International Journal of Forecasting 5 (1989) 559–583
North-Holland

Combining forecasts: A review and annotated bibliography

Robert T. CLEMEN *
College of Business Administration, University of Oregon, Eugene, OR 97403-1208, USA

Abstract: Considerable literature has accumulated over the years regarding the combination of forecasts. The primary conclusion of this line of research is that forecast accuracy can be substantially improved through the combination of multiple individual forecasts. Furthermore, simple combination methods often work reasonably well relative to more complex combinations. This paper provides a review and annotated bibliography of that literature, including contributions from the forecasting, psychology, statistics, and management science literatures. The objectives are to provide a guide to the literature for students and researchers and to help researchers locate contributions in specific areas, both theoretical and applied. Suggestions for future research directions include (1) examination of simple combining approaches to determine reasons for their robustness, (2) development of alternative uses of multiple forecasts in order to make better use of the information they contain, (3) use of combined forecasts as benchmarks for forecast evaluation, and (4) study of subjective combination procedures. Finally, combining forecasts should become part of the mainstream of forecasting practice. In order to achieve this, practitioners should be encouraged to combine forecasts, and software to produce combined forecasts easily should be made available.

Keywords: Forecast combination, Composite models, Forecast aggregation, Consensus, Forecast synthesis.

"In combining the results of these two methods, one can obtain a result whose probability law of error will be more rapidly decreasing."
Laplace (1818)

1. Introduction

Consider what we have learned about the combination of forecasts over the past twenty years. Models have been developed to find 'optimal' combinations of forecasts. Both simulation and empirical studies have been done to test the models. Bayesian interpretations have been presented. The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy. This has been the result whether the forecasts are judgmental or statistical, econometric or extrapolation. Furthermore, in many cases one can make dramatic performance improvements by simply averaging the forecasts. The ASA/NBER business outlook surveys have produced composite economic forecasts since 1968. These forecasts have been analyzed by Su and Su (1975); Zarnowitz (1984); Hafer and Hein (1985); and Zarnowitz and Lambros (1987). The technique has been put to use in practice by Robert J. Eggert of Blue Chip Economic Enterprises, who has published consensus macroeconomic forecasts since 1976. His forecasts are generally regarded as among the most accurate macroeconomic forecasts (see Bernstein and Silbert, 1984; Agnew, 1985a). Holden and Peel (1986b) note that *The Financial Times* regularly reports simple averages of economic forecasts, and they suggest that these aver-

* Scott Armstrong, Derek Bunn, John Guerard, Robin Hogarth, Chris Miller, and four anonymous referees provided helpful comments on earlier drafts of this paper. This research was supported in part by the National Science Foundation under Grant IST-8600788.

ages may indeed form the basis of individuals' expectations regarding economic phenomena.

Twenty years have passed since Reid's (1968, 1969) and Bates and Granger's (1969) studies appeared. These are considered by most forecasters to be the seminal works in the area of combining forecasts. In a sense, this is correct; these individuals were the first to develop a general analytical model specifically for combining forecasts in an optimal way and to apply their techniques in real world situations. However, a number of earlier studies, primarily in the fields of statistics and psychology, had focused on combining estimates (as opposed to forecasts per se) of unknown quantities. As the opening quote indicates, even Laplace might not have been surprised by our recent findings. One of the objectives of this review and bibliography is to point out some of these early contributions.

At the other end of our time horizon, we have seen in recent years an explosion in the number of articles on the combination of forecasts. The works listed in the bibliography comprise over 2000 journal pages and 11 books, monographs, and theses. Exhibit 1 shows the cumulative number of published works (based on those in the bibliography); note the increased steepness of the curve in the 1980's. The plethora of new studies has made it difficult to stay current with developments in the field, and so another motivation for this review is to catalog many of the recent developments. To a great extent, recent articles have focused on applying econometric theory and models to the combination of forecasts. Another important area of expansion has been the combination of probabilities and probability distributions. Finally, as forecast combination has gained acceptance, the technique has been applied in many different areas.

This review is not intended to be a critical review of the field, but rather a catalog of research, focusing on early papers, seminal papers, contributions from other disciplines, and forecasting theory and applications. For excellent and insightful critiques, the reader is referred to Bunn (1987, 1988) and Mahmoud and Makridakis (1988). Determining what to include and what not to include in the bibliography has been a problem. If an article was not in the mainstream of the forecasting literature, it was included only if it was deemed important for forecasters to know about. In a project such as this, there are always works that escape the author's attention, and such is undoubtedly the case here.

The next section provides a brief discussion of aspects of the literature on the theory of combining forecasts. It is convenient to think in terms of (1) work that is primarily related to the forecasting literature (multifarious as that literature is), (2) contributions from psychology, and (3) contributions from statistics and management science (other than forecasting). Each of these areas is discussed in turn. After discussion of theoretical issues, a brief survey of applications of forecast combination techniques is presented, followed by a discussion of future directions for research in forecast combination. The annotated bibliography follows, containing over 200 items. Throughout, the terms 'combined forecast' and 'composite

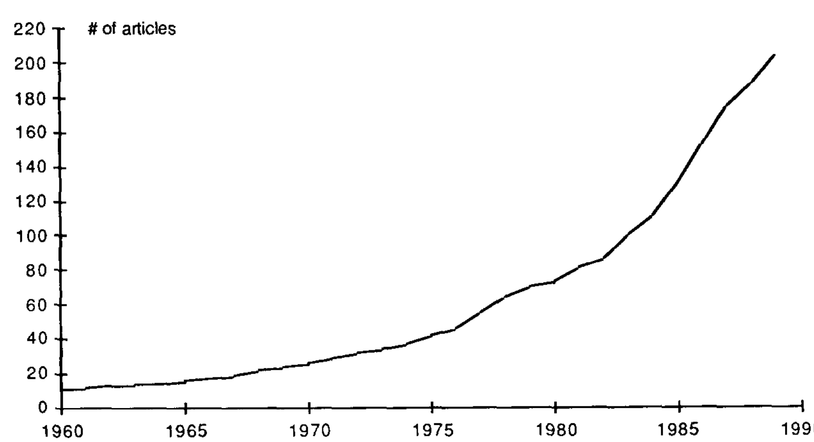

Exhibit 1. Graph showing the cumulative number of articles published on combined forecasts. The graph is based on the articles contained in the annotated bibliography (see pp. 568–583).

forecast' have been used synonymously, although some researchers have made a distinction (e.g., Reid, 1974, in his discussion of Newbold and Granger (1974) or Kang, 1986).

2. Combining forecasts: Theory

2.1. Contributions from forecasting

The work by Bates and Granger (1969) often is considered to be the seminal article on combining forecasts. In this paper, the authors developed and tested a number of techniques for combining point forecasts. Slightly earlier Crane and Crotty (1967) suggested combining forecasts through regression, and Zarnowitz (1967) commented on the superiority of the average of several forecasts of GNP. Nevertheless, it was the work of Bates and Granger (1969) and Reid (1968, 1969) that provided the initial impetus to the development of theory in the combination of forecasts. Following them came a stream of articles in *Operational Research Quarterly*, including articles by Dickinson (1973, 1975), Bunn (1975, 1977) and Öller (1978). These articles were primarily theoretical by nature, developing various details of the statistical models. Dickinson focused on the minimum-variance model, while Bunn developed the idea of a Bayesian combination on the basis of how likely one forecast is to outperform another. Bunn's approach reflected a practical, if not theoretically rigorous, approach to using multiple forecasts. Subsequent debate about the interpretation of outperformance probabilities versus 'veridical' probabilities (probabilities that specific models represent the true process) and the appropriateness of these different probabilities for combining forecasts can be found in French (1980, 1981) and Bunn (1981).

At roughly the same time, Nelson (1972) and Cooper and Nelson (1975) developed methods for studying the efficiency of forecasts. These methods essentially used a composite forecast as a benchmark. They showed that certain econometric forecasts were inefficient in the sense that combining the econometric forecast with an ARIMA model significantly reduced the forecasting error. Granger and Newbold (1973) addressed similar issues in terms of forecast evaluation. These studies led to a stream of research studying the rationality of economic forecasts; a brief review can be found in Ahlers and Lakonishok (1983). Still more recently Phillips (1988) provided a more complete theoretical development of the problems involved in such tests.

To a great extent, these early articles laid out the fundamental theory involved in combining forecasts. The thoughts of the authors are still very appropriate, as evidenced by the tendency of recent papers to duplicate the basic results. Researchers interested in the theory of combining are urged to consult the papers discussed here in order to obtain a full appreciation for the depth and variety of results obtained by these early authors.

The relative performance of a number of forecasting techniques was examined in a series of three empirical studies: Newbold and Granger (1974), Makridakis and Hibon (1979) and Makridakis et al. (1982, 1983). Newbold and Granger's results demonstrated that in practice one should ignore correlations in estimating combining weights. Weighted averages that depended on estimated correlations performed poorly. This paper was published with discussion in the *Journal of the Royal Statistical Society*. To a great extent, the discussion attacked the empirical/pragmatic approach that Newbold and Granger took. Such an approach, as argued by Bunn (1987, 1988), was at odds with the forecasting climate at the time, steeped as it was in classical statistical tradition. The emphasis had been on the identification of the underlying process, modeling it, estimating parameters, validating the model, and finally generating forecasts. The idea of combining forecasts implicitly assumed that one could not identify the underlying process, but that different forecasting models were able to capture different aspects of the information available for prediction. Several of the discussants pointed out that successfully combining an ARIMA forecast with another time series forecast was an indication that the ARIMA model was misspecified in the first place.

Makridakis and Hibon (1979) continued in the same spirit as Newbold and Granger, comparing the forecasting performance of various forecasting methods. In contrast, though, the results of this study showed that simple forecasting models generally outperform more sophisticated approaches. While Makridakis and Hibon did not examine composite models, their work was an important forerunner to the 'M-competition', a forecasting competition organized by Makridakis. The results of this competition were reported in Makridakis et al. (1982, 1983), with extensive commentary in Armstrong et al. (1983). In this competition, a large variety of time series forecasting methods were applied to 1001 different economic time series. Ex ante forecasts were calculated, and forecast performance was measured using various error summary measures. While the primary motivation for the competition was to compare the forecasting performance of various time series methods, two different combining schemes were studied. Both of these combinations performed well relative to the individual techniques, with the simple average having the better performance of the two. More importantly, as noted by Winkler in the commentaries (Armstrong et al. 1983), is the fact that the combinations were robust, performing well for most of the various types of series. In follow-up studies, Makridakis and Winkler (1983) discussed the impact of the number of forecasts included in a simple average, and Winkler and Makridakis (1983) addressed the problem of using 'optimal' weighted averages, using techniques from Newbold and Granger (1974) for estimating the weights. Their results reconfirmed Newbold and Granger's message that it is usually better to ignore the effects of correlations in calculating combining weights.

One of the more influential recent articles on forecast combination has been that of Granger and Ramanathan (1984), who pointed out that the conventional forecast combination methods could be viewed within a regression framework. They argued that the standard techniques were equivalent to constrained ordinary least squares (OLS) estimation in a regression model having the actual value as the response variable and the individual forecasts as the explanatory variables. Instead of constraining the combining weights to sum to one and forcing the regression through the origin, Granger and Ramanathan suggested running unconstrained least squares to obtain a better fit and presumably better forecasting performance.

Granger and Ramanathan's suggestion has been discussed and contested theoretically (Clemen, 1986; Trenkler and Liski, 1986; Bordley, 1986), empirically (Mills and Stephenson, 1985; Holden and Peel, 1986a), and on the basis of simulation experiments (Holmen, 1987). Furthermore, their suggestion, while timely, was not entirely original. Combining forecasts using regression techniques had been suggested by Crane and Crotty (1967) and Reinmuth and Geurts (1979). However, Granger and Ramanathan provided an important impetus for the use of more sophisticated econometric methods for combining forecasts. For example, Diebold (1988) explored serial correlation in combined forecasts, Diebold and Pauly (1987b) studied the possibility of using weighted least squares techniques to create time-varying parameter combination models, and Diebold and Pauly (1986, 1987a) combined these two techniques. Engle, Granger and Kraft (1984) used a bivariate ARCH model for combining inflation forecasts. Phillips (1987, 1988) discussed the theory involved in optimal linear composite forecasts. Guerard (1987) and Guerard and Beidleman (1987) applied robust-weighting techniques and ridge regression, respectively, to the combination of earnings forecasts, and Guerard and Clemen (1989) used latent root regression to combine macroeconomic forecasts. Bayesian techniques for including prior information in a forecast combination have been studied by Clemen and Winkler (1986) and Diebold and Pauly (1987c). Schmittlein, Kim and Morrison (1989) explained how to use Akaike's information criterion to decide which of several possible models to estimate. While it would be inappropriate to credit Granger and Ramanathan for inspiring all of these econometric-like studies, their paper did mark the beginning of such research efforts.

2.2. Contributions from psychology

Psychology and forecasting come together in the field of judgmental forecasting. For our purposes, this includes subjective judgments of unknown quantities whether they are immediately measurable or not. Psychological research has covered group consensus judgments as well as the mechanical combination of individual judgments. Various techniques have been developed (e.g., Delphi and Nominal Group Technique) to aid groups in arriving at a consensus. The psychological literature on group processing of information is extensive, and it is not our purpose to review it here. Lock (1987) provides a recent general review. For an introduction to the vast literature on

Delphi, see Linstone and Turoff (1975) and more recently Parenté and Andersen-Parenté (1987). A description of Nominal Group Technique can be found in Delbecq, Van de Ven and Gustafson (1975). Other important reviews of the group information-processing literature include Lorge et al. (1958) and Hill (1982). Exclusion of the group judgment literature does not indicate that I think this literature is unimportant. Indeed, many important real world decisions rely on information from a panel of experts; understanding how a panel of experts processes information and formulates a consensus might greatly improve our use of expert information. However, the vastness of the literature precludes it from a survey on combining forecasts. Hogarth (1977), Ferrell (1985) and Lock (1987) reviewed portions of this literature. We will content ourselves with literature contributing to the mechanical combination of individual judgments.

Psychologists had the jump on forecasters when it comes to thinking about the combination of judgments. Gordon (1924) reported the results of an experiment in which individuals assessed weights for objects. When she averaged the orders assigned by the individuals, the correlation with the actual order increased. The averaged orders were more accurate than those of the average group subject and were at least as accurate as those of the best individual judgment. This article induced a small debate that went on for about twenty years. The upshot was that Gordon's results were eventually shown, via results from test theory, to be a statistical rather than psychological artifact. Zajonc (1962) reviewed these early studies.

Beginning in the 1950's, considerable research effort was aimed at the development of statistical models of expert judgments. To a great extent, this stream of research was inspired by the publication in 1954 of Meehl's *Clinical versus Statistical Prediction*. The technique of modeling a clinician's judgments has been labeled 'bootstrapping' (not to be confused with a randomization method in statistics that goes by the same name). The idea is to run a regression with the expert judgment as the response variable and the criteria as the explanatory variables. The rather remarkable finding has been that the bootstrapped judgments virtually always outperform the original judgments in out-of-sample prediction. Dawes, Faust and Meehl (1989) provide a concise review of the bootstrapping literature.

One of the fortuitous by-products of the bootstrapping research was insight regarding the combination of judgments. After all, when one has judgments from a number of individuals, it would be natural to average those judgments. This was done by Goldberg (1965, 1970) and Wiggins and Kohen (1971) in their bootstrapping studies. All three articles presented evidence that in a clinical situation, averages of judgments were more accurate than the individual judgments. In itself, bootstrapping does not contribute to the mechanical combination of expert judgments. However, there is a connection between the mechanical combination of cues to form a judgment and the mechanical combination of individual judgments. Einhorn (1972) and Einhorn and Hogarth (1975) blended ideas from the bootstrapping literature with issues in the combination of forecasts. In particular, Einhorn and Hogarth provided a theoretical explanation of the generally strong performance of equally weighted combinations, whether those combinations involve forecasts in a composite forecasting approach or cues in a bootstrapping model. Einhorn (1974) discussed the connection between consensus and expertise.

More recent work has directly addressed the issue of combining multiple judgments. For example, Einhorn, Hogarth and Klempner (1977) developed baselines for the performance of group judgment; one of these baselines was the average of the individual judgments. Hogarth (1978) used test theory as a basis for discussing the selection of experts. His conclusions were that between 6 and 20 different forecasters should be consulted, and that the more the forecasters differed, the more should be included in the combination. Libby and Blashfield (1978), though, reported that the majority of the improvement in accuracy was achieved with the combination of the first two or three forecasts. Ashton (1986), in a study of forecasts by executives and sales managers, found that Hogarth's model provided an excellent approximation.

2.3. *Contributions from statistics and management science*

In this section, we focus on research that comes from a statistical or management science tradition.

However, the heading for this section is somewhat misleading. Most of the forecasting literature, both theoretical and empirical, is statistical by nature. Furthermore, much of it has appeared in management science journals. The early contributions by Bates and Granger, Dickinson, and Bunn, for example, appeared in *Operational Research Quarterly*. Thus, to a great extent the distinctions are blurred.

The opening quote shows that, indeed, combining estimates is not new. More than 170 year ago, Laplace considered combining regression coefficient estimates. In his work, described by Stigler (1973), he was able to derive and compare the properties of two estimators, one being least squares and the other a kind of order statistic. Based on the joint distribution of the two, he derived a combining formula. However, he concluded that not knowing the error distribution rendered the combination infeasible.

Aside from Laplace, the earliest statistical treatment of combining multiple estimates appears to have been that of Edgerton and Kolbe (1936). The authors found an 'optimal' combined estimate, but their optimality criterion was to minimize the sum of squares of the differences of the standard scores for the estimates. Independently, Horst (1938) derived a composite measure by maximizing the pairwise separation among the sample points. Both approaches are closely related to least squares, but compared to standard least squares techniques they seem awkward and foreign. Halperin (1961) provided a minimum-squared-error combination of estimates, and Geisser (1965) discussed the Bayesian equivalent, finding a decision maker's posterior distribution for a quantity given multiple dependent forecasts of the quantity. The combining forecast research of the early 1970's appears to have been done without cognizance of this early statistical research.

Another major contribution of the statistics

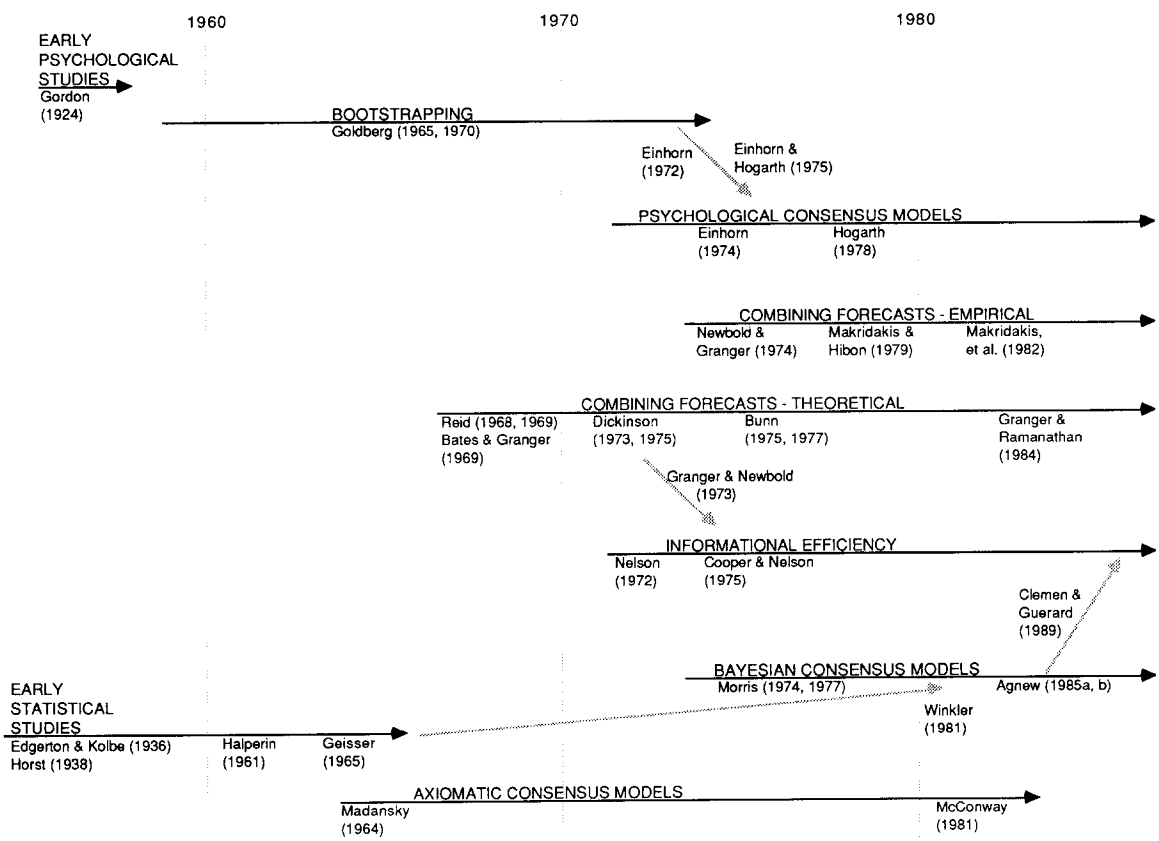

Exhibit 2. The historical development of the combining forecasts literature.

and management science literatures has to do with the combination of probability distributions. Treatment here is brief; access to this rich literature is provided in reviews by Genest and Zidek (1986) and French (1985). Contributions include axiomatic work (e.g., Madansky, 1964; McConway, 1981; Lehrer and Wagner, 1981) as well as Bayesian treatments (e.g., Winkler, 1968; DeGroot, 1974). More recently, and more pertinent to the study of combining forecasts, has been the focus on a systematic Bayesian paradigm developed by Morris (1974, 1977) in which a Bayesian decision maker views information (probabilities, probability distributions, or forecasts) from various sources simply as data to be used in updating prior information. While this approach is theoretically elegant, it is difficult to apply in practice, requiring the decision maker to assess a complicated multivariate likelihood function. Extension and applications of the theory include Winkler (1981), Bordley (1982), Agnew (1985a, b) and Clemen and Winkler (1985, 1986, 1987).

Finally, forecast combination methods have been inspired by operations research models for dealing with multiple objectives. Lawrence and Reeves (1981), Reeves and Lawrence (1982) and Gulledge et al. (1986) showed how multiple objective linear programming can be used to combine forecasts. In this approach, the objective function can involve the minimization of some composite of various error statistics. Wall and Correia (1989) take a mathematical programming approach to forecast combination; their optimization criterion is based on the decision maker's preferences over error distributions. Gupta and Wilton (1987, 1988) combined Bunn's outperformance approach with the mathematics of Saaty's (1980) analytic hierarchy process to create a novel subjective approach to forecast combination. Based on simulation and empirical results, this technique appears to be easy to use in practice and robust in terms of prediction performance. Since it requires assessment of odds of relative forecast performance for each possible pair of forecasts in the combination, the technique has proven effective in situations where few data are available.

Many individuals have contributed to our knowledge regarding the combination of forecasts. Exhibit 2 represents the evolution of thinking in this area and provides a graphical summary of the material discussed above. Of course, only a small selection of articles can be displayed on the diagram. Nevertheless, an effort has been made to (1) highlight contributions from the fields of psychology, statistics, and management science, and (2) identify a few of the articles that provide bridges between different streams of research.

3. Combining forecasts: Applications

As the notion of combining forecasts and the value of doing so has become more widespread, applications have increased. In this section we note some of these applications.

Meteorologists have long considered the potential for combining forecasts. Sanders (1963) discussed the possibility of probability forecasting in meteorology and reported an experiment in which probabilities from individual forecasters were averaged. Staël Von Holstein (1971) reported a similar but more extensive experiment. Winkler, Murphy and Katz (1977) studied the performance of various probability consensus models in the context of probability of precipitation forecasts. More recently, Clemen (1985), Clemen and Murphy (1986a, b) and Murphy, Chen and Clemen (1988) have applied forecast combination techniques to measure the informational contributions of the different components of the forecasting system.

Applications of combining forecasts in macroeconomic problems has been extensive. Reid (1968) was interested in predicting gross national product. As discussed above, the studies by Nelson (1972) and Cooper and Nelson (1975) were perhaps the first application of combination techniques to answer questions regarding econometric and time series forecasts. More recently, Nelson (1983) discussed the use of time series forecasts as benchmarks for macroeconomic forecast evaluation using composite forecasting techniques. Clemen and Guerard (1989) provided a related Bayesian approach for measuring information in economic forecasts. Some other economic forecasting applications include inflation (Engle, Granger and Kraft, 1984; Hafer and Hein, 1985), money supply (Figlewski and Urich, 1983; Mills and Stephenson, 1987), exchange rates (Bilson, 1983; Blake, Beenstock and Brasse, 1986; Guerard, 1989) and stock prices (Virtanen and Yli-Olli, 1987; Staël Von Holstein, 1972). Consensus forecasts of corporate earnings have been studied by Cragg and Malkiel (1968), Elton, Gruber and Gultekin (1981), Conroy and Harris (1987), Guerard (1987), Guerard and Beidleman (1987) and Newbold, Zumwalt and Kannan (1987). Sales forecasting using composite methods has been the focus of a number of studies including Doyle and Fenwick (1976), Moriarty and Adams (1984), Sewall (1981) and Schnaars (1986a, b). Studies by Brandt and Bessler (1983), Holt and Brandt (1985) and Guerard and Beidleman (1987) estimated economic benefits to be derived from the use of combined forecasts.

Application of combined forecasting has not been limited to meteorology and economics. Kaplan, Skogstad and Girshick (1950) studied the prediction of social and technological events. Schmitt (1954) considered a composite approach in forecasting city populations. Psychiatric diagnosis was the task of interest for Goldberg (1965, 1970) and Wiggins and Kohen (1971). Winkler (1971) reported an experiment involving prediction of football game outcomes. Rausser and Oliveira (1976) used a composite approach to forecast wilderness area use, and Mabert (1978) developed a combined forecast approach to predict check volume. Other applications include livestock prices (Bessler and Brandt, 1981; Brandt and Bessler, 1981, 1983; Bessler and Chamberlain, 1987), electrical demand (Bunn and Seigal, 1983; Bunn, 1987; Smith, 1989), tourism (Reinmuth and Geurts, 1979; Fritz, Brandon and Xander, 1984), insurance (Taylor, 1985), political risk (Bunn and Mustafaoglu, 1978), effects of the Oregon bottle bill (Geurts and Wheeler, 1980), population (Openshaw and Van Der Knaap, 1983), and sunspot cycles using the familiar Wolf data (Morris, 1977; Poskitt and Tremayne, 1986).

4. Future research directions

Many theoretical and empirical issues in the combination of forecasts have been addressed, and, to a great extent, many of the more fundamental issues have been resolved. However, a variety of issues remain to be addressed. Some of these will be discussed in the following paragraphs.

The empirical work has raised an issue that still deserves attention. What is the explanation for the robustness of the simple average of forecasts? In many studies, the average of the individual forecasts has performed best or almost best. Statisticians interested in modeling forecasting systems may find this state of affairs frustrating. The questions that need to be answered are (1) why does the simple average work so well, and (2) under what conditions do other specific methods work better? Some authors have speculated about instability of combining weights (Clemen and Winkler, 1986; Kang, 1986) and about the nonstationarity of the underlying system being forecast (Diebold and Pauly, 1986, 1987a, b). Initial work by Bunn (1985a) and Schmittlein, Kim, and Morrison (1988) may lead to empirically sound methods for deciding when to use different combining models including a simple average. On a related issue, Schnaars (1986a) and Russell and Adam (1987) provide guidance on forecast selection for composite forecasts.

From a conventional forecasting point of view, using a combination of forecasts amounts to an admission that the forecaster is unable to build a properly specified model. Trying ever more elaborate combining models seems only to add insult of injury, as the more complicated combinations do not generally perform all that well. It might be argued that studying forecast combinations may eventually help forecasters to specify underlying processes more appropriately and thus build better individual models. If several different models can be combined to obtain a better forecast, it should theoretically be possible to construct a single model that makes optimal use of the different kinds of information used by the component forecasts in the combination. However, the empirical result that more complicated univariate forecasting models do not always produce better forecasts (Armstrong, 1984) may reduce the value of such a strategy.

On the other hand, a Bayesian view leads to a considerably more optimistic perspective. The problem of combining forecasts can be viewed in a Bayesian sense as one in which a decision maker needs to make the best possible use of the multiple forecasts available. From this perspective, there are two intriguing avenues for research. The first of these is a matter of building Bayesian models to help decision makers in using multiple forecasts effectively. While a number of such models are available (see Genest and Zidek, 1986), a great opportunity exists for building models that can accomodate a wide variety of forecasts and even different kinds of forecast information, all of which may be pertinent to a decision maker facing uncertainty. The second area, proposed by Bunn (1988), asks whether a decision maker is best served by mechanically combining forecasts. Morris's (1974, 1977) Bayesian paradigm suggests that the decision maker should use the multiple forecasts to develop a posterior probability distribution for the event of interest and then use that distribution in subsequent decisions. A simpler and perhaps more practical first approximation for dealing with multiple forecasts is to use them in a sensitivity analysis mode. Is the decision to be taken sensitive to the forecasts? If not, combining the forecasts may not pay off. Furthermore, the sensitivity analysis may lead to new insights in the decision problem.

Another area which could provide fruitful involves using combined forecasting techniques in forecast evaluation. It is not enough to compare two individual forecasts. The question a decision maker needs to ask is: "If I have Forecast *A*, how much additional information can I get from Forecast *B*?" Alternatively, we can ask: "How much incremental information is provided by Forecast *B* in the combination of *A* and *B*?" The early work by Nelson (1972), Granger and Newbold (1973) and Cooper and Nelson (1975) forms the basis here. Recent work by Nelson (1984) and Clemen and Guerard (1989) shows how this approach can be applied. However, in those studies, forecasts were compared to single extrapolation forecasts. In contrast, Lupoletti and Webb (1986) suggested using a vector auto-regressive model as a benchmark against which to evaluate macroeconomic forecasts within a combination framework. Given the ease with which time series forecasts can be created now, it is a simple matter to create multiple extrapolation forecasts and average them. The strong performance of such averages was discussed above. With adequate data, one can determine whether any given forecast adds information (in a statistical sense) to a naive combination of extrapolation forecasts.

An additional area for research has been suggested by Flores and White (1988). In their review, they point out that much of the research to date has focused on both objective forecasts and mechanical combinations, and they suggest that it may be worthwhile to consider subjective forecasts as well as intuitive or subjective combinations. Some research has considered subjective forecasts (e.g., Edmundson, Lawrence and O'Connor, 1988), subjective combination weights (e.g., Ashton and Ashton, 1985), or both (Flores and White, 1989). However, a potentially fruitful research program might explore systematically the relationships between subjective forecast combinations and results from cognitive psychology on subjective judgments.

Combining forecasts has been shown to be practical, economical, and useful. Underlying theory has been developed, and many empirical tests have demonstrated the value of composite forecasting. We no longer need to justify this methodology. We do need to find ways to make the implementation of the technique easy and efficient. Dalrymple (1978, 1987) presented evidence that firms are gradually beginning to combine forecasts; in his 1987 survey, 40% of the firms frequently or usually combined forecasts. However, one might suspect that in many cases it is a matter of informal combination of judgmental estimates. Nowadays, with the advent of inexpensive forecasting software for personal computers, virtually any decision maker can generate multiple forecasts of a time series. Given the results of recent studies comparing automatic and non-automatic forecasting strategies (Carbone et al. 1983), such software can be extremely useful for decision makers with even a small amount of training. Subroutines to combine forecasts, if only through averaging, should be included with this software. The SIBYL-RUNNER program (see Makridakis et al. 1974) provided this possibility some time ago. More recently, Poulos, Kvanli, and Pavur (1987) described an automated forecasting system that generates a minimum-variance composite forecast from Box–Jenkins and Holt–Winters models. Finally, decision makers must be encouraged to use this and other software to create composite forecasts and to use these forecasts in making their decisions.

References

To avoid unnecessary duplication, only those references not included in the annotated bibliography below are given here.

Carbone, R., A. Andersen, Y. Corriveau and P.P. Corson, 1983, "Comparing for different time series methods the value of technical expertise, individualized analysis, and judgmental adjustment", *Management Science*, 29, 559–566.

Dawes, R.M., D. Faust and P.A. Meehl, 1989, "Clinical versus actuarial judgment", *Science*, 243, 1668–1673.

DeGroot, M.H., 1974, "Reaching a consensus", *Journal of the American Statistical Association*, 69, 118–121.

Delbecq, A.L., A.H. Van de Ven and D.H. Gustafson, 1975, *Group Techniques for Program Planning* (Scott Foresman, Glenview, IL).

Einhorn, H.J., 1974, "Expert judgment: Some necessary conditions and an example", *Journal of Applied Psychology*, 59, 562–571.

French, S., 1985, "Group consensus probability distributions: A critical survey", in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (Eds.), *Bayesian Statistics*, Vol. 2 (North-Holland, Amsterdam) 183–197.

Genest, C. and J.V. Zidek, 1986, "Combining probability distributions: A critique and annotated bibliography", *Statistical Science*, 1, 114–135.

Hill, G.W., 1982, "Group versus individual performance: Are $N+1$ heads better than one?", *Psychological Bulletin*, 91, 517–539.

Laplace, P.S. de, 1818, *Deuxième Supplément a la Théorie Analytique des Probabilités* (Courcier, Paris); reprinted (1847) in *Oeuvres Complètes de Laplace*, Vol. 7 (Paris, Gauthier-Villars) 531–580.

Lehrer, K. and C. Wagner, 1981, *Rational Consensus in Science and Society* (Reidel, Dordrecht).

Linstone, H. and M. Turoff, 1975, *The Delphi Method: Techniques and Applications* (Addison-Wesley, London).

Lorge, I., D. Fox, J. Davitz and M. Brenner, 1958, "A survey of studies contrasting the quality of group performance and individual performance", *Psychological Bulletin*, 55, 337–372.

Madansky, A., 1964, *Externally Bayesian Groups*, RM-4141-PR, The Rand Corporation Santa Monica, CA.

Makridakis, S. and M. Hibon, 1979, "Accuracy of forecasting: An empirical investigation (with discussion)", *Journal of the Royal Statistical Society, Series A*, 142, 97–145.

Makridakis, S., A. Hodgsdon and S.C. Wheelwright, 1974, "An interactive forecasting system", *American Statistician*, 28, 153–158.

McConway, K.J., 1981, "Marginalization and linear opinion pools", *Journal of the American Statistical Association*, 76, 410–414.

Meehl, P.E., 1954, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (University of Minnesota Press, Minneapolis).

Parenté, F.J. and J.K. Andersen-Parenté, 1987, "Delphi inquiry systems", in: G. Wright and P. Ayton (Eds.), *Judgmental Forecasting* (Wiley, New York) 129–156.

Stigler, S.M., 1973, "Laplace, Fisher, and the discovery of the concept of sufficiency", *Biometrika*, 60, 439–445.

Biography: Robert T. CLEMEN holds a Ph.D. in Business from Indiana University and is Associate Professor of Business at the University of Oregon. His research interests include decision analysis, decision theory, and forecasting, especially the use and aggregation of expert information. His articles have appeared in a variety of scholarly publications, including *Management Science, Journal of Business and Economic Statistics, Journal of Forecasting*, and the *International Journal of Forecasting*.

ANNOTATED BIBLIOGRAPHY

This annotated bibliography contains works that have contributed to knowledge regarding the combination of forecasts, either through theory or application. For the most part, entries are limited to published articles and books. A few particularly important theses, proceedings, and unpublished manuscripts have been included.

Compilation of this bibliography was greatly aided by other reviews, especially Mahmoud (1984), Armstrong (1985) and Genest and Zidek (1986). Consultation of the Social Science Citation Index for recent citations of key articles helped to identify recent articles that either apply forecast combination techniques or develop the theory. Thus, the bibliography is as up-to-date as possible. However, given the burgeoning nature of the forecast combination literature, new articles on forecast combination have undoubtedly appeared since this article was written.

Each entry contains a key that is intended to give some guidance to the nature and contents of the study. The abbreviations are:

F – Forecasting,
TS – Time Series,
R – Regression/Econometric models,
EO – Expert opinion or judgmental forecasts,
P – Psychology,
S – Statistics,
C – Classical,
BY – Bayesian,
T – Theoretical,
RV – Review,
SI – Simulation,
A – Application,
X – Experimental,
E – Economics,
BS – Business,
M – Meteorology,
O – Other.

Although this arbitrary classification scheme is self-explanatory for the most part, some explanation is in order. Classifying a work as **F**, **P**, or **S** depended to a great extent on whether the work fell into a broad mainstream of research in one of these disciplines. Of course, there is substantial overlap between the forecasting and statistical disciplines. If an article appeared to be primarily statistical and theoretical in nature, it was labeled **S**. If it appeared to be more aimed at practical forecasting, it was labeled **F**. Some articles are labeled with both. The difference between **A** and **X** is that **A** was reserved for studies in which real world data were used to demonstrate or test a forecasting technique, while **X** labels those studies in which data were generated experimentally (in ways other than computer simulation). Labels **E** and **BS** were used for the most part to distinguish between macroeconomic (**E**) and microeconomic (**BS**) applications or experiments.

1. Adams, A.J. (1978), " Modeling and forecasting seasonal time series of product sales across territories: A comparative study", Ph.D. Thesis, The University of Iowa. **F, TS, A, BS.** In this study of product sales forecasting, regression and Box-Jenkins forecasts were combined. The combined forecast substantially reduced the forecast error.

2. Agnew, C.E. (1985a), "Bayesian consensus forecasts of macroeconomic variables", *Journal of Forecasting*, 4, 363–376. **F, BY, A, E.** Forecasts from the Blue Chip Economic forecasters were combined using a sequential Bayesian weighting scheme. The author found that the Bayesian combination performed better than either the average or the median of the forecasts.

3. Agnew, C.E. (1985b), "Multiple probability assessments by dependent experts", *Journal of the American Statistical Association*, 80, 343–347. **S, BY, T.** Agnew extends Winkler's (1981) model to the case in which the experts provide information on multiple unknown quantities. His normal model could be of value in combining forecast research.

4. Ahlers, D., and Lakonishok, J. (1983), "A study of economists' consensus forecasts", *Management Science*, 29, 1113–1125. **F, A, E.** The data for this study were equal-weight consensus forecasts from Livingston's data set. The primary results were that the consensus forecasts were not efficient (did not incorporate all available information) and were unable to improve on simple models. Performance of consensus forecasts for ten economic variables varied considerably in terms of efficiency.

5. Anandalingam, G., and Chen, L. (1989), "Linear combination of forecasts: A general Bayesian model", *Journal of Forecasting*, 8, 199–214. **S, T.** The authors develop the general Bayesian normal model for combining biased and correlated forecasts. They include a comparison with the Kalman filter as well as an extension to the case when the biases are unknown.

6. Anandalingam, G., and Chen, L. (1989), "Bayesian forecast combination and Kalman filtering", *International Journal of Systems Science*, 20, 1499–1507. **S, T.** The authors show the equivalence of the Bayesian multi-normal combination model to the Kalman filter.

7. Armstrong, J.S. (1984), "Forecasting by extrapolation: Conclusions from 25 years of research (with discussion)", *Interfaces*, 14, no. 6, 52–66. **F, TS, RV.** Armstrong reviews empirical research on the accuracy of extrapolation forecasts. Of 39 studies, 28 showed that the sophisticated methods were no better than simpler methods. Combining extrapolation forecasts is suggested as a way to improve predictive performance, and some specific studies were discussed.

8. Armstrong, J.S. (1985), *Long Range Forecasting: From crystal ball to computer*, 2nd ed., (Wiley, New York). **F, TS, R, EO, P, A, X, T, E, BS, M, O.** Armstrong believes strongly in combining forecasts. References to combining forecasts can be found throughout the book.

9. Armstrong, J.S. (1986), "Research on forecasting: A quarter-century review, 1960–1984 (with discussion)", *Interfaces*, 16, no. 1, 89–109. **F, TS, R, EO, RV.** Armstrong discusses how empirical research in forecasting over the past 25 years has supported or failed to support conventional forecasting wisdom in 1960. With regard to combining forecasts, Armstrong cites eight applications which combined forecasts, achieving error reductions from 0% up to 23.4%, with an average of 6.6% for combinations of two methods.

10. Armstrong, J.S., Lusk, E.J., Gardner, E.S. Jr., Geurts, M.D., Lopes, L.L., Markland, R.E., McLaughlin, R.L., Newbold, P., Pack, D.J., Andersen, A., Carbone, R., Fildes, R., Newton, H.J., Parzen, E., Winkler, R.L. and Makridakis, S. (1983), "Commentary on the Makridakis time series competition (M-competition)", *Journal of Forecasting*, 2, 259–311. **F, TS, A, E.** This collection of commentaries includes reviews of the original *Journal of Forecasting* article by seven independent researchers, followed by replies submitted by the participants in the competition.

11. Ashton, A.H. (1985), "Does consensus imply accuracy in accounting studies of decision making?", *The Accounting Review*, 60, 173–185. **P, EO, A, BS.** Ashton used correlations in a psychometric way to measure accuracy and consensus of predictions made as part of accounting and auditing tasks. Her results showed that more consensus generally implied greater accuracy.

12. Ashton, A.H., and Ashton, R.H. (1985), "Aggregating subjective forecasts: Some empirical results", *Management Science*, 31, 1499–1508. **F, EO, A, BS.** The authors analyzed advertising sales forecasts made by executives working for *Time* magazine. In combining the forecasts, they tried (1) a weighted average scheme in which the weight was associated with the forecaster's position in the organizational chart, and (2) a weighting scheme based weights on the CEO's subjective assessment of relative precision of the forecasters. The latter improved on the performance of a simple average, but the former did not.

13. Ashton, R.H. (1986), "Combining the judgments of experts: How many and which ones?", *Organizational Behavior and Human Decision Processes*, 38, 405–414, **P, EO, A, BS**. Ashton tested Hogarth's 1978 model for determining how many and which experts to include in a consensus. The task was forecasting advertising sales at Time, Inc., and the subjects were executives, managers, and sales personnel for the firm. Hogarth's model provided an excellent approximation in this situation.

14. Bates, J.M., and Granger, C.W.J. (1969), "The combination of forecasts", *Operational Research Quarterly*, 20, 451–468. **F, TS, T, C, A, E**. This is the earliest paper to develop a general analytical model for combining multiple forecasts. Aside from the general model, the authors also present conditions under which an estimated combining weight has a beta distribution, and they argue that negative combining weights are anomalous.

15. Bernstein, P.L., and Silbert, T.H. (1984), "Are economic forecasters worth listening to?", *Harvard Business Review* (September–October, 1984), 32–40. **F, RV, A, E**. The authors present results showing the advantages of combining economic forecasts. Much of their argument is based on analysis of Blue Chip's consensus forecasts. The graphs are particularly intriguing.

16. Bessler, D.A., and Brandt, J.A. (1981), "Forecasting livestock prices with individual and composite methods", *Applied Economics*, 13, 513–522. **F, TS, R, EO, A, E**. The authors applied combined forecasting techniques to forecast livestock prices (cattle, hogs, and broilers). They consider econometric, ARIMA, and expert opinion forecasts. Minimum-variance and adaptive-weights combinations of the econometric and ARIMA methods outperformed the individual forecasts (in terms of MSE), but simple averages of these two forecasts worked better. The three-way average, including the expert opinion, performed better for the hog prices, but not the others.

17. Bessler, D.A., and Chamberlain, P.J. (1987), "On Bayesian composite forecasting", *OMEGA International Journal of Management Science*, 15, 43–48. **F, TS, R, A, E**. The authors demonstrate Bunn's outperformance combining method in the context of forecasting hog prices, showing that a prior distribution can be used to capture the decision maker's beliefs regarding the credibility of the individual forecasts. The combined forecast performed well relative to the individual forecasts.

18. Bhattacharya, M.N. (1980), "The prediction performance of the Bonn monetary model", *Applied Economics*, 12, 399–412. **F, TS, R, A, E**. Bhattacharya replicates Nelson's (1972) study on another econometric model.

19. Bilson, J.F.O. (1983), "The evaluation and use of foreign exchange rate forecasting services", in: R. Herring (Ed.), *Managing Foreign Exchange Rate Risk* (Cambrige, Cambridge University Press) 149–179. **F, A, E**. The author used a composite exchange rate forecast to construct and manage currency portfolios. The results of using the method were equivocal in terms of being able to earn excess returns in the foreign exchange market.

20. Bischoff, C. (1989), "The combination of macroeconomic forecasts", *Journal of Forecasting*, 8, 293–314. **F, R, TS, A, E**. The author combines econometric and ARIMA forecasts of six macroeconomic variables, using a number of different combination methods. This is one of the few articles that conclude that on average one of the individual forecasts (in this case the econometric one) was superior to the combined forecasts.

21. Blake, D., Beenstock, M., and Brasse, V. (1986), "The performance of UK exchange rate forecasters", *The Economic Journal*, 96, 986–999. **A, E**. The authors applied composite forecasting in a study of exchange rate forecasts.

22. Bohara, A., McNown, R., and Batts, J.T. (1987), "A re-evaluation of the combination and adjustment of forecasts", *Applied Economics*, 19, 437–445. **F, TS, A, E**. The authors studied the performance of combined and corrected forecasts of three macroeconomic variables. Their results show that sometimes a combination of forecasts can perform worse than an individual forecast. In particular, this can happen when one forecast is much more precise than the others.

23. Bonini, C.P., and Freeland, J.R. (1979), "Forecasting by smoothed regression: Development and application to predicting customer utility bills", in: S. Makridakis and S.C. Wheelwright (Eds)., *Forecasting*, Vol. 12, TIMS Studies in the Management Sciences (North-Holland, New York), 279–296. **F, TS, R, T, A, BS**. The authors develop a technique for merging exponential smoothing models and multiple regression. Regression coefficients are estimated adaptively.

24. Bopp, A. (1985), "On combining forecasts: Some extensions and results", *Management Science*, 31, 1492–1498. **F, TS, R, A, T, E**. If a forecast is accurate, a transinformation measure can be used to tell whether the series is easy to forecast or whether a sound method has been applied to forecast a difficult series. Combined forecasts came out ahead in the transinformation-measure contest.

25. Bordley, R.F. (1982), "The combination of forecasts: A Bayesian approach", *Journal of the Operational Research Society*, 33, 171–174. **F, S, BY, T**. Bordley develops Bayesian counterparts to the Bates and Granger minimum-variance combinations. The results are essentially the same as those in Winkler (1981) and Geisser (1965).

26. Bordley, R.F. (1986), "Technical note: Linear combination of forecasts with an intercept: A Bayesian approach", *Journal of Forecasting*, 5, 243–429. **F, S, BY, T**. This paper develops the Bayesian counterpart to the Granger and Ramanathan unconstrained combination. The intercept term is interpreted in terms of the decision maker's prior information.

27. Brandon, C., Fritz, R., and Xander, J. (1983), "Econometric forecasts: evaluation and revision", *Applied Economics*, 15, 187–201. **F, R, A, E**. Combining four econometric forecasts resulted in improved forecast performance. A variety of combinations were tried, but the sample was very small.

28. Brandt, J.A., and Bessler, D.A. (1981), "Composite forecasting: An application with U.S. hog prices", *American Jour-*

nal of Agricultural Economics, 63, 135–140. **F, TS, R, EO, A, E.** In an application to forecasting hog prices, the authors found that a simple average of econometric, ARIMA, and expert opinion forecasts performed the best. They compared this composite to individual forecasts as well as weighted averages of the econometric and ARIMA forecasts.

29. Brandt, J.A., and Bessler, D.A. (1983), "Price forecasting and evaluation: An application in agriculture", *Journal of Forecasting*, 2, 237–248. **F, TS, R, EO, A, E.** The authors show the economic impact of various forecasting strategies by calculating average prices obtained for hogs. The expert judgment forecast was the worst (worse even than a naive forecast), and the ARIMA model led to the best price performance, followed closely by the composite forecast, a simple average of the econometric, ARIMA, and expert opinion forecast. It would be interesting to know how a composite formed of only the econometric and ARIMA models would have performed.

30. Bunn, D.W. (1975), "A Bayesian approach to the linear combination of forecasts", *Operational Research Quarterly*, 26, 325–329. **F, BY, T.** This article presents Bunn's 'outperformance' model, in which a decision maker assesses and updates a subjective probability that one forecast will outperform another.

31. Bunn, D.W. (1977), "A comparative evaluation of the outperformance and minimum variance procedures for linear syntheses of forecasts", *Operational Research Quarterly*, 28, 653–660. **F, S, BY, SI.** Bunn's conclusion, on the basis of simulation experiments, is that the outperformance method is preferable to the minimum-variance method if there is little prior information (less than 10 observations and possibly less than 30).

32. Bunn, D.W. (1978), *The Synthesis of Forecasting Models in Decision Analysis* (Birkhauser, Basel). **F, S, T.** Bunn compares the minimum-variance combining method with his outperformance method. Computer program listings are provided for adapative estimation of combining weights.

33. Bunn, D.W. (1979a), "Composition of estimators for decision making", *Technological Forecasting and Social Change*, 13, 157–167. **F, BY, A, E.** Bunn presents the multinormal and outperformance combination methods. Both methods were used to forecast Hawaiian tourist visits.

34. Bunn, D.W. (1979b), "The synthesis of predictive models in marketing research", *Journal of Marketing Research*, 16, 280–283. **F, RV.** Bunn describes his outperformance combination procedure and illustrates the feasibility of the procedure with two cases.

35. Bunn, D.W. (1981), "Two methodologies for the linear combination of forecasts", *Journal of the Operational Research Society*, 32, 213–222. **F, SI.** Bunn compares veridical, outperformance, and variance-minimizing composite forecasts in a simulation experiment. The veridical approach performed poorly relative to the outperformance method for nearly exchangeable forecasts (see comment by French, 1981).

36. Bunn, D.W. (1985a), "Statistical efficiency in the linear combination of forecasts", *International Journal of Forecasting*, 1, 151–163. **F, BY, SI, A, E.** This article compares a variety of combination techniques, including equal weights, the standard multinormal model, outperformance, Bayesian model selection, and some variations of these. Bunn's theoretical discussion focuses on the sensitivity of the posterior variance to error correlations in the multinormal model. His simulation results suggest specific combination procedures depending on the amount of prior information one has.

37. Bunn, D.W. (1985b), "Forecasting electric loads with multiple predictors", *Energy*, 10, 727–732. **F, RV.** In this essay Bunn identifies a number of motivations for combining forecasts, including the idea that combining makes sense when more elaborate model building is not practical. Other motivations relate to credibility. The essay concludes with Bunn's response to various criticisms of composite forecasting.

38. Bunn, D.W. (1987), "Expert use of forecasts: Bootstrapping and linear models", in: G. Wright and P. Ayton (Eds.), *Judgmental Forecasting* (Wiley, New York) 229–241. **F, P, A, BS.** This chapter provides a brief introduction to some of Bunn's earlier work. Through an example involving electricity demand forecasting, Bunn relates bootstrapping (linear modeling of expert opinions) to forecast combination.

39. Bunn, D.W. (1988), "Combining forecasts", *European Journal of Operational Research*, 33, 223–229. **F, RV.** In this thoughtful review, Bunn questions the value of blindly pursuing more accurate forecasts through forecast combination.

40. Bunn, D.W. (1989), "Editorial: Forecasting with more than one model", *Journal of Forecasting*, 8, 161–166. **RV.** Bunn reviews forecast combination research, speculates on reasons for its popularity, and discusses methodological issues.

41. Bunn, D.W., and Kappos, E. (1982), "Synthesis of selection of forecasting models", *European Journal of Operational Research*, 9, 173–180. **F, BY, T, SI.** The authors considered the possibility that, under conditions of sparse data, selection of a forecasting model may be better than combining models. In their simulation, they studied selection and combination with veridical probabilities (probability that a model is correct) and outperformance probabilities. The combinations performed better than the selection of individual forecasts. The paper provides a concise comparison of veridical and outperformance probabilities.

42. Bunn, D.W., and Mustafaoglu, M.M. (1978), "Forecasting political risk", *Management Science*, 24, 1557–1567. **F, EO, T, A, O.** Judgments regarding political risk factors in developing countries from a number of experts were aggregated via a Bayesian method. These composite judgments were then used to generate probabilities of specific risk events.

43. Bunn, D.W., and Seigal, J.P. (1983), "Forecasting the effects of television programming upon electricity loads", *Journal of the Operational Research Society*, 34, 17–25. **F, R, A, E.** This paper reports a straightforward application of combining two forecasts. The authors used a variance-minimizing approach.

44. Bunn, D.W., and Topping, I. (1984), "Efficiency of the independence assumption in the combination of forecasts", *Operations Research Letters*, 3, 173–178. **F, T, SI**. This is a short and intriguing simulation study of the effect of not using the estimated correlation coefficient in combining two dependent forecasts. For small data sets (less than 30 observations), positive correlations, and roughly equal variances, the independence assumption improves forecast performance.

45. Chong, Y.Y., and Hendry, D.F. (1986), "Econometric evaluation of linear macro-economic models", *Review of Economic Studies*, 53, 671–690. **F, R, T**. The authors show how the concept of encompassing can be related to combining forecasts and used for forecast evaluation. The technique is closely related to the idea of informational efficiency (Nelson, 1972).

46. Clemen, R.T. (1985), "Extraneous expert information", *Journal of Forecasting*, 4, 329–348. **F, BY, T, A, M**. This paper develops the notion that some experts may provide information that is extraneous if viewed in the light of other information. The idea is developed along Bayesian lines, and is closely related to the idea of sufficiency. The approach was applied to a large sample of precipitation probability forecasts to determine whether one of two forecasts was extraneous. Clemen and Murphy (1986a, b) report a more extensive study along the same lines.

47. Clemen, R.T. (1986), "Linear constraints and the efficiency of combined forecasts", *Journal of Forecasting*, 5, 31–38. **F, R, T, A, E**. Clemen argues that imposing contraints on combining weights may be justified in terms of increased forecasting efficiency. The empirical study of GNP forecasts, using an adaptive technique to calculate combining weights, showed that the constrained forecast yielded slightly better results than the unconstrained combination. Trenkler and Liski (1986) extend Clemen's results.

48. Clemen, R.T., and Guerard, J. (1989), "Econometric GNP forecasts: Incremental information relative to naive extrapolation", *International Journal of Forecasting*, 5, 417–426. **F, R, BY, T, A, E**. This paper extends the 'equivalent independent experts' model of Clemen and Winkler (1985) and uses it to examine GNP forecasts relative to a naive extrapolation forecast (random walk with drift). The results indicate that one econometric forecast is worth obtaining for the current quarter. However, for more distant horizons the econometric forecasts add little. Also, it appears to be not worthwhile to obtain more than one econometric forecast.

49. Clemen, R.T. and Murphy, A.H. (1986a), "Objective and subjective precipitation probability forecasts: Statistical analysis of some interrelationships", *Weather and Forecasting*, 1, 56–65. **F, EO, A, M**. This paper asks two questions: (1) Do subjective probability of precipitation forecasts contain information not included in objective forecasts? (2) Do the subjective forecasts make full use of the information in the objective forecast? The answers are yes and no, respectively, based on analysis developed by Clemen (1985).

50. Clemen, R.T., and Murphy, A.H. (1986b), "Objective and subjective precipitation probability forecasts: Some methods for improving forecast quality", *Weather and Forecasting*, 1, 213–218. **F, EO, A, M**. The authors show that it is possible to improve the performance of National Weather Service probability of precipitation forecasts through calibration and combination techniques.

51. Clemen, R.T., and Winkler, R.L. (1985), "Limits for the precision and value of information from dependent sources", *Operations Research*, 33, 427–442. **S, BY, T**. The authors develop the notion of 'equivalent independent experts', and show how dependence among experts can reduce the aggregate amount of information. The theory is extended in Schmittlein, Kim and Morrison (1988), and applied by Clemen and Guerard (1989) and Murphy, Chen and Clemen (1988).

52. Clemen, R.T., and Winkler, R.L. (1986), "Combining economic forecasts", *Journal of Business and Economic Statistics*, 4, 39–46. **F, R, BY, A, E**. This paper reports the results of combining forecasts of real and nominal US GNP by four econometric forecasters. The results showed that equal weights performed well. A combination formula that ignored correlations also worked well, as did a Bayesian shrinkage model.

53. Clemen, R.T., and Winkler, R.L. (1987), "Calibrating and combining precipitation probability forecasts", in: R. Viertl (Ed.), *Probability and Bayesian Statistics* (Plenum, New York) 97–110. **F, S, BY, A, M**. Clemen and Winkler develop normal-log-odds models for calibrating and combining probability forecasts. Application of the models to a large set of probability forecast data indicated that the models work poorly relative to not calibrating (for individual forecasts) or simple averages (for multiple forecasts).

54. Conroy, R., and Harris, R. (1987), "Consensus forecasts of corporate earnings: Analysts' forecasts and time series methods", *Management Science*, 33, 725–738. **F, TS, EO, A, BS**. The authors' results, based on an extensive analysis of the Intitutional Brokers Estimate System (IBES), suggest that earnings forecasts can be improved though combination of analyst and time series forecasts.

55. Cooke, J.K. (1967), "Clinicians' decisions as a basis for deriving actuarial formulae", *Journal of Clinical Psychology*, 23, 1967, 232–233. **P, EO, X, O**. In an empirical examination of the ability to distinguish between psychiatric and non-psychiatric individuals, the authors found that a combination of judgments was superior to individual judgements.

56. Cooper, J.P., and Nelson, C.R. (1975). "The ex-ante prediction performance of the St. Louis and FRB-MIT-PENN econometric models, and some results on composite predictions", *Journal of Money, Credit and Banking*, 7, 1–32. **F, TS, R, C, A, E**. This paper develops the idea of a composite predictor, and looks at the incremental value of forecasts through the significance of t-statistics in the combining regression.

57. Cragg, J., and Malkiel, B. (1968), "The consensus and accuracy of some predictions of the growth in corporate earnings", *Journal of Finance*, 23, 67–84. **F, TS, EO, X, BS**. The primary conclusion of this paper is that analysts are not much

more successful at predicting earnings than are naive extrapolation forecasts. The authors do not combine forecasts, but they do address the extent to which the analysts' forecasts are correlated.

58. Crane, D.B., and Crotty, J.R. (1967), "A two-stage forecasting model: Exponential smoothing and multiple regression", *Management Science*, 13, B501–B507. **F, TS, R, A, BS.** The authors constructed a method for merging exponential smoothing models and multiple regression. The approach is adaptive: it reestimates regression coefficients as new data become available. The forecast results were acceptable (see Bonini and Freeland, 1979).

59. Dalrymple, D.J. (1978), "Using Box–Jenkins techniques in sales forecasting", *Journal of Business Research*, 6, 133–145. **F, TS, R, A, BS**, This is mainly an evaluation of Box–Jenkins forecasting. In terms of combining forecasts, Dalrymple concludes that instability of the models is one of the reasons for the improved performance that results from combining Box–Jenkins with other forecasts.

60. Dalrymple, D.J. (1987), "Sales forecasting practices", *International Journal of Forecasting*, 3, 379–391. **F, BS.** Dalrymple includes data regarding the use of combined forecasts in practice. Almost 40% of firms surveyed frequently or usually combine forecasts. Dalrymple contrasts his results with the claim by PoKempner and Bailey (1970) that combining was a common practice.

61. Dickinson, J.P. (1973), "Some statistical results on the combination of forecasts", *Operational Research Quarterly*, 24, 253–260. **F, S, C, T.** Dickinson considers the sampling distribution for combining weights, and concludes that the poor reliability of the weight estimates may reduce the practical usefulness of the combination procedure. His analysis foreshadows ideas of Kang (1986).

62. Dickinson, J.P. (1975), "Some comments on the combination of forecasts", *Operational Research Quarterly*, 26, 205–210. **F, S, C, T.** This paper continues the theoretical discussion, begun in Dickinson (1973) of statistical properties of the weight estimators. An explanation of negative weights is included.

63. Dickinson, J.P. (1988), "Company forecast accuracy for exponential smoothing models of earnings-per-share data for financial decision making: A comment", *Decision Sciences*, 19, 233–235. **F, T.** Dickinson argues that Brandon, Jarrett and Khumawala (1986) should have advocated the combination of forecasts. His analysis, however, ignores the error involved in estimating the correlation coefficient.

64. Diebold, F.X. (1988), "Serial correlation and the combination of forecasts", *Journal of Business and Economic Statistics*, 6, 105–111. **F, TS, T, SI.** Diebold assumes that the individual forecast errors are white noise and are contemporaneously and serially uncorrelated. In spite of this, he shows that regression-based combination methods can lead to serially correlated combined forecast errors. However, if the regression is constrained so that the weights sum to one, then it follows that the combined errors are not serially correlated.

65. Diebold, F.X., and Pauly, P. (1986), "The combination of forecasts", *Previsions et Analyse Économique*, 7, 7–31. **F, TS, R, S, C, T, SI.** This paper is a re-packaging of the authors' results on combining forecasts with time-varying weights (see Diebold and Pauly, 1987a, b).

66. Diebold, F.X., and Pauly, P. (1987a), "The combination of forecasts: A general approach", in: P. Hackl (Ed.), *Adaptive Estimation and Structural Change in Regression and Time Series Analysis* (North-Holland, Amsterdam). **F, TS, R, S, C, T, SI.** The authors combine the results from Diebold (1988) and Diebold and Pauly (1987b) to arrive at a very general time-varying-parameter approach to forecast combination.

67. Diebold, F.X., and Pauly, P. (1987b), "Structural change and the combination of forecasts", *Journal of Forecasting*, 6, 21–40. **F, S, C, T, SI.** The authors show how weighted least squares techniques can be used to model changes in the relative contributions of the forecasters (time-varying combining parameters).

68. Diebold, F.X., and Pauly, P. (1987c), "The use of prior information in forecast combination", Board of Governors of the Federal Reserve System, Special Studies Paper-Division of Research and Statistics, #218. **F, S, BY, T, X, E.** This paper follows up on Clemen and Winkler (1986), but looks at prior information about the combining weights in the unrestricted regression framework of Granger and Ramanathan (1984) and a more general framework involving forecasts of more than one variable. Shrinkage of the combining weights toward equal weights provides the best improvement in forecast accuracy.

69. Doyle, P., and Fenwick, I.A. (1976), "Sales forecasting – Using a combination of approaches", *Long-Range Planning*, 9, 60–64. **F, RV.** This is a very basic article written for practitioners. The authors argue for equal weights, weights based on past relative accuracy, or subjectively assessed weights. They illustrate the advantage of the simple average with an example.

70. Edgerton, H.A., and Kolbe, L.E. (1936), "The method of minimum variation for the combination of criteria", *Psychometrika* 1, 183–188. **P, S, C, T.** The authors take a very unusual approach to the derivation of combining weights. Their optimality criterion is to minimize the sum of squares of the differences of the standared scores for k different estimates of the same variable.

71. Edmundson, B., Lawrence, M., and O'Connor, M. (1988), "The use of non-time series information in sales forecasting: A case study", *Journal of Forecasting*, 7, 201–211. **F, EO, X, BS.** In a forecasting experiment involving a large multinational company, the authors found that the forecasts generated by the company for sales of eighteen products were more accurate (lower MAPE) than both time series and judgemental extrapolation forecasts as well as the average of these forecasts. Given this, the authors advocate including specific product information in the formulation of sales forecasts.

72. Einhorn, H.J. (1972), "Expert measurement and mechanical combination", *Organizational Behavior and Human Perfor-*

mance, 7, 86–106. **P, EO, E, O**. This paper blends ideas from the earlier bootstrapping literature with the notion of combining predictions from multiple judges.

73. Einhorn, H.J., and Hogarth, R.M. (1975), "Unit weighting schemes for decision making", *Organizational Behavior and Human Performance*, 13, 171–192. **P, S, T**. The authors study the performance of equally-weighted composites relative to regression weights. They discuss conditions under which equal weights outperform regression weights (see their figure 3). Reasons for the strong performance of equal weights are that there is no estimation error, no degrees of freedom are lost, and 'true' relative weights cannot be reversed.

74. Einhorn, H.J., Hogarth, R.M., and Klempner, E. (1977), "Quality of group judgment", *Psychological Bulletin*, 84, 158–172. **P, T, X, O**. The authors compare group judgments with four different theoretical models for the combination of individual information. The models include choosing a group member randomly, choosing the best member (post hoc), choosing the best but possibly making an error, and taking an average. These models provide baselines for evaluating the performance of group judgments.

75. Elton, E.J., Gruber, M.J., and Gultekin, M. (1981), "Expectations and share prices", *Management Science*, 27, 975–987. **EO, A, BS**. The authors use consensus forecasts of earnings per share (averages of 3 or more analyst forecasts) in a study of stock market efficiency.

76. Engle, R.F., Granger, C.W.J., and Kraft, D.F. (1984), "Combining competing forecasts of inflation using a bivariate ARCH model", *Journal of Economic Dynamics and Control*, 8, 151–165. **F, C, T, A, E**. The authors use a bivariate ARCH model to combine two forecasts. The ARCH model provides for time-varying parameters. However, their combination did not improve on the performance of a fixed-weight combination of inflation forecasts.

77. Eysenck, H.J. (1939), "The validity of judgments as a function of the number of judges", *Journal of Experimental Psychology*, 25, 650–654. **P, X, O**. In a critique of Gordon (1924), the author uses results from test theory to show that aggregating judgments from members of a population leads to higher correlation of those judgments with the judgments of an independent sample from the same population.

78. Falconer, R.T., and Sivesind, C.M. (1977), "Dealing with conflicting forecasts: The eclectic advantage", *Business Economics*, 12, 5–11. **F, RV, A, E**. This is a basic practitioner-oriented paper. The authors go out of their way to argue that composite forecasts only reduce uncertainty, but do not eliminate it. They emphasize the use of complementary techniques (e.g., combining time series and econometric forecasts).

79. Ferrell, W.R. (1985), "Combining individual judgments", in: G. Wright (Ed.), *Behavioral Decision Making*, (Plenum, New York), 111–145. **F, P, S, EO, RV**. This chapter in George Wright's reader is itself a review paper on models for combining judgments. It covers most of the basic results in mathematical and behavioral aggregation research.

80. Figlewski, S. (1983), "Optimal price forecasting using survey data", *Review of Economics and Statistics*, 65, 13–21. **F, EO, T, A, E**. Figlewski uses the Livingston data and combines forecasts of CPI. He develops both a 'diagonal' (zero correlation) and a 'single-index' model. The single index model restricts the nature of the covariance matrix, and is similar to the market model in finance. His results using the single-index model were a substantial improvement over the simple average forecast. His tables 2 and 3 suggest that correcting for bias would be appropriate, but the author argues against correcting for bias on rational expectations grounds.

81. Figlewski, S., and Urich, T. (1983), "Optimal aggregation of money supply forecasts: Accuracy, profitability and market efficiency", *Journal of Finance*, 28, 695–710. **F, EO, T, A, E**. In combining money supply forecasts, the simple average performed well. The authors derive a composite forecast for location-biased forecasters. It appears from their results that one could make money trading in Treasury bills and futures.

82. Fildes, R. (1985), "Quantitative forecasting – The state of the art: Econometric models", *Journal of the Operational Research Society, 36, 549–580.* **F, TS, R, RV**. Combining forecasts is mentioned briefly on pages 575–576. Fildes points out that an alternative to combining forecasts is to "ask how the various information sources can best be used" (p. 576), suggesting that it may be possible to use the forecasts in some way other than just combining them.

83. Fildes, R., and Fitzgerald, M.D. (1983), "The use of information in balance of payments forecasting", *Economica*, 50, 249–258. **F, EO, TS, X, E**. The authors examine the comparative accuracy of three judgmental forecasters and two ARIMA models as well as composite forecasts. The correlations among their three judgmental forecasters ranged from 0.79 to 0.95. The composite forecast improved on the individual forecasts in terms of RMSE, indicating that the ARIMA model included some information (in a statistical sense) not considered by the forecasters. The authors also consider some aspects of expectations formations and test for rational expectations of the forecasters.

84. Flores, B.E., and White, E.M. (1988), "A framework for the combination of forecasts", *Journal of the Academy of Marketing Science*, 16, 95–103. **RV**. Flores and White review the literature paying particular attention to the kinds of forecasts that have been combined and the methods that have been used. They note that most research has been in the area of systematic combinations of objective forecasts and call for more research effort involving intuitive combinations and/or subjective forecasts.

85. Flores, B.E., and White, E.M. (1989), "Subjective vs. objective combining of forecasts: An experiment", *Journal of Forecasting*, 8, 331–341. **F, X, EO, E**. The authors report on a small experiment in which subjects generated subjective forecasts and then created subjective consensus forecasts within 2- and 3-person groups. The performance improvement over the individual forecasts was slight, and there appeared to be very little difference in the performance of the subjective and objective combination methods.

86. French, S. (1980), "Outranking probabilities and the synthesis of forecasts", *Journal of the Operational Research Society*, 31, 545–551. **F, S, BY, T**. This reference actually comprises an interchange between French and Bunn, in which French attacks the outperformance-based composite forecast. Much of the discussion revolves around the appropriateness of the outperformance approach versus a veridical combination. This prompted the 1981 paper by Bunn in the same journal.

87. French. S. (1981), "Linear combination of forecasts – A comment", *Journal of the Operational Research Society*, 32, 937–938. **F, T**. French argues that the simulation study in Bunn (1981) was flawed, although this was refuted by Bunn (p. 1145) in a subsequent issue.

88. Fritz, R., Brandon, C., and Xander, J. (1984), "Combining time-series and econometric forecast of tourism activity", *Annals of Tourism Research*, 11, 219–229. **F, TS, R, A, O**. The authors combined econometric and time series forecasts of tourism. The improvement in accuracy due to combining was enhanced by an ad hoc modification of the weights to account for error variability relative to forecast variability.

89. Fuhrer, J., and Haltmaier, J. (1988), "Minimum variance pooling of forecasts at different levels of aggregation", *Journal of Forecasting*, 7, 63–73. **F, T, A, E**. The authors develop a statistical technique for pooling forecasts that are at different levels of aggregation.

90. Gardner, E.S., Jr. (1979), "A note on forecast modification based upon residual analysis", *Decision Sciences*, 10, 493–494. **F, TS, R, T**. This paper is a critique of Mabert (1978). Mabert's reply follows.

91. Geisser, S. (1965), "A Bayes approach for combining correlated estimates", *Journal of the American Statistical Association, Series A*, 60, 602–607. **S, BY, T**. This paper appears to be the earliest Bayesian approach to combining dependent forecasts or estimates.

92. Geurts, M.D., and Wheeler, G.E. (1980), "Converging conflicting research findings: The Oregon bottle bill case", *Journal of Marketing Research*, 17, 552–557. **S, EO, A, O**. This paper describes an application of DeGroot's (1974) consensus model. Two experts were asked to review research findings and report probability distributions for three variables. They were also asked to rate each other. These assessments formed the inputs to the consensus model.

93. Goldberg, L.R. (1965), *Diagnosticians Versus Diagnostic Signs: The Diagnosis of Psychosis vs. Neurosis from MMPI*, Psychological Monographs, 79. **P, EO, X, O**. In studying psychiatric judgments, Goldberg found that combining scores from multiple staff members provided predictions that were almost as good as that of the best judge in a group of 29 clinicians.

94. Goldberg, L.R. (1970), "Man versus model of man: A rationale, plus some evidence for a method of improving on clinical interferences", *Psychological Bulletin*, 73, 422–432. **P, EO, X, O**. This paper is concerned with bootstrapping, or modeling expert opinions using linear models. Goldberg also considered pooling of expert judgments.

95. Gordon, K. (1924), "Group judgments in the field of lifted weights", *Journal of Experimental Psychology*, 7, 398-400. **P, X, O**. The conclusion was that by averaging the orders (of weighted objects) assigned by individuals, the correlation with the actual order becomes higher. The group results were better than those of the average member, and were at least as good as those of the best members.

96. Gordon, K. (1935), "Further observations on group judgments of lifted weights", *Journal of Psychology*, 1, 105–115. **P, X, O**. Gordon responds to Stroop's (1932) criticism of her previous experiment.

97. Granger, C.W.J. (1989), "Combining forecasts – Twenty years later", *Journal of Forecasting*, 8, 167–173. **F, T, RV**. Granger selectively reviews some of the theoretical developments in the field of combining forecasts. Topics include nonstationarity, combining forecasts with different lead times, and combining confidence intervals.

98. Granger, C.W.J., and Newbold, P. (1973), "Some comments on the evaluation of forecasts", *Applied Economics*, 5, 35–47. **F, C, T**. This paper develops the idea of 'conditional' information contained in a forecast. That is, a decision maker can evaluate one forecast relative to another by asking whether the incremental forecast adds any information in a combined forecast. Their approach involves the use of incremental explanatory power (increases in R^2).

99. Granger, C.W.J., and Newbold, P. (1977), *Forecasting Economic Time Series* (Academic Press, London). **F, TS, T, A, E**. Chapter 7 is about combining forecasts. Much of it is a review of Newbold and Granger (1974).

100. Granger, C.W.J., and Ramanathan, R. (1984), "Improved methods of forecasting", *Journal of Forecasting*, 3, 197–204. **F, C, T, A, E**. The authors show that standard combining methods are equivalent to constrained regression where the combining weights are constrained to sum to one and the intercept is suppressed. They argue for an unconstrained combination of forecasts.

101. Greene, M.N., Howrey, E.P., and Hymans, S.H., (1985)", The use of outside information in econometric forecasting', in: Kuh and D. Belsley (Eds), *Model Reliability*, (Wiley, New York) 90–116. **F, S, C, T, A, E**. The authors use theory from combining forecasts to develop a model for incorporating outside information into a forecast. They show that in a nonlinear system, 'optimal' combining weights will vary over time.

102. Guerard, J.B. (1987), "Linear constraints, robust-weighting and efficient composite modeling", *Journal of Forecasting*, 6, 193–199. **F, TS, EO, A, BS**. This study is an application of combining techniques for forecasting earnings. Guerard introduces robust statistical estimation of the combining weights.

103. Guerard, J.B. (1989), "Composite model building for foreign exchange rates", *Journal of Forecasting*, 8, 315–329. **F, A, E**. Guerard presents an example in which biased regression methods are useful for combining highly collinear forecasts. The combined forecast was not able to improve on the forward interest rate.

104. Guerard, J.B., and Beidleman, C.R. (1986), "A new look at forecasting annual corporate earnings in the U.S.A.", *European Journal of Operational Research*, 23, 288–293. **F, A, BS**. In this study of corporate earnings forecasts, the authors found that by combining analyst forecasts with ARIMA (0, 1, 1) forecasts, average mean square forecasting error over 35 firms was reduced by 74.2%. This result seems unusual in light of other published research.

105. Guerard, J.B., and Beidleman, C.R. (1987), "Composite earnings forecasting efficiency", *Interfaces*, 17, 103–113. **F, TS, EO, A, BS**. The authors compare OLS and ridge regression for combining analysts' forecasts of earnings with time series forecasts. They used the Treynor index to study the extent to which economic profits could be earned using composite forecasts as a basis for portfolio management.

106. Guerard, J.B., and Clemen, R.T. (1989), "Collinearity and the use of latent root regression for combining GNP forecasts", *Journal of Forecasting*, 8, 231–238. **F, R, A, E**. The authors used latent root regression (LRR) to combine highly correlated economic forecasts. LRR did not work as well as OLS regresion or a simple average.

107. Guerrero, V.M. (1989), "Optimal conditional ARIMA forecasts", *Journal of Forecasting*, 8, 215–229. **F, TS, T**. Combining an ARIMA forecast with additional information is discussed within the context of creating a conditional ARIMA forecast.

108. Gulledge, T.R., Jr., Ringuest, J.L., and Richardson, J.A. (1986), "Subjective evaluation of composite econometric policy inputs", *Socio-Economic Planning Sciences*, 20, 51–55. **F, A, E**. This paper demonstrates the use of multiple objective linear programming to combine economic forecasts. The authors did not evaluate the performance of the combined forecast.

109. Gunter, S.I., and Aksu, C. (1989), "*N*-step combinations of forecasts", *Journal of Forecasting*, 8, 253–267. **F, TS, A, E**. The authors introduce the idea of combining different kinds of forecast combinations. In a test using GNP data, these combinations show slightly improved forecasting performance.

110. Gupta, S., and Wilton, P.C. (1987), "Combination of forecasts: An extension", *Management Science*, 33, 356–372. **F, SI**. The authors' method for combining forecasts is similar to Saaty's Analytic Hierarchy Process and Bunn's earlier outperformance technique, using a subjective assessment of the odds matrix. It guarantees non-negative weights and appears to be quite robust.

111. Gupta, S., and Wilton, P.C. (1988), "Combination of economic forecasts: An odds-matrix approach", *Journal of Business and Economic Statistics*, 6, 373–379. **F, A, E**. The authors test their odds-matrix method for combining forecasts using the GNP forecasts from Clemen and Winkler (1986). The results show that their method performs better than the individual forecasts, and is particularly useful when sample sizes are small.

112. Hafer, R.W., and Hein, S.E. (1985), "On the accuracy of time-series, interest rate, and survey forecasts of inflation", *Journal of Business*, 58, 377–398. **F, TS, A, E**. The analysis shows convincingly that neither an ARIMA nor an interest rate inflation forecast contains incremental information relative to the median inflation forecast from the ASA-NBER survey.

113. Hallman, J., and Kamstra, M. (1989), "Combining algorithms based upon robust estimation techniques and cointegrating restrictions", *Journal of Forecasting*, 8, 189–198. **F, R, A, E**. The authors consider incorporating cointegrating restrictions on the combining weights, resulting in improved forecast combinations for integrated series. They also use an encompassing procedure to rank forecasts.

114. Halperin, M. (1961), "Almost linearly-optimum combination of unbiased estimates", *Journal of the American Statistical Association*, 56, 36–43. **S, C, T**. This is an early statistical treatment of the combination of multiple dependent estimates of a single unknown parameter.

115. Hogarth, R. (1977), "Methods for aggregating opinions", in: H. Jungermann and G. de Zeeuw (Eds.), *Decision Making and Change in Human Affairs*, (Reidel, Dordrecht, Holland) 231–255. **P, RV**. Hogarth reviews literature in psychology and decision theory that is pertinent to the aggregation of opinions.

116. Hogarth, R.M. (1978), "A note on aggregating opinions", *Organizational Behavior and Human Performance*, 21, 40–46. **F, P, EO, T**. Using arguments from test theory, Hogarth concludes that one should use between 6–20 different forecasts. The more the forecasters differ, the more forecasters should be included in the combination.

117. Holden, K., and Peel, D.A. (1986a), "An empirical investigation of combinations of economic forecasts", *Journal of Forecasting*, 5, 229–242. **F, TS, R, A, E**. The authors perform a test of unconstrained (Granger and Ramanathan, 1984) vs. constrained (Clemen, 1986) forecast combination in which the constrained method performs slightly better. The simple average also performs well.

118. Holden, K., and Peel, D.A. (1986b), "Expectations formation, public forecasts and the wage equation", *Economic Modelling* (April), 129–134. **F, R, A, E**. The authors argue that agents may act on available consensus forecasts, and that such forecasts can serve as a reasonable proxy for price expectations in an econometric model.

119. Holden, K., and Peel, D.A. (1988a), "Combining economic forecasts", *Journal of the Operational Research Society*, 39, 1005–1010. **F, R, A, E**. The authors combined five different forecasts for growth and inflation in the UK. The combination involved constraining the weights and including a constant term, and the performance of the combined forecast was compared to the average of the forecasts.

120. Holden, K., and Peel, D.A. (1988b), "The accuracy of forecasts of the UK economy", *Prévision et Analyse Économique (Cahiers du GAMA)*, 7, no. 3, 35–52. **F, R, A, E.** The authors analyze forecasts of the UK economy reported in *The Investor's Chronicle*. Their analysis shows that a simple average outperformed an 'optimal' combination.

121. Holden, K., and Peel, D.A. (1989a), "Unbiasedness, efficiency and the combination of economic forecasts", *Journal of Forecasting*, 8, 175–188. **F, R, T, A, E.** The authors consider the problem of testing for biasedness of component forecasts. Their analysis leads to the conclusion that combining weights whould be constrained to sum to one and a constant term should be included to account for bias.

122. Holden, K., and Peel, D.A. (1989b), "A comparison of some inflation, growth and unemployment forecasts", *Journal of Economic Studies*, 15, 48–55. **F, R, A, E.** The authors analyze economic forecasts from three major econometric models of the UK economy. They find some forecasting performance improvement as a result of averaging the three forecasts.

123. Holmen, J.S. (1987), "A note on the value of combining short-term earnings forecasts", *International Journal of Forecasting*, 3. 239–243. **F, TS, SI.** This is a straightforward simulation study of Granger and Ramanathan's (1984) three combining methods. The unconstrained combination performed slightly better than the others.

124. Holt, M.T., and Brandt, J.A. (1985), "Combining price forecasting with hedging of hogs: An evaluation using alternative measures of risk", *Journal of Futures Markets*, 5, 297–309. **F, TS, R, A, E.** The authors ask whether it is possible to use forecasts and a hedging strategy in the hog market to earn excess profits beyond a straight cash strategy. The answer is affirmative, with a combined econometric-ARIMA forecast giving the most economic benefits.

125. Horst, P. (1938), "Obtaining a composite measure from a number of different measures of the same attribute", *Psychometrika*, 1, 53–60. **S, T.** Horst derives a formula for combining multiple measures. His criterion is obtaining maximum separation among the individual population members.

126. Kang, H. (1986), "Unstable weights in the combination of forecasts", *Management Science*, 32, 683–695. **F, SI, A, E.** In a study that uses both simulation and economic data, Kang shows that estimated combining weights can be very unstable. Kang thus argues that in practice a simple average is the best composite predictor.

127. Kaplan, A., Skogstad, A.L., and Girshick, M.A. (1950), "The prediction of social and technological events", *Public Opinion Quarterly*, 14, 93–110. **F, EO, X, O.** This paper reports the results of a study of individual and group predictions of events. The study is survey-based, asking experts for predictions concerning real upcoming events.

128. Klugman, S.F. (1947), "Group and individual judgments for anticipated events", *Journal of Social Psychology*, 26, 21–28. **P, X, O.** During World War II, soldiers predicted the dates for the ending of hostilities between the US and Germany and between the US and Japan. Combining the soldiers' predictions was substantially more accurate than the individual predictions.

129. Larréché, J.-C., and Moinpour, R. (1983), "Managerial judgment in marketing: The concept of expertise", *Journal of Marketing Research*, 20, 110–121. **F, P, X, BS.** Among a number of results, the authors found that averaging the individual initial judgments of a group's members gives better estimates than the group consensus. Also, the Delphi process was found to provide even better estimates.

130. Lawrence, K.D., and Geurts, M. (1984), "Converging conflicting forecast parameters on forecasting durable new product sales", *Euopean Journal of Operational Research*, 16, 42–47. **F, T, BS.** The authors develop a way to merge conflicting parameter estimates for a diffusion model used to forecast new product sales. The procedure is based on DeGroot's (1974) consensus model.

131. Lawrence, K.D., and Reeves, G.R. (1981), "Consensus time series forecasting", in: J. Morse (Ed.) *Organizations: Multiple Agents with Multiple Criteria* (Springer–Verlag, New York) 199–204. **F, TS, A, BS.** The authors use goal programming to determine optimal weights for combining forecasts. The goals can be flexibly defined to satisfy a variety of management objectives.

132. Lawrence, M.J., Edmundson, R.H., and O'Connor, M.J. (1985), "An examination of the accuracy of judgmental extrapolation of time series", *Internatinal Journal of Forecasting*, 1, 25–35. **F, TS, EO, A, E.** This paper is primarily about the accuracy of judgmental versus statistical extrapolation forecasting. In their experiment, they tried both graph-based and table-based judgments as well as a combination of the two. The data used were Makridakis's 111 series. In general, these subjective methods performed reasonably well relative to the statistical methods. Furthermore, the results suggested that the combination of judgmental approaches should be able to outperform the single judgmental methods.

133. Lawrence, M.J., Edmundson, R.H., and O'Connor, M.J. (1986), "The accuracy of combining judgemental and statistical forecasts", *Management Science*, 32, 1521–1532. **F, EO, TS, A, E.** Following up on the authors' earlier work (Lawrence, Edmunson, and O'Connor, 1985), judgmental and statistical forecasts are combined. Judgmental forecasts are found to contribute to improved forecasting performance. Mechanical combination was superior to judgmental combination or adjustment. The experiment was based on Makridakis's 111 series.

134. Libby, R. (1976), "Man versus model of man: Some conflicting evidence", *Organizational Behavior and Human Performance*, 16, 1–12. **P, X, BS.** Libby's results in this experiment pertaining to the prediction of business failures from financial ratio information conflicts with previous results. Judges outperformed the bootstrap model, the composite judge and most accurate judge outperformed the most accurate model, and the

average judge outperformed the average model. A reply by Goldberg and a rebuttal by Libby follow. Goldberg shows that, upon transforming the data, the conflict vanishes.

135. Libby, R., and Blashfield, R.K. (1978), "Performance of a composite as a function of the number of judges", *Organizational Behavior and Human Performance*, 21, 121–129. **P, EO, X, BS, O**. This paper reports the psychological counterpart to the Makridakis and Winkler (1983) finding that the majority of the improvement in forecasting accuracy is obtained in the combination of the first two or three forecasts.

136. Lock, A. (1987), "Integrating group judgments in subjective forecasts", in: G. Wright and P. Ayton (Eds.), *Judgmental Forecasting* (Wiley, New York), 109–127. **F, P, EO, RV**. This paper reviews literature on both aggregation of forecasts and group judgments. For group judgmental forecasting, the author proposes a 7-step approach that integrates principles from the group process literature.

137. Longbottom, J.A., and Holly, S. (1985), "The role of time series analysis in the evaluation of econometric models", *Journal of Forecasting*, 4, 75–87. **F, TS, R, A, E**. The authors compare forecasts from structural econometric and time series models, suggesting that such comparisons can provide insight for the respecification of the econometric model. They also suggest combining the two forecasts as a stopgap measure to improve performance while the task of respecifying the econometric model is under way.

138. Lupoletti, V.M., and Webb, R.H. (1986), "Defining and improving the accuracy of macroeconomic forecasts: Contributions from a VAR model", *Journal of Business*, 59, 263–285. **F, R, A, E**. The authors suggest using a VAR model as a benchmark against which to compare macroeconomic forecasts. The empirical results indicate that combining the VAR model with other macro models pays off.

139. Mabert, V.A. (1978), "Forecast modification based on residual analysis: a case study of check volume estimation", *Decision Sciences*, 9, 285–296. **F, R, TS, A, BS**. Mabert combined a simple adaptive response exponential smoothing technique with a regression model to track historical check volume. Critiqued by Gardner (1979).

140. Mahmoud, E. (1982), "Short-term forecasting: matching techniques to tasks. An integrated framework and empirical investigation", Ph.D. Thesis, State University of New York at Buffalo. **F, A, BS**. Mahmoud discusses the use of forecasting in managerial decision making. He shows that accuracy can be improved through careful evaluation of the data, selection of appropriate forecasting techniques, and combining forecasts.

141. Mahmoud, E. (1984), "Accuracy in forecasting: A survey", *Journal of Forecasting*, 3, 139–159. **F, RV**. This review article deals with a broad range of issues in forecasting and forecast accuracy. Pages 149–153 cover combined forecasts.

142. Mahmoud, E., and Makridakis, S. (1988), "The state of the art and future directions in combining forecasts", Unpublished manuscript, University of North Texas. **F, RV**. The authors provide an overview of the field of forecast combination. Of particular interest are appendix 1, which categorizes and gives specific findings for a large number of combining forecast studies, and appendix 2, which summarizes results regarding the relative efficiency of different combining methods.

143. Makridakis, S., and Winkler, R.L. (1983), "Averages of forecasts: Some empirical results", *Management Science*, 29, 987–996. **F, TS, A, E**. This paper presents empirical results from Makridakis' 1001 series pertaining to the performance of averages of forecasts. The results show the effect of including more forecasts in the average.

144. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition", *Journal of Forecasting*, 1, 111–153. **F, TS, A, E**. This paper reports the results of a large-scale forecasting competition. Two combination methods were tried, simple averages and weighted averages based on the error covariance matrix. Compared to other extrapolation techniques, the combinations performed well. The simple average performed better overall than the individual methods used in the combination.

145. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1983), *The Forecasting Accuracy of Major Time Series Methods*, (Wiley, London). **F, TS, A, E**. This book discusses the results of the M-competition. Much of it is a re-packaging of Makridakis and Hibon (1979) and Makridakis, et al., (1982).

146. Marwaha, S., and Peel, D.A. (1985), "Behavior of the Liverpool model with weight given to alternative public forecasts", *Economic Modelling* (January), 33–38. **F, R, A, E**. The authors studied the behavior of the Liverpool macroeconomic model when different kinds of forecasts were used to specify expectations in the model. In particular, they combined public forecasts for this purpose. Their results indicated a large degree of sensitivity of the model and subsequent forecasts to the inputs (forecasts) regarding expectations.

147. Mills, T.C., and Stephenson, M.J. (1985), "Forecasting contemporaneous aggregates and the combination of forecasts: The case of the U.K. monetary aggregates", *Journal of Forecasting*, 4, 273–281. **F, A, E**. The authors found that a linear combination of forecasts of aggregates improved out-of-sample prediction performance. Imposing constraints on the regression combining weights resulted in further (but slight) improvement.

148. Mills, T.C., and Stephenson, M.J. (1987), "A time series forecasting system for the UK money supply", *Economic Modelling, (July), 355–369.* **F, R, TS, A, E**. The authors combined up to three different forecasts of the UK money supply. The best combined forecasts had no constant term and constrained weights. Additional forecasting improvement was obtained by calculating weights recursively.

149. Mizrach, B., and Santomero, A.M. (1986), "The stability of money demand and forecasting through changes in regimes", *Review of Economics and Statistics*, 68, 324-328. **F, TS, A, E.** An OLS combination of three forecasts of money demand outperformed (1) any single forecast, and (2) a combination of two ARIMA forecasting models.

150. Moore, G.H. (1977), "The President's economic report: A forecasting record", *NBER Reporter* (April) 1977, 4–12. **F, EO, A, E.** Moore compares the ASA-NBER median forecasts with forecasts contained in the annual Economic Report of the President. The forecasts were very similar with highly correlated errors. The author stops short of suggesting that the President use the ASA-NBER survey rather than the government forecasts, but it is a possible conclusion.

151. Moriarty, M.M., and Adams, A.J. (1984), "Management judgement forecasts, composite forecasting models, and conditional efficiency", *Journal of Marketing Research*, 21, 239–250. **F, TS, EO, A, BS.** This is an application of Nelson (1972) and the Granger and Newbold (1973) idea of conditional efficiency (see also Clemen and Guerard, 1988). Moriarty and Adams conclude that in their situation, sales forecasting, the management judgment forecasts are conditionally efficient with respect to a composite forecast made up of the judgment forecast and a Box–Jenkins forecast. In the terms of Clemen (1985), they concluded that the management forecast was 'sufficient for' the Box–Jenkins forecast.

152. Morris, M.J. (1977), "Forecasting the sunspot cycle", *Journal of the Royal Statistical Society, Series A*, 140, 437–448. **F, TS, A, O.** Morris combines two forecasts of sunspot activity (the 'outburst' model and an autoregressive model) using minimum-variance weights. The composite forecast performs better than the individual forecasts.

153. Morris, P.A. (1974), "Decision analysis expert use", *Management Science*, 20, 1233–1241. **S, BY, T.** Morris proposes a Bayesian approach to the aggregation of information. The essence of the theory is that a Bayesian decision maker should treat expert testimony as data in a Bayesian updating problem. Assessment of the likelihood function for the information is a major practical difficulty.

154. Morris, P.A. (1977), "Combining expert judgments: A Bayesian approach", *Managment Science*, 23, 679–693. **S, BY, T.** Morris extends the theory presented in his earlier (1974) paper. Normal distribution results are presented. His analysis demonstrates the extent to which aggregation of expert opinions is related to problems of expert calibration.

155. Murphy, A.H., Chen, Y.-S., and Clemen, R.T. (1988), "Statistical analysis of interrelationships between objective and subjective temperature forecasts", *Monthly Weather Review*, 116, 2121–2131. **F, EO, A, M.** This paper applies the methods developed in Clemen and Winkler (1985) and Clemen and Guerard (1988) regarding incremental information to the study of objective and subjective temperature forecasts. The results are similar qualitatively to the results of Clemen and Murphy (1986a). The subjective forecast adds substantial information, but fails to use all of the information contained in the objective forecast.

156. Nelson, C.R. (1972), "The prediction performance of the F.R.B.-M.I.T.-PENN model of the U.S. economy", *American Economic Review*, 62, 902–917. **F, TS, R, A, E.** The author measures the incremental contribution of a forecast through the significance of t-statistics in a regression-based combination of forecasts. Essentially, Nelson asks whether the econometric model adds any information to an ARIMA forecast. The answer was that the econometric forecast was not fully 'rational', or did not include all available information; estimated coefficients in the combining equation were significant for both forecasts. In a post-sample test, ARIMA prediction errors were smaller.

157. Nelson, C.R. (1984), "A benchmark for the accuracy of econometric forecasts of GNP", *Business Economics*, 19, 52–58. **F, TS, R, A, E.** Nelson considers incremental information contained in forecasts of various macroeconomic variables relative to some benchmark (ARIMA) forecasts.

158. Newbold, P., and Granger, C.W.J. (1974), "Experience with forecasting univariate time series and the combination of forecasts (with discussion)", *Journal of the Royal Statistical Society, Series A*, 137, 131–149. **F, TS, T, A, E.** The authors tried a variety of techniques for combining time series forecasts of economic variables. Their results suggest calculating combining weights on the basis of relative precision of the forecasts.

159. Newbold, P., Zumwalt, J.K., and Kannan, S. (1987), "Combining forecasts to improve earnings per share prediction: An examination of electric utilities", *International Journal of Forecasting*, 3, 229–238. **F, TS, EO, A, BS.** In an application of composite forecasting, the authors estimated weights on the basis of cross-sectional data. To avoid negative weights, they performed a stepwise removal of the forecasts with the most negative t-values until all remaining forecasts had positive weights.

160. O'Brien, P.C. (1988), "Analysts' forecasts as earnings expectations", *Journal of Accounting and Economics*, 10, 53–83. **A, BS.** The most current earnings forecast was found to be more accurate than either the mean or median forecast, suggesting that analysts do incorporate information into their forecasts through time.

161. Oliveira, R.A. (1978), "Combining forecasts to predict property values for single-family residences: A comment", *Land Economics*, 54, 524–530. **F, T.** Oliveira critiques Wood (1976), who uses a composite forecast in a property-value prediction task. The critique is based primarily on arguments from Newbold and Granger (1974).

162. Öller, L.-E. (1978), "A method for pooling forecasts", *Journal of the Operational Research Society*, 29, 55–63. **F, T.** Öller develops an ad hoc way for experts to assign credibility weights to their own forecasts. These weights are then used to create composite forecasts.

163. Openshaw, S., and Van der Knaap, G.A. (1983), "Small area population forecasting: Some experience with British models", *Tijdschrift voor Economische en Sociale Geografie*, 74,

291–304. **F, TS, A, O**. The authors developed various extrapolation forecasts of populations for Dutch municipalities. They evaluated a combination of ARIMA and Holt–Winters forecasts, as well. The performance of the composite forecasts apparently was mediocre, being included among neither the best nor worst models.

164. Pereira, B., Coqueiro, R., and Perrota, A. (1989), "Experience in combining subjective and quantitative forecasts of open market rates", *Journal of Forecasting*, 8, 343–348. **F, A, E**. Subjective and statistical interest rate forecasts are combined, giving a 30% performance improvement.

165. Phillips, R.F. (1987), "Composite forecasting: an integrated approach and optimality reconsidered", *Journal of Business and Economic Statistics*, 5, 389–395. **F, S, C, T, A, E**. The author suggests estimating constrained combining weights when the constraints are not seriously violated (see Clemen, 1986). In contrast to most of the 'minimum-variance' combined forecast literature, Phillips's model does not require normally distributed errors.

166. Phillips, R.F. (1988), "Comments on testing for forecast and specification optimality using linear composites", *Journal of Forecasting*, 7, 131–137. **F, S, C, T**. This paper is a theoretical response to all of the 'optimality' theory (e.g., Cooper and Nelson 1975, Nelson 1972).

167. PoKempner, S.J., and Bailey, E. (1970), *Sales Forecasting Practices*, (The Conference Board, New York). **F, RV, BS**. This is a short practitioner-oriented essay about the state of sales forecasting in 1970, based on a survey of Conference Board members. On pages 4 and 8, the authors claim that many firms combine forecasts from more than one technique.

168. Poskitt, D.S., and Tremayne, A.R. (1986), "The selection and use of linear and bilinear time series models", *International Journal of Forecasting*, 2, 101–114. **F, T, A, O**. Bilinear ARMA models include terms that are the product of autoregressive and error terms. The authors develop such models and apply them to the familiar Wolf sunspot data. Combining forecasts, however, gave better performance than any single forecast within this family.

169. Poulos, L., Kvanli, A., and Pavur, R. (1987), "A comparison of the accuracy of the Box–Jenkins method with that of automated forecasting methods", *International Journal of Forecasting*, 3, 261–267. **F, A, E**. The authors designed and implemented an automated forecasting system that combines automatic Box–Jenkins and Holt–Winters forecasts. The automatic system compares favorably with the component forecasts when used to forecast the series from the M-competition.

170. Preston, M.G. (1938), "Note on the reliability and the validity of the group judgment", *Journal of Experimental Psychology*, 24, 462–471. **P, X, O**. On the basis of a card-sorting demonstration, Preston argues that the aggregation effect observed by Gordon (1924) was indeed a statistical artifact rather than psychological.

171. Rausser, G.C., and Oliveira, R.A. (1976), "An econometric analysis of wilderness area use", *Journal of the American Statistical Association*, 71, 276–285. **F, TS, R, S, A, O**. The authors calculated both constrained least-squares weights and adaptive weights to combine econometric and Box–Jenkins forecasts. The adaptive weights were based only on squared errors and ignored covariances. The least-squares weight proved to be more accurate than the adaptive weights. Compare this result with the results of Clemen and Winkler (1986) and Newbold and Granger (1974), who found the opposite. Possible explanations are differences in sample sizes and correlations.

172. Reeves, G.R., and Lawrence, K.D. (1982), "Combining multiple forecasts given multiple objectives", *Journal of Forecasting*, 1, 271–279. **F, TS, T, A, BS**. The authors develop a straightforward multiobjective-linear programming approach to the combination of forecasts. The model incorporates under- and overachievement scores for the combined forecast at each period of time, and permits the forecaster to consider such objectives as: (1) the minimization of total forecast error over all time periods, or (2) the minimization of the maximum forecast error in any individual time period.

173. Reid, D.J. (1968), "Combining three estimates of gross domestic product", *Economica*, 35, 431–444. **F, A, E**. This is one of the very early papers that reports the results of an experiment to combine forecasts 'optimally' on the basis of error variances and covariances. While he does not derive expressions for the combinations in this paper, Reid does calculate combining weights by finding a constrained minimum–variance linear combination of three forecasts.

174. Reid, D.J. (1969), "A comparative study of time series prediction techniques on economic data", Ph.D. thesis (University of Nottingham, Nottingham). **F, TS, T, A, E**. Chapter 7 from this thesis develops the mathematics for combining more than two forecasts. The methods are used to combine multiple forecasts for a number of different economic variables under conditions of nonstationarity. The results show that in some cases it is possible to obtain small forecast improvements through combinations, although often the combination is not as good as the best individual forecast.

175. Reinmuth, J.E., and Geurts, M.D. (1979), "A multideterministic approach to forecasting", in: Makridakis, S., and Wheelwright, S.C. (Eds.) *Forecasting*, Vol. 12, TIMS Studies in the Management Sciences (North-Holland, New York) 203–211. **F, A, E, BS**. This paper suggests using regression to combine multiple forecasts. The Hawaiian tourist forecast errors were negatively correlated; thus the strong performance of the combination (see Crane and Crotty (1967) for an even earlier use of regression for combining forecasts).

176. Ringuest, J.L., and Tang, K. (1987), "Simple rules for combining forecasts: Some empirical results", *Socio-Economic Planning Sciences*, 21, 239–243. **F, A, E**. The authors test a simple average of forecasts, the median forecast, and a focus forecast on four macroeconomic variables. They argue that the mean and median performed relatively poorly because of the tendency for the individual forecasts to cluster either above or below the actual value. The focus forecast performed well, however.

177. Rowse, G.L., Gustafson, D.H., and Ludke, R.L. (1974), "Comparison of rules for aggregating subjective likelihood ratios", *Organizational Behavior and Human Performance*, 12, 274–285. **P, X, O.** In an experimental study, the authors found that equal weighting of subjects' likelihood ratios performed better than other subjectively assessed weights and better than a behaviorally determined consensus.

178. Russell, T.D., and Adam, E.E., Jr, (1987), "An empirical evaluation of alternative forecasting combinations", *Management Science*, 33, 1267–1276. **F, TS, A, E.** This paper develops an ad hoc approach to selection of the best forecast for inclusion in a combination. The selection criterion depends on a ranking scheme based on calculation of various error statistics.

179. Sanders, F. (1963), "On subjective probability forecasting", *Journal of Applied Meteorology*, 2, 191–201. **F, EO, X, M.** Sanders discusses the use of probability forecasting in meteorology. In an experiment, the average of judgments from two forecasters was found to perform better than either individual forecast.

180. Schmidt, J.R. (1979), "Forecasting state retail sales: Econometric vs. time series models", *Annals of Regional Science*, 13, 91–101. **F, TS, R, A, E.** The author developed an ARIMA model for predicting net taxable retail sales in Nebraska. The ARIMA model performed better than an econometric model, and a minimum-variance combination of the two performed slightly better. A simple average was not tried, and the data set was small.

181. Schmitt, R.C. (1954), "An application of multiple correlation to population forecasting", *Land Economics*, 30, 277–279. **F, R, A. O.** In an experiment involving forecasts of city populations, the author found that a two-variable regression equation did not perform as well as the 'ratio method' (forecasting city population as a ratio of forecasted state population). The differences in accuracy between the two methods appear to be small, but it was noted that a combination of the two techniques did not improve on the ratio method.

182. Schmittlein, D.C., Kim, J., and Morrison, D.G. (1989), "Combining forecasts: Operational adjustments to theoretically optimal rules", Working paper (The Wharton School, University of Pennsylvania). **S, BY, T, SI.** The authors do two things in this paper. First, they provide an interpretation of the Clemen–Winkler (1985) model in terms of the Mahalanobis distance of the forecasts from the origin. Second, they explain how to use Akaike's information criterion to decide which of several possible models to estimate, or how many parameters to include in the covariance matrix.

183. Schnaars, S.P. (1986a), "An evaluation of rules for selecting an extrapolaton model on yearly sales forecasts", *Interfaces*, 16, 100–107. **F, TS, A, BS.** Schnaars studied the performance of six rules for selecting among seven different extrapolation models. Four rules selected single forecasts, one averaged all forecasts, and the last averaged only those forecasts chosen by the first four selection rules. In a study of 103 sales forecast series, all of the selection procedures improved on the performance of the individual models, and the second combination (combination of selected models) performed the best.

184. Schnaars, S.P. (1986b), "A comparison of extrapolation models on yearly sales forecasts", *International Journal of Forecasting*, 2, 71–85. **F, TS, A, BS.** Schnaars replicated the findings of the M-competition in a study of unit sales forecasting. Specifically, combinations performed better than individual forecasts, and a weighted average did not perform significantly better than a simple average. All but one of the series had less than 30 data points.

185. Sessions, D.N. (1987), "The combining of forecasts using recursive techniques with non-stationary weights", Ph.D. Thesis (New York University). **F, EO, R, T, A, E.** This thesis provides more details regarding the theory and tests reported in Sessions and Chatterjee (1989).

186. Sessions, D.N., and Chatterjee, (1989), "The combining of forecasts using recursive techniques with non-stationary weights", *Journal of Forecasting*, 8, 239–251. **F, EO, R, T, A, E.** The authors test a variety of combination methods for nonstationary data. Their results varied depending on the data set. For the Livingston data, the 'optimal' methods actually performed fairly well. However, this was not the case for GNP data from a number of econometric models, for which the average performed well.

187. Sewall, M.A. (1981), "Relative information contributions of consumer purchase intentions and management judgment as explanators of sales", *Journal of Marketing Research*, 18, 249–253. **F, EO, A, BS.** The author used forecast combination techniques to study the incremental information contained in consumer intention surveys and management forecasts.

188. Smith, D. (1989), "Combination of forecasts in electricity demand prediction", *Journal of Forecasting*, 8, 349–356. **F, A, BS.** Box–Jenkins and spectral decomposition forecasts of electricity demand are combined.

189. Staël Von Holstein, C.-A.S. (1971), "An experiment in probabilistic weather forecasting", *Journal of Applied Meteorology*, 10, 635–645. **F, X, M.** The author tried aggregating probability forecasts in various ways.

190. Staël Von Holstein, C.-A.S. (1972), "Probabilistic forecasting: An experiment related to the stock market", *Organizational Behavior and Human Performance*, 8, 139–158. **F, EO, X, E.** The author used a variety of weighting techniques in determining the consensus distribution. Aggregation rules that attempted to identify the best assessors and give them more weight performed better than the other combination techniques. Scoring rules were used to measure the quality of the assessments.

191. Stroop, J.R. (1932), "Is the judgment of the group better than the average member of the group?", *Journal of Experimental Psychology*, 15, 550–560. **P, X, O.** Stroop argues that Gordon's (1924) interpretation of her results is flawed because

the results follow from a mathematical property of aggregating ratings and calculating correlation coefficients.

192. Su, V., and Su, J. (1975), "An evaluation of the ASA/NBER business outlook survey forecasts", *Explorations in Economic Research*, 2, 588–618. **F, A, E.** The first seven years of the ASA/NBER survey forecasts are evaluated relative to econometric and extrapolation forecasts. The survey forecasts performed significantly better than the extrapolation forecasts. Median as well as mean forecasts were studied; the difference between the two may be related to business cycle developments.

193. Taylor, G.C. (1985), "Combination of estimates of outstanding claims in non-life insurance", *Insurance: Mathematics and Economics*, 4, 81–91. **F, A, O.** The author develops mathematically a general class of combined forecasts and shows how to apply the models in an insurance context.

194. Thorndike, R.L. (1983), "The effect of discussion upon the correctness of group decisions when the factor of a majority influence is allowed for", *Journal of Social Psychology*, 9, 343–362. **P, X, O.** Thorndike asked his subjects to make individual predictions as well as group predictions. The average individual prediction was correct 61.9% of the time, while the combination from groups of four to six people was correct 64.4% of the time.

195. Trenkler, G., and Liski, E.P. (1986), "Note: Linear constraints and the efficiency of combined forecasts", *Journal of Forecasting*, 5, 197–202. **F, S, C, T.** The authors extend the results in Clemen (1986) by developing a technique for testing the appropriateness of constrained combination weights.

196. Virtanen, I., and Yli-Olli, P. (1987), "Forecasting stock market prices in a thin security market", *OMEGA International Journal of Management Science*, 15, 145–155. **F, R, TS, A, E.** The authors used a minimum-variance composite model of econometric and ARIMA forecasts to study stock prices on the Helsinki Stock Exchange.

197. Wall, K.D., and Correia, C. (1989), "A preference based method for forecast combination", *Journal of Forecasting*, 8, 269–292. **F, TS, T, A, E.** The authors take a mathematical programming approach to forecast combination, basing the combination on the decision maker's preferences over different possible error distributions.

198. Wiggins, N., and Kohen, E. (1971), "Man vs. model of man revisited: The forecasting of graduate school success", *Journal of Personality and Social Psychology*, 19, 100–106. **P, EO, X, O.** The authors tested Goldberg's (1970) hypothesis regarding bootstrapping in a clinical situation involving diagnosis of the MMPI. Averaging individual judgments improved accuracy.

199. Winkler, R.L. (1968), "The consensus of subjective probability distributions", *Management Science*, 15, B61–B75. **S, BY, T.** Winkler presents a variety of ad hoc methods for combining probabilities and probability distributions. The methods are compared in a small-scale experiment.

200. Winkler, R.L. (1971), "Probabilistic prediction: Some experimental results", *Journal of the American Statistical Association*, 66, 675–685. **S, EO, BY, A, O.** Winker studied the performance of individual and combined probability assessments regarding the outcomes of football games.

201. Winkler, R.L. (1981), "Combining probability distributions from dependent information sources", *Management Science*, 27, 479–488. **S, BY, T.** Winkler discusses the combination of probability distributions within a Bayesian framework. This work is based on Morris (1974, 1977) and Geisser (1965). A small example used sportswriters predictions of football point spreads.

202. Winkler, R.L., and Makridakis, S. (1983), "The combination of forecasts", *Journal of the Royal Statistical Society, Series A*, 146, 150–157. **F, TS, A, E.** The authors found that the best combinations were those that ignored correlations among errors. These weighted averages outperformed a simple average, but the differences in accuracy were small.

203. Winkler, R.L., Murphy, A.H., and Katz, R.W. (1977), "The consensus of subjective probability forecasts: Are two, three, ..., heads better than one?", *Preprint volume Fifth Conference on Probability and Statistics*, Nov. 15–18, 1977, Las Vegas, Nevada (American Meteorological Society, Boston, MA) 57–62. **F, S, A, M.** The authors present a variety of probability-consensus models and measure the performance of these models in terms of quadratic score for a set of probability forecasts. They also examine the impact of the number of forecasts included in the consensus. The simple average performed well compared to models based on past performance data, and performance improved in general as more forecasts were included in the consensus.

204. Wolfe, H.D. (1966), *Business Forecasting Methods* (Holt, Rinehart and Winston, New York). **F, RV.** Wolfe suggests an 'eclectic' approach of using more than one forecasting method and being sure that the methods are as independent as possible.

205. Wood, S. (1976), "Combining forecasts to predict property values for single-family residences", *Land Economics*, 52, 221–229. **A, O.** Wood shows how a two-forecast composite can be used to predict property values in a real estate appraisal model.

206. Zajonc, R.B. (1962), "A note on group judgments and group size", *Human Relations*, 15, 177–180. **P, RV** Zajonc reviews the early psychology literature on the aggregation of individual judgments.

207. Zarnowitz, V. (1967), "*An appraisal of short-term economic forecasts*" (National Bureau of Economic Research, New York). **F, A, E.** On pages 123–125, the author briefly discusses the superiority of the average of forecasts of GNP, comparing the average to the distribution of the individual forecasts.

208. Zarnowitz, V. (1984), "The accuracy of individual and group forecasts from business outlook surveys", *Journal of Forecasting*, 3, 11–26. **F, EO, A, E.** The author analyses con-

sensus forecasts from the ASA-NBER surveys. The combined forecasts resulted in substantial performance improvements.

209. Zarnowitz, V., and Lambros, L.A. (1987), "Consensus and uncertainty in economic prediction", *Journal of Political Economy*, 95, 591–621. **F, EO, A, E**. The authors used the NBER–ASA data to study the relationship between consensus and uncertainty. Is the degree of dispersion among forecasters indicative of greater forecast uncertainty? The evidence would indicate that there is at most a weak relationship between the two constructs (contrast these results with Einhorn's (1974) theory regarding consensus and expertise – see also Ashton, 1985).