximation to the distribution of the least

squares AR(1) parameter estimator by parametric bootstrap.  After that, generate R future values

of the series, each time drawing an AR(1) parameter from its sampling distribution and an

innovation from the appropriate normal distribution.

e.  The "real sample" of data ends with observation $y_T$, and the optimal point forecast

depends only on $y_T$.  It would therefore seem desirable that all of your R

"bootstrap samples" of data also end with $y_T$.  Do you agree?  How might you

enforce that property while still respecting the AR(1) dynamics?  (This is tricky.)

* Remarks, suggestions, hints, solutions:  For a particular sample path, which is all we have in any

practical application, it seems compelling to enforce the condition that all the bootstrap samples of

data end with $y_T$, as did the actual sample.  This can be done for an AR(1) process with Gaussian

disturbances by generating realizations with $y_T$ as the *initial* value, and then reversing the

realization. (The linear Gaussian AR(1) process is time reversible.)

      f. Can you think of a way to assemble the results thus far to produce a density forecast

          that acknowledges both innovation distribution uncertainty and parameter

          estimation uncertainty? (This is very challenging.)

* Remarks, suggestions, hints, solutions: First obtain an approximation to the distribution of the

least squares AR(1) parameter estimator by non-parametric bootstrap. After that, generate R

future values of the series, each time drawing an AR(1) parameter from its sampling distribution

and an innovation from the empirical distribution of the residuals.

**Chapter 10 Problems and Complements**

1. (Serially correlated disturbances vs. lagged dependent variables)  Estimate the quadratic trend

model for log liquor sales with seasonal dummies and three lags of the dependent variable

included directly.  Discuss your results and compare them to those we obtained when we instead

allowed for AR(3) disturbances in the regression.

* Remarks, suggestions, hints, solutions:  The key point is to drive home the intimate relationship

between regression models with AR(p) disturbances and regression models with p lags of the

dependent variable.

2. (Assessing the adequacy of the liquor sales forecasting model trend specification)  Critique the

liquor sales forecasting model that we adopted (log liquor sales with quadratic trend, seasonal

dummies, and AR(3) disturbances).[4]

 a.  If the trend is not a good approximation to the actual trend in the series, would it

 greatly affect short-run forecasts?  Long-run forecasts?

* Remarks, suggestions, hints, solutions:  Misspecification of the trend would likely do more harm

to long-run forecasts than to short-run forecasts.

 b.  Fit and assess the adequacy of a model with log-linear trend.

* Remarks, suggestions, hints, solutions:  The fitting is easy, and the assessment can be done in

many ways, such as by comparing the actual and fitted values, plotting the residuals against

powers of time, seeing which trend specification the SIC selects, etc.

 c.  How might you fit and assess the adequacy of a *broken* linear trend?  How might you

---

 [4] I thank Ron Michener, University of Virginia, for suggesting parts d and f.

decide on the location of the break point?

* Remarks, suggestions, hints, solutions:  Broken linear trend can be implemented by including

appropriate dummy variables, with the breakpoint selected based on prior knowledge or by

minimizing the sum of squared residuals.  (Be careful of data mining, however.)  The broken

linear trend model could be compared to other trend models using the usual criteria, such as SIC.

3.  (Improving non-trend aspects of the liquor sales forecasting model)

        a.  Recall our earlier argument from Chapter 8 that best practice requires using a $\chi^2_{m-k}$

           distribution rather than a $\chi^2_m$ distribution to assess the significance of Q-statistics

           for model residuals, where m is the number of autocorrelations included in the

           Box-Pierce statistic and k is the number of parameters estimated.  In several places

           in this chapter, we failed to heed this advice when evaluating the liquor sales

           model.  If we were instead to compare the residual Q-statistic p-values to a $\chi^2_{m-k}$

           distribution, how, if at all, would our assessment of the model's adequacy change?

* Remarks, suggestions, hints, solutions:  Because the $\chi^2_{m-k}$ distribution is shifted left relative to

the $\chi^2_m$ distribution, it is likely that more of the Q-statistics will appear significant.  That is, the

evidence against adequacy of the model will be increased.

        b.  Return to the log-quadratic trend model with seasonal dummies, allow for ARMA(p,q)

           disturbances, and do a systematic selection of p and q using the AIC and SIC.  Do

           AIC and SIC select the same model?  If not, which do you prefer?  If your

           preferred forecasting model differs from the AR(3) that we used, replicate the

           analysis in the text using your preferred model, and discuss your results.

* Remarks, suggestions, hints, solutions:  Regardless of whether the selected model differs from

the AR(3), the qualitative results of the exercise are likely to be unchanged, because the AR(3)

provides a very good approximation to the dynamics, even if it is not the "best."

      c. Discuss and evaluate another possible model improvement:  inclusion of an additional

          dummy variable indicating the number of Fridays and/or Saturdays in the month.

          Does this model have lower AIC or SIC than the final model used in the text?  Do

          you prefer it to the one in the text?  Why or why not?

* Remarks, suggestions, hints, solutions:  It's a good idea!

4.  (CUSUM analysis of the housing starts model)  Consider the housing starts forecasting model

that we built in Chapter 6.

      a. Perform a CUSUM analysis of a housing starts forecasting model that does not account

          for cycles.   (Recall that our model in Chapter 6 did not account for cycles).

          Discuss your results.

* Remarks, suggestions, hints, solutions:  It is likely that the joint hypothesis of correct model

specification and parameter stability will be rejected.  We know, however, that the model is

presently incorrectly specified, because housing starts have a cyclical component.  Thus, the fact

that the CUSUM test rejects does not necessarily imply parameter instability.

      b. Specify and estimate a model that *does* account for cycles.

* Remarks, suggestions, hints, solutions:  This could be done either by including lagged

dependent variables or serially correlated disturbances.

      c. Do a CUSUM analysis of the model that accounts for cycles.  Discuss your results and

          compare them to those of part a.

* Remarks, suggestions, hints, solutions:  It's much less likely that the CUSUM will reject, now

that we've made a serious attempt at model specification. Ultimately, there is little evidence of

parameter instability in the model.

5. (Model selection based on simulated forecasting performance)

      a. Return to the retail sales data of Chapter 5, and use recursive cross validation to select

            between the linear trend forecasting model and the quadratic trend forecasting

            model. Which do you select? How does it compare with the model selected by

            the AIC and SIC?

* Remarks, suggestions, hints, solutions: The crucial point, of course, is not the particular model

selected, but rather that the students get comfortable with recursive estimation and prediction.

      b. How did you decide upon a value of $T^*$ when performing the recursive cross validation

            on the retail sales data? What are the relevant considerations?

* Remarks, suggestions, hints, solutions: $T^*$ should be large enough such that the initial

estimation is meaningful, yet small enough so that a substantial part of the sample is used for out-

of-sample forecast comparison.

      c. One virtue of recursive cross validation procedures is their flexibility. Suppose that

            your loss function is not 1-step-ahead mean squared error; instead, suppose it's an

            asymmetric function of the 1-step-ahead error. How would you modify the

            recursive cross validation procedure to enforce the asymmetric loss function?

            How would you proceed if the loss function were 4-step-ahead squared error?

            How would you proceed if the loss function were an average of 1-step-ahead

            through 4-step-ahead squared error?

* Remarks, suggestions, hints, solutions: We would simply modify the procedure to compare the

appropriate asymmetric function of the 1-step-ahead error, or 4-step-ahead squared error. We might even go farther and use the relevant loss function in estimation.

6. (Seasonal models with time-varying parameters: forecasting AirSpeed passenger-miles) You work for a hot new startup airline, AirSpeed, modeling and forecasting the miles per person ("passenger-miles") traveled on their flights through the four quarters of the year. During the past fifteen years for which you have data, it's well known in the industry that trend passenger-miles have been flat (that is, there is no trend), and similarly, there have been no cyclical effects. It is believed by industry experts, however, that there are strong seasonal effects, which you think might be very important for modeling and forecasting passenger-miles.

     a. Why might airline passenger-miles be seasonal?

* Remarks, suggestions, hints, solutions: Travel, for example, increases around holidays such as Christmas and Thanksgiving, and in the summer.

     b. Fit a quarterly seasonal model to the AirSpeed data, and assess the importance of seasonal effects. Do the t and F tests indicate that seasonality is important? Do the Akaike and Schwarz criteria indicate that seasonality is important? What is the estimated seasonal pattern?

* Remarks, suggestions, hints, solutions: The students should do t tests on the individual seasonal coefficients, as well as an F test of the hypothesis that the seasonal coefficients are identical across seasons. The AIC and SIC can be used to compare models with and without seasonality. It is a good idea to have the students plot and discuss the estimated seasonal pattern, which is just the set of four seasonal coefficients.

     c. Use recursive procedures to assess whether the seasonal coefficients are evolving over

time.  Discuss your results.

* Remarks, suggestions, hints, solutions:  Compute and graph the recursive seasonal parameter

estimates.  Also do a formal CUSUM analysis.

      d.  If the seasonal coefficients are evolving over time, how might you model that evolution

         and thereby improve your forecasting model? (Hint:  Allow for trends in the

         seasonal coefficients themselves.)

* Remarks, suggestions, hints, solutions:  If we allow for a linear trend in each of the four

seasonal coefficients, then we need to include in the regression not only four seasonal dummies,

but also the products of those dummies with time.

      e.  Compare 4-quarter-ahead extrapolation forecasts from your models with and without

         evolving seasonality.

* Remarks, suggestions, hints, solutions:  I've left this to you!

7.  (Formal models of unobserved components)  We've used the idea of unobserved components

as informal motivation for our models of trends, seasonals, and cycles.  Although we will not do

so, it's possible to work with formal unobserved components models, such as

$$y_t = T_t + S_t + C_t + I_t,$$

where T is the trend component, S is the seasonal component, C is the cyclical component, and I

is the remainder, or "irregular," component, which is white noise.  Typically we'd assume that

each component is uncorrelated with all other components at all leads and lags.  Typical models

for the various components include:

Trend

$$T_t = \beta_0 + \beta_1 TIME_t \qquad \text{(deterministic)}$$

$$T_t = \beta_1 + T_{t-1} + \varepsilon_{1t} \qquad \text{(stochastic)}$$

Seasonal

$$S_t = \sum_{i=1}^{s} \gamma_i D_{it} \qquad \text{(deterministic)}$$

$$S_t = \frac{1}{1 - \gamma L^s} \varepsilon_{2t} \qquad \text{(stochastic)}$$

Cycle

$$C_t = \frac{1}{(1-\alpha_1 L)} \varepsilon_{3t} \qquad \text{(AR(1))}$$

$$C_t = \frac{1 + \beta_1 L + \beta_2 L^2}{(1-\alpha_1 L)(1-\alpha_2 L)} \varepsilon_{3t} \qquad \text{(ARMA(2,2))}$$

Irregular

$$I_t = \varepsilon_{4t}.$$

8. (The restrictions associated with unobserved-components structures)  The restrictions associated with formal unobserved-components models are surely false, in the sense that real-world dynamics are not likely to be decomposable in such a sharp and tidy way.  Rather, the decomposition is effectively an accounting framework that we use simply because it's helpful to

do so.  Trend, seasonal and cyclical variation are so different -- and so important in business,

economic and financial series -- that it's often helpful to model them separately to help ensure that

we model each adequately.  A consensus has not yet emerged as to whether it's more effective to

exploit the unobserved components perspective for intuitive motivation, as we do throughout this

book, or to enforce formal unobserved components decompositions in hopes of benefitting from

considerations related to the shrinkage principle.

9.  (Additive and multiplicative unobserved-components decompositions)  We introduced the

formal unobserved components decomposition,

$$y_t = T_t + S_t + C_t + I_t,$$

where T is the trend component, S is the seasonal component, C is the cyclical component, and I

is the remainder, or "irregular," component.  Alternatively, we could have introduced a

*multiplicative* decomposition,

$$y_t = T_t \ S_t \ C_t \ I_t.$$

      a.  Begin with the multiplicative decomposition and take logs.  How does your result

            relate to our original additive decomposition?

* Remarks, suggestions, hints, solutions:  Relationships multiplicative in levels are additive in

logs.

      b.  Does the exponential (log-linear) trend fit more naturally in the additive or

            multiplicative decomposition framework?  Why?

* Remarks, suggestions, hints, solutions:  The log-linear trend is additive in logs; hence it fits

more naturally in the multiplicative framework.

10. (Signal, noise and overfitting) Using our unobserved-components perspective, we've discussed trends, seasonals, cycles, and noise. We've modeled and forecasted each, with the exception of noise. Clearly we *can't* model or forecast the noise; by construction, it's unforecastable. Instead, the noise is what *remains* after accounting for the other components. We call the other components signals, and the signals are buried in noise. Good models fit signals, not noise. Data mining expeditions, in contrast, lead to models that often fit very well over the historical sample, but that fail miserably for out-of-sample forecasting. That's because such data mining effectively tailors the model to fit the idiosyncracies of the in-sample noise, which improves the in-sample fit but is of no help in out-of-sample forecasting.

    a. Choose your favorite trending (but not seasonal) series, and select a sample path of
       length 100. Graph it.

* Remarks, suggestions, hints, solutions: The series selected should have a visually obvious trend.

    b. Regress the first twenty observations on a fifth-order polynomial time trend, and allow
       for five autoregressive lags as well. Graph the actual and fitted values from the
       regression. Discuss.

* Remarks, suggestions, hints, solutions: Numerical instabilities may be encountered when fitting the model. Assuming that it is estimated successfully, it will likely fit very well, because of the high-ordered trend and high-ordered autoregressive dynamics.

    c. Use your estimated model to produce an 80-step-ahead extrapolation forecast.
       Graphically compare your forecast to the actual realization. Discuss.

* Remarks, suggestions, hints, solutions:  The forecast will likely be very poor.  The data were overfitted, the telltale sign of which is good in-sample fit and poor out-of-sample forecast performance.

## Chapter 11 Problems and Complements

1. (Econometrics, time series analysis, and forecasting)  As recently as the early 1970s, time series analysis was mostly univariate and made little use of economic theory.  Econometrics, in contrast, stressed the cross-variable dynamics associated with economic theory, with equations estimated using multiple regression.  Econometrics, moreover, made use of simultaneous systems of such equations, requiring complicated estimation methods.  Thus the econometric and time series approaches to forecasting were very different.[5]

As Klein (1981) notes, however, the complicated econometric system estimation methods had little payoff for practical forecasting and were therefore largely abandoned, whereas the rational distributed lag patterns associated with time-series models led to large improvements in practical forecast accuracy.[6]  Thus, in more recent times, the distinction between econometrics and time series analysis has largely vanished, with the union incorporating the best of both.  In many respects the VAR is a modern embodiment of both econometric and time-series traditions.  VARs use economic considerations to determine which variables to include and which (if any) restrictions should be imposed, allow for rich multivariate dynamics, typically require only simple estimation techniques, and are explicit forecasting models.

---

[5] Klein and Young (1980) and Klein (1983) provide good discussions of the traditional econometric simultaneous equations paradigm, as well as the link between structural simultaneous equations models and reduced-form time series models.  Wallis (1995) provides a good summary of modern large-scale macroeconometric modeling and forecasting, and Pagan and Robertson (2002) provide an intriguing discussion of the variety of macroeconomic forecasting approaches currently employed in central banks around the world.

[6] For an acerbic assessment circa the mid-1970s, see Jenkins (1979).

* Remarks, suggestions, hints, solutions: I included this complement to address the common question as to the distinction between the "time series" and "econometric" approaches to forecasting. My view is that, although historically there certainly was a distinction, it has now largely vanished as the two literatures have intermingled and complemented each other.

2. (Forecasting crop yields) Consider the following dilemma in agricultural crop yield forecasting:

> The possibility of forecasting crop yields several years in advance would, of course, be of great value in the planning of agricultural production. However, the success of long-range crop forecasts is contingent not only on our knowledge of the weather factors determining yield, but also on our ability to predict the weather. Despite an abundant literature in this field, no firm basis for reliable long-range weather forecasts has yet been found. (Sanderson, 1953, p. 3)

a. How is the situation related to our concerns in this chapter, and specifically, to the issue of conditional vs. unconditional forecasting?

* Remarks, suggestions, hints, solutions: The situation described is one in which reliable conditional forecasting is relatively easy, but reliable unconditional forecasting is notoriously difficult.

b. What variables other than weather might be useful for predicting crop yield?

* Remarks, suggestions, hints, solutions: Knowledge of movements in fertilizer use, irrigation, harvesting technology, etc.

c. How would you suggest that the forecaster should proceed?

* Remarks, suggestions, hints, solutions: Decide for yourself!

3. (Regression forecasting models with expectations, or anticipatory, data) A number of surveys exist of anticipated market conditions, investment intentions, buying plans, advance commitments,

consumer sentiment, and so on.

  a. Search the World Wide Web for such series and report your results.  A good place to

     start is the Resources for Economists page mentioned in Chapter 1.

  b. How might you use the series you found in an unconditional regression forecasting

     model of GDP?  Are the implicit forecast horizons known for all the anticipatory

     series you found?  If not, how might you decide how to lag them in your

     regression forecasting model?

* Remarks, suggestions, hints, solutions:  Try forecasting models with GDP regressed on lagged

values of the anticipatory variable.  Sometimes the implicit forecast horizon of the anticipatory

variable is unknown, in which case some experimentation may help determine the best lag

structure.

  c. How would you test whether the anticipatory series you found provide incremental

     forecast enhancement, relative to the own past history of GDP?

* Remarks, suggestions, hints, solutions:  Do a Granger causality test.

4.  (Business cycle analysis and forecasting:  expansions, contractions, turning points, and leading

indicators[7])  The use of anticipatory data is linked to business cycle analysis in general, and

leading indicators in particular.  During the first half of this century, much research was devoted

to obtaining an empirical characterization of the business cycle.  The most prominent example of

this work was Burns and Mitchell (1946), whose summary empirical definition was:

  Business cycles are a type of fluctuation found in the aggregate economic activity of

  nations that organize their work mainly in business enterprises:  a cycle consists of

---

[7] This complement draws in part upon Diebold and Rudebusch (1996).

expansions occurring at about the same time in many economic activities, followed by

similarly general recessions, contractions, and revivals which merge into the expansion

phase of the next cycle. (p. 3)

The comovement among individual economic variables was a key feature of Burns and Mitchell's

definition of business cycles. Indeed, the comovement among series, taking into account possible

leads and lags in timing, was the centerpiece of Burns and Mitchell's methodology. In their

analysis, Burns and Mitchell considered the historical concordance of hundreds of series, including

those measuring commodity output, income, prices, interest rates, banking transactions, and

transportation services, and they classified series as leading, lagging or coincident. One way to

define a leading indicator is to say that a series x is a leading indicator for a series y if x causes y in

the predictive sense. According to that definition, for example, our analysis of housing starts and

completions indicates that starts are a leading indicator for completions.

Leading indicators have the potential to be used in forecasting equations in the same way

as anticipatory variables. Inclusion of a leading indicator, appropriately lagged, can improve

forecasts. Zellner and Hong (1989) and Zellner, Hong and Min (1991), for example, make good

use of that idea in their ARLI (autoregressive leading-indicator) models for forecasting aggregate

output growth. In those models, Zellner *et al*. build forecasting models by regressing output on

lagged output and lagged leading indicators; they also use shrinkage techniques to coax the

forecasted growth rates toward the international average, which improves forecast performance.

Burns and Mitchell used the clusters of turning points in individual series to determine the

monthly dates of the turning points in the overall business cycle, and to construct composite

indexes of leading, coincident, and lagging indicators. Such indexes have been produced by the

National Bureau of Economic Research (a think tank in Cambridge, Mass.), the Department of

Commerce (a U.S. government agency in Washington, DC), and the Conference Board (a

business membership organization based in New York).[8]  Composite indexes of leading indicators

are often used to gauge likely future economic developments, but their usefulness is by no means

uncontroversial and remains the subject of ongoing research.  For example, leading indexes

apparently cause aggregate output in analyses of ex post historical data (Auerbach, 1982), but

they appear much less useful in real-time forecasting, which is what's relevant (Diebold and

Rudebusch, 1991).

* Remarks, suggestions, hints, solutions:  For more on business cycle analysis and its relation to

forecasting, see the 1999 Diebold and Rudebusch book.

5.  (Subjective information, Bayesian VARs, and the Minnesota prior)  When building and using

forecasting models, we frequently have hard-to-quantify subjective information, such as a

reasonable range in which we expect a parameter to be.  We can incorporate such subjective

information in a number of ways.  One way is informal judgmental adjustment of estimates.

Based on a variety of factors, for example, we might feel that an estimate of a certain parameter in

a forecasting model is too high, so we might reduce it a bit.

     Bayesian analysis  allows us to incorporate subjective information in a rigorous and

replicable way.  We summarize subjective information about parameters with a probability

distribution called the prior distribution, and as always we summarize the information in the data

with the likelihood function.  The centerpiece of Bayesian analysis is a mathematical formula

---

[8] The indexes build on very early work, such as the Harvard "Index of General Business
Conditions."  For a fascinating discussion of the early work, see Hardy (1923), Chapter 7.

called Bayes' rule, which tells us how to combine the information in the prior and the likelihood to form the posterior distribution of model parameters, which then feed their way into forecasts.

The Minnesota prior (introduced and popularized by Robert Litterman and Christopher Sims at the University of Minnesota) is commonly used for Bayesian estimation of VAR forecasting models, called Bayesian VARs, or BVARs. The Minnesota prior is centered on a parameterization called a random walk, in which the current value of each variable is equal to its lagged value plus a white noise error term. Thus the parameter estimates in BVARs are coaxed, but not forced, in the direction of univariate random walks. This sort of stochastic restriction has an immediate shrinkage interpretation, which suggests that it's likely to improve forecast accuracy.[9] This hunch is verified in Doan, Litterman and Sims (1984), who study forecasting with standard and Bayesian VARs. Ingram and Whiteman (1994) replace the Minnesota prior with a prior derived from macroeconomic theory, and they obtain even better forecasting performance.

6. (Housing starts and completions, continued) Our VAR analysis of housing starts and completions, as always, involved many judgement calls. Using the starts and completions data, assess the adequacy of our models and forecasts. Among other things, you may want to consider the following questions:

a. Should we allow for a trend in the forecasting model?

* Remarks, suggestions, hints, solutions: There is probably no need. The graph indicates, however, that there may be a very slight downward trend, and some authors have argued that recent demographic shifts have produced such a trend. If interest centers on very long-term forecasting, additional analysis may be fruitful; otherwise, trend can probably be safely ignored.

---

[9] Effectively, the shrinkage allows us to recover a large number of degrees of freedom.

b. How do the results change if, in light of the results of the causality tests, we exclude

lags of completions from the starts equation, re-estimate by seemingly-unrelated

regression, and forecast?

* Remarks, suggestions, hints, solutions:  See for yourself!

c. Are the VAR forecasts of starts and completions more accurate than univariate

forecasts?

* Remarks, suggestions, hints, solutions:  See for yourself!

7. (Nonlinear regression models I:  functional form and Ramsey's test)  The idea of using powers

of a right-hand-side variable to pick up nonlinearity in a regression can also be used to test for

linearity of functional form, following Ramsey (1969).  If we were concerned that we'd missed

some important nonlinearity, an obvious strategy to capture it, based on the idea of a Taylor series

expansion of a function, would be to include powers and cross products of the various x variables

in the regression.  Such a strategy would be wasteful of degrees of freedom, however, particularly

if there were more than just one or two right-hand-side variables in the regression and/or if the

nonlinearity were severe, so that fairly high powers and interactions would be necessary to

capture it.  In light of this, Ramsey suggests first fitting a linear regression and obtaining the fitted

values, $\hat{y}_t$, t = 1, ..., T.  Then, to test for nonlinearity, we run the regression again with powers of

$\hat{y}_t$ included.  There is no need to include the first power of $\hat{y}_t$, because that would be redundant

with the included x variables.  Instead we include powers $\hat{y}_t^2$, $\hat{y}_t^3$, ..., $\hat{y}_t^m$, where m is a maximum

power determined in advance.  Note that the powers of $\hat{y}_t$ are linear combinations of powers and

cross products of the x variables -- just what the doctor ordered.  Significance of the included set

of powers of $\hat{y}_t$ can be checked using an F test or an asymptotic likelihood ratio test.

* Remarks, suggestions, hints, solutions: It's useful for the students to know about Ramsey's test, so I have included it, but it is worth pointing out that the old strategy of including powers of the x variables still has its place.

8. (Nonlinear regression models II: logarithmic regression models) We've already seen the use of logarithms in our studies of trend and seasonality. In those setups, however, we had occasion only to take logs of the left-hand-side variable. In more general regression models, such as those that we're studying now, with variables other than trend or seasonals on the right-hand side, it's sometimes useful to take logs of *both* the left- and right-hand-side variables. Doing so allows us to pick up multiplicative nonlinearity. To see this, consider the regression model,

$$y_t = \beta_0 \, x_t^{\beta_1} \, e^{\varepsilon_t}.$$

This model is clearly nonlinear due to the multiplicative interactions. Direct estimation of its parameters would require special techniques. Taking natural logs, however, yields the model

$$\ln y_t = \ln \beta_0 + \beta_1 \ln x_t + \varepsilon_t.$$

This transformed model can be immediately estimated by ordinary least squares, by regressing log y on an intercept and log x. Such "log-log regressions" often capture nonlinearities relevant for forecasting, while maintaining the convenience of ordinary least squares.

* Remarks, suggestions, hints, solutions: Students often seem mystified by the log-log regression. This complement tries to make clear that it is simply one way to allow for nonlinearity, which may

or may not be appropriate, depending on the specifics of the problem at hand.  Multiplicative

nonlinearities are naturally handled by modeling in logs.

9. (Nonlinear regression models III:  neural networks)  Neural networks amount to a particular

nonlinear functional form associated with repeatedly running linear combinations of inputs

through nonlinear "squashing" functions.  The 0-1 squashing function is useful in classification,

and the logistic function is useful for regression.

The neural net literature is full of biological jargon, which serves to obfuscate rather than

clarify.  We speak, for example, of a "single-output feedforward neural network with n inputs and

1 hidden layer with q neurons."  But the idea is simple.  If the output is y and the inputs are x's,

we write

$$y_t = \Phi(\beta_0 + \sum_{i=1}^{q} \beta_i h_{it}),$$

where

$$h_{it} = \Psi(\gamma_{i0} + \sum_{j=1}^{n} \gamma_{ij} x_{jt}), \ i = 1, \ ..., \ q$$

are the "neurons" ("hidden units"), and the "activation functions" $\Psi$ and $\Phi$ are arbitrary, except

that $\Psi$ (the squashing function) is generally restricted to be bounded.  (Commonly $\Phi(x)=x$.)

Assembling it all, we write

$$y_t = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i \Psi\left(\gamma_{i0} + \sum_{j=1}^{n} \gamma_{ij} x_{jt}\right)\right) = f(x_t; \ \theta),$$

which makes clear that a neural net is just a particular nonlinear functional form for a regression

model.

To incorporate dynamics, we can allow for autoregressive effects in the hidden units.  A dynamic ("recurrent") neural network is

$$y_t = \Phi(\beta_0 + \sum_{i=1}^{q} \beta_i h_{it}),$$

where

$$h_{it} = \Psi(\gamma_{i0} + \sum_{j=1}^{n} \gamma_{ij} x_{jt} + \sum_{r=1}^{q} \delta_{ir} h_{r,t-1}), \ i = 1, \ ..., \ q.$$

Compactly,

$$y_t = \Phi\left( \beta_0 + \sum_{i=1}^{q} \beta_i \Psi\left( \gamma_{i0} + \sum_{j=1}^{n} \gamma_{ij} x_{jt} + \sum_{r=1}^{q} \delta_{ir} h_{r,t-1} \right) \right).$$

Recursive back substitution reveals that y is a nonlinear function of the history of the x's.

$$y_t = g(x^t; \theta),$$

where $x^t = (x_t, ..., x_1)$ and $x_t = (x_{1t}, ..., x_{nt})$.

The Matlab Neural Network Toolbox implements a variety of networks.  The toolbox manual is itself a useful guide to the literature on the practical aspects of constructing and forecasting with neural nets.  Kuan and Liu (1995) use a dynamic neural network to predict foreign exchange rates, and Faraway and Chatfield (1995) provide an insightful case study of the efficacy of neural networks in applied forecasting.  Ripley (1996) provides a fine and statistically-informed (in contrast to much of the neural net literature) survey of the use of neural nets in a variety of fields.

* Remarks, suggestions, hints, solutions:  Neural nets are interesting and topical; hence this

complement. It's important, however, to make the students aware of both the use and abuse of neural nets.

10. (Spurious regression) Consider two variables y and x, both of which are highly serially correlated, as are most series in business, finance and economics. Suppose in addition that y and x are completely unrelated, but that we don't know they're unrelated, and we regress y on x using ordinary least squares.

a. If the usual regression diagnostics (e.g., $R^2$, t-statistics, F-statistic) were reliable, we'd expect to see small values of all of them. Why?

* Remarks, suggestions, hints, solutions: Because there is in fact no relationship between y and x.

b. In fact the opposite occurs; we tend to see large $R^2$, t-, and F-statistics, and *a very low Durbin-Watson statistic*. Why the low Durbin-Watson? Why, given the low Durbin-Watson, might you *expect* misleading $R^2$, t-, and F-statistics?

* Remarks, suggestions, hints, solutions: The DW is low because y is highly serially correlated, and not explained by x, which means that the residual is highly serially correlated. The residual serial correlation wreaks havoc with the t and F statistics, and hence with $R^2$, which is a simple function of the F statistic.

c. This situation, in which highly persistent series that are in fact unrelated nevertheless appear highly related, is called spurious regression. Study of the phenomenon dates to the early twentieth century, and a key study by Granger and Newbold (1974) drove home the prevalence and potential severity of the problem. How might you insure yourself against the spurious regression problem? (Hint: Consider allowing for lagged dependent variables, or dynamics in the regression

disturbances, as we've advocated repeatedly.)

* Remarks, suggestions, hints, solutions: The answer is given by the hint. The key is to incorporate some way of modeling the dynamics in y in the event that x doesn't explain any or all of them.

11. (Comparative forecasting performance of VAR and univariate models) Using the housing starts and completions data on the book's website, compare the forecasting performance of the VAR used in this chapter to that of the obvious competitor: univariate autoregressions. Use the same in-sample and out-of-sample periods as in the chapter. Why might the forecasting performance of the VAR and univariate methods differ? Why might you expect the VAR completions forecast to outperform the univariate autoregression, but the VAR starts forecast to be no better than the univariate autoregression? Do your results support your conjectures?

* Remarks, suggestions, hints, solutions: The VAR completions forecast benefits from conditioning on the history of starts, whereas the univariate completions forecast does not.

## Chapter 12 Problems and Complements

1. (Forecast evaluation in action)  Discuss in detail how you would use forecast evaluation techniques to address each of the following questions.

  a.  Are asset returns (e.g., stocks, bonds, exchange rates) forecastable over long horizons?

* Remarks, suggestions, hints, solutions:  If sufficient data are available, one could perform a recursive long-horizon forecasting exercise (using, for example, an autoregressive model), and compare the real-time forecasting performance to that of a random walk.

  b.  Do forward exchange rates provide unbiased forecasts of future spot exchange rates at all horizons?

* Remarks, suggestions, hints, solutions:  Check whether the forecast error, defined as the realized spot rate minus the appropriately lagged forward rate, has zero mean.

  c.  Are government budget projections systematically too optimistic, perhaps for strategic reasons?

* Remarks, suggestions, hints, solutions:  If revenue is being forecast, optimism corresponds to revenue forecasts that are too high on average, or forecast errors (actual minus forecast) that are negative on average.

  d.  Can interest rates be used to provide good forecasts of future inflation?

* Remarks, suggestions, hints, solutions:  One could examine forecasting models that project inflation on lagged interest rates, but for the reasons discussed in the text it's preferable to begin with a simple inflation autoregression, and then to ask whether including lagged interest rates provides incremental predictive enhancement.

2. (Forecast error analysis) You are working for a London-based hedge fund, Thompson Energy Investors, and your boss has assigned you to assess a model used to forecast U.S. crude oil imports. On the last day of each quarter, the model is used to forecast oil imports at horizons of 1-quarter-ahead through 4-quarters-ahead. Thompson has done this for each of 80 quarters and has kept the corresponding four forecast error series, which appear on the book's web page.

    a. Based on a correlogram analysis, assess whether the 1-quarter-ahead forecast errors are white noise. (Be sure to discuss all parts of the correlogram: sample autocorrelations, sample partial autocorrelations, Bartlett standard errors and Ljung-Box statistics.) Why care?

    b. Regress each of the four forecast error series on constants, in each case allowing for a MA(5) disturbances. Comment on the significance of the MA coefficients in each of the four cases and use the results to assess the optimality of the forecasts at each of the four horizons. Does your 1-step-ahead MA(5)-based assessment match the correlogram-based assessment obtained in part a? Do the multi-step forecasts appear optimal?

    c. Overall, what do your results suggest about the model's ability to predict U.S. crude oil imports?

3. (Combining Forecasts) You are a managing director at Paramex, a boutique investment bank in Paris. Each day during the summer your two interns, Alex and Betsy, give you a 1-day-ahead forecast of the Euro/Dollar exchange rate. At the end of the summer, you calculate each intern's series of daily forecast errors. You find that the mean errors are zero, and the error variances and covariances are $\hat{\sigma}^2_{AA} = 153.76$, $\hat{\sigma}^2_{BB} = 92.16$ and $\hat{\sigma}^2_{AB} = .2$.

    a. If you were forced to choose between Alex's forecast and Betsy's forecast, which

        would you choose? Why?

* Remarks, suggestions, hints, solutions: Betsy's (much lower error variance).

    b. If instead you had the opportunity to combine the two forecasts by forming a weighted

        average, what would be the optimal weights according to the variance-covariance

        method? Why?

* Remarks, suggestions, hints, solutions: Just plug into the formula in the text.

    c. Is it guaranteed that a combined forecast formed using the "optimal" weights calculated

        in part b will have lower mean squared prediction error? Why or why not?

* Remarks, suggestions, hints, solutions: Improvement is guaranteed in theory (i.e., when using

population variances and covariances) but not in practice (due to variance and covariance

estimation error). Even in practice, however, combination often produces improvements.

4. (Quantitative forecasting, judgmental forecasting, forecast combination, and shrinkage)

Interpretation of the modern quantitative approach to forecasting as eschewing judgement is most

definitely misguided. How is judgement used routinely and informally to modify quantitative

forecasts? How can judgement be formally used to modify quantitative forecasts via forecast

combination? How can judgement be formally used to modify quantitative forecasts via

shrinkage? Discuss the comparative merits of each approach. Klein (1981) provides insightful

discussion of the interaction between judgement and models, as well as the comparative track

record of judgmental vs. model-based forecasts.

* Remarks, suggestions, hints, solutions: Judgement is used throughout the modeling and

forecasting process. It is used informally to modify quantitative forecasts when, for example, the

quantitative forecast is used as the input to a committee meeting, the output of which is the final

forecast. Judgement can be formally used to modify quantitative forecasts via forecast

combination, when, for example, an "expert opinion" is combined with a model-based forecast.

Finally, shrinkage often implicitly amounts to judgmental adjustment, because it amounts to

coaxing results into accordance with prior views.

5. (The algebra of forecast combination) Consider the combined forecast,

$$y^{c}_{t+h,t} = \omega y^{a}_{t+h,t} + (1-\omega)y^{b}_{t+h,t}.$$

Verify the following claims made in the text:

a. The combined forecast error will satisfy the same relation as the combined forecast;

that is,

$$e^{c}_{t+h,t} = \omega e^{a}_{t+h,t} + (1-\omega)e^{b}_{t+h,t}.$$

b. Because the weights sum to unity, if the primary forecasts are unbiased then so too is

the combined forecast.

* Remarks, suggestions, hints, solutions: A linear combination (with weights summing to unity)

of zero-mean forecast errors also has zero mean. Thus the combined forecast error has zero

mean, which is to say that the combined forecast is unbiased.

c. The variance of the combined forecast error is

$$\sigma^{2}_{c} = \omega^{2}\sigma^{2}_{aa} + (1-\omega)^{2}\sigma^{2}_{bb} + 2\omega(1-\omega)\sigma^{2}_{ab},$$

where $\sigma^{2}_{11}$ and $\sigma^{2}_{22}$ are unconditional forecast error variances and $\sigma^{2}_{12}$ is their

covariance.

* Remarks, suggestions, hints, solutions: Begin with the expression for the combined forecast

error, and take the variance of each side. Recall that the variance of ax+by is

$a^2$var(x)+$b^2$var(y)+2ab cov (x,y).

    d. The combining weight that minimizes the combined forecast error variance (and hence

        the combined forecast error MSE, by unbiasedness) is

$$\omega^* = \frac{\sigma^2_{bb} - \sigma^2_{ab}}{\sigma^2_{bb} + \sigma^2_{aa} - 2\sigma^2_{ab}}.$$

* Remarks, suggestions, hints, solutions: We need to find the value of the combining weight that

minimizes the variance of the combined forecast error. We find the minimum by differentiating

the expression for the variance of the combined forecast error with respect to the combining

weight, setting it equal to zero to form the first-order condition, and solving the first-order

condition for the combining weight. (A check of the second-order condition reveals that the

critical point obtained is in fact a minimum.)

    e. If neither forecast encompasses the other, then

$$\sigma^2_c < \min(\sigma^2_{aa}, \sigma^2_{bb}).$$

* Remarks, suggestions, hints, solutions: The combined forecast error can't be *bigger* than

$\sigma^2_c < \min(\sigma^2_{aa}, \sigma^2_{bb})$, because we could always put a weight of 1 on one forecast and 0 on the

other, in which case $\sigma^2_c = \min(\sigma^2_{aa}, \sigma^2_{bb})$. If any other weight is chosen, it must produce

$\sigma^2_c < \min(\sigma^2_{aa}, \sigma^2_{bb})$.

    f. If one forecast encompasses the other, then

$$\sigma_c^2 = \min(\sigma_{aa}^2, \sigma_{bb}^2).$$

* Remarks, suggestions, hints, solutions:  If one forecast encompasses the other, then we *do* put a weight of 1 on one forecast and 0 on the other.

6.  (The mechanics of practical forecast evaluation and combination)  On the book's web page you'll find the time series of shipping volume, quantitative forecasts, and judgmental forecasts used in this chapter.

> a.  Replicate the empirical results reported in this chapter.  Explore and discuss any variations or extensions that you find interesting.

> b.  Using the first 250 weeks of shipping volume data, specify and estimate a univariate autoregressive model of shipping volume (with trend and seasonality if necessary), and provide evidence to support the adequacy of your chosen specification.

> c.  Use your model each week to forecast two weeks ahead, each week estimating the model using all available data, producing forecasts for observations 252 through 499, made using information available at times 250 through 497.  Calculate the corresponding series of 248 2-step-ahead recursive forecast errors.

> d.  Using the methods of this chapter, evaluate the quality of your forecasts, both in isolation and relative to the original quantitative and judgmental forecasts. Discuss.

> e.  Using the methods of this chapter, assess whether your forecasting model can usefully be combined with the original quantitative and judgmental models.  Discuss.

* Remarks, suggestions, hints, solutions:  This problem provides valuable training in the

mechanics of forecast evaluation and combination, as well as the recursive estimation and forecasting methods introduced in Chapter 9, beginning with simple replication of the results in the text, and proceeding into uncharted territory.

7. (What are we forecasting?  Preliminary series, revised series, and the limits to forecast accuracy)  Many economic series are revised as underlying source data increase in quantity and quality.  For example, a typical quarterly series might be issued as follows.  First, shortly after the end of the relevant quarter, a "preliminary" value for the current quarter is issued.  A few months later, a "revised" value is issued, and a year or so later the "final revised" value is issued.  For extensive discussion, see Croushore and Stark (2001).

    a. If you're evaluating the accuracy of a forecast or forecasting technique, you've got to decide on what to use for the "actual" values, or realizations, to which the forecasts will be compared.  Should you use the preliminary value?  The final revised value?  Something else?  Be sure to weigh as many relevant issues as possible in defending your answer.

* Remarks, suggestions, hints, solutions:  My view is that, other things the same, we're trying to forecast the truth, not some preliminary estimate of the truth, so it makes sense to use the final revised version.  Occasionally, however, data undergo revisions so massive (due to redefinitions, etc.) that it may be appropriate to use a preliminary release instead.

    b. Morgenstern (1963) assesses the accuracy of economic data and reports that the great mathematician Norbert Wiener, after reading an early version of Morgenstern's book, remarked that "economics is a one or two digit science."  What might Wiener have meant?

* Remarks, suggestions, hints, solutions: There is a great deal of measurement error in economic

statistics. Even our "final revised values" are just estimates, and often poor estimates. Hence it

makes no sense to report, say, the unemployment rate out to four decimal places.

    c. Theil (1966) is well aware of the measurement error in economic data; he speaks of

        "predicting the future and *estimating* the past." Klein (1981) notes that, in

        addition to the usual innovation uncertainty, measurement error in economic data -

        - even "final revised" data -- provides additional limits to measured forecast

        accuracy. That is, even if a forecast were perfect, so that forecast errors were

        consistently zero, *measured* forecast errors would be nonzero due to measurement

        error. The larger the measurement error, the more severe the inflation of measured

        forecast error. Evaluate.

* Remarks, suggestions, hints, solutions: It's true. Measurement error in economic data places

bounds on attainable forecast accuracy.

    d. When assessing improvements (or lack thereof) in forecast accuracy over time, how

        might you guard against the possibility of spurious assessed improvements due not

        to true forecast improvement, but rather to structural change toward a more

        "forecastable" process? (On forecastability, see Diebold and Kilian, 2001).

* Remarks, suggestions, hints, solutions: One possibility is not to assess the evolution of

accuracy directly, but rather to assess the evolution of accuracy relative to a benchmark, such as a

random walk.

8. (Ex post vs. real-time forecast evaluation) If you're evaluating a forecasting model, you've

also got to take a stand on precisely what information is available to the forecaster, and when.

Suppose, for example, that you're evaluating the forecasting accuracy of a particular regression model.

    a. Do you prefer to estimate and forecast recursively, or simply estimate once using the full sample of data?

    b. Do you prefer to estimate using final-revised values of the left- and right-hand side variables, or do you prefer to use the preliminary, revised and final-revised data as it became available in real time?

    c. If the model is explanatory rather than causal, do you prefer to substitute the true realized values of right-hand side variables, or to substitute forecasts of the right-hand side variables that could actually be constructed in real time?

* Remarks, suggestions, hints, solutions: Each of the sub-questions gets at an often-neglected issue in forecast evaluation. The most credible (and difficult) evaluation would proceed recursively using only that data available in real time (including forecasts rather than realized values of the right-hand-side variables).

These sorts of timing issues can make large differences in conclusions. For an application to using the composite index of leading indicators to forecast industrial production, see Diebold and Rudebusch (1991).

9. (What do we know about the accuracy of macroeconomic forecasts?) Zarnowitz and Braun (1993) provide a fine assessment of the track record of economic forecasts since the late 1960s. Read their paper and try to assess just what we really know about:

    a. comparative forecast accuracy at business cycle turning points vs. other times

* Remarks, suggestions, hints, solutions: Turning points are especially difficult to predict.

b. comparative accuracy of judgmental vs. model-based forecasts

* Remarks, suggestions, hints, solutions: It's hard to make a broad assessment of this issue.

c. improvements in forecast accuracy over time

* Remarks, suggestions, hints, solutions: It's hard to make a broad assessment of this issue.

d. the comparative forecastability of various series

* Remarks, suggestions, hints, solutions: Some series (e.g., consumption) are much easier to predict than others (e.g., inventory investment).

e. the comparative accuracy of linear vs. nonlinear forecasting models.

* Remarks, suggestions, hints, solutions: See Stock and Watson (1999).

Other well-known and useful comparative assessments of U.S. macroeconomic forecasts have been published over the years by Stephen K. McNees, a private consultant formerly with the Federal Reserve Bank of Boston. McNees (1988) is a good example. Similarly useful studies for the U.K.. with particular attention to decomposing forecast error into its various possible sources, have recently been produced by Kenneth F. Wallis and his coworkers at the ESRC Macroeconomic Modelling Bureau at the University of Warwick. Wallis and Whitley (1991) is a good example. Finally, the Model Comparison Seminar, founded by Lawrence R. Klein of the University of Pennsylvania and now led by Michael Donihue of Colby College, is dedicated to the ongoing comparative assessment of macroeconomic forecasting models. Klein (1991) provides a good survey of some of the group's recent work, and more recent information can be found on the web at http://www.colby.edu/economics/faculty/mrdonihu/mcs/.

10. (Forecast evaluation when realizations are unobserved) Sometimes we never see the realization of the variable being forecast. Pesaran and Samiei (1995), for example, develop

models for forecasting ultimate resource recovery, such as the total amount of oil in an underground reserve. The actual value, however, won't be known until the reserve is depleted, which may be decades away. Such situations obviously make for difficult accuracy evaluation! How would you evaluate such forecasting models?

* Remarks, suggestions, hints, solutions: Most forecast evaluation techniques naturally proceed by examining the forecast errors, or some other function of the actual and forecast values. Because that's not possible in the environment under consideration, one would evidently have to rely on assessing the theoretical underpinnings of the forecasting model used and compare then with those of alternative models (if any).

11. (Forecast error variances in models with estimated parameters) As we've seen, computing forecast error variances that acknowledge parameter estimation uncertainty is very difficult; that's one reason why we've ignored it. We've learned a number of lessons about optimal forecasts while ignoring parameter estimation uncertainty, such as:

    a. Forecast error variance grows as the forecast horizon lengthens.

    b. In covariance stationary environments, the forecast error variance approaches the

        (finite) unconditional variance as the horizon grows.

Such lessons provide valuable insight and intuition regarding the workings of forecasting models and provide a useful benchmark for assessing actual forecasts. They sometimes need modification, however, when parameter estimation uncertainty is acknowledged. For example, in models with estimated parameters:

    a. Forecast error variance needn't grow monotonically with horizon. Typically we *expect*

        forecast error variance to increase monotonically with horizon, but it doesn't *have*

to.

  b. Even in covariance stationary environments, the forecast error variance needn't

    converge to the unconditional variance as the forecast horizon lengthens; instead,

    it may grow without bound.  Consider, for example, forecasting a series that's just

    a stationary AR(1) process around a linear trend.  With known parameters, the

    point forecast will converge to the trend as the horizon grows, and the forecast

    error variance will converge to the unconditional variance of the AR(1) process.

    With estimated parameters, however, if the estimated trend parameters are even

    the slightest bit different from the true values (as they almost surely will be, due to

    sampling variation), that error will be magnified as the horizon grows, so the

    forecast error variance will grow.

Thus, results derived under the assumption of known parameters should be viewed as a

benchmark to guide our intuition, rather than as precise rules.

* Remarks, suggestions, hints, solutions:  Use this complement to warn the students that the

population results used as a benchmark are just that -- a benchmark, and nothing more -- and may

be violated in realistic conditions.

12.  (The empirical success of forecast combination)  In the text we mentioned that we have

nothing to lose by forecast combination, and potentially much to gain.  That's certainly true in

population, with optimal combining weights.  However, in finite samples of the size typically

available, sampling error contaminates the combining weight estimates, and the problem of

sampling error may be exacerbated by the collinearity that typically exists between

$y_{t+h,t}^{a}$ and $y_{t+h,t}^{b}$.  Thus, while we hope to reduce out-of-sample forecast MSE by combining,

there is no guarantee.  Fortunately, however, in practice forecast combination often leads to very good results.  The efficacy of forecast combination is well-documented in Clemen's (1989) review of the vast literature, and it emerges clearly in the landmark study by Stock and Watson (1999).

* Remarks, suggestions, hints, solutions:  Students seem to appreciate the analogy between forecasting combination and portfolio diversification.  Forecast combination essentially amounts to holding a portfolio of forecasts, and just as with financial assets, the performance of the portfolio is superior to that of any individual component.

13.  (Forecast combination and the Box-Jenkins paradigm)  In an influential book, Box and Jenkins (latest edition, Box, Jenkins and Reinsel, 1994) envision an ongoing, iterative process of model selection and estimation, forecasting, and forecast evaluation.  What is the role of forecast combination in that paradigm?  In a world in which information sets can be instantaneously and costlessly combined, there is no role; it is always optimal to combine information sets rather than forecasts.  That is, if no model forecast-encompasses the others, we might hope to eventually figure out what's gone wrong, learn from our mistakes, and come up with a model based on a combined information set that *does* forecast-encompass the others.  But in the short run -- particularly when deadlines must be met and timely forecasts produced -- pooling of information sets is typically either impossible or prohibitively costly.  This simple insight motivates the pragmatic idea of forecast combination, in which forecasts rather than models are the basic object of analysis, due to an assumed inability to combine information sets.  Thus, forecast combination can be viewed as a key link between the short-run, real-time forecast production process, and the longer-run, ongoing process of model development.

* Remarks, suggestions, hints, solutions:  It is important to stress that forecast encompassing tests

complement forecast combination, by serving as a preliminary screening device. If one model

forecast-encompasses the others, then it should be used, and there's no need to proceed with

forecast combination.

14. (Consensus forecasts) A number of services, some commercial and some non-profit,

regularly survey economic and financial forecasters and publish "consensus" forecasts, typically

the mean or median of the forecasters surveyed. The consensus forecasts often perform very well

relative to the individual forecasts. The Survey of Professional Forecasters is a leading consensus

forecast that has been produced each quarter since the late 1960s; currently it's produced by the

Federal Reserve Bank of Philadelphia. See Zarnowitz and Braun (1993) and Croushore (1993).

* Remarks, suggestions, hints, solutions: Consensus point forecasts are typically reported.

Interestingly, however, the Survey of Professional Forecasters also publishes consensus density

forecasts of inflation and aggregate output, in the form of histograms. Have the students check

out the Survey of Professional Forecasters on the Federal Reserve Bank of Philadelphia's web

page.

15. (The Delphi method for combining experts' forecasts) The "Delphi method" is a structured

judgmental forecasting technique that sometimes proves useful in very difficult forecasting

situations not amenable to quantification, such as new-technology forecasting. The basic idea is

to survey a panel of experts anonymously, reveal the distribution of opinions to the experts so

they can revise their opinions, repeat the survey, and so on. Typically the diversity of opinion is

reduced as the iterations proceed.

      a. Delphi and related techniques are fraught with difficulties and pitfalls. Discuss them.

* Remarks, suggestions, hints, solutions: There is no guarantee that the iterations will converge,

and even if they do, it's not clear why we should have confidence in the final forecast.

      b.  At the same time, it's not at all clear that we should dispense with such techniques;

          they may be of real value.  Why?

* Remarks, suggestions, hints, solutions:  They at least provide a structured framework for

attempting to reach consensus on difficult matters.

## Chapter 13 Problems and Complements

1.  (Modeling and forecasting the deutschemark / dollar exchange rate)  On the book's web page you'll find monthly data on the deutschemark / dollar exchange rate for the same sample period as the yen / dollar data studied in the text.

   a.  Model and forecast the deutschemark / dollar rate, in parallel with the analysis in the text, and discuss your results in detail.

   b.  Redo your analysis using forecasting approaches without trends -- a levels model without trend, a first-differenced model without drift, and simple exponential smoothing.

   c.  Compare the forecasting ability of the approaches with and without trend.

   d.  Do you feel comfortable with the inclusion of trend in an exchange rate forecasting model?  Why or why not?

* Remarks, suggestions, hints, solutions:  The idea of this entire problem is to get students thinking about the appropriateness of trends in financial asset forecasting.  Although we included a trend in the example of Chapter 10, it's not clear why it's there.  On one hand, some authors have argued that local trends may be operative in the foreign exchange market.  On the other hand, if asset prices were really trending, then they would be highly predictable using publicly available data, which violates the efficient markets hypothesis.

2.  (Housing starts and completions, continued)  As always, our Chapter 11 VAR analysis of housing starts and completions involved many judgement calls.  Using the starts and completions data, assess the adequacy of our models and forecasts.  Among other things, you may want to

consider the following questions:

    a. How would you choose the number of augmentation lags? How sensitive are the

        results of the augmented Dickey-Fuller tests to the number of augmentation lags?

* Remarks, suggestions, hints, solutions: Our standard information criteria (AIC, SIC) may be

used, as can standard t-tests on the augmentation lag coefficients. Often Dickey-Fuller test results

will be robust for some reasonable range of augmentation lags. In the present application it's not

clear what we mean by a "reasonable range" of augmentation lags; the information criteria select a

small number, and t-testing selects a larger number. Within that range the test results do vary, but

certain considerations suggest focusing on the tests with more augmentation lags. In particular,

for a fixed sample size, a large number of augmentation lags reduces distortion of the size of the

augmented Dickey-Fuller test but also reduces power. It turns out, however, that with the

comparatively large number of lags selected by t-testing we nevertheless easily reject the unit

root hypothesis, so power seems to be good. Using fifteen augmentation lags for starts, we reject

the unit root with p<.01, and using twenty-four augmentation lags for completions we reject the

unit root with p<.05. All told, the evidence indicates that the dynamics of housing starts and

completions, although highly persistent, are nevertheless stationary, which suggests specifying the

VAR in levels. Moreover, even if housing starts and completions were I(1), specification in levels

may still be acceptable and is fine asymptotically, and the sample size at hand is quite large.

    b. When performing augmented Dickey-Fuller tests, is it adequate to allow only for an

        intercept under the alternative hypothesis, or should we allow for both intercept

        and trend?

* Remarks, suggestions, hints, solutions: It depends on whether we feel that trend is likely to be

operative under the alternative. In the case of housing starts and completions, trend seems not to play a major role, so it's probably not necessary to include trend.

     c. Should we allow for a trend in the forecasting model?

     d. Does it make sense to allow a large number of lags in the augmented Dickey-Fuller tests, but not in the actual forecasting model?

\* Remarks, suggestions, hints, solutions: Yes. Lags are included in the augmented Dickey-Fuller regression and in the forecasting model for very different reasons. In particular, the parsimony/shrinkage principle indicates that the best forecasting model may involve only a few lags, even if the "true" model involves a large (or infinite) number of lags.

     e. How do the results change if, in light of the results of the causality tests, we exclude lags of completions from the starts equation, re-estimate by seemingly-unrelated regression, and forecast?

     f. Are the VAR forecasts of starts and completions more accurate than univariate forecasts?

3. (ARIMA models, smoothers, and shrinkage) From the vantage point of the shrinkage principle, discuss the tradeoffs associated with "optimal" forecasts from fitted ARIMA models vs. "ad hoc" forecasts from smoothers.

\* Remarks, suggestions, hints, solutions: To the extent that the underlying process for which a smoother is optimal is *not* the process that generates the data, the smoother will generate suboptimal forecasts. ARIMA models, in contrast, tailor the model to the data and therefore may produce forecasts closer to the optimum. The shrinkage principle, however, suggests that imposition of the restrictions associated with smoothing may produce good forecasts, even if the

restrictions are incorrect, so long as they are not too egregiously violated.

4. (Using stochastic-trend unobserved-components models to implement smoothing techniques in a probabilistic framework) In the text we noted that smoothing techniques, as typically implemented, are used as "black boxes" to produce point forecasts. There is no attempt to exploit stochastic structure to produce interval or density forecasts in addition to point forecasts. Recall, however, that the various smoothers produce optimal forecasts for specific data-generating processes specified as unobserved-components models.

      a. For what data-generating process is exponential smoothing optimal?

* Remarks, suggestions, hints, solutions: Random walk plus white noise, which is ARIMA(0,1,1). I generally don't expect students to figure this out for themselves, but rather to come upon it through background reading (e.g., Harvey's 1989 book).

      b. For what data-generating process is Holt-Winters smoothing optimal?

* Remarks, suggestions, hints, solutions: Again, encourage the students to do the necessary background reading, in particular Harvey's 1989 book.

      c. Under the assumption that the data-generating process for which exponential smoothing produces optimal forecasts is in fact the true data-generating process, how might you estimate the unobserved-components model and use it to produce optimal interval and density forecasts? Hint: Browse through Harvey (1989).

* Remarks, suggestions, hints, solutions: One could use the methods (and specialized software) developed by Harvey. Alternatively, one could fit an ARIMA(0,1,1) model and proceed in the usual way.

      d. How would you interpret the interval and density forecasts produced by the method of

part c, if we no longer assume a particular model for the true data-generating process?

* Remarks, suggestions, hints, solutions:  Strictly speaking, their validity is contingent upon the truth of the assumed model.  If the misspecification is not too serious, however, they may nevertheless provide useful and reasonably accurate quantifications of forecast uncertainty.

5.  (Automatic ARIMA modeling)  "Automatic" forecasting software exists for implementing the ARIMA and exponential smoothing techniques of this and previous chapters without any human intervention.

   a.  What are do you think are the benefits of such software?

* Remarks, suggestions, hints, solutions:  Human judgement and emotion can sometimes be harmful rather than helpful.  Automatic forecasting software eliminates reliance on such judgement and emotion.

   b.  What do you think are the costs?

* Remarks, suggestions, hints, solutions:  Forecasting turns into a "black box" procedure, and the user may emerge as the servant rather than the master.

   c.  When do you think it would be most useful?

* Remarks, suggestions, hints, solutions:  One common situation is when a multitude of series (literally thousands) must be forecast, and frequently.

   d.  Read Ord and Lowe (1996), who review most of the automatic forecasting software, and report what you learned.  After reading Ord and Lowe, how, if at all, would you revise your answers to parts a, b and c above?

* Remarks, suggestions, hints, solutions:  You decide!

6.  (The multiplicative seasonal ARIMA (p,d,q) x (P,D,Q) model)  Consider the forecasting

model,

$$\Phi_s(L^s) \; \Phi(L)(1 - L)^d(1 - L^s)^D y_t \; = \; \Theta_s(L^s) \, \Theta(L) \; \varepsilon_t$$

$$\Phi_s(L^s) \; = \; 1 \; - \; \varphi_1^s L^s \; - \; ... \; - \; \varphi_P^s L^{Ps}$$

$$\Phi(L) \; = \; 1 \; - \; \varphi_1 L \; - \; ... \; - \; \varphi_p L^p$$

$$\Theta_s(L^s) \; = \; 1 \; - \; \theta_1^s L^s \; - \; ... \; - \; \theta_Q^s L^{Qs}$$

$$\Theta(L) \; = \; 1 \; - \; \theta_1 L \; - \; ... \; - \; \theta_q L^q.$$

    a.  The standard ARIMA(p,d,q) model is a special case of this more general model.  In

        what situation does it emerge?  What is the meaning of the ARIMA (p,d,q) x

        (P,D,Q) notation?

* Remarks, suggestions, hints, solutions:  The standard ARIMA(p,d,q) model emerges when

$\Theta_s(L) \; = \; \Phi_s(L) \; = \; 1$ and D=0.  p, d, and q refer to the orders of the "regular" ARIMA lag

operator polynomials, as always, whereas P, D and Q refer to the orders of seasonal ARIMA lag

operator polynomials.

    b.  The operator $(1-L^s)$ is called the seasonal difference operator.  What does it do when

        it operates on $y_t$?  Why might it routinely appear in models for seasonal data?

* Remarks, suggestions, hints, solutions:  $(1-L^s)y_t \; = \; y_t \; - \; y_{t-s}$, which makes it natural for the

seasonal difference operator to appear in seasonal models.

    c.  The appearance of $(1-L^s)$ in the autoregressive lag operator polynomial moves us into the realm of stochastic seasonality, in contrast to the deterministic seasonality of Chapter 6, just as the appearance of (1-L) produces stochastic as opposed to deterministic trend.  Comment.

* Remarks, suggestions, hints, solutions:  Just as $(1-L)$ has its root on the unit circle, so too does $(1-L^s)$ have twelve roots, all on the unit circle.

    d.  Can you provide some intuitive motivation for the model?  Hint:  Consider a purely seasonal ARIMA(P,D,Q) model, shocked by serially correlated disturbances.  Why might the disturbances be serially correlated?  What, in particular, happens if an ARIMA(P,D,Q) model has ARIMA(p,d,q) disturbances?

* Remarks, suggestions, hints, solutions:  Inspection reveals that a purely seasonal ARIMA(P,D,Q) model with ARIMA(p,d,q) disturbances is of ARIMA (p,d,q) x (P,D,Q) form. The notion that a seasonal ARIMA(P,D,Q) model might have ARIMA(p,d,q) disturbances is not unreasonable, as shocks are often serially correlated for a variety of reasons.  On the contrary, it's *white noise* shocks that are special and require justification!

    e.  The multiplicative structure implies restrictions.  What, for example, do you get when you multiply $\Phi_s(L)$ and $\Phi(L)$?

* Remarks, suggestions, hints, solutions:  It's easiest to take a specific example.  Suppose that $\Phi_s(L) = (1-\Phi L^{12})$ and $\Phi(L) = (1-\varphi L)$.  Then

$$\Phi_s(L)\Phi(L) = 1 - \varphi L - \Phi L^{12} + \varphi\Phi L^{13}.$$

The degrees of the seasonal and nonseasonal lag operator polynomials add when they are

multiplied, so the product is a lag operator polynomial of degree 12+1=13. It is, however, subject

to a number of *restrictions* associated with the multiplicative structure. Powers of L from 2

through 11 don't appear (the coefficients are restricted to be 0) and the coefficient on $L^{13}$ is the

product of the coefficients on L and $L^{12}$. The restrictions promote parsimony.

      f. What do you think are the costs and benefits of forecasting with the multiplicative

          ARIMA model vs. the "standard" ARIMA model?

* Remarks, suggestions, hints, solutions: The multiplicative model imposes restrictions, which

may be incorrect. If the restrictions are strongly at odds with the dynamics in the data, they will

likely hurt forecasting performance. On the other hand, the restrictions promote parsimony,

which the parsimony/shrinkage principle suggests may enhance forecast performance, other things

the same.

      g. Recall that in Chapter 10 we analyzed and forecasted liquor sales using an ARMA

          model with deterministic trend. Instead analyze and forecast liquor sales using an

          ARIMA (p,d,q) x (P,D,Q) model, and compare the results.

7. (The Dickey-Fuller regression in the AR(2) case) Consider the AR(2) process,

$$y_t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} = \varepsilon_t.$$

    a. Show that it can be written as

$$y_t = \rho_1 y_{t-1} + \rho_2(y_{t-1} - y_{t-2}) + \varepsilon_t,$$

       where

$$\rho_1 = -(\varphi_1 + \varphi_2)$$

$$\rho_2 = \varphi_2.$$

* Remarks, suggestions, hints, solutions:  Just substitute the expressions for $\rho_1$ and $\rho_2$ and rearrange.

   b.  Show that it can also be written as a regression of $\Delta y_t$ on $y_{t-1}$ and $\Delta y_{t-1}$.

* Remarks, suggestions, hints, solutions:  Subtract $y_{t-1}$ from each side of the expression in part a.

   c.  Show that if $\rho_1 = 1$, the AR(2) process is really an AR(1) process in first differences;

      that is, the AR(2) process has a unit root.

* Remarks, suggestions, hints, solutions:  Note that the coefficient on $y_{t-1}$ in the representation obtained in part b is $(\rho_1 - 1)$.  But $\rho_1 = -(\varphi_1 + \varphi_2)$, so the coefficient on $y_{t-1}$ is really

$-(\varphi_1 + \varphi_2) - 1$.  But in the unit root case, $\varphi_1 + \varphi_2 = -1$, so the coefficient on $y_{t-1}$ is 0, which is to say

that the AR(2) process is really an AR(1) in first differences.

8.  (Holt-Winters smoothing with multiplicative seasonality)  Consider a seasonal Holt-Winters smoother, written as

   (1)  Initialize at t=s:

$$\bar{y}_s = \frac{1}{s} \sum_{t=1}^{s} y_t$$

$$T_s = 0$$

$$F_j = \frac{y_j}{\left(\dfrac{1}{s}\displaystyle\sum_{t=1}^{s} y_t\right)}, \quad j = 1, 2, \ldots, s$$

(2) Update:

$$\bar{y}_t = \alpha\left(\frac{y_t}{F_{t-s}}\right) + (1-\alpha)\left(\bar{y}_{t-1}+T_{t-1}\right), \quad 0<\alpha<1$$

$$T_t = \beta\left(\bar{y}_t-\bar{y}_{t-1}\right) + (1-\beta)\,T_{t-1}, \quad 0<\beta<1$$

$$F_t = \gamma\left(\frac{y_t}{\bar{y}_t}\right) + (1-\gamma)\,F_{t-s}, \quad 0<\gamma<1$$

$$t = s+1, \ldots, T.$$

(3) Forecast:

$$\hat{y}_{T+h,T} = \left(\bar{y}_T + hT_T\right)F_{T+h-s}, \quad h = 1, 2, \ldots, s,$$

$$\hat{y}_{T+h,T} = \left(\bar{y}_T + hT_T\right)F_{T+h-2s}, \quad h = s+1, s+2, \ldots, 2s,$$

etc.

a. The Holt-Winters seasonal smoothing algorithm given in the text is more precisely called Holt-Winters seasonal smoothing with additive seasonality. The algorithm given above, in contrast, is called Holt-Winters seasonal smoothing with multiplicative seasonality. How does this algorithm differ from the one given in the text, and what, if anything, is the significance of the difference?

* Remarks, suggestions, hints, solutions: The key difference is in the first updating equation, which now involves division rather than subtraction of the seasonal. Division is appropriate if the seasonality is multiplicative, whereas subtraction is appropriate if it is additive.

b. Assess the claim that Holt-Winters with multiplicative seasonality is appropriate when the seasonal pattern exhibits increasing variation.

* Remarks, suggestions, hints, solutions: Multiplicative seasonality corresponds to additive seasonality in logs, and the logarithm is a common variance-stabilizing transformation.

c. How does Holt-Winters with multiplicative seasonality compare with the use of Holt-Winters with additive seasonality applied to logarithms of the original data?

* Remarks, suggestions, hints, solutions: Again, multiplicative seasonality corresponds to additive seasonality in logs.

9. (Cointegration) Consider two series, x and y, both of which are I(1). In general there is no way to form a weighted average of x and y to produce an I(0) series, but in the very special case where such a weighting does exist, we say that x and y are cointegrated. Cointegration corresponds to situations in which variables tend to cling to one another, in the sense that the

cointegrating combination is stationary, even though each variable is nonstationary. Such situations arise frequently in business, economics, and finance. To take a business example, it's often the case that both inventories and sales of a product appear I(1), yet their ratio (or, when working in logs, their difference) appears I(0), a natural byproduct of various schemes that adjust inventories to sales. Engle and Granger (1987) is the key early research paper on cointegration; Johansen (1995) surveys most of the more recent developments, with emphasis on maximum likelihood estimation.

    a. Consider the bivariate system,

$$x_t = x_{t-1} + v_t, \qquad v_t \sim WN$$

$$y_t = x_t + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

        Both x and y are I(1). Why? Show, in addition, that x and y are cointegrated. What is the cointegrating combination?

* Remarks, suggestions, hints, solutions: x is a random walk, which we showed to be the most fundamental I(1) process. Y equals x plus white noise, and adding white noise to an I(1) process cannot eliminate its I(1) property. Immediately, y-x, is I(0) and in fact white noise, so the cointegrating combination is y-x.

    b. Engle and Yoo (1987) show that optimal long-run forecasts of cointegrated variables obey the cointegrating relationship exactly. Verify their result for the system at hand.

* Remarks, suggestions, hints, solutions:  Here the optimal forecast of x at any horizon is $\hat{x}=x_t$,

and the optimal forecast of y is $\hat{y}=\hat{x}$.

10.  (Error-correction)  In an error-correction model, we take a long-run model relating I(1)

variables, and we augment it with short-run dynamics.  Suppose, for example, that in long-run

equilibrium y and x are related by y=bx.  Then the deviation from equilibrium is z=y-bx, and the

deviation from equilibrium at any time may influence the future evolution of the variables, which

we acknowledge by modeling $\Delta$x as a function of lagged values of itself, lagged values of $\Delta$y, *and*

*the lagged value of z, the error-correction term*.  For example, allowing for one lag of $\Delta$x and one

lag of $\Delta$y on the right side, we write equation for x as

$$\Delta x_t = \alpha_x \Delta x_{t-1} + \beta_x \Delta y_{t-1} + \gamma_x z_{t-1} + \varepsilon_{xt}.$$

Similarly, the y equation is

$$\Delta y_t = \alpha_y \Delta x_{t-1} + \beta_y \Delta y_{t-1} + \gamma_y z_{t-1} + \varepsilon_{yt}.$$

So long as one or both of $\gamma_x$ and $\gamma_y$ are nonzero, the system is very different from a VAR in first

differences; the key feature that distinguishes the error-correction system from a simple VAR in

first differences is the inclusion of the error-correction term, so that the deviation from equilibrium

affects the evolution of the system.

      a.  Engle and Granger (1987) establish the key result that existence of cointegration in a

              VAR and existence of error-correction are equivalent -- a VAR is cointegrated if

              and only if it has an error-correction representation.  Try to sketch some intuition

              as to why the two should be linked.  Why, in particular, might cointegration imply

error correction?

* Remarks, suggestions, hints, solutions: Cointegration between series x and y implies that they cling together in the long run, in the sense that they are unlikely to deviate very far from the cointegrating combination (because it is stationary with zero mean). Hence the current deviation from the cointegrating combination is informative about the likely future paths of x and y, which is to say that error correction is operative.

      b. Why are cointegration and error correction of interest to forecasters in business,

          finance, economics and government?

* Remarks, suggestions, hints, solutions: Many economic and financial situations are consistent with the notion that although series of interest may be approximately I(1), ratios or spreads of those variables may be approximately I(0). Three- and six-month treasury bill rates, for example, may be I(1), yet the spread between them may be I(0). Hence if the spread gets too far out of line, economic forces operate to pull it back. This is precisely the idea behind spread or convergence trades in financial markets.

      c. Evaluation of forecasts of cointegrated series poses special challenges, insofar as

          traditional accuracy measures don't value the preservation of cointegrating

          relationships, whereas presumably they *should*. For details and constructive

          suggestions, see Christoffersen and Diebold (1998).

11. (Forecast encompassing tests for I(1) series) An alternative approach to testing for forecast encompassing, which complements the one presented in Chapter 12, is particularly useful in I(1) environments. It's based on forecasted h-step *changes*. We run the regression

$$\left(y_{t+h} - y_t\right) = \beta_a\left(y_{t+h,t}^a - y_t\right) + \beta_b\left(y_{t+h,t}^b - y_t\right) + \varepsilon_{t+h,t}.$$

As before, forecast encompassing corresponds to coefficient values of (1,0) or (0,1). Under the null hypothesis of forecast encompassing, the regression based on levels and the regression based on changes are identical.

12. (Evaluating forecasts of integrated series) The unforecastability principle remains intact regardless of whether the series being forecast is stationary or integrated: the errors from optimal forecasts are not predictable on the basis of information available at the time the forecast was made. However, some additional implications of the unforecastability principle emerge in the case of forecasting I(1) series, including:

    a. If the series being forecast is I(1), then so too is the optimal forecast.

    b. An I(1) series is always cointegrated with its optimal forecast, which means that there exists an I(0) linear combination of the series and its optimal forecast, in spite of the fact that both the series and the forecast are I(1).

    c. The cointegrating combination is simply the difference of the actual and forecasted values -- the forecast error. Thus the error corresponding to an optimal forecast of an I(1) series is I(0), in spite of the fact that the series is not.

Cheung and Chinn (1999) make good use of these results in a study of the information content of U.S. macroeconomic forecasts; try to sketch their intuition. (Hint: Suppose the error in forecasting an I(1) series were *not* I(0). What would that imply?)

* Remarks, suggestions, hints, solutions: If the error in forecasting an I(1) series were I(1) rather than I(0), then the forecast error would be forecastable, so the forecast couldn't have been optimal.

13. (Theil's U-statistic) Sometimes it's informative to compare the accuracy of a forecast to that

of a "naive" competitor. A simple and popular such comparison is achieved by the U statistic, which is the ratio of the 1-step-ahead MSE for a given forecast relative to that of a random walk forecast $y_{t+1,t}=y_t$; that is,

$$
U = \frac{\sum_{t=1}^{T} \left(y_{t+1} - y_{t+1,t}\right)^2}{\sum_{t=1}^{T} \left(y_{t+1} - y_t\right)^2} \ .
$$

One must remember, of course, that the random walk is not necessarily a naive competitor, particularly for many economic and financial variables, so that values of U near one are not necessarily "bad."

The U-statistic is due to Theil (1966, p. 28), and is often called "Theil's U-statistic." Several authors, including Armstrong and Fildes (1995), have advocated using the U statistic and close relatives for comparing the accuracy of various forecasting methods across series.

* Remarks, suggestions, hints, solutions: It is important to emphasize that macroeconomic series (e.g., consumption) and financial series (e.g., asset prices) are often well-approximated by random walks. Thus, particularly in macroeconomic and financial environments, U-statistics near one do not necessarily indicate poor forecast performance.

## Chapter 14 Problems and Complements

1. (Removing conditional mean dynamics before modeling volatility dynamics)  In the application in the text we noted that NYSE stock returns appeared to have some weak conditional mean dynamics, yet we ignored them and proceeded directly to model volatility.

   a. Instead, first fit autoregressive models using the SIC to guide order selection, and then fit GARCH models to the residuals.  Redo the entire empirical analysis reported in the text in this way, and discuss any important differences in the results.

* Remarks, suggestions, hints, solutions:  Neglected conditional mean dynamics may masquerade as conditional variance dynamics, so it is important that they be adequately modeled first.  The conditional mean "model" used in the text was simply a constant term, which is arguably adequate to the extent that the true conditional mean dynamics, if any, are negligible, which is effectively what the students are asked to assess in this problem.

   b. Consider instead the simultaneous estimation of all parameters of AR(p)-GARCH models.  That is, estimate regression models where the regressors are lagged dependent variables and the disturbances display GARCH.  Redo the entire empirical analysis reported in the text in this way, and discuss any important differences in the results relative to those in the text and those obtained in part a above.

* Remarks, suggestions, hints, solutions:  It's usually better to do things in one step rather than two...

2. (Variations on the basic ARCH and GARCH models).  Using the stock return data, consider

richer models than the pure ARCH and GARCH models discussed in the text.

    a. Estimate, diagnose and discuss a threshold GARCH(1,1) model.

    b. Estimate, diagnose and discuss an EGARCH(1,1) model.

    c. Estimate, diagnose and discuss a component GARCH(1,1) model.

    d. Estimate, diagnose and discuss a GARCH-M model.

* Remarks, suggestions, hints, solutions:  This problem will build students' understanding and skill in estimating and interpreting variations on the basic GARCH model.  Equally importantly, it will emphasize the numerical delicacy of the optimizations, and the care that must be taken.

3.  (Empirical performance of pure ARCH models as approximations to volatility dynamics) Here we will fit pure ARCH(p) models to the stock return data, including values of p larger than p=5 as done in the text, and contrast the results with those from fitting GARCH(p,q) models.

    a. When fitting pure ARCH(p) models, what value of p seems adequate?

    b. When fitting GARCH(p,q) models, what values of p and q seem adequate?

    c. Which approach appears more parsimonious?

* Remarks, suggestions, hints, solutions:  Pure ARCH models typically require *very* large p, whereas the GARCH(1,1) is often adequate.

4.  (Direct modeling of volatility proxies)  In the text we fit an AR(5) directly to a subset of the squared NYSE stock returns.  In this exercise, use the *entire* NYSE dataset.

    a. Construct, display and discuss the fitted volatility series from the AR(5) model.

    b. Construct, display and discuss an alternative fitted volatility series obtained by

             exponential smoothing, using a smoothing parameter of .10, corresponding to a

             large amount of smoothing, but less than done in the text.

c. Construct, display and discuss the volatility series obtained by fitting an appropriate

   GARCH model.

d. Contrast the results of parts a, b and c above.

e. Why is fitting of a GARCH model preferable in principle to the AR(5) or exponential

   smoothing approaches?

* Remarks, suggestions, hints, solutions:  GARCH is preferable to exponential smoothing because

it chooses the amount of smoothing optimally via likelihood optimization.

5. (GARCH volatility forecasting)  You work for Xanadu, a luxury resort in the tropics.  The

daily temperature in the region is beautiful year-round, with a mean around 76 (Fahrenheit!) and

no conditional mean dynamics.  Occasional pressure systems, however, can cause bursts of

temperature volatility.  Such volatility bursts generally don't last long enough to drive away

guests, but the resort still loses revenue from fees on activities that are less popular when the

weather isn't perfect.  In the middle of such a period of high temperature volatility, your boss gets

worried and asks you make a forecast of volatility over the next ten days.  After some

experimentation, you find that daily temperature $y_t$ follows

$$y_t \,|\, \Omega_{t-1} \sim N(\mu, \ \sigma_t^2),$$

where $\sigma_t^2$ follows a GARCH(1,1) process,

$$\sigma_t^2 \ = \ \omega \ + \ \alpha \varepsilon_{t-1}^2 \ + \ \beta \sigma_{t-1}^2.$$

a. Estimation of your model using historical daily temperature data yields $\hat{\mu}=76$,

   $\hat{\omega}=3$, $\hat{\alpha}=.6$, and $\hat{\beta}=0$.  If yesterday's temperature was 92 degrees, generate point

   forecasts for each of the next ten days conditional variance.

b. According to your volatility forecasts, how many days will it take until volatility drops

enough such that there is at least a 90% probability that the temperature will be

within 4 degrees of 76?

    c. Your boss is impressed by your knowledge of forecasting, and asks you if your model

can predict the next spell of bad weather. How would you answer him?

6. (Assessing volatility dynamics in observed returns and in standardized returns) In the text we

sketched the use of correlograms of squared observed returns for the detection of GARCH, and

squared standardized returns for diagnosing the adequacy of a fitted GARCH model.

Examination of Ljung-Box statistics is an important part of a correlogram analysis. McLeod and

Li (1983) show that the Ljung-Box statistics may be legitimately used on squared observed

returns, in which case it will have the usual $\chi^2_m$ distribution under the null hypothesis of

independence. Bollerslev and Mikkelson (1996) argue that one may also use the Ljung-Box

statistic on the squared standardized returns, but that a better distributional approximation is

obtained in that case by using a $\chi^2_{m-k}$ distribution, where k is the number of estimated GARCH

parameters, to account for degrees of freedom used in model fitting.

7. (Allowing for leptokurtic conditional densities) Thus far we have worked exclusively with

conditionally Gaussian GARCH models, which correspond to

$$\varepsilon_t = \sigma_t v_t$$

$$v_t \overset{iid}{\sim} N(0, 1) \, ,$$

or equivalently, to normality of the standardized return, $\varepsilon_t / \sigma_t$.

    a. The conditional normality assumption may sometimes be violated. However,

Bollerslev and Wooldridge (1992) show that GARCH parameters are consistently estimated by Gaussian maximum likelihood even when the normality assumption is incorrect. Sketch some intuition for this result.

* Remarks, suggestions, hints, solutions: Does consistency of ordinary least squares in a linear regression require normality of the disturbance? What is the relationship between the ordinary least squares estimator and the Gaussian maximum likelihood estimator?

b. Fit an appropriate conditionally Gaussian GARCH model to the stock return data. How might you use the histogram of the standardized returns to assess the validity of the conditional normality assumption? Do so and discuss your results.

* Remarks, suggestions, hints, solutions: Visually examine it, together with the usual statistics such as skewness and kurtosis.

c. Sometimes the conditionally Gaussian GARCH model does indeed fail to explain all of the leptokurtosis in returns; that is, especially with very high-frequency data, we sometimes find that the conditional density is leptokurtic. Fortunately, leptokurtic conditional densities are easily incorporated into the GARCH model. For example, in Bollerslev's (1987) conditionally Student's-t GARCH model, the conditional density is assumed to be Student's t, with the degrees-of-freedom d treated as another parameter to be estimated. More precisely, we write

$$\varepsilon_t = \sigma_t v_t$$

$$v_t \overset{iid}{\sim} \frac{t_d}{std(t_d)} \; .$$

What is the reason for dividing the Student's t variable, $t_d$, by its standard

deviation, $std(t_d)$? How might such a model be estimated?

* Remarks, suggestions, hints, solutions: We divide the Student's t variable by its standard

deviation to give it a unit variance. Estimation may proceed by maximum likelihood, with d

viewed as another parameter to be estimated.

8. (Optimal prediction under asymmetric loss) In the text we stressed GARCH modeling for

improved interval and density forecasting, implicitly working under a symmetric loss function.

Less obvious but equally true is the fact that, under *asymmetric* loss, volatility dynamics can be

exploited to produce improved *point* forecasts, as shown by Christoffersen and Diebold (1996,

1997). The optimal predictor under asymmetric loss is not the conditional mean, but rather the

conditional mean shifted by a time-varying adjustment that depends on the conditional variance.

The intuition for the bias in the optimal predictor is simple -- when errors of one sign are more

costly than errors of the other sign, it is desirable to bias the forecasts in such a way as to reduce

the chance of making an error of the more damaging type. The optimal amount of bias depends

on the conditional prediction error variance of the process because, as the conditional variance

grows, so too does the optimal amount of bias needed to avoid large prediction errors of the more

damaging type.

9. (Multivariate GARCH models) In the multivariate case, such as when modeling a *set* of

returns rather than a single return, we need to model not only conditional variances, but also

conditional *covariances*.

    a. Is the GARCH conditional variance specification introduced earlier, say for the i-th

        return,

$$\sigma_{it}^2 = \omega + \alpha \varepsilon_{i,t-1}^2 + \beta \sigma_{i,t-1}^2 \, ,$$

still appealing in the multivariate case? Why or why not?

* Remarks, suggestions, hints, solutions: It is potentially quite restrictive, because in general $\sigma_{it}^2$ should depend on lagged squared returns and lagged variances for *every* asset in the system.

   b. Consider the following specification for the conditional covariance between **i-th** and j-th returns:

$$\sigma_{ij,t} = \omega + \alpha \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta \sigma_{ij,t-1} \, .$$

   Is it appealing? Why or why not?

* Remarks, suggestions, hints, solutions: It is potentially quite restrictive, because in general $\sigma_{ij,t}$ should depend on lagged return cross products and lagged variances for *every* asset in the system.

   c. Consider a fully general multivariate volatility model, in which every conditional variance and covariance may depend on lags of every conditional variance and covariance, as well as lags of every squared return and cross product of returns. What are the strengths and weaknesses of such a model? Would it be useful for modeling, say, a set of five hundred returns? If not, how might you proceed?

* Remarks, suggestions, hints, solutions: The unrestricted multivariate model is quite general, but without imposition of restrictions it would contain so many parameters as to make its estimation impossible. Possibilities include the simple restrictions entertained in parts a and b, as well as more subtle factor structures consistent with financial economic theory, as for example in Diebold and Nerlove (1989) and the subsequent literature.