

Note: To search for specific words and phrases in the Acrobat Reader 3.0, use the Find button (binoculars) and the Find Again button (binoculars with a right arrow).

Forecasting Literature 1978 to 1985: Annotations

by J. Scott Armstrong

The following books and articles were discussed in the second edition of *Long-Range Forecasting*, published in 1985. In the few cases where unpublished sources are cited, instructions are provided on how to obtain a copy of the book or article. An attempt was also made to choose the most readable source when multiple sources were available. The list has been pruned to eliminate studies that were about side issues, and papers that are no longer relevant .

This updated bibliography is based primarily on research from 1977-1985 and contains more than 270 items. Some references prior to 1977 are also included, these having been overlooked in the first edition of *Long-Range Forecasting*.

Summaries are provided for most of the items. These summaries were sent to the authors of the papers. I asked the authors whether they were accurate and fair. Almost all authors replied, often sending corrections, and suggestions.

Abdel-Kalik, A. Rashad and El-Sheshai, Kamal N., (1980), "Information choice and utilization in an experiment on default prediction," *Journal of Accounting Research*, vol. 18, pp. 325-342.

Ackoff, Russell L.,(1983), "Beyond prediction and preparation," *Journal of Management Studies*, vol. 1, pp. 59-69.

Adam, Everette E., Jr. and Ebert, Ronald J., (1976). Comparison of human and statistical forecasting," *AIIE Transactions*, vol 8, no. 1, pp. 120-127.

In this experiment, 240 graduate business students made subjective extrapolations for six patterns of simulate~data. The accuracy of these forecasts was compared with that from three extrapolation models. Compared with the models, the human forecasters tended to be biased and were more strongly influenced by random noise in the data. An exponential smoothing model with trend and seasonal components was more accurate than the intuitive extrapolations, except in cases where the data pattern was characterized by trend and low seasonality; here there were no significant differences.

Ahlburg, Dennis A., (1984), "Forecast evaluation and improvement using Theil's decomposition," *Journal of Forecasting*, vo1. 3 pp. 345-351.

This paper discusses the use of Theil's decomposition and presents an analysis of data on annual housing starts. The mechanical adjustment provided major improvements in accuracy for the two-quarter-ahead forecast, and minor improvements for eight-quarters-ahead.

Ahlers, David and Lakinishok, Josef, (1983), "A study of economists' consensus forecasts," *Management Science*, vol. 29, pp. 1113-1125.

This study examines the economic forecasts from J.A. Livingston's survey of about 50 well-known business, public, and academic economists. (See also Keen 1981). The study covers the period from the first survey, in 1947, up to 1978; separate analyses are also provided for 1947-1960, 1961-1969, and 1970-1978. Forecast accuracy was examined for ten macroeconomic variables and for two forecast horizons, 7 and 13 months. Some conclusions supported previous research findings, such as: (1) economists underestimate changes; (2) economists are too optimistic; (3) economists do better than a "no-change" model. (This is my conclusion from the data presented; the authors concluded that this improvement was not significant); (4) economists did no better than a simple trend extrapolation; and (5) forecasts of turning points are of little value in comparison with naive forecasts such as "always predict that the indicator will move in the direction that it generally moves." One surprising conclusion by the authors is that the quality of the economists' forecasts improved over time, but this conclusion was based on only three time periods.

Ajzen, Icek, (1977), "Intuitive theories of events and the effects of base-rate information on prediction," *Journal of Personality and Social Psychology*, vol. 35, pp. 303-324.

Albrecht, W. Steve, Lookabill, Larry L. and McKeown, James, (1977), "The time series properties of annual earnings," *Journal of Accounting Research*, vol. 15, pp. 226-244.

This study compares forecast errors of Box-Jenkins and no-change models.

Alexander, Elmore R., III and Wilkins, Ronnie D., (1982), "Performance rating validity: The relationships of objective and subjective measures of performance," *Group and Organization Studies*, vol. 7, pp. 485-498.

Many previous studies have attempted to predict successful job performance, where job performance was based on subjective measures. Typically, these subjective measures were performance ratings by the workers' immediate supervisor. But are these performance ratings valid indicators of job performance? Alexander and Wilkins reviewed prior research. While performance ratings were related to actual performance in a number of laboratory experiments, this might have been an artifact of the research design. In most of these studies, all things except performance were controlled; thus, there was no other basis for the ratings than performance. Alexander and Wilkins suggested that the interaction between a worker and supervisor may be a relevant variable that was excluded from these studies. It is important, then, to test the validity of subjective ratings in a field setting. They did this, using data on 130 vocational rehabilitation counselors from 23 different groups. Objective measures of output on this job are provided quarterly to the supervisors by the State of Tennessee. The correlations between the subjective ratings and the objective measures were positive, but low (over four criteria, r^2 ran from .01 to .08). In short, subjective ratings of performance are suspect.

Alumbaugh, Richard V., Crigler, M.A. and Dightman, C. R., (1978), "Comparison of multivariate techniques in the prediction of juvenile post-parole outcome," *Educational and Psychological Measurement*, vol. 38, pp. 97-106.

This study compared factor analysis vs. stepwise regression vs. stepwise discriminant function to select variables for predicting recidivism for 579 juvenile cases during the 15-month period after release from parole. The authors used cross-validation and concluded that the stepwise discriminant function was best. They say the poor showing of factor analysis conflicts with a study by Alumbaugh in 1969.

Anderson, Craig A.,(1983a), "Imagination and expectation: The effect of imagining behavioral scripts on personal intentions," *Journal of Personality and Social Psychology*, vol. 45, pp. 293-305.

The subjects in these experiments (114 college students) were asked to prepare behavioral scenarios by drawing cartoons relating to six types of behavior: blood donation, tutoring, taking a new part-time job, running for student-government office, changing academic major, and taking a trip over spring break. The main hypothesis was that in scenarios where the subject is the main character, the subject would change behavioral intentions. Increased intentions were expected for scenarios where the self-as(main(character performed the behavior, decreased intentions if the self-as-main-character did not. This hypothesis draws upon prior research on "availability:" The more often the subject imagined the behavior, the greater the expected change in intentions. The hypothesis was supported, and the study did a convincing job of ruling out competing hypotheses. The second experiment replicated these findings and obtained evidence that the changes persisted over a three-day period. Some questions remain unanswered: Would the changes in intentions lead to changes in behavior? Could these results be applied to a business executive writing scenarios about possible strategic actions she might take for the organization? This is a well-designed and clearly written study on an important topic.

Anderson, Craig A.,(1983b), "Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs," *Journal of Experimental Social Psychology*, vol. 19, pp. 93-108.

Archibald, Robert and Gillingham, Robert, (1980), "An analysis of the short-run consumer demand for gasoline using house-hold survey data," *Review of Economics and Statistics*, vol. 62, pp. 622-628.

Arkes, Hal R. et al.,(1981), "Hindsight bias among physicians weighing the likelihood of diagnoses," *Journal of Applied Psychology*, vol. 66, pp. 252-254.

Hindsight bias was found in this study of 75 physicians. Compared to a control group, physicians given information that an unlikely outcome had occurred were more likely to say they would have predicted that outcome. Implications for forecasting: *Users of forecasts are likely to feel that preparers of forecasts did a poor job when unusual events occur.*

Armstrong, J. Scott, (1979), "Advocacy and objectivity in science," *Management Science*, vol. 25, pp. 423-428.

Armstrong, J. Scott,(1980), "Unintelligible management research and academic prestige," *Interfaces*, vol. 10, pp. 80-86.

This paper shows that journals that are more difficult to read are regarded as more prestigious. Also, readers rated authors as more competent when their papers were written in a complex manner. So, should your forecasting report be clear?

Armstrong, J. Scott,(1981), "Review of *Mail and Telephone Surveys* by Don A. Dillman," *Journal of Business*, vol. 54, pp. 622-625.

Armstrong, J. Scott,(1982), "Strategies for implementing change: An experiential approach," *Group and Organization Studies*, vol. 7, pp. 457-475.

How can one implement new procedures, such as new forecasting methods? The Compu-Heart Case was presented to 16 undergraduate seniors at the Wharton School. They were asked to describe their plans for implementation. Each subject worked individually. Only one subject (6%) suggested a procedure that resembled the Delta Technique. A role playing version of the case was then presented to 15 groups of executives from health care providers. Only one group (7%) used a procedure that resembled the Delta Technique. This group was successful at implementing change while all other groups failed. A different group of subjects was then given brief instruction (five to ten minutes) on the use of the Delta Technique. Of these 14 groups, two encountered difficulty in applying the rules and were unsuccessful in their change efforts. The other 12 groups (86%) were all successful in gaining commitment to change.

Armstrong, J. Scott,(1982), "The forecasting audit," in Spyros Makridakis and Steven C. Wheelwright (Eds.), *The Handbook of Forecasting: A Manager's Guide*. New York: Wiley, pp. 535-552.

Armstrong, J. Scott,(1982), "The value of formal planning for strategic decisions: Review of empirical research," *Strategic Management Journal*, vol. 3, pp. 197-211.

A review of research from organizational behavior supported the guidelines suggested for formal corporate planning: that is, use an explicit approach for setting objectives, generating strategies, evaluating strategies, monitoring results, and obtaining commitment. A review was made of all published field research on the evaluation of formal planning. Formal planning was superior in 10 of the 15 comparisons drawn from 12 studies, while informal planning was superior in only two comparisons. Although this research did not provide sufficient information on the use of various aspects of the planning process, mild support was provided for having participation by stakeholders. Formal planning tended to be more useful where large changes were involved, but, beyond that, little information was available to suggest when formal planning is most valuable.

Armstrong, J. Scott,(1982), "Research on scientific journals: Implications for editors and authors," *Journal of Forecasting*, vol. 1, pp. 83-104.

I reviewed the empirical research on the communication of research findings. From this, I developed guidelines for journals. Many of these guidelines were adopted by the *International Journal of Forecasting*,

Armstrong, J. Scott,(1983), "Strategic planning and forecasting fundamentals," in Kenneth Albert (Ed.), *The Strategic Management Handbook*, New York: McGraw Hill, pp. 2-1 to 2-32.

This paper presents viewpoints on planning, shows how forecasting relates to planning and presents checklists for practitioners.

Armstrong, J. Scott, (1983), "Relative accuracy of judgmental and extrapolative methods in forecasting annual earnings," *Journal of Forecasting*, vol. 2, pp. 437-447.

Analyzes previously published studies on annual earnings forecasts. Comparisons of forecasts produced by management, analysts and extrapolative techniques indicated that: (1) management forecasts were superior to professional analyst forecasts (the mean absolute percentage errors were 15.9 and 17.7, respectively, based on five studies using data from 1967-1974), and (2) judgmental forecasts (both management and analysts) were superior to extrapolation forecasts on 14 of 17 comparisons from 13 studies using data from 1964-1979 (the mean absolute percentage errors were 21.0 and 28.4 for judgment and extrapolation, respectively).

Armstrong, J. Scott,(1983), "The importance of objectivity and falsification in management science," *Journal of Management*, vol. 9, pp. 203-216.

Written in response to Boal and Willis (1983). Ian Mitroff has a reply in the same issue.

Armstrong, J. Scott, (1983), "Cheating in management science" (with commentary), *Interfaces*, vol. 18(August), pp. 20-29.

If you observe what you believe to be cheating, my advice is that you do not call it "cheating." Try to replicate the findings and report your results only as a "failure to replicate."

Armstrong, J. Scott,(1984), "Forecasting by extrapolation: Conclusions from 25 years of research," *Interfaces*, vol. 14, pp. 52-61.

Provides a discussion on the results presented in Appendix J. Commentary on this paper by Robert U. Ayres, Carl Christ, and J. Keith Ord follows on pages, 61-66 of the same issue.

Armstrong, J. Scott,(1984), "Do judgmental researchers use their own research? A review of *Judgment Under Uncertainty: Heuristics and Biases*," *Journal of Forecasting*, vol. 3, pp. 236-239.

A review of Kahneman, Slovic and Tversky (1982).

Armstrong, J. Scott (1985), "Research on forecasting: A quarter-century review, 1960-1984" (with commentary), *Interfaces*, 16, (January/February), 89-109.

This paper summarizes my opinions on the major advances in forecasting. Substantial progress has been made over this 25 year period.

Armstrong, J. Scott and Lusk, Edward J., (1988), "Research on the accuracy of alternative extrapolation models: Analysis of a forecasting competition through open peer review," *Journal of*

Forecasting, vol. 2, pp. 259-262, with seven commentaries followed by replies by each of the original six authors on pages 268-311.

This set of papers discusses what can be learned from the M-Competition (Makridakis, et al., 1982) and what research should be done in the future.

Aschenbrenner, K. Michael and Kasubek, W. (1978), "Challenging the Cushing Syndrome: Multiattribute evaluation of cortisone drugs," *Organizational Behavior and Human Performance*, vol. 22, pp. 216-234.

Decomposed estimates of dangerousness for seven cortisone drugs were obtained from five physicians. The overall ratings based on separate ratings of six side effects led to substantial agreement among the physicians. In contrast, the global ratings led to much disagreement among the physicians.

Ascher, William (1978), *Forecasting: An Appraisal for Policy Makers and Planners*. Baltimore, MD: Johns Hopkins University Press.

Ascher looks at forecasting in population, economics, energy, transportation, and technology. He asks, for example, whether forecasting is getting more accurate over time? (In most areas his answer seemed to be "No.") He also assessed whether forecast accuracy differs by method or by source; I found it difficult to draw conclusions about these issues from the information presented in the book.

Ascher, William and Overholt, William H. (1983), *Strategic Planning and Forecasting*. New York: Wiley.

This book, which focuses on political forecasting, discusses some worthwhile topics such as how to present forecasts, the relationships between forecasting and planning, how to organize the forecasting and planning functions, and how to choose a forecasting method. For the most part, these sections draw primarily upon the authors' experience and the prevailing opinions of experts, rather than upon empirical evidence.

Assmus, Gert (1984), "New product forecasting," *Journal of Forecasting*, vol. 3, pp. 121-138.

A state of the art review of models for new product forecasting. It describes attributes and costs of some of the more popular commercial models.

Auerbach, Alan J. (1982), "The index of leading indicators: Measurement without theory, thirty-five years later," *Review of Economics and Statistics*, vol. 64, pp. 589-595.

This study reaches some favorable conclusions about the use of leading indicators in forecasting. Equal weighting of the 12 leading indicators did better than regression weights.

Avison, William R. and Nettler, Gwynn, (1976), "World views and crystal balls," *Futures*, vol. 8 (February), pp. 11-21.

Babich, George and Goodhew, John (1978), "Short-term econometric forecasting and seasonal adjustment," *Economic Record*, vol. 54, pp. 229-236.

This paper compares forecasts using deseasonalized data with those from a model that used dummy variables to estimate seasonality on one and two-period ahead forecasts for

14 variables over eight periods. Little difference was found, though the deseasonalized approach tended to be more accurate.

Baker, Earl Jay (1979), "Predicting response to hurricane warnings: A re-analysis of data from four studies," *Mass Emergencies*, vol. 4, pp. 9-24.

This study reviews survey research data from three serious hurricanes in the United States and produces some interesting findings: (1) if people believed the hurricane forecasts, they were more likely to evacuate, but this relationship was weak; also, the relationship held up only for very short-range forecasts (less than three hours before landfall); (2) attention devoted by a person to monitoring the hurricane forecasts was unrelated to whether that person evacuates or not; (3) public education about the dangers and proper responses to hurricane forecasts was not related to evacuation behavior. In general, the studies failed to identify strong predictors of evacuation behavior. Encouragingly, the strongest predictor was the degree of risk to which respondents were exposed.

Baker, Earl J.; West, S.G., Moss, D.J. and Weyant, J. M. (1980), "Impact of offshore nuclear power plants: Forecasting visits to nearby beaches," *Environment and Behavior*, vol. 12, pp. 867-407.

To forecast the effect of offshore nuclear plants upon the visits to beaches, the authors spread their budget among many methods including: prior research, studies of analogous situations (beaches near land-based nuclear plants), surveys of experts, intentions surveys, and attitude surveys. Each approach suffered from serious problems, yet when considered as a group they provided a convincing picture. Whereas the intentions survey indicated that about one-quarter of the tourists would avoid beaches with offshore nuclear plants, the other methods suggested one-quarter was a substantial overestimate. The authors concluded that 5 to 10 percent is a reasonable estimate. Floating nuclear plants were not so important as finding a clean and uncrowded beach with nice facilities. (This is one of the few cases where I think it is fortunate that data do not exist to test predictive ability.)

Baker, H. Kent and Tralins, Stanley M. (1976), "An analysis of published financial forecasts," *Atlanta Economic Review*, vol. 26, pp. 42-46.

Bar-Hillel, Maya and Fischhoff, Baruch (1981), "When do base rates affect predictions?" *Journal of Personality and Social Psychology*, vol. 41, pp. 671-680.

A reply to Manis, *et al.*(1980).

Barrett, Gerald V., Phillips, James S. and Alexander, Ralph A. (1981), "Concurrent and predictive validity designs: A critical reanalysis," *Journal of Applied Psychology*, vol. 66, pp. 1-6.

This paper argues that for, personnel predictions, the superiority of tests of predictive validity (outside sample time period), has been overestimated relative to concurrent validity. See also Guion and Cranny (1982).

Bass, Bernard M. (1977), "Utility of managerial self-planning on a simulated production task with replications in 12 countries," *Journal of Applied Psychology*, vol. 62, pp. 506-509.

This is an important study of planning involving experiments with 1416 managers from 12 countries. Groups that developed their own plans were more effective than those that were presented with plans. As groups gained experience with self-planning, their efficiency improved still more. Modest nationality differences were found, Americans gaining most from self-planning, Germans the least.

Basu, Shankar and Schroeder, Roger G. (1977), "Incorporating judgments in sales forecasts: Application of the Delphi method at American Hoist and Derrick," *Interfaces*, vol. 7(May), pp. 18-27.

This study describes how the authors used 23 experts in a three-round Delphi study covering three months. The authors claim that this procedure provided accurate sales forecasts for one and two year horizons.

Becker, Lawrence J. (1978), "Joint effect of feedback and goal setting on performance: A field study of residential energy conservation," *Journal of Applied Psychology*, vol. 68, pp. 428-433. Significant gains were achieved when families set high goals for conservation *and* when they also received feedback about their performance in relation to those goals. Those that either did not set high goals or did not receive feedback did not change their use of energy significantly.

Beyth-Marom, Ruth (1982), "How probable is probable? A numerical translation of verbal probability expressions," *Journal of Forecasting*, vol. 1, pp. 257-269.

This experiment took place in a professional forecasting organization accustomed to giving verbal probability assessments ("likely," "probable," etc.). It highlights the communication problems caused by verbal probability expressions. Experts in the organization were first asked to give a numerical translation to 30 different verbal probability expressions, most of which were taken from the organization's own published political forecasts. In a second part of the experiment, the experts were given 15 paragraphs selected from the organization's political publications, each of which contained at least one verbal expression of probability. Subjects were again asked to give a numerical translation to each verbal probability expression. The results indicated that (1) there was a high variability in the interpretation of verbal probability expressions, and (2) the variability is even higher in the problem context.

Beyth-Marom, Ruth and Arkes, Hal R. (1988), "Being accurate but not necessarily Bayesian: Comments on Christensen-Szalanski and Beach," *Organizational Behavior and Human Performance*, vol. 31, pp. 255-257.

See Christensen-Szalanski and Beach (1982).

Binroth, W., Burshstein, I., Haboush, R.K. and Hartz, J.R., (1979), "A comparison of commodity price forecasting by Box-Jenkins and regression-based techniques," *Technological Forecasting and Social Change*, vol. 14, pp. 169-180.

This paper analyzed three months-ahead *ex ante* forecasts of rubber commodity prices. It used 72 months of data to develop the models, then forecasted over 26 months. Box-

Jenkins was slightly more accurate than the econometric model, but the difference did not appear to be significant.

Boje, David M. and Murnighan, J. Keith, (1982), "Group confidence pressures in iterative decisions," *Management Science*, vol. 28, pp. 1187-1196.

Individual estimates were compared with ones from group face-to-face interaction and from groups with only written feedback. The experiment involved four estimation problems (two almanac questions and two on heights and weights of people). Participants were 324 undergraduates in group sizes of 3, 7, and 11. The confidence of group members went up over the three rounds in the Delphi-like procedure, which is a typical result. These gains in confidence were unrelated to accuracy, also not surprising. However, no gain was found in the accuracy of the later rounds, which is mildly surprising in light of previous research where small gains were found. Group members preferred the face-to-face interaction and thought it the most effective; however, it was the *least* accurate, a finding that agrees with previous research. I enjoy studies with surprising results. Imagine, then, how pleased I was to find this conclusion in their study... "group size had no significant effects on accuracy..." This conclusion conflicts with prior research.

Bopp, Anthony E. and Durst, Mitchell (1978), "Estimated importance of seasonal adjustment on energy forecasts," *Atlantic Economic Journal*, vol. 6, pp. 53-59.

Bopp, Anthony E. and Neri, John A. (1978), "The price of gasoline: Forecasting comparisons," *Quarterly Review of Economics and Business*, vol. 18 (Winter), pp. 23-34.

This study makes *ex ante* forecasts for the 18-month period from January 1975 to June 1976 to compare Box-Jenkins, simple econometric, and simultaneous equations models. For one-month ahead, Box-Jenkins tended to be most accurate, but as forecast horizon lengthened, it became least accurate. However, there were too few comparisons to draw statistically significant conclusions.

Borgida, Eugene (1978), "Scientific deduction-evidence is not necessarily informative: A reply to Wells and Harvey," *Journal of Personality and Social Psychology*, vol. 36, pp. 477-482.

This paper extends Nisbett and Borgida (1975).

Borgida, Eugene and Nisbett, Richard E. (1977), "The differential impact of abstract vs. concrete information on decisions," *Journal of Applied Social Psychology*, vol. 7, pp. 258-271.

Borman, Walter C. (1982), "Validity of behavioral assessment for predicting military recruiter performance," *Journal of Applied Psychology*, vol. 67, pp. 3-9.

Sixteen experienced recruiters assessed 57 soldiers entering the Army's recruiter school. Their assessment ratings were compared with subsequent performance in short training episodes. First impressions, ratings based on a structured interview, and scores on a paper and pencil test of personality and vocational interests, each correlated near zero with the training performance. But an assessment program using role playing, in-basket and the preparation of a short recruiting speech correlated highly with the criteria. Statistical com-

posites of the assessment ratings were less expensive and slightly more valid than clinical judgments based on consensus among the assessors.

Branbon, Charles, Fritz, Richard and Xander, James (1988), "Econometric forecasts: Evaluation and revision," *Applied Economics*, vol. 15, pp. 187-201.

Brandon, Charles H., Jarrett, Jeffrey K. and Khumawala, Saleha (1983), "Revising forecasts of accounting earnings: A comparison with the Box-Jenkins method," *Management Science*, vol. 29, pp. 256-263.

Brandt, Jon A. and Bessler, David A. (1988), "Price forecasting and evaluation: An application in agriculture," *Journal of Forecasting*, vol. 2, pp. 237-248.

Seven forecasting methods were used to make one-quarter ahead *ex ante* forecasts of U.S. hog prices for 24 quarters from 1976 to 1981. The errors (MAPES) were: ARIMA (7.96), expert judgment (8.61), econometric (9.98), "no change" (10.07), simple exponential smoothing (10.16), and Holt-Winters (10.28). A combination of ARIMA, econometric and expert judgment was best (7.27). Interestingly, expert forecasts alone did worse than a strategy of never hedging (always selling for cash in the market). The differences in accuracy in this study do not appear to be statistically significant.

Braun, Michael A. and Srinevasan, V. (1975), "Amount of information as a determinant of consumer behavior towards new products," Proceedings. Chicago: American Marketing Association, pp. 373-378.

Study of intentions to purchase a new product, the Gillette TRAC II twin blade razor.

Brown, Lawrence D. and Rozeff, Michael S. (1978), "The superiority of analyst forecasts as measures of expectations: Evidence from earnings," *Journal of Finance*, vol. 33, pp.1-16.

Buhmeyer, Kenneth J. and Johnson, Alan H. (1978), "Predicting success in a physician-extender training program," *Psychological Reports*, vol. 42, pp. 507-513 .

I hereby announce this paper as the winner of the 1978 "Tom Swift Award for Data Abuse."

Bunn, Berek W. (1979), "The synthesis of predictive models in marketing research," *Journal of Marketing Research*, vol. 16, pp. 280-28.

This paper used two small examples to illustrate the value of combining forecasts.

Bunn, Derek W. and Seigal, Jeremy P. (1983), "Forecasting the effects of television programming upon electricity loads," *Journal of the Operational Research Society*, vol. 34, pp. 17-25.

Small validation sample used to compare econometric and subjective forecasts of peak electricity loads. Mixed results were obtained, so no firm conclusions could be drawn.

Burns, Michael and Pearl, Judea (1981), "Causal and diagnostic inferences: A comparison of validity," *Organizational Behavior and Human Performance*, vol. 28, pp. 379-394.

The results of this experiment were surprising; better predictions did not result when causal reasoning was used to decompose problems for a group of subjects. Non-causal approaches worked just as well.

Burrows, Paul (1971), "Explanatory and forecasting models of inventory investment in Britain," *Applied Economics*, vol. 3, pp. 275-289.

Examined forecasts for seven quarters from 1967II to 1968IV. For *ex ante* forecasts, an econometric model did better than an extrapolation model. But for *ex post* forecasts, the extrapolation did better than the econometric model.

Camerer, Colin (1981), "General conditions for the success of boot-strapping models," *Organizational Behavior and Human Performance*, vol. 27, pp. 411-422.

Camerer examined theoretical arguments and empirical evidence and concluded that boot-strapping models make better predictions than experts in nearly all practical situations in which data on the criteria are missing or vague.

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

Students with limited training were able to obtain as accurate forecasts when using Box-Jenkins methods as were experts. Subjective adjustments did not improve the accuracy of the extrapolations.

Carbone, Robert and Armstrong, J. Scott (1982), "Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners," *Journal of Forecasting*, vol. 1, pp. 215-217.

Carbone, Robert and Gorr, Wilpen L. (1985), "Accuracy of judgmental forecasting of time series." *Decisions Sciences*, 16, 153-160.

Carky, Kenneth J. (1978), "The accuracy of estimates of earnings from naive models," *Journal of Economics and Business*, vol. 30, No. 3, pp. 182-193.

Cattin, Philippe (1980), "Estimation of the predictive power of a regression model," *Journal of Applied Psychology*, vol. 65, pp. 407-414 .

Cattin compared standard regression estimates of predictive power with measures obtained from cross-validation. He used theoretical arguments and simulation with two criteria (mean square error and R^2). The formulas, which are less expensive, were often adequate.

Cattin, Philippe and Wittink, Dick R. (1982), "Commercial use of conjoint analysis: A survey," *Journal of Marketing*, vol. 46 (Summer), pp. 44-53.

This paper reports on a survey of commercial uses of conjoint analysis in determining customer preferences for products. The first commercial application was in 1971. Since then, usage has grown dramatically. The most important application of conjoint analysis is to

predict preferences for new products. The authors' survey goes through the various steps in conjoint analysis to determine which techniques are used most often. For example, to develop attributes of a product, some projects used the direct opinions of management while protocols were less popular. To obtain data, the most common approach was to ask customers to choose among products that were described in terms of all key attributes rather than to rely on comparisons of two factors at a time. Most commonly, the question was cast in terms of "intention to buy" rather than preference. The analysis of the data was typically based on some form of regression analysis. Cattin and Wittink encourage research firms to share their experiences with the research community; to date, few published studies have tested the predictive validity of conjoint analysis.

Cerf, Christopher and Navasky, Victor (1984), *The Experts Speak*. New York: Pantheon Books. A collection of inaccurate statements and forecasts by experts.

Cerullo, Michael J. and Avila, Alfonso (1975), "Sales forecasting practices: A survey," *Managerial Planning*, vol. 24 (Sept.-Oct), pp. 33-39.

This survey of 110 of the *Fortune* 500 companies yielded replies from 56 companies. Judgmental methods proved to be most popular, as 89% of the companies reported their use. In comparison, 52% said they used extrapolation, 30% econometric, 24% regression, and 20% input-output. The most common approach was to ask each member of the sales force to make a forecast and then to have a group of executives adjust this. The *perceived* accuracy of the forecast was about the same for those who used causal methods as for those who used naive methods. Most firms (77%) said they did not know how much they spend on forecasting. Few firms (4%) used outside consultants. Few firms (9%) forecasted sales beyond one year. (That seems surprising, doesn't it?) About 57% used computers in the forecasting process. Finally, 98% thought that forecasting with causal methods should be taught at business schools..

Chatfield, Christopher (1978), "The Holt-Winters forecasting procedure," *Applied Statistics*, vol. 27, pp. 264-279.

Christensen-Szalanski, Jay J. J. and Beach, Lee Roy (1982), "Experience and the base-rate fallacy," *Organizational Behavior and Human Performance*, vol. 29, pp. 270-278 and Christensen-Szalanski, Jay, J. J. and Beach, Lee Roy (1988), "Believing is not the same as testing," *Organizational Behavior and Human Performance*, vol. 31, pp. 258-261.

In the Christensen-Szalanski and Beach (1982) experiment, decision makers who *experienced* the relationship between the base rate (i.e., the frequency with which an event occurred in a series of trials) and diagnostic information used this relationship when they made judgments. However, when given the necessary *theoretical* information, they did not use the base rate effectively. (In other words, people may not use Bayes rule; but with experience, they can come close to the Bayesian solution.) Beyth-Marom and Arkes (1988) challenge this interpretation. Rather than use Bayes theorem, they suggest that the subjects made direct estimates of the proportions. Christensen-Szalanski and Beach (1988), however, say that this is compatible with their interpretation.

Clarke, Mike, Dix, Martin and Goodwin, Phil (1982), "Some; issues of dynamics in forecasting travel behavior: A discussion paper," *Transportation*, vol. 11 (June), pp. 153-172.

Cocozza, Joseph J. and Steadman, Henry J. (1978), "Prediction in psychiatry: An example of misplaced confidence in experts," *Social Problems*, vol. 25, pp. 265-276.

In this study of 256 defendants, psychiatrists were asked to predict which defendants might be violent. The psychiatrists seemed unaware of how they made their ratings. For example, only 11.5% said their rating was related to the violence of the crime of which the defendant was charged. Yet 73% of those charged with a violent crime were rated as dangerous, a much higher figure than for cases where the crime was not violent. A three-year follow-up indicated that there was no difference in violence between those predicted to be dangerous and those predicted not to be dangerous. This agreed with some prior research. Still, courts use psychiatrists, and in 87% of these cases they followed the psychiatrists' recommendations.

Coggin, T. Daniel and Hunter, John E. (1982-88), "Analysts' EPS forecasts nearer actual than statistical models," *Journal of Business Forecasting*, vol. 1 (Winter), pp. 20-23.

This paper reports on a study of one-year-and two-year-ahead forecasts of annual earnings per share for 149 companies in 1978 and 1979. In addition, one-year ahead forecasts were made for another sample of 180 companies for 1979. The accuracy of combined judgmental forecasts by analysts (at least three analysts, but typically about 12 analysts) was compared with the accuracy of three extrapolation models using the mean square percentage error as the criterion (which I converted to root mean square error here to aid in understanding). *Conclusions:* (1) the judgmental forecasts were significantly better than the best of the three extrapolations (RMSE of 33% vs. 37.2% respectively for the one-year-ahead EPS forecasts); (2) even the typical analyst was significantly better than the best of the three extrapolations (33.8 % vs. 37.2 % respectively); (3) the simpler the extrapolation method, the more accurate the forecast-especially for the two-year horizon; and (4) the forecast error increased substantially for the two-year-ahead forecast (an increase of 27 % for the root mean square percentage error for the best extrapolation method).

Collins, Daniel W. (1976), "Predicting earnings with sub-entity data: Some further evidence," *Journal of Accounting Research*, vol. 14, pp. 163-177.

This study examined data on 96 firms for 1968, 1969, and 1970 with two econometric models, five extrapolation models, and two segmented econometric models. The segmented econometric models was more accurate for both sales and profit forecasts. Silhan (1983) points out that Collins' tests were flawed because different data sets were used for the segmented models as compared to the other models.

Cosier, Richard A. (1978), "The effects of three potential aids for making strategic decisions on prediction accuracy," *Organizational Behavior and Human Performance*, vol. 22, pp. 295-306.

Here is an interesting experiment showing that the Devil's Advocate approach led to better predictions than an approach using an expert who argued in favor of a plan.

Cox Eli P. (1980), "The optimal number of response alternatives for a scale: A review," *Journal of Marketing Research*, vol. 17, pp. 407-422.

Cummins, J. David and Griepentrog, Gary L. (1985), "Forecasting automobile insurance paid claim costs using econometric and ARIMA models," *International Journal of Forecasting*, (in press) .

Currim, Imran S. (1981), "Using segmentation approaches for better prediction and understanding from consumer mode choice models," *Journal of Marketing Research*, vol. 18, pp. 301-309.

The basic proposition of this paper is that segmentation of consumers should allow one to make better predictions because different groups behave differently. That is what we call "common sense" in marketing. Sometimes, of course, our common sense is wrong; hence, the present study seemed like a worthwhile undertaking. The proposition was tested on the prediction of transportation mode choice (e.g., auto or bus) between two geographical points. But it was not the actual choice, merely the mode the consumers *say* they would take if they happened to make that hypothetical trip. The two segmentation schemes, one using 10 "benefit segments" and the other one with nine "situational segments," did *not* yield more accurate predictions of overall market shares for five possible mode choices for a hold-out sample of about 170 subjects. The average error for the two segmented models was identical to that of the aggregate model. These results were surprising and disappointing.

Dalrymple, Douglas J. (1975), "Sales forecasting: Methods and accuracy," *Business Horizons*, vol. 18, pp. 69-78.

This paper reports on a survey of 500 firms with 175 replies. One finding: Systematic records of forecast accuracy were kept by 61% of the firms that replied.

Dalrymple, Douglas J. (1978), "Using Box-Jenkins in sales forecasting," *Journal of Business Research*, vol. 6, pp. 138-145.

Dalrymple, Douglas J. (1985), "Sales forecasting practices in businesses: Results from a 1983 U.S. Survey," Working paper, Graduate School of Business, Indiana University, Bloomington, Indiana 47401.

Responses were received from 134 business firms, a 16% return from a survey of 850 firms in the United States. About 60% of the respondents were manufacturers, 23% were in distribution, and 14% in retailing. This is the first study I have found that has assessed the use of *combined forecasts* in business:

- 20 % do it "usually"
- 19 % "frequently"
- 29 % "occasionally"
- 32 % did not use this strategy.

Also of interest were the results on the use of *upper and lower confidence intervals* when presenting forecasts. They are not widely used:

- not used 48 %
- occasional 29 %

frequent 11 %
usual 10 %

The survey contains many useful findings on the practice of forecasting.

Dalrymple, Douglas J. and King, Barry K. (1981), "Selecting parameters for short-term forecasting techniques," *Decision Sciences*, vol. 12, pp. 661-669.

Parameters for extrapolation models are generally selected to reduce the error for a one-period-ahead forecast horizon. Often, however, the forecasts are made for horizons beyond one period. This study asks whether it would be worthwhile to select parameters for the specific forecast horizon. Good question. This is apparently the first study on this issue. The authors examined data for 25 business time series ("mostly monthly" they say). Using cumulative MAPEs for forecast horizons from 1 to 12 periods ahead, their search for optimum parameters led to *no gain* when using either exponential smoothing or trend regression. Dalrymple and King found some benefit for this procedure when using moving averages, but I did not draw the same conclusion from their data. Surprisingly, one-period-ahead searches seem adequate for n-period ahead forecasts. The paper also presents evidence showing an increase in error as the forecast horizon increases. While it did help to use more historical data for the parameter search, their conclusion was unintentionally overstated by a misprint (their p. 668), where they say eight periods of data were optimal for trend regression for a one-period-ahead forecast vs. 27 for a 12-period-ahead forecast. (It should have been 18 periods not eight, for the one-period horizon).

Danos, Paul and Imhoff, Eugene A. (1988), "Factors affecting auditors' evaluations of forecasts," *Journal of Accounting Research*, vol. 21, pp. 473-494 .

An experiment consisting of four realistic corporate cases was presented to 81 auditors from the Big Eight accounting firms. The results suggested that auditors tended to have more confidence in forecasting systems that: (1) had centralized financial planning systems; (2) rewarded the managers for accurate forecasts; and, (3) did not make large revisions from the initial to the final forecast. However, the auditors' confidence in the forecasts was significantly increased in cases where the forecasters had a good track record in predicting income statement data.

Daub, Mervin (1981), "The accuracy of Canadian short-term economic forecasts revisited," *Canadian Journal of Economics*, vol. 14, pp. 499-507.

Daub compared errors in predicting annual changes in Canadian GNP in the 1970s with those from 1957-69. Forecast errors in the 1970s were *smaller*.

Daub, Mervin and Peterson, E. (1981), "The accuracy of a long-term forecast: Canadian energy requirements," *Energy Research*, vol. 5, pp. 141-154.

Daub and Peterson analyze the accuracy of a 10-year energy forecast made in Canada in 1966. As we now know, the early 1970s were a period of high turbulence due, first, to environmental concerns, and then to the OPEC petroleum crisis; as a result, it became much more difficult to forecast in general, and especially to forecast energy. According to Daub and Peterson's study, however, the preceding statement is false. The Canadian National Energy Board's forecasts, based on elaborate data to supplement judgmental proce-

dures, did not deteriorate over time. Surprisingly, the error did not grow over the forecast horizon as one would expect in times of turbulence. This finding corresponds to that reported in forecasts of 1,001 time series in Makridakis, et al. (1981) and to Daub (1981). It was also interesting that simple extrapolations based on the previous 5 to 10 years did better than forecasts by the five-member Canadian Board.

Dawes, Robyn M. (1979), "The robust beauty of improper linear models in decision making," *American Psychologist*, vol. 34, pp. 571-582.

This paper is a follow-up on Dawes (1974). Dawes was aware of only four universities (U. of Illinois, NYU, U. of Oregon, and UC Santa Barbara) that adopted bootstrapping and, even in these places, it was used only for initial screening. However, large state universities with the need to allocate spaces in a politically acceptable manner have been moving in the direction of using linear models.

Dawes, Robyn M. (1980), "Apologia for using what works," *American Psychologist*, vol. 85, p. 678.

This critiques Pritchard (1980).

Dennis, John D. (1978), "A performance test of a run-based adaptive exponential forecasting technique," *Production and Inventory Management*, vol. 19, pp. 43-46.

The conclusions of this study were challenged by Ekern (1981).

Dielman, Terry E. (1985), "Regression forecasts when disturbances are autocorrelated," *International Journal of Forecasting*, vol. 1.

This study uses a Monte Carlo simulation to assess the quality of forecasts obtained from regression models with various degrees of autocorrelation in the error term. Deilman concludes that it is important to correct for autocorrelation, especially for very short-range forecasts.

Dillman, Don A. (1978), *Mail and Telephone Surveys*. New York: Wiley. A useful set of guidelines for conducting mail and telephone surveys. Reviewed by Armstrong [1981].

Dipboye, Robert L. (1982), "Self-fulfilling prophecies in the selection-recruitment interview," *Academy of Management Review*, vol. 7, pp. 579-586.

This review paper concludes that interviewers are strongly influenced by prior information.

Dorans, Neil and Drasgow, Fritz (1978), "Alternative weighting schemes for linear prediction," *Organizational Behavior and Human Performance*, vol. 21, pp. 316-345.

They examined six approaches to estimating relationships, and tested them on simulated data with three variables. Equal weights performed well across different sample sizes. Regression (OLS) was poorest for small samples in the cross-validation ($n < 30$), but best for large samples. Equal weights are appropriate when (1) sample size is small or moderate; (2) good *a priori* information exists on the direction or the relationship, and (3) positive (not negative) intercorrelations exist among predictors.

Downs, Sylvia, Farr, Robert M. and Colbeck, L. (1978), "Self-appraisal: A convergence of selection and guidance," *Journal of Occupational Psychology*, vol. 51, pp. 271-278.

People seem to be able to make good predictions about how successful they will be on a job if they have been given a realistic preview.

Duan, Naihua, Manning, W.G., Morris, C. N. and Newhouse, J. P. (1983), "A comparison of alternative models for the demand for medical care," *Journal of Business and Economic Statistics*, vol. 1, pp. 115-126.

Improved forecasts were obtained by segmenting into two or into four subproblems. The overall problem involved forecasting the effects of different insurance plans. Segmentation seemed especially appropriate because there were such large differences in forecasted behavior for those who had consumed health services and those who had not. The paper is difficult to read as exemplified by this sentence (p. 120): "Moreover, when the normal assumption indeed holds, the nonparametric smearing estimate has high efficiency relative to the parametric normal transformation factor $\exp(a/2)$ for a wide range of parameter values."

Duda, Richard O. and Shortliffe, Edward H. (1983), "Expert systems research," *Science*, vol. 220, pp. 261-268.

This paper reviews research on expert systems: (computer systems designed to make diagnoses that rival those of experts). The authors view expert systems as a subset of "knowledge based systems" which, in turn, is a subset of "artificial intelligence." They briefly describe applications of expert systems for medicine, geology, computer design, and chemistry. These programs seemed to perform well in comparison with experts. It appears that the ability of the program to tell the user *why* various questions are being asked is thought to be important for acceptance by the potential user.

Ebert, Ronald J. and Kruse, Thomas E. (1978), "Bootstrapping the security analyst," *Journal of Applied Psychology*, vol. 63, pp. 110-119.

Eggleton, Ian R.C. (1982), "Intuitive time-series extrapolation," *Journal of Accounting Research*, vol. 20, pp. 68-102.

This paper reports on a laboratory experiment. Subjects were presented with 12 sets of two-digit numbers which were referred to as "monthly production costs." After viewing the series for 15 seconds, each subject forecasted the next observation and provided a confidence interval. Findings: (1) subjects were conservative relative to extrapolation models. That is, they predicted smaller changes than the commonly used extrapolation methods; and (2) the confidence intervals were sensitive to the historical variance.

Ehrenberg, Andrew S.C. (1981), "The problem of numeracy," *American Statistician*, vol. 35 (May), pp. 67-70.

Presents six rules for improving the presentation of data in tables: (1) round to two significant digits, (2) provide row or column averages, (3) arrange the numbers to be compared in a column rather than a row, (4) order the rows and columns by size, (5) use layout to guide the eye and facilitate comparisons, and (6) give verbal summaries about major pat-

terns and exceptions. The first five rules are easy to implement on a personal computer with a spreadsheet program.

Einhorn, Hillel J. and Hogarth, Robin M., (1982), "Prediction diagnosis and causal thinking in forecasting," *Journal of Forecasting*, vol. 1, pp. 23-36.

This paper discusses theory and prior research.

Ekern, Steinar, (1981), "Adaptive exponential smoothing revisited," *Journal of the Operational Research Society*, vol. 82, pp. 775-782.

Replicates and challenges the studies by Dennis (1978) and Whybark (1972).

Elliott, J. Walter and Baer, Jerome R., (1979), "Econometric models and current interest rates: How well do they predict future rates?" *Journal of Finance*, vol. 34, pp. 975-986.

Elliott, J. W. and Uphoff, H. L., (1972), "Predicting the near term profit and loss statement with an econometric model: A feasibility study," *Journal of Accounting Research*, vol. 10, pp. 259-274.

An econometric model provided more accurate *ex post* forecasts of nine variables for one company than did three extrapolation models. However, these were based upon only four monthly forecasts.

Elstkin, Arthur S., Shulman, A.S. and Spafka, S.A., (1978), *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, Mass.: Harvard University Press.

Simple "low fidelity" role-playing ("simulations") provided similar results to those from the more realistic ("high fidelity") and more expensive role-playing.

Falconer, Robert T. and Sivesind, C.M., (1977), "Dealing with conflicting forecasts: The eclectic advantage," *Business Economics*, vol. 12, No. 4, pp. 5-11.

This paper examined combined *ex post* econometric and naive forecasts of U.S. personal income for forecasts up to a six quarter horizon. "Composite" (combined) forecasts, weighted by the relative Root Mean Square Errors, helped in all cases, but especially for the longer forecast horizon.

Field, Hubert S. and Holley, William H., (1982), "The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases," *Academy of Management Journal*, vol. 25, pp. 392-406.

Assume that a worker claims to have received unfair discrimination in termination or promotion, and files suit against your organization. Your organization's action had been based on a performance appraisal. How could you predict your chances of success? This paper reviews relevant research in the field to suggest which factors are important (e.g., an organization that uses specific written instructions will be more successful in defending itself: also, organizations are more successful when they rely on evaluations of *behavior* rather than *personal traits*). The paper also examines 66 new cases in an effort to develop better predictors. Unfortunately, this aspect of the paper did not involve any predictions; it focused on an explanation of historical results. The ability to explain was modest; a linear

discriminant function explained only 39% of the variability in the verdicts. Some surprises here: (1) firms that presented evidence on the validity and reliability of their performance appraisal systems received *no* better treatment from the courts, (2) industrial organizations fared *less well* than nonindustrial organizations (such as universities). The study implies that understanding the prejudices of the court will allow one to make better predictions about the outcome. In addition, the results imply actions that organizations can take to reduce the likelihood of losing verdicts.

Fildes, Robert (with Dews, D. and Howell, S.), (1981), *A Bibliography of Business and Economic Forecasting*. Farnborough, Hants, England: Gower Publishing. (A revised edition was published in 1984 by the Manchester Business School under same title with the subtitle *Part 2, 1979-1981*.) My summary draws upon the review by Everette S. Gardner, *Journal of Forecasting*, 1 (1982), (320-321).

This is a comprehensive reference source on forecasting methods. Many of the studies in Fildes bibliography are only indirectly related to forecasting. More than 4,000 items are indexed in the 1981 edition, mostly articles from 40 journals over the period 1971-1978. Fifteen economics journals were searched. Other areas include: statistics, management science, and marketing-six journals from each field; general business-four; accounting-two; and finance-one. Some articles prior to 1971 were included, along with a selection of books. The bibliography is directed largely toward economic model building. The key words are so extensive that they compose a mini-abstract for most references listed. Here is an illustration under "judgmental forecasting:"

YA2657 O'Carroll, F.M.

"Subjective probabilities and short-term economic forecasts: an empirical investigation," *Appl. Stats.*, vol. 26, 1977, pp. 269-278. APPL-MACRO: INTERNATIONAL FINANCE, EXCHANGE RATES* APPL-MACRO, UK*APPL-SECTOR: FINANCL, STOCK PRICE INDEX*: APPL-SECTOR: PRODUCTION, PETROLEUM* PRICE-SECTOR: PRODUCTION*JUDGMENTAL FORECASTS-SPECIALIST*ERROR DISTRIBUTION-LOGNORMAL* EVALUATION-JUDGMENTAL FORECASTS*'JUDGMENTAL FORECASTS-UNCERTAINTY

This example shows that macro applications may be located using key words for sector, variable, or country. The key words also indicate that the judgmental forecasts were made by a specialist in the field, using probabilities (key worded as "uncertainty"), and that the article evaluates the effectiveness of the forecasts. A statistical problem related to the lognormal error distribution is also discussed. The key words are based on 14 dimensions or categories of knowledge. These include: (1) applications, (2) variables to forecast, (3) types of models, (4) model interpretation, (5) model estimation, (6) statistical problems, (7) uses and users, (8) forecast effectiveness, (9) forecast monitoring and evaluation, (10) how to develop and select a model, (11) data-related problems, (12) the effects of certain independent variables, (13) the theory underlying a model, and (14) implementation problems. More than 500 examples of applications of forecasting are classified by firm and industry, with subcategories by product. Many of these applications are also cross-referenced to implementation problems and how the forecasts were used. The coverage in inventory control, manpower planning, and portfolio selection is particularly thorough. The listings of comparisons among alternative forecasting methods should be valuable. One

can choose any major forecasting method and find references that compare it with other methods. For example, 60 comparisons are listed between ARIMA methods and one or more of the following: autoregressive, causal, decomposition, distributed lag, exponential smoothing, judgmental, and others. Many of these comparisons were not evident from the titles or abstracts of the papers, and this reflects the care that went into this bibliography. Users of this bibliography will be able to spice up their lectures with some of the more exotic references listed. Some classroom examples include papers on forecasting productivity in the Israeli diamond industry (good results), the population of colored foxes in Labrador (also good), and earthquakes (shaky results). Although Fildes does not evaluate the references, some are labeled either basic or advanced, according to mathematical complexity. Basic references can be used by beginners in the field. Articles labeled as advanced have little general value because of their inaccessibility. But many difficult papers on such topics as spectral analysis and statistical testing of simultaneous equations models were not labeled as advanced. The average user will find that papers with advanced labels are incomprehensible. This reference work should help unify the field of forecasting. Because of the extensive literature on forecasting, I believe that it is important to use this book if one expects to do a thorough literature review. The book's problem, however, is similar to that faced by Alice in *Through the Looking Glass*. The White Queen's advice was: "Now, here, you see, it takes all the running you can do, to keep in the same place." Its place now is as the leading source book. The 1979-1981 update expanded the coverage from 40 to 70 journals and added 1500 references to the original 4000.

Fildes, Robert, (1982), "Forecasting: The issues," in S. Makridakis and S.C. Wheelwright (Eds.). *The Handbook of Forecasting*. New York: Wiley.

Fildes, Robert and Fitzgerald, M. Desmond, (1983), "The use of information in balance of payments forecasting," *Economica*, vol. 50, pp. 249-258.

Fildes and Fitzgerald examined the performance of three economists who each month made one-month-ahead forecasts of the U.K. balance of payments. The period from July 1975 to December 1978 was studied. Some findings were consistent with prior evidence—for example, the combined forecasts of the three economists were better than those by the average forecaster (RMSE of 176 vs. 185). Fildes and Fitzgerald also examined an extrapolation (ARIMA) model and found it equal in accuracy to the average judgmental forecaster (RMSE 184 vs. 185). Then they combined the extrapolation and judgmental forecasts and found little improvement over the combined forecast of three judges (RMSE of 173 vs. 176). Finally, they concluded that bootstrapping models did not improve the forecast accuracy of any of the three judges. But the sample size used (three judges) is too small to allow us to conclude that bootstrapping was less accurate.

Fildes Robert and Howell, Syd, (1979), "On selecting a forecasting model," in Spyros Makridakis and Steven C. Wheelwright (eds.), *Forecasting*. New York: North Holland, pp. 297-312.

This literature review, among other things, presents evidence favoring simplicity in econometric models.

Fildes, Robert and Lusk, Edward J., (1984), "The choice of a forecasting model," *Omega*, vol. 12, pp. 427-435.

Fink, Edward B., Braden, W. and Qualls, C.B., (1982), "Predicting pharmacotherapy outcome by subjective response," *Journal of Clinical Psychiatry*, vol. 43, pp. 272-275.

How effective will a drug be for a patient? One way to improve this prediction is ask the patient about the benefit of the drug after 24 hours of use. Patients' responses were related to clinical improvement after 8 to 21 days in this experiment. What seems most surprising is that such predictive measures are not used routinely.

Finkel, Sidney R. and Tuttle, Donald L., (1971), "Determinants of the aggregate profits margins," *Journal of Finance*, vol. 26, pp. 1067-1075.

Forecasts (apparently *ex post*) were made for total corporate profits in the U.S. economy. The forecasts were made for the quarters from 1968-I to 1970-II, using a single starting point. An econometric model that had four independent variables was more accurate than a naive model that was based on a weighted moving average.

Fischer, Gregory W., (1981), "When oracles fail-A comparison of four procedures for aggregating subjective probability forecasts," *Organizational Behavior and Human Performance*, vol. 28, pp. 96-110.

This paper contrasts four methods for obtaining forecasts from a group of experts: (1) statistical average of the individual forecasts, (2) face-to-face discussion to reach consensus, (3) Delphi, and (4) Estimate-Talk-Estimate (E-T-E). Fischer's review of the literature, along with a reanalysis of an important E-T-E study, provided little basis to suggest one method is more accurate than another. Fischer then used the four methods to aggregate opinions on a simple problem, estimating grade point averages of 10 randomly selected students given sex, high school GPA, and SAT scores. The four aggregation methods produced estimates of comparable accuracy. Fischer concluded that in terms of accuracy "it makes little or no difference how one aggregates conflicting opinions of experts." He suggested cost and acceptability are likely to be relevant criteria. These methods differ on cost (P1 being the least expensive) and on acceptability by the group (#2 offering the highest). Fischer omitted mention of Dalkey (1969) who found that method 1 (averaging) was slightly superior to group discussion for simple problems, and of Hall, Mouton, and Blake (1963), who found method 2 (consensus) was superior to unstructured discussion in making predictions in what may have been a more complex case. My conclusion from these studies is that structured group process is superior to unstructured group process, but that a variety of structured approaches yield similar accuracy.

Fischer, Gregory W., (1982), "Scoring-rule feedback and the overconfidence syndrome in subjective probability forecasting," *Organizational Behavior and Human Performance*, vol. 20, pp. 352-369.

Subjects used information on the sex, SAT scores, and high school grades of 40 college freshmen to predict first-year grades. The feedback on outcome had no effect on overconfidence. Incentives did lead to better scores, but only because subjects were less likely to assign extremely low probabilities, which were heavily penalized. These results must be

viewed with caution, because the task proved to be so difficult for the subjects; by assuming an equal probability of being in each of the four categories (a strategy of “pure ignorance”), the subjects would have *improved* their predictions.

Fischhoff, Baruch and MacGregor, Don, (1982), “Subjective confidence in forecasts,” *Journal of Forecasting*, vol. 1, pp. 155-172.

Forecasts have little value to decision makers unless it is known how much confidence to place in them. Those expressions of confidence have, in turn, little value unless forecasters are able to assess the limits of their own knowledge accurately. Previous research has shown patterns in the judgments of individuals who have not received special training in confidence assessment: Knowledge generally increases as confidence increases. However, it increases too swiftly, with a doubling of confidence being associated with perhaps a 50% increase in knowledge. With all but the easiest of tasks, people tend to be overconfident about how much they know. These prior results were derived from studies of judgments of general knowledge. The present study found that they also pertained to confidence in *forecasts*; indeed, the confidence-knowledge curves observed here were strikingly similar to those observed previously. The only deviation was the discovery that a substantial minority of judges never expressed complete confidence in any of their forecasts; these individuals also proved to be better assessors of the extent of their own knowledge. Apparently confidence in forecasts is determined by processes similar to those that determine confidence in general knowledge. Decision makers can use forecasters’ assessments in a relative sense, in order to predict when they are more or less likely to be correct. However, they should be hesitant to take confidence assessments literally. Someone is more likely to be right when she is “certain” than when she is “fairly confident,” but there is no guarantee that the supposedly certain forecast will come true. The paper includes a table summarizing 37 studies that have tried to reduce overconfidence.

Fischhoff, Baruch; Slovic, Paul and Lichtenstein, Sarah, (1977), “Knowing with certainty: The appropriateness of extreme confidence,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 3, pp. 552-564.

Overconfidence was found with a variety of stimulus materials and response modes. Lectures on how to assess probabilities and how to avoid extreme probability predictions did little to reduce overconfidence.

Fischhoff, Baruch; Slovic, Paul and Lichtenstein, Sarah, (1978), “Fault trees: Sensitivity of estimated failure probabilities to problem representation,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, pp. 330-344.

Fault trees involve the causal decomposition of a complex event as a way to assess its likelihood. In this paper, the event is “car fails to start.” Subjects were asked to predict the likely causes for failure. When likely causes were omitted, subjects assigned higher probabilities to the potential causes remaining, and they made small but insufficient increases in the “other” category of possible causes. It was just as likely for people with more expertise to overlook causes that had been omitted. Increasing the amount of detail about the potential causes had little impact, except the probabilities for a cause could be increased by presenting it as two branches rather than one. Fault trees can be used for prediction prob-

lems with mechanical systems (e.g., to predict the likelihood of a failure at a nuclear plant) and for other problems involving multiple causality (e.g., what is the probability that two people will remain married for the next 30 years or that a firm will continue for the next 20 years).

Foster, George, (1977), "Quarterly accounting data: Time series properties and predictive ability results," *Accounting Review*, vol. 52, pp. 1-21.

This study is based on sales and earnings forecasts for 69 firms. It compares the accuracy of six extrapolation methods.

Fralicx, Rodney and Raju, Namburg S., (1982), "A comparison of five methods for combining multiple criteria into a single composite," *Educational and Psychological Measurement*, vol. 42, pp. 823-827.

Canonical correlations have been suggested in forecasting problems where a number of criteria are of interest and a number of predictors are available. The canonical weights determine the index that best predicts a criterion index. Canonical correlation is a method that is often used when theory is lacking. Theoretically, there is no reason to expect that a canonical index will be valid. This paper tests the validity of the canonical index for the formulation of a job performance index. The canonical index was compared with four alternative weighting schemes: managements' subjective weights, equal weights, unit weights, and principal components factor weights. The alternatives yielded *nearly identical weights* for judging the overall performance of 117 bank tellers based on eight performance criteria (e.g., customer relations, attention to detail). In contrast, the canonical weights (which used the eight performance criteria as well as 13 predictor variables such as memory and arithmetic ability) *had almost no correlation to the other methods*. It is distressing that the canonical index bore no relation to methods with high face validity (managements' subjective weights and equal weights).

Frank, Werner, (1969), "A study of the predictive significance of two income measures," *Journal of Accounting Research*, vol. 7, pp. 128-136.

Exponential smoothing was more accurate than a moving average and also more accurate than a regression against time.

Gaeth, Gary J. and Shanteau, James, (1984), "Reducing the influence of irrelevant information in experienced decision makers," *Organizational Behavior and Human Performance*, vol. 33 pp. 263-282.

Lectures were not effective in getting judges to ignore irrelevant information in an experiment where 12 judges rated the composition of soil samples. However, experiential learning was effective. Their errors were then noted (negative feedback). The judges were given advice and were then asked to make judgments. Further active training was then provided with an emphasis on positive feedback for good responses.

Gardner, Everette S., Jr., (1979), "Box-Jenkins vs. multiple regression: Some adventures in forecasting the demand for blood tests," *Interfaces*, vol. 9 (August), pp. 49-54.

On page 54 of Gardner's paper regression models #23 and #24 should say "*Dependent* variable lagged one period" instead of "*Independent* variable lagged one period."

Gardner, Everette S., Jr., (1979), "A note on forecast modification based upon residual analysis," *Decision Sciences*, vol 10, pp. 493-494.

Gardner, Everette S., Jr., (1988), "Automatic monitoring of forecast errors," *Journal of Forecasting*, vol. 2, pp. 1-21.

This paper evaluates a variety of automatic monitoring schemes to detect biased forecast errors. Backward cumulative sum (CUSUM) tracking signals have been recommended in previous research to monitor exponential smoothing models. This research shows that identical performance can be had with much simpler tracking signals. The smoothed-error signal is recommended for $a = 0.1$, although its performance deteriorates badly as a is increased. For higher a values, the simple CUSUM signal is recommended. Comments by the referees were published along with this paper. See also Gardner (1985a)

Gardner, Everette S., Jr., (1983), "Evolutionary operation of the exponential smoothing parameter: Revisited," *Omega*, vol. 11, pp. 621-623.

Gardner, Everette S., Jr., (1984), "The strange case of the lagging forecasts," *Interfaces*, vol. 14, (May – June), pp. 47-50.

Apparently it is difficult to explain exponential smoothing without making some type of error. Gardner found 23 books and articles with errors in model formulations for smoothing a linear trend.

Gardner, Everette S., Jr., (1985), "CUSUM vs. Smoothed-error forecast monitoring schemes: Some empirical comparisons," *Journal of the Operational Research Society*, vol. 36, pp. 43-47.

Gardner, Everette S., Jr., (1985), "Exponential smoothing: The state of the art," *Journal of Forecasting*, vol. 4, pp. 1-28 (commentary follows, pp. 29-38)

A comprehensive review of the literature.

Gardner, Everette S., Jr. and Dannenbring, David G., (1980), "Forecasting with exponential smoothing: Some guidelines for model selection," *Decision Sciences*, vol. 11, pp. 370-383.

This simulation experiment compared different extrapolation methods (Holt; Gilchrist; Montgomery; Simple Smoothing; Whybark; Trigg and Leach; Roberts and Reed; and Chow) to predict for 9,000 simulated time series (variations in levels, trends and random error). Used a variety of error measures (e.g., MAD, MSE, MAE). Adaptive models generated unstable forecasts, even when average demand was stable. This is an important paper.

Gardner, Everette S., Jr. and McKenzie, E, (1985), "Forecasting trends in time series," *Management Science*, 31, 1237-1246.

This paper analyzes data from the M-Competition and demonstrates procedures for automatic dampening of trend factors.

Geurts, Michael D., (1982), "Forecasting the Hawaiian tourist market," *Journal of Travel Research*, vol. 21 (Summer), pp. 18-21.

Replacement of outliers by the estimated values led to dramatic improvements in accuracy in forecasts for the two months following each of four atypical periods occurring over a two year period.

Glantz, Michael H., (1982), "Consequences and responsibilities in drought forecasting: The case of Yakima, 1977," *Water Resources Research*, vol. 18, pp. 3-18

This paper describes a U.S. Bureau of Reclamation forecast of a drought. The forecast led to actions to save crops. As one farmer put it, "Drought is when the government sends you a report telling you there's no water." However, the forecast was wrong. (No confidence interval was published, but the actual flow was much different from the forecast.) It appears, in this case, that the objective forecasting methods performed well but that subjective adjustments were made. The subjective adjustments led to the prediction of an extreme event. Attempts are being made to sue the government for malpractice in forecasting. Some questions: Will such legal actions lead to a greater reliance on objective methods of forecasting? Should good practice in forecasting require that confidence limits also be published with the forecast? Should forecasters intentionally bias forecasts if the loss function seems asymmetric (e.g., the cost of a drought might be seen as greater than the cost of a flood)?

Gomez-Mejia, Luis R., Page, Ronald C. and Tornow, Walter W., (1982), "A comparison of the practical utility of traditional, statistical, and hybrid job evaluation approaches," *Academy of Management Journal*, vol. 25, pp. 790-809.

This paper compares the predictive accuracy and the acceptability of different methods for classifying job levels for managers. It first presents an interesting literature review. Three statistical methods were then compared with three judgmental methods and with a "hybrid" method. The various methods were calibrated on the same samples and compared on a cross-validation sample of 150 managers. Some interesting conclusions resulted. *First*, a factor analysis of 235 potential predictor items, followed by a stepwise regression on the factors, was inferior to a direct stepwise regression on the variables (cross-validation r^2 of 38% and 58% respectively). *Second*, statistical procedures based on stepwise regression offered no advantage over traditional methods, such as "assign points to key factors and calculate a score." *Third*, regression weights on the variables selected by the experts (their "hybrid model") yielded improvements over the subjective weights (cross-validation of 64% and 55% respectively). And *fourth*, compensation practitioners rated this hybrid model as clearly the most acceptable, the traditional approaches were next most acceptable, and the purely statistical approaches were the least acceptable. My summary: *Use prior theory and judgment to develop a model*, then estimate relationships.

Gould, John P. and Wauo, R. (1973), "The neoclassical model of investment behavior: Another view," *International Economic Review*, vol. 14, pp. 33-48.

The authors examined one-quarter-ahead *ex post* forecasts for eight quarters in 11 industries. No clear-cut winner for accuracy of econometric versus extrapolation forecast accuracy.

Gray, Clifton W.,(1979), "Ingredients of intuitive regression," *Organizational Behavior and Human Performance*, vol. 28, pp. 30-48.

In this experiment, 44 subjects made predictions on a task with a single variable. Their intuitive regressions provided better predictions than did their direct predictions.

Green, Paul E. and Wind, Yoram,(1975), "New way to measure consumers' judgments," *Harvard Business Review*, vol. 53(July/August), pp. 107-117.

One of the *Harvard Business Review's* more popular papers, this provides a short and clear description of conjoint analysis.

Greer, Charles R. and Armstrong, Daniel, (1980), "Human resource forecasting and planning: A state of the art investigation," *Human Resource Planning*, vol. 3, pp. 67-78.

Survey of personnel managers at 300 firms drawn randomly from the *1979 College Placement Annual*. Used one follow-up and obtained a 29% response rate.

Gregory, W. Larry; Cialdini, R.B. and Carpenter, Kathleen M., (1982), "Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so?" *Journal of Personality and Social Psychology*, vol. 43, pp. 89-99.

Griffith, John R. and Wellman, B.T., (1979), "Forecasting bed needs and recommending facilities plans for community hospitals: A review of past performance," *Medical Care*, vol. 17, pp. 293-803 .

The authors examined forecasts for six hospitals. The forecasts, prepared by consultants between 1967 and 1971, covered the need for beds in 1975. The clients were often dissatisfied with the consultants' forecasts. They should have used them, however; the formal forecasts were more accurate than the intuitive forecasts used by the decision makers in the hospitals.

Guion, Robert M. and Cranny, C.J., (1982), "A note on concurrent and predictive validity designs: A critical reanalysis," *Journal of Applied Psychology*, vol. 67, pp. 239-246 .

A comment on Barrett, Phillips and Alexander (1981).

Hagerman, Robert L. and Ruland, William, (1979), "The accuracy of management forecasts and forecasts of simple alternative models," *Journal of Economics and Business*, vol. 31, pp. 172-179.

Studies 98 one-year earnings forecasts from the *Wall Street Journal*. Compares five extrapolation, one judgment, and one econometric method.

Hamill, Ruth, Wilson, T.D. and Nisbett, R.E., (1980), "Insensitivity to sample bias: Generalizing from atypical cases," *Journal of Personality and Social Psychology*, vol. 39, pp. 578-589.

Vivid examples have a strong impact on people's attitudes—much stronger it seems than carefully prepared statistical summaries from large samples. This generalization is drawn from prior research they cite as well as from the two clever experiments reported in this paper. The experiments have implications for the presentation of forecasts, as well as for the use of information to support a forecast. Of particular interest for the *presentation of a*

forecast is the use of scenarios. A vivid scenario would be expected to appear to be likely. According to this study, an event described in a scenario will be regarded as more likely even if the scenario was identified as being atypical or unlikely. This might be useful if one is trying to make a case for contingency planning. But scenarios may be dangerous if used to make predictions. Alternatively, vivid scenarios may help to improve estimates in cases where people seriously underestimate the probability. (Perhaps this is the intention of Ground Zero Demonstrations?) In the *presentation of data*, the choice of examples seems to influence people's attitudes more than the statistical information, even when the example is identified as being atypical. To avoid bias in presentation, one should select typical examples. The dangers of atypical examples should be recognized: even more powerful than "lying with statistics" is the opportunity of "lying with examples."

Hanke, John E., (1984), "Forecasting in the business schools," *Journal of Forecasting*, vol. 3, pp. 229-234.

Hanke sent a survey to 620 member institutions of the American Assembly of Collegiate Schools of Business. Responses were received from 52 % of the schools. Forecasting courses were offered in 60 % of the schools. Regression analysis is the most important technique that is taught in the courses, 83 % of the classes use a project, and, surprisingly, judgment methods are hardly ever taught.

Hatjoulis, G. and Wood, D., (1979), "Economic forecasts—An analysis of performance," *Business Economist*, vol. 10 (Spring), pp. 6-21.

In general, the econometric models were superior to extrapolation models (either a no-change or an exponential smoothing model).

Hill, Gareth and Fildes, Robert, (1984), "The accuracy of extrapolation methods: An automatic Box-Jenkins package (SIFT)," *Journal of Forecasting*, vol. 3, pp. 319-323.

Hinrichs, John R., (1969), "Comparison of real-life assessments of potential with situational exercises, paper and pencil ability tests, and personality inventories," *Journal of Applied Psychology*, vol. 53, pp. 425-432.

Concludes that an "assessment center" evaluation may be unnecessary if a reliable and relevant employment history is available. However, assessment centers might be useful when no job history exists for an individual.

Hinrichs, John R. (1978), "An eight-year follow-up of a management assessment center," *Journal of Applied Psychology*, vol. 63, pp. 596-601.

More evidence that an inexpensive prediction based on a review of personnel files did as well as the assessment center in predicting advancement for a group of sales persons. Not surprising is that those who were promoted in this organization were described as upwardly mobile.

Hirsch, Albert; Grimm, B.T. and Narasimham, G.V.L. (1974), "Some multiplier and error characteristics of the BEA Quarterly Model," *International Economic Review*, vol. 15, pp. 616-631.

Hogarth, Robin M. (1978), "A note on aggregating opinions," *Organizational Behavior and Human Performance*, vol. 21, pp. 40-46.

How many experts should you use in forecasting a given variable? Hogarth, using theoretical arguments, concludes that one should use at least 6, but no more than 20 experts. You should tend toward the higher side of this range if your experts differ among one another in their forecasts and if they can make good forecasts. A good rule of thumb is to use ten experts.

Hogarth, Robin M. and Makridakis, Spyros (1981), "Forecasting and planning: An evaluation," *Management Science*, vol. 27, pp. 115-138.

This paper organizes and reviews research on the judgmental aspects of forecasting and planning. Contains 175 references.

Holmes, David S., et al. (1980), "Biorhythms: Their utility for predicting post-operative recuperative time, death and athletic performance," *Journal of Applied Psychology*, vol. 65, pp. 233-236.

The marketplace declares biorhythms to be a winner! Unfortunately, research studies do not agree; they find no evidence that biorhythms improve forecast accuracy. Holmes, et al., add three more competent studies.

Hopwood, William S., Newbold, Paul and Silhan, Peter A. (1982), "The potential for gains in predictive ability through disaggregation: Segmented annual earnings," *Journal of Accounting Research*, vol. 20, pp. 724-732 .

This study examined forecasts for fictitious conglomerates (constructed by averaging across two to five actual firms), by extrapolating the composite earnings directly, and comparing the forecasts with those built up from separate extrapolations of each of the components. Minor gains in accuracy were obtained when this was done for the Box-Jenkins and exponential smoothing methods. However, the most accurate forecasts were provided by the no change forecasts (where the issue of segmentation was irrelevant).

Howrey, E.P., Klein, L.R. and McCarthy, M.D., (1974), "Notes on testing the predictive performance of econometric models," *International Economic Review*, vol. 15, pp. 366-383.

A version of the Wharton Econometric Model was found to be more accurate than an autoregressive model in quarterly forecasts of real GNP, 1955-1966. The superiority of the econometric model increased as the forecast horizon was increased. The authors discuss some of the problems in drawing inferences from comparisons of the forecast accuracy of alternative models.

Inciardi, James A. (1977), "The parole prediction myth," *International Journal of Criminology and Penology*, vol. 5, pp. 285-244.

Johnson, Timothy E. and Schmitt, Thomas G. (1974), "Effectiveness of earnings per share forecasts," *Financial Management*, vol. 3, pp. 64-72.

Uses data on annual income from 150 industrial companies for 1962-1971 to compare forecast accuracy of no-change, moving average, regression against time, exponential smoothing (with and without trend), adaptive exponential smoothing, and triple smoothing. Basic conclusions: not much difference in accuracy among the various methods, and last year's earnings provided a good forecast of next year's earnings.

Johnson, W. Bruce (1983), "Representativeness in judgmental predictions of corporate bankruptcy," *Accounting Review*, vol. 58, pp. 78-97.

Kahneman, Daniel; Slovic, Paul and Tversky, Amos (Eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, England: Cambridge University Press.

A useful collection of papers dealing mostly with shortcomings in human judgment, most of which have been previously published. (For a more detailed review see Armstrong, 1984.)

Kahneman, Daniel and Tversky, Amos, (1979), "Intuitive prediction: Biases and corrective procedures," in Spyros Makridakis and Steven Wheelwright (eds.), *Forecasting*. Amsterdam: North Holland.

Provides a good introduction to the work of Kahneman and Tversky and how it relates to forecasting.

Kalton, Graham and Schuman, Howard (1982), "The effect of the question on survey responses: A review," *Journal of the Royal Statistical Society: Series A*, vol. 145, Pt 1, pp. 42-73.

In addition to providing an overview of recent research on the topic of response error, this paper describes some recent efforts to reduce the errors. These include the use of instructions to the respondent, feedback to the respondent, and the gaining of commitment from the respondent to provide accurate answers.

Kalwani, Manohar and Silk, Alvin J. (1982), "On the reliability and predictive validity of purchase intention measures," *Marketing Science*, vol. 1, pp. 243-286.

In general, intentions understate actual purchase rates.

Keen, Howard, Jr. (1981), "Who forecasts best? Some evidence from the Livingston survey," *Business Economics*, vol. 16 (September), pp. 24-29.

In June and December of each year since 1946, Joseph A. Livingston, a business journalist for the *Philadelphia Inquirer*, has been publishing forecasts of business variables based on a survey of about 50 experts. (Details on the Livingston survey are available from: Research Department, Federal Reserve Bank of Philadelphia, Philadelphia, Pa. 19105) Keen analyzed forecasts from 1971 to 1978 in an effort to tell which forecasters were best: those from academia, banking, or business? No consistent differences were found in the forecasts of nominal GNP, real GNP, consumer prices, and unemployment when considering size of error and turning points. Another issue Keen examined was whether the Livingston forecasts were better than the nochange model for 6 and 12-month-ahead-

forecasts. They were, with the exception of forecasts for the industrial stock price index. This is reassuring and is consistent with findings from previous studies. (See also Ahlers and Lakinishok (1983).)

Kenny, Peter B. and Durbin, James,(1982), “Local trend estimation and seasonal adjustment of economic and social time series” (with discussion), *Journal of the Royal Statistical Society: Series A*, vol. 145, Pt 1, pp. 1-41.

Keren, Gideon and Newman, J. Robert (1978), “Additional considerations with regard to multiple regression and equal weighting,” *Organizational Behavior and Human Performance*, vol. 22, pp. 143-164.

Suggests that measurement errors in the criterion (dependent) variable are more damaging to the model than errors in measuring causal variables. Also discusses ridge regression and says that it “shrinks” the estimated relationships towards the origin (it mitigates the estimate of the forecasted relationship). Finally, this paper presents results of a simulation study with three predictor variables and one criterion. Ridge regression was more accurate than OLS, and OLS was more accurate than unit weights.

Kerr, Norbert L., Nerenz, David R. and Herrick, David, (1979), “Role playing and the study of jury behavior,” *Sociological Methods and Research*, vol. 7, pp. 337-355.

This experiment used 117 mock juries and 108 real juries (at least they *thought* they were real) in a case involving student discipline. Prior to the role playing, 48 % of the *individuals* in the mock jury thought the defendant was guilty. For six-person juries, assuming the majority would prevail, this means that 40 % of the juries would have been expected to reach a verdict of guilty. But in the real trial, the defendant was never found guilty (0 of 10 juries, though there were 8 hung juries). These results were matched by the mock juries (guilty in 1 of 12 juries with 8 hung juries).

Koriat, Asher, Lichtenstein, Sarah and Fischhoff, Baruch (1980), “Reasons for confidence,” *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, pp. 107-118.

People tend to think of the reasons to support a given decision or forecast: this leads to overconfidence. This study traces its roots to an idea of Ben Franklin’s: Making an explicit list of the reasons that contradicted their answers in a test of knowledge led subjects to provide more realistic estimates of confidence in their answers. They also found a slight (but not significant) tendency for the resulting answers to be more accurate. This is an important study.

Kreilkamp, Karl (1971), “*Hindsight* and the real world of science policy,” *Science Studies*, vol. 1, pp. 43-66.

Larcker, David F. and Lessig, V. Parker (1983), “An examination of the linear and retrospective process tracing approaches to judgment modeling,” *Accounting Review*, vol. 58, pp. 58-77.

This study asks people to describe, after the fact, how they made decisions to buy stocks. The authors refer to this as “retrospective process tracing models” an unfortunate term in my opinion. How about calling it a “memory model”? Buy/no-buy decisions were made

for 45 stocks by 31 subjects. Each stock was described by six relevant and obvious variables. An indirect bootstrapping, done by discriminant analysis, matched the actual decision in 73 % of the cases. The memory model (done immediately after the completion of all stock decisions) was significantly better ($p < .05$), and it matched the actual for 85% of the decisions. The gain came at some cost, as the memory model required % hour with each subject. The authors caution that the results may not be applicable to more complex problems or to problems where irrelevant variables are present. They recommend a combined use of discriminant models, memory models, and “concurrent process tracing.” It is a thorough study and the literature review brings together a number of relevant findings from accounting, marketing, and psychology. Although this is a long-winded paper with much jargon, it is important and will be rewarding to those who manage to stay awake.

Larson, James R., Jr. and Reenan, A.M. (1979), “The equivalence interval as a measure of uncertainty,” *Organizational Behavior and Human Performance*, vol. 28, pp. 49-55.

Questions 60 subjects on number of marbles in cartons and similar tasks. Confidence intervals for a judge became large as the judge’s accuracy decreased. Uncertainty intervals (based on the range “outside of which they were reasonably certain that the correct answer did not lie) contained the correct answer about 60% of the time.

Lawrence, Michael J. (1988), “An exploration of some practical issues in the use of quantitative forecasting models,” *Journal of Forecasting*, vol. 2, pp. 169-179.

A small survey of a convenience sample of firms in Australia indicated that computer-based forecasting systems are not widely used and, in fact, a number of established systems have been discarded, due to poor accuracy. Other problem areas mentioned as contributing to the abandonment of forecasting systems include the difficulty of manually reviewing the computer forecasts and the effort required to review carefully the forecast database to adjust for extraordinary events.

Lawrence, Michael J., Edmundson, R.H. and O’Connor, M.J.,(1985), “An examination of the accuracy of judgmental extrapolation of time series,” *International Journal of Forecasting*, vol. 1, pp. 29-35.

Ledolter, Johannes and Abraham, Bovas (1981), “Parsimony and its importance in time series forecasting,” *Technometrics*, vol. 23, pp. 411-414.

Leigh, Thomas W., MacKay, David B. and Summers, John O. (1984), “Reliability and validity of conjoint analysis and self-explicated weights: A comparison,” *Journal of Marketing Research*, vol. 21, pp. 456-462.

The technology conjoint analysis is highly developed by marketing researchers. Nevertheless, Leigh, et al. were unable to find a single study that tested the predictive validity of this approach in comparison to the direct approach. They then made a test by obtaining data from 122 business students about their preferences for hand held calculators. Predictions from a variety of indirect approaches were compared with those from a simple and low cost direct approach. The predicted behavior was the choice of a

calculator from a list of 10 in a lottery. Few differences were found among the 12 different indirect approaches that were examined, so these were compared, as a group, with the direct approach. The direct approach proved to be slightly more accurate (36.3 % correct predictions vs. 34.9 %, where chance would be 10 %), a result that was statistically significant. The direct approach was also more reliable based on a test-retest with the same subjects. An unfortunate problem with this study is that the direct bootstrapping always followed the indirect bootstrapping for each subject.

Libby, Robert and Blashfield, Roger K., (1978), "Performance of a composite as a function of the number of judges," *Organizational Behavior and Human Performance*, vol. 21, pp. 121-129.

Based on three empirical studies, they show significant gains in accuracy obtained when going from one judge to using an average based on three judges. According to the authors, the optimum number of judges is likely to be between five and nine.

Libert, G.,(1984), "The M-Competition with a fully automatic Box-Jenkins procedure," *Journal of Forecasting*, vol. 8, pp. 325-328.

Lichtenstekin, Sarah and Fischhoff, Baruch,(1980), "Training for calibration," *Organizational Behavior and Human Performance*, vol. 26, pp. 149-171.

Complex paper describing two experiments.

Lindley, D.V., (1982), "The improvement of probability judgments," *Journal of the Royal Statistical Society: Series A*, vol. 145, pp. 117-126.

I was not able to learn much from this study, with the exception of pages 122 and 128 where the responses of a given subject to 500 almanac-type questions are discussed.

Lord, Charles G., Ross, Lee and Lepper, Mark R.,(1979), "Biased assimilation and attribute polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, vol. 37, pp. 2098-2109.

Subjects in this study were provided with evidence for and against capital punishment. The methods used in the "study" on capital punishment were rated higher when the results agreed with the subject's prior opinion.

Louis, Arthur N., (1978), "Should you buy biorhythms," *Psychology Today*, vol. 11 (April), pp. 93-96.

No. Biorhythms did not help in this study for predictions in baseball and boxing.

Lyon, Don and Slovic, Paul,(1976), "Dominance of accuracy information and neglect of base rates in probability estimation," *Acta Psychologica*, vol. 40, pp. 287-298.

One of the primary tasks of the forecaster is to help the client make better forecasts. As a result of this effort, however, the client may be misled and put too much weight on the conclusions presented by a forecaster. Lyon and Slovic use three problems, including the Blue and Green Cab problem, to show how people can be easily misled; in this case, subjects used new information and ignored prior knowledge of base rates.

Mabert, Vincent A. (1976), "Statistical versus sales force-executive opinion short range forecasts: A time series case study," *Decision Sciences*, vol. 7, pp. 310-318.

This study compared judgmental forecasts by one company against Winters' exponential smoothing, Brown's harmonic model, and Box-Jenkins on four-week forecasts for five years (1968-1972). Successive updating was used so that 65 monthly (actually four week) forecasts were obtained for each model. The extrapolation methods were cheaper and more accurate than the judgmental forecasts. Few differences were found among the extrapolation methods.

Mabert, Vincent A. (1978), "Forecast modification based upon residual analysis: A case study of check volume estimation," *Decision Sciences*, vol. 9, pp. 285-296.

Looked at daily forecasts of check volume in a commercial bank. Two forecasting procedures were evaluated. First, a dummy variable regression model was used to estimate check volume forecast. A second approach used regression, then used exponential smoothing on the residuals, was used to see if improved results could be obtained. This second approach utilized exponential smoothing in an attempt to adapt to systematic forecast errors that were identified. At best, the latter approach improved the forecasts marginally. For a critique see Gardner (1979) followed by Mabert's reply.

Makridakis, Spyros, 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*, vol. 1, pp. 111-153.

This study of the comparative accuracy of 21 methods for *ex ante* forecasts of 1001 time series is one of the most important works that has been done on forecasting methods. It is often referred to as the M-Competition. This paper presents the results on the accuracy on each method. Makridakis, et al. provide more detail on each of the methods. A discussion by outside commentators, as well as by the original authors can be found in Armstrong and Lusk (1983).

Makridakis, Spyros, et al. (1984), *The Forecasting Accuracy of Major Time Series Methods*. Chichester: Wiley.

This book provides a detailed report on the M-Competition (Makridakis et al., 1984) by having each of the original authors explain each method.

Makridakis, Spyros and Hibon, Michele (1979), "Accuracy of forecasting: An empirical investigation" (with discussion), *Journal of the Royal Statistical Society: Series A*, vol. 142, pp. 97-145.

A study of the comparative accuracy of 111 time series that varied by country, time period, industry, company, and time intervals. A forerunner of Makridakis et al. (1983). Contains numerous interesting conclusions – e.g., adaptive parameters were not useful for exponential smoothing.

Makridakis, Spyros and Wheelwright, Steven C. (1979), *Forecasting. (TIMS Studies in Management Science, Volume 12)*. New York: North Holland.

Contains 21 papers on a variety of topics in forecasting. The contributions were reviewed by 73 referees. An appendix reviews 40 books on forecasting.

Makridakis, Spyros and Wheelright, Steven C. (1987. 2nd ed.), *The Handbook of Forecasting: A Manager's Guide*. New York: Wiley.

“Handbook” is perhaps not a descriptive title. Instead, this is a collection of papers on a wide variety of subjects in forecasting.

Makridakis, Spyros and Winkler, Robert L. (1983), “Averages of forecasts: Some empirical results,” *Management Science*, vol. 29, pp. 987-996.

Manegold, James G. (1981), “Time series properties of earnings: A comparison of extrapolative and component models,” *Journal of Accounting Research*, vol. 19 (Autumn), pp. 360-373.

Examined one-year-ahead earnings forecasts for 27 firms for 1974 and 1975 and two-year-ahead forecasts for 1975.

Manis, Melvin, Dovalina, Ismael, Avis, Nancy E. and Cardoze, Steven (1980), “Base rates can affect individual predictions,” *Journal of Personality and Social Psychology*, vol. 38, pp. 231-248, and Manis, Melvin; Avis, Nancy E. and Cardoze, Steven, (1981), “Reply to Bar-Hillel and Fischhoff,” *Journal of Personality and Social Psychology*, vol. 41, pp. 681-683.

Judgmental forecasters should consider what typically happens in a situation (the “base rate”) as well as specific information available about the case in hand. Although the specific information should be considered only if it is valid and reliable, earlier research showed that even irrelevant specific information led people to ignore the base rate. This paper reports on four experiments and reexamines previous studies. From these, the authors defined a set of conditions in which judges showed much sensitivity to base rates. Their study seemed convincing to me until I read the Bar-Hillel and Fischhoff (1981) paper, which reinterpreted Manis et al. (1980) and concluded that their results were consistent with previous research: Base rates are important when the subject does not receive information on representativeness (evidence that the subject of the prediction fits a stereotype). Manis et al. (1981) is a well-reasoned reply to Bar-Hillel and Fischhoff. The combination of articles helps to specify the conditions under which forecasters should not trust their intuitions when interpreting base rates plus specific information. An interesting finding in Manis et al. (1980) was that subjects did not seem to be aware of the occasions on which they used base rates in making their predictions.

Marks, Robert E. (1980), “The value of ‘almost’ perfect weather information to the Australian tertiary sector,” *Australian Journal of Management*, vol. 5, pp. 67-85.

Marks provides a clear discussion on how to assess the *potential* value of improved forecasts. He then applies this, using a mail survey of 131 corporations in Australia. These “tertiary” (or service) sector corporations consisted mostly of electricity, gas, water, construction, transport, and communications firms. Responses, received from 46% of the firms surveyed, were used to estimate the value of improved forecast accuracy as a percentage of revenues for each type of corporation. These were then compared with estimates from the United States. The paper integrates much previous research, but one

relevant omission was Schnee (1977). This would have been interesting because Schnee concluded that the costs of more accurate weather forecasting exceeded its potential benefits, even if the forecasts were perfect. Marks did not consider the costs of better weather forecasts, only the potential gross benefits. The savings are *potential* because, if people do not act on these forecasts, they are of no value. Judging from studies, such as Baker (1979), it appears that weather forecasts often are not used effectively.

McClain, John O. (1974), "Dynamics of exponential smoothing with trend and seasonal terms," *Management Science*, vol. 20, pp. 1300-1304.

McClain's theoretical analysis leads him to conclude that Brown's exponential smoothing model responds appropriately to changes in the data.

McIntyre, Shelby H. (1981), "An experimental study of the impact of judgment-based models," *Management Science*, vol. 28, pp. 17-33.

Formal processing of information by each of 96 judges did *not* lead to better predictions in a management game, although it did improve their decision making.

McIntyre, Shelby H., Montgomery, D.B., Srinivasan, V. and Weitz, B.A. (1983), "Evaluating the statistical significance of models developed by stepwise regression," *Journal of Marketing Research*, vol. 20, pp. 1-11.

This paper discusses how to test for statistical significance when stepwise regression is used. Tables are provided for more realistic tests of significance than those typically used.

McLeavey, Dennis W., Lee, T.S. and Adam, E.E., Jr. (1981), "An empirical evaluation of individual item forecasting models," *Decision Sciences*, vol. 12, pp. 708-714.

The importance of replication is highlighted by this study. It replicates a study by Adam (1973) and finds that two of the seven models in the original paper were in error. However, the general results were similar when seven models were used to make one-period and 12-period forecasts for five different simulated demand patterns. It is not easy to make generalizations from this study, but here are mine: For one-period-ahead forecasts, a two period moving average performed well for constant, trend and seasonal patterns, for a combination of all these patterns, and for a step function. None of the five exponential models produced significant improvements, and the adaptive smoothing model was less accurate. Double exponential smoothing performed well for all demand patterns on both one-period-ahead and twelve-period-ahead forecasts. Which model was most accurate depended upon the demand pattern and the forecast horizon, as well as upon the noise level.

McWhorter, Archex, Narasimham, G.V.L. and Simonds, R.R. (1977), "An empirical examination of the predictive performance of an econometric model with random coefficients," *International Statistical Review*, vol. 45, pp. 243-255.

This study examined ex post forecasts using 16 quarters of validation data from 1971.I to 1970.IV. The forecasts, made for 1 to 4 quarters in the future, were based on data from 1950.I to 1970.IV. A variety of extrapolation and econometric models were used. The extrapolation forecasts proved to have much lower errors, about % as large as those from the econometric models. (I suspect that this is due to problems in the estimation of the current

status.). OLS performed well in comparison with a more complex approach (three-stage least squares). Exponential smoothing with trend did about the same as the more complex ARIMA and better, it appears, than Kalman filter methods.

Meade, Nigel (1984), "The use of growth curves in forecasting market development—A review and appraisal," *Journal of Forecasting*, vol. 3, pp. 429-451.

Mentzer, John T. and Cox, James E., Jr. (1984), "Familiarity, application, and performance of sales forecasting techniques," *Journal of Forecasting*, vol. 3, pp. 27-36.

Reports on a survey sent to forecasting managers in 500 U.S. companies. Usable replies were received from 32 % of the companies.

Milstein, Robert M., et al. (1980), "Prediction of interview ratings in a medical school admission process," *Journal of Medical Education*, vol. 55, pp. 451-453 and Milstein, Robert M., et al. (1981), "Admissions decisions and performance during medical school," *Journal of Medical Education*, vol. 56, pp. 77-82.

Millions of dollars are spent each year on personal interviews for admission to medical school. Milstein's 1980 study found that differences between interviewer and interviewee were of major importance: The greater the difference, the lower the prediction of success. The 1981 study examined 24 applicants who were interviewed and accepted at Yale's School of Medicine, but who went elsewhere to medical school (AYEs). They were compared with 27 applicants interviewed and rejected by Yale who also went elsewhere to medical school (NAYs). No differences were found between the medical school performance of AYEs and NAYs. That is, for predictive purposes, the interview was worthless. The conclusion is consistent with the research on personnel selection in business.

Mitchell, Terry W. and Klimoski, Richard J. (1982), "Is it rational to be empirical? A test of methods for scoring biographical data," *Journal of Applied Psychology*, vol. 67, pp. 411-481 .

Biographical data on 88 variables (e.g., education, work experience, family background) were used to predict success in obtaining a real estate license for 698 prospective applicants. The researchers compared two methods. The first method, which they call "empirical," used a nontheoretical approach to weighting (the weighted application blank). The weights were obtained from a subsample. The second method, called "rational," used the same subsample, it started with subjective weights for each variable, followed by a factor analysis to yield six factors. The six factors were then entered into a regression model. (I do not agree with the authors who claim this approach to be "rational" and to provide a "better understanding.") The empirical approach provided more accurate forecasts in the cross-validation sample. This result is consistent with the few previously published empirical results: *factor analysis of predictor variables has not been shown to have any demonstrable value in forecasting.*

Moore, Geoffrey H. (1983), *Business Cycles, Inflation and Forecasting*. Cambridge, Mass.: Ballanger.

This book examines the behavior of macroeconomic variables during the course of business cycles in the United States. It also includes an update on leading indicators in the United States, Canada, United Kingdom, West Germany, Italy, France, and Japan.

More, Roger A. and Little, Blair, (1980), "The application of discriminant analysis to the prediction of sales forecast uncertainty in new product situations," *Journal of the Operational Research Society*, vol. 31, pp. 71-77.

Is it possible to assess the sales forecast uncertainty for a new product introduction? More and Little address this important question, more or less. They present a conceptual model relating the error in the first year's sales forecast to marketing task similarity and marketing task complexity. (Complexity was a function of buyer-risk, distribution difficulty, and competitive advantage.) Data from 185 new product situations were collected by personal interviews and self-administered questionnaires from 152 Canadian firms. The discriminant function did somewhat better than chance in identifying the high risk introductions (over 20% error in unit sales) when tested on a hold-out sample. This test was biased because the respondents knew the outcome. Further, a more revealing comparison than testing against chance would be to test against the currently used subjective methods.

Morris, John D. (1982), "Ridge regression and some alternative weighting techniques: A comment on Darlington," *Psychological Bulletin*, vol. 91, pp. 203-210.

Some researcher have advocated ridge regression as a way to obtain better estimates of parameters and, presumably, better predictions. While advocates have used theoretical arguments, Darlington (1978), in a widely cited paper, provided empirical evidence supporting a ridge regression by showing that it led to more accurate predictions in hold-out samples. Rozeboom (1979) challenged the applicability of Darlington's results because they depend on knowing the optimal value of a key constant (k) in the ridge regression, and because Darlington did not consider the effects of such practical considerations as sample sizes. Morris reanalyzed Darlington's results, using Darlington's simulated data, by estimating the constant k from the sample data. He contrasted the predictive validity on hold-out samples with that obtained from four other estimation procedures. Darlington's recommended one-parameter ridge regression technique was found *never* to be superior to the other methods. The best results were nearby always provided by either ordinary least squares or by equal weights. Furthermore, the differences among the predictive validities of the various methods were small, so one might question the practical significance of these alternative approaches.

Morris, John D. (1981), "Updating the criterion for regression predictor variable selection," *Educational and Psychological Measurement*, vol. 41, pp. 777-780.

Proposals have been made that predictor variables should be selected on how well they perform on the cross-validated sample rather than on the calibration sample. (As a proponent of *theory* as the proper way to select predictor variables, I have not been among those making such suggestions.) This paper reviews evidence on the use of cross-validated selection procedures. He described a program that can use forward or backward stepwise procedures or it can examine all possible combinations of variables to obtain the

best cross-validated model. Will such procedures improve our ability to forecast? That is a good topic for further research. (My guess is “No.”)

Morris, M.J. (1977), “Forecasting the sunspot cycle,” *Journal of the Royal Statistical Society: Series A*, vol. 140, pp. 437-478.

Combinations of two different extrapolation models led to a 50 % reduction in forecast error. However, the weights were selected after the fact.

Moskowitz, Herbert, Weiss, Doyle L., Cheng, K.K. and Reibstein, David J. (1982), “Robustness of linear models in dynamic multivariate predictions,” *Omega*, vol. 10, pp. 647-661.

This paper shows how, in some cases, unit weights and equal weights are not identical. (Unit weights differed from equal weights due to the dynamic nature of their short-range production planning problems.) They concluded that equal weights were better than human decisions and almost as good as regression weights. Some interesting ideas surrounded by a complex writing style. For a related paper see Remus and Jenicke (1978).

Moyer, R. Charles (1977), “Forecasting financial failure: A reexamination,” *Financial Management*, vol. 6 (Spring), pp. 11-17.

This study attempts to predict bankruptcy in a sample of 27 bankrupt and 27 non-bankrupt firms. Predictive accuracy did not increase when the number of predictor variables was increased from two to five.

Murphy, Allan H. and Brown, Barbara G. (1984), “A comparative evaluation of objective and subjective weather forecasts in the United States,” *Journal of Forecasting*, vol. 3, pp. 369-393.

Comparisons of objective and subjective forecasts of precipitation occurrence indicate that the latter are more accurate for the short lead-times (12-24 hours), but they are about the same for longer lead-times (e.g., 36-48 hours). Objective forecasts of cloud cover are more accurate than subjective forecasts for all lead-times. Forecasts accuracy improved over the period from 1971 to 1982, especially for the objective forecasts. Many weather forecasters make a subjective analysis of the data before examining the objective forecasts not to be unduly influenced by the latter. However, little study has been done on the subjective forecasting process and forecasts that start with the objective forecasts may be just as accurate. This is an excellent review.

Murphy, Allan H. and Daan, Harald (1984), “Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zierikzee experiment,” *Monthly Weather Review*, vol. 112, pp. 413-428.

Examines short-range forecasts of wind, fog, and rain that were prepared by four forecasters in 1981 and 1982.

Murphy, Allan H., Lichtenstein, S., Fischhoff, B. and Winkler, R.L. (1980), “Misinterpretations of precipitation probability forecasts,” *Bulletin of the American Meteorological Society*, vol. 61, pp. 695-701.

Managers often tell me that it introduces too much complexity to use probabilities or distributions when presenting forecasts in their organizations. *Other* managers will be

confused, they say. In this survey of 79 residents of Eugene, Oregon, the authors show that the general public has a good understanding of the meaning of precipitation probability forecasts. They even *preferred* probability forecasts.

Narasimhan, Chakravarthi and Sen, Subrata (1988), "New product models for test market data," *Journal of Marketing*, vol. 47 (Winter), pp. 11-24.

Reviews the more widely known models for predicting the success of new products based on test market data. The earlier models tend to rely on the extrapolation of trial and repeat purchase sales. The more recent models also examine the impact of marketing variables.

Nelson, Charles R. (1984), "A benchmark for the accuracy of econometric forecasts of GNP," *Business Economics*, vol. 19 (April), pp. 52-58.

This paper compares extrapolation and econometric forecasts of quarterly GNP from 1976-1982 for one-, two-, three-, and four-quarters-ahead. The econometric forecasts were five of those published by the Conference Board in its *Statistical Bulletin* (Chase, Conference Board, DRI, Kent, and Michigan). The econometric forecasts were based on slightly more recent data than the extrapolations and, as a normal course, they were subjectively adjusted. So what do you predict as more accurate for one-quarter forecasts? . . . for four-quarter ahead forecasts? The answers: The extrapolation method had a higher error for one-quarter-horizon, but lower errors for the longer horizons, especially four-quarters-ahead. (No statistical tests were provided.) Nelson also examined whether a combination of the forecasts would be better than any single forecast. The answer seems to be "yes." (I say "seems" because the weights were selected in retrospect.)

Neslin, Scott A. (1981), "Linking product features to perceptions: Self-stated versus statistically revealed importance weights," *Journal of Marketing Research*, vol. 18, pp. 80-86.

The statistical approach of inferring the model (judgmental bootstrapping) was superior to asking people to state their rules (expert systems) in predicting individual perceptions for an ambulatory health service. This finding emerged from answers given by 112 respondents.

Nevin, John R. (1974), "Laboratory experiments for estimating consumer demand: A validation study," *Journal of Marketing Research*, vol. 11, pp. 261-268.

A lab experiment and a questionnaire each produced good estimates of market share and of market share changes for Coke, Pepsi, and RC, but not such good estimates for brands of coffee. Price elasticity estimates were highest in the questionnaire, followed by simulated shopping and test market methods.

Nisbett, Richard E. and Wilson, T.D. (1977), "Telling more than we can know: Verbal reports on mental processes," *Psychological Review*, vol. 84, pp. 231-259.

Contains clever studies suggesting that people are often not aware of how they make decisions or predictions.

Page, Carl V. (1977), "Heuristics for signature table analysis as a pattern recognition technique," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 7, No. 2 (February), pp. 77-86.

Cross-classifications were slightly more accurate than regressions in predicting for a cross-validation sample (74% vs. 70% correct predictions). Unfortunately, this paper is filled with jargon.

Palmore, Erdman (1979), "Predictors of successful aging," *Gerontologist*, vol. 19, pp. 427-431. Measures of current status (initial health and happiness) were important in predicting health changes over a 9½ year period for older people. .

Parente, Frederick J., Anderson, J.K., Myers, P. and O'Brien, T. (1984), "An examination of factors contributing to Delphi accuracy," *Journal of Forecasting*, vol. 8, pp. 178-182. Draws a distinction between forecasting *if* an event will occur and *when* it will occur. Additional rounds of polling in Delphi did not help to improve "if" predictions, but they did improve predictions for "when." Feedback in Delphi did not improve the accuracy of either *if* or *when* predictions.

Parker, Barnett R. and Srinivasan, V. (1976), "A consumer preference approach to the planning of rural primary health-care facilities," *Operations Research*, vol. 24, pp. 991-1025. Note in particular-pages 1009-1025, which deal with the forecasting issues. They provide tests of face validity, cross-validity, and predictive validity and concluded that differential weights were superior to equal weights in predicting individual patient's choices of health care facilities.

Pencavel, John H. (1971), "A note on the predictive performance of wage inflation models of the British economy," *Economic Journal*, vol. 81, pp. 113-119.

A "constant change" extrapolation (based on last year's change) was more accurate than five econometric models for 1962-1967. (The econometric forecasts were *ex post*.) The relative accuracy of the five econometric models varied each year. That is, a high rank in one year was not more likely to be followed by a high rank the next year.

Perry, Paul (1979) "Certain problems in election survey methodology," *Public Opinion Quarterly*, vol. 43, pp. 312-325.

Perry describes various technical advances that have contributed to the improved accuracy of intentions forecasts.

Peters, Lawrence H., Jackofsky, Ellen F. and Salter, James R., (1981), "Predicting turnover: A comparison of part-time and full-time employees," *Journal of Occupational Behavior*, vol. 2, pp. 89-98.

This study attempted to predict employee turnover in a telephone sales job over the 12 months following hiring. Separate analytical models were developed for full and part-time workers. Predictions were based on items from a survey taken two months after the employees started work. Demographic variables were similar for each group with the exception that part-time workers lived closer to their places of employment. Key variables in this study were all derived from previous literature on turnover, and included were job satisfaction, thoughts of quitting, expectation of finding alternative employment, job search behavior, and intention to quit. These five variables all helped to predict turnover

for full-timers-but none of them helped to predict turnover among part-timers! As shown in this study, it is frequently useful to segment a problem and then to develop a model for each segment. The segmentation in this study might be thought of in terms of the *importance of the decision*. It is generally easier to predict how people will behave for important decisions.

Pfaff, Philip (1977), "Evaluation of some money stock forecasting models," *Journal of Finance*, vol. 32, pp. 1639-1646.

Compares four extrapolation models with eight econometric models, all estimated from 1947-1960 quarterly data, in making one to six-quarter-ahead *ex post* forecasts of the money stock over the 1961 to 1970 period. The extrapolation models were superior to the econometric models, despite the fact that the latter were recognized by academics to be the leading models in the field. The RMSE for the econometric models were cut in half for one-quarter-ahead forecasts by merely adjusting the forecast to compensate for the previous quarter's forecast error. This adjustment was of less value as the forecast horizon increased to six-quarters-ahead where it was of no value. (Instead of this mechanical error adjustment, a lagged dependent variable could be added as a predictor variable.) A decomposed extrapolation, based on extrapolations of its five components, was more accurate than a global extrapolation for the medium-range (six-quarter-ahead) forecasts, but not so for the very short-range (one-quarter-ahead).

Phelps, Ruth H. and Shanteau, James (1978), "Livestock judges: How much information can an expert use?" *Organizational Behavior and Human Performance*, vol. 21, pp. 209-219.

This study of the judging of female breeding pigs showed that experts were capable of using about ten pieces of information when these variables are not correlated with one another.

Pristo, L.J. (1979), "The prediction of graduate school success by the canonical correlation," *Educational and Psychological Measurement*, vol. 39, pp. 929-933.

The correlation between actual and predicted success in the validation sample was *negative*.

Pritchard, David A. (1980), "Apologia for clinical/configural decision making," *American Psychologist*, vol. 35 (July), pp. 676-678.

See also Dawes (1980) and Remus (1980) on pages 678-680 of the same issue.

Rausser, Gordon C. and Oliveira, Ronald A. (1976), "An econometric analysis of wilderness area use," *Journal of the American Statistical Association*, vol. 71, pp. 276-285.

This study examined *ex post* short term forecasts using alternative criteria for accuracy. The econometric model was more accurate than Box-Jenkins, and a combined forecast was even more accurate.

Read, Stephen J. (1983), "Once is enough: Causal reasoning from a single instance," *Journal of Personality and Social Psychology*, vol. 45, pp. 323-334.

A single concrete example had a significant impact on people's predictions in this experiment. The results suggest that predictions by political decision makers may be unduly influenced by single historical events rather than by generalizations from a broad range of situations. The tendency to rely heavily on a single event was higher for more complex situations.

Reibstein, David J. and Traver, Phillis A. (1982), "Factors affecting coupon redemption rates," *Journal of Marketing*, vol. 46 (Fall), pp. 102-113.

Good illustration of the use of prior research to develop a forecasting model. The Logit model, which transformed the dependent variable from Y to $\ln \frac{Y}{1-Y}$, has often been proposed for situations where Y varies between 0 and 1. Thus, it looked relevant for this study, where the task was to predict the percentage of coupons redeemed. Interestingly, however, the Logit did not provide a better fit than a regression against Y , nor did it do better on the validation sample. (The latter comparison was based on personal communication with Reibstein.)

Reilly, Richard R. and Chao, Georgia T. (1982), "Validity and fairness of some alternative employee selection procedures," *Personnel Psychology*, vol. 35, pp. 1-62.

This systematic and impressive review of the literature includes 41 unpublished papers and 107 published papers. It is well written; but, given the immense material that is covered, be well-rested before you attempt to read it. The authors examined alternatives to standardized tests for predicting which job applicants will be successful. The alternatives were biographical data, peer evaluation, interviews, self-assessments, reference checks, academic achievement, expert judgment, and projective techniques. (Which ones would you predict to be most valid?) Of these methods, only the biographical data and peer evaluations had validities comparable to those achieved by using standardized tests; the other methods had little validity and some involved high costs. However three methods appeared promising, although the evidence was limited. One method is the "miniaturized training test," applicable for people without prior experience. The applicant is rated on ability to learn key components of the job in a short training exercise. Second is a structured "situational interview," where job candidates are asked how they would behave in given situations. Third, in "unassembled examinations," job candidates use structured guidelines to assemble a portfolio of verifiable past accomplishments relevant to the job at hand. As implied by the title, Reilly and Chao also examine the extent to which each method avoids prejudice.

Reinmuth, James E. and Geurts, Michael D. (1979), "Multideterministic approach to forecasting," in S. Makridakis and S.C. Wheelwright (eds.), *Forecasting*. New York: North Holland.

Large reductions in error were achieved by combining extrapolative forecasts in this *small sample* study involving retail sales in Salt Lake City.

Remus, William E. (1980), "Measure of fit for unit rules," *American Psychologist*, vol. 35, pp. 678-680. See also Pritchard (1980).

In trying to replicate Dawes' (1971) study of graduate admissions, Remus claims different results. He obtained 58% correct predictions for unit rules and 74% for regression.

Remus, William E. and Jenicke, Lawrence O. (1978), "Unit and random linear models in decision making," *Multivariate Behavioral Research*, vol. 13 (April), 215-221.

Examined a simulated production scheduling problem and found that "unit rules" and "random" coefficients led to higher costs than these obtained using judgmental decisions. For a related paper, see Moskowitz, et al. (1982).

Ricketts, Donald E. and Barrett, Michael J. (1973), "Corporate operating income forecasting ability," *Financial Management*, vol. 2 (Summer), pp. 53-62.

Extrapolation of components followed by aggregation was no more accurate than extrapolation of aggregate corporate income. (In fact, it was slightly worse, but the difference was not significant.)

Riggs, Walter E. (1983) "The Delphi technique: An experimental evaluation," *Technological Forecasting and Social Change*, vol. 23, pp. 89-94.

Forecasts for two college football games were obtained from eight traditional groups and eight Delphi groups. Each group had four or five students. The forecasts were made four weeks before the games were played. One game was an intense rivalry well known to the students, while the other was less well known:

	Mean Absolute Error For:	
	Rivalry	Other Game
Delphi (Round 2)	2.2	13.9
Traditional	5.6	17.0

Delphi was significantly more accurate ($p < .05$).

Robertson, Ivan and Downs, Sylvia (1979), "Learning and the prediction of performance: Development of trainability testing in the United Kingdom," *Journal of Applied Psychology*, vol. 64, pp. 42-50.

Work sample and trainability tests were found to be superior to written tests for predicting success at semi-skilled manual labor jobs.

Robertson, Ivan T. and Kandola, R.S. (1982), "Work sample tests: Validity, adverse impact and applicant reaction," *Journal of Occupational Psychology*, vol. 55, pp. 171-183.

This paper examined the validity of psychomotor work samples, job-related information, situational decision making, and group discussion as predictors of job performance, job progress, and training. The conclusions, based on over 60 empirical studies, showed that each of the four methods was of roughly equal validity across all criteria. For the specific criterion of job performance, the psychomotor work samples had the highest predictive validity, followed by group discussion, situational decision making, then job related information tests. When compared with traditional (pencil and paper) psychological tests,

work sample tests appeared to have a less adverse impact (i.e., they are not so biased against minorities). Furthermore, work sample tests allow *applicants* to make better predictions of how they would perform a given job. Finally, applicants preferred work samples as a predictive and selection technique.

Rohrbaugh, John (1979), "Improving the quality of group judgment: Social judgment analysis and the Delphi Technique," *Organizational Behavior and Human Performance*, vol. 24, pp. 73-92.

This experiment pitted groups that met face-to-face (and discussed the logic of their judgment, as well as the judgment itself) against Delphi groups. Grade point averages of prospective freshmen were predicted by the subjects (172 psychology students). The face-to-face meeting, with some structure, did no better than the Delphi procedure of simply averaging the responses.

Roose, Jack E. and Doherty, Michael E. (1976), "Judgment theory applied to the selection of life insurance salesmen," *Organizational Behavior and Human Performance*, vol. 16, pp. 231-249.

Sixteen agency managers made predictions on the potential success for 200 salespeople who had been hired. A validation sample of another 160 salespeople was used.

Conclusions: (1) insight was poor and not related to the managers' experience; (2) commensurate information was weighted too heavily; (3) bootstrapping yielded a small gain for the average judge, but was of little value for the consensus judge; and (4) unit weights did better than bootstrapping.

Rosenberg, Richard D. and Rosenstein, Eliezer (1980), "Participation and productivity: An empirical study," *Industrial and Labor Relations Review*, vol. 33, pp. 355-367.

An analysis of records from 262 meetings between workers and managers from 1969 to 1975 showed that participation led to increases in productivity.

Rozeboom, W.W. (1978), "Estimation of cross-validated multiple correlation: A clarification," *Psychological Bulletin*, vol. 85, pp. 1348-1851.

Rothe, James T. (1978), "Effectiveness of sales forecasting methods," *Industrial Marketing Management*, vol. 7, pp. 114-118.

Interviewees from 52 firms were asked about forecasting for production, finance, marketing, purchasing, inventory, and personnel. Opinion techniques were the most popular, as 96% of respondents reported using them. Exponential smoothing was used by 14%, and 6% used regression. About half of the firms kept historical records on accuracy. Only one firm had examined the cost due to inaccurate forecasts. None of the respondents knew how much was being spent on forecasting in their firm. This study addressed many useful questions. Read with care, however, as the conclusions sometimes go beyond the evidence.

Rudelius, William, Dickson, G.W. and Hartley, S.W. (1982), "The little model that couldn't: How a decision support system for retail buyers found limbo," *Systems, Objectives, Solutions*, vol. 2, pp. 115-124.

Interesting description of a high quality solution that, the authors say, was designed for use by retail buyers without any concern for the implementation *process*. Although it succeeded in meeting the buyers' needs, it was not actually implemented by the firm. Later, the authors talked to newly hired executives in the firm. The authors were asked to design a new model, and the requirements were the same as for the model their colleagues had discarded earlier. The moral, say the authors, is to begin by paying explicit attention to the implementation process.

Ruland, William (1980), "On the choice of simple extrapolative model forecasts of annual earnings," *Financial Management*, vol. 9 (Summer), pp. 30-37.

This study compared the forecast accuracy of eight extrapolative models. Simple models were just as accurate.

Rush, Howard and Page, William (1979), "Long-term metals forecasting: The track record: 1910-1964," *Futures*, vol. 11, pp. 321-337.

The authors examined 372 forecasts and coded them. (Coding was not easy because some original sources did not provide sufficient information.) Judgmental methods were commonly used up to 1939 (about 50 % of the published forecasts) and even more so after 1939 (65 %). Explicit references to uncertainty were found in 22 % of the forecasts published before 1939, but in only 8 % afterwards.

Schmitt, Neal (1978), "Comparison of subjective and objective weighting strategies in changing task situations," *Organizational Behavior and Human Performance*, vol. 21, pp. 171-188.

A partial replication of Cook and Stewart (1975). Subjects (112 students) were asked to make predictions of academic success based on three or four variables (contrived data). After practicing on 20 "applicants," subjects made predictions for 30 new "applicants." Interesting results: (1) subjects performed better when they did *not* receive feedback on whether the prediction was right or wrong, (2) three subjective weighting schemes were tried and found to be of equal accuracy, (3) regression against predicted outcomes (judgmental bootstrapping) was more accurate than the direct bootstrapping; and (4) equal weights provided good forecasts.

Schnaars, Steven P. (1984), "Situational factors affecting forecast accuracy," *Journal of Marketing Research*, vol. 21, pp. 290-297.

Schnaars, Steven P. and Bavuso, R. Joseph (1985), "A comparison of extrapolation models on very short-term forecasts," *Journal of Business Research*, (in press).

Schnee, Jerome E. (1977), "Predicting the unpredictable: The impact of meteorological satellites on weather forecasting," *Technological Forecasting and Social Change*, vol. 10, pp. 299-307.

Schott, Kerry (1978), "The relations between industrial research and development and factor demands," *Economic Journal*, vol. 88, pp. 85-106.

Ex ante forecasts were more accurate than the *ex post* forecasts from some econometric models.

Schreuder, Hein and Klaassen, Jan (1984), "Confidential revenue and profit forecasts by management and financial analysts: Evidence from the Netherlands," *The Accounting Review*, vol. 59, pp. 64-77.

This study extends the research on the relative accuracy of management and analysts in forecasting next year's annual earnings by examining *confidential* forecasts by a sample of Dutch firms for 1980. Firms were asked to file these confidential forecasts with a notary, with many safeguards provided against misuse. The authors concluded that management was not more accurate. This finding, however, is based on a small sample (38 companies for one year). Furthermore, the *direction* of the results favored management (MAPE of 102.9 vs. 139.4 respectively for management and analysts), a result that seems consistent in relative terms with my meta-analysis of previous studies (Armstrong, 1983b). Schreuder and Klassen also examined sales forecasts. Again the management errors were a bit smaller than those of analysts (MAPEs of 6.7 vs. 7.7); these results were not significantly different. As might be expected, when the sales forecast was too high (low), the profit forecast tended to be too high (low), but there were many exceptions (38 %). Management and analysts estimated 50 % and 100 % (!) confidence intervals. Consistent with prior research, these confidence intervals were too narrow: 56 % of the revenue and 72 % of the profit forecasts fell outside the 50 % confidence intervals; 35 % of the revenue forecasts and 89% of the profit forecasts fell outside the 100 % confidence intervals.

Sewall, Murphy A. (1981), "Relative information contributions of consumer purchase intentions and management judgment as explanators of sales," *Journal of Marketing Research*, vol. 18, pp. 249-253.

The study is better than the title. It examined U.S. mail order sales from a 1979 catalog for 44 women's blouses priced from \$5 to \$20. Predictions were made by a buyer for the mail order house, the normal procedure used in deciding on initial orders. Consumer intentions were then obtained from 600 women shoppers in shopping malls. The intentions were obtained with a 5-point rating scale in response to a set of photographs. Four different methods were considered for summarizing the ratings for each blouse (median, Thurstone, mean, and "fraction in top two categories"). Here are the questions: (1) Which provides the best predictions, the expert (buyer) or the intentions (shoppers) survey? (2) Does it matter how the rating scale is summarized in the intentions survey? The answer to (1) was that each provided useful information for prediction, and the predictive ability of the experts was about equal to that of the intentions survey. Sewall (personal communication) suggests that the combined use of expert and intentions information will improve predictions. He said that it allowed for a 15 % reduction in inventory order ing errors in this case. For (2), the method used to summarize the rating scale did not affect the accuracy of the predictions.

Sherman, Steven J. (1980), "On the self-erasing nature of errors of prediction," *Journal of Personality and Social Psychology*, vol. 39, pp. 211-221.

This is an interesting and important study relevant to planning, scenarios, and implementation. It is based on the self-fulfilling prophecy. If people are asked how they will respond in a given situation, they tend to cast themselves in a responsible and

favorable manner. Then, if presented with that situation or a similar situation, they tend to live up to their predictions.

Shocker, Allan and Srinivasan, V. (1979), "Multiattribute approaches for product evaluation and generation: A critical review," *Journal of Marketing Research*, vol. 16, pp. 159-180.

Good review of the research on methods to predict preferences for products in the concept phase. See especially their summary Table 1. They cite studies that examined estimation procedures other than ordinary regression analysis.

Silhan, Peter A. (1983), "The effects of segmenting quarterly sales and margins on extrapolative forecasts of conglomerate earnings: Extension and replication," *Journal of Accounting Research*, vol. 21, pp. 341-347.

This study used quarterly data on income for 60 firms with one-quarter and one-year ahead *ex ante* forecasts for 1976-1978. Supports Kinney (1971) and Collins (1976). An excellent study.

Smith, David E. (1974), "Adaptive response for exponential smoothing: Comparative system analysis," *Operational Research Quarterly*, vol. 25, pp. 421-435.

Smith, Gary and Brainard, William (1976), "The value of a *priori* information in estimating a financial model," *Journal of Finance*, vol. 31, 1299-1322.

Examines *ex post* forecasts over an eight-quarter forecast horizon using RMSE as the criterion for accuracy. Forecasts were made for six variables for banks and four for savings and loan institutions. Extrapolation models were more accurate than econometric models for short-run forecasts, but their performance deteriorated rapidly and seemed worse for the eight-quarter-ahead forecasts. Models based solely on prior information were generally more accurate than those estimated by standard regression analysis. The combination of prior information and data, done in a rigorous manner here, performed well overall. The paper addresses many important issues, but it is difficult to read.

Smith, M.C. (1976), "A comparison of the value of trainability assessments and other tests for predicting the practical performance of dental students," *International Review of Applied Psychology*, vol. 25, pp. 125-130.

Good description of trainability tests (work sample used to see how long it takes an applicant to learn). The key rules for such a test are that it be (1) based on crucial elements of the job, (2) use skill and knowledge that can be imparted only during a short learning period, and (3) be sufficiently complex to allow for a range of observable errors to be made by the applicants. Presents evidence on validity of this method.

Smyth, David J. (1983), "Short-run macroeconomic forecasting: The OECD performance," *Journal of Forecasting*, vol. 2, pp. 37-49.

Econometric forecasts for Canada, France, West Germany, Italy, Japan, the United Kingdom, and the United States are published on a regular basis in the OECD's *Economic Outlook*. This paper analyzes the accuracy of the OECD annual forecasts. The forecasts were compared with those generated by a naive model using mean-absolute error, the

root-mean-square error, the median-absolute error, and Theil's inequality coefficient. The OECD forecasts of real GNP changes were significantly superior to those generated by a random walk process; however, the OECD price changes and current balance of payments forecasts were not significantly more accurate than those obtained from the naive model. The OECD's forecasting performance has neither improved nor deteriorated over time.

Sparkes, John R. and McHugh, A.K. (1984), "Awareness and use of forecasting techniques in British industry," *Journal of Forecasting*, vol. 3, pp. 37-42.

Received 76 replies (25 %) from a survey mailed to 300 British manufacturing firms. These firms seemed less familiar with objective methods than did the U.S. firms surveyed by Mentzer and Cox (1984).

Stewart, Thomas R. and Glantz, Michael H. (1985), "Expert judgment and climate forecasting: A methodological critique of 'Climate Change to the Year 2000,'" *Climatic Change*, vol. 7, No. 1.

Stewart and Glantz use the existing research on judgmental forecasting to evaluate a widely distributed expert-opinion study by the U.S. National Defense University. This study concluded that climate changes would be small, but, as noted by Stewart and Glantz, the study was not well-designed in light of the research findings on judgmental forecasting.

Teigen, Karl Halvor (1988), "Studies in subjective probability III: The unimportance of alternatives," *Scandinavian Journal of Psychology*, vol. 24, pp. 97-105.

Timmers, Han and Wagenaar, Willem A. (1977), "Inverse statistics and misperception of exponential growth," *Perception and Psychophysics*, vol. 21, pp. 558-562.

Judges tend to greatly underestimate exponential growth.

Traugott, Michael W. and Tucker, Clyde (1984), "Strategies for predicting whether a citizen will vote and estimation of electoral outcomes," *Public Opinion Quarterly*, vol. 48, pp. 380-348.

A segmentation approach (eight segments) and a regression (using the logit function) produced almost identical forecasts as to who would vote in the 1980 U.S. presidential election.

Tversky, Amos and Kahneman, Daniel (1981), "The framing of decisions and the psychology of choice," *Science*, vol. 211, pp. 458-458.

Seemingly inconsequential changes in the formulation of choice problems can cause major shifts in the preferences of people.

Tversky, Amos and Kahneman, Daniel (1983), "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychological Review*, vol. 90, pp. 293-315.

An interesting set of experiments on the conjunction fallacy (*A* and *B* seems more likely than *B* alone, because *A* seems to be a plausible reason). Incidentally, I was a subject in one of these studies and I would not be surprised to find that I was guilty of this fallacy.

Wagenaar, Willem A. (1978), "Intuitive prediction of growth," in Dietrich F. Burkhardt and William H. Ittelson (Eds.), *Environmental Assessment of Socioeconomic Systems*. New York: Plenum (1978).

This study shows how frequent reference to the latest data led to poorer forecasts in cases of exponential growth. People involved closely with exponential growth would be less likely to be able to predict change. Subjects seem to look at *differences* rather than *ratios* in their subjective forecasts. Mathematical training did not improve accuracy. The following steps were helpful: to (1) observe the process *less* frequently, and (2) use an inverse representation of growth (e.g., instead of people per square mile, try to predict square miles per person). For this inverse representation, the large differences occur early, rather than late, in the sequence.

Wagenaar, Willem A. and Sagaria, Sabato D. (1975), "Misperception of exponential growth," *Perception and Psychophysics*, vol. 18, pp. 416-422.

Subjects were presented with exponential growth series and were told that "nothing will stop the growth." Their intuitive predictions were highly conservative. Surprisingly, it did not help when the data were presented to the subjects in graphic form.

Wagenaar, Willem A., Schreuder, R. and Van der Heijden, A.H.C. (1985), "Do TV-pictures help people to remember the weather forecast?" *Ergonomics*.

Wagenaar, Willem A. and Timmers, Han (1978), "Extrapolation of exponential time series is not enhanced by having more data points," *Perception and Psychophysics*, vol. 24, pp. 182-184.

Subjects were provided with 3, 5 and 7 observations in an exponentially growing series. (All subjects received the same first and last observations.) Those who received more observations made less accurate forecasts.

Wagenaar, Willem A. and Timmers Han (1979), "The pond-and-duckweed problem: Three experiments in the misperception of exponential growth," *Acta Psychologica*, vol. 43, pp. 289-251.

This study used a computer display screen to display the growth process.

Wagenaar, Willem A. and Visser, Jenny G. (1979), "The weather forecast under the weather," *Ergonomics*, vol. 22, pp. 909-917.

Their experiment provided useful guidelines on how to present forecasts effectively: (1) group the forecast information into meaningful blocks, and (2) present current status first (i.e., "what is the weather now?" This was seldom included in the weather reports they analyzed), and (3) shorten the message.

Warshaw, Paul R. (1980), "Predicting purchase and other behaviors from general and contextually specific intentions," *Journal of Marketing Research*, vol. 17, pp. 26-33.

Werner, Paul D., Rose, Terrence L. and Yesavage, Jerome A. (1983), "Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness," *Journal of Consulting and Clinical Psychology*, vol. 51, pp. 815-825.

The mass media, movies, and courts assume that it is possible to predict who will be violent. But beyond the obvious factor that those who have been violent in the past are more likely to be violent in the future, predictive ability is low, as had been shown previously. This study adds further evidence. They asked 30 experts (15 psychologists and 15 psychiatrists) to make predictions about physical violence occurring in the first week of hospitalization for 40 newly admitted mental patients. The judges received information on 19 variables about each patient, but they did not meet the patient. The findings: (1) individual judges had modest *reliability* ($r = .42$), and reliability was greatly increased by using a composite of 15 judges ($r = .93$), (2) experience, including experience in a similar situation, *did not* yield better predictions, (3) ratings by individual judges did not have *significant predictive validity* (mean $r = .12$, with only 2 of 30 judges doing better than chance), and (4) the composite of 80 judges tended to be more accurate, but the gain was unexpectedly small ($r = .17$ for composite vs. the mean r of $.12$). Furthermore, it was not statistically significant (vs. chance). A step-wise regression of the actual violence versus the original variables revealed a different set of factors. Possibly the judges were using the wrong variables? This paper provides an interesting application of the Brunswick Lens Model to the problem.

Whybark, D. Clay (1972), "A comparison of adaptive forecasting techniques," *Logistics and Transportation Review*, vol. 8, No. 3, pp. 13-26.

Although the differences were small, adaptive parameters apparently led to improvements. This conclusion was later challenged in a re-analysis of the data by Ekern (1981).

Wilton, Peter C. and Pessemier, Edgar A. (1981), "Forecasting the ultimate acceptance of an innovation: The effects of information," *Journal of Consumer Research*, vol. 8, pp. 162-171.

This paper used a multivariate probit model to predict the stated choices for a subcompact electric vehicle for 196 individuals in a hold-out sample. The probit model is an extension to the ordinary regression model that overcomes the problems of heteroscedasticity and negative forecasts when using dummy variables for the dependent variable (such as "1 = Buy" and "0 = Do Not Buy"). The results were not impressive when compared with chance. Actual market behavior was also used as a criterion and here the probit model predictions appeared to be of some value.

Wimsatt, Genevieve B. and Woodward, John T. (1970), "Revised estimates of new plant and equipment expenditures in the United States, 1947-1969: Part II," *Survey of Current Business*, vol. 50, pp. 19-39.

For business as a whole, annual expectations correctly predicted the direction of change in investment expenditures in 20 of 21 years (including four years when there was a decline: 1949, 1954, 1958, and 1960). It missed 1950 (big change due to the Korean peace action). The accuracy of forecasts of quarterly changes was also impressive.

Wind, Yoram, Mahajan, Vijay and Cardozo, Richard N., (Eds.) (1981) *New Product Forecasting*. Lexington, Mass.: Lexington Books.

A comprehensive set of readings (12 previously published and 10 prepared for this book) on approaches to new product forecasting. It is an important (and profitable) field and this

book presents the state of the art. As pointed out by the editors, validation studies are in dire need in this area. The readings cover forecasting at the various stages of new product development: concept testing, pretest-market, test-market, and early sales. Melvyn Hirst provides an extensive review of this book, along with additional references, in the *Journal of Forecasting*, vol. 2 (1983), 85-87.

Wood, Gordon (1978), "The knew-it-all-along effect," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, pp. 345-353.

Winkler, Robert L. and Makridakis, Spyros (1983), "The combination of forecasts," *Journal of the Royal Statistical Society: Series A*, vol. 146, part 2, pp. 150-157.

Examines weighting schemes for a large number of time series, many different methods, and several time horizons. See also Makridakis and Winkler (1983).

Wood, Gordon (1978), "The knew-it-all-along effect," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, pp. 345-353.

This experiment shows that once the outcome is known, subjects have difficulty remembering their prior beliefs.

Wright, George and Whalley, Peter (1983) "The supra-additivity of subjective probability" in B.P. Stigum and F. Wenstop (Eds.), *Foundations of Utility and Risk Theory with Applications*. London: D.Reidel.

When subjects were asked to estimate the probability of two mutually exclusive and exhaustive events, their probabilities would generally sum to 1. As the number of possibilities were increased, the sum of the probabilities increased:

Possible Outcomes	Total Probabilities
5	1.70
6	1.65
6	1.70
7	2.13
16	3.04

Yetton, Philip and Bottger, Preston (1983), "The relationships among group size, member ability, social decision schemes, and performance," *Organizational Behavior and Human Performance*, vol. 32, pp. 145-147.

This study used the NASA Lost-on-the-Moon exercise. For nominal groups, accuracy improved as group size increased to five people. For interacting groups, accuracy improved up to four people.

Zarnowitz, Victor (1979), "An analysis of annual and multi-period quarterly forecasts of aggregate income, output, and the price level," *Journal of Business*, vol. 52, pp. 1-33.

This paper examines errors in forecasting GNP from 1959-1976 (annually) and 1970-1975 (quarterly) using forecast horizons of 1 to 8 quarters.

Zarnowitz, Victor (1984), "The accuracy of individual and group forecasts from business outlook surveys," *Journal of Forecasting*, vol. 3, pp. 11-26.

The group mean forecast was more accurate than the typical group member for six economic variables over different forecast horizons.

Zukier, Henry (1982), "The dilution effect: The role of the correlation and the dispersion of predictor variables in the use of diagnostic information," *Journal of Personality and Social Psychology*, vol. 48, pp. 1163-1174.

This experiment asks subjects to predict grade point averages of students. When they were given irrelevant information, along with the relevant information, they became more conservative.