October 1978

Forecasting with Econometric Methods: Folklore Versus Fact

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Postprint version. Published in *Journal of Business*, Volume 51, Issue 4, October 1978, pages 549-564.
Publisher URL: http://www.jstor.org/browse/00219398?config=jstor

Forecasting with Econometric Methods: Folklore Versus Fact

Abstract

Evidence from social psychology suggests that econometricians will avoid evidence that disconfirms their beliefs. Two beliefs of econometricians were examined: (1) Econometric methods provide more accurate short-term forecasts than do other methods; and (2) more complex econometric methods yield more accurate forecasts. A survey of 21 experts in econometrics found that 95% agreed with the first statement and 72% agreed with the second. A review of the published empirical evidence yielded little support for either of the two statements in the 41 studies. The method of multiple hypotheses was suggested as a research strategy that will lead to more effective use of disconfirming evidence. Although this strategy was suggested in 1890, it has only recently been used by econometricians.

Comments

Postprint version. Published in *Journal of Business*, Volume 51, Issue 4, October 1978, pages 549-564.
Publisher URL: http://www.jstor.org/browse/00219398?config=jstor

J. Scott Armstrong*
University of Pennsylvania

Forecasting with Econometric Methods: Folklore versus Fact

Evidence from social psychology suggests that econometricians will avoid evidence that disconfirms their beliefs. Two beliefs of econometricians were examined: (1) Econometric methods provide more accurate short-term forecasts than do other methods; and (2) more complex econometric methods yield more accurate forecasts. A survey of 21 experts in econometrics found that 95% agreed with the first statement and 72% agreed with the second. A review of the published empirical evidence yielded little support for either of the two statements in the 41 studies. The method of multiple hypotheses was suggested as a research strategy that will lead to more effective use of disconfirming evidence. Although this strategy was suggested in 1890, it has only recently been used by econometricians.

Introduction

This paper is concerned with the use of econometric methods for forecasting in the social sciences. Although this is not the only use of econometric methods, it is one of the ways they are used; it is also the use that can most easily be validated. The paper examines only the predictive validity of econometric models. The importance of predictive validity has long been recognized by econometricians. Christ (1951) stated, "The ultimate test of an econometric model . . . comes with checking its predictions."

"Econometric methods" are defined in this paper as quantitative approaches that attempt to use causal relationships in forecasting. In particular, they refer to models based on regression analysis. This definition conforms to common usage of the term "econometric methods." "Folklore" is used here to reflect what econometricians believe, as judged by what they do. "Fact" is based upon published empirical studies.

The first part of this paper draws upon evidence from social psychology to explain why folklore persists. Most of the evidence is based upon the behavior of people in general. How-

* The Stockholm School of Economics provided time and money for this project. Permission was granted by John Wiley & Sons to include sections from *Long-Range Forecasting: From Crystal Ball to Computer* (New York: Wiley-Interscience, 1978).

(*Journal of Business,* 1978, vol. 51, no. 4)

0021-9398/78/5104-0006$01.30

ever, there is evidence to suggest that scientists act as other people when testing their favored hypotheses.

Two examples of the discrepancy between folklore and fact are provided in the second part of the paper. These are only two of a number of possible examples, but they deal with two important questions. First, do econometric methods provide the most accurate way to obtain short-range forecasts? Second, do complex econometric methods provide more accurate forecasts than simple econometric methods?

The third part of the paper describes the method of multiple hypotheses. This method should help to overcome folklore.

The Persistence of Folklore

Folklore persists because people who hold viewpoints on an issue tend to perceive the world so as to reinforce what they already believe; they look for "confirming" evidence and avoid "disconfirming" evidence. There is much literature on this phenomenon, commonly known as "selective perception."

The tendency for intelligent adults to avoid disconfirming evidence was demonstrated by Wason (1960, 1968). He provided three numbers (2, 4, 6) to subjects, and they were asked to determine what rule had been used to generate the three numbers. In order to gain additional information, the subjects were encouraged to generate other series of three numbers. The experimenter provided feedback on whether or not each new series was in agreement with the rule. What happened? The typical subject would think of a rule and then generate series that were consistent with that rule. It was unusual for a subject to try a series that was inconsistent with his own rule. Subjects who were told that their rules were incorrect were allowed to generate additional series. The majority of these subjects maintained the same rule that they had previously but stated it in different terms. (It is like magic; it will work if one can pronounce it correctly!)

In cases where disconfirming evidence is thrust upon people, they tend to remember incorrectly. Fischhoff and Beyth (1975), for example, found that subjects tended to remember their predictions differently if the outcome was in conflict with their prediction.

Wason's studies dealt with situations in which the person had no stake and no prior emotional attachment. When one has invested effort in supporting a particular viewpoint, the tendency to avoid disconfirming evidence would be expected to be stronger. The reward system in science encourages researchers to devote their energies to one viewpoint. The scientist gains recognition by being an advocate of a particular approach or theory. In such a case, the scientist can be expected to avoid disconfirming evidence.

Studies of scientists indicate that they are biased in favor of their own hypothesis. They interpret evidence so that it conforms to their beliefs. For example, in Rosenthal and Fode (1963), experimenters were provided with two equivalent samples of rats, but they were told that one sample was gifted and the other was disadvantaged. In the subsequent "scientific tests," the gifted rats learned tasks more quickly than did the disadvantaged rats.

The above studies dealt with individuals rather than with groups. What happens when group pressures are involved—for example, when someone submits an article to be evaluated by his peers in the "marketplace of ideas?" What happens when learned societies, such as the Econometric Society, are formed to promote the advancement of the science? As the group pressures become stronger, one would expect stronger efforts to avoid evidence that disconfirms the group's opinions. Substantial literature shows how group judgment distorts reality. The study by Asch (1965) showed that most subjects would agree with the group that a given line B was longer than another line A, even though the reverse was obviously true.

In fact, the peer review process was studied in an experiment by Mahoney (1977). A paper was sent to 75 reviewers. Some reviewers received the paper along with results that were supportive of the commonly accepted hypothesis in this group. Other reviewers received a copy of the identical study except that the results were reversed so that they disconfirmed the prevailing hypothesis. Reviewers with confirming results thought the study was relevant and methodologically sound. Reviewers with disconfirming results thought the study was not relevant and that the methodology was poor. The confirming paper was recommended for publication much more frequently.

The studies cited above provide only a portion of the evidence. Other relevant studies include Pruitt (1961), Geller and Pitz (1968), Chapman and Chapman (1969), Rosenthal and Rosnow (1969), and Greenwald (1975). This evidence implies that scientists avoid disconfirming evidence. This tendency is stronger when the position is adopted by a group.

It is not surprising then, that great innovations in science have often met with resistance. (Barber [1961] describes some important historical examples.) There is little reason to expect that "modern science" is different. For illustration, one might examine the treatment of Immanuel Velikovsky, a case that is being followed closely by sociologists (de Grazia 1966). This treatment was not the result of a lack of interest or a lack of time; rather it was an active attempt to suppress Velikovsky's theories and to discredit him.

Social scientists are expected to be more prone to group opinion than are physical scientists. Thus, they would experience serious difficulties in adopting new findings. Are econometricians also resistant to innovations? In a critique of what is being done by econometricians, Bassie (1972) implies that they are. He claims that econometricians display much conformity to their preconceptions.

Two examples from econometrics are examined below. These examples were selected because they represent an important part in the life of an econometrician—and also because there seem to be discrepancies between folklore and fact. (Additional examples can be found in Armstrong [1978*a*].)

Short-Range Forecasting

Most textbooks on econometrics discuss short-range forecasting. Although seldom stated, the implication is that econometric methods provide more accurate short-range forecasts than other methods. Brown (1970, p. 441) asserted that econometric models were originally designed for short-range forecasting. Kosobud (1970, pp. 260–61), in a paper on short-range forecasting, referred to " . . . the growing body of evidence on the predictive value of econometric models." In a review of a book on short-range economy-wide forecasting, Worswick (1974, p. 118) said that "the value of econometric models in short-term forecasting is now fairly generally recognized." Various econometric services sell short-range forecasts, and one of their claims is improved accuracy. The press publishes short-range forecasts from well-known econometric models with the implication that these models will provide accurate forecasts.

Survey of Econometricians

In order to go beyond the indirect evidence cited in the preceding paragraph, a questionnaire was mailed to experts in econometrics in late 1975. The survey was based on a convenience sample. Of 56 questionnaires that were sent out, 21 were completed. An additional eight were returned incomplete by respondents who said they lacked the necessary expertise. Thus, replies were received from over 40% of the experts. The respondents were from some of the leading schools in econometrics—for example, M.I.T., Harvard, Wharton, Michigan State—and from well-known organizations that sell econometric forecasts. Many of the respondents are recognized as leading econometricians. (A listing of the sample was provided to the editors of the *Journal of Business*.)

The questionnaire asked, "Do econometric methods generally provide more accurate or less accurate forecasts than can be obtained from competitive methods for short-term forecasting in the social sciences? Or is there no difference in accuracy?'' A set of definitions was also provided.[1]

The results of the survey, presented in table 1, were that 95% of the experts agreed that predictions from econometric models are more accurate.

Respondents were asked how much confidence they had in their opinion on accuracy. Confidence was rated on a scale from 1 (''no confidence'') to 5 (''extremely confident''). (If the question was not clear to respondents, they were instructed to report a low level of confidence.) The average response was about 4. No one rated confidence lower than 3.0. Those who responded with ''significantly more accurate'' had the highest confidence level.

Another question asked how the respondent would rate himself '' . . . as an expert on applied econometrics.'' Eight respondents rated themselves as ''very much of an expert,'' six as ''fairly expert,'' four as ''somewhat of an expert,'' and two felt that they were ''not much of an expert'' (there was one nonresponse on this question). Those who rated themselves as more expert felt that econometric methods were more accurate: Five of the eight who rated themselves as ''very much of an expert'' felt that econometric methods were significantly more accurate, a rating that was significantly higher than the ratings by the other respondents ($P < .05$ using the Fisher Exact Test).

TABLE 1 **Survey of Experts on Accuracy of Short-Range Econometric Predictions** ($N = 21$)

Econometric Predictions Rated	Percentage
Significantly more accurate	33
Somewhat more accurate	62
No difference (or undecided)	0
Somewhat less accurate	5
Significantly less accurate	0

1. These definitions were as follows: ''(a) 'Econometric methods' include all methods which forecast by explicitly measuring relationships between the dependent variable and some causal variables. (b) 'Competitive methods' would include such things as judgment by one or more 'experts' or extrapolation of the variable of interest (e.g., by relating the variable to 'time' such as in autoregressive schemes). (c) By 'do,' we mean that comparisons should be made between methods which appear to follow the best practices which are available at the current time. In other words, the methods should each be applied in a competent manner. (d) 'Short-term' refers to time periods during which changes are relatively small. Thus, for forecasts of the economy, changes from year to year are rather small, almost always less than 10%. For some situations, however, one-year changes may be substantial. (e) 'Forecasts' refer to unconditional or 'ex ante' forecasts only. That is, none of the methods shall use any data drawn from the situation which is being forecast. Thus, for time series, only data prior to time t could be used in making the forecasts. (f) The 'social sciences' would include economics, psychology, sociology, management, etc. In short, any area where the behavior of people is involved.''

In general, the survey supported the anecdotal evidence. Experts are confident that short-range econometric predictions are more accurate than predictions from other methods.

Empirical Evidence

Turning to "fact," an examination was made of all published empirical studies that I could find in the social sciences. This survey was conducted primarily by examining references from key articles and by searching through journals. Respondents to the expert survey were asked to cite evidence, but this yielded few replies.[2] Finally, early drafts of this paper were presented at conferences and were circulated for comments over a period of 4 years; this approach did lead to additional studies. The studies are summarized below.

Christ (1951, 1956) provided disconfirming evidence on the accuracy of econometric predictions. In the 1951 study, econometric forecasts were better than "no change" forecasts on six occasions and worse on four. These were conditional or ex post forecasts; nevertheless, the results were not encouraging. The reaction to these findings was similar to previously mentioned occasions when disconfirming evidence was thrust upon scientists. Two of the discussants for Christ's paper were Lawrence Klein and Milton Friedman. Klein, whose model had been examined by Christ (1951, p. 121), stated that ". . . a competent forecaster would have used an econometric model . . . far differently and more efficiently than Christ used his model." Friedman, however, was receptive. He said (Christ [1951], p. 112) that additional evidence would tend to strengthen Christ's conclusion and that " . . . the construction of additional models along the same general lines [as Klein's model] will, in due time, be judged failures."

Additional evidence on the predictive validity of econometric methods since Christ's papers is described here. Most of these studies are recent. Some are only of a suggestive nature because they compare ex post predictions of econometric models with ex ante predictions from alternative methods. Comparisons between extrapolations and ex post econometric forecasts were made by Kosobud (1970), Cooper (1972), Nelson (1972), Elliott (1973), Granger and Newbold (1974), Narasimham, Castellino, and Singpurwalla (1974), Levenbach, Cleary, and Fryk (1974), and Ibrahim and Otsuki (1976).[3] Extrapolations provided better forecasts than the econometric methods in all studies

2. Two of the respondents who rated themselves highly as experts and who had the highest confidence in their ratings stated that they were not aware of any empirical evidence on this issue.

3. One of these papers (Cooper 1972) had econometricians as discussants. The discussion was emotional and much effort was given to showing how the econometric forecasts might have been revised to yield a more favorable comparison. No attempt was made to show how the extrapolations might have been improved.

except Kosobud's and Levenbach's. In Levenbach's, there was a tie for the 1-year forecast, and the econometric model was better for the 2-year forecast. None of these eight studies claimed to find a statistically significant difference. A comparison between ex post econometric forecasts and judgmental forecasts was carried out by Kosobud (1970); Fair (1971); Haitovsky, Treyz, and Su (1974); and Rippe and Wilkinson (1974). Although the econometric forecasts were superior in all but Rippe and Wilkinson, none of these studies reported on statistical significance. However, sufficient data were provided in the Rippe and Wilkinson study to allow for such a test; my analysis of their results indicated that the econometric forecasts were significantly poorer than the judgmental forecasts. Thus, the analyses of 13 ex post studies with 14 comparisons did not provide evidence that econometric methods were superior.

To obtain direct evidence on the short-range predictive validity of econometric methods, a review was made of studies involving ex ante or unconditional forecasts. To qualify for inclusion, a study must have compared econometric and alternative methods where each was carried out in a competent manner. The question of when a method was competently applied created some difficulty. The major effect of this restriction was to rule out studies where the alternative model was a "no-change" extrapolation. Some studies were retained (e.g., Ash and Smyth 1973), although the alternative models could have been improved.

In all, 12 studies involving 16 comparisons were found. These studies are summarized in table 2. The criteria were taken from each study. In other words, they were the most appropriate criteria in the opinion of the researchers who did each study. Efforts were made to test for statistical significance where this had not been done in the published study. In general, serious difficulties were encountered; most of these studies did not provide sufficient data (e.g., Naylor, Seaks, and Wichern 1972), others failed to use comparable time periods, and still others suffered from small sample sizes. The most striking result was that *not one study was found where the econometric method was significantly more accurate.* Nor did the econometric method show any general superiority: Six comparisons showed the econometric method to be superior, three suggested no difference, and seven found that it was inferior.

To guard against biases that may have been held by the author and to ensure that this study could be replicated, two research assistants coded a sample of three studies. The coding was done independently (i.e., the coders did not meet each other) and it was done blindly (i.e., the coders were not aware of the hypotheses in this study). In each of the four comparisons from these studies (Vandôme 1963; Markland

TABLE 2 **Accuracy of Econometric Methods for Short-Term Forecasting**

Relative Accuracy of Econometric Methods	Source of Evidence	Forecast Situation	Alternative Forecasting Method	Criteria for Accuracy (RMSE = root mean square error; MAPE = mean absolute percentage error)	Test of Statistical Significance
Significantly more accurate ($P <$ .05)	. . .	. . .	. . .	. . .	. . .
More accurate	Sims (1967)	Dutch economic indicators	Extrapolation	RMSE	None
	Ash and Smyth (1973)	U.K. economic indicators	Extrapolation	Theil's *U*	None
	McNees (1974)	U.S. economic indicators	Extrapolation	RMSE	None
	McNees (1974)	U.S. economic indicators	Judgmental	RMSE	None
	Haitovsky et al. (1974, table 7.3)	U.S. economic indicators	Judgmental	Average absolute error	None
	Christ (1975)	U.S. economic indicators	Extrapolation	RMSE	None
No difference	Sims (1967)	Norwegian economic indicators	Extrapolation	RMSE	None
	Ridker (1963)	Norwegian economic indicators	Extrapolation	(Five criteria used)	None
	Christ (1975)	U.S. economic indicators	Judgmental	RMSE	None
Less accurate	Vandome (1963)	U.K. economic indicators	Judgmental	Percentage changes	None
	Vandome (1963)	U.K. economic indicators	Extrapolation	MAPE	Armstrong*
	Naylor et al. (1972)	U.S. economic indicators	Extrapolation	Average absolute error	None
	McNees (1975)	U.S. economic indicators	Judgmental	Theil's *U*	None
	Cooper and Nelson (1975)	U.S. economic indicators	Extrapolation	RMSE/Theil's *U*	Armstrong*
	Liebling, Bidwell, and Hall (1976)	Nonresidential investment	Judgmental	MAPE	None
Significantly less accurate ($P <$ .05)	Markland (1970)	Inventory control	Extrapolation	Coefficient of variation	Armstrong*

*Details on these tests can be obtained from Scott Armstrong, Wharton School, University of Pennsylvania.

1970; Naylor et al. 1972) there was perfect agreement among the author and the two raters. In addition, five of the ex post prediction studies were coded (Kosobud 1970; Fair 1971; Cooper 1972; Elliott 1973; Granger and Newbold 1974). The only exception to perfect agreement occurred when one of the coders classified the econometric models as superior to extrapolation in Granger and Newbold. The agreement between the two raters and me on eight out of nine comparisons provides evidence that the ratings were reliable. (A copy of the instructions to the coders can be obtained from the author.)

The 16 comparisons of predictive validity were in agreement with the 14 ex post comparisons. Econometric forecasts were not found to be more accurate.[4]

Simple versus Complex Econometric Methods

"Progress" in econometric methods appears to be reflected by an increase in complexity in the methods used to analyze data. Leser (1968) noted long-term tendencies toward the use of more variables, more equations, more complex functional forms, and more complex interactions among the variables in econometric models. This increase in complexity can be observed by examining various issues of *Econometrica* since 1933 or by examining textbooks. The inference is that, because more complex procedures provide more realistic ways to represent the real world, they should yield more accurate forecasts.

Some researchers imply that complexity will lead to greater accuracy. For example, Suits (1962, p. 105) states " . . . clearly the fewer the equations the greater must be the level of aggregation and the less accurate and useful the result." Of course, not all econometricians believe this. Bassie (1958, p. 81) proposed a general rule, "the more a function is complicated by additional variables or by nonlinear relationships, the surer it is to make a good fit with past data and the surer it is to go wrong sometime in the future."

4. These results do not imply that econometric methods are of no value in short-range forecasting. A number of studies (e.g., Granger and Newbold [1974] and Cooper and Nelson [1975]) suggest that econometric forecasts can be combined with other types of forecasts to yield forecasts that are superior to any one of the components. Econometric methods are also valuable because they provide greater accuracy in long-range economic forecasting. Three studies on long-range forecasting met the criteria stated for table 2 (O'Herlihy et al. 1967; Armstrong 1968; Armstrong and Grohman 1972). Econometric methods were superior to other methods in each study. Furthermore, the relative superiority of the econometric method increased as the forecast horizon increased in two studies (Armstrong and Grohman 1972; Christ 1975). This finding conflicts with the viewpoints of many econometricians, however. For example, Wold (quoted as a discussant in NATO [1967], p. 48) implied that econometric methods are more appropriate for short-range than long-range forecasting because "the longer the forecast span, the more the actual course of events will be affected by minor influencing factors that are too numerous to be taken into account in a causal model."

Survey of Econometricians

To gain further information on whether experts believe that increased complexity in econometric models leads to more accurate forecasts, my previously mentioned mail survey asked: "Do complex methods generally provide more accurate or less accurate forecasts than can be obtained from less complex econometric methods for forecasting in the social sciences?—or is there no difference in accuracy?"[5] As shown in table 3, there was substantial agreement on the value of complexity; 72% of the experts agreed and only 9% disagreed. The experts were confident in their ratings on the value of complexity. The average confidence level was 4.0 (where 5 = "extremely confident").

Many factors could affect the relationship between complexity and accuracy. For example, Schmidt (1971), working with psychological data, found simple unit weights to be superior to regression weights for small sample sizes where there were many predictors. Furthermore, the relationship may not be a linear one; that is, complexity up to a modest level might be desirable, and beyond that it could be undesirable.

A specific question asked the experts to make any qualifications they felt important in assessing the relationship. Most respondents did qualify their answers, but it was difficult to find factors that were mentioned by more than one person.

Empirical Evidence

To assess the value of complexity in econometric methods, an examination was made of all published empirical evidence that I could find in the social sciences. Some studies provided indirect evidence on the value of complexity. McLaughlin (1973) examined the accuracy of forecasts from 12 econometric services in the United States. These forecasts were made by models that differed substantially in complex-

TABLE 3 **Survey of Experts on Complexity and Accuracy** (N = 21)

Complex Methods Rated	Percentage
Significantly more accurate	5
Somewhat more accurate	67
No difference (or undecided)	19
Somewhat less accurate	9
Significantly less accurate	0

5. The definitions were the same as provided in the footnote 1. Complexity was defined as follows: " 'Complexity' is to be thought of as an index reflecting the methods used to develop the forecasting model: (1) the use of coefficients other than 0 or 1 (2) the number of variables (more variables being more complex) (3) the functional relationship (additive being less complex than multiplicative; nonlinear more complex than linear) (4) the number of equations (5) whether the equations involve simultaneity."

ity (although all were complex). There were no reliable differences in accuracy among these models: The rankings of accuracy for the models in 1971 were negatively correlated (0.6 Spearman rank correlation) with those for 1972. If there are no reliable differences, then no differences would be found between accuracy and complexity. I reanalyzed data from the study by Jorgenson, Hunter, and Nadiri (1970) and found a perfect negative correlation between complexity of the four models (ranked by the number of variables in the model) and the stability of the regression coefficients from one period to the next (ranked by Jorgenson et al.); this lack of stability for more complex methods would suggest a loss in predictive validity. Friend and Taubman (1964) asserted their simple model was superior to more complex models (unfortunately they did not include the data from their study and, furthermore, the study only examined ex post predictive validity). Fair (1971) found little difference between his simple model and the more complex Wharton model in a test of ex post predictive validity.

Direct evidence on the value of complexity was sought by using only studies with ex ante forecasts. Each of the models, whether simple or complex, was done in a competent manner. The results of this literature survey are summarized in table 4.

To determine whether the coding of the studies in table 4 was reliable, eight of the 11 studies (all but Johnston and McNeal [1964], Grant and Bray [1970], and McNees [1974]) were independently coded by two research assistants. The coding was blind in that the assistants were unaware of the hypotheses. Discrepancies were noted on only two of these studies; one assistant coded Dawes and Corrigan (1974) to show that more complex methods were superior, and the other assistant reported complexity to be superior in Wesman and Bennett (1959). The studies in table 4 suggest that complexity and accuracy are not closely related. No study reported a significant positive relationship between complexity and accuracy. Overall, seven comparisons favored less complexity and four favored more complexity.

The 11 studies that assessed predictive validity directly were in agreement with the five studies that provided indirect evidence: Added complexity did not yield improvements in accuracy. The empirical evidence does not support the folklore in this area.

Multiple Hypotheses: An Alternative Research Strategy

The first part of this paper suggested that econometricians often act as advocates; they attempt to find evidence to support their viewpoint. Furthermore, group opinion is often used to judge truth. Under such conditions, it is likely that beliefs will persist even if unsupported by empirical evidence.

TABLE 4 **Accuracy of Simple vs. Complex Methods**

Relative Accuracy of Complex Methods	Source of Evidence	Forecast Situation	Criterion for Accuracy	Nature of Comparison	Test of Statistical Significance
Significantly more accurate ($P < .05$)	...	...	...	...	...
More accurate	Stuckert (1958)	Academic performance	Percent correct	Unit weights vs. regression	None
	McNees (1974)	GNP	Theil coefficient; RMSE, mean absolute error	Small vs. large models	None
	Grant and Bray (1970)	Personnel	Correlation coefficient	Unit weights vs. regression	Armstrong*
	Johnston and McNeal (1964)	Medicine	Correlation coefficient	Unit weights vs. regression	Authors
No difference	...	...	...	...	...
Less accurate	Dawes and Corrigan (1974)	Academic performance, simulated data, psychiatric ratings	Correlation coefficient	Unit weights vs. regression	Armstrong*
	Lawshe and Schucker (1959)	Academic performance	Percent correct	Unit weights vs. regression	None
	Reiss (1951)	Criminology	Percent correct	Few vs. many causal variables	None
	Wesman and Bennett (1959)	Academic performance	Correlation coefficient	Unit weight vs. regression	None
	Scott and Johnson (1967)	Personnel selection	Percent correct, correlation coefficient	Unit weights vs. regression	None
Significantly less accurate ($P < .05$)	Claudy (1972)	Simulated data (typical of psychological data)	Correlation coefficient	Unit weights vs. regression	Armstrong*
	Summers and Stewart (1968)	Political judgments	Correlation coefficients	Linear vs. nonlinear models	Armstrong*

*Details on these tests can be obtained from Scott Armstrong, Wharton School, University of Pennsylvania.

An alternative to the use of advocacy is to adopt the method of multiple hypotheses (Chamberlin [1890] 1965; Platt 1964). Here, each scientist examines two or more reasonable hypotheses (or methods) at the same time. The role of the scientist is to determine which of the methods is most useful in the given situation. When two or more reasonable hypotheses are studied, it is less likely that the scientist will feel a bias in favor of "his" hypothesis—they are all "his" hypotheses. The orientation of the scientist is changed from one where he seeks to confirm a hypothesis to one where he seeks to disconfirm one or more hypotheses. Because the various hypotheses are tested within each study, there is less need to rely upon the opinions of other experts. The method of multiple hypotheses should help researchers to make more effective use of disconfirming evidence.

Although the method of multiple hypotheses would appear to be less prone to selective perception, and thus superior to the use of advocacy, surprisingly little evidence is available on this issue. This evidence, summarized in Armstrong (1978*b*), provides modest support for multiple hypotheses over advocacy. Most surprising again was the lack of evidence to support advocacy, the research strategy that appears to be most common among social scientists.

Conclusions

Certain hypotheses about econometric methods have been accepted for years despite the lack of evidence. Ninety-five percent of the experts agreed that econometric methods are superior for short-range forecasting. An examination of the empirical literature did not support this belief: Econometric forecasts were not shown to be significantly better in any of the 14 ex post and 16 ex ante tests. Furthermore, there was no tendency toward greater accuracy over these 30 tests. Similarly, 72% of the experts felt that complexity contributed to accuracy, but the examination of the literature did not support such a belief: Complex models were not significantly better in any of the five indirect and 11 direct tests.

Thrusting disconfirming evidence upon others provides an ineffective way of changing attitudes. Econometricians are more likely to be convinced by their own studies. The use of the method of multiple hypotheses provides a rational way for econometricians to test their beliefs.

In one sense the situation is encouraging. Twenty-three studies using the method of multiple hypotheses were found (see tables 2 and 4). These studies are becoming more common; the oldest study was published in 1951 and almost half were published since 1970. This trend in research strategy should be useful in distinguishing folklore from fact.

References

Armstrong, J. Scott. 1968. Long-range forecasting for international markets: the use of causal models. Pp. 222–27 in Robert L. King (ed.), *Marketing and the New Science of Planning*. Chicago: American Market Association.

Armstrong, J. Scott. 1978*a*. *Long-Range Forecasting: From Crystal Ball to Computer*. New York: Wiley-Interscience.

Armstrong, J. Scott. 1978*b*. Advocacy and objectivity in science. Wharton School Working Paper. Mimeographed. Philadelphia: Wharton School of Management, University of Pennsylvania.

Armstrong, J. Scott, and Grohman, Michael, C. 1972. A comparative study of methods for long-range market forecasting. *Management Science* 19:211–21.

Asch, S. E. 1965. Effects of group pressure upon the modification and distortion of judgments. In Harold Proshansky and Bernard Seidenberg (eds.), *Basic Studies in Psychology*. New York: Holt, Rinehart & Winston.

Ash, J. C. K., and Smyth, D. J. 1973. *Forecasting the United Kingdom Economy*. Farnborough: Saxon House.

Barber, Bernard. 1961. Resistance by scientists to scientific discovery. *Science* 134:596–602.

Bassie, V. Lewis. 1958. *Economic Forecasting*. New York: McGraw-Hill.

Bassie, V. Lewis. 1972. A note on scientific method in forecasting. Pp. 1211–18 in B. G. Hickman (ed.), *Econometric Models of Cyclical Behavior*. Vol. 2. Studies in Income and Wealth no. 36. New York: Columbia University Press.

Brown, T. Merritt. 1970. *Specification and Uses of Econometric Models*. New York: St. Martin's.

Chamberlin, T. C. (1890) 1965. The method of multiple working hypotheses. *Science* 148:754–59.

Chapman, Loren J., and Chapman, J. R. 1969. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74:271–80.

Christ, Carl F. 1951. A test of an econometric model for the United States, 1921–47. Pp. 35–107 in *Conference on Business Cycles*. New York: National Bureau for Economic Research. (Discussion on pp. 107–29.)

Christ, Carl F. 1956. Aggregate econometric models. *American Economic Review* 46:385–408.

Christ, Carl F. 1975. Judging the performance of econometric models of the U.S. economy. *International Economic Review* 16 (1975): 54–74.

Claudy, John G. 1972. A comparison of five variable weighting procedures. *Educational and Psychological Measurement* 32:311–22.

Cooper, J. Phillip, and Nelson, Charles R. 1975. The ex ante prediction performance of the St. Louis and FRB-MIT-PENN econometric models and some results on composite predictors. *Journal of Money, Credit and Banking* 7:1–32.

Cooper, Ronald L. 1972. The predictive performance of quarterly econometric models of the United States. Pp. 813–926 in Bert G. Hickman (ed.), *Econometric Models of Cyclical Behavior*. Vol. 2. Studies in Income and Wealth no. 36. New York: Columbia University Press. (Discussion on pp. 926–47.)

Dawes, Robyn M., and Corrigan, Bernard. 1974. Linear models in decision making. *Psychological Bulletin* 81:95–106.

De Grazia, Alfred; Juergens, Ralph E.; and Steccini, Livio C. 1966. *The Velikovsky Affair*. Hyde Park, N.Y.: University Books.

Elliott, J. W. 1973. A direct comparison of short-run GNP forecasting models. *Journal of Business* 46:33–60.

Fair, Ray C. 1971. *A Short-Run Forecasting Model of the United States*. Lexington, Mass.: Heath.

Fischhoff, Baruch, and Beyth, Ruth. 1975. I knew it would happen: remembered probabilities of once-future things. *Organizational Behavior and Human Performance* 13:1–16.

Friend, Irwin, and Taubman, P. 1964. A short-term forecasting model. *Review of Economics and Statistics* 46:229–36.

Geller, E. Scott, and Pitz, Gordon F. 1968. Confidence and decision speed in the revision of opinion. *Organizational Behavior and Human Performance* 3:190–201.

Granger, C. W. J., and Newbold, P. 1974. Economic forecasting: the atheist's viewpoint. In G. A. Renton (ed.), *Modelling the Economy*. London: Heinemann.
Grant, Donald L., and Bray, Douglas W. 1970. Validation of employment test for telephone company installation and repair occupations. *Journal of Applied Psychology* 54:7–15.
Greenwald, Anthony G. 1975. Consequences of prejudice against the null hypothesis. *Psychological Bulletin* 82:1–20.
Haitovsky, Yoel; Treyz, G.; and Su, V. 1974. *Forecasts with Quarterly Macroeconomic Models*. New York: Columbia University Press.
Ibrahim, I. B., and Otsuki, T. 1976. Forecasting GNP components using the method of Box and Jenkins. *Southern Economic Journal* 42:461–70.
Johnston, Roy, and McNeal, B. F. 1964. Combined MMPI and demographic data in predicting length of neuropsychiatric hospital stay. *Journal of Consulting Psychology* 28:64–70.
Jorgenson, Dale W.; Hunter, J.; and Nadiri, M. I. 1970. The predictive performance of econometric models of quarterly investment behavior. *Econometrica* 38:213–24.
Kosobud, Richard F. 1970. Forecasting accuracy and uses of an econometric model. *Applied Economics* 2:253–63.
Lawshe, C. H., and Schucker, R. E. 1959. The relative efficiency of four test weighting methods in multiple prediction. *Educational and Psychological Measurement* 19:103–14.
Leser, C. E. V. 1968. A survey of econometrics. *Journal of the Royal Statistical Society* 131, series A: 530–66.
Levenbach, Hans; Cleary, J. P.; and Fryk, D. A. 1974. A comparison of ARIMA and econometric models for telephone demand. *Proceedings of the American Statistical Association: Business and Economics Statistics Section,* pp. 448–50.
Liebling, Herman I.; Bidwell, Peter T.; and Hall, Karen E. 1976. The recent performance of anticipations surveys and econometric model projections of investment spending in the United States. *Journal of Business* 49:451–77.
McLaughlin, Robert L. 1973. The forecasters' batting averages. *Business Economics* 3:58–59.
McNees, Stephen K. 1974. How accurate are economic forecasts? *New England Economic Review* (November-December), pp. 2–19.
McNees, Stephen K. 1975. An evaluation of economic forecasts. *New England Economic Review* (November-December), pp. 3–39.
Mahoney, Michael J. 1977. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1:161–75.
Markland, Robert E. 1970. A comparative study of demand forecasting techniques for military helicopter spare parts. *Naval Research Logistics Quarterly* 17:103–19.
Narasimham, Gorti; Castellino, V. F.; and Singpurwalla, Nozer D. 1974. On the predictive performance of the BEA quarterly econometric model and a Box-Jenkins type ARIMA model. *Proceedings of the American Statistical Association: Business and Economics Section,* pp. 501–4.
NATO. 1967. *Forecasting on a Scientific Basis*. Lisbon: NATO.
Naylor, Thomas H.; Seaks, T. G.; and Wichern, D. W. 1972. Box-Jenkins methods: an alternative to econometric models. *International Statistical Review* 40:123–37.
Nelson, Charles R. 1972. The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review* 5:902–17.
O'Herlihy, C.; Fane, G.; Gwilliam, K. M.; and Ray, G. F. 1967. Long-term forecasts of demand for cars, selected consumer durables and energy. *National Institute Economic Review* 40:34–61.
Platt, John R. 1964. Strong inference. *Science* 146:341–53.
Pruitt, D. G. 1961. Informational requirements in making decisions. *American Journal of Psychology* 74:433–39.
Reiss, Albert J., Jr. 1951. The accuracy, efficiency and validity of a prediction instrument. *American Journal of Sociology* 56:552–61.
Ridker, Ronald G. 1963. An evaluation of the forecasting ability of the Norwegian national budgeting system. *Review of Economics and Statistics* 45:23–33.
Rippe, R. D., and Wilkinson, M. 1974. Forecasting accuracy of the McGraw-Hill anticipations data. *Journal of the American Statistical Association* 69:849–58.

Rosenthal, Robert, and Fode, K. L. 1963. The effect of experimenter bias on the performance of the albino rat. *Behavioral Science* 8:183–89.
Rosenthal, Robert, and Rosnow, Ralph L. 1969. *Artifact in Behavioral Research.* New York: Academic Press.
Schmidt, Frank L. 1971. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement* 31:699–714.
Scott, Richard D., and Johnson, R. W. 1967. Use of the weighted application blank in selecting unskilled employees. *Journal of Applied Psychology* 51:393–95.
Sims, Christopher A. 1967. Evaluating short-term macroeconomic forecasts: the Dutch performance. *Review of Economics and Statistics* 49:225–36.
Stuckert, R. P. 1958. A configurational approach to prediction. *Sociometry* 21:225–37.
Suits, Daniel B. 1962. Forecasting and analysis with an econometric model. *American Economic Review* 52:104–32.
Summers, David A., and Stewart, Thomas R. 1968. Regression models of foreign policy beliefs. *Proceedings of the American Psychological Association,* pp. 195–96.
Vandome, Peter. 1963. Econometric forecasting for the United Kingdom. *Bulletin of the Oxford University Institute of Economics and Statistics* 25:239–81.
Wason, P. C. 1960. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12:129–40.
Wason, P. C. 1968. On the failure to eliminate hypotheses—a second look. Pp. 165–74 in P. C. Wason and P. N. Johnson-Laird (eds.), *Thinking and Reasoning.* Baltimore: Penguin.
Wesman, A. G., and Bennett, G. K. 1959. Multiple regression vs. simple addition of scores in prediction of college grades. *Educational and Psychological Measurement* 19:243–46.
Worswick, G. D. N. 1974. Preston, R. S., *The Wharton Annual and Industry Forecasting Model. Journal of Economic Literature* 12:117–18.