Man vs. Model? The Role of Judgment in Forecasting

Stephen K. McNees

Vice President and Economist, Federal Reserve Bank of Boston. *The author is indebted to the modelers for graciously providing the history of their forecasts. He also thanks James Clark for valued research assistance. This article will also appear in the* International Journal of Forecasting's *special issue on macroeconomics.*

Two apparently conflicting themes have developed on the role of judgment in forecasting. On the one hand, the literature from the field of psychology is replete with studies in which individuals' biases, prejudices, unfounded optimism, and undue conservatism undermine the accuracy of their predictions. Some interpret this evidence of fallibility as indicating that any form of subjective adjustment is "unscientific," and bound to impair predictive accuracy. Their counsel is "Don't trust your common sense" (Armstrong 1985, p. 86).

In contrast, economists commonly postulate the economic "rationality" of the individual. Many argue that much of psychologists' evidence has been gathered in unrealistic, "artificial" environments in which the participants were untrained and lacked incentives to do their best. In addition, they point to evidence showing that predictive accuracy in macroeconomics derives mainly from individuals' adjustments to their models (Evans, Haitovsky, and Treyz 1972; Haitovsky and Treyz 1972). Critics of large-scale macromodels take the fact that the models are adjusted as prima facie evidence that the models are "incredible" and "provide no useful information" (Sims 1980, p. 3 and Lucas 1976, p. 20). Many model builders themselves insist that their models are fragile and unsuitable for use without "tender loving care." Thus, at least in macroeconomics, it is uncommon for forecasters to attribute their forecasting errors to their own judgment in order to exonerate their models.

The primary purpose of this article is to review some recent evidence on the value of judgmental adjustments in macroeconomic forecast accuracy. With the notable exception of Wallis's (1989, pp. 52–55) studies of macroeconometric models of the United Kingdom, much of the evidence on which the conventional impression rests was gathered more than two decades ago and based on a fairly small sample of models very different from the current generation. Are judgmental adjustments still required to keep macromodels within reasonable bounds? Or, as psychologists suggest, do forecasters exaggerate the value of their intuitive insights, undervaluing the importance of their models?

Those who characterize judgment as "bad" and concurrently observe that models are even less accurate without adjustments often draw the cynical conclusion that because nothing works, all models are "bad" and anyone's forecast is as good as anyone else's. This article concludes that such a view misinterprets the evidence from both psychology and economics. The fallacy follows from thinking of either judgment or models in absolute terms as "good" or "bad." What psychologists have demonstrated is that models, even highly imperfect ones, are more accurate than "pure," unreasoned expertise or global intuition. The moral to be drawn is that some systematic procedure, or "model," helps to integrate disparate information. What macroeconomic forecasting evidence suggests is that those who do use models, as opposed to pure intuition or expertise, are also often aware of some of the limitations of applying their models in practice. More often than not, model builders can adapt their models to bring them closer to actual outcomes. The best forecasts are made, not by abandoning models or by abandoning judgment, but by blending both sources of information. The observed mixture does not appear to be the optimal blend, however.

I. Adjusted versus Unadjusted Macroforecasts

The most prominent macroeconomic forecasts are generated by forecasters who adjust their purely mechanical model simulations in an attempt to improve their accuracy. The practice of adjusting models has been severely criticized on both theoretical and empirical grounds, with the main inference drawn that the underlying models are so unreliable that any success achieved in forecast accuracy must derive from the personal insights of the forecaster who adjusts the model.

One of the earliest empirical criticisms was Nelson's (1972) demonstration that a simple univariate time series or ARIMA model could produce more accurate one-quarter-ahead forecasts than the ex post simulations of a large-scale macroeconometric model. Nelson's conclusion was based on a fairly small sample collected more than twenty years ago. Since 1976, Nelson has used simple univariate equations to mechanically generate forecasts of nominal GNP and the GNP implicit price deflator (IPD) and thus, implicitly, real GNP. Nelson refers to these equations as the Benchmark (BMARK) model because they are intended to be used as a standard of comparison for assessing macroeconomic forecasts. In early 1988, Frederick Joutz of George Washington University extended the BMARK model and now issues regular forecasts of several more variables.

Table 1 compares the accuracy of the mechanically generated BMARK forecasts with the average of four prominent adjusted forecasts—those by Data Resources, Inc. (DRI), Georgia State University (GSU), the Research Seminar in Quantitative Economics at the University of Michigan (RSQE), and Wharton Econometric Forecasting Associates (WEFA). For all three variables and four horizons the BMARK forecast was less accurate. In fact, the BMARK forecast was always the least accurate of all the forecasters, with a single exception. The one-year-ahead forecasts of the implicit price deflator (IPD) showed BMARK to be virtually identical to two of the adjusted forecasts. Contrary to some previous evidence, the margin of superiority of the adjusted forecasts is greatest at the shortest horizon and decreases as the horizon lengthens. Moreover, the mechanically generated BMARK forecasts of IPD are nearly as accurate as the adjusted forecasts of a three-quarter horizon. Finally, note that these results pertain to cumulative changes and that BMARK fore-

Table 1
Evaluation of BMARK Model Forecasts against Four Adjusted Forecasts, by Variable and Forecast Horizon,[a] 1976:II–1989:IV

Variable	Forecast Horizon (quarters) 1	2	3	4	Average for All Horizons
Implicit GNP Price Deflator	86	92	96	94	92
Real GNP	78	84	85	88	84
GNP	74	80	85	90	82
All Variables	80	86	89	91	86

[a][(Mean RMSE/BMARK RMSE) * 100] where mean RMSE is the average of the RMSEs of the four adjusted forecasts: Data Resources, Inc. (DRI); Georgia State University (GSU); the Research Seminar in Quantitative Economics at the University of Michigan (RSQE); and Wharton Economic Forecasting Associates (WEFA).

Table 2
Evaluation of BVAR Model Forecasts against Four Adjusted Forecasts, by Variable and Forecast Horizon,[a] 1980:II–1989:IV

Variable	Forecast Horizons (quarters) 1	2	3	4	5	6	7	8	Average for All Horizons
Unemployment Rate	80	92	98	107	121	136	142	128	113
Real GNP	90	98	102	107	114	111	100	83	101
Money Supply, M1	99	102	102	99	98	97	97	95	98
Business Fixed Investment, Real	94	84	80	82	84	85	87	91	86
GNP	77	77	77	79	80	81	80	77	78
Treasury Bill Rate	81	85	80	79	78	75	74	74	78
Implicit GNP Price Deflator	84	64	56	51	48	44	46	47	55
All Variables	86	86	85	86	89	90	89	85	87

[a][(Mean RMSE/BVAR RMSE) * 100] where mean RMSE is the average of the RMSEs of the four adjusted forecasts: DRI, GSU, RSQE, and WEFA.

casts of quarterly changes are as accurate as other forecasts (Nelson 1984; McNees 1988).

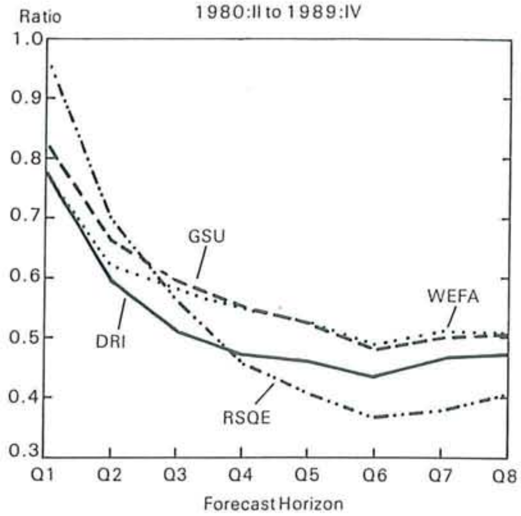
A decade ago, Litterman (1979) developed a six-variable statistical model, which he calls a Bayesian vector autoregression or BVAR model. He used the BVAR model to generate forecasts mechanically. Initially, the BVAR model produced relatively accurate forecasts of the severe recession of 1981–82 but highly inaccurate forecasts of the inflation rate (McNees 1986). In response to the poor inflation performance, the model was expanded to incorporate three additional variables hoped to contain predictive information on inflation (Litterman 1986). In August 1987, Sims (1989) modified the statistical procedures on which the model is based.

This evolution of the model illustrates why it is a mistake to think of a model used mechanically to forecast as "free of human judgment." On the contrary, the BVAR model embodies novel and sophisticated techniques that have evolved over nearly a decade. Nevertheless, the BVAR forecasts, as well as the BMARK forecasts, are as close examples of a "pure model" forecast as one is likely to see. In contrast to traditional macroeconometric models, they do not require an explicit set of external assumptions to generate their forecasts. (Indeed, it has been argued that a BVAR model is incapable of generating forecasts conditional on fixed assumptions about, for example, the future path of macroeconomic policy instruments.) It is the combination of not requiring explicit input assumptions and the modeler's refraining from adjusting the mechanically generated forecast that places the BVAR model forecasts at the "pure model/no judgment" end of the spectrum.

Table 2 contrasts the accuracy of the BVAR model forecasts with the average of the four adjusted forecasts. (The BVAR forecasts enjoy some advantage in that five of the more recent ones were made later in the quarter when more high-frequency and revised data were available. All other forecasts were issued soon after the release of the preliminary GNP data for the prior quarter.) For four of the seven variables examined, the BVAR forecasts are distinctly inferior to the others. This is particularly true of the forecasts of the rate of inflation as measured by the IPD. (See Chart 1.) The forecasts of the narrow definition of the money stock are roughly as accurate as the adjusted forecasts. For the other two variables, the relative performance of the BVAR model forecast depends on the forecast horizon. Chart 2 presents the ratios of the RMSEs of each of the four adjusted unemployment rate forecasts to the RMSE of the BVAR model. The BVAR model's one-quarter-ahead forecast of the unemployment rate is less accurate than the others but its five- through eight-quarter-ahead forecasts were the most accurate, often by a sizable margin. Chart 3 provides comparable information for the real GNP forecasts. For this variable, the BVAR forecasts were the most accurate, by small margins, for the four-quarter through six-quarter horizons, but the least accurate by fairly small margins for the shortest and longest forecast horizons. This result contrasts sharply with the superior performance of the BVAR model in the early 1980s. For the longer horizons, the BVAR model's real GNP forecasts were somewhat less accurate in the late 1980s than in the early 1980s.

Chart 1

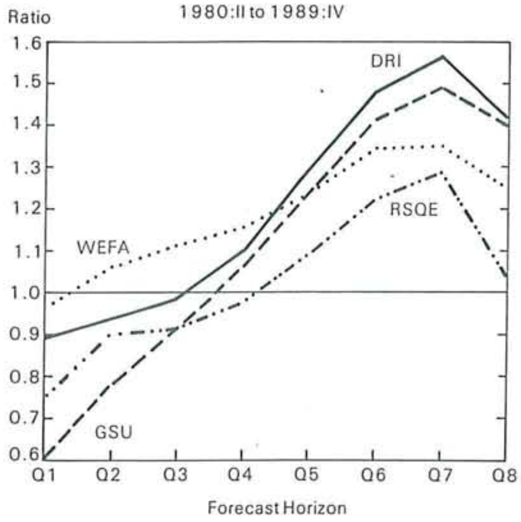
Implicit GNP Price Deflator: Ratio of RMSEs of Adjusted Forecasts to RMSEs of BVAR Forecasts



RMSE= Root mean squared error.

Chart 2

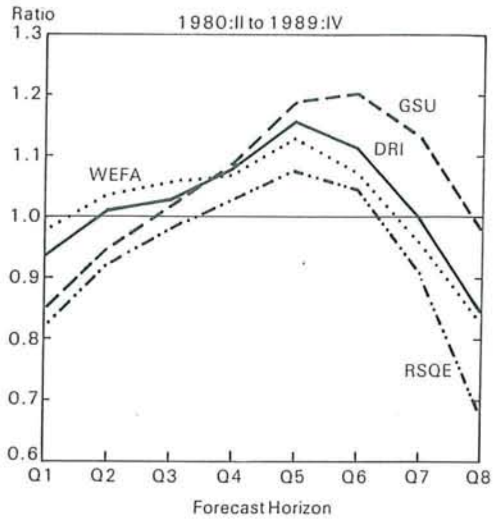
Unemployment Rate: Ratios of RMSEs of Adjusted Forecasts to RMSEs of BVAR Forecasts



RMSE = Root mean squared error.

Chart 3

Real GNP: Ratios of RMSEs of Adjusted Forecasts to RMSEs of BVAR Forecasts



RMSE = Root mean squared error.

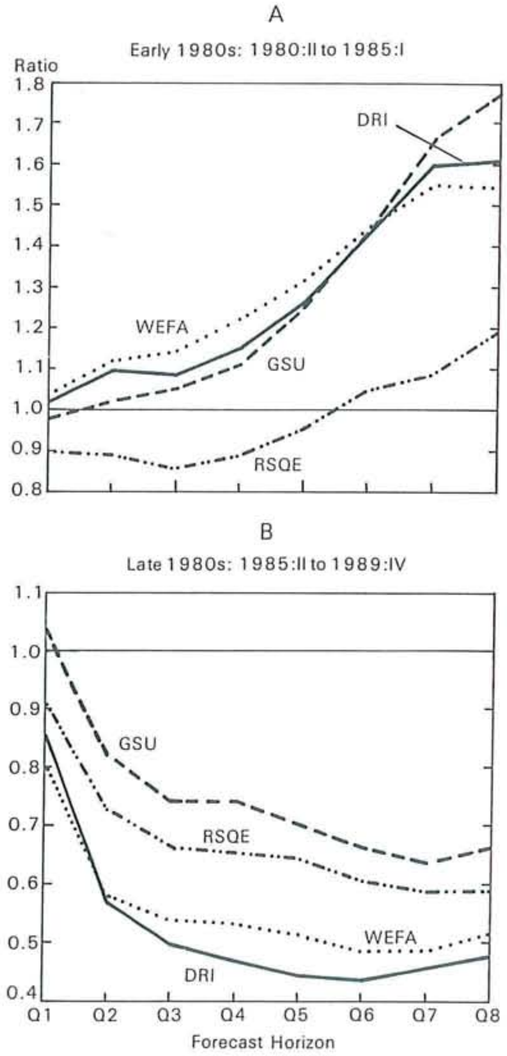
In sharp contrast, most other forecasters' real GNP forecasts have been far more reliable in the late 1980s than they were in the early 1980s.

The discrepancy between the relative accuracy of the BVAR model in the first and second halves of the 1980s is most clearly illustrated by its performance on business fixed investment measured in constant dollars. The top panel in Chart 4 shows that in the early 1980s the BVAR model's real capital spending forecasts were the most accurate by a sizable margin. In sharp contrast, its more recent forecasts have been the least accurate, also by a sizable margin. This clearly illustrates how the past, even when it appears fairly unambiguous, need not be a reliable guide to the future.

In 1983, Fair started to issue regular forecasts generated by his "structural" macroeconometric model. Unlike the statistical BMARK and BVAR models, Fair's model is based on an explicit theory of economic behavior. In addition, the model user must select specific assumptions about external conditions, such as fiscal policy and economic developments outside the United States, in order to generate a forecast. Unlike most other large-scale macroecono-

Chart 4

Business Fixed Investment: Ratios of RMSEs of Adjusted Forecasts to RMSEs of BVAR Forecasts



RMSE = Root mean squared error.

metric model users, Fair consistently refrains from adjusting his model's forecasts once he provides fairly mechanical assumptions about external variables. Thus, the Fair model represents minimal forecaster input from the user of a conditional, "structural" model.

Table 3 contrasts the forecast performance of the Fair model with the average of the adjusted forecasts of the four other models. The Fair model forecasts of three of the five variables examined—the Treasury bill rate, nominal GNP, and the unemployment rate—are distinctly less accurate than the other forecasts. Its relative performance for real GNP and inflation is more complicated to describe. The Fair inflation performance depends critically on the horizon of the forecast. As shown in Chart 5, Fair's one-quarter-ahead inflation forecast is nearly as accurate as those of three forecasters and more accurate than the fourth forecast. Four to eight quarters ahead, the accuracy of Fair's inflation forecasts is similar to two and inferior to two other forecasters. Only one forecaster clearly dominates Fair's inflation forecasts at all horizons.

Chart 6 shows that the relative accuracy of Fair's real GNP forecasts depends on the horizon as well as which individual forecaster is used for comparison. Specifically, the Fair model was uniformly more accurate than the GSU forecasts, especially at longer horizons, quite similar to the other three forecasters' up through six quarters, and more accurate at the longer horizons.

Overall, looking at all variables, it seems clear that the adjusted forecasts tend to be more accurate than those generated mechanically. There are, however, some significant exceptions to this generalization: for the longer horizons, the most accurate real GNP forecasts came from the Fair model and the most accurate unemployment rate forecasts were generated with the BVAR model. Even the simple, univariate BMARK model provides a fairly demanding standard of comparison, especially for IPD forecasts. Nevertheless, the historical record does suggest that adjusted forecasts are usually the most accurate, despite instances of success of mechanically generated forecasts.

II. Do Adjustments Improve Forecast Accuracy?

The preceding comparisons do not isolate the role of judgment per se because the forecasters who refrain from making adjustments use different models from those used by forecasters who do adjust their models. The general superiority, though not total dominance, of the adjusted forecasts could be due

Table 3
Evaluation of Fair Model Forecasts against Four Adjusted Forecasts, by Variable and Forecast Horizon,[a] 1983:II–1989:IV

Variable	Forecast Horizons (quarters) 1	2	3	4	5	6	7	8	Average for All Horizons
Real GNP	99	107	108	109	115	122	132	136	116
Implicit GNP Price Deflator	103	100	94	88	84	83	81	82	89
Unemployment Rate	68	76	74	74	76	80	90	97	79
GNP	89	86	79	73	68	66	64	64	74
Treasury Bill Rate	40	58	67	67	68	66	67	67	63
All Variables	80	85	85	82	82	84	86	89	84

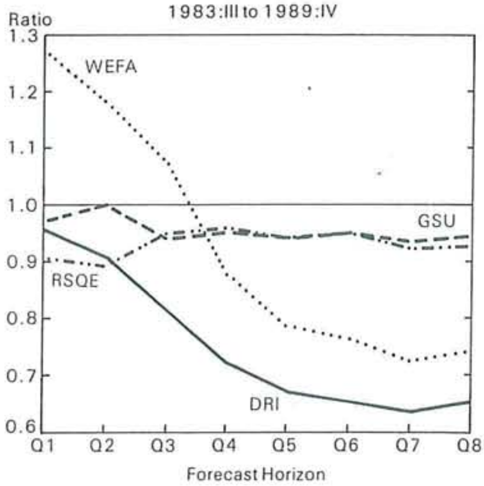
[a][(Mean RMSE/FAIR RMSE) * 100] where mean RMSE is the average of the RMSEs of the four adjusted forecasts: DRI, GSU, RSQE, and WEFA.

either to the adjustments *or* to the superiority of those models that were judgmentally adjusted. To distinguish between the two, one needs both adjusted and unadjusted forecasts from the *same* model. Typically, those who refrain from making adjustments do not accompany their mechanically generated forecast with their own judgment about where the model may go wrong and those who adjust their models do not also provide their model's "mechanical" forecast, prior to adjustment.

Fortunately, four prominent macroeconometric forecasters (not the same group considered above) who do adjust their models' forecasts have provided data on both their publicized (adjusted) and mechan-

Chart 5

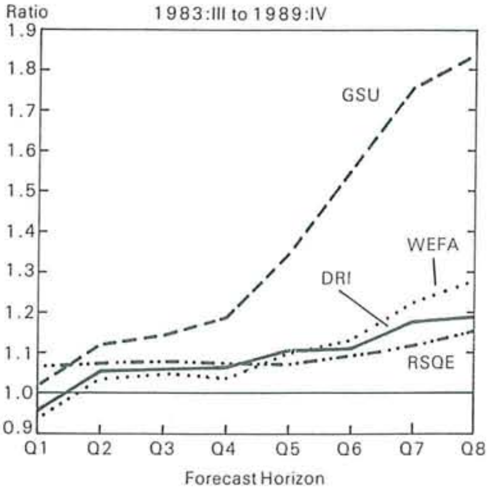
Implicit GNP Price Deflator: Ratios of RMSEs of Adjusted Forecasts to RMSEs of Fair Forecasts



RMSE = Root mean squared error.

Chart 6

Real GNP: Ratios of RMSEs of Adjusted Forecasts to RMSEs of Fair Forecasts



RMSE = Root mean squared error.

Table 4
Impact of Judgment on Forecasting Accuracy

	Forecast Horizon					
	One Quarter Ahead		Four Quarters Ahead		Eight Quarters Ahead	
	Judgment Improved (Percent) (1)	Lower RMSE (2)	Judgment Improved (Percent) (3)	Lower RMSE (4)	Judgment Improved (Percent) (5)	Lower RMSE (6)
Forecaster 1	68*	9 of 13	49	8 of 13	33	5 of 13
Forecaster 2	66*	15 of 19	62	16 of 19	65	15 of 19
Forecaster 3	62*	16 of 20	47	9 of 20	53	10 of 20
Forecaster 4	55	14 of 19	66	15 of 19	64	15 of 19
Short-term interest rates	90*	3 of 3	77	3 of 3	60	3 of 3
Long-term interest rates	78*	4 of 4	78	4 of 4	76	3 of 4
Federal deficit	76*	3 of 3	71	2 of 3	56	3 of 3
Consumer prices	76*	4 of 4	72	4 of 4	61	4 of 4
Corporate profits	74*	3 of 3	56	2 of 3	75	3 of 3
Nominal GNP	66*	4 of 4	47	2 of 4	65	2 of 4
Labor compensation	65*	2 of 3	52	2 of 3	58	1 of 3
Consumption	64*	2 of 4	50	2 of 4	60	3 of 4
Exports	63	2 of 2	47	1 of 2	30	1 of 2
Residential investment	63	4 of 4	72	4 of 4	60	4 of 4
State and local purchases	61	4 of 4	66	3 of 4	53	2 of 4
Unemployment rate	60	3 of 4	48	2 of 4	50	2 of 4
Business fixed investment	59	4 of 4	51	2 of 4	50	2 of 4
GNP implicit price deflator	59	2 of 4	72	4 of 4	85	4 of 4
Nominal net exports	57	2 of 3	57	3 of 3	50	2 of 3
Narrow money stock	56	1 of 2	58	2 of 2	50	2 of 2
Real GNP	55	3 of 4	50	1 of 4	50	1 of 4
Import price deflator	54	1 of 3	52	1 of 3	56	1 of 3
Productivity	53	2 of 2	59	2 of 2	57	1 of 2
Change in business inventories	46	1 of 4	40	2 of 4	48	1 of 4
Imports	36*	0 of 3	28	0 of 3	44	0 of 3
Total	62*	54 of 71	57	48 of 71	58	45 of 71

*Significantly different from 50 at 90 percent confidence level for one-quarter-ahead forecast. Because the four- and eight-quarter-ahead forecasts are not independent, no statistical test was applied. For each variable there are 11 to 13 one-quarter-ahead forecasts, 8 to 10 four-quarter-ahead forecasts and 4 to 6 eight-quarter-ahead forecasts.

ical (unadjusted) forecasts. (The mechanical forecasts were generated using a predetermined, fixed rule for taking account of recent residuals. Because these adjustment rules were not varied over time by the forecaster, it is appropriate to think of them not as an ad hoc adjustment but rather as part of the model.) The impact of their adjustments on the accuracy of their forecasts is described in this section.

It is important to emphasize that these data measure the importance of forecasters' adjustments of their models, not judgment in some absolute sense. Adjustments are most valuable when a forecaster is especially astute or a model especially poor. Adjustments can be harmful either when the adjuster has no information or when the model is quite reliable by itself. The results describe the net interaction between forecasters and models, roughly apportioning the forecast accuracy between the two factors.

The simplest way to measure the impact of adjustments on forecast accuracy is to count the number of times that adjusted forecasts were more accurate than those generated mechanically. Column 1 in Table 4 shows that 62 percent of all one-quarter-ahead forecasts were more accurate with adjustments. Because this is based on a large number (841) of observations, one can infer with a high level of confidence that the superiority of adjusted forecasts is significant in the statistical sense. The table shows that the proportion of improvements ranged among forecasters from a low of 55 percent (which is not significantly different from 50 percent at a 90 percent level of confidence) by forecaster 4 to a high of 68 percent for forecaster 1. Although adjustments clearly help all forecasters on average, note that improvement is by no means assured. The table also shows, for all forecasters combined, the differences among variables. Adjustments improved 90 percent of the one-quarter-ahead forecasts of short-term interest rates. This nearly universal improvement may well reflect the fact that actual interest rate data are contemporaneously available while these models, like most macroeconometric forecasting models, are based on quarterly data. Of the variables where the adjusted forecasts were most often more accurate, several are available on at least a monthly basis.

Adjustments do not always enhance accuracy, however. Mechanically generated forecasts of the change in business inventories were usually better, and those of real imports significantly better, than the adjusted ones. Data for these two variables are not available on an accurate, current basis. Indeed, few reliable data on these variables are available when the preliminary GNP data are released. For these variables, forecasters would appear to do better to rely on their models until the "actual" data are released.

The frequency of release of actual data does not, of course, explain all of these results. Model adjustments improve forecasts of corporate profits even though the actual data only become available with a considerable lag. Similarly, even though preliminary money stock data are available weekly, the adjusted forecasts are closer to the final, actual figures only slightly more often than mechanically generated model forecasts. It is interesting also to observe that for several variables including, most notably, real GNP, mechanical forecasts are more accurate nearly as often as the publicized, "official" forecasts.

The assumptions behind the mechanical forecast, as well as the model itself, reflect the modeler's judgments.

A simple ordinal, better-or-worse comparison provides no indication of how much adjustments improve accuracy. Conceivably, model adjustments could frequently help a little but occasionally hurt a lot. Column 2 of Table 4 shows that this is not so. It compares the RMSEs of the adjusted and mechanical forecasts. (The same overall result holds for comparisons of the mean absolute forecast errors.) Of seventy-one forecaster/variable combinations, three-quarters of the RMSEs of the adjusted forecasts are lower. This same result holds across forecasters though not across variables. For eleven of the twenty-one variables studied, the RMSEs of the adjusted forecasts were uniformly smaller. For only three variables—real imports, the import price deflator, and the change in business inventories—were the mechanical forecasts more accurate for more than half of the forecasts.

The discussion so far has pertained solely to one-quarter-ahead forecasts. The remainder of Table 4 describes forecasts with longer horizons. Any conclusions drawn from them are necessarily even more tentative because they are based on fewer, overlapping (and therefore not independent) time periods. Specifically, the data were collected over a three-year period containing eleven to thirteen one-quarter-ahead observations, depending on the forecaster. Because the actual outcomes for the most recent forecasts of longer horizons are still unknown, there have been very few independent outcomes and little assurance that what has been observed so far will hold up in the future.

Taken at face value, this evidence shows adjustments are less helpful for the longer horizons than for the short horizon. Whether measured by the frequency of improvement or the relative size of RMSEs, this conclusion holds. For example, for eighteen of twenty-one variables the average RMSE of the adjusted forecast is lower than that of the mechanical one-quarter-ahead forecast. Of the four-quarter-ahead forecasts, the corresponding figure is fourteen of twenty-one, and the number falls to only eleven of twenty-one for the eight-quarter-ahead forecasts. Broadly speaking, forecasters would appear to have more "extra-model" information about the current quarter, including some high-frequency actual data, than they do about longer forecast horizons.

This general tendency can be attributed entirely to two (#1 and #3) of the four forecasters. For them, adjustments impaired forecast accuracy of their four-and eight-quarter-ahead forecasts more frequently than it improved it. Nevertheless, the RMSEs were reduced by adjustments for about one-half of the variables. In contrast, the judgment of the other two forecasters (#2 and #4) improved their longer-term forecasts at least as often as it did their one-quarter-ahead forecasts. This suggests that different forecasters may attach differing importance to their short-term and long-term forecasts. Short-term forecasts are more important for trading financial assets or for those who judge a forecaster by his most recent forecast of the preliminary data. Longer-term forecasts are more important for longer-range planning, including macroeconomic policymaking.

The impact of model adjustments on a particular variable clearly tends to be uniform across horizons. The adjusted forecasts of short-term interest rates, consumer prices, and residential investment always had lower RMSEs than the mechanical forecasts. With a single exception in each case, the same holds true for long-term interest rates, corporate profits, and the federal government deficit. At the other extreme, adjustments always impaired forecasts of real imports and usually impaired forecasts of the change in business inventories.

For a few variables, however, the degree of improvement from adjustments depends heavily on the horizon of the forecast. Specifically, while the one-quarter-ahead forecasts of short-term interest rates and the federal deficit were improved most frequently, the improvement of their eight-quarter-ahead forecasts was only about as frequent as for other variables. Adjustments improved the one-quarter-ahead forecasts of exports and the GNP deflator about as often as they did all other variables. But, at an eight-quarter horizon, adjustments improved GNP deflator forecasts more often than any other variable (85 percent) while export forecasts were improved less often than any variable (only 30 percent). Forecasters have the ability to keep their models' long-run forecasts of the GNP deflator from running off track, but no such ability for forecasts of real exports.

In this context, it may be worth stressing again that we are not dealing with "pure" judgment or "pure" models. These data measure the relative contribution of judgmental adjustments to mechanically generated, model-based forecasts. Just as the forecaster's adjustments may be influenced by the mechanical forecast, the assumptions behind the mechanical forecast, as well as the model itself, reflect the modeler's judgments.

III. Are Forecasters Too Timid or Too Aggressive? Actual and Optimal Adjustments

The fact that forecasts are usually more accurate when models are adjusted does not imply that the forecasts we see reflect an ideal blend between model and judgment. In some instances, larger adjustments would have improved accuracy even more, but the forecasters "underadjusted" their models. On other occasions, greater reliance on the model relative to the forecasters' own insights would have improved the accuracy of their forecasts; they "overadjusted" their models. This section documents whether forecasters are "too timid" or "too aggressive" about the value of their own adjustments of their models.

The impact of model adjustments can be summarized by the ratio of the adjustments to the error of the mechanically generated model forecast, which will here be called J.

In symbols,

$$(1)\quad J = \frac{\text{Published forecast} - \text{Mechanical forecast}}{\text{Actual outcome} - \text{Mechanical forecast}} = \frac{\text{Adjustment}}{\text{Mechanical forecast error}}$$

or, rearranging,

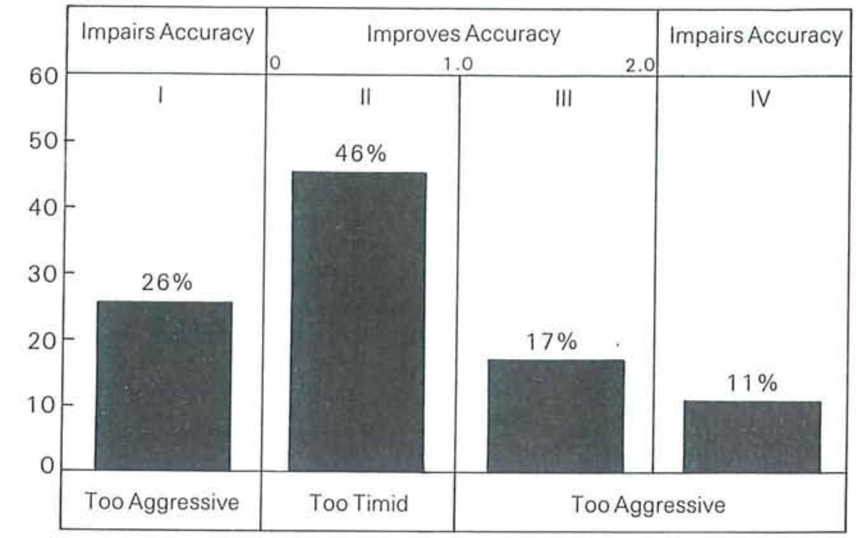
$$\text{Error of Published forecast} = (1 - J)\ \text{Error of mechanical forecast.}$$

Note that $J = 1$ when the adjustment exactly offsets the mechanical forecast error so that the published forecast is perfect. When $0 < J < 1$, the adjustment offsets part, but not all, of the mechanical error, so that the published forecast is more accurate than the mechanical forecast. In these cases, the forecaster was "too timid"; a larger adjustment would have enhanced accuracy. In all other cases, when $J < 0$ or $J > 1$, the forecaster was "too aggressive"; a larger adjustment would have impaired accuracy. When $1 < J < 2$ the adjustment more than offsets the "model error" but still improves upon the mechanical forecast. Thus, adjustments improve forecast accuracy when $0 < J < 2$.

When $J > 2$, the adjustment is in the correct direction but too large, rendering the published forecast error larger (and of opposite sign) than the mechanical forecast error. When J is negative, the adjustment is in the wrong direction, compounding the "pure model" error. Thus when J is either negative or exceeds 2, judgment impairs forecast accuracy.

Chart 7

Distribution of J for All One-Quarter-Ahead Forecasts



$$J = \frac{\text{Adjustment}}{\text{Mechanical Forecast Error}} = \frac{\text{Published Forecast} - \text{Mechanical Forecast}}{\text{Actual} - \text{Mechanical Forecast}}$$

Chart 7 presents a frequency distribution of the 841 Js of the one-quarter-ahead forecasts of all forecasters and all variables. Nearly two-thirds (63 percent) of the Js fall into the $0 < J < 2$ region, reflecting the fact that adjustments improved the forecasts more often than not. However, more than one-quarter of the adjustments were in the wrong direction and about one in ten were so large that they actually impaired the accuracy of the forecast. Less than half (46 percent) of the Js fall between 0 and 1, indicating forecasters are usually not "too timid" in adjusting their models.

The top panel of Table 5 summarizes the distribution of Js for each forecaster and three horizons. At each horizon, the forecasters as a group were "too aggressive." More often than not, a smaller adjustment would have improved accuracy more than the actual adjustment that was made. This tendency toward overly aggressive adjustments is mainly attributable to two of the four forecasters. It is clearly evident in forecaster 1, whose frequency rose from 63 percent at the one-quarter horizon to 75 percent at the eight-quarter horizon. Forecaster 3 also tended to

Table 5

Optimal Adjustments by Forecast Horizon

	Forecast Horizon (quarters)		
	1	4	8
	Too Aggressive (Percent of forecasts)		
Forecaster 1	63*	65	75
Forecaster 2	48	54	52
Forecaster 3	55	64	60
Forecaster 4	55	49	46
All	54*	57	55
	Optimal Weight > (<)1**		
Forecaster 1	3 (8)	1 (8)	2 (8)
Forecaster 2	3 (7)	2 (7)	5 (9)
Forecaster 3	5 (7)	2 (13)	2 (11)
Forecaster 4	2 (10)	9 (6)	8 (5)
All	13 (32)	14 (34)	17 (33)

*Significantly different from 50 at the 90 percent confidence level for one-quarter-ahead forecasts. The four- and eight-quarter ahead forecasts are not independent and no statistical test was applied. There are approximately 200 one-quarter-ahead forecasts, 150 four-quarter-ahead forecasts and 80 eight-quarter-ahead forecasts.

**Significantly different from 1 at the 50 percent confidence level. See Table 4 for total number of variables for each forecaster.

overadjust his model, with frequencies ranging from 55 percent to 64 percent. The other two forecasters do not exhibit a clear tendency to overadjust more frequently than they underadjust. Some forecasters clearly tend to be overly confident of the value of their own adjustments relative to their models.

The tendency to overadjust was most evident among the several variables where, as we have already seen, adjustments impaired accuracy. The clearest case (not shown in the table) was real imports, where adjustments were excessive in from 69 percent to 80 percent of the cases, depending on the horizon. Forecasters also tended to be overly aggressive in adjusting estimates of the change in business inventories and the import price deflator. The few counterexamples of underadjustment, or "timidity," were for corporate profits and the federal deficit one quarter ahead, and short-term interest rates at the longer horizons.

More often than not, forecasters could improve accuracy by placing less weight on their own adjustments relative to their mechanically generated model forecasts.

We have learned that forecasters' adjustments improve accuracy more than they impair accuracy but that their adjustments are too large more often than they are too small. We turn now to what an ideal adjustment would have been, or, more concretely, the optimal, fixed weight that the forecasters should have placed on adjustments in order to maximize the accuracy (based on a quadratic loss function) of their forecasts.

In symbols, we calculate the W that minimizes the expression

$$(\text{Mechanical forecast} - \text{Actual outcome} + W \cdot \text{Adjustment})^2.$$

A weight of 1 implies a perfect blend between model and adjustments. A weight greater than 1 implies that the forecasters were too timid, too "anchored" to their models, too skeptical about the predictive value of their own insights. In contrast, weights of less than 1 imply the forecasters were too aggressive, overadjusted their models, and had undue confidence in their own extra-model insights. A negative weight suggests the adjustments would have to be reversed in order to improve accuracy.

The optimal weights were calculated by regressing the error of the mechanical forecast on the forecaster's adjustment. (The following conclusions also hold when a constant is included in the regression.) A clear majority of these weights were less than 1, confirming the impression of overadjusting. Many of the estimated weights were, however, quite close to the ideal of 1. In order to allow for the imprecision in the estimation of the ideal weight, the bottom panel in Table 5 summarizes only those Ws that were statistically significantly different from 1 at a 50 percent level of confidence. The table again confirms the predominance of weights less than 1, showing that better forecasts would have been obtained by sticking closer to the model. At the one-quarter horizon, thirty-two weights were significantly less than 1 and only thirteen significantly more than 1. At the four-quarter horizon the proportion increases—thirty-four were less than 1 and fourteen greater than 1. The prevalence of overadjusting is overwhelming if one counts only those weights that are significantly different from 1 at a 90 percent level of confidence. More often than not, forecasters could improve accuracy by placing less weight on their own adjustments relative to their mechanically generated model forecasts.

IV. Conclusion

This article has argued that the man (or judgment) versus model dichotomy is a false one. The question is not whether human judgment is either always "good" or always "bad." Everyone is all too familiar with examples of each. The important questions are: Under what circumstances is judgment most likely to incorporate information with predictive value above and beyond that which has already been incorporated in a formal model? How can the imperfect information from judgment be combined with the imperfect information in models to maximize predictive accuracy? Models and judgment are not mutually exclusive, and can be complementary.

Social psychologists have carefully and extensively documented and replicated examples of the fallibility of judgment. Their findings recommend a healthy skepticism toward those who offer forecasts based purely on experience or expertise with no explicit form of systematic reasoning or model. They do not, however, show that judgments are in some sense devoid of predictive content.

At the same time, the limitations of models of an entire economy seem so painfully obvious that it is difficult to imagine that anyone could seriously expect them to incorporate all information with predictive content. The economy is more or less continually buffeted by a variety of events if not entirely unique, at least without close historical precedent. Most recently, it has been drought, earthquake, freeze, and abnormal warming. A decade ago, rather than natural disasters it was strikes, oil shocks, and the phasing in and out of wage, price, and credit controls. At other times, institutional and regulatory changes have led to changes in the very definition of macroeconomic concepts. Such events have implications for the economic future, implications not well captured by standard models which, of necessity, describe the "normal," past behavior of the economy. The evidence presented here broadly confirms the conclusion that individuals adjust their models to compensate in part for their models' deficiencies, thereby improving the accuracy of their forecasts.

The fact that adjustments usually enhance forecast accuracy does not imply that the adjustments that are actually made are optimal for maximizing accuracy. At one extreme, the unadjusted forecasts of real imports were always more accurate, and those of the change in business inventories usually more accurate, than the widely publicized adjusted ones.

More generally, with the benefit of hindsight we can show that the adjustments that were made were typically too large. The prevailing tendency to place too much weight on the specific circumstances and too little on the model is what Kahneman and Tversky (1982, p. 416) have called "the major error of intuitive prediction." It is important to stress that this tendency has been observed in a highly realistic situation. In contrast to some of the previous experiments, all of the subjects were highly trained and informed, had a major incentive to do their best, could eventually learn the actual outcome, and were relatively immune from the treatment effects that contaminate the interpretation of many social science experiments.

While it would be a mistake to ask forecasters to refrain from adjusting their models, it is also a mistake to accept the adjustments that are made at face value, especially when the adjustments appear without any explanation of the reasoning behind them. On occasion, forecasters may have objectives other than simply maximizing the accuracy of their forecasts. This possibility strongly suggests that both forecasters and forecast users should be aware of the differences between their adjusted and unadjusted forecasts. The first step in learning how to extract the most predictive information from both model and forecaster is to be clear about how the two may differ.

References

Armstrong, J. Scott. 1985. *Long-Range Forecasting, from Crystal Ball to Computer.* New York: John Wiley & Sons, Inc.

Evans, Michael K., Yoel Haitovsky and George I. Treyz. 1972. "An Analysis of the Forecasting Properties of U.S. Econometric Models." In *Econometric Models of Cyclical Behavior*, Bert G. Hickman, ed., pp. 949–1139. National Bureau of Economic Research, Studies in Income and Wealth, no. 36. New York: NBER.

Fair, Ray C. 1984. *Specification, Estimation, and Analysis of Macroeconomic Models.* Cambridge, MA: Harvard University Press.

Hacking, Ian. 1984. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference*, esp. pp. 10–17. New York: Cambridge University Press.

Haitovsky, Yoel and George I. Treyz. 1972. "Forecasts with Quarterly Macroeconomic Models, Equation Adjustments, and Benchmark Predictions: The U.S. Experience." *Review of Economics and Statistics*, vol. 54, August, pp. 317–25.

Kahneman, Daniel and Amos Tversky. 1982. "Intuitive prediction: Biases and corrective procedures." In *Judgment Under Uncertainty: Heuristics and Biases*, Kahneman, Slovic and Tversky, eds., pp. 414–21. New York: Cambridge University Press.

Litterman, Robert B. 1979. "Techniques of Forecasting Using Vector Autoregressions." Working Paper 115, Federal Reserve Bank of Minneapolis.

———. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics*, vol. 4, pp. 25–38.

Lucas, Robert E. Jr. 1976. "Econometric policy evaluation: a critique." In *The Phillips Curve and Labor Markets*, K. Brunner and Allan H. Meltzler, eds., Carnegie-Rochester Conferences on Public Policy, 1.

McNees, Stephen K. 1986. "The Accuracy of Two Forecasting Techniques: Some Evidence and an Interpretation." *New England Economic Review*, March/April, pp. 20–31.

———. 1988. "How Accurate Are Macroeconomic Forecasts?" *New England Economic Review*, July/August, pp. 15–36.

Nelson, Charles R. 1972. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." *The American Economic Review*, vol. LXII, no. 5, December, pp. 902–917.

———. 1984. "A Benchmark for the Accuracy of Econometric Forecasts of GNP." *Business Economics*, April, pp. 52–58.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica*, vol. 48, January, pp. 1–48.

———. 1989. "A Nine-Variable Probabilistic Macroeconomic Forecasting Model." Discussion Paper 14, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

Wallis, K. F. 1989. "Macroeconomic Forecasting: A Survey." *Economic Journal*, vol. 99, pp. 28–61.