View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Lancaster University Management School
Working Paper
1999/002

The State of Macroeconomic Forecasting

Robert Fildes and Herman Stekler

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

©Robert Fildes and Herman Stekler
All rights reserved. Short sections of text, not to exceed
two paragraphs, may be quoted without explicit permission,
provided that full acknowledgement is given.

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/
LUMS home page: http://www.lums.lancs.ac.uk/

The State of Macroeconomic Forecasting

Lancaster University EC2/99

George Washington University, Center for Economic Research

Discussion Paper No. 99-04

Robert Fildes,

Lancaster University, UK

Herman Stekler,

George Washington University, USA

Revised: November 1999

The State of Macroeconomic Forecasting

Robert Fildes, Department of Management Science,
Lancaster University, Lancaster, UK, LA1 4YX
Email: R.Fildes@Lancaster.ac.uk

Herman Stekler, Department of Economics,
George Washington University, 2201 G ST, NW, Washington DC 20052, USA
Email: hstekler@gwis2.circ.gwu.edu.

ABSTRACT

Macroeconomic forecasts are used extensively in industry and government The historical accuracy of US and UK forecasts are examined in the light of different approaches to evaluating macro forecasts. Issues discussed include the comparative accuracy of macroeconometric models compared to their time series alternatives, whether the forecasting record has improved over time, the rationality of macroeconomic forecasts and how a forecasting service should be chosen. The role of judgement in producing the forecasts is also considered where the evidence unequivocally favors such interventions. Finally the use of macroeconomic forecasts and their effectiveness is discussed. The conclusion drawn is that researchers have paid too little attention to the issue of improving the forecasting accuracy record. Areas where improvements would be particularly valuable are highlighted.

Keywords: econometric models, macroeconomic forecasting, rationality, forecast evaluation, structural breaks, cliometrics, industry structure

The State of Macroeconomic Forecasting

Robert Fildes

Department of Management Science,

Lancaster University, UK, LA1 4YX

H.O. Stekler

Department of Economics

George Washington University, Washington DC 20052

Considerable intellectual activity within the economics profession is devoted to the production, interpretation and analysis of forecasts of major economic variables. Since such forecasts are important to both government planning and industry, it is material to determine how well we, as a profession, have performed this activity and what lessons may lead to improvements. Given the number of analyses that have included predictions of one economic variable or another and the range and depth of the macroeconomic forecasting industry (as surveyed by Fildes, 1995), it is impossible, in one paper, to review all aspects of the field. We will, therefore, concentrate on surveying and analyzing the predictions of short-run GDP (GNP), with particular emphasis upon real GDP and inflation forecasts for the US and the UK, bringing in data from other countries only when necessary in order to have a common set of information about cyclical conditions, etc. We focus on GDP and inflation forecasts because these much predicted variables are of interest to the entire profession. Unfortunately, this emphasis precludes an analysis of the characteristics of the forecasts of other variables and leaves many questions unanswered, e.g. which GDP components were the hardest to predict and which contributed the most to the inaccuracy of the aggregate forecasts?

While many techniques have been used to make short-run GDP forecasts, the focus of this article will be on quantitative methods. In examining methods designed to provide quantitative estimates

of GDP, we will consider time series analyses, econometric models (as well as forecasts made using judgmental techniques), and the contribution that expert judgment has made to the modeling process. As the survey evidence in Fildes (1995, p.6) showed, this range of methods covers those used by macroeconomic forecasting services. Since other articles have compared the structure of many of the econometric models used in macroeconomic forecasting, for example Klein (1991) for the US and Wallis and his co-workers in the UK (1984,1985,1986,1987) and other researchers have dealt with the history and structure of the industry (see Bodkin et al, 1991; Daub, 1987; Smith, 1994; and Den Butter and Morgan, 1998), this paper will only deal with the forecasting record of these models.

It is generally agreed that a good economic forecast should provide certain information; the direction of movement of the economy should be predicted, and if there is an expected reversal in direction, the timing of the turn. In addition, the forecast should also indicate the magnitude of the change and the length of time over which a movement is expected to persist. By their very nature, quantitative forecasts contain all of these elements and they can therefore be evaluated to determine whether they have satisfactorily met these criteria. Thus, one would expect that the myriad evaluations of these forecasts would have provided answers to questions such as: What is the magnitude of the average forecast error (relative to the outcome)? Were the dates of the recessions predicted accurately? Were the major inflationary cycles forecast? Was there a bias in the predicted direction of the change in GDP (or GNP)? Why did the errors occur? etc. These questions are of importance to both the users and providers of macro forecasts. The former group needs to understand the strengths and limitations of the predictions while the latter requires these data in order to develop techniques that will improve the quality of the forecasts.

In addition to these fundamental questions about the quality of quantitative forecasts, we shall review other topics that previous evaluations have (or should have) been addressed. These include: (1) Do judgmental adjustments improve the accuracy of the econometric and time series predictions? (2) Is one method or forecaster consistently superior to another? (3) Are the forecasts superior to readily available and simpler standards of comparison? and (4) Is it desirable to combine forecasts from different sources? Some of these questions can be answered quite easily; others will require a more extensive discussion.

In summary, this survey will primarily report on the findings of previous evaluations of the accuracy record of macroeconomic forecasts and the factors that have affected accuracy although some new data analysis is included. A secondary aim is to develop an agenda for the improvement of macroeconomic forecasting and an increased understanding of its limitations. In so doing difficult economic theoretic issues are raised but for the purposes of this article these are subordinated to the focal question of how macroeconomic forecasts can be made more immediately useful.

1. Measuring Accuracy

1.1 The Base Line Data

The National Income Accounts of both the US and the UK are revised periodically. The first question that arises is whether it matters which measure of the outcome (or 'actuals') is used in any evaluation of accuracy. We illustrate this point in Table 1 where we calculate the effect of using different data 'vintages' to evaluate the accuracy of US macroeconomic forecasts. These calculations were derived from several studies published by McNees (1976a, 1979, 1985). He has always used the final data, defined as those available at the time that the evaluation is undertaken, as the basis of comparison with the forecasts. McNees argues that it is the forecaster's responsibility to predict what actually happened, as measured by the final numbers, rather than to estimate what the statistical agencies report initially. Thus if the revised data show that GNP grew at x% rather than the preliminary published estimate of y%, the forecast error would be changed by the difference x-y.

Table 1 Mean Absolute Errors (% in Growth Rates): Different Data Vintages

	Real GNP			**Change in GNP Deflator**		
Evaluation	Data Vintage			Data Vintage		
Period	1976	1979	1985	1976	1979	1985
1971.2-1976.4	1.86	1.80	1.55	2.02	1.93	1.70
1971.2-1979.2	-	1.43	1.31	-	1.67	1.53

Source: McNees (1976a, 1979, 1985)

Table 1 shows that when the National Income Accounts data available in 1985 are used as the final figures, the errors of both the GNP and inflation predictions are lower than if either of the earlier sets of numbers is used as the 'actuals'.

One can argue that the results of forecast evaluations should not change with each revision of the National Income Accounts. Like McNees, Zarnowitz has settled on one vintage of actual data for use in his evaluation analyses. In his evaluations, however, the actuals with which the forecasts were compared are the first revised figures, i.e. those published 45 days after the end of the quarter to which they refer. Zarnowitz argues that individuals should not be judged on their ability to predict how the statistical agencies will revise the data in the future, this is especially true if there are benchmark revisions that could not have been anticipated when the forecast was issued. In any event, Zarnowitz and Braun (1992) have shown that the quantitative results do not vary substantially with the vintage of the data, as long as there have not been any benchmark revisions. Moreover, the qualitative results remain consistent. Melliss (1997) in a review of UK evidence also found no consistent effect from the particular data vintage used although in recent years, where GDP has been revised upward and forecasts have tended to underestimate growth, the errors have increased.

In contrast to the US studies, many authors who evaluated the forecasts of other countries generally did not indicate which data sets were used in their analyses. However, a number of forecasters such as stock market analysts are vitally interested in the preliminary figures and the shocks that the publication of those numbers might have on financial markets. Thus, in any evaluation the objectives of both the forecaster in releasing the forecast and the use to which it is put should be considered in deciding which set of actuals should be used. [1]

1.2 Qualitative Measures

The choice of measures to use in evaluating the accuracy of macroeconomic forecasts is dependent upon the questions that are under investigation. When focussing on the qualitative characteristics of forecasts, simple measures are frequently used. First, to see whether the cyclical

[1]As an example, Zellner (1986a, 1986b) has demonstrated that if a user has an asymmetric loss function, a biased prediction may be optimal.

turns are predicted, one can count the number of turning points that were predicted, the number that were not forecast, and the number of times that turns that were predicted failed to occur. Second, it is possible to look for systematic errors, such as over and underestimates of the rate of growth or of the inflation rate, and determine whether they are associated with particular phases of the business cycle.

1.3 Descriptive Statistics

In order to move beyond qualitative results, it is necessary to develop quantitative measures that summarize the raw errors. Let $\hat{Y}_t(k)$ represent the k period ahead forecast of Y_{t+k} made at time t for period t+k, and $e_Y(t,k)$ is the corresponding k period ahead forecast error. It will then be possible to compare different methods and forecasters and to determine whether there are any statistically significant differences among them. A large number of error measures have been proposed. Among the most frequently used in the analysis of macroeconomic forecasting errors are: the mean error, ME, that measures bias. For lead time k based on time origins T_0+1, . . T_0+n. it is defined as

$$ME = \frac{1}{n}\sum_{j=1}^{n} e_Y(T_0 + j, k)$$

The root mean square error (RMSE)

$$\text{RMSE} := \sqrt{\frac{1}{n}\sum_{j=1}^{n} e^2{}_Y(T_0 + j, k)}$$

and the mean absolute error (MAE)

$$\text{MAE} := \frac{1}{n}\sum_{j=1}^{n} \left| e_Y(T_0 + j, k) \right|$$

While these measures are related to the loss functions being minimized (with the RMSE related to quadratic loss), all have a number of limitations. First, by themselves, they do not provide meaningful information and, moreover, they have been subject to intense criticisms. The RMSE is particularly affected by outliers that are common in economic data. Neither of the measures is naturally scale independent except when applied to percentage changes (Armstrong and Collopy, 1992; Fildes, 1992) and they involve averaging errors over observations that have different degrees of variability (Fair, 1980; Jenkins, 1982; Diebold, in Baillie et al, 1993). The issue of scaling is also critical when a number of data series are being analyzed. Clements and Hendry

(1993) have proposed a generalization of the RMSE that takes into account the intercorrelations between errors when more than one macroeconomic series is being analyzed and to ensure invariance to linear transformations. The practicality of their measure has been questioned because typical macroeconomic forecast evaluations are based on small samples of non-normal data (Armstrong and Fildes, 1995).[2] Nor is invariance a necessarily desirable measure (Baillie et al, 1993). We must, therefore, present a caveat: the error measures that are used in forecast evaluations should be reliable when used with small samples and be robust to the potential lack of normality, providing a fair characterization of the error distribution of $e_Y(t,L)$, a task no single measure can achieve. Although there are several articles that deal with this distribution of errors (Dawid, 1984; Diebold et.al. 1998), this issue has often been disregarded in most forecast evaluations.

Despite such substantial criticisms, the RMSE and MAE descriptive statistics have invariably been used as standards for judging the quality of predictions. For example, the RMSE can be compared with the standard deviation of the series being forecast. The MAE can be compared with the mean absolute change of the predicted variable. These measures show how the errors are related to the variability of the series that is being forecast.

Typically, the discussion of error measures has focussed on the point forecast alone. Recently, attention has also been given to the uncertainty around the published forecasts (for example, the historical error statistics are also published with the UK Treasury forecasts). The question of whether estimates of the uncertainty in a point forecast are well calibrated (i.e. the estimated probability distribution matches the realized distribution, see Diebold, Gunther and Tay, 1998) has not often been examined as there are little data available. In general, Chatfield (1993) has argued, both model based and judgmental estimates tend to underestimate the uncertainty in the forecasts. Nevertheless, for many users an accurate measure of future uncertainty may be as valuable as the point forecast itself and this poses a second order forecasting problem (Tay and Wallis, 1999).

[2] Lahiri and Teigland (1987) demonstrate the non-normality of the forecast errors for GNP and inflation. Also see Granger's (1995) analysis of Zarnowitz and Braun's (1992) data. However, Holden and Thompson (1997), Melliss and Whittaker (1998) examining UK data and Öller and Barot using OECD data found less of a problem.

1.4 Naive and Time Series Benchmarks

A better approach to evaluating performance is to compare these descriptive statistics with similar statistics obtained from a naive standard or time series model (e.g. from within the ARIMA class). The rationale for using such series is that forecasters should perform at least as well as the simplest time series models from which predictions could have been derived (Nelson, 1972, 1984). The use of a standard also permits an analyst to test whether the errors of a set of forecasts are significantly smaller than those of the benchmark.

A naive standard that is frequently used involves the comparison of a set of forecast errors with those obtained from a no-change (random walk) naive model. Theil's (1966) U coefficient is used in this comparison.

$$U = \sqrt{\frac{\sum (Y_{t+k} - \hat{Y}_t(k))^2}{\sum Y_{t+k}^{\ 2}}} .$$

This statistic formalizes the process of comparing a set of predictions against some standard. If U is less than 1, the forecasts which are being evaluated have smaller errors than those which would have been generated by the naive model, but this result does not guarantee that the former are significantly better than the latter.

This naive model (N1) serves as an appropriate minimum standard of comparison for variables specified in terms of changes or growth such as real GNP, because they experience both positive and negative changes. However, a different basic standard is required for variables such as nominal GNP and inflation, which almost always show positive rates of growth. In this case, the naive model (N2) extrapolates the same change as was observed in the last period. The U coefficient can still be used with this naive model; the errors that N2 would have generated now enter the denominator of U.

The naive models establish the minimum level of accuracy that a set of forecasts should have. Moore (1969), therefore, suggested that the accuracy of a particular set of forecasts should be compared with the predictions obtained from time series extrapolations. From the 1970s on evaluations have adopted even more demanding benchmarks by using time series models from the

ARIMA and Vector Autoregressive (VAR) class as standards against which the forecasts are compared (see for example, Holden, 1995 in a special issue on the topic).

There is, however, a problem with using these techniques as benchmarks. They are usually estimated with the latest available data, and not the numbers which existed when the forecasts were made. Whatever benchmark is used in the evaluation of forecasts, the difference between the two sets of errors should be tested for statistical significance. Granger and Newbold (1977) had proposed a useful statistic for making MSE comparisons. A new statistic developed by Diebold and Mariano (1995) permits a statistical comparison when other quantitative error measures are reported. This test was further extended by Harvey et al., (1997)[3]. Ashley (1998) has developed a new measure, based on bootstrap inference, that does not require large samples.

1.5 Rationality Tests

A comparison of the accuracy of any set of forecasts with the predictions of naive or benchmark standards merely shows whether one method had lower errors than another but do not indicate how to improve upon the observed record. Rationality tests, as well as the encompassing tests and combining procedures that are discussed below, determine whether it would have been possible to improve upon the observed forecasting accuracy or whether they are already optimal with regard to a particular information set (Wallis, 1989).

A variety of rationality tests have been proposed. The simplest and the one that is most commonly used tests for bias and weak form informational efficiency. This test is based on the regression:

$$Y_{t+k} = \alpha + \beta \hat{Y}_t(k) + e_{t+k}$$

where Y_t and $\hat{Y}_t(k)$ are the actual and k period ahead predicted values at time t respectively. The test is a joint test of the hypothesis that the coefficients α and β do not differ significantly from 0 and 1, respectively, and that the errors, e_{t+k} be MA (k-1). For k=1, they must not be serially

[3] These statistics may also be used to compare the relative accuracy of any two forecasters or methods.

correlated. A rejection of this hypothesis would clearly indicate that the forecasts are not rational either because they are biased and or because they have not incorporated the information contained in past errors. If the errors are autocorrelated this is a rejection of weak form efficiency because, in principle, an improvement in forecast accuracy is clearly possible. A slightly more restricted condition is to test:

$$Y_{t+k} - \hat{Y}_t(k) = \alpha + e_{t+k}$$

with null hypothesis: $\alpha=0$ since the joint hypothesis (of $\alpha=0$, $\beta=1$) is sufficient for unbiasedness but not a necessary condition (Holden and Peel, 1990).

An improvement in accuracy is also possible if the forecast error is correlated with any information; (including the past values of the variable being forecast) that was known at the time that the forecast was prepared. This correlation indicates that past information explains the

current error, and, consequently, that this knowledge has not been used efficiently. This test has been called the orthogonality test and is based on the regression:

$$e_Y(t-k,k) = \gamma + \sum_i \sum_{j>k} \delta_{i,j} X_i(t-j) + e_t$$

where the Xs are possible variables, which might have affected the forecast. The joint hypothesis is that γ and all the $\delta_{i,j}$ are not significantly different from zero. A rejection of this null hypothesis indicates non-rationality and the possibility of improving forecast accuracy by using the information contained in those variables, which had non-zero coefficients.[4]

A further test for weak form efficiency is to require that forecast revisions for a fixed outcome are independent of past available information (Nordhaus, 1987, Clements, 1997).

A caveat should be noted. Jeong and Maddala (1991), using financial data, show how the acceptance or rejection of the rationality hypothesis depends upon the statistical assumptions and procedures that are used. For example, the regression tests are not valid in the presence of unit

[4] It has been suggested that these rationality tests be conducted on rolling regressions which also take into account the difficulty in forecasting particular periods (i.e. heteroscedastic forecast errors). Finally if there is a regime change the forecasts may be rationally biased as learning occurs. It is also possible that the forecaster's loss function is asymmetric and the predictions would still be rational even if the mean error is non-zero.

roots and co-integration tests should be used (Pain, 1994). They also point out that errors in variables could affect the results.

2. The Record:

In this section we summarize the results of the many studies that have evaluated the accuracy of macroeconomic forecasts. We examine both the turning point errors and various quantitative measures of forecast accuracy.

2.1 Turning Point Forecasts; Systematic Error

2.1.1 The US Forecasts

Over the past 30 years, Zarnowitz (1967, 1979, 1984, 1992a), Zarnowitz and Braun (1992), McNees (1976b, 1986, 1988, 1992a), and McNees and Ries (1983) have provided the most complete and continuing analysis of US macroeconomic forecasts. The recent analyses of Zarnowitz and McNees are based on different data sets, with the former focussing on the record of professional macroeconomic forecasters responding in the ASA/NBER surveys, latterly called the Survey of Professional Forecasters (Croushore, 1993). McNees examines the predictions primarily generated from various econometric models, but he also includes the median ASA/NBER predictions in some of his analyses. The results that we report are robust with respect to both the time periods, which are covered, and the methods, which were used to make the predictions. Zarnowitz's analyses of the ASA/NBER forecasts covered the predictions over the entire period from the inception of the surveys in 1969 to the date of his study while the time periods of McNees' analyses varied from study to study.

One of the most disturbing findings is that, in the United States, recessions were generally not forecast prior to their occurrence. This is true for both annual and quarterly estimates. (McNees, 1976b, 1988, 1992a; McNees and Ries, 1983, Zarnowitz, 1992a). For example, both the 1974 and 1981 peaks were not recognized even as they occurred. While missing the actual turning points, the forecasts did indicate that the economy would be slowing down. The forecasters were thus successful in distinguishing between periods of rapid growth on one hand and slow and negative growth on the other hand. The question then becomes: when did the forecasters recognize that the 'slowdowns' had turned into recessions? Although this question has not been examined

intensively, the evidence that is available suggests that this recognition sometimes occurred after the quarter in which the economy had its peak (Schnader and Stekler, 1990; McNees, 1992b; Stekler, 1994).[5]

Economists made few predictions of peaks which did not occur, with the false turns of 1978-79 an obvious exception.[6] On the other hand, they were willing to forecast the end of recessions (Zarnowitz, 1992), even to the extent of predicting such turns too quickly (McNees, 1976b, 1987). The factors that might produce such an asymmetric performance in recognizing turning points have not been adequately explored. However, Stekler (1972) has suggested one possible explanation; forecasters, using a Bayesian type decision process, had zero priors that a recession would occur. Within this decision making process, other explanations are also possible, i.e. the economists assigned differential costs to the two types of turning point errors: predicting a recession which did not occur and failing to predict one which did happen (Schnader and Stekler, 1998).

Other types of systematic errors in predicting the rate of growth of the US economy have been observed. Zarnowitz and Braun (1992) indicated that growth was generally overestimated during slowdowns and recessions while underestimates occurred during recoveries and booms. Anderson (1983) found a similar pattern in the forecasts of the National Association of Business Economists. Originally, it had been argued that the underestimates of growth resulted from optimal forecasts (Mincer and Zarnowitz, 1969; Hatanaka, 1975; Samuelson, 1976) given that the variance of the predictions should be less than that of the outcomes. However, the existence of these underestimates cannot be attributed to this theoretical explanation. Stekler (1975) showed that during his sample period, in which there were underestimates, the variance of the forecasts

[5] Fels and Hinshaw (1968) directly addressed this issue of when turning points were recognized. Stekler (1972) developed a model that might explain why forecasters failed to predict or even recognize such events.

[6] While there was no recession in the 1978-79 time frame, real GNP did decline for two quarters during this period. Thus, the economists foresaw economic weaknesses but recessions did not develop. The question that may be asked is why forecasters were so willing to predict a recession in this period. A possible explanation is that economists had badly missed the date of the previous (1973-1975) recession which was associated with an oil shock and did not wish to make a similar error when the second oil shock occurred, i.e. the forecasters had over-compensated for this last observed error, an example of the so-called 'anchoring heuristic' often observed when human judgment is used to supplement model-based forecasts (Bolger and Harvey, 1998).

exceeded that of the outcomes[7], while Zarnowitz (1979) argued that these underestimates occurred because forecasters failed to recognize the strength of booms perhaps because of their use of a 'regression towards the mean' heuristic which would lead to a lack of rationality in their forecasts.

In terms of predicting the US inflation rate, the record is not much better. The inflation rate was generally underpredicted when it was rising and overpredicted when it was declining (Zarnowitz and Braun, 1992; McNees and Ries, 1983). Particularly large errors occurred during the periods when prices were rising rapidly during the 1970s and early 1980s. Zarnowitz (1979) argues that the inflation forecasts appeared to be projections of the last observed value. When turning points occurred, the extrapolative projections always forecast them with a lag. The major evaluations do not focus on the length of time that it took for forecasters to recognize that inflation had either accelerated or decelerated.

On the basis of this qualitative information, what can we conclude about this predictive record? The forecasters did best when economic conditions were relatively stable; they performed less well when the economy was subject to major perturbations which was when accurate forecasts were most needed.

2.1.2 UK Forecasts and Other Evidence

Although institutionalized macroeconomic forecasting has a long history in the UK, for many years this activity received little academic attention (Cairncross, 1969). In addition to the forecasts issued by the government and a number of research organizations, there are, as in the US, a substantial number of commercial forecasting organizations but these predictions have not been evaluated extensively. The major non-commercial UK macroeconomic forecasting organizations have, however, benefited from an extended and detailed study by Wallis and his colleagues at the Economic and Social Research Council funded Macroeconomic Modelling Bureau, Warwick University. This research has differed from that of Zarnowitz and McNees because the UK Bureau's emphasis has been primarily with the validity and behavior of the models rather than the accuracy of the forecasts (Smith, 1990).

[7] Smyth and Ash (1981), using UK and OECD data obtained the opposite result. Using OECD data Smyth and Ash (1985) also showed that the ratio of the variance of the outcomes to that of the predictions increased with the length of the forecast horizon.

Moreover, the UK studies did not specifically determine whether the forecasters had predicted cyclical turning points in advance of their occurrence; rather the analyses have focussed on the magnitude of the forecast errors. The limited evidence regarding forecasters' ability to predict these turns suggests that they generally were not able to do this. (Barker, 1985; Wallis, 1989; Britton and Pain, 1992). Some preliminary work by Mills and Pepper (1999) confirms this finding. They suggest that the UK forecasters not only failed to predict the cyclical turns of the 1979-82 and 1989-92 recessions in advance, but in addition their ability to identify a peak (or trough) even at the time of its occurrence was limited. Barker's (1985) analysis of the 1979-82 recession indicated that those forecasters who had a longer term perspective (using annual data) were more successful at longer term forecasts than those, like the London Business School, the UK Treasury and the National Institute of Economic and Social Research, that focussed on the year ahead.

The inflation surges of 1979-80 and 1989-90 were also not predicted until a number of months after prices had started rising rapidly; a similar result was observed when inflation tapered off (Mills and Pepper, 1999). As in the US there is a general tendency to overpredict inflation when it is falling and to underpredict it when it is rising (Wallis, 1989; Melliss, 1997).

As for other countries, Ash et al (1998) have analyzed whether the OECD forecasts made with horizons of 6, 12, and 18 months correctly predicted the direction of change of a number of variables. While rationality could not be rejected, tests of the value of the predictions, defined by Stekler (1994) as 'changing the user's prior distribution about the direction of change in the economy' showed that few of the GNP forecasts that were made 12 or 18 months ahead had this property, a conclusion confirmed by Őller and Barot (1999). Thus, they could not be considered significantly better, in terms of predicting the direction of change, than the naive no change model. In addition Anderson (1997) indicated that the OECD predictions tended to underestimate changes in the trends of both real output growth and inflation. However, in periods when inflation was decelerating, the rate of inflation was overpredicted.

We can thus conclude that the UK and OECD evidence are in accord with the findings that were obtained from the evaluations of the US forecasts.

2.2 Quantitative Results

Although the qualitative findings with regard to turning points and systematic bias are crucial to understanding the forecasting record over cyclical periods, most evaluations have been quantitative in nature and have not separated the results into the cyclical phases.

2.2.1. US Forecasts

The US forecasting record analyzed by Zarnowitz and McNees have been generated by different techniques, cover different time periods, and involve estimates over different horizons, i.e. annual and quarterly spans from 0 to n periods. In the latter category, both the forecasts made over a span of 0-n quarters and the predictions made for a particular quarter k periods in advance have been examined.

One-Year Forecasts

Zarnowitz (1992a) indicates that annual forecasts of real GNP growth, made for various sub-periods of 1962-1989, had mean absolute errors slightly in excess of 1% of GNP. These errors were at least 25% of the mean absolute change in GNP. McNees (1988) obtained a similar result in his analysis of the Michigan model's forecasts. The MAE of the real GNP forecasts was 1.3% with a corresponding mean absolute change of 3.7%. Similarly, Zarnowitz showed that the inflation predictions had mean absolute errors of 1.0-1.4%, which, depending on the time period, should be compared with the mean absolute changes of 4.2- 5.9% in the inflation rate.

Diebold et al (1997) examined the record of the Survey of Professional Forecasters who along with giving point forecasts of inflation also provide a probability density function. Their average estimates prove not to be well calibrated: “negative inflation surprises occur less often than expected” and in more recent data this tendency extends to both tails. This adds to the result noted elsewhere, that inflation is underestimated when it is rising, overestimated when falling.

Quarterly Estimates and Forecasts

There are a number of questions one can ask about the quality of multi-period quarterly forecasts: What is the magnitude of the errors? What is the relationship between the length of the forecasting lead and the size of the errors? Do errors cumulate or offset each other over time? In answering these questions it is important to recognize that we might reach different conclusions if we examine the forecasts made over a span of the next n quarters as compared to forecasting activity in a specific quarter was predicted n periods in advance. The n period span involves the predicted change from the last known observation t-1 to t+n-1; the annual prediction with n=4 is an example[8].

The results of the major studies show that the accuracy of both the predictions of GNP growth expected n quarters in the future and the forecasts of the growth that is expected to occur over a span of n periods improve as the length of the prediction horizon decreases. A substantial improvement occurs when the task of forecasting switches from predicting what will happen in the next quarter to estimating the level of activity of the current period. Table 2 shows that the MAE of the median ASA/NBER forecast of the change in real GNP for the current quarter was 2.36% (annual rate calculated according to footnote 8) rising to 3.04% for the subsequent period and 3.68% for the change expected three-quarters in advance. Similar results hold for the inflation predictions. When predictions are made every month, accuracy improves substantially as the first month's actual data (for the current quarter) become available (McNees, 1988; Kolb and Stekler, 1989).

[8] A difficulty arises in interpreting the results of the major published studies. Although both McNees and Zarnowitz transform their data into growth rates, McNees converts these data into annual rates of growth; Zarnowitz does not make this conversion. An approximate adjustment to the published figures in the latter's studies is to multiply all quarterly errors by 4; all semiannual data by 2; etc.

Table 2 MAE and RMSE of median ASA/NBER Forecasts of GNP, Real GNP and the GNP Price Deflator, Span and Intra-quarter Forecasts

Mean Absolute Error

Variable	**Span**					**Intra-quarter Changes**			
	0-1	0-2	0-3	0-4	0-5	1-2	2-3	3-4	4-5
GNP	.56	1.02	1.49	1.84	2.31	.73	.81	.82	.86
Real GNP	.59	1.00	1.41	1.82	2.26	.76	.82	.92	.94
Price deflator	.38	.72	1.08	1.55	2.07	.49	.54	.57	.62

Root Mean Square Error

Variable	**Span**					**Intra-quarter Changes**			
	0-1	0-2	0-3	0-4	0-5	1-2	2-3	3-4	4-5
GNP	.74	1.33	1.93	2.45	3.06	.97	1.10	1.11	1.14
Real GNP	.77	1.32	1.90	2.57	3.12	.98	1.15	1.26	1.31
Price deflator	.47	.91	1.43	2.06	2.79	.61	.70	.74	.81

Source: Zarnowitz and Braun (1992)

These results can be interpreted in the following way. The accuracy of forecasts of the change expected to occur in a specific period made with a lead of three-quarters is not likely to be significantly better than the prediction made a year in advance unless substantial new information has become identified. It is only when actual data for the current period become available that a significant improvement in accuracy might be expected. However, the projections made for the current quarter cannot be considered pure forecasts. Rather they are a combination of estimates based on the newly available data and forecasts of the activities for which information has not yet become available (McNees, 1988). Various model-based approaches have been proposed for incorporating current monthly information in the quarterly forecast (Howrey, 1978; Corrado and Greene, 1988).

The increase in accuracy as the forecasting horizon decreases is not monotonic. This is especially true for the forecasts made over a span of n quarters. The median ASA/NBER forecasts of the growth of real GNP have a MAE of 2.36% (annual rate) for the current quarter but only a MAE of 1.82% over a four-quarter span. In other words, for this variable, the longer time span was predicted more accurately than the shorter one. However, the inflation forecasts do not show this tendency. Thus, for some variables the errors made in forecasting the changes in each quarter of a span may offset each other while the errors made in predicting other variables apparently accumulate (Joutz, 1988; McNees, 1988). The reason that longer span forecasts might be more accurate than the shorter span ones is that the predictions capture the trends that prevail over the longer period but do not precisely determine the dynamics of the quarterly movements which offset each other.

2.2.2. The UK and Other Evidence

The UK forecast evaluations have examined predictions made by the UK Treasury and by the model based forecasting groups such as the National Institute of Economic and Social Research, Cambridge Econometrics, Liverpool University, London Business School (LBS), and Oxford Economic Forecasting. The UK Treasury forecasts have been given the most severe scrutiny, in part because of the issue of public accountability and the role that they might play in setting government policy. Burns (1986) presented a report on the Treasury's record, concentrating (as we do) on growth and inflation. Subsequently a report was prepared and commented on for the Treasury and Civil Service Committee (1991) which again evaluated the Treasury's record and concluded that the Treasury forecasts are comparable in accuracy to the other major providers, both public and private. This conclusion has been given further confirmation by a comparative analysis of the Treasury with the National Institute, supplemented by data on the CBI (Confederation of British Industry) and LBS group forecasts (Britton and Pain, 1992). Table 3 summarizes this and more recent evidence.

Forecasting Group	**GDP**		**Inflation**	
	1986-1990	1990-1998	1986-1990	1990-1998
Independent Average	1.2	1.02 (.91)	1.6	1.21 (.80)
Selected Independents	1.0	.95 (1.0)	1.4	1.21 (.70)
Independent consensus	NA	1.06 (1.1)	NA	1.06 (.70)
City Average	1.0	.95 (.74)	1.7	1.21 (.50)
City consensus	NA	.94 (1.09)	NA	1.01 (.60)
Treasury	.8	1.15 (1.09)	1.5	1.075 (.70)
Average Outcomes	3.05	1.47	6.32	3.72
Standard deviation Outcomes	1.61	1.93	2.68	2.47
Mean Absolute deviation	1.20	1.49	2.06	1.50
Naïve Forecast	1.35	1.72	1.74	1.70

Table 3 UK Forecast Errors – Mean (median) Absolute Error

(Source: UK Treasury Compilation of Forecasts and Treasury and Civil Service Committee, (1991), forecasts made for the March budget)

N.B. GDP is based on preliminary figures, average estimates of GDP. RPI (Retail price index, 4th quarter) is used for inflation apart from those forecasting groups who forecast averaged inflation over the year. Since 1993 the Treasury have only forecast RPI (excluding interest charges on housing – MIPS) and this has been used for to calculate the Treasury error statistics.

The different forecasting groups have been categorized into 'Independent' – typically units with an academic, not-for-profit institutional base, 'Selected independents' – a Treasury selected sub-set chosen to reflect established opinion and includes some City-based forecasters, and those forecasting groups based in the City of London. The two consensus forecasts are calculated from the average forecasts of the respective groups. The division raises the question of whether there is any difference in performance. Burrell and Hall (1994) conclude their analysis of a similar database by stating that, when forecasting GDP, "on balance neither group has been superior"., although there are periods when one group outperforms the other. In fact, the evidence provided by Britton and Pain (1992) for three macro forecasting groups suggests no consistent 'best performer' for growth, but inflation was apparently best forecast by one of the independents, the CBI, but this strong performance has not persisted in the latter part of the 1990s. The 'selected independents' are no better performer than the 'city average'.

The record when forecasting inflation is complex for it is overpredicted when low and underpredicted when high (in contrast to Burrell and Hall's conclusion, when using a shorter data base, that it is underestimated). In the years 1992-1998 when inflation was low (averaging 2.8%), the City forecasters' average error was –0.6%. When inflation was high the City forecasters were more pessimistic and tended to be more successful (Burrell and Hall, 1994).

Overwhelmingly, the individual forecasters tended to make the same mistakes with errors year by year either all positive or all negative. For example, around 90% of forecasters overpredicted growth in 1990-92 and over predicted inflation for 1991-1994. No single year since 1990 saw the forecasters getting both variables broadly right. Moreover, neither the consensus forecast of the independents nor the City consensus did particularly well. The conclusion we must draw is that despite adhering to different economic theories and using different econometric methodologies they share in making the same mistakes – or at the very least, miss out on the same large, unpredictable shocks, a result that accords with US evidence.

The Treasury's record, despite intense scrutiny, has emerged with a clean bill of health relative to its competitors. Its record is therefore of particular interest and recently, Melliss (1997) and Melliss and Whittaker (1998) have updated and extended the earlier analyses. Over the period 1971-1996 the mean absolute annual forecast error was 1.45% for real GDP (based on latest outturns) and ranged from 1.82% over the 70s to a low of 1.20% for the years 1979-1985. The mean absolute change in GDP over the whole period was 2.10%. However, when standardized this does not compare favorably with other OECD countries (Őller and Barot, 1999). The inflation record had a mean outturn of 8.52% with average absolute error of 1.95% with errors ranging from 3.02% during the 70s to .65% during 1993-1996 compared to mean changes of 13.85% and 2.68% respectively. A rough measure is that for both variables the MAE is approximately 25% of the actual absolute error, approximately the same as in the US.

There is little or no correlation between performance on GDP and performance on inflation, a finding in accord with McNees' (1992a) analysis of the US record.

Extending the countries under consideration, Őller and Barot[9] (1999) have examined the GDP forecasting record of thirteen national forecasting institutes in Europe together with forecasts of the same counties produced by the OECD. The accuracy of the two sets of institutions proved similar. For the period 1971-1995 the average absolute error (compared to the MAD) is 1.43% (1.66%) and ranges from 1.05% (1.46%) for France to 2.12% (2.37%) for Finland.

Quarterly estimates

Less evidence has accumulated on short-term forecasts for the UK with Melliss the only researcher to give a full account (but only of the Treasury forecasts). A summary of his results is given below. They are compatible with the evidence of Holden and Peel (1985) who analyzed the National Institute Forecasts.

Span of	**GDP**		**Inflation (RPI)**	
Forecast horizon	**MAE**	**Theil**	**MAE**	**Theil**
1 Quarter	2.72	1.09	1.36	.31
2	1.90	.87	1.58	.41
4	1.45	.73	1.95	.60
8	1.32	.80	2.73	.69
Year 2 period	-	-	3.51	1.19
12	1.39	.90	-	-

Table 4: Quarterly Error Statistics for the UK Treasury (annualized rates): 1971-1996

Source: Melliss and Whittaker (1998)

Using the MAE as a measure, the Treasury forecasts exhibit the same pattern as was observed in the US predictions. While the absolute forecast error increases with the forecast span for both GDP and the RPI, the annualized errors decrease for GDP while increasing for RPI.

Summary

Burns (1986) provides a convenient comparison of US and UK forecasters, the results of which a reproduced in Table 5 below.

[9] This extends the data based previously analyzed by Ash, Smyth and Herevi (1990).

Source	**76-83 Cumulative Growth**			
Quarters ahead	**2**	**4**	**6**	**8 Quarters**
US (McNees & Ries, 83)	2.4 (3.6)	1.3 (2.4)	1.0 (2.5)	0.9 (2.4)
Average: 6 forecasters				
UK (HM Treasury)	2.4 (4.1)	1.8 (3.3)	1.4 (2.7)	1.3 (2.9)
	76-83 Cumulative inflation			
US (McNees & Ries, 83)	2.0 (2.4)	2.1 (2.6)	2.5 (3.3)	2.9 (4.0)
Average: 6 forecasters				
UK (HM Treasury)	1.7 (4.8)	2.1 (4.4)	2.3 (5.0)	2.2 (5.4)

Table 5 Comparison of US and UK Forecasting Accuracy – MAE%, annualized rates (Burns, 1986)

N.B. Figures in parentheses show the average absolute change over the period, e.g. MAE for the random walk forecast.

Judged by these absolute standards Burns concluded that there was little if any difference in performance between forecasting performance in the two countries (a controversial issue at the time he drew this conclusion when the UK Treasury was under attack). Daub (1987) when comparing Canadian with US forecasting practice drew the same conclusion, explaining the similarities by the fact that innovations in forecasting practice have occurred because of an "efficient-like market in forecasting technology; and the economies are tied together by trade and financial institutions that transmit expectations related thinking". However, Őller and Barot (1999) demonstrate that the quality of the performance in forecasting GDP across 13 European countries is very variable suggesting the importance of economic structure.

3. Judging the Forecasts

There are a number of ways that these forecasts might be judged. First, did the users find them useful? A more conventional criterion, based on quantitative error measures is to determine whether the forecasts have become more accurate over time and whether they were better than alternative predictions that would have been available to the users.

3.1 Qualitative Results

Given the number of forecasters and forecasting services, one way of judging the forecasts is to determine whether the users found them useful. The provision of macroeconomic forecasting services has proved financially successful although there has been some recent reduction in the number of external sources offering macroeconomic forecasts and a number of internal company forecasting groups have also been closed down (Fildes, 1995). In the UK business community two surveys in 1987 and 1994 revealed that the major uses of these macroeconomic forecasts were for medium term financial decision making, financial instrument appraisal, direct investment appraisal, and sales forecasting. It is thus not surprising that when the respondents were asked to identify the information that they most needed for making decisions, they wanted forecasts about economic growth if they were concerned with real variables and predictions about monetary variables when concerned about financial decisions. These surveys also examined the perceived adequacy of the forecasting services and the criteria for selecting a service. The overwhelming impression from the 1994 survey results was that macroeconomic forecasts were of value to the companies. Accuracy was seen as important but the decision requirements (the availability of forecasts of the particular variable of interest for the appropriate time horizon) were dominant, an observation also made by Daub's Canadian respondents (Daub, 1987).[10]

A different view of the value of economic forecasts has been expressed by McCloskey (1983, 1992) which is often summarized by 'if you're so smart, why aren't you rich?'. The core of the argument is that if forecasts had value, those producing the forecasts would utilize them directly themselves without resorting to intermediaries. This view cannot easily be translated into a criticism of those who forecast GDP and are not necessarily in a position to directly benefit from their personal expertise. This criticism is more applicable to interest or exchange rate forecasts (whose accuracy record we have not examined but is typically no better than those derived from time series models, Meese and Rogoff, 1983; Fildes et al, 1985). Perhaps more interestingly in the light of McCloskey's critiques, even the publicly available forecasts of interest rates, despite their poor relative accuracy record, can still prove financially valuable (Leitch and Tanner, 1995).

[10] While the level of dissatisfaction in 1994 was modest it had increased compared to the earlier 1987 survey reflecting perhaps the greater difficulty experienced by forecasters through the recession of the 1990s.

The value of forecasts cannot solely be judged financially. They provide an "organizational tool helping to ensure consistent thinking about the future" by their use as a framework on which to build organizational plans, thereby "spreading the responsibility of senior management", in addition to their PR and marketing function both within and outside the organization (Daub, 1987, pp.72-73).

While the business community has apparently felt well served by the forecasting services industry, governments have been less pleased, often sheltering behind forecast errors to justify their policy mistakes. Cairncross (1969) and Burns (1986), who were both senior economic advisors to the UK government, point out that economic forecasts and their link to policy options are just one of the determinants of government policy. In the United States the published official forecasts have typically been optimistic and the errors have been larger and more biased than the private sector predictions and the unpublished forecasts of the Federal Reserve System (Ulan et al, 1994). The political consequences of policy will dominate most government decisions and as such a publicly announced forecast becomes part of the political process. (See Grunberg and Modigliani (1954) for a discussion of self-fulfilling government predictions). Despite this, there are countervailing pressures on any modeler, whether operating in the public or private sector, to develop the best performing model; bias can be injected when the model is operational. However, the models also serve another, more subtle purpose for government where they offer a framework for sharing knowledge between the modelers and the policy makers, thereby creating value. The models embody 'tacit expertise' offering a forum for consistent discussion. The process of interaction leads to "a new mental picture of how the world is likely to change" (Den Butter and Morgan, 1998).

3.2. Quantitative Results

Rather than relying only on these qualitative data, we now examine the quantitative evidence regarding the quality of the forecasts. First we determine whether the accuracy of the forecasts has improved over time and whether the macro forecasts compare favorably with predictions obtained from naive standards and/or time series benchmarks. In a later section, we also look at the rationality of the GNP and inflation forecasts.

3.2.1 Have the Forecasts Improved Over Time?

Whether the forecasts have improved over time is still an open question: the conclusions depend both on the time periods that are being compared and on the measure of accuracy that is used. The ASA/NBER quarterly surveys produced mixed results: Zarnowitz (1979) indicates that accuracy has improved over time, but Zarnowitz and Braun (1992) do not reach such a definite conclusion. They note that in some cases the RMSEs in the second half of the sample period were smaller than the similar discrepancies for the first half of the period; in other cases the opposite effect was noted. However, Zarnowitz and Braun have not baselined their comparison against the volatility of the variables in the two periods. McNees (1988), in contrast, used a baseline approach and concluded that the Michigan Model's annual forecasts of real GNP had improved over a 35-year period. This analysis showed that the ratio of the RMSEs to the standard deviation of the actual changes declined monotonically from the 1950s to the 1980s.[11]

There is also mixed evidence about this question in the UK studies. First Burns (1986) and more recently Melliss (1997) analyzed this topic by standardizing MAE by the corresponding index of variation (a measure of the difficulty of forecasting). Burns felt the results supported the conclusion that there had been an improvement in accuracy while Melliss noted that in the late 80s and early 90s there were signs of deterioration. The conclusions of Britton and Pain (1992) are mixed. They noted that there was no noticeable deterioration in the quality of the GDP forecasts, but that the domestic demand errors had become larger. In neither case, however, were the errors adjusted for the inherent difficulties in forecasting those variables.

In Öller and Barot's analysis of 13 European counties they could not find any evidence of improvements in accuracy when they standardised the error statistics (using RMSE/SD and MAE/MAD).

In determining whether the forecasts have improved over time, various descriptive statistics have been used to measure the difficulty of forecasting over different time periods. This method of

[11] While McNees indicates that "it is difficult to make a case that nominal GNP forecasts have become more accurate over time", the same ratios that were used for the real GNP forecasts also decline for both the nominal GNP and inflation predictions. This type of ratio for determining whether accuracy has improved over time was first used by Smyth and Ash (1981). Keen's (1987) analysis of the industrial production projections in the Livingston surveys also yields mixed results regarding the trend in accuracy over time.

standardization may not capture the difficulty of preparing forecasts in any specific time period. A more appropriate approach is to consider performance relative to a suitable benchmark forecasting method (e.g. a univariate ARIMA model or VAR, see below) which, as Fildes (1992) shows, has the potential to eliminate heteroscedastic effects arising from the peculiarities of the particular period under study.

3.2.2 Comparison with Naive Standard; U coefficients

In comparing a set of forecasts against a naive standard, the usual procedure is to calculate Theil's U coefficient. If it is less than 1 in value, the forecasts have smaller MSEs than would have been obtained from a simple no-change naive model and the forecasts have at least met the minimum standard of not being inferior to naive predictions. Where the variable is monotonic (e.g. nominal GNP) U will inevitably be less than one and, to compensate, the denominator is sometimes replaced by an extrapolation of last period's changes.

The results are in general agreement: all of the US forecasts have U coefficients that are less than one, thus indicating that they are clearly not inferior to the naive extrapolations from these minimum standards of comparison. However the comparison is not stringent and some of the Us are close to or larger than 0.9. (McNees and Ries, 1983; Zarnowitz, 1992a). Moreover, U is only a descriptive statistic. Consequently, studies that only present values of U do not indicate whether or not a particular set of forecasts is significantly better than those obtained from the naive model. No US studies have examined longer horizons.

In the UK the UK Treasury's forecasts for GDP growth had Theil's U <1 for forecast horizons as far ahead as 2 years (averaging .60 for the 12 quarters ahead forecasting and remaining robust across sub-periods for growth, Melliss, 1997). Relative performance of forecasts of GDP growth over the period improved as the forecasting horizon lengthened up to a year ahead.

For inflation, Melliss' study yielded Us that were <1 only up to one year ahead. There was less consistency in the longer-term inflation forecasts where in periods of increasing inflation, 1971-79, and 1985-89 consistent underprediction led to poor relative performance longer term. The U coefficients for the NIESR forecasts of GDP and inflation were also less than one (Britton and Pain, 1992). Holden and Thompson's (1997) examination of four macro forecasting groups

yielded Us close to 1 for GDP (average of .79) for horizons longer than one year and these coefficients increased for the longer horizons. For inflation, with one exception, (Cambridge) the Us are substantially lower than 1 with average .56 but, again, they increased with the length of the forecast horizon.

3.3. Comparison with Time Series Benchmarks

The naive models are not very stringent standards against which forecasts should be compared. Time series models, such as ARIMA, VARS and Bayesian VARS (BVARs), are relatively simple to construct and can be used to generate predictions. Consequently, forecast evaluations have used these models as benchmarks to judge the accuracy of economic predictions.

3.3.1 US Forecasts

Nelson (1972) was among the first to compare the forecasts of an ARIMA model with those generated by econometric models. The evidence in that study indicated that the performance of a major model was not markedly superior to that of the ARIMA benchmark. Other studies that compared the forecasts of ARIMA models with the predictions of econometric models for a wider class of variables included Cicarelli and Narayan (1980) and Dhrymes et al (1988). Unfortunately, these studies yielded conflicting evidence. Nevertheless, Nelson's results led him to prepare ex-ante real time predictions using ARIMA models. These forecasts are now generated by Fred Joutz and have been used as a benchmark with which other predictions are compared (Nelson, 1984; Joutz, 1988; McNees, 1992a).

The Nelson and Joutz studies showed that from 1976.2 to 1985.4 the structural forecasts of growth rates and inflation at the shorter horizons were more accurate than those of the ARIMA models[12]. As the horizon lengthened to three and four quarters the accuracy of the Benchmark predictions not only improved relative to the others but also were among the best. This was explained as possible poor dynamic specifications of the structural models that were corrected in the short run projections by judgmental adjustments. An additional explanation is that the use of structural models requires forecasts of exogenous variables that may have been difficult to predict.

[12] On the other hand, Cicarelli and Narayan (1980) show that for a portion of the 1970s, the ARIMA models were superior to eleven econometric models. While the ARIMA models were selected ex post, even a naive model (N2) also outperformed these structural models.

However, McNees' more recent results showed that from 1986.1 through 1991.3 structural models forecast the rate of growth of real GNP more accurately than the Benchmark time series alternative. Moreover, the relative accuracy of the structural models increased as the time horizon lengthened. On the other hand, a comparison of the time series predictions of the implicit GNP deflator with those obtained from structural models showed that at all horizons the former was more accurate than about half of the model based forecasts.

It is possible to provide only a partial explanation of the differences between the Nelson and Joutz results on the one-hand and McNees' findings on the other. First, the time periods of the two evaluations were not identical. In addition, McNees evaluated the forecasts over a span of n quarters while the other studies considered the accuracy of predictions for a particular quarter made n periods in advance. There have been few tests to determine whether the differences in accuracy were statistically significant. Perhaps most important, developments in econometric dynamics were being given increasing attention in the models over this period which should have led model based forecasts to gain the upper hand (Wallis, 1989; Diebold, 1998).

Zarnowitz and Braun (1992) compared the ASA/NBER forecasts with this Benchmark. They found that the accuracy of the two sets of predictions was comparable. We can, therefore, conclude that for real growth and inflation the magnitude of the errors is similar and that an autoregressive model is an appropriate standard of comparison even if it might not be the model of choice for forecasting.

Forecasts may also be compared with the predictions obtained from multivariate time series models. Again the conclusions are in conflict. Lupoletti and Webb (1986) compared the ex post forecasts of a VAR with the ex ante forecasts of three forecasting services for the period 1970-1983. The VAR was estimated via rolling regressions although the model was based on after the fact-revised data, which were not available to the forecasting services that were forced to make their forecasts in real-time. For this period, over all horizons, all the forecasting services predicted real GNP better than did the VAR. The advantage was greatest at the short horizons, which is particularly significant given that the VARs had a data advantage over the forecasting services. All of the professional forecasters also predicted inflation more accurately one quarter

ahead and made comparable errors for the longer horizons.[13] In fact, Webb (1995) showed that inflation forecasts from VARs were significantly worse than naive predictions.

Litterman (1986) and McNees (1986) both compared the accuracy of ex ante forecasts obtained from structural models with those generated by one particular BVAR model for the period 1980-85. The results of the two studies were similar even though they differed in both the numbers of structural models, which were compared, and the analytical procedures that were used. The real GNP forecasts of the structural models were more accurate for the one-quarter ahead horizon, but thereafter the BVAR predictions made substantially smaller errors with the relative difference increasing with the length of the forecast horizon. On the other hand, the inflation predictions of the structural models were substantially superior to those of the BVAR with the relative performance of the structural models improving the longer the horizon. (This finding is similar to that found in the univariate time series-structural model comparisons.)

Zarnowitz and Braun (1992) also used several VARs to evaluate the relative accuracy of the ASA/NBER predictions. While the final data used to construct those VARs differed from the numbers used in real time by individual forecasters, the accuracy of the forecasts of the VARs was similar to that of the ex ante ASA/NBER surveys.

On the basis of these studies we must conclude that for the US, time series models provide competitive benchmarks for forecast evaluation and encompassing tests, sometimes outperforming more complex structural alternatives.

3.3.2 UK Forecasts and Other Evidence

In the UK the ARIMA model class again provided strong competition to the performance of the structural models of the 1970s and provided prima facie evidence that the London Business School Model was dynamically mis-specified (Longbottom and Holly, 1985). While the structural models performed better in predicting real growth rates and inflation, the ARIMA models made better forecasts for other series that are not included in this review.

[13] The results of this study also indicated that the forecast accuracy of ARIMA models and this VAR was comparable.

Wallis (1989), Holden and Broomhead (1990) and Holden (1997) have all compared VARs with the forecasts from various modeling services. Wallis indicated that the forecasts of the VARs were not clearly superior to those obtained from structural models. However, the benchmark has to be carefully specified because BVARs generally outperform their unconstrained alternatives and some approaches to specification are better than others (Allen and Fildes, 2000). In a comparison with structural models Holden and Broomhead (1990) added supplementary evidence showing that (constrained) BVARs outperformed their unconstrained alternatives[14]. Even using the BVARS, Holden (1997) found few instances where they outperformed their econometric competitors, either in magnitude or direction. However, combining BVAR forecasts with the econometric forecasts showed they contained information (presumably their dynamic structure) that was not in the structural model forecasts.

Artis and Zhang (1990) compared BVAR forecasts for the G-7 group of countries with those produced by the IMF in the World Economic Outlook (WEO). To make the comparison fair relies on two critical factors: (i) assumptions about the information that a forecaster might have about current quarter conditions and (ii) the data that are used in building the BVAR. The superior performance of the WEO is partially attributable to knowledge of current economic activity as Table 6 shows because the BVAR's performance improved when it used an additional quarter's information. On the other hand, the BVAR was based on a single set of revised data, which would not have been available to forecasters at the time they made their operational forecasts.

	Current year/ one-step ahead		**Year ahead/ two-step-ahead**	
Model	BVAR	BVAR+	BVAR	BVAR+
Growth	1/7	2/7	2/7	3/7
Inflation	2/7	5/7	3/7	3/7

Table 6 The Ratio of G-7 Countries for which the BVAR forecasts were superior to the IMF's, 1980-1987.

N.B. The BVAR+ forecasts include an extra quarter of data.

[14] There is a model selection problem for the multivariate models as well as for the univariate time series equations.

The results show that the BVAR models provide a rigorous benchmark with which to evaluate conventional macro forecasts produced by the WEO which Artis and Zhang note are heavily weighted towards the use of judgement rather than formal methods.

Zellner and his colleagues have extended the concept of what constitutes a benchmark model by (1) adding additional variables to an autoregressive model, (2) estimating the models over a cross-section of OECD countries, and (3) using time varying parameter models (Garcia-Ferrer et al, 1987, Zellner and Hong, 1989). All these extensions improved the fit and accuracy of the benchmark time series model. When the one-year ahead OECD forecasts were compared with the predictions of this final model, the time series model was superior. (The comparison is again not quite fair in that this year's growth data would not have been available to the operational forecasters of the OECD; only preliminary estimates would have been available. The models were again based on a single set of revised data.)

3.4. Summary of the Results

In judging the forecasts, qualitative appraisals have indicated that users found the macroeconomic predictions useful. The quantitative findings indicated that all of the model-based forecasts outperformed naive models, but that the predictive accuracy of time series models approximated that of the structural models and of the ASA/NBER surveys. In any event, the time series models are useful benchmarks for judging other types of forecasts. In addition, surveys have shown that macro forecasts are used as inputs into more disaggregate forecasts and decisions through conditional predictions. How useful they are is an open question. In an attempt to answer this Ashley (1983,1988) has proposed a MSE based criterion for a macro forecast to improve on the accuracy of a disaggregate autoregressive model. In an analysis of US forecasting services few variables met this criterion beyond a year ahead: broadly cumulative GNP and inflation. UK annual data were more satisfactory. Empirical evidence, summarized in Allen and Fildes (2000) suggest that, ex ante, macro forecasts are in fact helpful in disaggregate forecasting. In summary, macro model based forecasts are useful both directly and indirectly in forecasting and planning compared to the alternatives.

4. Could the Forecasts Have Been Improved? Rationality Tests

The naive and time series comparisons merely provide some benchmarks for judging a set of predictions. It is possible to ask another question that provides evidence on the limitations of the forecasting services: Could the forecasts have been improved? Suppose that ex post rationality studies show that the forecasts were biased or had not taken into account information available to them. Then, in principle, the accuracy of the ex ante predictions could have been improved if this information had been taken into account (although this would not necessarily be possible in practice, as the optimal weights would be unknown).

The results obtained from these studies differ depending on a number of factors, including whether the analysis was performed on the "consensus" forecast obtained from a survey or from the individual forecasters who constituted the cross section; the date that the studies were undertaken; and the econometric methods that were used in the analysis. For example, Rich (1989) pointed out that most of the cited studies had analyzed aggregate or pooled forecasts. On the other hand, Zarnowitz (1985), Keane and Runkle (1990), Batchelor and Dua (1991) and Davies and Lahiri (1995, 1999) examined the predictions of the individuals in the survey. Batchelor and Dua (1991) summarize the different approaches that have been taken in testing the rationality of the expectations or forecasts from the Livingston, ASA-NBER, and Blue Chip surveys. While the aforementioned studies analyzed the rationality of forecasts obtained from surveys, McNees (1978) and Joutz (1988) tested the forecasts of particular econometric models for rationality.

Studies of the rationality of UK forecasts have concentrated on the institutional producers of econometric model based forecasts, (Melliss, 1997; Melliss and Whittaker; 1998, - the UK Treasury; Holden and Peel, 1985; Britton and Pain; 1992, - NIESR; Holden and Thompson, 1997, - LBS, Liverpool, Cambridge, NIESR; Mills and Pepper, 1999, - LBS, NIESR, UK Treasury). However, a recent study by Egginton (1999) has analyzed various individual City of London based forecasters.

4.1.1 Results: Inflation - US

In the US studies in which aggregated or consensus forecasts were examined, the hypothesis that the inflation predictions were rational was rejected in most instances.[15] Rich's (1989) analysis of the SRC data and the Batchelor-Dua study of the consensus Blue Chip forecasts are exceptions. The former study found that the mean forecast did not violate the unbiasedness, efficiency, and orthogonality conditions. The latter were only unbiased.

Zarnowitz (1985), using disaggregated forecasts, also rejected the rationality hypothesis for the inflation predictions of many of the individuals in the ASA/NBER survey. Keane and Runkle (1990), however, showed that a smaller number were not rational. The difference was that Keane and Runkle used the Generalised Method of Moments estimation technique and the early actual data, while Zarnowitz used the final revised numbers. In turn Bonham and Cohen (1995) and Davies and Lahiri (1999) have challenged the Keane and Runkle methodology and conclusions. Davies and Lahiri found that 30 of 45 forecasters in their ASA/NBER sample failed the rationality test.

Batchelor and Dua show that the participants in the Blue Chip survey made unbiased inflation forecasts, but that most failed one or more of the other rationality tests. In particular, the individuals failed to take account of published information and of the information contained in their forecasts of other variables. The performance of these forecasters in predicting inflation is notably worse than their record in forecasting real growth (and other variables). Batchelor and Dua provide two possible explanations of their results. First, 'irrational' forecasters tended to place more importance upon non-traditional economic theories than did the 'rational' forecasters. They also placed greater reliance upon econometric models and assigned lower weights to judgmental adjustments. Using a different methodology, Davies and Lahiri (1999) found that more than half of the Blue Chip forecasters showed significant bias.

The McNees (1978) and Joutz (1988) analyses of the inflation forecasts of individual econometric models also yield conflicting results. McNees showed that some of the one-quarter ahead

[15] Hafer and Resler (1982), however, divided the entire Livingston survey into six subsets depending upon the professional affiliation of the survey participants. They found that the rationality hypothesis was not rejected for the consensus forecasts of only one of these subsets.

forecasts of inflation are unbiased but that most of the multi-period forecasts fail this test. Moreover, the results also depend on the econometric technique used to estimate the relationship between the predicted and actual outcomes. On the other hand, Joutz showed that, for a longer time period, all six of the econometric models in his sample produce unbiased inflation predictions, i.e. weak form informational efficiency.

In addition, Joutz' study examined a large number of series that would have been available to the modelers and shows that the information contained in these series was not used efficiently in a significant number of cases.[16] In fact, he suggested that these inflation forecasts could have been improved if important variables such as the degree of capacity utilization and labor market conditions had been more fully specified.

4.1.2 Results: Inflation – UK and Other Evidence

The UK evidence is broadly supportive of unbiasedness (Holden and Thompson, 1997, Melliss, 1997, Mills and Pepper, 1999); however there were sub-periods where it is rejected. In certain periods both the National Institute forecasting team (Britton and Pain, 1992) and the Treasury (Melliss , 1997; Melliss and Whittaker, 1998) had consistent over and underestimates. Melliss also tested for strong form informational efficiency, orthogonality, and examined whether past values of other variables or past forecast errors could have improved forecast accuracy. While there were many examples where such relationships were observed in particular sub-periods for different forecast horizons, no consistent effects were observed. Specifically, Melliss (1997) did find some evidence to support the view that longer-term forecasts of inflation could have been improved by including a capacity utilization variable. (This is similar to Joutz' results for the US.) In addition, Mills and Pepper found increasing irrationality for longer horizons. Although Eggington (1999) argued that individual short-term City forecasters were both biased and inefficient, the anomalies were small. By analyzing the successive updating of fixed event forecasts, Melliss and Whittaker (1998) again find evidence of irrationality whereby forecasts have been inadequately updated.

[16] The models in this sample have a smaller number of rejections of this rationality test if revised rather than preliminary data are used perhaps because the objectives of the forecasters were the prediction of the 'final' values. For the UK Britton and Pain (1992, pp.18-19) found data revisions were not significant influences.

There is only limited evidence about the rationality of the forecasts made in other nations. Kirchgassner (1993) examined the 6 month and 2nd half-year ahead predictions of growth, inflation and other real and nominal macroeconomic indicators made by various German Research Institutes. While he found no evidence for irrationality in the 6-month predictions, rationality was consistently rejected for the 2 step ahead predictions. Ash et al (1990) examined the rationality of OECD forecasts of a number of variables for the G-7 members. The inflation forecasts could have been improved by using additional available information for Japan, Germany and France.

4.2.1. Results: Growth Forecasts - US

Although there have been fewer analyses of the rationality of real growth forecasts, the evidence regarding the rationality of the US real GNP predictions is also mixed. Zarnowitz (1985) concludes that most of the individuals in the NBER/ASA surveys made rational forecasts, while Batchelor and Dua (1990) show that the predictions of only 7 of 19 participants of the Blue Chip survey passed all of the rationality tests.[17] The Blue Chip consensus forecast passed all but one of the tests.

McNees' (1978) study of three econometric models yielded mixed results similar to those associated with the inflation forecasts. Joutz (1988) discovered that the hypothesis of unbiasedness was rejected for only two of the six models, but that information was often used inefficiently. The use of monetary data could have improved these forecasts.

4.2.2. Results: Growth Forecasts – UK and Other Evidence

Bias is not a particularly common feature of the growth forecasts (Melliss, 1997 analyzing the UK Treasury 1-8 quarters ahead; Holden and Thompson, 1997 analyzing 4 macroeconomic model institutes 0 to 4 years ahead). However when efficiency is considered through the orthogonality criterion, Melliss 's (1997) evidence on growth is mixed (as with inflation) with no consistent pattern emerging, the results often being period specific. He concludes "forecast inefficiency has been a common feature" and the failure in the orthogonality criterion was most common when the rate of growth preceding the forecast was included. This points to poor dynamic specification.

[17] Using a different methodology, Davies and Lahiri (1995) found that the forecasts of half of the Blue Chip participants were biased.

Ash et al (1990) examined the rationality of OECD forecasts for the G-7 members. The forecasts were "relatively free of bias", but they often failed efficiency and consistency tests. In contrast to inflation, the errors in growth forecasts were non-systematic. Of the ten variables forecast Ash et al (1990) showed Government consumption to be by far the most suspect, perhaps because of its explicit political nature (Bretschneider and Gorr, 1992). They summarize the results by pointing out that only 14 of 84 series pass all rationality criteria and this drops to 5/60 when North America is excluded. In more recent work (using a similar database) Ash et al (1998) found the evidence supportive of rationality based on analyzing directional forecast evidence although they regard this as only a weak test.

Öller and Barot (1999) examined the one year ahead growth forecasts made for 13 European countries and found very little evidence of either bias or weak form inefficiency. This contrasts with Kirchgassner's results. This difference may be explained by Öller and Barot's analysis of the annual errors while Kirchgassner examined the six month (efficient) and the 2nd half year (inefficient) errors.

4.3. Conclusions regarding Rationality

The results of these studies show that a number of individuals and models have generated rational forecasts. These studies also demonstrate that other forecasters or models were not able to do this, perhaps acknowledging conflicting objectives such as the reputational value of differing from the consensus[18] (Batchelor and Dua, 1990, Spencer and Huston, 1993) or the desirability of smoothing their forecasts over time (Scotese, 1994). Performance over different sub-periods (and different economic conditions) is, however, inconsistent. A meta-analysis about the findings of rationality studies (Goldfarb and Stekler, 1998) shows that the overall conclusions varied from year to year and the time horizon of the forecast. Inflation changes (both on the upside and downside) are regularly mis-forecast (Diebold et al, 1997). Performance is also country specific with North American forecasts more likely to be rational than in other parts of the world including

[18] Burrell and Hall (1994) argue that UK City forecasters should be more likely than academic forecasters to recognize the possible benefits of differing from the consensus but fail to find any support for this in their data.

the remainder of OECD. For the UK, growth and inflation are both "poorly forecast with both serious bouts of inflation and deep recession underpredicted" (Mills and Pepper, 1999).

Nevertheless, the use of rationality tests, especially those that relate the forecast errors to known information, can be viewed as important diagnostic checks to determine why the errors occurred and to improve the forecasting process and the quality of subsequent predictions. It would also be of interest to determine why there are apparent differences between the characteristics of the inflation and real GNP predictions.

5. Choosing between Forecasters (or Methods) - or a Combination

Our discussion has suggested that most forecasting methods yield similar results. A number of studies have explicitly examined this issue of whether there is a best model, method or forecaster. Zarnowitz and Braun (1992) indicate that, in the US, the forecasting performance of any one individual relative to another is highly variable over time. A similar conclusion holds for the UK (Holden and Peel, 1986; Holden and Thompson, 1997). The clear implication is that there is no consistent best performer. McNees' (1992c) study supports this conclusion. It may be more appropriate to ask an alternative question: Are there some forecasters who are better or worse on average? (See Stekler, (1987) and Batchelor (1990) for a discussion of an appropriate methodology for analyzing this question). McNees (1992c) notes that while there is no forecaster or method, which is, best all of the time, there were a few forecasting services which consistently performed poorly.

The panel of forecasts summarized by the UK Treasury is broken into various affiliated groups, City, Independent and 'Selected Independent' and while early researchers had summarized the evidence as showing little difference between groups, for both growth and inflation it would seem that the 'Selected Independents' outperform the other groups.

Given that there is no "best" forecast, it is natural to determine whether a combination of forecasts might produce better results. The combined forecast is of the form:

$$Y_t^* = \alpha_t + \sum_i \beta_{it} \hat{Y}_{it}$$

where Y_t^* is the combined forecast of the variable obtained from the i individual sources. While considerable research has been undertaken to determine how best to choose the coefficients, α_t and β_{it}, the evidence suggests that the simple approach of averaging the individual predictions works well (Lupoletti and Webb, 1986; Clemen and Winkler, 1986; Clemen, 1989). Non-linear combinations are also possible.

In most cases, the combination of predictions reduces the out-of-sample errors. Although the seminal article on forecast combination is considered to be the one by Bates and Granger (1969), Zarnowitz (1967) had earlier noted that the average of several GNP forecasts was better than the individual predictions. Similarly, Nelson (1972) and Cooper and Nelson (1975) showed that when econometric forecasts were combined with ARIMA estimates, the combined forecast had smaller errors than were obtained from the models alone. More recent evidence also suggest that combining typically leads to improved accuracy. (Lupoletti and Webb, 1986; Clemen and Winkler, 1986). Clemen's (1989) review article cites additional studies that confirm these results.[19] The benefits of combining are greatest when forecasts that are generated from different methods or theories are combined (Batchelor and Dua, 1995). In the same vein, Graham (1996) pointed out that when the forecasts of economists are correlated, the effective number of independent forecasts is reduced.

Interpretation of the results on the effectiveness of combining is sometimes controversial where the combination underperforms the best of the model-based forecasts, as is the case in for example, Holden and Thompson (1997). However, the dominance of a particular model compared to the combination presupposes it was identifiable a priori as the best performer.

Some econometricians argue against combining forecasts. The model builders visualize the process of using linear regressions to combine forecasts merely as tests of the specification of the alternative models. They argue that the 'true' model should be able to explain the results of other models-i.e. by using encompassing tests. (See Chong and Hendry, 1986; Diebold, 1989). This test

[19] The combination of poor individual forecasts may still yield a poor combined prediction. See Lupoletti and Webb (1986).

may be performed by regressing the actual values on the forecasts of the two models, i.e.

$$\hat{Y}_t = \alpha + \beta_1 \hat{Y}_{1t} + \beta_2 \hat{Y}_{2t}$$

and then determining whether the vector of coefficients $(\alpha, \beta_1, \beta_2)$ is (0,1,0), in which case Model 1 encompasses Model 2, or whether it is (0,0,1), in which case the second encompasses the first. If any other result is obtained, there is useful information in both models and both models are mis-specified[20]. (This test is similar to the conditional efficiency rationality test where a second forecast is included in a rationality test for efficiency to see if it adds any information.) Hence, some model builders advocate that more effort should be devoted to obtaining the true model specification rather than combining the forecasts - although current users of the forecasts would no doubt prefer improved forecasts. In fact the incremental information in the model-based forecasts may well change relative to an extrapolative benchmark as the time horizon lengthens (Clemen and Guerard, 1989).

An alternative approach is to regress the forecast change on the predicted changes of the two models and perform the same tests on the coefficients. Fair and Shiller, (1990) used this approach in demonstrating that the one quarter ahead forecasts of the 1976 version of the Fair model encompassed those of a variety of time series and VAR models while the Fair model's four quarter ahead forecast did not.

6. The Role of Judgment and Forecast Error Decomposition

The role that judgment plays in the forecasting process has long been a subject of analysis. In the 1960s the question that was of interest was whether judgmental or model based forecasts were more accurate. This question is not posed any longer, for forecasting methods are rarely used in their pure form. Instead they are used in informal combinations, with all individuals using at least some quantitative techniques, and with most modelers using judgment to modify their model-based forecasts.

The reasons why a model builder may choose to modify the model can be considered under four headings: (1) structural breaks, where changes have occurred in a particular equation that

post-date the sample data on which the model was estimated; (2) parameter estimation errors, where part of a recently observed residual may be due to a change in a parameter during the post-estimation sample, for example a change in the institutional framework within which wages are determined, e.g. an incomes policy which leads to a constraint on a parameter in the model; (3) measurement errors, which typically arise when data used in the model based forecasts are subject to major revision, and (4) current information derived from sources outside the model such as monthly data that give an insight into possible inadequacies in the model based forecast.

$$ey = Y - f(X,0)$$

$$- \{f(X, Ad) - f(X,0)\}$$

$$+ \{f(\hat{X}, Ad) - f(X, Ad)\}$$

$$Y - f(\hat{X}, Ad)$$

The residuals of particular equations are modified judgmentally to take account of factors that may be considered either as signs of shortcomings in the model or of data peculiarities. In practice, the residuals are modified in a number of distinct ways (Young, 1982; Turner, 1990). Often non-zero residuals are incorporated into the final forecast, using either mechanical rules or subjective estimates. Todd (1992) has pointed out that the reasons that justify such an intervention should be stated explicitly and Moss et al. (1994) have taken this further by proposing the use of an expert systems approach that codifies the basis for making such adjustments to the model.

In order to determine both the effectiveness of these judgmental adjustments and to suggest procedures for improving the accuracy of model based forecasts, it is necessary to decompose the forecast error into various sources. The forecast f depends on the econometric specification, the exogenous variables, *X*, and the corresponding forecasts of the exogenous variables, $\hat{X}$, and any adjustment, Ad, the forecaster might make to the model output Forecast error can therefore be decomposed into:

	Model error (*ey*)
Less	Contribution of residual adjustments
Plus	
	Contribution of exogenous variable errors

[20] See Harvey et al (1998) for a test that is robust to non-normality, a common problem.

= Forecast error

.

In the absence of exogenous variables in the system, it would be possible to determine the value of judgment merely by comparing the ex ante forecasts, $f(\hat{X}, Ad)$ which incorporated judgmental adjustments, with the actual ex post unadjusted model projections, *f(X,0)*. However, the effects of predicting the exogenous variables also make it more difficult to evaluate the role of judgment in model based forecasts. To take these effects into account Wallis and Whitley (1991) proposed that the ex post mechanical forecasts derived from the actual values of the exogenous variables be compared with mechanical forecasts based on the assumed values of the exogenous variables used in the ex ante predictions, i.e. $f(X,0) - f(\hat{X},0)$. This leads us to conclude that in any evaluation of ex ante forecasting accuracy, in the typical case where the forecaster has used judgment in model building, as the error decomposition shows, the error deriving from the model cannot be uniquely distinguished from the judgmental effects.[21]

In addition, such decompositions raise a number of substantial issues for evaluating forecasts and determining how best to improve their accuracy. Of the various sources of error, the question of which is most important, is moot. In particular, the accuracy of a particular set of macroeconomic forecasts arises from (i) the quality of the model including its structure and its robustness to mis-specification, (ii) the judgement of the modeler or (iii) the forecasts of the conditioning variables. Analyses of these issues show that the exogenous assumptions are used by model builders to subjectively influence their forecasts. This finding is obtained from studies that have substituted the ex post actual values of the exogenous variables for the ex ante estimates with a concomitant deterioration in forecasting accuracy. They also show that the raw forecasts differ by more than their adjusted counterparts – the judgmental adjustments lead to convergence. This (together with the residual adjustments) explains the often quoted conclusion that ex ante macro forecasts are more accurate than the corresponding ex post forecasts. (Christ, 1975, was among the first to reach this conclusion.)

[21]The problem becomes even more complex if, as Hendry (1997) proposes, the model error decomposition incorporates various mis-specifications including parameter and intercept mis-specification over the forecast period.

In summary, the role of judgmental adjustments in the development of macroeconomic models is controversial, with Fair, among others, highly critical of subjective intervention. In forecasting applications, however, where the aim is improved accuracy it is hard to argue against subjective intervention as a matter of principle. Rather, the issue is whether the adjustments improve on the mechanical forecasts. Fair (1984), Zarnowitz (1992a, pp.408-410) and McNees (1982, 1990) have concerned themselves with whether judgmental adjustments are effective, concluding that the evidence is weakly in their favor. However, the adjustments made were not optimal. As McNees (1990) notes, the prevailing tendency is "to place too much weight on the specific circumstances and too little on the model. " In other words the forecasters overcompensated. Wallis (1986, 1987) and Wallis (1989) analyzed their effectiveness and concluded, "the adjustments reduce the size of forecast errors". Donihue's (1993) analysis of the Michigan model yielded the same result. The general picture is one of 'value-added by forecasters to the accuracy of forecasts'. Such a conclusion can also be found in the more general forecasting literature (Armstrong and Collopy, 1998). In summary, any evaluation of a macro model's forecasts cannot easily be disentangled from the judgment of the people running the model.

7. Discussion and Suggestions for Improving Forecasting

Based on this survey of the published results of forecast evaluations, what can we conclude about the state of macroeconomic forecasting? We have found that most forecasters fail to predict recessions in advance and sometimes fail to recognize them contemporaneously. Forecasters also seem to make systematic errors such as underestimating growth during periods of economic expansion, overestimating it during declines, underpredicting inflation when it is accelerating and overpredicting when it is decelerating.[22]

In terms of the quantitative measures of forecast errors we have found mixed results. Despite the relatively large errors observed, almost all forecasts are superior to the predictions that could have been obtained from naive models and often better than those generated by time series models. In most comparisons, however, the results are not tested for statistical significance, and we can not determine whether one set of forecasts is significantly better than another ex post nor can we identify ex ante the future best performers. For significant periods, particularly for inflation,

[22] It has been suggested that such lags associated with regime changes might be associated with "rational learning".

inflation forecasts are 'irrational' and could be improved by the efficient use of available information. In addition, there is conflicting evidence as to whether the forecasts have improved over time.

There is one result about which there is general agreement; namely no one forecasting method or one model or one individual does best all of the time. This has led to one effective suggestion for improving the quality of the forecasts: combine the predictions of different individuals or methods. We also found that there was agreement that the use of judgmental adjustments improved the quality of model forecasts. This finding is in accord with the generally accepted practices of model builders.

Both theoretical and empirical studies have shown economic forecasting to be potentially valuable (and market arguments show that it is of value to organizations). The economic benefits of improved forecast accuracy have been examined in a number of contexts, for example, in electricity supply (Price and Sharp, 1986) and steel (Daub, 1974). Both studies concluded that improved accuracy, if achievable, would have been extremely valuable. However, this leaves open the question of the value of specific macro forecasts of GNP. After reporting that even poor forecasts of interest rates (relative to naïve competitors) had value through various trading rules (Leitch and Tanner, 1995) they go on to argue that directional GNP forecasts are similarly valuable, a point supported by Öller and Barot (2000). But this field of study has received too little attention from the academic community. Exceptions are the financial valuation models, which directly integrate ex ante forecasting. This survey has reinforced the beliefs expressed by Leontief (1971) and Mayer (1980) among others that that our models are theoretically elegant representations of abstract economies and that the profession gives too little attention to the feedback the real world offers to highlight the inadequacies of such models. Even Chow tests for instability, a common enough regression diagnostic, are rarely reported (Allen and Fildes, 2000) and the issue of how to respond to a 'rejection' barely addressed.

Given these findings about the quality of macroeconomic forecasts, it is important to note what has either already been done to improve their accuracy or what further research has been suggested. First, what has already been done? The major modeling failures, in particular those caused by the oil shocks of the 1970s led to various theoretical enhancements of the models to

include supply side variables. Some forecast failures, for example, the recession of 1979-1981 in the UK was explainable ex post by existing models, however other events, such as the fall of oil prices have led to a reappraisal of the price response function (Wallis, 1987). The Lucas critique and the notion of rational expectations led to the incorporation of forward-looking expectations into most macro models. The critique itself, once the issue of the modeling of such expectations based decision rules is addressed, does not in itself explain forecast failure (Wallis, 1989).

Such economic theoretic developments embodied in the macro models have been matched by statistical innovations. The use of time series benchmarks had pointed out some inadequacies of structural models especially their dynamic specifications. The comparison of model forecasts with those generated by time series; thus led to improved model specifications as well as a more refined evaluation methodology based on co-integration.

The effect of these innovations should have been improved performance (relative to some suitable benchmark). Diebold (1998) has taken a sanguine view of the success of the modeling effort, describing the developments in economic theory and time series econometrics as a successful tradition. His " . .future of macroeconomic forecasting" is much like his past – more theoretical improvements derived from 'dynamic stochastic equilibrium models', joined with improvements in estimation for such models (rather than the current reliance on calibration). The evidence we report contradicts him – the past history which we recount is disappointing and there is little reason to think that the directions he suggests for the modeling effort are likely to be productive in that they follow a path that has been much explored without noticeable returns. (If such a comment reads overly pessimistically, it is a counterbalance to Diebold's enthusiasm for technological fixes following the established paradigm of research).

Others in the profession have shown more interest in new methodologies for reducing forecasting errors. These approaches are based on abandoning the premise that there is a 'correct' structural model that is stable over time. A number of valuable insights about methods for improving econometric forecasts have been provided by Granger (1995) and Clements and Hendry (1998a, 1998b) in their on-going research program. Hendry (1997) and Clements and Hendry (1998b) have defined the sources of quantitative forecast errors under the assumption that a structural change may occur and that econometric models are being used to generate those predictions. The

key source of forecasting failure, they argue, is deterministic shifts (in the level or trend equilibrium value) and not parameter instability, which is of lesser significance. An inadequate model (not 'data-congruent' in their terms) can therefore perform substantially better than the 'correct model' if it is insensitive to any structural break[23]. These studies have improved our understanding of why structural breaks can lead to models that ex post reflected the data generation process, but when used ex ante perform relatively poorly. They then suggest a number of econometric techniques that can improve the accuracy of forecasts in the presence of structural breaks. Ironically, the use of judgmental adjustments, a procedure that has long been utilized by model builders and, as summarized above, does indeed improve the accuracy of forecasts because the adjustment is in effect an 'intercept correction' that has the potential to correct for a structural break. In a world where such structural breaks are common in the forecast period, non-causal models will often perform well and the imposition of unit roots in conflict with the in-sample evidence improves accuracy as the model immediately corrects for the intercept shift.

Others (e.g. Ormerod in Granger at al., 1998) suggest that the failures we document arise from the non-linearity of economic time series. The problem that arises is that data observed in one part of a non-linear system is not informative as to behavior elsewhere (Ormerod and Campbell, 1998). This issue of identifiability is compounded by the potential dimensionality of the problem where, as Keynes remarked, the multitude of related factors, their interaction and the niceties of timing all affect the pace with which the business cycle develops. Attempts to capture elements of this have not, so far, proved successful (Birchenhall et al., 1999). Swanson and White (1997) show that, in general, a flexible specification (where parameters and variables are permitted to enter and exit from a single equation VAR) is more helpful than the inclusion of non-linear effects. The theoretical arguments as to why non-linear (perhaps chaotic) processes may be important are also controversial with Clements and Hendry (1998a) dismissing them as unconvincing. They argue that well-behaved near linear models can be found equivalent to the non-linear representations and these simpler models are suitable for forecasting.

[23] As a consequence a model, despite being outperformed by a non-structural alternative, might still be useful for policy analysis focused on a particular parameter unaffected by the break.

This summary of current developments leads to some suggestions about the types of research that should be undertaken to improve the accuracy of macroeconomic forecasts. In the area of economic methodology we have two suggestions:

(1) Develop a methodology for estimating parameter shifts[24] and an ex post explanation of why such shifts occur. Despite the profession's documentation that systematic forecast errors have occurred, we do not have specific information about the particular forecast failures. For example, the errors made in forecasting the UK recession of the early 1990s are not explainable by current models (Britton and Pain, 1992). This suggests that these models have omitted variables and structural change parameters which remain to be identified.

(2) Develop ex ante approaches for identifying structural change and forecasting its effects. Linked with (1) above this would in part be a cliometric history of forecasting failure interpreted through existing and past macro models. An example of such an approach is found in Wallis (1987).

To expand on these recommendations, if intercept adjustments (or revised estimates of other parameters in the model) are to be made, how will a forecaster know when to make them, i.e. how will the structural change or regime change be recognized in advance or even contemporaneously? The comparative analysis of differences in specific forecasts between alternative models, as described above, will sometimes be informative and lead to model improvement. But the 'core assumptions' that lead most forecasters to make the same types of mistake (Ascher, 1978) have not been investigated. What is required is a thorough understanding of the intellectual or cognitive processes that forecasters use in making forecasts and refining the models which have been changed (though not apparently improved much) in response to events such as the oil shocks of the 70s and methodological innovations throughout the period.

In order to obtain this understanding of forecasting processes, it would be necessary to build and test models of forecaster behavior. There are observable heuristics being adopted, for example, the tendency to produce damped forecasts of inflation towards the mean and the lack of efficiency in inflation forecasts is reminiscent of the 'regression fallacy'. Some efforts have already been directed towards identifying such rules, e.g. Ehrbeck and Waldman (1996) and Laster et al (1996).

[24] Clements and Hendry (1998b, Ch.12) discuss various tests but we are unconvinced that they can be used to help

These models postulate that individuals might be motivated by factors other than accuracy when they prepare and issue their forecasts. The perceptions of their clients and their response to the consensus forecast (Batchelor and Dua, 1990,1991) and the need to avoid sudden adjustments (even if new information has just become available) lead to apparent economic irrationality in the forecasts. If forecasters are so motivated, further research is required to determine the arguments of their loss functions and whether the losses are asymmetric (as they typically are in financial markets). The institutional framework in which the forecasts are produced (e.g. academic, governmental, financial sector) as well as the human capital invested in the forecasting organization may also affect accuracy (Daub, 1987; Smith, 1994; Baimbridge and Whyman, 1997).

A third suggestion as to an appropriate research agenda has, therefore, a rather different focus. Performance can be improved by making more effective use of the modelers' expertise, codifying and debiasing their knowledge through:
(3) Conducting further research into the role of judgement and how best to incorporate additional information sources within the model's framework in order to overcome at least some of the problems posed by structural change. These might include socio-economic variables such as consumer confidence although its effects are clearly non-linear (Batchelor and Dua, 1998).

Throughout, the timeliness and accuracy of data has proved important in determining forecast accuracy. Our final point is to reiterate the importance of the incorporation and improvement of real time, accurate data, perhaps as disaggregate indicator variables of the current state of the economy, into the modeling process. Even now, consistent revisions have led to persistent and unnecessarily large errors, emphasizing the continuing validity of Morgestern's (1963) conclusion on the destructive impact of data errors and their substantial effects on forecasting accuracy (Öller and Barot, 2000).

In summary, this paper has argued that macroeconomic forecasts need to be improved – they serve a wide variety of purposes within both public and private sector organizations and the value of any improvements has been shown to be high. However, the current distribution of research

real-time forecasting.

effort is not appropriate[25]. A research program for improving such forecasts was suggested and it is our belief that success in any of these areas should also lead to an improved understanding of economic processes. Unfortunately, as Smith (1994) has remarked, the industry structure in the UK is such as to lead to sub-optimal investment in research aimed at evaluating and improving the quality of the macroeconomic forecasting industry. The picture in the US (Klein, 1991,Wallis, 1993) is equally gloomy.

Acknowledgement: The authors would like to thank Ken Holden and staff in the UK Treasury for help in data collection, also the many commentators on the original draft, many of whom had themselves made contributions to researching macroeconomic forecasting. Only through discussion and debate can progress be made.

[25] Interestingly, the UK Macroeconomic Modelling Bureau is not now funded to examine the forecasting record of the major UK service providers. The funding agency thought it would be of limited academic interest. The US macro modelers have found it particularly difficult to carry out convincing comparative studies (Wallis, 1993).

REFERENCES

Allen P. Geoffrey and Robert Fildes "The principles of econometric forecasting" in J. Scott Armstrong (ed.), *The Principles of Forecasting* (1999), forthcoming, Kluwer Academic Publishers, Norwell, MA.

Anderson Evan E. "An analysis of NABE multiperiod quarterly forecasts" *Business Economics* Sept (1983): 11-17.

Anderson Palle S. "Forecast errors and financial developments" Working Paper No.51 (1997) Bank For International Settlements, Basle, Switzerland.

Armstrong J. Scott and Fred Collopy "Error measures for generalizing about forecasting methods: empirical comparisons" *International Journal of Forecasting* **8** (1992): 69-80.

Armstrong J. Scott and Fred Collopy "Integration of statistical methods and judgement for time series forecasting: principles from empirical research" in Wright George and Paul Goodwin (ed.), *Forecasting with Judgement* (1998), pp.99-137. Wiley, Chichester, UK.

Armstrong J. Scott and Robert Fildes "On the selection of error measures for comparisons among forecasting methods" *Journal of Forecasting* **14** (1995): 67-71.

Artis M.J. and W. Zhang "BVAR forecasts of the G-7" *International Journal of Forecasting* **6** (1990): 349-362.

Ascher William *"Forecasting: an appraisal for policy makers"* (1978) John Hopkins University Press, New York.

Ash J.C.K., D.J. Smyth and S.M Heravi, "The accuracy of OECD forecasts of the international economy" *International Journal of Forecasting* **6** (1990): 379-392.

Ash J.C., D.J Smyth and S.M. Heravi, "Are OECD forecasts rational and useful: a directional analysis" *International Journal of Forecasting* **14** (1998): 381-391.

Ashley R. "On the usefulness of macroeconomic forecasts as inputs to forecasting models" *Journal of Forecasting* **2** (1983): 211-223.

Ashley R. "On the optimal use of non-optimal forecasts" *International Journal of Forecasting* **4** (1988): 363-376.

Baillie R.T et al "On the limitations of comparing mean square forecast errors: Commentary" *Journal of Forecasting* **12** (1993): 639-667.

Baimbridge M. and P. Whyman "Institutional macroeconomic forecasting performance of the UK" *Applied Economic Letters* **4** (1997): 373-376.

Barker Terry "Forecasting the economic recession in the U.K. 1979-1982: a comparison of the model-based Ex Ante forecasts" *Journal of Forecasting* **4** (1985): 133-151.

Batchelor Roy "All forecasters are equal" *Journal of Business and Economic Statistics* **8** (1990): 143-144.

Batchelor Roy and Pami Dua "Forecaster ideology, forecasting method and the accuracy of economic forecasts" *International Journal of Forecasting* **6** (1990): 311-316.

Batchelor Roy and Pami Dua "Blue chip rationality tests" *Journal of Money Credit and Banking* **23** (1991): 692-705.

Batchelor Roy and Pami Dua "Forecaster diversity and the benefits of combining forecasts" *Management Science* **41** (1995): 68-74.

Batchelor Roy and Pami Dua "Improving macro-economic forecasts; the role of consumer confidence" *International Journal of Forecasting* 14 (1998): 71-81.

Bates J.M. and C. W.J. Granger "The combination of forecasts" *Operational Research Quarterly* **20** (1969): 451-468.

Birchenhall C.R., H. Jessen, D.R. Osborn and P. Simpson "Predicting U.S. business-cycle regimes" *Journal of Business and Economic Statistics* **17** (1999): 313-323.

Bodkin R., L.R. Klein and K. Marwah "*A History of Macroeconomic Model Building*" (1991) Edward Elgar, Aldershot, UK.

Bolger F. and N. Harvey "Heuristics and biases in judgemental forecasting" in Wright George and Paul Goodwin (ed.), *Forecasting with Judgement* (1998), pp.99-137. Wiley, Chichester, UK.

Bonham Carl and Richard Cohen "Testing the rationality of price forecasts" *American Economic Review* **85** (1995): 284-289.

Bretschneider Stuart and Wilpen Gorr "Economic, organizational and political influences on biases in forecasting state sales tax receipts" *International Journal of Forecasting* **7** (1992): 457-466.

Britton Andrew and Nigel Pain "*Economic forecasting in Britain - Report series no.4*" (1992) National Institute of Economic and Social Research, London.

Burns Terence "The interpretation and use of economic predictions" *Proceedings of the Royal Society, Series A* **407** (1986): 103-125.

Burrell Andrew and Stephen Hall "A comparison of macroeconomic forecasts" *Economic Outlook* (1994) London Business School, London.

Cairncross Alex "Economic forecasting" *Economic Journal* **79** (1969): 797-812.

Chatfield C. "Calculating interval forecasts" *Journal of Business and Economic Statistics* **11** (1993): 121-135.

Chong Y. Y. and David F. Hendry "Econometric evaluation of linear macro-econometric models" *Review of Economic Studies* **53** (1986): 671-90.

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

Cicarelli James and Jack Narayan "The performance of eleven economic forecasting models in the 1970s" *Business Economics* (1980): 12-16.

Clemen Robert T. "Combining forecasts: a review and annotated bibliography" *International Journal of Forecasting* **5** (1989): 559-83.

Clemen Robert T. and John B. Guerard "Economic GNP forecasts: incremental information relative to naive extrapolation" *International Journal of Forecasting* **5** (1989): 417-426.

Clemen Robert T. and Robert L. Winkler "Combining economic forecasts" *Journal of Business and Economic Statistics* **4** (1986): 39-46.

Clements Michael P. "Rationality and the role of judgement in macroeconomics forecasting" *Economic Journal* **105** (1995) 410-420.

Clements Michael P. "Evaluating the rationality of fixed-event forecasts" *Journal of Forecasting* **16** (1997): 225-239.

Clements Michael P. and David F. Hendry "On the limitations of comparing mean square forecast errors" *Journal of Forecasting* **12** (1993): 617-637.

Clements Michael P. and David F. Hendry "Forecasting economic processes - a reply" *International Journal of Forecasting* **14** (1998a): 139-143.

Clements Michael P. and David F. Hendry "*Forecasting Economic Time Series: The Marshall lectures on Economic Forecasting*" (1998b) Cambridge University Press, Cambridge, UK.

Cooper J. Philip and Charles R. Nelson "The ex-ante prediction performance of the St. Louis and FRB-MIT-Penn econometric models and some results on composite predictions" *Journal of Money Credit and Banking* **7** (1975): 1-32.

Corrado Carol and Mark Greene "Reducing uncertainty in short term predictions; linkage of monthly and quarterly models" *Journal of Forecasting* **7** (1988): 77-102.

Croushore D. "The survey of professional forecasters" *Federal Reserve Bank of Philadelphia Business Review* Nov/Dec (1993): 3-15.

Daub M. "*Canadian Economic Forecasting: In a World where All's Unsure*" (1987) McGill-Queens University Press, Kingston and Montreal, Canada.

Daub Mervin "On the cost to the firm of aggregate prediction errors" *Journal of Business* **47** (1974): 11-22.

Davies Anthony and Kajal Lahiri "A new framework for analyzing survey forecasts using three-dimensional panel data" *Journal of Econometrics* 68 (1995): 205-227.

Davies Anthony and Kajal Lahiri "Reexamining the rational expectations hypothesis using panel data on multiperiod forecasts" in Hsiao, C. et al (ed.), *Analysis of Panels and Limited Dependent Variable Models* (1999), pp.. Cambridge University Press, Cambridge, UK.

Den Butter F.A.G. and M.S. Morgan "What makes the models-policy interaction successful?" *Economic Modelling* **15** (1998): 443-475.

Dhrymes Phoebus J. and Stavros C. Peristiani "A comparison of the forecasting performance of WEFA and ARIMA time series methods" *International Journal of Forecasting* **4** (1988): 81-101.

Diebold Francis X. "Forecast combination and encompassing: reconciling two divergent literatures" *International Journal of Forecasting* **6** (1989): 589-592.

Diebold Francis X. "The past, present and future of macroeconomic forecasting" *Journal of Economic Perspectives*, **12**, (1998): 175-192.

Diebold Francis X., T.A. Gunter and A.S. Tay "Evaluating density forecasts with applications to financial risk management" *International Economic Review* **39** (1998): 863-883.

Diebold Francis X. and Jose A. Lopez "Forecast evaluation and combination" in Maddala, G.S. and C. R. Rao(ed.), *Handbook of Statistics 14: Statistical Methods in Finance* (1996), pp.241-268. North-Holland, Amsterdam.

Diebold Francis X. and Robert Mariano "Comparing predictive accuracy" *Journal of Business and Economic Statistics* **13** (1995): 253-263.

Diebold Francis X., Anthony S. Tay and Kenneth F. Wallis, "Evaluating density forecasts of inflation: the survey of professional forecasters" ESRC Macroeconomic Modelling Bureau: Discussion Paper 48 (1997) University of Warwick, Coventry, UK.

Donihue Michael R. "Evaluating the role judgment plays in forecast accuracy" *Journal of Forecasting* **12** (1993): 81-92.

Egginton Don M. "Testing the efficiency and rationality of City forecasts" *International Journal of Forecasting* **15** (1999): 57-66.

Ehrbeck Tilman and Robert Waldman "Why are professional forecasters biased? Agency versus behavioral explanations" *Quarterly Journal of Economics* **111** (1996): 21-40.

Fair Ray C. "Estimating the expected predictive accuracy of econometric models" *International Economic Review* **21** (1980): 355-378.

Fair Ray C. "*Specification, estimation and analysis of macroeconometric models*" (1984) Harvard University Press, Cambridge, MA.

Fair Ray C. and Robert J. Shiller "Comparing information in forecasts from econometric models" *American Economic Review* **80** (1990): 375-89.

Fels Rendigs and C. Elton Hinshaw "*Forecasting and recognizing business cycle turns*" (1968) National Bureau Of Economic Research, New York.

Fildes Robert "The evaluation of extrapolative forecasting methods with discussion" *International Journal of Forecasting* (1992): 81-111.

Fildes Robert (ed.) "*World Index of Economic Forecasts (Fourth Edition)*" (1995) Gower, Aldershot, Hampshire, UK.

Fildes Robert, A. Nasralla and N. Tessaromatis, "Interest rate forecasting and corporate profitability" *Business Economist* **17** (1) (1985): 17-32.

Garcia-Ferrer Antonio, R.A. Highfield and Palm F. and A. Zellner, "Macroeconomic forecasting using pooled international data" *Journal of Business and Economic Statistics* **5** (1987): 53-67.

Goldfarb Robert and H.O. Stekler "*Testing for rationality: a meta analysis*" mimeo (1998).

Graham John R. "Is a group of economists better than one? Than none?" *Journal of Business* **69** (1996): 193-232.

Granger C.W.J. "Can we improve the perceived quality of economic forecasts" *Journal of Applied Econometrics* **11** (1995): 455-473.

Granger C.W.J. and Paul Newbold "*Forecasting economic time series*" (1977) Academic Press, San Francisco.

Granger C.W.J., Paul Ormerod and Ron Smith, "Comments on "Forecasting economic processes"" *International Journal of Forecasting* **14** (1998): 133-137.

Grunberg E. and F. Modigliani "The predictability of social events" *Journal Of Political Economy* **62** (1954): 465-478.

Hafer R.W. and D.H Resler Jn. "On the rationality of inflation-forecasts - a new look at the Livingston data" *Southern Economic Journal* **48** (1982):1049-1056.

Harvey David, Stephen Leybourne and Paul Newbold, "Testing the equality of prediction mean squared errors" *International Journal of Forecasting* **13** (1997): 281-291.

Harvey David, Stephen Leybourne and Paul Newbold, "Tests for forecast encompassing" *Journal of Business and Economic Statistics* **16** (1998): 254-259.

Hatanaka M. "The underestimation of variations in the forecast series" *International Economic Review* **16** (1975): 151-160.

Hendry David F. "The econometrics of macroeconomic forecasting" *Economic Journal* **107** (1997): 1330-1357.

Hendry David F. and J.A. Doornik "The implications for econometric modelling of forecast failure" *Scottish Journal of Political Economy* **44** (1997): 437-461.

Holden Ken "A comparison of forecasts from UK economic models and some Bayesian vector autoregressive models" *Journal of Economic Studies* **24** (1997): 243-257.

Holden Ken "Vector autoregression modelling and forecasting" *Journal of Forecasting* **14** (1995): 410-420.

Holden Ken and Albert Broomhead "An examination of vector autoregressive forecasts of the UK economy" *International Journal of Forecasting* **6** (1990): 11-23.

Holden Ken and D.A. Peel "An evaluation of quarterly National Institute forecasts" *Journal of Forecasting* **4** (1985): 227-234.

Holden Ken and D.A. Peel "An empirical investigation of combinations of economic forecasts" *Journal of Forecasting* **5** (1986): 229-242.

Holden Ken and D.A. Peel "On testing for unbiasedness and efficiency of forecasts" *Manchester School* **58** (1990): 120-127.

Holden Ken and John Thompson "Combining forecasts, encompassing and the properties of UK macroeconomic forecasts" *Applied Economics* **29** (1997): 1447-1458.

Howrey E.P. "The use of preliminary data in econometric forecasting" *Review of Economics and Statistics* **60** (1978): 193-200.

Jenkins Gwilym M. "Some practical aspects of forecasting in organizations" *Journal of Forecasting* **1** (1982): 3-21.

Jeong Jjinook and G.S. Maddala "Measurement errors and tests for rationality" *Journal of Business and Economic Statistics* **9** (1991): 431-439.

Joutz Frederick L. "Informational efficiency tests of quarterly macroeconometric forecasts from 1976 to 1985" *Managerial and Decision Economics* **9** (1988): 311-320.

Keane Michael P. and David F. Runkle "Testing the rationality of price forecasts: new evidence from panel data" *American Economic Review* **80** (1990): 714-735.

Keen Howard "Economists and their forecasts: have the projections been that bad?" *Business Economics* (1987): 37-40.

Kirchgassner G. "Testing weak rationality of forecasts with different time horizons" *Journal of Forecasting* **12** (1993): 541-558.

Klein Lawrence R. (ed.) "*Comparative Performance of U.S. Econometric Models*" (1991) Oxford University Press, Oxford.

Kolb R.A and H.O. Stekler "The lead and accuracy of macroeconomic forecasts" *Journal of Macroeconomics* **12** (1990): 111-123.

Lahiri Kajal and Christie Teigland "On the normality of probability distributions of inflation and GNP forecasts" *International Journal of Forecasting* **3** (1987): 269-279.

Laster David, Paul Bennett and In Sun Geoum, "*Rational bias in macroeconomic forecasts*" Research Paper **9617** (1996) Federal Reserve Bank Of New York, New York.

Leitch G. and J.E. Tanner "Professional economic forecasts: are they worth their costs?" *Journal of Forecasting* **14** (1995): 143-157.

Leontief W. "Theoretical assumptions and nonobserved facts" *American Economic Review* **61** (1971): 1-7.

Litterman Robert B. "Forecasting with Bayesian vector autoregressions - five years of experience" *Journal of Business and Economic Statistics* **4** (1986): 25-38.

Longbottom J.A. and S. Holly "The role of time series analysis in the evaluation of econometric models" *Journal of Forecasting* **4** (1985): 75-87.

Lupololetti V.M. and R.H. Webb "Defining and improving the accuracy of macroeconomic forecasts: contributions from a VAR model" *Journal of Business* **59** (1986): 263-285.

Mayer T. "Economics as hard science: realistic goal or wishful thinking" *Economic Enquiry* **18** (1980): 165-178.

McCloskey D. "The rhetoric of economics" *Journal of Economic Literature* **21** (1983): 481-517.

McCloskey D. "The art of forecasting from ancient to modern times" *Cato Journal* **12**: Spring/S (1992): 23-48.

McNees Stephen K. "An evaluation of economic forecasts: extension and update" *New England Economic Review* (1976a): 30-44.

McNees Stephen K. "The forecasting performance in the 1970s" *New England Economic Review* July-August (1976b): 1-13.

McNees Stephen K. "The rationality of economic forecasts" *American Economic Review and Proceedings* **68** (1978): 301-317.

McNees Stephen K. "The forecasting record for the 1970s" *New England Economic Review* Sept-Oct (1979): 1-21.

McNees Stephen K. "The role of macroeconomic forecasting and policy analysis in the US" *Journal of Forecasting* **1** (1982): 37-48.

McNees Stephen K. "Which forecast should you use?" *New England Economic Review* (1985): 36-42.

McNees Stephen K. "Forecasting accuracy of alternative techniques: a comparison of US macroeconomic forecasts" *Journal of Business and Economic Statistics* **4** (1986): 5, 15 & 23.

McNees Stephen K. "Consensus forecasts: tyranny of the majority" *New England Economic Review* (1987): 15-21.

McNees Stephen K. "How accurate are macroeconomic forecasts?" *New England Economic Review* (1988): 15-26.

McNees Stephen K. "The role of judgment in macroeconomic forecasting accuracy" *International Journal of Forecasting* **6** (1990): 287-299.

McNees Stephen K. "How large are economic forecast errors?" *New England Economic Review* (1992a): 25-33.

McNees Stephen K. "The 1990-91 recession in historical perspective" *New England Economic Review* Jan-Feb (1992b): 3-22.

McNees Stephen K. "The uses and abuses of 'consensus' forecasts" *Journal of Forecasting* **11** (1992c): 703-710.

McNees Stephen K. and John Ries "The track record of macroeconomic forecasts" *New England Economic Review* (1983): 5-18.

Meese R. and K. Rogoff "Exchange rate models of the seventies: do they fit out of sample?" *Journal of International Economics* **14** (1983): 3-24.

Melliss Chris "The Treasury forecasting record: an evaluation" ESRC Macroeconomic Modelling Bureau (1997) University of Warwick, Coventry, UK.

Melliss Chris and Rod Whittaker "The Treasury forecasting record: some new results" *National Institute Economic Review* **164** (1998): 65-79.

Mills Terence C. and Gordon Pepper "Assessing the forecasters" *International Journal of Forecasting* **15** (1999): 247-257.

Mincer Jacob and Victor Zarnowitz "The evaluation of economic forecasts" in Mincer, Jacob (ed.), *Economic Forecasts And Expectations* (1969), National Bureau of Economic Research, New York.

Moore Geoffrey "Forecasting short-term economic change" *Journal of the American Statistical Association* **64** (1969): 1-22.

Morgenstern Oskar "*On the accuracy of economic observations* (2nd ed.)" (1963) Princeton University Press, Princeton, NJ.

Moss Scott, Michael Artis and P. Ormerod, "A smart automated macroeconometric forecasting system" *Journal of Forecasting* **13** (1994): 299-312.

Nelson Charles R. "The prediction performance of the FRB-MIT Penn model of the US economy" *American Economic Review* **62** (1972): 902-917.

Nelson Charles R. "A benchmark for the accuracy of econometric forecasts of GNP" *Business Economics* (1984): 52-58.

Nordhaus, W.D. "Forecasting efficiency: concepts and applications" *Review of Economics and Statistics* **69** (1987): 667-674

Oller Lars-Erik and Bharat Barot "Comparing the accuracy of European GDP forecasts" National Institute of Economic Research (1999), Stockholm, Sweden.

Pain Nigel "Cointegration and forecast evaluation: some lessons from National Institute forecasts" *Journal of Forecasting* **13** (1994): 481-494.

Price D.H.R. and J.A. Sharp "A comparison of the performance of different univariate forecasting methods in a model of capacity acquisition in UK electricity supply" *International Journal of Forecasting* **2** (1986): 333-348.

Rich Robert W. "Testing the rationality of inflation forecasts from survey data: another look at the SRC expected price change data" *Review of Economics and Statistics* (1989): 682-686.

Samuelson Paul A. "Optimality of sluggish predictors under ergodic probabilities" *International Economic Review* **17** (1976): 1-17.

Schnader M.H. and H.O. Stekler "Evaluating predictions of change" *Journal of Business* **63** (1990): 99-107.

Schnader M.H. and H.O. Stekler "Sources of turning point forecast errors" *Applied Economics Letters* (1998): Forthcoming.

Scotese C.A. "Forecast smoothing and the optimal under-utilization of information at the Federal-Reserve" *Journal of Macroeconomics* **16** (1994): 653-670.

Smith R "The Warwick ESRC macroeconomic modelling bureau: an assessment" *International Journal of Forecasting* **6** (1990): 301-309.

Smith R "The macromodelling industry: structure, conduct and performance" in Hall, Stephen (ed.), *Applied Economic Forecasting* (1994), pp.68-88. Harvester Wheatsheaf, London.

Smyth David and J.C.K. Ash "The underestimation of forecasts and the variability of predictions and outcomes" *Bulletin of Economic Research* **33** (1981): 37-44.

Smyth David and J.C.K. Ash "Multiperiod forecasts and the variability of predicted and actual changes" *Economics Letters* **19** (1985): 141-143.

Spencer R.W. and J.H. Huston "Rational forecasts - on confirming ambiguity as the mother of conformity" *Journal of Economic Psychology* **14** (1993): 697-709.

Stekler H. O. "An analysis of turning point errors" *American Economic Review* **62** (1972): 724-729.

Stekler H. O. "Why do forecasters underestimate?" *Economic Enquiry* **13** (1975): 445-449.

Stekler H. O. "Who forecasts better?" *Journal of Business and Economic Statistics* **5** (1987): 155-158.

Stekler H. O. "Are economic forecasts valuable?" *Journal of Forecasting* **13** (1994): 493-505.

Swanson Norman R. and Halbert White "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models" *International Journal of Forecasting* **13** (1997): 439-461.

Tay A.S. and K.F. Wallis "Density forecasting: A survey" mimeo (1999).

Theil Henri "Applied economic forecasting" *Economic Forecasting* (1966) North Holland Publishing Company, Amsterdam.

Todd R.M "Algorithms for explaining forecast revisions" *Journal of Forecasting* **11** (1992): 675-685.

Treasury and Civil Service Committee *"Memoranda on official economic forecasting, 532-i*" **11** (1992): 675-685.

Turner D.S. "The role of judgement in macroeconomic forecasting" *Journal of Forecasting* **9** (1990): 315-345.

Ulan Michael, William G. Dewald and James B. Bullard, *"U.S. Official forecasts of group of seven economic performance, 1976-90*" mimeo (1994).

Wallis Kenneth F. (ed.) "*Models of the UK economy*" (1984) Oxford University Press, Oxford.

Wallis Kenneth F. (ed.) "*Models of the UK economy*" (1985) Oxford University Press, Oxford.

Wallis Kenneth F. (ed.) "*Models of the UK economy*" (1986) Oxford University Press, Oxford.

Wallis Kenneth F. (ed.) "*Models of the UK economy*" (1987) Oxford University Press, Oxford.

Wallis Kenneth F. "Macroeconomic forecasting: a survey" *Economic Journal* **99** (1989): 28-61.

Wallis Kenneth F. "Comparing macroeconometric models: A review article" *Economica* **60** (1993): 225-237.

Wallis Kenneth F. and John D. Whitley "Sources of error in forecasts and expectations: U.K. economic models, 1984-1988" *Journal of Forecasting* **10** (1991): 231-253.

Webb Roy H. "Forecasts of inflation from VAR models" *Journal of Forecasting* **14** (1995): 267-285.

Young Richard M. "Forecasting with an econometric model: the issue of judgmental adjustment" *Journal of Forecasting* **1** (1982): 189-204.

Zarnowitz Victor "*An appraisal of short-term economic forecasts*" (1967) National Bureau of Economic Research, New York.

Zarnowitz Victor "An analysis of annual and multiperiod quarterly forecasts of aggregate income, output, and the price level" *Journal of Business* **52** (1979): Jan-33.

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

Zarnowitz Victor "Rational expectations and macroeconomic forecasts" *Journal of Business and Economic Statistics* **3** (1985): 293-311.

Zarnowitz Victor "*Business Cycles: Theory, History, Indicators and Forecasting*" (1992) National Bureau of Economic Research Studies in Business Cycles Vol. 27, University of Chicago Press, Chicago.

Zarnowitz Victor and Philip Braun "Twenty two years of the NBER-ASA Quarterly Outlook Surveys: aspects and comparisons of forecasting performance" *NBER Working Paper* **3965** (1992).

Zellner Arnold "Bayesian estimation and prediction using asymmetric loss functions" *Journal of the American Statistical Association* **81** (1986a): 446-451.

Zellner Arnold "Biased predictors, rationality and the evaluation of forecasts" *Economics Letters* **21** (1986b): 45-48.

Zellner Arnold and C. Hong "Forecasting international growth rates using Bayesian Shrinkage and other procedures" *Journal of Econometrics* **40** (1989): 183-202.