LONG-RANGE FORECASTING
From Crystal Ball to Computer
Seven
EXTRAPOLATION
METHODS

Contents

Data for Extrapolations .......... 152
Historical Data .................. 153
Simulated Data .................. 155
Analyzing the Data ............... 156
Markov Chains ................... 157
Exponential Smoothing .......... 160
Moving Averages ................. 172
Box-Jenkins ..................... 174
Regressions ..................... 177
Comparisons of Extrapolation Methods .......... 177
Cycles and Fancy Curves ........ 179
Combined Forecasts .............. 183
Assessing Uncertainty .......... 186
Summary ....................... 187
A trend is a trend is a trend,
But the question is, will it bend?
Will it alter its course
Through some unforeseen force
And come to a premature end?

Cairncross (1969)

Extrapolation is second only to judgmental forecasting in popularity. It can be used for short-range or long-range forecasting. The quotation from Cairncross, however, suggests that the errors increase when extrapolation is used for long-range forecasting.

This chapter describes the types of data that are available and explains when they may be used for extrapolations. The selection of data is important, as is the decision of how to analyze the data.

Techniques for analyzing data are then described. Attention is given to Markov chains, a method used when the pattern of events is of interest, and to exponential smoothing, a method for analyzing time series. The Box-Jenkins approach, moving averages, and regressions are also considered as ways to analyze time series. Some simple and fancy curves are then described briefly. These discussions would be long and complex except for one thing: the value of simplicity in extrapolation. Spellbinding rain dances exist for extrapolating data, but the empirical evidence suggests that you should not learn these dances. Comparable results can be achieved with simple methods. It is your choice: you can learn the fox-trot or you can learn to tap dance in scuba gear.

Eclectic research is reexamined. A strategy of combining forecasts from different methods is suggested, and the empirical evidence is reviewed.

This chapter concludes with a discussion of uncertainty. How can extrapolation techniques be used to estimate uncertainty, and what evidence exists regarding the value of these techniques?

DATA FOR EXTRAPOLATIONS

The basic strategy of extrapolation is to find data that are representative of the event to be forecast. The assumption is made that the future event will conform to these data. To ensure that this assumption is reasonable, it often helps to use the systems approach to define the
elements of the system that will be stable. It is not necessary to refer to the study on the prediction of transistors from (LRF pp. 19-20). Nor will I refer to the study of the bituminous coal market by Hutchesson (1967), for if I did, you would ask, “Why didn’t he look at the market for energy before examining the market for bituminous coal?” and I would have to say, “Well, he didn’t, and I told you so.” The extrapolations were bad, as you would plainly see if the following results were shown:

Hutchesson (1967) analyzed U.S. per capita consumption of bituminous coal from 1880 to 1920. Two extrapolations were made, one from a logistic curve and the other from a Gompertz curve. The forecasts of pounds per capita for 10, 20, and 30 years in the future are shown, along with the actual consumption.

<table>
<thead>
<tr>
<th>Year</th>
<th>Forecast Horizon</th>
<th>Logistic Forecast</th>
<th>Gompertz Forecast</th>
<th>Actual Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930</td>
<td>10</td>
<td>15,100</td>
<td>14,300</td>
<td>7400</td>
</tr>
<tr>
<td>1940</td>
<td>20</td>
<td>18,500</td>
<td>21,700</td>
<td>6200</td>
</tr>
<tr>
<td>1950</td>
<td>30</td>
<td>21,200</td>
<td>32,700</td>
<td>6100</td>
</tr>
</tbody>
</table>

Sometimes the choice of data is obvious (e.g., forecasts of automobile sales in the United States may be obtained from historical data on automobile sales). Other times the selection of data is not so obvious. This is especially true in situations involving large changes (e.g., forecasting automobile sales in the Middle East).

In the following sections, four sources of data are described: historical data, analogous situations, laboratory simulations, and field simulations. The advantages and disadvantages of these sources are examined.

**Historical Data**

Extrapolations are typically based upon historical data for the event that is of interest. If we have direct experience with this event over a historical period, it provides an obvious way to forecast. For example, today’s weather provides a good forecast of tomorrow’s weather.

The accuracy of extrapolations is affected by two major conditions:
the accuracy of the historical data and the extent to which underlying conditions will change in the future.

Alonso (1968) described how errors in measurement can lead to large errors in forecasting change even if the underlying change process remains stable. He presented an example in which the current population was estimated to be within 1% (i.e., ± 1%) of actual, and the underlying change process was known perfectly. A two-period forecast was then calculated, and the forecast of change had a confidence interval of ± 37%. (This illustration is described in Appendix B of LRF.)

If, in addition to measurement error, the underlying process can change, the difficulties of forecasting change are compounded. Alonso's example assumed that, if the effects of the change in the underlying process were known within ± 20%, the compound forecast error from measuring current status and from estimating change would be ± 41% (see Appendix B).

The impact of measurement errors is important because real-world data are often inaccurate. Mistakes, cheating, shifting definitions, missing data, and abnormal events cause serious problems. Because accurate data are often not available when needed, recent historical data must be approximated.

If the time lag in the collection of data is large, and additional historical data cannot be obtained, the implication is that extrapolations may be more appropriate for medium-range than for short-range forecasts. On the other hand, because the underlying process changes more in the long range, extrapolations seem more appropriate to the short- rather than to the long-range.

When historical data cannot be obtained, as often occurs when large changes are expected, one might consider analogous situations. For example, if a new educational program was being introduced into a school system, one would look for evidence on other school systems. Evidence would be desired concerning schools where the situation before the change was similar to that in the school of current interest and where the proposed change was similar. Similar situations would arise in predicting the spread of an innovation in farming (Rogers, 1983); forecasting sales for a new product (Claycamp and Liddy, 1969); predicting the vote on a proposal for a new county health board; or forecasting acceptance of a new form of mass transit in Philadelphia. The last two were projects in which I was involved, and it was easy to identify analogous situations. For the county health board, similar proposals had been considered in the previous election in neighboring counties. For the mass transit system, similar systems had been introduced in the Netherlands.
Stein's paradox (Efron and Morris, 1977) says that the use of analogous data can improve predictions. Stein presented evidence that the overall league batting average early in the season can be used to improve the prediction of a baseball player's final batting average. The less information available on a given player, the more weight should be given to the league average. (This probably sounds like a paradox only to a trained statistician. But then, they averaged Volkswagens and batting averages—and that was a bit strange.)

Simulated Data

In cases where no actual situations are available, a simulation could be used to generate data. The simulation could be done in a laboratory or in a realistic situation; the latter is referred to as a field test.

Laboratory simulations frequently offer lower costs, better control over the changes, and more confidentiality than a field test. Such simulations have been used commercially for many years in marketing where, for example, new products can be tested in simulated stores. For a description of marketing simulations see WIND [1982, pp. 422-429]. The laboratory simulation strategy has also been employed successfully in personnel predictions via the "work sample:"

Studies on the validity of work samples are found in DOWNS et al. [1978], HINRICH [1969, 1978], REILLY and CHAO [1982], ROBERTSON and KANDOLA [1982] and SMITH [1976].

Field tests are used in areas such as marketing, social psychology, and agriculture. Descriptions of test marketing go back many years (e.g., Harris, 1964); WIND [1982, pp. 398-422] summarizes the literature. NEVIN [1974] tested the predictive validity of consumer laboratory experiments. An example of a field test in social psychology is provided by the study of obedience among nurses (Hofling et al., 1966), which was discussed in Chapter 6. This study illustrates the difficulties involved in field tests of behavior that is important to people. The subjects were upset when they learned about the experiment. (It's much easier to restrict your field experiments to less important matters, like preference for a new brand of beer.) Field tests offer greater realism than do laboratory experiments, and this is generally an important advantage.

Simulated data, though advantageous for assessing large changes,
may be seriously influenced by the researcher’s biases. Anyone who has worked on government-sponsored research knows this; the sponsors become committed to making the simulation a success.

An outstanding case demonstrating bias due to the sponsor is that of Lysenko, a Russian agricultural scientist (Medvedev, 1969), who tested hypotheses such as these: “Heredity is the property of the living body to demand certain environmental conditions and to react in a certain way to them”; and “No special hereditary substance exists, anymore than does the substance of combustion, phlogiston, or the substance of heat, caloric.” Lysenko conducted field experiments that supported his theories and led to some strange methods of farming in Russia. Lysenko had a big advantage: Stalin was his buddy. Still, the effect of bias on both laboratory and field experiments is enormous—even for non-Russians.

The effects of bias on the part of the researcher or the organization funding the research are so powerful that they can be observed without carefully controlled experiments. For example, I found myself under a great deal of pressure when I obtained results suggesting that the U.S. Department of Transportation should end its funding of a project involving a new form of transportation. The pressure in this case came, not from the Department of Transportation, but from many of the recipients of roughly $1 million per year. No one was killed, but it was exciting. More details are provided in ARMSTRONG 1983d.

A summary of the types of data is provided in Exhibit 7-1. The data sources are rated against five criteria. The ratings were done subjectively by the author. Crude though they are, there are certain implications:

1. No one type of data is best for all situations.
2. For long-range extrapolations (large changes), data for the assessment of current status should be different from that used for predicting change.

The second implication seems reasonable, yet seldom is it used.

ANALYZING THE DATA

For simulated data, one can assume that future behavior will be like that in the simulation. An alternative procedure is to assume that the future behavior will be an average of the current behavior, without change, and the behavior observed in the simulation. This is an ex-
tension of the principle of conservatism; to the extent that you are uncertain about the evidence, you should forecast small changes from the current situation. This tempering of the prediction would seem more appropriate for laboratory data than for field data. For example, predictions based on a laboratory experiment in agriculture should be used with caution. Again, one could refer to the Lysenko case; the results claimed for a controlled test did not hold up when the method was adopted by farmers.

Although conceptually simple, projection from the simulated data to the forecast situation encounters numerous difficulties. Gold (1964) assessed the extent of these difficulties in test marketing. One must be concerned whether the test market is representative of the actual situation that will prevail in the total market. Gold’s study indicated that, even if the test were perfect, substantial errors occur in generalizing from the sample observations to the test market, and from the test market to the total market.

The next section considers the use of Markov chains, a method for examining data organized by events rather than by time. Techniques are then examined for analyzing time series data.

**Markov Chains**

Markov chains, named in honor of a Russian mathematician, have been the focus of much interest by researchers. Numerous books and articles have touted Markov chains as a superior forecasting technique. One of the earliest practical applications was to predict changes in the occupational status of workers (Blumen, Kogan, and McCarthy, 1955). Another early application was predicting doubtful accounts in finance.
Extrapolation Methods

(Cyert, Davidson, and Thompson, 1962). Ezzati (1974) applied Markov chains to forecasts of the home heating market. Unfortunately, these studies are only descriptive; they provide no evidence on the value of Markov chains.

Markov chains use the recent pattern of behavior as a basis for forecasting. Behavior in the future is forecasted from knowledge of the current state of an element and from an analysis of how the element moves from one state to another. This statement may be vague, so here is an example of a simple Markov chain.

Assume that the task is to forecast the type of automobile purchased by families, considering large domestic (L), small domestic (S), and small foreign (F) cars. Also, assume that all cars can be classified within the scheme. Now suppose that three families report the following purchasing sequences over a 15-year period:

Family 1. LLLLLLSLFLLL
Family 2. FSSSSSSSSSL
Family 3. FFFFFFFFSFLLL

This says, for example, that the first car purchased by family 2 was foreign, the next eight were small cars, and the last purchase was a large car.

If these families are representative of the total population of car-buying families, what predictions can be made about the future market shares for L, S, and F? The first step in using Markov chains for prediction is to develop a transition matrix. This summarizes the data by indicating the fraction of times that the behavior in one trial will change (move to another state) in the next trial. Our data on automobile purchases are used to develop the transition matrix in Exhibit 7-2.

This transition matrix indicates that 80% of those who purchased a large domestic car (L) will buy the same type the next time they

<table>
<thead>
<tr>
<th>Exhibit 7-2 EXAMPLE OF A TRANSITION MATRIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand Purchased at Trial t + 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Brand</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>
Analyzing the Data

shop, while 10% will switch to a small domestic car (S), and 10% to a small foreign car (F).

If the process underlying Exhibit 7-2 remains stable, and if the sample of families is representative of the entire population, the transition matrix can be used to forecast changes in market shares. One can multiply the current market shares (say 70% for L, 20% for S, and 10% for F) by the transition matrix to determine how market shares will change during the next buying cycle (about 1.4 years in this example). This yields a prediction for the next time period of 62% for L, 23% for S, and 15% for F. If forecasts were desired for the long run, the process could be repeated many times. For this example, the long-range solution approaches 50% for L, 30% for S, and 20% for F. (This was obtained by calculating $MP = M$, where $M$ is the matrix of shares (L, S, and F), and $P$ is the transition matrix. The calculations for this example are shown in Appendix F.)

Remember that an assumption of stability is made when Markov chains are used. This means that it becomes risky to obtain long-range forecasts in cases where efforts are made to change the transition matrix. For example, an American firm's success in convincing the U.S. government to restrict the entry of foreign cars would affect the entries in the last column of Exhibit 7-2.

Markov chains seem reasonable for some problems. For example, they are widely used for personnel predictions. This technique and similar ones have been recommended frequently for predictions in marketing when people are assumed to go through various states in using a product (e.g., trial, repeat purchase, and adoption). Unfortunately, despite many publications on Markov and related models, little research on their predictive value was found. Akers and Irwin (1967) claim that Markov chains have been of no demonstrable value in demographic forecasting. I suspect there may be unpublished cases in marketing where Markov chains led to no improvements because I have met some disappointed users.

In the following, two relevant empirical studies are summarized. They involve short-range forecasting and suggest that Markov models have no substantial advantage over much simpler extrapolation methods.

Armstrong and Farley (1969) examined the selection of supermarkets by each family, using data from 45 families in the Chi-
cago Tribune panel. Data from 6 months were used to estimate the transition matrix. Forecasts were made of the percentage of trips made to the favorite store during the following 6 months for each family. Forecasts from the Markov model were compared to an extrapolation that assumed nothing would change. The Markov model provided forecasts that were correlated to the actual changes ($r = .4$). Although this was significantly better than the "no-change" model, it reduced the standard error of the forecast by only 8%. When the data were analyzed by family, the Markov model was correct on forecasts of change for 19 families, and incorrect for 16 (no change was predicted for the other 10 families); this result is likely to have arisen by chance.

Barclay (1963) used Markov chains to predict the percentage of families that would purchase a new product. The Markov forecast, drawn from Chicago Tribune panel data from March to September 1956, was that the purchase rate during 27 bimonthly periods from 1957 to mid-1961 would be 10.8%. I reanalyzed Barclay's data and found that an extrapolation based on the average purchase rate from April through September 1956 yielded a forecast that 16.7% of the families would purchase the product. The Markov model always predicted on the low side, and the average absolute error was 6.6%. In contrast, the no-change model was low on 17 forecasts, was high on 9, and had an average absolute error of 2.5%.

Exponential Smoothing

Exponential smoothing is relatively simple, intuitively pleasing, inexpensive, and well known. It draws upon the philosophy of decomposition. Time series data are assumed to be made up of some basic components, the average, trend, seasonality, and error. The first three components are illustrated in Exhibit 7-3. The x's represent the historical data. Dashed line $A$ represents an estimate of the updated average; dashed line $B$ represents the forecast obtained when the trend is added to the estimate of current status; and dashed line $C$ represents the forecast when the current status, trend, and seasonality components are combined.
Exponential smoothing is similar to a moving average, but it places more weight on the most recent data. The weight on early periods drops off exponentially, so that the older the data, the less their influence. This weighting makes good sense and is supported by empirical studies (e.g., Ash and Smyth, 1973; Pashigian, 1964).

Exponential smoothing can be used, along with decomposition, to analyze historical data. A description of the major steps follows. To help in this discussion, the steps are listed by time priority in Exhibit 7-4. This listing goes beyond the exponential smoothing technique to consider all of the steps involved in obtaining the forecast. The description draws heavily from Brown (1959b) and Winters (1960). More recent work [e.g., see GARDNER, 1985b, for a summary] provides more efficient approaches. No doubt you will rely upon one of the many software packages now available. However, the framework in this chapter, which draws heavily upon Brown, is a good one to illustrate the various issues.

Clean Data. The first step is to review the data to remove obvious errors. When many data are involved, this can be done by computer
Exhibit 7-4  STEPS IN FORECASTING WITH EXPONENTIAL SMOOTHING

1. Clean data
2. Deseasonalize data
3. Select smoothing factors
4. Calculate new average
5. Calculate new trend
6. Estimate current status
7. Calculate forecast
8. Update seasonal factors

(Return to step 3)

by establishing limits for the data and identifying observations that go outside these limits; appropriately enough, these are called outliers. The presence of outliers can affect the estimate of the mean, trend, or seasonality. If possible, outliers should be examined to determine whether they are due to an error (recording or keypunch errors are common causes of outliers) or a large change in the environment. When no explanation is found, it may be wise to temper the influence of outliers. There are many ways to do this. One approach, windsorizing, reduces the outliers to equal the most extreme observation about which you feel confident (Tukey, 1962). GEURTS [1982] claims that this step is likely to have a major influence on accuracy.

Another way to check for errors is to compare measures of a given quantity obtained in different ways. For example, to estimate the number of Japanese cameras sold in the United States in 1956, one could obtain estimates of the number of cameras that the United States claimed were imported from Japan, and the number of cameras that Japan claimed to have exported to the United States. A large difference would suggest the possibility that one of the measures was in error. If it were not clear which measure was in error, the two estimates could be averaged. Incidentally, in this particular example, Japan claimed that 160,180 cameras were exported to the United States, while the United States claimed to have imported 819,374 cameras from Japan (Armstrong, 1968a). I love the way they present six significant digits when there is so much uncertainty about the first digit.

Problems often arise when changes are made in definitions (e.g., Do the data on a U.S. economic series include Hawaii and Alaska in the
complete series? Have the data been adjusted to compensate for inflation?). For short-term forecasts, there is a concern over matters such as the number of working days in a period. Once the problems of shifting definitions are identified, adjustments can be made.

**Deseasonalize Data.** Quite frequently the variable of interest depends upon the time of year. Monthly seasonal factors can be especially useful in economic forecasting. We know, although our local electric company may not, that the demand for electricity is high in the summer. What the electric company calls a “crisis,” we call “July.” Even Arthur Strimm knows this (see Exhibit 7-5).

Seasonal factors can be calculated in many ways. This can be done by using the Census X-11 program\(^1\), by applying regression analysis where the months are represented by dummy variables\(^2\), by calculating the relationship between each month and a corresponding moving average, or by relating each period to the average for the year. If the last approach is used, the trend must be removed before the seasonality factors are calculated.

The seasonal factors can be stated in either multiplicative (e.g., demand in January is 85% of that in the typical month in a year) or additive form (e.g., demand in January is 20,000 units below the average). Multiplicative factors are often used in economics. They are most appropriate when:

1. The data are based on a ratio scale, that is, the scale has a zero point and the intervals between points on the scale are meaningful.
2. Measurement error is low.
3. The trend is large.

---

### Exhibit 7-5  **ON THE VALUE OF SEASONAL FACTORS**

*Source.* “Miss Peach” by Mell Lazarus. Courtesy of Mell Lazarus and Field Newspaper Syndicate.
If conditions 1 and 2 are not met, consider additive factors.

Seasonal factors increase the error if there is a great deal of uncertainty in estimating these factors (Groff, 1973; Nelson, 1972). For example, if only 1 year of data were available, how could one tell whether the variations from month to month were random fluctuations or real seasonal effects? As the number of data increases, the reliability of the seasonality factors increases. It would be desirable, then, to place less weight on unreliable seasonal factors and more weight on highly reliable seasonal factors. Given a multiplicative factor, you could calculate a modified seasonal factor, $\bar{S}_j^*$:

$$\bar{S}_j^* = K_s + (1 - K_s)\bar{S}_j$$

where $K_s$ is a dampening factor (0 < $K_s$ < 1), $j$ is the period of the year, $S_j$ is the raw seasonal factor, and $\bar{S}_j$ is the smoothed seasonal factor. The weight $K_s$ could be set subjectively. A high $K_s$ means more dampening. The $K_s$ would be close to 1.0 if there were few data and little prior expectation of seasonality, and close to 0.0 for many data and a high prior expectation of seasonality. Then weight the seasonal factor by the square root of the number of data, as follows:

$$K_s = \frac{1}{\sqrt{d}}$$

where $d$ is the number of years of data. Do you think we could call this Armstrong's modifier for seasonality? If data for enough years are available, say five or more, $K_s$ could vary according to the standard deviation of the seasonality estimate.

The use of modified seasonal factors is consistent with the fact that some studies find seasonal factors useful, some find them to be of no value, while others find them to be detrimental (Nelson, 1972).

The final step in the calculation of seasonal factors is to normalize so that the average factor is exactly 1.0 (or 0.0 if additive factors are used). An easy way to normalize is to multiply each seasonal factor by the ratio of the number of periods to the total of the seasonal factors. Thus, for monthly seasonal factors, the correction would be to multiply each seasonal factor by $(12/\Sigma \bar{S}_j^{**})$. The modified and normalized seasonal factor can be designated by $\bar{S}_j^{**}$.

Given the seasonal factors, the data can be deseasonalized. Exponential smoothing can be done on the deseasonalized data.
Select Smoothing Factors. Users of exponential smoothing often invest much time and energy in selecting the optimal smoothing factors. Of prime concern is alpha (\(a\)), the smoothing factor for the average:

\[
\bar{Y}_t = a \left( \frac{Y_t}{S_j} \right) + (1 - a)\bar{y}_{t-1}
\]

where \(\bar{Y}\) is the calculated average, \(Y\) is the reported value, \(a\) is the smoothing factor, and \(t\) is the time period.

The smoothing factor determines how much weight is to be placed on the most recent data: the higher the factor, the heavier the weight. For example, a factor of 0.2 would mean that 20% of the new average is due to the latest observation, and the other 80% is due to the previous average. The weights on each period drop off exponentially. Thus the latest period is weighted by \(a\), the period before that by \(a(1 - a)\), and the observation two periods ago by \(a(1 - a)^2\); data of \(d\) periods ago would be weighted by \(a(1 - a)^d\).

How should the smoothing factor be selected? Traditionally, historical data have been used to determine which factors provide optimal forecasts (e.g., see Berry and Bliemel, 1972). DALRYMPLE and KING [1981] examined the selection of factors to minimize the error for \(h\) periods ahead. These procedures seem reasonable, yet I have been unable to find empirical evidence that they lead to improved forecasts. The evidence suggests that the accuracy of forecasts is not highly sensitive to the choice of the optimal smoothing factor. One need only select factors that are in the general region of the optimum. This has been my experience. Support may be found in Winters (1960).

You can use judgment initially to select the smoothing factors. I must tell you that I have no empirical evidence to support the following procedure. It is based upon my belief that the procedure should make sense to the analyst. Keep three things in mind:

1. If the process is unstable, use a high \(a\) so that the smoothing will quickly adjust to the new situation.
2. If the measurement error is high, use a low \(a\) to increase reliability, that is, to dampen the effects of transitory or unusual fluctuations.
3. If the time periods are short, use a low \(a\). Monthly data require a lower \(a\) than do quarterly data, and weekly data a lower \(a\) than monthly data.
Rather than thinking directly about $\alpha$, one could rephrase the question to ask how many historical observations should be included in the average. Ayres (1969, p. 100) suggested a rule of thumb that the extrapolation forward (by $h$ periods) should not exceed the time span of the historical data (i.e., $d \geq h$). In my opinion, this rule understates the need for data in short-term forecasting and is too stringent for long-range forecasting. I suggest a rule of thumb that requires more data for short-term forecasts and fewer for long-term forecasts. This can be accomplished by a square root function. Therefore Armstrong’s rule of thumb suggests that the number of periods of historical data is equal to four times the square root of the number of periods to be forecast, that is,

$$d = 4 \sqrt{h}$$

For a 1-period forecast, this rule calls for 4 periods of historical data; for a 16-period forecast, 16 periods of historical data; and for a 100-period forecast, 40 periods of historical data. Limited evidence is presented in Exhibit 7-6; the time period that yielded the lowest error in each study was compared with recommendations from Armstrong’s and Ayres’ rules of thumb. Armstrong’s rule underestimates the need for data and Ayres’ rule underestimates it even more. If, as Ayres suggests, people follow his rule of thumb, a simple way to improve extrapolations would be to increase the number of historical data. This is a conclusion that Dorn (1950) reached in his review of forecasts on populations: demographers had been using too few data. SCHNAARS [1984], however, concluded that more data does not improve accuracy by a significant amount.

After an appropriate time span has been selected, Brown (1959a, b) states that the number of periods of data in an exponentially smoothed average can be approximated by

$$d = \frac{2 - \alpha}{\alpha}$$

One can select an appropriate time span, $d$, and then calculate the corresponding $\alpha$ by recasting the above formula:

$$\alpha = \frac{2}{d + 1}$$

To the extent that you are uncertain about $\alpha$, it seems best to err on the high side [GARDNER, 1985b].
<table>
<thead>
<tr>
<th>Study</th>
<th>Forecast Horizon (periods)</th>
<th>Situation</th>
<th>Optimum Span by ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ash &amp; Smyth (1973)</td>
<td>1</td>
<td>Macroeconomic</td>
<td>Ayres' Rule 1</td>
</tr>
<tr>
<td>HAGERMAN &amp; RULAND [1979]</td>
<td>1</td>
<td>Annual earnings</td>
<td>Armstrong's Rule 4</td>
</tr>
<tr>
<td>DALRYMPLLE &amp; KING [1983]</td>
<td>1</td>
<td>&quot;Business&quot;</td>
<td>Actual Test 5</td>
</tr>
<tr>
<td>Cragg &amp; Malkiel (1968)</td>
<td>2, 3</td>
<td>Annual earnings</td>
<td>2.5</td>
</tr>
<tr>
<td>Orson (1968)</td>
<td>6</td>
<td>Electric power</td>
<td>6</td>
</tr>
</tbody>
</table>

Exhibit 7-6  OPTIMAL HISTORICAL TIME SPAN FOR EXTRAPOLATION
Typically, $\alpha$ is expected to be between 0 and 1. Indeed, the way I described the model makes an $\alpha$ of more than 1.0 seem counter-intuitive. By using a bit of algebra, however, we can recast the basic formula for exponential smoothing in a different, but equivalent form:

$$\bar{Y}_t = \bar{Y}_{t-1} + \alpha e_t$$

where $e_t$ is the difference between the latest (deseasonalized) observation and the previous average. What this says is that we adjust the new average in light of the most recent change. If one expects the data to be highly auto-correlated, then an $\alpha$ greater than 1.0 might be considered. Still, this strikes me as risky.

Judgmental smoothing factors should be adequate. If the forecast is of particular importance, however, you should conduct a computerized search for the best smoothing constants to be sure that you are near the optimum. Many computer packages have this facility.

This discussion has covered the selection of an appropriate time span and, hence, an appropriate $\alpha$ for exponential smoothing of the average. A similar analysis can also be used in selecting a smoothing factor, beta ($\beta$), for the trend calculations, and also a gamma ($\gamma$) factor to smooth seasonality.

**Calculate New Average.** The formula for calculating the new average was presented in the preceding subsection. Brown's formulation is especially sensitive to starting values. A starting value must be selected for $\bar{Y}_{t-1}$. If a substantial amount of historical data exists (well in excess of $(2 - \alpha)/\alpha$ observations), the decision on the starting value is less important. You can use a range of starting values (highest, most likely, lowest) to test sensitivity.

**Calculate New Trend.** The trend is calculated from period-to-period changes in the new average. The trend can be calculated in a manner analogous to the calculation of the new average, that is, the new smoothed trend is the weighted average of the most recent trend in the average and the previous smoothed trend:

$$\bar{G}_t = (\beta)(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - \beta)(\bar{G}_{t-1})$$

where $\beta$ is the smoothing factor and $\bar{G}$ is the smoothed trend or growth factor. McClain [1974], using theoretical arguments, suggests that it is best to use $\alpha = \beta$. 
The starting value for $\bar{G}$ is often set at zero, but the analyst is advised to make a "best guess" here. This is most important where few historical data exist.

A decision must be made whether to use an exponential or an additive trend. The exponential trend is appropriate when:

1. All observations are greater than zero
2. Data are ratio scaled
3. Measurement error is small
4. Growth is expected to be geometrical

However, for long-range forecasts, exponential trends are risky—even if the above conditions are met. Little direct evidence is available on this issue, though. Elton and Gruber (1972) found additive and exponential trends to produce equivalent accuracy in their forecasts of the earnings of firms.

When uncertainty exists about the trend, it may be useful to estimate a dampened or modified trend. The reasoning here is the same as that used for the modified seasonal factors; it is desirable to place less weight on inaccurate estimates. For a multiplicative factor, calculate a modified trend factor, $\bar{G}_t^*$:

$$\bar{G}_t^* = K_G + (1 - K_G)\bar{G}_t$$

where $K$ is a dampening factor ($0 < K_G < 1$). $K$ could be set subjectively or by a search; it would be 1.0 if the trend estimates are of no value and 0.0 for highly accurate trend estimates.

GARDNER and McKENZIE (1985), in a study of data from the M-Competition [MAKRIDAKIS et al. 1982], found significant improvements in accuracy due to dampening of the trend. Their scheme estimated a parameter that automatically increased the dampening for erratic trends. Consistent trends (noted for about ⅓ of their quarterly and monthly series) were not dampened at all. The dampening was especially valuable for longer forecast horizons. Dampening also helped to avoid large errors.

Some analysts have proposed that the acceleration also be calculated. In other words, the data on the trend in the changes can be exponentially smoothed. This procedure, generally called triple exponential smoothing (because average, trend, and acceleration are smoothed), might be adequate for short-term forecasts. It is risky, however, for longer term forecasts, because it violates Occam's razor (i.e., use the simplest method). Eight studies were found on this topic:
Elton and Gruber (1972) forecasted firms' earnings up to three years; models with acceleration did poorer in this study than those without acceleration. Acceleration models were less accurate over a variety of forecast horizons in the M-competition [MAKRIDA-KIS et al., 1982]. For short-range forecasts, Markland (1970) found an advantage to triple smoothing, Groff (1973) and Torfin and Hoffman (1968) found no difference, and Davies and Scott (1973), SCHNAARS [1984], and SCHNAARS and BAVUSO [1985] found that acceleration models did worse.

**Estimate Current Status.** The smoothed average typically lags the most recent changes. Thus, some adjustment should be made to bring the average up to date. The approach discussed here develops a substantial lag so Brown (1959a, b) suggested the following correction:

$$\bar{Y}_t^* = \bar{Y}_t + \left( \frac{1 - \alpha}{\alpha} \right) \bar{G}_t$$

where $\bar{Y}_t^*$ is current status. In other words, the smoothed trend, $\bar{G}_t$, is multiplied by $(1 - \alpha)/\alpha$. Thus, if $\alpha = 0.4$, the trend would be multiplied by 1.5.

**Calculate Forecast.** The forecast may be obtained for 1 to $h$ periods into the future by taking the estimate of current status, calculating the trend over the forecast horizon, and incorporating seasonality. If additive trends and multiplicative seasonality are assumed, the forecast is the product of the current status plus trend times seasonality:

$$F_{j,t+h} = (\bar{Y}_t^* + h \bar{G}_t^*) \bar{S}_{j,t}^*$$

where $F$ is used to represent the forecast.

**Update Seasonal Factors.** Exponential smoothing can be applied to the seasonal factors using gamma ($\gamma$) as a smoothing factor, where $\gamma$ is selected on the basis of the number of observations for the given period of the year (e.g., the number of January's), meaning that $\gamma$ will nearly always differ from $\alpha$ and $\beta$:

$$\bar{S}_{j,t} = \gamma \left( \frac{Y_t}{\bar{Y}_t} \right) + (1 - \gamma)(\bar{S}_{j,t-1})$$

where $J$ is the number of periods in the year (e.g., 12 for monthly data).
Analyzing the Data

Thus the new smoothed seasonal factor is the weighted sum of the latest information on seasonality (e.g., the ratio of the latest observation to the smoothed average) and the previous smoothed estimate of seasonality. Each time new seasonal factors are calculated, it is necessary to dampen and normalize.

This procedure requires much data and small measurement error. A simpler approach, which gives adequate results, is to recalculate seasonal factors once a year (e.g., by using the Census X-11 program).

Return and Repeat Process. As each new observation comes in, it is a simple matter to update the exponentially smoothed forecast. If there are a number of forecasts, this can be done more cheaply by computer. Frequent updating is important for accuracy.

The smoothing factors are usually held constant. Brown (1959a, b), however, suggested that these factors be increased if large changes are expected; this makes sense. Harrison and Davies (1964) and Trigg and Leach (1967) suggested that the factors be increased if systematic errors are observed over time; this also makes sense. Arguing against this is that the added complexity of adaptive parameters might not yield significant gains in accuracy. In fact, the 12 studies to date offer little evidence favoring adaptive parameters:

SMITH [1974], Chow [1965], WHYBARK [1972], and DENNIS [1978] all found that adaptive parameters improved accuracy. But the findings by Chow, Whybark, and Dennis were not confirmed when replicated [(GARDNER, 1983b; EKERN, 1981]. Torfin and Hoffman [1968], Raine [1971], Adam [1973], Harris and Adam (1975), and Dancer and Gray [1977] found little difference in accuracy. GARDNER and DANNENBRING [1980], MABERT [1978] and McLEAVEY, LEE, and ADAM [1981] found poorer accuracy with adaptive parameters.

Many details have been glossed over in this description. The early works in this area are still useful (e.g., Brown, 1959a, b and Winters, 1960), but the forecasting texts listed on page 78 of LRF offer better descriptions. GARDNER [1985b] provides a comprehensive summary of the literature.

The important thing, I believe, is to find some way to implement the guidelines in this section (Exhibit 7-4). Use Gerstenfeld’s Law of Trying. To test this claim, I asked a research assistant to use the guidelines to develop a computer program. The output looked reason-
able and was verified by hand calculations on a sample of data. Also I presented the monthly data on international air travel from 1949 to 1958 and asked students to provide monthly forecasts for 1959 and 1960. Eight project reports were prepared. The details of the procedure varied greatly. Before any assessment was made of accuracy, I coded the extent to which each report conformed to the guidelines presented here. These ratings were crude because the reports did not always provide adequate descriptions. The mean absolute percentage error of the forecasts was about 5% for the four reports that followed the guidelines, and about 11% for the four reports that departed from the guidelines.

Numerous alternatives to exponential smoothing can be used. One possibility is subjective or “eyeball” extrapolations. ADAM and EBERT [1976], and CARBONE and GORR [1985] found objective methods to be a bit more accurate; however, LAWRENCE, 1983; LAWRENCE et al. 1985 found the accuracy of subjective extrapolations to be good relative to objective methods:

In LAWRENCE et al. [1985], three alternative judgmental methods were compared with extrapolative methods in forecasts for the 111 time series used in MAKRIDAKIS, et al. [1982]. No information was available to the judges other than the historical data. In general, the judgmental extrapolations (“eyeballing”) were at least as accurate as the extrapolation methods. For longer lead times, the quantitative extrapolations were superior. (I believe this may be due to the dampening of the trend factor that occurs with judgmental extrapolations, as shown in EGGLETON [1982].) Surprisingly, subjects provided with tables gave slightly more accurate forecasts than those provided with graphs.

Moving Averages

As noted previously, moving averages are similar to exponential smoothing. The only conceptual difference is that each observation is weighted equally. Thus a four-period moving average would be

\[ Y_t = \frac{Y_{t-3} + Y_{t-2} + Y_{t-1} + Y_t}{4} \]

As a result, moving averages tend to lag current status even more than do exponentially smoothed averages. The average age of the data in the four-period example above would be two periods; this compares
with 1.5 for the corresponding exponentially smoothed model. It is possible, however, to weight more recent observations more heavily, and corrections can be made for the lag.

The advantages of moving averages in comparison with exponential smoothing are that the former are easier to understand and that seasonality can be removed easily from the data by using an average that incorporates complete cycles. Thus, for quarterly data, the moving average could be based on 4, 8, or 12 periods.

An excellent computer program, Census Program X-11, has been available for moving-average analyses for many years (Shiskin, 1965). It has provisions for seasonality, trend, adjustments of outliers, working or trading day adjustments, and differential weighting of observations.

A disadvantage of moving averages is that many data need to be retained for the analysis. A 12-period average requires carrying the 12 periods of historical data in addition to the summary statistics. This disadvantage can become costly when many items are being forecast (e.g., inventory control).

The M-Competition showed moving averages to be less accurate than exponential smoothing [MAKRIDAKIS et al. 1982], as did FRANK [1969]. Other evidence includes the following studies, three of which favor exponential smoothing, and one that shows no difference.

Winters (1960) developed an exponentially weighted moving average with multiplicative seasonal factors to obtain a forecast one period ahead. He examined three sets of data—monthly sales of cooking utensils manufactured by a subsidiary of Alcoa, bimonthly sales of a type of paint manufactured by Pittsburgh Plate Glass, and the number of cellars excavated each month by Admiral Homes. Data were available for five to seven years. Exponential smoothing was slightly superior to a one-year moving average, but the latter did not include a trend estimate.

Kirby (1966) compared three extrapolation methods in short-range forecasting for company sales data. The forecasts were for 23 sewing machine products sold in five countries. The forecast horizon ranged from 1 to 6 months. A period of 7 ½ years was examined, using exponential smoothing and unweighted moving averages. The accuracies of the exponential smoothing and the moving averages were comparable for the 6-month horizon. As
the forecast horizon was shortened, however, the exponential smoothing did slightly better. Kirby also developed forecasts from artificial data by imposing various types of error upon the original data. He found, for example, that with more random error the moving average performed better than the exponential smoothing, and that with a strong cyclical component the exponential smoothing did relatively better than the moving average. Overall, however, the differences among the various extrapolation methods examined by Kirby were small; this was the major conclusion.

Elton and Gruber (1972) found exponentially smoothed forecasts to be better than moving averages in one-, two-, and three-year forecasts of earnings. The relative advantage of exponential smoothing increased slightly as the forecast horizon was lengthened from one to three years. In this case, however, the moving average, did not have a trend factor.

Adam (1973) compared a two-period moving average with an exponentially smoothed average. One-period and 12-period forecasts were obtained for five simulated demand patterns. The moving averages yielded forecasts that were equivalent in accuracy to those obtained with exponential smoothing. (See also the replication by McLeavy, Lee, and Adam [1981].)

Box–Jenkins

Another approach to the analysis of time series is the Box–Jenkins (Box and Jenkins, 1976). It has been one of the most popular dances in the history of forecasting, and the book is among those most frequently stolen from libraries. (Stealing books is actually a form of defensive education for required courses. Although you may not be able to understand the textbook, you can make things more difficult for others by stealing it.)

Some of my best friends are Box–Jenkins experts, but I even have trouble understanding the interpreters. However, the number of interpretations continues to grow. It is interesting that some experts argue that other experts do not understand Box–Jenkins well enough (e.g., see discussants in Chatfield and Prothero, 1973). On the other hand,
CARBONE et al. [1983] found that 18 hours of classroom training were sufficient for learning it. Once learned, Box–Jenkins is expensive to apply. LUSK and NEVES [1984] found that it required about 25 minutes for the typical series though substantially less time was required for simple series.

It may not be necessary to become an expert in Box–Jenkins in order to obtain good results. HILL and FILDES [1984] and LIBERT [1984] each used automatic procedures and obtained accuracy that was comparable to Andersen’s for the 111 series in the M-Competition. The automatic procedure was also faster (about two minutes of computer time per series). But it does not strike me as a good idea to use a complex method that you do not understand, especially when comparable results can be achieved with simple models that people do understand.

In general, Box–Jenkins uses the most recent observation as the starting value (Nelson, 1973) and then analyzes recent forecasting errors to determine the proper adjustments for future time periods. It is analogous to control procedures used in chemical processing, where an examination is made of the desired state and the actual state, and an adjustment based on the difference is made in the process. Usually the adjustment serves to compensate for only part of the error. For example, if the forecast was 10% low in the last period, the forecast for the next period might be adjusted upward by some fraction of this error (e.g., by 5%). This description suggests that Box–Jenkins is more appropriate for short-range than long-range forecasts.

The Box–Jenkins method allows for much flexibility in the selection of a model. This appeals to analysts. On the other hand, it calls for much subjectivity on the part of the analyst. Granger and Newbold (1974) consider this need for subjective inputs to be a disadvantage relative to exponential smoothing. Chatfield and Prothero (1973) suggest that the need for subjective inputs is so great that instead of Box–Jenkins one might use BFE (bold freehand extrapolation). In fact, LAWRENCE et al. [1985] have tried that! In a test of the reliability of the Box–Jenkins method, LUSK and NEVES [1984] obtained substantially different B-J models from those identified by Andersen for 111 time series in the M-Competition (though the overall accuracy did not differ much).

The detailed procedure becomes quite complicated. Systematic procedures are used to determine the best model for the historical data. These examine autoregression with previous values of the time series, differencing of successive values, and moving averages. A good description of these procedures is provided in MAKRIDAKIS, WHEEL-WRIGHT, and McGEE [1983, Chapter 9].
The Box—Jenkins procedures have had an immense impact on forecasting. Box—Jenkins is clearly the most widely discussed and the most highly cited work in forecasting. Theoretically, it sets the standard. Practitioners and researchers tend to believe that it is the most accurate approach [FILDES and LUSK, 1984]. But the empirical evidence has been less convincing.

Results favorable to Box—Jenkins were obtained by Newbold and Granger (1974), REID [1975], BROWN and ROZEFF [1978], and DALRYMPLE [1978]. No differences were found in Chatfield and Prothero (1973), Groff (1973), Geurts and Ibrahim (1975), MA-BERT [1976], CHATFIELD [1978], KENNY and DURBIN [1982], ALBRECHT et al. [1977], MAKRIDAKIS and HIBON [1979], and MAKRIDAKIS et al. [1982]. Less accuracy was found in BRAN-DON, JARRETT, and KHUMAWALA [1983].

Thus, in only four of the fourteen comparative studies that I found was it reported that a higher degree of accuracy was achieved by using Box—Jenkins procedures. On the other hand, Box—Jenkins was less accurate only once. The most impressive evidence was provided in the large scale comparisons by MAKRIDAKIS and HIBON [1979] and MAKRIDAKIS et al. [1982, 1984]. (For the latter study, see also the discussion in ARMSTRONG and LUSK [1983].) Some of the comparative studies are described here:

A discussant to Chatfield and Prothero (1973) obtained forecasts from a simple naive model that were comparable to those from Box—Jenkins.

Groff (1973) examined monthly data for 63 time series representing factory shipments of automotive parts and drug items. There were also five simulated time series. The 1-month and 6-month forecasting errors from the best of ten Box—Jenkins models were slightly greater than the errors from exponentially smoothed models. However, overall, there was little difference among the forecasting errors of the various methods.
Newbold and Granger (1974) compared 1- to 8-month forecasts by exponential smoothing and by Box–Jenkins for 106 economic time series. The Box–Jenkins forecasts were superior on 60% of the comparisons. On the average, Box–Jenkins errors were 80% as large as those from exponential smoothing. The Box–Jenkins forecasts did relatively better for the shorter range forecasts.

Geurts and Ibrahim (1975) compared Box–Jenkins and exponential smoothing for 1-month forecasts of visitors to Hawaii. Each model was developed using data from 1952 to 1969, and the forecasts were made for the 24 months of 1970 and 1971. The forecast accuracy (Theil's $U^2$) for the two methods was almost identical.

**Regressions**

Still another method of analyzing time series data is to run a regression, using time as the independent variable. This method, which weights all of the historical data equally, provides estimates of both current status and trend. If an additive trend factor is used,

$$Y = a + bt$$

The forecasting accuracy of regressions against time appears to be slightly inferior to that of exponential smoothing.

Wagle, Rappoport, and Downes (1968) found little difference, although it is difficult to draw any firm conclusions because their description is inadequate. Kirby (1966), found regression to be slightly inferior to exponential smoothing. Hirschfeld (1970) found regression to be less accurate than exponential smoothing in predicting the growth rates of 10 white males from age 6 to 16. Finally, Elton and Gruber (1972) found regression to be inferior to exponential smoothing for earnings forecasts, especially as the forecast horizon increased from 1 to 3 years.
Exhibit 7-7 RANKINGS OF EXTRAPOLATION METHODS
(1 = highest ranking)

<table>
<thead>
<tr>
<th>Method</th>
<th>Cost</th>
<th>Understandability</th>
<th>Forecast Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential smoothing</td>
<td>1</td>
<td>2.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Moving averages</td>
<td>2</td>
<td>1</td>
<td>3.5</td>
</tr>
<tr>
<td>Box-Jenkins</td>
<td>4</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>Regressions</td>
<td>3</td>
<td>2.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Comparisons of Extrapolation Methods

These primary methods for extrapolation of time series data are listed in Exhibit 7-7, where they are ranked on the criteria of cost, understandability, and accuracy. Exponential smoothing is a less expensive forecasting method because it does not require storage of historical data, as do moving averages and regression; Box-Jenkins requires many diagnostic tests to keep up with the changes. Moving averages are easy to understand, exponential smoothing and regressions are a bit more difficult, and Box-Jenkins procedures are still harder to understand because of their complexity. The rankings on accuracy are based on the previously presented evidence (including Appendix J). They are heavily influenced by the M-Competition and by my judgment. Exponential smoothing is expected to be most accurate overall. Under ideal conditions, moving averages can approximate exponential smoothing; however, they use data inefficiently for start-up. Regressions suffer from their sensitivity to outliers and from errors in measuring the current status. Note, however, that the differences in accuracy for short-range forecasting are usually small (except for regression). For longer term forecasts, these differences are expected to become larger; thus, I expect exponential smoothing, with dampened trend, to be superior. Quite likely, the relative accuracy of the methods depends upon the situation. This issue is discussed in the commentary to the M-Competition [see ARMSTRONG and LUSK, 1983], but to date I do not believe we have good guidelines.

One conclusion that emerges from the research on extrapolation is simplicity. Beyond the fairly obvious steps that were recommended as early as 1960 (clean the data, adjust for trading days, deseasonalize, and then estimate the average and the trend), methodological sophistication has been of little value. This statement is based on my review
of the empirical evidence. Of the studies that I found, 18 showed no gain in accuracy for sophistication beyond the 1960-type procedures, nine studies found simpler methods to be more accurate, and only five studies found improved accuracy for sophisticated methods. (These results were abstracted from the studies listed in Appendix J, LRF pp. 494–495. I did not include studies that had been challenged nor the studies that compared exponential smoothing to moving averages.)

The need for simplicity is especially important where few historical data exist or where the historical data are unstable. Interestingly, instability can be judged just as well subjectively as statistically:

SCHNAARS [1984] compared sales forecasts generated by six extrapolation methods for 98 annual time series. The methods ranged from simple (next year will be the same as this year) to complex (curvilinear regression). A significant omission was the simple exponential smoothing model (i.e., without trend). The forecast horizon ranged from 1 to 5 years and successive updating was used so that almost 1500 forecasts were examined. This is an interesting study. The simplest model performed well, especially where there were few historical data and where the historical series seemed unstable. Stability was assessed just as effectively by subjects who looked at historical scatter plots as by using autocorrelation or runs statistics. Models that squared the time variable were especially inaccurate. (It has always struck me as a bit strange to raise time to some power in an attempt to improve forecasting. It is a dangerous practice and I strongly recommend against it.)

SCHNAARS and BAVUSO [1985] compared the short-range predictive accuracy of seven extrapolation models on 15 economic indicators, simple smoothing was included as one of the models. The random walk (no change from last period) was, in general, the most accurate of the methods examined. The most complex methods produced the least accurate forecasts.

Cycles and Fancy Curves

Once upon a time in the land of Academia, there lived two mystics named Dewey and Dakin (1947). They believed in cycles. Not the two-
wheeler kind that you ride only once every 3 years after you grow up and then feel sore for a week. No, Dewey and Dakin believed in big panoramic cycles that explained man’s existence. They drew inspiration from astronomy, where cycles had been used successfully. They claimed that the world is so complex, relative to man’s ability for dealing with complexity, that a detailed study of causality is a hopeless task. The only way to forecast, they said, is to forget about causality and, instead, to find past patterns or cycles of behavior. These cycles should then be projected without asking why they exist. Dewey and Dakin believed that economic forecasting should be done only through mechanical extrapolation of the observed cycles. They emphasized that “the forecasts are written by the data themselves.”

Much of the Dewey and Dakin book is devoted to a description of how cyclical components can describe historical events. The ability of cycles to fit data for events that have already occurred is not, however, the primary concern. How useful are cycles in forecasting? Well, Dewey and Dakin (1947, p. 277) did make some forecasts. These forecasts were made in 1946 and covered the period from 1947 to 1952. They expected the economy to trend downward and to reach a low in 1952. As noted in Schoeffler (1955), these forecasts proved to be drastically wrong. The economic growth from 1947 to 1952 exceeded all previous peacetime records in the United States. (I’m not sure why Schoeffler called this peacetime—unless we were actively waging peace in Korea.)

With the exception of seasonality cycles, I have been unable to find empirical evidence on the predictive value of cycles. I even obtained information from the Foundation for the Study of Cycles, which is associated with the University of Pittsburgh. (It publishes the Journal of Interdisciplinary Cycle Research and Cycles.)

The added complexity of cycles makes this method risky for anything but short-range forecasting (Cowden, 1963, thought so too). Small errors in estimating the length of the cycle can lead to large errors in long-range forecasting if the forecast gets out of phase with the true situation. Of course, if little uncertainty exists about the length of a cycle, use this information. For example, the attendance at the Olympic games follows a four year cycle ... usually.

What have economists learned from this apparently fruitless quest? Well, spectral analysis is a hot topic. This is a complex approach that is similar to a regression against time, except that now the independent variables are sines or cosines or various powers of time. (See Brown, 1963, pp. 396–401 or Chan and Hayya, 1976, for descriptions of spectral analysis.)

Spectral analysis is a dance in a different dimension. It is a sophis-
ticated approach to the study of cycles. Advocates of spectral analysis apparently assume that previous failures with cycles resulted because the latter did not represent the complexity of the data.

Unlikely. A good rule to follow in extrapolation methods is to keep it simple. I have been unable to find any evidence that spectral analysis provides better forecasts.

Spectral analysis offers complex curves with no theory. There are a number of curves, both simple and complex, that can be justified on some *a priori* grounds. Discussions of various curves may be found in Harrison and Pearce (1972), Gilchrist (1976), and Johnston (1984). Despite many years of experience with various growth curves in market forecasting, it is difficult to provide generalizations. MEADE [1984] assesses the empirical evidence on this topic.

Remember to use complex forms only where there are good *a priori* reasons for doing so. An example of the use of *a priori* reasons in the selection of a curve is presented in the following. Despite all of this good thinking, a simple rule of thumb proved just as effective for prediction:

Armstrong (1975b), in a study on mail surveys, made predictions for the percentage reduction in nonresponse that would be achieved by monetary incentives. Four assumptions were made about the relationships:

1. The curve should go through the origin to reflect, by definition, the fact that no reduction in nonresponse is obtained with no incentive.
2. The curve should approach an asymptote (or limit). Obviously, the limit in this case cannot exceed 100%. An asymptote of 100% was selected under the assumption that “everyone has his price.” (This decision was subject to much uncertainty.)
3. There should be diminishing marginal returns: each additional amount of money should have less impact.
4. The relationship should be simple if possible.

One way to capture these assumptions is to fit the data to the following functional form:

\[ Y = l - \frac{l}{e^{bx}} \]
where $Y =$ the reduction in nonresponse bias

$l =$ the asymptote

$x =$ the monetary incentive

$b =$ the parameter to be estimated

$e =$ the base for natural logs*

This functional form, which did not turn out to be so simple, can be transformed to

$$\ln (l - Y) - \ln (l) = bx$$

Assuming an asymptote of 100% gives

$$\ln (100 - Y) - \ln (100) = bx$$

The $b$ was estimated via regression analysis, using 24 observations, to be 0.64. Thus the reduction in the percentage of nonresponse, $R$, can be predicted from

$$R = 100 - \frac{100}{e^{0.064x}}$$

The accuracy of this complex curve was matched by a simple extrapolation that said, "There is a 1% reduction in nonresponse for each 1 cent in incentive up to 40 cents." You can't imagine how disappointing such results are for an analyst! Analysts are like politicians; they would rather make simple things complex.

Studies in demography have made extensive use of curves for extrapolations. According to a review by Hajnal (1955), crude extrapolations have worked as well as complex ones. My impression from Dorn's (1950) review is that the more thinking the demographers put into their extrapolations, the more complex the methods become and the poorer the resulting forecasts.

Curves have been used in fields other than demography and survey research, of course, but the results are about the same. In contrast to short-range forecasting, the curves provide widely differing long-range forecasts. One conclusion is that simple curves are as good as complex

*This is one time I have lied in the glossary. In other cases, $e$ represents error.
ones. Another conclusion, when using curves for long-range forecasts, is "Be careful." Mark Twain already said this (*Life on the Mississippi*, chapter 17, passed along via Daly, 1963):

In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. That is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oölitic Silurian Period, just a million years ago next November, the Lower Mississippi River was upward of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-rod. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long, and Cairo and New Orleans will have joined their streets together, and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

P.S. Don’t use fancy curves to decide when to invest your money in the stock market (Levy, 1971b).

**COMBINED FORECASTS**

This chapter’s adventure in eclectic research was inspired by Levins (1966), a biologist. He suggested that, rather than building one master model of the real world, one might build several simple models that, among them, would use all of the data available, and then average them. This implies that gains can be achieved by using a set of simple extrapolation methods and combining the forecasts. This strategy has received impressive support:

Bates and Granger (1969) tested combined forecasts of the international air travel market. They started with the premise that the gains are greatest if the forecasts are based upon different information, rather than merely upon different assumptions about the form of the relationship. They used five different methods to obtain 1-month forecasts for the period 1951–1960. Nine different combined forecasts were developed from the five original sets of forecasts using various weighting schemes to combine pairs of forecasts. As an example, forecasts for the months in 1953 were
obtained by exponential smoothing and also an extrapolation by Box and Jenkins (taken from Barnard, 1963); the error variances of these forecasts were 196 and 188, respectively. A combined (average) forecast from these two methods had an error variance of only 150. Overall, the error variances of the combined forecasts were smaller than the error variance of either of the two components in all but one case (where the combined forecast and the best component tied). Therefore the combined forecasts were superior to the average of the individual forecasts. The gains from combining were greatest when the two original forecasts were relatively independent of one another. Additional evidence is available from MORRIS [1977], REINMUTH and GERUTS [1978], and BUNN [1979]. This evidence also favors combining.

MAKRIDAKIS and WINKLER [1983] examined the value of combining extrapolation forecasts using data that were selected from the 1001 series in the M-Competition [MAKRIDAKIS, et al., 1982]. Significant gains in accuracy were achieved as more forecasts were combined. Combinations of two methods led to a 7% reduction in MAPE. Significant error reductions occurred for the first five methods, after which the reductions were small. Combinations were more important for data measured on a shorter time interval (monthly data) than longer time (yearly data), presumably due to the lower reliability of estimate for the shorter time period. (See also WINKLER and MAKRIDAKIS [1983].)

Because the gains due to combining forecasts are expected to increase as the similarity between the forecasts decreases, and because extrapolation forecasts tend to diverge as the forecast horizon is lengthened, the gains for combined forecasts should be greater when the forecast horizon is longer. Long-range forecasts were examined in studies by Ogburn and by Armstrong:

Ogburn's study (1946, pp. 113–145) of the air travel market demonstrates that various extrapolation methods can provide widely varying predictions when large changes are expected. Using data through 1943, he extrapolated the 1953 revenue passenger miles (RPM) on U.S. domestic airlines by various methods. His forecasts
are presented here. I added a combined forecast based on an average of Ogburn’s five extrapolations, then compared the forecasts with actual RPMs of 14.9 billion in 1953.

<table>
<thead>
<tr>
<th>Method</th>
<th>RPM Forecast</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Straight line projection</td>
<td>4.0</td>
<td>73</td>
</tr>
<tr>
<td>b. Logistic curve</td>
<td>5.6</td>
<td>62</td>
</tr>
<tr>
<td>c. Parabola</td>
<td>6.0</td>
<td>60</td>
</tr>
<tr>
<td>d. Gompertz curve</td>
<td>13.0</td>
<td>13</td>
</tr>
<tr>
<td>e. Constant percentage growth</td>
<td>30.0</td>
<td>101</td>
</tr>
<tr>
<td>Average error for a, b, c, d, e</td>
<td>—</td>
<td>62</td>
</tr>
<tr>
<td>Combined forecast (all of above)</td>
<td>11.7</td>
<td>21</td>
</tr>
</tbody>
</table>

The combined forecast was more accurate than the forecast obtained with Ogburn’s favored method, which was the parabola, and also better than the average MAPE for Ogburn’s five methods, 21% vs. 62%.

In Armstrong (1968a), three extrapolations were used for a 6-year backcast of 1954 camera sales for 17 countries. The extrapolations, developed using data from 1965–1960, were done by a method assuming no changes in sales, one that forecasted a constant percentage trend within each country, and one that applied the worldwide market trend to each country. The two combined forecasts were more accurate.

<table>
<thead>
<tr>
<th>Extrapolation Method</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. No change</td>
<td>67</td>
</tr>
<tr>
<td>b. Constant trend within country</td>
<td>51</td>
</tr>
<tr>
<td>c. Constant world market trend</td>
<td>37</td>
</tr>
<tr>
<td>Average errors for a, b, and c</td>
<td>52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combined Forecasts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of a &amp; b</td>
<td>43</td>
</tr>
<tr>
<td>Average of a &amp; b &amp; c</td>
<td>33</td>
</tr>
</tbody>
</table>
I used equal weights in my reanalysis of these studies. When you are uncertain as to what weights to use, use equal weights. If you have strong evidence about which forecasts are likely to be most accurate, use this information to establish weights. Bates and Granger (1969) present a good discussion on how to weight, but studies to date have found few differences among weighting schemes.

**ASSESSING UNCERTAINTY**

A common approach to estimating uncertainty with extrapolation methods is to examine how closely a particular method fits the historical data. In other words, what is the typical error when comparing predicted and actual values for the data used to develop the model? The most important problem with this approach is that future environmental change makes the estimates less useful.

The estimation of uncertainty can be substantially improved if data are withheld and used to simulate the actual forecasting situation. These “let’s pretend” forecasting errors are used as estimates of the uncertainty for the future. For example, to estimate the uncertainty involved in a two-year forecast, you would like to know how well the method has done on other two-year forecasts. Williams and Goodman found this approach to be superior to the usual approach of examining the fit to historical data. Newbold and Granger also found this approach to be useful in estimating uncertainty:

Williams and Goodman (1971) examined monthly data on phones for homes and businesses in three cities in Michigan. The first 24 months of data were analyzed by a regression on the first differences of the data. Seasonality was also estimated. Forecasts were made for an 18-month horizon. The model was then updated, and another forecast was calculated; this procedure was repeated for 144 months of data. When the standard error for the historical data was used to establish confidence limits, the actual values were contained within the 95% confidence limits for about 84% of the forecasts. When confidence intervals were based instead on the forecast errors, the actual values were contained in the 95% confidence intervals for about 90% of the forecasts.
Newbold and Granger (1974, p. 161) examined 1-month forecasts for 20 economic series covering 600 forecasts. Of their Box–Jenkins forecasts, 93% fell within the 95% confidence intervals; of their regression forecasts, 91% fell within these limits.

The measures of uncertainty can be used to develop tracking signals. These can monitor the forecasting system and indicate when it is no longer appropriate. For discussions on tracking signals, see Gardner [1983a, 1985a].

**SUMMARY**

Historical data are useful for extrapolation if they are timely and accurate, and if the underlying process is expected to be stable in the future. If historical data are not available, a situation that occurs for large changes, one might examine historical data from analogous situations. If analogous situations do not exist, it may be necessary to use simulated data from either laboratory or field tests. Exhibit 7-1 ranked these four types of data (historical, analogous situations, laboratory simulation, and field test) as to appropriateness for estimating current status and for making short-range and long-range forecasts. Rankings were also provided on cost and the effects of researcher bias.

The method of Markov chains has been widely studied. It seems appropriate when there are various states of behavior that are related to one another. Unfortunately, little evidence could be found that Markov chains provide more accurate forecasts.

Exponential smoothing offers an inexpensive and simple approach for extrapolation. Exhibit 7-4 outlined an eight-step procedure for using exponential smoothing in forecasting. There were few surprises in this description, although the following points are worth repeating:

**Step 1. Clean Data**
(a) Remove or tone down the outliers.
(b) Combine multiple measures for a variable.
(c) Adjust for trading days.

**Step 2. Deseasonalize Data**
(a) Use multiplicative seasonal factors if you have ratio-scaled data, low measurement error, and large trends in the data. Otherwise, use additive factors.
(b) Use dampened seasonal factors rather than no seasonal or full seasonal factors. Increase emphasis on seasonality when you can obtain good estimates.

**Step 3. Select Smoothing Factors**
Forecast accuracy is not highly sensitive to the selection of smoothing factors. Use judgment to select a historical time span, and then derive a smoothing factor. Alternatively, search for the factors that minimize the forecast error in the historical data.

**Step 4. Calculate New Average**
The selection of starting values is important if data are limited. Use sensitivity testing to assess this problem.

**Step 5. Calculate New Trend**
(a) Use multiplicative factors if you have ratio-scaled data, low measurement error, and large trends in the data. Otherwise, use additive trends. In either case, dampen the trend.
(b) Generally speaking, ignore acceleration.

**Step 6. Estimate Current Status**
Adjust the new average to correct for lag.

**Step 7. Calculate Forecast**

**Step 8. Update Seasonal Factors**

**Return to Step 4 and Repeat.**
Keep things simple, especially if the historical series is short or unstable. Do not use adaptive smoothing constants.

Alternatives to exponential smoothing were discussed. The method of moving averages is similar to exponential smoothing, but it is slightly more expensive. Box–Jenkins is more difficult to understand and more expensive to use than exponential smoothing. From a theoretical viewpoint, it is most appropriate for short-range forecasts. These methods, along with regressions against time, were ranked on cost, understandability, and accuracy in Exhibit 7-7.

The use of various curves was examined. Complex curves, if used at all, should be limited to cases where uncertainty is low. To the extent that uncertainty is high, one should use simpler curves. For this reason, spectral analysis is particularly poor for long-range forecasting.