LONG-RANGE FORECASTING
From Crystal Ball to Computer
## Ten

**BOOTSTRAPPING AND OTHER COMBINED METHODS**

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guidelines for Combining Methods</td>
<td>272</td>
</tr>
<tr>
<td>Bootstrapping</td>
<td>274</td>
</tr>
<tr>
<td>Direct Bootstrapping</td>
<td>276</td>
</tr>
<tr>
<td>Indirect Bootstrapping</td>
<td>279</td>
</tr>
<tr>
<td>Accuracy of Bootstrapping</td>
<td>279</td>
</tr>
<tr>
<td>When to Use Bootstrapping</td>
<td>283</td>
</tr>
<tr>
<td>Econometric Methods Within Segments</td>
<td>284</td>
</tr>
<tr>
<td>Leading Indicators</td>
<td>287</td>
</tr>
<tr>
<td>Combined Forecasts</td>
<td>288</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>292</td>
</tr>
<tr>
<td>Summary</td>
<td>293</td>
</tr>
</tbody>
</table>
George Bernard Shaw is reported to have received in the mail a proposal for marriage in which a lady said, "With my beauty and your brains we could have the perfect child." Shaw replied, "But what if the child had my looks and your brains?"

Chapters 6 through 9 described various types of forecasting methods. The best approach to a given forecasting problem may, however, call for a combination of methods. The story about George Bernard Shaw is relevant: when different methods are combined, it is not always clear that they are being combined to utilize their best features. This chapter presents guidelines to help ensure that the combination will be beneficial.

After describing the general guidelines, detailed consideration is given to bootstrapping, the most important of the combined methods. This method has important implications for forecasters.

Brief descriptions are provided of econometric methods within segments and of leading indicators. The chapter concludes with a discussion of combined forecasts that rely upon eclectic research; very different methods are combined.

GUIDELINES FOR COMBINING METHODS

A major consideration in the combination of methods is which method to use first. Two rules seem important in specifying time priorities. The first (and more important) is that subjective methods should precede objective methods. Auditors of financial forecasts believe this to be good practice, according to DANOS and IMHOFF [1983] and so do weather forecasters [MURPHY and BROWN, 1984]. The second rule is that segmentation should precede extrapolation or econometric methods. These rules are summarized in Exhibit 10-1.

Exhibit 10-1 PRIORITIES FOR COMBINING METHODS

<table>
<thead>
<tr>
<th>Extrapolation</th>
<th>Econometric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgement</td>
<td>Segmentation</td>
</tr>
</tbody>
</table>
A substantial amount of evidence was presented earlier (especially in Chapter 4) that subjective methods should precede objective methods. (The subjective inputs were described under the designation “the a priori analysis.”) This recommendation does not lessen the value of the individual’s subjective inputs. Instead, it utilizes these inputs in a more effective way.

Arguments for using subjective inputs after the objective analysis have come from “practical” econometricians and from businesspeople. They are convinced that their subjective adjustments improve the accuracy of the forecasts. Hence businesspeople persist in such wasteful and prejudicial practices as using personnel managers to interview prospective employees (to make predictions about their probable success) after objective data have been received from these interviewees. Businesspeople and econometricians revise sales forecasts that have been produced by objective methods. Not only is this bad practice, but it also leads to poor forecasts. For a vivid example, see GLANTZ [1982].

A study by Strong provided evidence consistent with this viewpoint:

<table>
<thead>
<tr>
<th>Reply</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not use judgment</td>
<td>6.8</td>
</tr>
<tr>
<td>Used judgment</td>
<td>8.5</td>
</tr>
<tr>
<td>Emphasized use of judgment</td>
<td>9.6</td>
</tr>
</tbody>
</table>

Some evidence has been presented by econometricians to suggest that subjective adjustments improve short-term economy-wide forecasts. Reinmuth and Geurts (1972) claimed to have found advantages to subjective adjustments, but their study was based on a single forecast (it is bad practice to generalize from a sample of one). CARBONE, et al. [1983] found no advantage to subjective adjustments of extrapolations. In Harris (1963), subjective adjustments made things worse. As suggested in Chapter 8, such adjustments can correct for errors in the estimate of current status, and they do not help to forecast change.
The recommendation that segmentation should precede extrapolation and econometric methods is based on common sense.

The discussions of methods in Chapters 6 through 9 suggested another important consideration in the combination of methods: different methods should be considered for different parts of the problem. In particular, different methods are of value in estimating current status, in forecasting short-range changes, and in forecasting long-range changes. Subjective methods are most appropriate for current status. The Wizard of Id illustrates this (Exhibit 10-2). Extrapolations are most relevant for short-range changes. Segmentation methods serve best for long-range changes. Finally, econometric methods are useful for estimating current status and for long-range forecasting. Exhibit 10-3 suggests which methods are appropriate for each time span. (Further empirical evidence relating to these recommendations is presented in Chapter 15.)

BOOTSTRAPPING

Frederick W. Taylor, the father of scientific management, created a bit of a sensation with his studies in the early 1900s. He applied scientific management to blue-collar jobs. The idea was that, by observing how a job was done, one could find ways to improve job performance. For example, he showed that a good way to find out how to improve a shoveler's shoveling is to observe him as he shovels. Scientific management, as defined by Taylor, involved telling others how to work
more efficiently. You can imagine how happy that made management. Labor was a bit less enthusiastic about the idea.

The idea that scientific management could be applied to managers is something else again. Richard Cyert and James March, well-respected members of the academic community, presented the radical notion that economists might learn something by watching people make decisions rather than by assuming what the rational person would do (Cyert and March, 1963). The Wizard agrees, as you can see in Appendix I of LRF.

What about you doctors, lawyers, generals, stockbrokers, priests, psychologists, and management scientists? Could your job be studied scientifically ... and be improved? More specifically, what about the predictions that you make; could a computer replace you for these predictions? The answers are “almost certainly” and “yes.” (You thought maybe I was talking about someone else? Perhaps someone who works for you could be replaced by a computer, but . . .)

Incidentally, Taylor (1911) did not think his method would be applicable to thinking-type jobs. He thought it would be most useful for low-level occupations such as pig-iron handling, where the ideal worker, Taylor said, “is so stupid that the word ‘percentage’ has no meaning to him, and he must consequently be trained by a man more intelligent than himself. . . .” Taylor knew how to get those managers on his side.

Bootstrapping assumes that the decision process used by people can be made explicit. In a sense, one can develop a “model of man” (or a “model of woman” if you are in a womanagement science). This isn’t surprising, but it is surprising that the model of man is generally

---

**Exhibit 10-3 METHODS FOR ESTIMATING CURRENT STATUS AND FORECASTING CHANGE**

<table>
<thead>
<tr>
<th>Estimating Current Status</th>
<th>Short-Range Forecasts</th>
<th>Middle-Range Forecasts</th>
<th>Long-Range Forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judgmental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Econometric</td>
<td>Econometric</td>
<td></td>
<td>Segmentation</td>
</tr>
</tbody>
</table>
superior to the man himself at forecasting! Thus the method is called bootstrapping: the forecaster can lift himself up by his own bootstraps. The concept almost seems impossible because what comes out of the model is superior to what went into the model.

Bootstrapping follows the guidelines on the combination of forecasts. Subjective methods are used before objective methods. The subjective method in this case provides the rules, and the objective method applies the rules. This is an intuitively pleasing combination (Einhorn, 1972a, presents evidence in favor of such an approach). The decision maker thinks that he knows how to make the forecasts, and he wants things done his way; in fact, his inputs are used to do this. The objective model does the repetitive work by applying the judge’s rules without getting tired or irritable. This consistency seems to outweigh the losses that occur in going from a complex judgmental method to a simple objective method.

There are two approaches to bootstrapping: direct and indirect. These approaches are described below. A discussion follows on the effectiveness of bootstrapping and the situations in which bootstrapping is appropriate.

### Direct Bootstrapping

Following Frederick Taylor’s approach, direct bootstrapping involves a translation of the judge’s rules into a model. In other words, the subjective process is made explicit and operational.

Often it is sufficient to ask the judges what rules they are using. Evidence that such an approach can successfully replicate expert’s decisions is provided in finance by LARCKER and LESSIG [1983], and by the following two studies:

---

Kort (1957) analyzed 14 U.S. Supreme Court decisions on “right to counsel cases,” from 1932 to 1947. There were 9 “pro” and 5 “con” decisions. The judges provided detailed records on why they decided as they did. Using the factors that the judges said were important, Kort was able to develop a mechanical weighting scheme to forecast similar cases from 1947 to 1956. His forecasts were correct for the 12 cases that were examined.
Amstutz (1967) programmed the decision rules used by the skilled union members of the American Medical Association (commonly known as doctors) as they prescribed drugs for patients. A panel of experts was unable to tell which drugs had been prescribed by the doctors and which had been prescribed by the computer (this is the Turing test).

Sometimes judges are not aware of how they make decisions. In such cases, one might ask the judge to think aloud while they make predictions. The researcher records this thinking and translates it into specific rules. Often the rules are stated as questions that can be answered by “yes” or “no.” (Some people call this the use of protocols, but I don’t think Webster would use that term.) This procedure was used to match the decisions made by a professional investor in the Clarkson study.

Clarkson (1962) developed a model of an investment trust officer who selected stocks. The stocks selected by the computer program were similar to those selected by the trust officer. I do not believe that this small sample (one subject) study was ever replicated.

The “think aloud” approach is more expensive than just asking people what rules they use. Clarkson’s study was a major undertaking, although he analyzed only one trust officer. Compare this with McClain (1972), who found that bootstrapping information could be obtained in about 3 hours by asking a doctor directly how she made decisions. The first attempt then, in direct bootstrapping, should be to ask the judge to describe the rules she uses. This strategy is widely used in research on experts. In the natural sciences and engineering, the term expert systems is used, rather than bootstrapping. For a review of the expert systems literature, see DUDA and SHORTLIFFE [1983].

Indirect Bootstrapping

Indirect bootstrapping starts with the judge’s forecasts and works backward to infer what rules the judge used to make these forecasts.
Indirect bootstrapping has had a long career. It started in 1923 when Henry A. Wallace (1923) developed a model of people who rated the quality of corn. (This is the same Wallace who later became Secretary of Agriculture and then Vice President of the United States.)

The basic approach to indirect bootstrapping is to develop a model where the judge’s forecasts are themselves predicted by a quantitative model. In other words, the model uses the judge’s forecasts as the dependent variable, and the variables that the judge used serve as the causal variables. Generally, this step is accomplished by regression analysis. The procedure is the same as described in Chapter 8, and the model looks the same:

\[ Y' = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n \]

except that \( Y' \) represents the judge’s forecasts rather than the actual outcomes.

Research on bootstrapping has tried to capture the complexity of the judge’s rules. This line of research yielded little. For example, Cook and Stewart (1975) examined seven different ways of obtaining the weights in a bootstrapping model and found that it did not make much difference which one was used. SCHMITT (1978) replicated this study. The conclusion is to use a simple linear model. This conclusion was supported by Goldberg (1968a, 1971), Heeler, Kearney, and MeHaffey (1973), Slovic, Fleissner, and Bauman (1972), and Wiggins and Hoffman (1968).

Indirect bootstrapping can generally be competitive in cost with direct bootstrapping. Evidence is mixed on the relative accuracy of these approaches. LARCKER and LESSIG (1983) and LEIGH, MacKAY, and SUMMERS (1984) found a slight advantage for the direct approach; Summers, Taliaferro, and Fletcher (1970) found no difference; GRAY (1979), NESLIN (1981), and SCHMITT (1978) found indirect bootstrapping to be superior. Perhaps the key consideration is whether the forecaster has a good awareness of the process. For example, CO-COZZA and STEADMAN (1978) found that psychiatrists did not have a good understanding of how they made predictions about the potential dangerousness of defendants in court cases. In such cases, the indirect approach is preferable.

Applications of indirect bootstrapping are found in many areas. One of the most popular uses has been to represent the judgments of consumers. For descriptions of this work, typically referred to as part of “conjoint analysis,” see CATTIN and WITTINK (1982) and GREEN and WIND (1975).
Accuracy of Bootstrapping

The evidence cited above suggests that the bootstrapping model provides a good representation of the judge's decisions. But what about forecasting? Does the bootstrapping model provide good forecasts? Much research has been done in this area. It has been carried out independently by researchers with different backgrounds who were studying different problems. The only thing that has not changed is the conclusion: the bootstrapping model is at least as accurate as the judge. In most cases, the bootstrapping model is more accurate than the judge. Sometimes there are ties. Never has the typical judge been significantly better.

In view of the surprising nature of these findings and their importance, an extensive summary of the evidence is provided. All the bootstrapping models were indirect, unless otherwise indicated:

Yntema and Torgerson (1961) provided pictures of 180 ellipses to six subjects. The ellipses were various combinations of six sizes, six shapes, and five colors. Subjects were asked to judge the "worth" of each ellipse. The problem had been constructed so that worth always increased with size, thinness, and brownness, although these were not linear relationships. The subjects were trained with 180 ellipses on each of 11 days. The order of the ellipses was varied each day, and the subjects got feedback after each trial on the correct worth. On the twelfth day, the judge evaluated all 180 ellipses with no feedback. The ellipses were also evaluated using a bootstrapping model for each judge. The average $R^2$ between the judge's evaluation and the true worth was .71; the average $R^2$ between the bootstrapping model's prediction and the true worth was .79. In addition, a deductive bootstrapping model was constructed by asking the judges what weights they placed on size, shape, and color; this bootstrapping model performed as well as the inductive bootstrapping (average $R^2 = .79$). Shepard (1964) reported on a follow-up by Pollack (1962) that successfully replicated the Yntema-Torgerson results.

Bowman (1963) examined ice cream, chocolate, candy, and paint companies. A regression analysis of management's decisions on
production and work force would have led to improvements over
the decisions actually made in three of the four situations. He
presented the following data on costs:

<table>
<thead>
<tr>
<th>Costs</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ice Cream</td>
</tr>
<tr>
<td>Actual</td>
<td>105.3</td>
</tr>
<tr>
<td>Using Bootstrapping Model</td>
<td>102.3</td>
</tr>
</tbody>
</table>

It is difficult to understand how Bowman arrived at his findings
due to the description being incomplete. Similar results were ob-
tained for a brewery in a study by J. R. M. Gordon (described
briefly in Buffa and Taubert, 1972).

Kleinmuntz (1967) developed a direct bootstrapping model. He
coded the rules used by one expert to identify the best-adjusted
and worst-adjusted subjects who sought counseling at college.
Data from a personality inventory (the MMPI) were used. The
bootstrapping model did as well as the best of eight clinicians in
predictions based on five cross validation samples from different
colleges.

Kunreuther (1969) developed a bootstrapping model for short-
range production forecasting in an electronics firm. The model,
developed partly from direct and partly from indirect bootstrap-
ping, was a simple two-variable model. According to Kunreuther,
this model would have enabled the firm to carry a 25\% smaller
inventory, while improving service to customers.

Michael (1969, 1971) used direct bootstrapping (which he called
a heuristic method) to forecast sales for items in a mail order
catalogue. He studied the rules used by a man who had been
making these forecasts for a company, and developed a boot-
Bootstrapping

strapping model to forecast sales for the coming season for 42 items. He concluded that experts were unable to distinguish which forecasts had been made by the bootstrapping model and which by the judge. The accuracies of the two forecasts were also comparable; in fact, the bootstrapping model did slightly better than the judge.

In Goldberg (1970), 29 judges used scores from the MMPI to differentiate between psychotics and neurotics among a sample of 861 patients. A bootstrapping model was developed for each judge, using a portion of the sample, and the rest of the sample was used for testing. The bootstrapping models were more accurate than 86% of the judges.

Wiggins and Kohen (1971) asked 98 graduate students in psychology to forecast first-year grade-point averages for 110 students entering graduate school. A bootstrapping model was developed for each judge. The bootstrapping model was superior for each of the 98 judges; furthermore, the typical bootstrapping model was better than the best of the 98 judges, and also better than the combined forecast by the 98 judges.

Dawes (1971) examined the admissions decisions for the Ph.D. program in psychology at the University of Oregon. There were six categories for rating applicants: (1) reject now; (2) defer rejection but looks weak; (3) defer; (4) defer acceptance but looks strong; (5) accept now; (6) offer fellowship. The committee used traditional measures in making its judgments, such as scores on Graduate Record Examination (GRE), quality index (QI) of school where undergraduate degree was received, grade point average (GPA), letters of recommendation, and record of work experience. A simple bootstrapping model (a regression of admissions committee decisions vs. GRE, QI, and GPA) reproduced the committee’s decisions; the $R^2$ was .78, and none of the applicants rated in the lower 55% by the bootstrapping model were actually admitted by the admissions committee. Faculty evaluations after
the first year were used as a measure of student success in the program. The bootstrapping model provided a more accurate ranking of student success than that provided by the admissions committee (Spearman rank correlation coefficients of .51 and .10, respectively). DAWES [1979] provided a follow-up.

Moskowitz and Miller (1972) provided another example in the field of production management. Eighty-six managers were presented with a simulated production problem and were asked to make production and work force decisions for one and three periods in the future. The forecasting error was varied during the experiment from low to medium to high. The bootstrapping model led to better decisions than the manager himself had made for both forecast horizons and for all three levels of forecast error. In no case was the manager superior to his model. Carter and Hammer (1972) replicated this study with similar results; in addition, they found that the combined decision from a number of bootstrapping models was even more effective. Moskowitz (1974) presented similar results, but this report appears to draw upon the same data as Moskowitz and Miller (1972). MOSKOWITZ et al. [1983] added further support.

Evidence from personnel forecasting was provided by a study of 16 managers by ROOSE and DOHERTY [1976]. Bootstrapping yielded a small gain (vs. the average judge) in the accuracy of predictions for the success of 160 salesmen.

EBERT and KRUSE [1978] developed bootstrapping models for five analysts who forecasted returns for securities using information on 22 variables. Bootstrapping was more accurate than the analysts for 18 of 25 comparisons, and it was as accurate as an econometric model.

In ABDEL-KHALIK [1980], mechanical models made more accurate predictions of defaults on loans than did 29 bank officers.
CAMERER [1981] concluded from his review that “bootstrapping will improve judgments under almost any realistic task conditions.”

Somebody (actually it was Libby, 1976a) once thought that he had found an occasion when the judges defeated their models. The case involved the prediction of bankruptcy for 60 large industrial corporations. Goldberg (1976) showed, however, that Libby’s result was due to severe skewness in the data. When the data were normalized, the percentage of times that the model beat the judge went from 23 to 72%. Another challenge comes from the small sample study by FILDES and FITZGERALD [1983].

The success of the bootstrapping model was not often sensitive to the type of regression analysis used, nor was it sensitive to the estimates of the magnitude of the relationship. That is a familiar story, because the same thing was observed in Chapter 8 for econometric models. Furthermore, Dawes and Corrigan (1974) reanalyzed the data from Yntema and Torgerson (1961), Goldberg (1970), Wiggins and Kohlen (1971), and Dawes (1971) and found that a unit weights model did better than the bootstrapping model, though REMUS [1980] presented contradictory evidence. This too has a familiar ring to it. You do not even have to bother to run a regression. Merely specify signs for the relationships and you are in business. William of Occam wins again!

When to Use Bootstrapping

Bootstrapping offers advantages in a number of situations.

1. It is inexpensive (Robinson, Wahlstrom, and Mechan, 1974). This is relevant for jobs with repetitive predictions (e.g., doctors, lawyers, stockbrokers, judges, university administrators).
2. It provides a first step for the introduction of objective forecasting
models. Let's face it. Managers are in favor of change when it affects other people, but they are like everyone else when the change strikes home. They may not take kindly to suggestions that they can be replaced by a quantitative model, but the resistance might be lower if they can be sure that their model will be used to make the forecasts.

3. It provides a quantitative model in cases where no data exist for the dependent variable. You could use "what if" questions to develop a direct bootstrapping model with signs and unit weights. If the judge was unsure about answering abstract questions, you could create hypothetical data for the situation and ask the judge to make predictions. These predictions would be used to develop an indirect bootstrapping model. Such an approach was used by Christal (1968) to develop a model for officer promotions in the U.S. Air Force. Bootstrapping can help to predict the success of new products (e.g., PARKER and SRINIVASAN [1976]), the best treatment for someone with lower back pain, the results of a proposed organizational change, or the outcome of a new social program.

4. It gives the decision maker a better understanding of the rules that she has been using. Experience sometimes fails to provide this insight [ROOSE and DOHERTY, 1976]. This may also highlight areas of prejudicial decision making.

5. If the forecasts are currently made by an individual rather than a group, bootstrapping is likely to produce bigger gains in accuracy [ROOSE and DOHERTY, 1976].

Exhibit 10-4 summarizes this advice on when to use bootstrapping.

<table>
<thead>
<tr>
<th>Exhibit 10-4 WHEN TO USE BOOTSTRAPPING— AND WHY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Repetitive forecasts —— cheaper than judgment</td>
</tr>
<tr>
<td>2. First step toward objective method —— judge keeps control</td>
</tr>
<tr>
<td>3. No objective data on dependent variable —— allows for quantification</td>
</tr>
<tr>
<td>4. To examine judgmental forecasts —— reveals prejudice</td>
</tr>
</tbody>
</table>

**ECONOMETRIC METHODS WITHIN SEGMENTS**

One common strategy for forecasting is to group the data into homogeneous segments, and then to develop an econometric model within each segment. This strategy has been used for many years. For ex-
ample, Seashore (1961) found that women's grades in high school and college were more predictable than men's. It is also found in the large scale econometric models. The combination of econometric and segmentation approaches is most useful when:

1. There is interaction (more specifically, a variable affects different groups in different ways).
2. The objective data are limited.

The effects of interaction are handled by segmentation, but this quickly depletes the data. Econometric methods use the remaining data more efficiently within the segments. The approach is illustrated by the hypothetical example in Exhibit 10-5.

Exhibit 10-5 ILLUSTRATION OF ECONOMETRIC MODELS WITHIN SEGMENTS: FORECASTING SALARIES

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Race</th>
<th>Education</th>
<th>Econometric Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Over 65</td>
<td></td>
<td></td>
<td>$Y_1 = a_1$</td>
</tr>
<tr>
<td></td>
<td>18-65</td>
<td>Female</td>
<td>College</td>
<td>$Y_2 = a_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Noncollege</td>
<td>$Y_3 = a_3 + b_4x_4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>White</td>
<td>$Y_4 = a_4$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nonwhite</td>
<td>$Y_5 = a_5 + b_6x_6$</td>
</tr>
<tr>
<td></td>
<td>Over 65</td>
<td>Male</td>
<td>College</td>
<td>$Y_6 = a_6 + b_1x_1$</td>
</tr>
<tr>
<td></td>
<td>18-65</td>
<td></td>
<td>Noncollege</td>
<td>$Y_7 = a_7 + b_2x_2 + b_3x_3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>White</td>
<td>$Y_8 = a_8 + b_4x_4 + b_5x_5$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nonwhite</td>
<td>$Y_9 = a_9 + b_6x_6 + b_7x_7$</td>
</tr>
</tbody>
</table>

Note:

$Y =$ yearly salary

$x_1 =$ percentage of jobs in city with mandatory retirement age of 65 or less

$x_2 =$ size of city

$x_3 =$ age

$x_4 =$ years of seniority in latest job

$x_5 =$ union membership (1 = yes, and 0 = no)

$a_i =$ constants
Salomon and Brown (1964) claimed that an approach using econometric models within each of 15 market segments led to more accurate forecasts than a "traditional method." Unfortunately, they did not provide sufficient information on what they did. Seven studies did allow for a comparison, and the results support the use of econometric methods within segments:

Elton and Gruber (1971) used econometric methods within segments to forecast the growth rate in earnings per share for 180 firms. They used a factor analysis of recent earnings data to group firms with homogeneous patterns of earnings growth. One-year forecasts were then made for 1964 earnings for firms in each of the ten homogeneous groups by using an econometric model. These forecasts were superior in accuracy to those provided by mechanical extrapolations of earnings by firm. However, another way of grouping (by SIC code) was also used, and these forecasts were inferior.

Kinney (1971) compared earning forecasts generated by two models:

1. Total earnings for the firm were multiplied by the forecasted percentage growth in GNP.
2. Earnings within each market segment of the firm were multiplied by the predicted percentage growth in the appropriate industry.

Model 2, the econometric method within segments, was superior for one-year forecasts. For 24 companies in 1968, the MAPE for the aggregate data model (model 1) was 17.2, while that for the model by segment was 13.4. For 19 companies in 1969, the respective figures were 11.6 and 11.2. Overall, the superiority of the model by segment was statistically significant ($p < .05$).

Raubenheimer and Tiffin (1971) developed an econometric model to predict job success for clerical workers. A model using 224 subjects had an $R^2$ of 6% when predicting success for 110 new subjects. The authors then broke the sample into three segments. The $R^2$ for each segment in the 110-subject validation sample was
Leading Indicators

7, 16, and 23%; that is, predictive ability was better for each segment than for the total sample.


Another approach to using segmentation and econometric methods is to develop an econometric model after accounting for the effects of the segmentation. The dependent variable is the deviation between the actual value and the mean value for the segment, and the causal variables are the ones not used in the segmentation. In other words, the econometric model is used to predict residuals from the segment mean for each observation. This is an econometric method across segments. It assumes that individuals in all segments respond in the same manner to the variables in the econometric model. Evidence on the value of the approach is lacking, but an example is provided by Palmore:

Palmore (1969), using a sample of people between the ages of 60 and 94, started with actuarial tables from insurance companies. These broke the population into homogeneous segments on the basis of age, sex, and race. A regression analysis was used to explain differences in a longevity quotient that had been obtained by dividing the number of years a person lived by the expected number of years for a person in that segment. The econometric model explained only a small portion of the variance in these differences ($R^2 = .17$). (As a sidelight, it was found that the factor most closely related to longevity for old men was “work satisfaction.” Involuntary retirement may lead to early death.)

LEADING INDICATORS

Leading indicators are obtained by exploring the data until you happen to find some variables whose trends precede similar trends in other variables. I would have no qualms about this if the search were for
causal variables. But the idea of just using whatever happens to work strikes me as strange. This is not an extrapolation method because it uses other variables, and it is not an econometric model because there is no concern for causality. What is it then? Let me tell you by way of an analogy how leading indicators relate to extrapolation. An extrapolation model is like looking for a lost horse by following the direction in which it started out. Leading indicators are concerned with factors that often vary according to the presence or absence of horses—like flies, for example. If you can find the flies, maybe you can then find the horse nearby. Of course, you might merely find some horse shit.

I would not discuss leading indicators were it not for all of the historical interest in this approach. Possibly, leading indicators can serve as additional measures of current status. Most variables in the economy move together, so you can gain reliability in your assessment of current status by grouping a set of indicators. It is better to group variables that are expected to measure the variable of interest, however. Do not use noncausal leading indicators to forecast change. This is contrary to theory, and empirical evidence is lacking, except for the favorable results in AUERBACH [1982].

COMBINED FORECASTS

Eclectic research suggests the use of methods that are as different as possible. Use each method independently, then combine the forecasts. This procedure has been recommended in business forecasting for many years (National Industrial Conference Board, 1963, Chapter 8, and Wolfe 1966, p. 21). Reichard (1966, p. 202) and PoKempner and Bailey (1970) claimed that combining is common practice among business forecasters. However, DALRYMPLE's [1985] survey found that most firms either did not use combined forecasts or they used them infrequently. Combining is illustrated in Exhibit 10-6.

As an example of the use of Exhibit 10-6, consider a 10-year forecast of air travel in the United States. A judgmental method, such as a mail survey of experts outside of the airline industry, might be used for F1; a simple extrapolation of the percentage change over the past 13 years could be used for F2; a segmentation study based on a survey of traveling habits for the various types of air travelers might be used for F3; and an econometric model employing population, income per capita, prices of air travel, prices of substitutes, and speed of service might be used for F4.

The combined forecast is a weighted average of the different fore-
Combined Forecasts

Exhibit 10-6 COMBINED FORECASTS FROM DIFFERENT METHODS

Judgement method \( F_1 \)

Extrapolation method \( F_2 \)

Segmentation method \( F_3 \)

Econometric method \( F_4 \)

Combined forecast

\[ w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 \]

Note: The w's, the relative weights on various forecasts, should sum to 1.0.

casts, with the weights reflecting the confidence that the researcher has in each method. The weights should be selected before generating the forecasts to reduce the possibility of bias by the researcher. The use of a simple mechanical rule, such as providing equal weights for each forecast, is adequate in most situations.

Although the concern has been primarily with unintentional bias, it should be recognized that researchers, like politicians, doctors, professors, and businesspeople, apparently cheat once in a while (ARM-STRONG [1983d]). The cheating occurs because of organizational pressures, as in the testing of an airbrake by B. F. Goodrich (Heilbroner, 1972); because of peer group pressure (e.g., Dr. William Summerlin of Sloan-Kettering Institute, who painted laboratory mice to make them appear as if they had received successful skin grafts, or Sir Isaac Newton, who was expected to obtain more precise results; see Westfall, 1973); or because of payoffs. There is a fine line between payoffs and consulting. Often it is obvious that we, as consultants, are being asked to give not our expert opinion, but instead the forecast desired by the organization.

The combined forecast, done independently by different researchers, provides a safeguard against cheating. Using different researchers offers the following advantages:

1. You are more likely to include an honest researcher.
2. You make it difficult for cheating to occur. It is not easy for different researchers to come up with the same mistakes.
Of course, there are disadvantages if you already know what forecast you want. (But if you did, you would not have read this far.)

To assess the value of combined forecasts, I searched the literature for studies that combined forecasts from different methods. FALCONER and SIVESIND [1977] found significant advantages as their combination of an econometric and an extrapolative forecast was 51% better than the average of the components and 44% better than the best component. NELSON [1984] also found gains in accuracy from weighting macroeconomic forecasts from extrapolation and econometric models. RAUSser and OLIVEIRA [1976] found that a combined forecast based on an average of an econometric and extrapolation forecast was better than either component. However, these studies used ex post forecasts. Schmitt (1954) claimed similar gains in population forecasting, but few details were provided. Granger and Newbold (1974) examined forecasts for three quarterly macroeconomic series in the United Kingdom; a combination of Box-Jenkins and econometric forecasts was generally more accurate than the average error of the components for one-period forecasts, but the weights were found retrospectively.

The most appropriate tests were those that used ex ante forecasts and provided sufficient descriptions to allow for a test of combined forecasts. In most of these studies it was necessary to reanalyze the results. In doing so, I used equal weights for each forecast. The studies are summarized here (see also the study by FILDES and FITZGERALD, 1983):

My reanalysis of a 10-year forecast of air travel passengers in the United States by the Port of New York Authority (1957) found that a simple unweighted combination of a segmentation and an econometric forecast did no better than the average forecast error.

I reanalyzed Okun's (1960) study of housing starts in the United States, where six one-year forecasts had been presented. There were two intentions forecasts and two extrapolations. This allowed for the calculation of four different combined forecasts. Small improvements were found for three of the four combined forecasts (the fourth forecast had the same error as the average for the components).
I reanalyzed Levine's (1960) study on forecasting capital expenditures by U.S. firms. Five different one-year forecasts were examined. A combined forecast was developed by averaging the Commerce Department–SEC judgmental forecast and the extrapolation based on no change from the preceding year; the MAPE was 6.4%. This compared with an average MAPE of 6.8% for the two components. When the combination was based on an average of the McGraw-Hill and the no-change forecasts, the MAPE was 5.9 vs. an average forecast error of 6.2 for the components.

In Vandome's (1963) quarterly forecasts of the United Kingdom's economy, an econometric forecast missed by 6.2% and an extrapolation model missed by 5%, for an average error of 5.6%. A combined forecast missed by 5.2%.

A reanalysis of O'Herlihy et al.'s (1967) five-year forecasts for five product areas used a combination based on an econometric model and a five-year extrapolation of growth in percentage terms. There was no improvement of combined over average forecasts for car exports, consumer durables, or energy, but the five-year combined forecast of domestic auto demand was better than the average (7.1% vs. 10.1%), and the combined forecast of coal consumption, with its 1% error, was better than either the econometric (+6.8%) or extrapolation (−8.7%) model.

In my study of the camera market (Armstrong, 1968b), a combined forecast was developed using the average of an extrapolation model and an econometric model. Six-year backcasts for 17 countries were examined; improvements over the average backcast were modest as the MAPE was reduced from 33% to 31.6%.

In BRANDON, FRITZ, and XANDER (1983), I first calculated the root mean square error. After obtaining an average RMSE of the four econometric forecasts, I combined it with a Box–Jenkins forecast. This combined forecast had an RMSE that was 8.6% less than its components.
Bootstrapping and Other Combined Methods

Exhibit 10-7 ERROR REDUCTION BY COMBINING FORECASTS FROM DIFFERENT METHODS
(All ex ante forecasts using Mean Absolute Percentage Error unless noted otherwise)

<table>
<thead>
<tr>
<th>Study</th>
<th>Situation</th>
<th>Error Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port of N.Y. Authority (1957)</td>
<td>Air travel</td>
<td>0.0</td>
</tr>
<tr>
<td>FILDES AND FITZGERALD (1983)</td>
<td>Balance of payments</td>
<td>1.7&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Armstrong (1968b)</td>
<td>International photographic market</td>
<td>4.2</td>
</tr>
<tr>
<td>BRANDON, FRITZ AND XANDER (1983, p. 195)</td>
<td>U.S. GNP</td>
<td>4.8&lt;sup&gt;b,c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Levine (1960)</td>
<td>Capital expenditure</td>
<td>5.4&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Okun (1960)</td>
<td>Housing starts</td>
<td>6.2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Vandome (1963)</td>
<td>National economy (UK)</td>
<td>7.1</td>
</tr>
<tr>
<td>O'Herlihy et al. (1967)</td>
<td>Five product categories (UK)</td>
<td>23.4</td>
</tr>
<tr>
<td><strong>Unweighted Average</strong></td>
<td></td>
<td><strong>6.6</strong></td>
</tr>
</tbody>
</table>

<sup>a</sup>RMSE (Root Mean Square Error)
<sup>b</sup>Different combinations were presented; I used equal weights.
<sup>c</sup>All calculations were done using RMSE (not MSE as in this paper)

A summary of the evidence from the eight studies is provided in Exhibit 10-7. Combining is a powerful strategy! Combinations of two methods reduced the error by over 6%. It never hurt accuracy. Combinations of forecasts from three methods are expected to yield even further improvements, but only two studies were available on this: The error reduction in Rosenzweig (1957) was 0.0 and in BRANDT and BESSLER [1983] it was 17.9%.

UNCERTAINTY

The agreement among various forecasts can also be used as a measure of uncertainty. However, this is a crude measure, especially because the errors are usually positively correlated across the methods.
The best approach to a given forecasting problem may call for a combination of different methods. There are many ways in which the methods can be combined, but few rules exist to guide the researchers in such combinations. Two rules were suggested:

1. Subjective methods should precede objective methods.
2. Segmentation should precede extrapolation or econometric methods.

Rule 1 is supported by evidence. Rule 2 seems to be obvious. Exhibit 10-1 provided a summary of these hypotheses. It was also suggested that different methods be used for estimating current status and forecasting change. These recommendations were summarized in Exhibit 10-3.

Bootstrapping, which involves the development of a model by using the judge's rules, offers substantial advantages. It is cheaper than judgmental forecasts for repetitive decisions; it offers advantages for implementation because the decision makers can be sure that their rules are being used; and it allows you to develop a quantitative model when no data exist for the dependent variable. A substantial body of empirical evidence indicated that bootstrapping is almost always as accurate as the typical judge, and generally more accurate. The particular way in which the bootstrapping model is developed is of little importance. You may use either indirect or direct methods because the performance of the model is highly dependent on the signs of the relationships, but is not sensitive to the magnitude of relationships. You should choose the cheapest and most acceptable way; this is likely to be the use of equal weights. Advice on when to use bootstrapping was provided in Exhibit 10-4.

The use of econometric methods within segments seems of value when interaction is high and when a moderate amount of historical data is available. The use of econometric methods across segments offers a sensible procedure in this same situation.

The method of leading indicators was briefly considered. From a theoretical viewpoint, there is little to recommend this method. Little empirical evidence was found to suggest that it is useful for forecasting change, although it may help to estimate current status.

Combining forecasts from different methods offers a powerful strategy for improving accuracy. Evidence from eight studies showed a reduction in the size of the error by over 6%.