COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 717-R

EMPIRICAL TESTS OF THE RATIONALITY OF ECONOMIC FORECASTERS: A FIXED HORIZONS APPROACH

William D. Nordhaus

and

Steven Durlauf

August, 1984

Revised: November, 1984

EMPIRICAL TESTS OF THE RATIONALITY OF ECONOMIC FORECASTERS:

A FIXED HORIZONS APPROACH

— William D. Nordhaus
Steven N. Durlauf

Section 1. Introduction*

The rationality of economic agents has become one of the most hotly debated issues in macroeconomic analysis. The potential implications of the rational expectations (RE) assumption for such issues as stabilization policy are too well known to require enumeration. Whereas an extraordinarily rich body of theoretical work on RE has been produced, examinations of the validity of the RE assumption have been far less extensive. The stringent implications of RE for policy efficacy demand that the nature of information processing itself be scrutinized. A major task for empirical work is thus to test rationality as an operational assumption. This paper attempts a new examination of the empirical validity of rationality which we hope will provide a number of new insights into the nature of information processing.

Previous attempts to test the rationality of expectations have all relied upon the same general model for testing. These tests are all based upon fairly weak forms of RE. In essence, such tests examine whether forecast errors by economic agents are unpredictable. A time series of expectations is obtained either from surveys of economic agents or from data realizations.[1] Then, the following regression is run:

(1) $$e(t) = c + bP(t) + v(t)$$

where $e(t)$ equals expectational errors

c equals a constant

$P(t)$ equals prediction of the event

$v(t)$ equals regression residual.

Unbiasedness in expectations requires that $c = 0$ and $b = 0$, i.e., the expected value of the forecast error is zero.

In slightly more sophisticated versions of rationality testing, one may test models of the form:

$$(2) \qquad e(t) = B' X(t) + v(t)$$

where $e(t)$ = forecast error

$X(t)$ = column vector of variables observable at time forecast is made

B' = row vector of coefficients.

Fully efficient information processing requires that expectational errors be independent of all information available at time $t-1$, i.e. B' equals zero. If the information set defined by $X(t)$ consists of a constant and the forecast itself, this general efficiency test is identical to the unbiasedness test described above. If there exist elements of the information set which are independent of both the constant and the forecast yet which are not independent of the expectational error, then the general efficiency test may fail even when the expected value of a prediction equals the actual event. When the information set is restricted to a constant and the current or lagged set of forecasts, the tests examine what is commonly called the "weak version" of RE. When other, publicly available regressors are included, the tests examine the "semi-strong" form of RE. Thus all rationality tests are reducible to regressions of forecast errors on agents' information sets where these sets are defined as observable time series.[2]

A number of studies over the last few years have employed this framework to analyze expectational data sets. These tests have in general provided mixed results. In analyzing the rationality of the Livingston Surveys, Gramlich (1983), Figlewski and Wachtel (1981) and Pesando (1975) have all rejected RE in finding that unbiasedness of the survey expectations for inflation, output growth and interest rates is inconsistent with the data. The RE hypothesis has found more comfort from such work as Mullineaux (1978) and Mishkin (1981). The Mullineaux results demonstrated the heteroscedasticity in the error structures across agents may bias survey rationality tests towards rejection. Mishkin constructed an artificial expectations series for holding returns on long bonds by assuming that these expected holding returns differed from the known holding returns on short bonds exclusively by a constant risk premium. (Holding returns on a bond are known if the maturity is shorter than the observation period, thereby eliminating the potential for capital gains.) The difference between long and short realized holding returns thus will be equal to the expectational errors plus a constant. Analyzing the expectational error on government bonds, Mishkin found that these expected holding returns did in fact fulfill the unbiasedness and efficiency conditions for rationality.

This paper seeks to improve upon these previous studies in two important respects. First, the objects of our scrutiny will be the forecasts of several major forecasters in the American economy, including Data Resources Incorporated, the UCLA forecasting model, Wharton Economic Forecasting associates, Chase Econometrics and the Eggert Consensus. Utilization of the forecasts of these groups, we think, will allow us to avoid the Scylla and Charybdis of irrelevant and imprecise expectational measures. On the one hand, the use of massive surveys along the lines of the Livingstone series is likely to bias the tests of rationality improperly towards rejection (in the sense of examining the rationality of macroeconomically irrelevant expectations).[3] If some agents have irrational expectations and others rational ones, arbitrage opportunities will make the rational agents' expectations relevant from a policy perspective.[4] On the other hand, the use of artificial measures of expectations will induce a larger variance in the error of equation (1) thereby reducing the power of rationality tests and incline the hypothesis tests towards acceptance.

The use of major forecasters as the representative agents will ensure that the agents examined are those most likely to have the most accurate expectations. As the success of these services is presumably a function of accuracy, and given the differing nature of risk premia across the customers of these services, these forecasts are the most likely to be impervious to adjustments consistent with the internal interests of the forecaster yet inconsistent with pure statistical rationality.[5] In other words, individual agents might have asymmetric loss functions which generate deviations from statistical rationality. As a forecaster caters to many different customers, it is likely the customers would prefer an unbiased forecast which the buyer can then adjust. With disparate buyers, a forecaster could not perform the adjustment, as disparate buyers are likely to have different loss functions. Hence, one would expect the performance of these forecasters to define an upper bound on the expectational accuracy of agents in the economy as a whole.

Second, we hope in this paper to develop a new methodology for empirically analyzing expectational rationality. This new methodology should act as a complement to the standard rationality tests. In fact, some basis may exist for believing these new tests possess greater power in testing forecast rationality. Section II below will develop this methodology in detail. Briefly, rather than examine the relationship between a series of one step ahead or n-step ahead forecasts and their realizations, we will examine the behavior of a sequence of forecasts of the same event. Rationality requires that these forecast adjustments be unpredictable over time. Analysis of these forecast adjustments will provide insights into the structure of information processing as new information is made available over time.

The organization of this paper is as follows: Section II will develop the methodology concerning the relationship between forecast adjustments and rationality. Section III will present a number of tests of the rationality of the macroeconomic forecasters based upon the results just derived, including an analysis of partial adjustment models.[6] Section IV will analyze the potential sources of irrationality in the forecasts. Optimal forecasts will be derived based upon the implicit time series properties of the actual forecasts. Section V will provide a summary and conclusions.

Section II: The Martingale Property of Fixed-Horizons Forecasts[7]

The analysis in this paper will focus on the critical property of optimal forecasts: a sequence of rational forecasts of a given model made over subsequent time periods must follow a martingale. Formally,

$$E_{t-1}(P_t) = P_{t-1} \tag{3}$$

where E_{t-1} represents the mathematical expectation at $t-1$. A time series with this property is of course a martingale. Correspondingly, the adjustments in forecasts of a given event must be unpredictable at t-1 . First, we shall prove this property formally, and then provide some intuition as to why this result holds.

Proof That Fixed-Horizon Forecasts Fluctuate Randomly

Define a sequence of information sets ϕ_i such that $\phi_{i-1} \subset \phi_i$. Thus information increases with i . For notational purposes, define the realization of the event we wish to forecast as $E(Y|\phi_T)$ where T is the time at which the event is actually realized. Let

$$E(Y|\phi_T) - E(Y|\phi_{T-1}) = \eta_1 .$$

If the forecast error is unpredictable, then

(4) $$E(\eta_1|f(\phi_{T-1})) = 0 \quad \text{for any function} \quad f(\ ) .$$

Now, consider

(5) $$E(Y|\phi_T) - E(Y|\phi_{T-2}) = \eta_2 .$$

Again, rationality requires

$$E(\eta_2|f(\phi_{T-2})) = 0 \quad \text{for any function} \quad f(\ ) .$$

Given these definitions, we can consider the forecast adjustment

$$E(Y|\phi_{T-1}) - E(Y|\phi_{T-2}) .$$

Add and subtract $E(Y|\phi_T)$ from forecast adjustment, yielding

(6) $$-E(Y|\phi_T) + E(Y|\phi_{T-1}) - E(Y|\phi_{T-2}) + E(Y|\phi_T) = -\eta_1 + \eta_2 .$$

But $E(-\eta_1 + \eta_2|f(\phi_{T-2})) = 0$ by assumption of rationality.

Because we could have replaced T-1 and T-2 with T-i and T-j for any $i < j$ in the equation, this shows that $E(-\eta_i + \eta_j | f(\phi_{T-j})) = 0$. Hence the conditional forecast of P_{T-i} is P_{T-j} at time (T-j) earlier than (T-i) ; that is, the forecasts of the event P_T are a martingale.

Intuition

The intuition for this result is quite straightforward. The forecast at t of a particular event is superior (in a minimum variance sense) to that at t-1 as more information is available at t . If one could predict the movements between t-1 and t , one would immediately incorporate this information into the t-1 forecast in order to move the forecast toward the superior one.

Again, we emphasize that the nature of the data set which we shall employ differs from the data employed in the standard test. Rather than looking at a time sequence of terminal and near terminal forecasts (which we define as rolling horizon tests), we shall be examining the relationship of the sequence of a large number of forecasts for a small number of independent events (which we define as fixed horizon tests). Therefore, our results will not so much test the unbiasedness of expectations over time but rather the ways that informational shocks affect forecasts. Thus, our results are complementary to the standard time series analysis.

The full stochastic model of forecasts provides some insights into the sources of statistical bias which may plague the fixed horizon and rolling horizon analyses. Specifically, the critical issues from a regression perspective concern the stationarity and normality of the forecast adjustments under the null hypothesis that the forecasts are in fact rational. If the underlying model of the economy is changing over time, the forecast adjustments are likely not to be stationary. This problem of heteroscedasticity can adversely affect the behavior of hypothesis tests, as noted above. The fixed horizon test will be more susceptible to heteroscedasticity if the flow of information shocks is lumpy in the sense that different time intervals (defined by proximity to the event realization) will contain different amounts of information. For example, there may be relatively little information about real GNP growth accruing each month during the period two or three years before the event. On the other hand, in the year immediately prior to the event, a great deal of information may become available.

This "fixed horizon" approach to rationality testing, we believe, is particularly interesting for a number of reasons. First, by analyzing rationality on a year-by-year basis, we shall be able to identify those periods when information processing is particularly inefficient. The relationship between these periods and the stability of economic structures (such as government policy) during these periods may potentially shed light on the adaptability of agents' expectations to new stochastic environments --an ability which lies at the core of both the Lucas critique of standard econometric inference and the Sargent/Wallace policy ineffectiveness theorems. Second, the fixed horizon approach to expectations formation may prove to be a more powerful test of rationality than the standard time series tests.[9] If the statistical inefficiency in the expectation formation of agents lies in lags in the processing of information, for example, then it is likely that agents' expectations will appear to be rational as the time gap between forecasts and actual events becomes narrow. However, the inefficacy of policy in many models requires rationality of expectations over quite long horizons.[10] In general, it seems much more plausible that information processing be inefficient yet unbiased in predicting an uncertain event than that a sequence of forecasts fulfill a set of internally consistent statistical conditions yet be biased even as the forecast times coverage with the event in question.

Thus, we find that the standard rolling horizon analyses of rationality are actually a subset of the general properties of optimal forecasts. Our tests of rationality will augment the current battery of rationality tests most frequently seen in the literature.

Section III. Parametric and Nonparametric Tests of Martingales

A. Description of Data

As mentioned in the introduction, the object of our concern will be the predictions of a number of leading macroeconomic forecasters: DRI, Wharton, Chase, UCLA, and the Eggert Consensus. These forecasters span the range of academic and business concerns. The specific forecasts which we shall be concentrating on are the predictions of annual growth of real GNP and the GNP deflator for the US in the years 1980, 1981 and 1982. Some attention will be paid later on to a number of other macroeconomic aggregates and other forecasters.

The construction of the sequence of forecasts was done by Eggert Economic Enterprises, Inc. For each month of the sample period, Eggert receives at the beginning of the month the forecast for the annual variable in question. Eggert then constructs a consensus forecast by taking an unweighted average of all forecasts which have been received for the month. The time lag between the actual construction of a forecast and the month for which the forecast is reported appears to be uniform both across forecasters and across time periods.[11] Thus we may take the underlying forecasts to be roughly contemporaneous--which will prove essential in the subsequent estimations.

B. Interpretation of the Data

In analyzing the martingale property of these forecasts, it is useful to view the adjustment of forecasts graphically. In Figures 1, 2, 3 and 4 we find graphs of the adjustments in forecasts for the year 1982. Under the assumption of rationality, these adjustments must have zero mean and be independent over time. As is visually clear, these properties are not borne out by the data.

A number of features of these graphs deserve particular mention. First, the 1982 forecast adjustments appear to be overwhelmingly negative. The adjustments clearly do not have a zero mean. All the forecasts took a sharp plunge in late 1981. The plunge coincided with both a loss of confidence in the rosy forecast of the Reagan Administration (which, indeed, had forecast errors far worse than those of the private forecasters); and with a sharp decline in Industrial Production starting in September 1981. Barring a dramatic outlying event, the clear interpretation of the forecast errors of 1981-1982 is that most major forecasters misunderstood the extent to which monetary policy was exercising a contractionary effect on the economy. The extraordinary optimism was not, however, translated into a real expansion.

A second important observation is that for 1982, the consensus forecast generally appears less rational than the individual forecasters, in terms of deviations from the martingale hypothesis. The inferiority of the consensus is consistent with the possibility that the majority of forecasters are guilty of ignoring potentially useful elements of their information sets. This fact, of course, does not tell us anything about the relative accuracy of the consensus versus the major forecasters, especially if each individual forecaster has access to some idiosyncratic piece of information concerning the economy. We shall return to the role of the consensus as an aggregator of information later below.

C. Regression Tests of Rationality

We now turn to examine whether the visual impression of irrationality is borne out by formal statistical tests. In this section we shall test the hypothesis that the forecasts are a martingale; that is, whether the forecast revisions from one month to the next have zero mean and are unpredictable on the basis of then-available information.

In the regression tests, we shall make the standard assumptions of identical and independently distributed (i.i.d.) normal errors. We thus examine a regression of the form:

$$e(t) = B'X(t) + v(t) \tag{7}$$

where $e(t)$ = forecast adjustment, $X(t)$ = set of variables observable at time $t-1$ forecast is made, and $v(t)$ are the i.i.d. errors. The rationality hypothesis assumes that all coefficients, B', are zero.

A key aspect of any test of rationality is the set of variables --the $X(t)$ --to employ when examining the independence of forecast adjustments. Two particular variables will be examined below; these variables are ones that are specific to the fixed-horizon approach and cannot easily be tested in the customary rolling-horizon models. First, we examine whether there is a constant adjustment value; second, we test for a relationship between current and past forecast adjustments. These two central tests will focus on whether the <u>forecast adjustments</u> themselves are rational--that is, whether forecasters tend to smooth their forecasts to minimize the complaint of "jumpiness" or "inconsistency" in forecasting.

If forecasters experience some loss of credibility when their forecasts jump around a great deal, they may tend to smooth out forecast adjustments over more than one forecasting period; in this case, the adjustments would tend to have some serial correlation, and the rationality hypothesis would be violated.

The interpretation of the significance of these variables is related to the structure of the underlying adjustment process. For example, if the forecast is generated by the minimization of a quadratic loss function, then only the lagged forecast adjustment will be significant. Specifically, suppose that the forecaster is penalized when his forecasts change rapidly. For example, the forecaster might perceive the following loss function:

$$L = a(p(t) - p(t-1))^2 + (1-a)(p(t) - E_t(p(t))^2 \tag{8}$$

where $p(t)$ = actual forecast at t

$E_t p(t)$ = mathematical expectation of $p(t)$ at t

a = weight.

First order conditions for a minimal loss imply that the current forecast will equal

$$p(t) = ap(t-1) + (1-a)E_t(p(t)) . \tag{9}$$

Differencing this equation will yield

$$p(t) - p(t-1) = e(t) = ae(t-1) + (1-a)\Delta E_t[p(t)] \tag{10}$$

Note that $e(t)$ corresponds exactly to the notation for forecast changes used earlier. However, $\Delta E_t[p(t)]$ is not observable. But by rationality it is orthogonal to the information set available at $t-1$, of which

$e(t-1)$ is an element. Therefore, equation (10) represents a consistent regression which will reveal the coefficient of adjustment $(1-a)$.

In what follows, we find some indication that the irrationality in forecasts takes the form of a non-zero constant in equation (7). This behavior is somewhat unorthodox, so we explore briefly possible reasons for adjusting forecasts by a constant amount rather than proportionally, as in the norm for quadratic loss functions.

One possibility is that the loss function itself is of some unusual form. Suppose, for example, that the loss function is quadratic in deviations from the mathematical mean and linear in absolute magnitude of adjustment. Further, suppose that the adjustments each period do not impose any cost if they are below a certain magnitude. This might arise if a forecaster's clientele were accustomed to small changes, but became alarmed at large changes. The loss function, L', would then resemble:

$$L' = \begin{cases} a(p(t) - E_t[p(t)])^2 + b|p(t) - p(t-1)| & \text{, if } \mathrm{abs}(p(t) - p(t-1)) < k \\ a(p(t) - E_t[p(t)])^2 & \text{, otherwise} \end{cases}$$

where a, b are weights and abs( ) is the absolute value operator.

The behavior of this system will be highly nonlinear. However, we can get some clues as to its likely behavior. For a $E_t[p(t)]$ series which is stationary and normal, the $p(t)$ series will be similarly normal for small movements over time. Once a large shock hits, the $p(t)$ forecast will adjust to the point where

$$2a|p(t) - E_t[p(t)]| = b .$$

After the initial large adjustment, the subsequent adjustments will be equal in magnitude to k at most.

Thus in periods where there is a single large information lump, such as a large oil-price shock, a sharp change in monetary operating procedures, or the failure of supply side economies to live up to its promises, one would find a large adjustment, followed by a sequence of smaller shocks in the same direction. A forecast path of this kind would imply an adjustment path which is dominated by adjustments with an absolute valued mean equal to k.

The results of our regression tests are presented in Tables 1 through 4. Table 1 shows regression tests on the constant terms, while Table 2 indicates those regressions for which the constant terms are significant at the 5% level. It is apparent that the coefficients are significant for nearly every forecaster for the years 1978 and 1982. These years are significant in that they represent the years in our sample when the entire sequence of observations follows a shift in presidential administration. Apparently, the ability of forecasters to rationally predict conditions during a shift in regime--a major assumption in the policy neutrality literature--comes into clear question. It is striking that in both cases the new administration generated excessively optimistic forecasts of GNP growth, as can be gleaned from the output constant coefficient estimates which are reported in the upper half of Table 1. Evidently, forecasters are willing to follow the tide of public opinion --at least in times of the optimism which a new administration ushers in.

Another striking result concerning the output coefficient estimates is the within year similarity across forecasters. The 1982 adjustment rates all are near -.25 whereas the 1978 estimates seem to center around .1. The -.25 level would seem to be plausible in terms of a fixed cost to forecast changes, as discussed above. However, the small magnitude of the 1978 coefficients makes it unlikely that a fixed barrier to the magnitude of forecast adjustments generated this result. After all, the 1979-1982 forecast adjustments very frequently are of a magnitude greater than .1. The smoothing implicit in the 1978 regressions most probably is generated by some sort of asymmetric loss function which is dissimilar in form from those discussed above.

The adjustments in inflation forecasts provide even more dramatic verification of the absence of rationality in forecast adjustments. Irrationality appears frequently in the 1979 and 1980 forecast adjustments in addition to the past regime change years. With the exception of 1982, when the coefficients were negative, the forecasters have tended to systematically underestimate the magnitude of inflation growth.

One interesting result is the general consistency of the magnitude of the inflation constant across the different years as well as the different forecasters. The magnitude of .1 is relatively consistent with the hypothesis of a fixed adjustment cost: we do not see years where every adjustment is of a substantially greater magnitude.

The poor performance of the rationality of the inflation forecasts corresponds, of course, to the generally poor forecasting performance of inflation. A possible explanation of the poor inflation regression results may be that in the absence of a well accepted and useful theory of inflation behavior, forecasters have relied upon intuition rather than statistical procedures in constructing inflation forecasts. If this is so, the presence of statistical irrationality may be one of the inadvertent results.

Table 3 reports the results of regressions where a pure partial adjustment model is used in place of a constant adjustment model. A number of features stand out. First, there is marked evidence of irrationality; 14 of 50 tests show a significant effect of the lagged forecast adjustment on the current period's forecast adjustment. Partial adjustment tends to appear more powerfully for real GNP growth than for inflation. Smoothing of forecast adjustments is most apparent for the Consensus. In recent years, the adjustment is less than fifty percent. The other four forecasters have (an unweighted) average rate of adjustment of 32 percent per month. Moreover, 24 of the 25 real GNP forecast adjustments are positive, indicating that there is indeed some implicit cost of having jumpy forecasts.

The results for inflation are less clear cut. The Consensus indicates a definite pattern of partial adjustment of forecasts, with somewhat less than one-half of the adjustment occurring in the first month. DRI also shows some evidence of partial adjustment, while the other three forecasters have a mixed sign and significance pattern.

A couple of additional points should be made in terms of the significance of the constants in our regressions. To check on the problem of nonstationarity in the adjustment structure, the regressions were rerun dropping firstly the largest observation in absolute magnitude and secondly the largest three observations in absolute magnitude. In neither case did the significance of the coefficients waver. If a nonstationarity were induced due to information shocks with a far larger magnitude in some periods than others --we failed to detect it. Actually, the relative robustness of the sample mean hypothesis tests follows directly from the form of the t-statistic. Non-stationarity really means that the sample variance improperly estimates the true variance of the sample mean. Increasing the variance of a single observation by a factor of say 10 will not appreciably underestimate this value when we have 25 or so observations as is apparent from the definition of the sample variance.

Clearly, the rationality estimates for a given year are independent neither across forecasters nor across inflation and output growth. However, estimation of dependent regressions jointly using Zellner's SUR techniques failed to appreciably affect the coefficient results, and thus is not reported.

Finally, the pooling of the data on forecasts of a single event failed to provide any new information. The 1982 forecast adjustments completely dominated the pooled regressions. Partial adjustment significance was completely lost. This should not have been unexpected due to the fact that the magnitude of the 1982 forecast adjustment was much larger and much more irrational (in the sense of coefficient magnitude) than other years.

Our single event tests, then, reveal strong evidence of irrationality for the 1982 and 1978 output growth forecasts and overwhelming evidence of irrationality for the inflation forecasts across all years.

D. The Role of the Consensus Forecast

In recent years, many consumers of forecasts have come to rely upon a portfolio of forecasts (such as the Eggert "consensus forecast") rather than upon a single forecast. In the strict sense, the existence of a wide variety of different forecasts is inconsistent with formal models of rational expectations. Most rigorous models assume that there is a "true" model, and that all agents know the true model. In such a world, of course, all econometric models would be the same; and, unless some people had different information than others, all forecasts would be the same.

In reality, there is a wide variety of forecasting models and of actual forecasts. Retrospective analyses, such as those of McNees, indicate that no forecaster has a dominant model; some do better in some periods, others in other years.

Given the diverse forecasts available, we ask in this section whether in some sense the Consensus is aggregating information of individual forecasters. If such aggregation is occurring, then we may detect a pattern of forecast adjustment to lagged changes in the Consensus forecast.

We can make these points formally by examining the exact relationship between the consensus and the sufficient statistic for processing information. The conditional probability density function for an outcome y may be written

$$f_t(y|\phi_{t-1}) \ ,$$

where ϕ_{t-1} represents the realization of all relevant variables at $t-1$. If the complete set of information ϕ_t can be reduced down to a specific piece of information or sufficient statistic-- $T(\phi_{y-1})$ --which forecasters at time t employ, then

$$f_t(y|\phi_{t-1}) = f_t(y|T(\phi_{t-1}))$$

Our interest in the consensus lies in determining a relationship between the consensus and the $T(\phi_{t-1})$ variable. Ideally, the consensus may function as an approximation of the sufficient statistic and therefore be of use to a forecaster. The presence of idiosyncratic information on the part of any forecaster will find its way into the consensus and thus make the consensus useful to other forecasters.

We make these general observations in order to provide a background for the next set of experiments which we shall examine. Specifically, we shall accept for the moment that each forecaster is conditionally rational. Thus, we would like to find out something about the sources of their forecast discrepancies. An appealing hypothesis is that the consensus forecast functions as an indirect yet powerful aggregator of information across agents. One way of verifying this proposition is to test whether the individual forecasts behave as martingales after the lagged concensus forecast is included in the regression.

For these tests, we assume that the consensus is an optimally constructed sufficient statistic or aggregator of all individual forecasts. Suppose that at time $t-1$, the consensus were the sufficient statistic for the probability distribution of the variable to be forecasted, conditional on all information at time $t-1$. Further, suppose that the consensus is public knowledge with a 1 period lag. Then the optimal predictor of a forecast at time t would be the lagged consensus rather than the lagged own forecast. This follows, because from the perspective of $t-1$, there does not exist any old information beyond the consensus which could provide potential benefit to the forecast. Even if the consensus is not a perfect aggregator, if the consensus provides useful information, then the only value of the own lagged forecast to predicting the current forecast is in adjusting the weight of the own forecasts in the consensus weighting scheme. Therefore a natural regression to examine is:

$$p(t) = c + b_1^* p(t-1) + b_2^* \mathrm{CONS}(t-1)$$

where $p(t)$ equals own forecast $\mathrm{CONS}(t-1)$ equals consensus forecast. The significance of the lagged consensus will give us some insight into the extent to which the forecasts function as bootstraps for one another.

Table 5 reports the number of times where a regression including both a constant and the lagged consensus generates failure of the martingale hypothesis. Notice that we are now using levels rather than changes as our variables. As is clear, the inflation forecasts provide strong evidence of the failure of the martingale hypothesis for most years. The output regressions with the exception of 1982 are less inconsistent with the martingale hypothesis.

Table 6, providing the actual regression results, is essential in determining whether the martingale hypothesis fails due to inclusion of the consensus, or rather, due to the inclusion of the constant and lagged forecast adjustment. (Running the level regression runs the risk of the lagged consensus proxying for the omitted constant term.)

In only 7 of the 32 cases is the lagged consensus variable statistically significant by itself. In 16 of the 32 cases the lagged own forecast is statistically significant. In only 2 cases--the 1982 Wharton output growth rate and the 1979 Wharton inflation prediction--are the two variables both significant. In the remaining 11 cases the multicollinearity between the two variables is too great to allow determination of which lagged forecast is driving the regression results. Thus, there is some evidence to support the hypothesis of the consensus functioning as a sufficient statistic.

Two final points should be made. First the 1982 consensus regressions have overwhelmingly rejected the consensus variable as a significant explanatory variable. In the other years, the presence of the consensus appears to be responsible for the failure of the martingale hypothesis. This would appear to make sense in that the 1982 consensus predictions were overwhelmingly the most inaccurate of the years we scrutinized. In 1982 the consensus (based upon a RMSE criterion) was outdistanced by virtually all the forecasters for both inflation and output growth. The other years, however, the consensus was usually more accurate than our major forecasters. Evidently, the presence of erroneous information in the 1982 consensus invalidated its usefulness as an information source.

Second, how are we to interpret the significant coefficients on the lagged consensus? They are not, it must be emphasized, evidence of irrationality of forecasters, for strictly speaking they were not available at the time of the last period's forecast. There are two possible interpretations at this point. First, it may be that individual forecasters have private information that is included in the forecast. In this case, the past adjustment of other people's forecast is useful information for each forecaster. In this case; the consensus is acting as an aggregator of private information.

A second possibility is that forecasters are simply acting in a herdlike fashion--desiring not to find themselves isolated from the rest of the forecasting pack. "There's safety in common forecasting errors," might run such behavior.

The tests we have run do not allow us to discriminate between these two hypotheses. The results suggest, however, that the "herd" hypothesis may have some validity, for the following reason: If people tend to follow the herd, they will be misled when the herd is thundering off in the wrong direction. And 1982 was the year when the herd (i.e. the consensus) performed very poorly. An examination of Table 6 indicates that the coefficients on the consensus were not markedly below the average values in 1982; indicating that people were paying about as much attention to the consensus in 1982 as they did in other years. While this test is only impressionistic, it does suggest that at least part of the dependence of forecasts on the consensus may have been due to forecasters' desires to stay in the middle of the forecasting pack.

Section IV: Optimal Forecasts and Irrationality Criteria

In this section, we shall consider the issue of optimal forecasts based upon the actual sequence of forecasts which we have observed. How would one construct an optimal forecast from the individual forecasts without *a priori* knowledge of the stochastic process generating the forecasts.

There is a wide variety of possible filters to use in constructing an optimal forecast. Two candidates come to mind. First, we shall consider a rolling regression where we assume that

$$p_t = p_{t-1} + c$$

represent the underlying structure. At each point in time, c will represent the sample mean of the forecast adjustments up to t. Thus:

$$p_t^* = p_t + \hat{\mu}(t-T) \quad \text{where} \quad \hat{\mu} = \sum_{i=1}^{t} \Delta p_t / t \ .$$

For the 1982 consensus, this sequence would generate an average squared error as reported as "Filter #1" in Table 7.

The average squared error in the GNP growth equation is improved by about 30%, whereas the inflation errors are improved by only 10%.

A second filter tries to correct for possible misspecification in the estimated forecast adjustment equation. In the forecast sequence, it is apparent that observations fluctuate randomly about some mean until a sequence of information shocks hit in the fall of 1981. If these observations were equilibrium observations, then the rolling regression procedure would erroneously adjust these observations. In order to isolate periods of "irrational" adjustments we postulate the following form:

$$p^* = p_t + \delta_{ij}\hat{\mu}(t-T) \ , \quad \delta_{ij} = \begin{cases} 1 & \text{if 3 consecutive adjustments have the same sign} \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\mu} = \text{sample mean of last three adjustments.}$$

Thus we only extrapolate sample means generated in periods where the adjustments are unidirectional. The column labeled "Filter #2" in Table 7 reports the average mean squared errors of this approach.

Again, this filter provides significant, although unspectacular improvements in the average squared errors. The improvement is slightly inferior to Filter #1.

To understand the relationships between the two filters we have prepared, we have reproduced the forecast sequences along with the actual forecast sequence for real GNP in Table 8. As is clear from the tables, the two filters increase the rate at which information shocks are assimilated into the forecasts. The second filter possesses a tendency to overshoot the realization as is evidenced by observations near the end of 1981. Evidently the degree of information assimilation which occurs after 3 consecutive periods of unidirectional adjustment is sufficiently high to render full extrapolation an overestimate of the degree of irrationality.

In summary, by using straightforward filters that recognize the irrationality of the consensus forecast, we see that it is possible to reduce the variance of the forecast by a healthy margin--and that, after all, is the bottom line in the forecasting business. Will we soon see someone selling an "adjusted consensus forecast" that takes into account the irrationality of the existing consensus?

V. Conclusions

This paper has examined the behavior of 4 major forecasters and the forecast consensus. Using a new technique of "fixed-horizon" models, we first showed that forecast adjustments in this framework would fluctuate randomly. We then examined approximately 1200 forecast adjustments over the 1978-1982 period to examine the statistical properties of forecast adjustments.

The evidence is clear that there are marked and significant elements of statistical irrationality for these major forecasters. Information shocks tend to get processed slowly; hence, after a shock, it takes a few months before the information is completely incorporated into new forecasts. Indeed, for the consensus forecast, only 50 percent of the new information relating to the 1982 recession was incorporated within 5 months; for large forecasters, the average time for incorporation of new information was 1-1/2 months.

The reasons for the deviation of forecasts from statistical rationality are unclear. The pattern of adjustments is consistent with forecasters being averse to "inconsistency," i.e. rapid changes in forecasts. There may also be evidence that forecasters tend to move toward the consensus in herd-like fashion.

FOOTNOTES

*We are grateful to Anne Jenkins and Ellen McGratten for superb research assistance. We thank Sam Ouliaris for helpful comments.

[1]By construction of expectations, we refer to techniques such as equating expectations to least square forecasts generated by the researcher.

[2]The use of the terms unbiasedness and efficiency differs slightly from the definitions usually employed, such as in Mincer and Zarnowitz (1969). The standard definitions are that $c = 0$ constitutes unbiasedness and $b = 1$ constitutes efficiency. Our use of the terms unbiasedness and efficiency are designed to distinguish between misuse of available information and optimal and suboptimal sufficient statistics. If the information sets consists solely of the prediction and a constant, the unbiasedness test is examining whether given information constraints of the forecaster, the information is correctly used. On the other hand, a more natural definition of efficiency revolves around whether the expectational errors can be reduced in variance by use of other publically available information not used by the forecaster. As we do not know what information is actually employed by forecasters outside of their actual forecasts and a constant, the unbiasedness/efficiency dichotomy we employ is appropriate.

[3]This point is discussed in more detail by Mishkin (1981). The main intuition is that the prevailing market equilibrium is determined by the most rational agents just as the level of risk premia on assets is determined by the presence of some risk neutral agents.

One can posit a number of reasons for this phenomenon. First, the expectations which are reported are <u>ex ante</u> ones and not necessarily the the ones which agents actually employ when engaging in market transactions. This would be especially likely if prices in actual markets function as sufficient statistics for the aggregation of information. Alternatively, if each agents' behavior is observable by all members of the market, then information will be transmitted across agents and disparate expectations disappear. Second, in the presents of futures markets and risk neutral rational agents, the market equilibrium would necessarily be determined by those agents. This explanation particularly applies to the area of interest rates.

[4]See Berger and Krane (1982) for a full discussion of this problem in the context of expectational models of the term structure of interest rates. This problem is exactly equivalent to the increased coefficient variance associated with the use of instruments rather than actual variables in regression.

[5]By this argument, we mean to deal with the problem that announced expectations on the parts of agents may not reflect their true mathematical expectations but rather the mathematical expectations adjusted by some sort of risk premia. For example, firms may adjust profit forecasts downwards in order to avoid unfulfilled dividend predictions. The aim of a forecasting service, on the other hand, is presumably to provide information to disparate agents whose internal adjustment decisions are likely to vary. In addition, some customers may desire unbiased expectations. Thus the forecaster will not be in a position to perform the desired adjustment for its customers, and thus will release unbiased forecasts.

[6]By partial adjustment models, we refer to models of expectation formation whereby agents process information subject to some lag structure. For example, permanent changes in government policy are temporarily perceived as transitory adjustments.

[7]An exception to this use of time series tests of rationality is the work of Berger and Krane (1983). They independently derive a set of tests which are equivalent to the fixed-horizons tests which we derive below. Their paper should be read as a complement to the work which we present.

[8]This proof is merely a restatement of the well known property of double conditional expectations. See Chung (1974) for the most general proof. The technique we employ is different from the method of proof employed by Samuelson (1965) in proving the martingale nature of futures prices. We feel our use of expectation operators unlocks the intuition of the result.

[9]The following example shows why the fixed horizon test may be a more powerful test than the rolling-horizon test. Suppose that a forecaster is forecasting an event 3 periods into the future. The event is a white noise process with mean zero, i.e., $E(x_t) = 0$. The forecaster obtains no useful information until two periods ahead of the event, at which time he is able to make a perfect forecast of the event. I.e., $E_{t-3}(x_t) = 0$, but $E_{t-2}(x_t) = x_t$.

The forecaster smooths the forecast so that the forecast adjustments in period (t-2) and (t-1) are equal. That is, the forecast of event x_t made in period t-j (denoted by ${}_{t-j}p_t$) is:

$$
{}_{t-2}p_t = \frac{1}{2}x_t
$$

$$
{}_{t-1}p_t = x_t
$$

An examination of the forecasts will show that any rolling-horizon test will be a martingale [because $E({}_{v}p_{t+\theta} - {}_{k}p_t) = 0$ for all $\theta \neq 0$]. However, the fixed-horizon test will definitely reject the martingale test [because $E({}_{t-1}p_t - {}_{t-2}p_t) = \frac{1}{2}x_t \neq 0$].

[10]The importance of long range foresight in macroeconomic models can be easily seen in such areas as rational expectations models of hyperinflation.

[11] The definition of cross-sectional forecast accuracy which we employ is the RMSE of the sequence of forecasts from the eventual result. This calculation implicitly attaches greater weight to the earlier forecasts, as the expected forecast error is declining over time. (As one approaches the event, the potential variance in prediction will become quite small.) We feel that this weighting is justified as the value of forecast accuracy will be higher when this accuracy can be attained early. Investment decisions which require knowledge of GNP growth will be best augmented by early accuracy. Also, the informational value of the macro forecasts will diminish as the event draws near as information concerning the event moves more and more into the realm of free publicly available information such as government estimates.

FIGURE 1

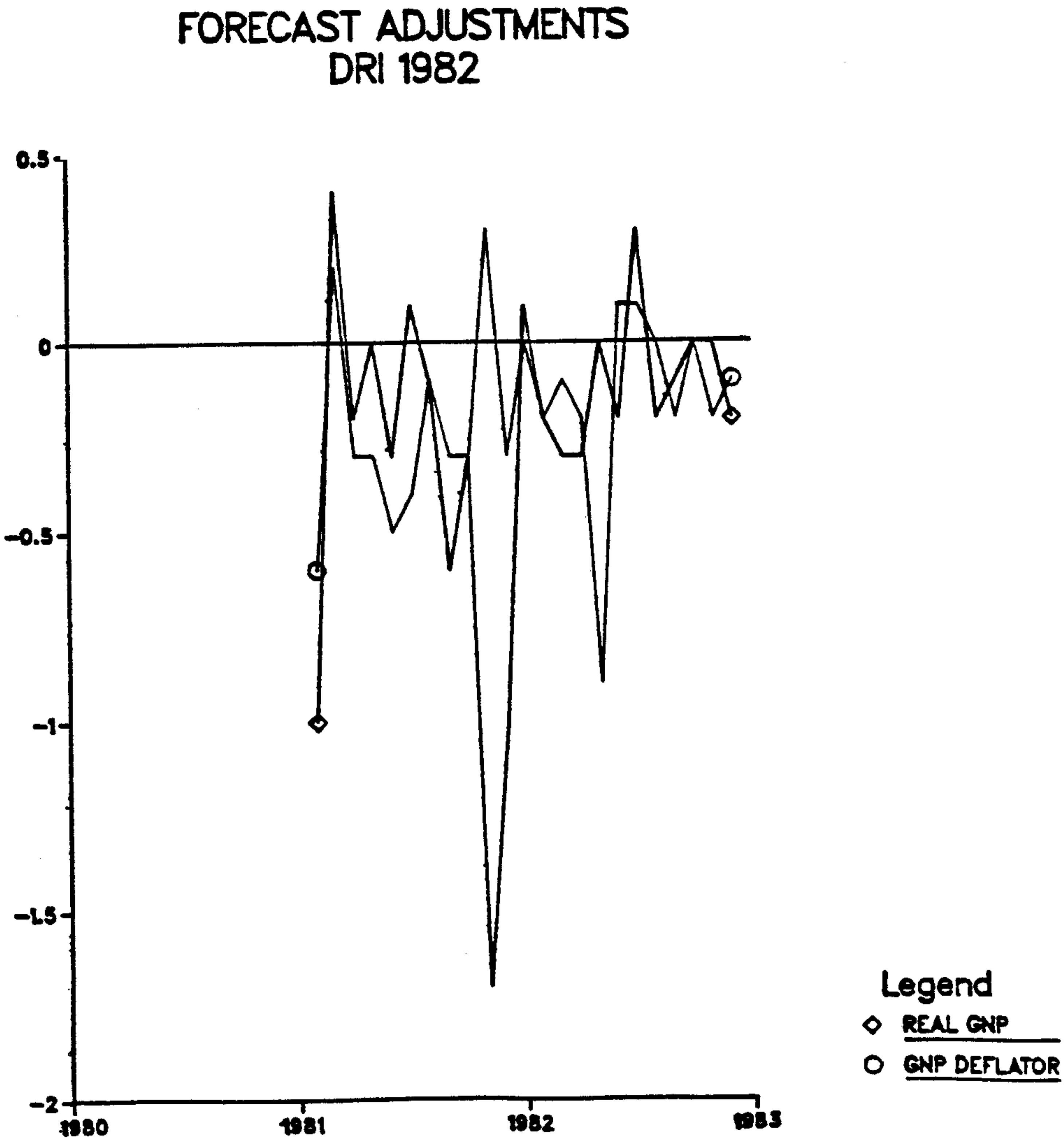
FORECAST ADJUSTMENTS
DRI 1982
0.5
0
-0.5
-1
-1.5
-2
1980
1981
1982
1983
Legend
REAL GNP
GNP DEFLATOR

FIGURE 2

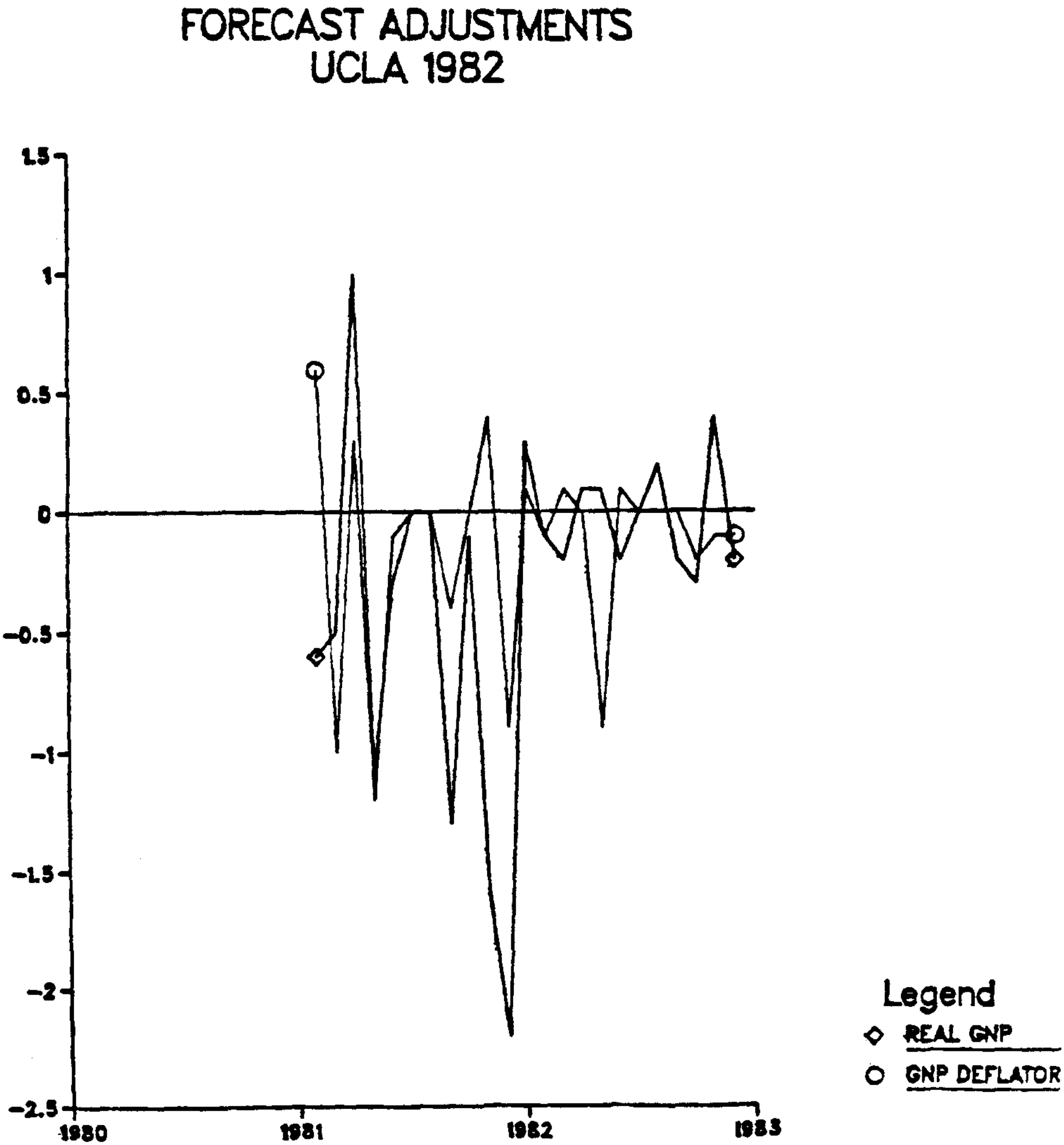
FORECAST ADJUSTMENTS
UCLA 1982
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1980
1981
1982
1983
Legend
REAL GNP
GNP DEFLATOR

FIGURE 3

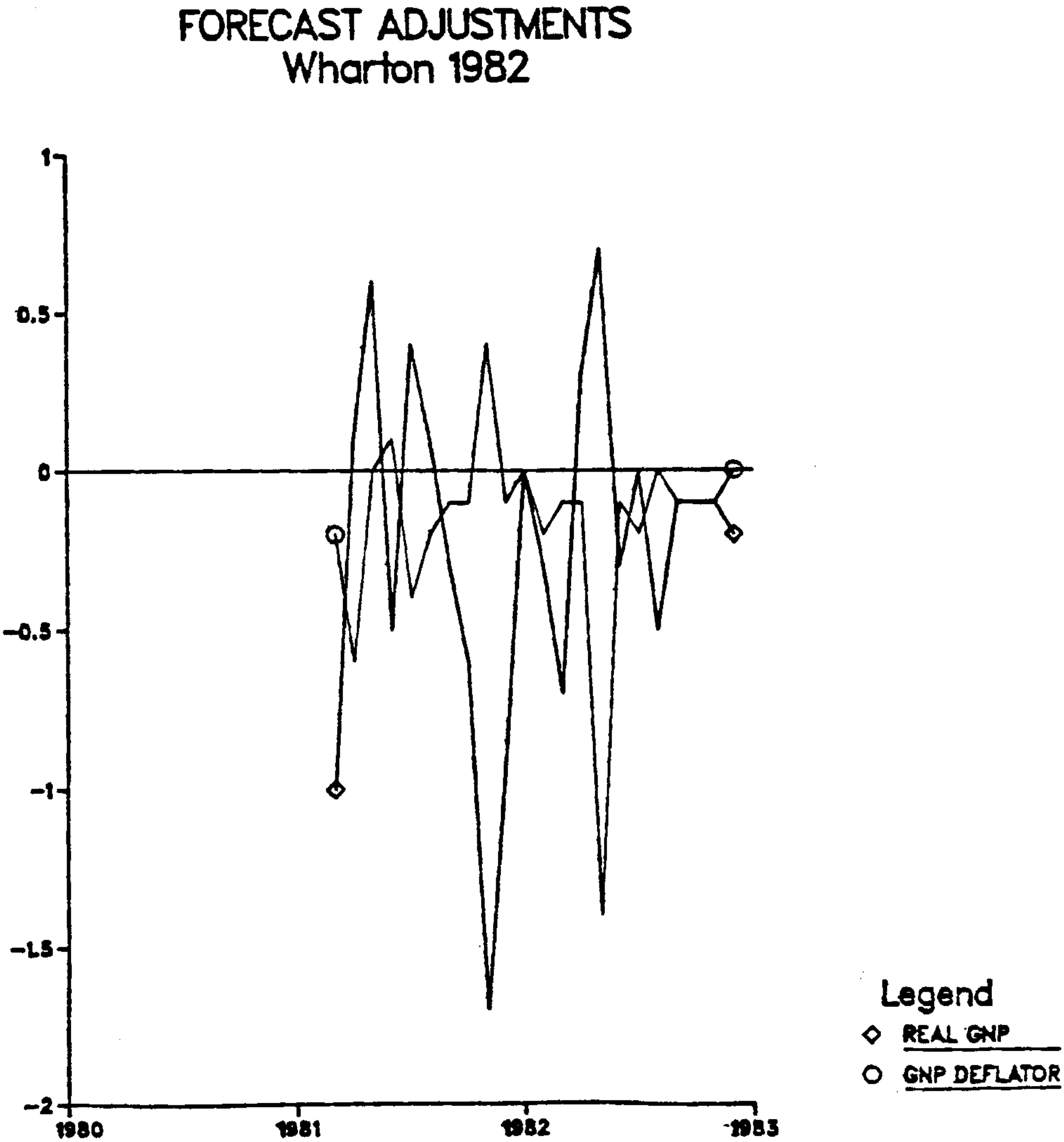
FORECAST ADJUSTMENTS
Wharton 1982
1
0.5
0
-0.5
-1
-1.5
-2
1980
1981
1982
1983
Legend
REAL GNP
GNP DEFLATOR

FIGURE 4

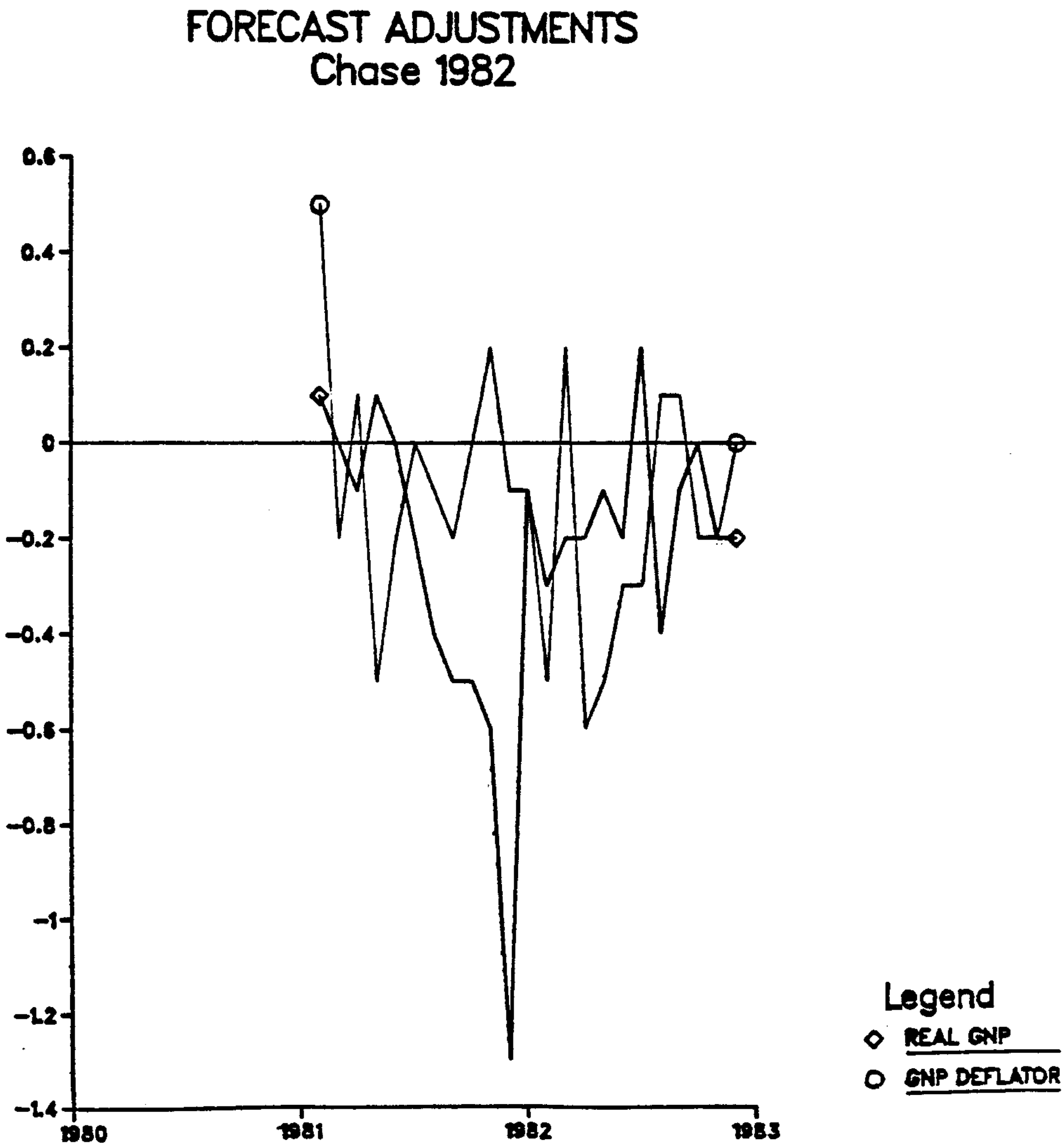
FORECAST ADJUSTMENTS
Chase 1982
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
1980
1981
1982
1983
Legend
REAL GNP
GNP DEFLATOR

TABLE 1. Linear Orthogonality Tests: Estimates of Means Adjustments*

Real GNP Growth	1982	1981	1980	1979	1978
Consensus	-.23 (.05)	-.03 (.07)	-.12 (.09)	-.06 (.06)	-.05 (.02)
DRI	-.25 (.09)	-.03 (.11)	-.15 (.10)	-.09 (.09)	-.08 (.03)
UCLA	-.30 (.14)	.06 (.10)	-.09 (.15)	-.14 (.10)	-.18 (.005)
Wharton	-.23 (.10)	-.09 (.14)	-.10 (.10)	-.10 (.09)	-.10 (.07)
Chase	-.22 (.06)	-.03 (.08)	-.12 (.08)	-.08 (.09)	.10 (.08)
Inflation					
Consensus	-.12 (.03)	.039 (.046)	.08 (.035)	.12 (.03)	.08 (.03)
DRI	-.19 (.06)	-.10 (.10)	.10 (.07)	.14 (.04)	.08 (.03)
UCLA	-.14 (.05)	-.04 (.08)	.10 (.10)	.10 (.10)	.08 (.06)
Wharton	-.12 (.06)	.04 (.10)	.14 (.10)	.13 (.06)	.06 (.05)
Chase	-.12 (.05)	.05 (.08)	.17 (.05)	.13 (.06)	.10 (.04)

*Table shows estimates constant terms and standard errors of constants. The equation is $\Delta_T p_{it} = c + \varepsilon_t$, where $\Delta_T p_{it}$ is change in forecaster i's forecast for real GNP growth or inflation in year T, when the forecasts are made in month t and t-1 ; c is a constant. Figures in parentheses are standard errors of the coefficients. The number of observations ranges from 12 to 24 for each cell.

TABLE 2. Significance Tests for Constants*

GNP Growth	1982	1981	1980	1979	1978
Consensus	x				x
DRI	x				x
UCLA	x				x
Wharton	x				x
Chase	x				
Inflation					
Consensus	x		x	x	x
DRI	x			x	x
UCLA	x				
Wharton	x	x		x	
Chase	x		x	x	x

<u>24 out of 50 cells significant at 5 percent level</u>.

*The significance tests are standard t-tests for the regressions shown in Table 1.

TABLE 3. Linear Orthogonality of Partial Adjustment of Forecasts*

Real GNP Growth	1982	1981	1980	1979	1978
Consensus	.81 (.13)	.52 (.18)	.57 (.19)	.49 (.20)	.11 (.26)
DRI	.3 (.2)	.003 (.2)	.56 (.18)	.3 (.2)	.24 (.23)
UCLA	.15 (.21)	.1 (.2)	.24 (.23)	-.06 (.22)	1.9 (.25)
Wharton	.3 (.2)	.12 (.27)	.07 (.22)	.26 (.22)	.24 (.23)
Chase	.57 (.18)	.38 (.19)	.33 (.18)	.17 (.22)	.17 (.23)
Inflation					
Consensus	.52 (.18)	.27 (.19)	.48 (.19)	.62 (.19)	.36 (.23)
DRI	.16 (.19)	.16 (.20)	.02 (.20)	.24 (.22)	.42 (.22)
UCLA	-.33 (.2)	-.05 (.25)	.2 (.18)	-.04 (.03)	.8 (.6)
Wharton	.15 (.21)	-.4 (.2)	-.05 (.2)	-.02 (.2)	.15 (.26)
Chase	.16 (.19)	.2 (.2)	.04 (.22)	-.18 (.23)	.23 (.23)

*Table shows results of a regression of forecast adjustments on lagged forecast adjustments; i.e. $\Delta_T p_{i,t} = a_T \Delta p_{i,t-1}$, where $\Delta_T p_{i,t}$ is defined in the legend to Table 1 and $\Delta_T p_{i,t-1}$ is the lagged forecast adjustment. Table reports the coefficient and standard error on a. Other variables conventions, and statistics are given in Table 1.

TABLE 4. Lagged Adjustment Significance Tests*

GNP Growth	1982	1981	1980	1979	1978
Consensus	x	x			
DRI			x		
UCLA					
Wharton	x	x	x	x	
Chase					
Inflation					
Consensus	x		x	x	x
DRI					x
UCLA					
Wharton		x			
Chase					

14 out of 50 cells at 5 percent level

*The significance tests are standard t-tests for the regression shown in Table 3.

TABLE 5. Linear Orthogonality Test including the Consensus Forecast*

GNP Growth	1982	1981	1980	1979	1978
DRI	×				
UCLA	×		×		
Wharton	×		×		×
Chase	×				
Inflation					
DRI	×		×	×	×
UCLA	×		×	×	
Wharton	×	×	×	×	
Chase	×	×		×	

Rationality fails in 21 of 40 cases (shown by ×).

*Table shows results of a regression of forecast adjustments as follows:

$${}_{T}p_{i,t} = c + \beta_1 \, {}_{T}p_{i,t-1} + \beta_2 CONS_{t-1}$$

where ${}_{T}p_{i,t}$ is defined in Table 1; c is a constant; ${}_{T}p_{i,t-1}$ is the lagged forecast; and $CONS_{t-1}$ is the consensus forecast for period $(t-1)$. Other variables, conventions, and statistics are given in Table 1. n.a. is not available.

TABLE 6. Lagged Consensus Tests--Regression Results

GNP Growth	DRI		UCLA		Wharton		Chase	
GNP Growth	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
1982	.63	.33	.8	.17	.47	.44	.47	.52
	(.33)	(.29)	(.2)	(.28)	(.23)	(.22)	(.38)	(.38)
1981	.45	.28	.41	.30	.7	.04	.28	.59
	(.3)	(.31)	(.37)	(.36)	(.3)	(.35)	(.53)	(.56)
1980	.85	-.14	-.05	.91	.025	.93	.4	.27
	(.38)	(.54)	(.7)	(.44)	(.3)	(.41)	(.29)	(.21)
1979	.84	-.03	.43	.37	1.1	-.4	.8	-.05
	(.38)	(.4)	(.36)	(.63)	(.34)	(.5)	(.2)	(.3)
1978	.83	-.07	2.3	-.5	.46	.3	.7	-.5
	(.15)	(.32)	(.2)	(.5)	(.2)	(.4)	(.1)	(.46)
Inflation								
1982	.8	.13	.51	.44	.57	.46	.47	.61
	(.2)	(.24)	(.24)	(.27)	(.28)	(.34)	(.25)	(.28)
1981	.48	.22	.85	-.31	-.27	.63	.25	.51
	(.22)	(.36)	(.19)	(.24)	(.29)	(.32)	(.27)	(.37)
1980	.2	.6	-.17	.67	.29	.21	.7	.29
	(.3)	(.35)	(.24)	(.29)	(.25)	(.25)	(.2)	(.29)
1979	.46	.53	.36	.59	.45	.48	-.04	.9
	(.3)	(.3)	(.35)	(.37)	(.21)	(.21)	(.24)	(.23)
1978	-.16	.77	.45	1.00	.65	.46	1.01	-.06
	(.20)	(.20)	(6.1)	(.97)	(.26)	(.29)	(.16)	(.24)

*Table shows results of a regression of forecast adjustments as follows:

$$_Tp_{i,t} = c + \beta_1 \, _Tp_{i,t-1} + \beta_2 CONS_{t-1}$$

where $_Tp_{i,t}$ is defined in Table 1; c is a constant; $_Tp_{i,t-1}$ is the lagged forecast; and $CONS_{t-1}$ is the consensus forecast for period (t-1) . Other variables, conventions, and statistics are given in Table 1.

TABLE 7. 1982 Forecast Accuracy: Averaged Squared Error of Forecast under Alternative Filters for Consensus Forecast

	Filter #1	Filter #2	Actual
GNP Growth	7.1	7.9	10.6
Inflation	2.47	2.55	3.2

*Average squared error = $\frac{1}{n}\sum_{t=1}^{n}(\text{prediction}(t) - \text{truth})^2$

$P_t = \alpha + \beta P_{t-1}$ = underlying forecast structure

Sample: 1981:2 1982:12

TABLE 8. Comparison of Consensus Forecast with Filtered Forecasts, Real GNP growth for 1982

Forecast Date	Actual	Filter #1	Filter #2
1981:2	3.6	3.6	3.6
3	3.6	4.8	3.7
4	3.7	3.7	3.6
5	3.6	3.2	3.5
6	3.5	2.8	1.7
7	3.4	2.3	1.1
8	3.2	1.3	-.1
9	2.9	.7	-1.1
10	.26	.4	-1.9
11	2.2	-1.1	-4.8
12	1.2	-1.8	-7.2
1982:1	.5	-2.5	-5.7
2	.3	-2.4	-2.7
3	0	-2.5	-2.1
4	-.3	-2.5	-2.0
5	-.6	-2.6	-1.9
6	-.7	-2.1	-1.8
7	-.8	-1.9	-1.3
8	-.9	-1.9	-1.6
9	-1.2	-1.9	-1.6
10	-1.4	-1.7	-1.7
11	-1.5	-1.7	-1.7
12	-1.7	-1.7	-1.7

BIBLIOGRAPHY

Berger, Allen N. and Spencer D. Krane, "Testing the Efficiency of Econometric Forecasts," mimeo, 1983.

________, "Why Indirect Tests of the Fair Game-Efficient Markets Hypothesis Are Losers," Working Paper, University of California, Berkeley, July 1982.

Carlson, John A., "Study of Price Forecasts," Annals of Economic and Social Measurement, 6 (Winter 1977), pp. 27-56.

Chung, Kai Lai, A Course in Probability Theory, New York: Academic Press, 1974.

Figlewski, Stephen and Paul Wachtel, "The Formation of Inflationary Expectations," Review of Economics and Statistics, 63 (February 1981), pp. 1-10.

Gramlich, Edward M., "Models of Inflation Expectation Formation," Journal of Money, Credit and Banking, 15:2 (May 1983), pp. 155-173.

Granger, Clive W. J. and Paul Newbold, Forecasting Economic Time Series, New York: Academic Press, 1977.

McNees, Stephen K., "The Recent Record of Thirteen Forecasters," New England Economic Review (September/October 1981), pp. 5-20.

Mincer, Jacob and Victor Zarnowitz, "The Evaluation of Economic Forecasts," in Economic Forecasts and Expectations, New York: National Bureau of Economic Research, 1969.

Mishkin, Fredric S., "Are Market Forecasts Rational?" American Economic Review, 71 (June 1981), pp. 295-306.

Mullineaux, Donald J., "On Testing for Rationality: Another Look at the Livingston Price Expectations Data," Journal of Political Economy, 86 (1978), pp. 329-336.

Pearce, Douglas K., "Comparing Survey and Rational Measures of Expected Inflation," Journal of Money, Credit and Banking, 11 (November 1979), pp. 447-456.

Pesando, James E., "A Note on the Rationality of the Livingston Price Expectations," Journal of Political Economy, 83 (August 1979), pp. 849-858.

Samuelson, P. A., "Proof that Properly Anticipated Prices Fluctuate Randomly," Industrial Management Review, 6:2 (1965), pp. 41-50.