Research Program on Forecasting

EVALUATING A VECTOR OF THE FED'S FORECASTS

Tara M. Sinclair
The George Washington University
Washington, DC 20052
USA
H.O. Stekler
The George Washington University
Washington, DC 20052
USA
Warren Carnow
The George Washington University
Washington, DC 20052
USA

RPF Working Paper No. 2012-002
http://www.gwu.edu/~forcpgm/2012-002.pdf

Original Version
January 30, 2012

Updated
April 26, 2013

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

EVALUATING A VECTOR OF THE FED'S FORECASTS

Tara M. Sinclair*, H.O. Stekler, and Warren Carnow
Department of Economics
The George Washington University
Monroe Hall #340
2115 G Street NW
Washington, DC 20052

JEL Codes: C5, E2, E3

Keywords: Federal Reserve, Forecast Evaluation, real GDP Expenditure Approach, Business Cycle, Mahalanobis Distance

Abstract

In this paper we present a multivariate analysis of the Federal Reserve's forecasts. First, we use existing univariate methods to evaluate each of the Fed's forecasts of the ten major expenditure categories of real GDP which have not previously been evaluated in the literature. Second, we apply a recently developed methodology to jointly evaluate the vector of these forecasts. Finally, we use the same methodology to determine whether the Fed's forecasts of GDP growth, inflation, and unemployment taken together present an accurate overall view of the economic situation and compare the Fed's forecasts to those of the Survey of Professional Forecasters. We find that the Fed's forecasts were generally consistent with the overall conditions that actually occurred. We also find that the Fed's forecasts and those of the Survey of Professional Forecasters are quite similar overall.

*Corresponding author: tsinc@gwu.edu. The authors thank Blake Saville for excellent research assistance and gratefully acknowledge generous support from the Institute for International Economic Policy.

EVALUATING A VECTOR OF THE FED'S FORECASTS

This paper evaluates two key sets of forecasts, the ten components of US real GDP and three major macroeconomic indicators, prepared by the staff of the Board of Governors of the Federal Reserve System. Most evaluations have used a univariate methodology that separately examined the Fed's forecasts of select variables such as GDP, inflation, and unemployment. (Clements et al., 2007; Joutz and Stekler, 2000; Romer and Romer, 2000; Sims, 2002; and Stekler, 1994.)

There have, however, been a small number studies that have considered some multivariate characteristics of the Fed's forecasts. Sinclair, Stekler and Kitzinger (SSK, 2010) examined the joint *directional* forecasts of GDP and inflation using contingency tables. Sinclair, Gamber, Stekler, and Reid (SGSR, 2012) calculated the costs of jointly misestimating GDP and inflation within the context of a Taylor type rule. These studies, however, did not develop a general approach for jointly evaluating quantitative forecasts.

Focusing on the rationalizability of the forecasts, Caunedo et al (2013) jointly test the rationality of the Federal Reserve's forecasts of inflation, unemployment, and output growth. They use the methodology of Komunjer and Owyang (2012) where forecast errors in a multivariate framework are used to derive the weights of a utility function. Their approach differs from the one that we present below. Our method focuses on forecast comparison rather than on rationalizability.

This paper will examine two topics that have not been addressed before. The first, an evaluation of the Fed's forecasts of the components of real GDP, has not been performed before using either the univariate or multivariate approaches. Past evaluations of the Fed's real GDP growth rate forecasts have only focused on the headline GDP projections, and, to the best of our knowledge, the Fed's forecast of the ten main components of GDP have never been evaluated. The second topic involves a multivariate analysis of the Fed's forecasts of growth, inflation, and unemployment. We undertake this multivariate analysis because these forecasts are produced and used *jointly*. This approach enables us to determine whether the Fed's forecasts provided an accurate comprehensive view of the various sectors of the economy. This view is especially important if the forecast is used in making policy and suggests that the forecasts should be evaluated jointly in a multivariate framework. We next present the rationale for this approach.

Consider a large database of forecasts prepared by a number of individuals/organizations. The database would likely consist of forecasts made for a number of variables over a number of horizons over a period of time. How should one evaluate these forecasts? There is not a simple answer because there are a number of ways of doing this analysis. They range from the simple univariate single horizon method to the more complex methods which aggregate across the various dimensions of the data.

The database of forecasts in general will have four dimensions: (1) the number of variables (J) that are predicted, (2) the number of horizons/periods (H) for which each variable is predicted, (3) the number of times (T) that the predictions are made, and (4) the number of forecasters (N). The traditional procedure for evaluating forecasts involves calculating a scalar descriptive statistic such as mean-squared error (MSE) which describes the *average* accuracy of the T forecasts of *each* variable that were made for each forecast horizon. This approach yields NHJ descriptive statistics, one for each forecaster, at each horizon, for each variable.

Recent research has proposed several different procedures that have been used to aggregate across the various dimensions and reduce the number of descriptive statistics. The appropriate procedure for aggregating depends on the question that is being investigated. For example, Eisenbeis, Waggoner and Zha (2003) aggregated across variables for a single time period and a single horizon for each forecaster. Their procedure created a ranking of the average quality of the Wall Street Journal Forecasters across multiple variables for each forecaster, for each period, for a single horizon.[1]

On the other hand, Clements, Joutz, and Stekler (2007) and Davies and Lahiri (1995, 1999) do not pool across variables.[2] For each variable, they pool across horizons. One difference between those two studies is that Clements et al. evaluate only one forecaster (the Fed) whereas Davies and Lahiri consider the forecasts of multiple forecasters (from the Blue Chip surveys and the Survey of Professional Forecasters).

In this paper we are interested in the *overall* quality of one organization's forecasts of a number of different variables over time for a particular horizon. To illustrate this issue, we start with the simplest case: an evaluation of the organization's forecasts of next period's value of one variable, say the growth rate of real GDP. These forecasts have been made T times. The traditional univariate procedure involves calculating MSE which describes the *average* accuracy of the T forecasts of real GDP growth. Now let us assume that the organization also prepares forecasts of inflation and unemployment. Traditionally, we would have calculated MSEs for each of these additional variables. If the MSE of one variable were "small", while that of one or both of the others were "large", how would we evaluate the overall quality of this forecast? What do we learn by saying that the errors made in forecasting one variable were large while those made in forecasting the other variable were small?

[1] Eisenbeis, Waggoner, and Zha also produced an average ranking of the forecasters over time.
[2] Davies, Lahiri, and Sheng (2011) provide a useful summary of the framework used in these papers.

In order to determine whether the individual produced a "good" *overall* forecast, we would need to obtain an error measure from a multivariate evaluation that aggregated across the variables. For each forecast horizon, this aggregation is accomplished by (1) creating a vector of forecasts, (2) creating a vector of outcomes, and (3) measuring the distance between the two vectors. This methodology reduces the number of error measures to one for each forecast horizon. We can then test the statistical significance of this distance.

Our approach is based on the methodology that Sinclair and Stekler (2011) utilized to analyze early GDP component estimates from the Bureau of Economic Analysis. That methodology determined whether, for each quarter, the vector of the first vintage of BEA estimates of all the major GDP components was similar to a vector of a later vintage of BEA estimates of the same components. To determine whether the two sets of estimates are related, it was necessary to compare the difference between the two vectors. Sinclair and Stekler utilized the Mahalanobis measure for estimating the relationship of two vectors. This measure, which is well established in the natural sciences, is a generalization of the Euclidean distance and allows for the interdependence of the vectors.[3] In order to test whether there was a difference between the two vintages of estimates, they focused on the difference between the mean vectors relative to the common within-group variation. [4]

In this paper we will utilize the same methodology to analyze the Fed's forecasts. One vector will consist of the forecasts of all the variables that the Fed made at one time that refer to a particular point in time. The other vector will be comprised of the actual outcomes for those variables. In addition, we will also apply the new vector generalization of the Holden and Peel

[3] See Abdi (2007) for a discussion of different distance measures.
[4] Sinclair, Stekler, and Carnow (2012) applied this methodology to the median forecasts of the Survey of Professional forecasts for GDP growth, unemployment, and inflation.

(1990) test for unbiasedness that Sinclair and Stekler developed. This will enable us to determine whether taking into account the revisions to other variables might have improved the forecasts.

This paper makes several contributions to the forecast evaluation literature. First, we evaluate the Fed's forecasts of the ten main components of GDP. Previous analyses of the Fed's forecasts have only focused on the headline GDP projections. Second, we use a new methodology to jointly evaluate the vector of these forecasts. Finally, we use the same methodology to determine whether the Fed's forecasts of GDP growth, inflation, and unemployment taken together present an accurate overall view of the economic situation and compare the Fed's forecasts to those of the Survey of Professional Forecasters.

The rest of the paper proceeds in this way: We first describe the data and the new methodology and then evaluate the Fed's forecasts of ten major expenditure categories of GDP. As a second example, we also conduct a multivariate analysis of the forecasts of three variables, the GDP growth rate, the rate of inflation, and the unemployment rate, that together describe the overall condition of the economy. Finally, using the same methodology, we compare the Fed's forecasts of those three variables with the median predictions of the same variables obtained from the Survey of Professional Forecasters.

I. Data

The projections used in this analysis are the Federal Reserve's Greenbook forecasts for (1) the growth rates of the ten major components of GDP that the Fed staff predicted from 1986.3 through 2004.4,[5] and (2) the growth rates of real output, the inflation rate, and the unemployment

---

[5] The Greenbook data are only available with a 5-year lag. We obtained our dataset from the PDF files on the Federal Reserve Bank of Philadelphia Website: http://www.philadelphiafed.org/econ/forecast/greenbook-data/index.cfm. The beginning dates are chosen based on when the Fed first began forecasting these variables. The ten components are: Durable consumption, non-durable consumption, services consumption, non-residential investment, residential investment, private inventories, exports, imports, federal government spending, and state and local government spending.

rate for the period 1965.4 – 2005.4.[6] In each quarter the Fed staff makes multiple predictions for many quarters into the future. We use the last set of forecasts made in each quarter and analyze the projections that are made for the current quarter and one quarter ahead.

We focus on short horizons because the Fed staff has, at times, based its Greenbook forecasts on an assumed (possibly varying) path for monetary policy. At other times, however, the Fed has assumed that monetary policy would remain unchanged over its forecast horizon (see Reifschneider and Tulip, 2007, for further discussion). Since the assumed path for monetary policy associated with each Greenbook forecast is not known, a possible complication arises when analyzing longer-term forecasts. The current quarter and one-quarter ahead forecasts are too short of a time horizon, however, to be affected by the Fed's future path for monetary policy. Therefore, regardless of whether the Fed assumes a constant path or a time-varying path for monetary policy, the current and one-quarter-ahead forecasts will be unaffected by those assumptions.

The actual values were the real-time data published approximately 90 days after the end of the quarter to which they refer. The use of the real-time data avoids definitional and classification changes.

II. Methodology

A. Single Variable Analysis

We first analyze the forecast errors of each variable separately and focus on two topics: directional accuracy and systematic error.

1. Directional Accuracy

[6] For output growth we use GNP from 1965 to 1991 and GDP from 1992 on. The last forecast in the fourth quarter of 1991 was the first forecast of GDP. The inflation rate is based on the implicit price deflator through the first quarter of 1996, then the chain-weighted price index from 1996.2 on.

A desirable characteristic of any forecast is that it should provide a correct picture of the direction in which the economy is moving. Thus the signs of the predicted changes of each of the ten components will be compared with the sign of the actual changes reported 90 days after the end of the quarter to which they refer.

2. Systematic Error (Bias)

Even if there are not a substantial number of differences in signs between the forecast and actual changes, there may still be a systematic error- a bias. We use two approaches to determine whether the forecasts are systematically related to the actual data. First, we test this relationship using the Mincer-Zarnowitz (1969) regression. We then question whether there are systematic errors related to the state of the economy. These two tests are applied *individually* to each of the ten components of GDP.[7] Customarily, the Mincer-Zarnowitz (1969) regression has been used to test for bias in the forecasts of a single variable:

$$A_t = \beta_0 + \beta_1 F_t + e_t, \qquad (1)$$

where A_t and F_t are the actual real-time data and the Fed's forecasts, respectively. For a test of informational efficiency, the null hypothesis is: $\beta_0 = 0 \; and \; \beta_1 = 1$ .A rejection of this hypothesis indicates that the forecasts are biased and/or inefficient. The Wald test and the F distribution are used to test this null.[8]

Recent research has shown that forecasts sometimes contain systematic errors (Joutz and Stekler, 2000, Hanson and Whitehorn, 2006). Forecasters overestimated the rate of growth during slowdowns and recessions and underestimated it during recoveries and booms. Similarly, inflation was under-predicted when it was rising and over-predicted when it was declining. In

[7] These tests for the growth rate and inflation forecasts were already done in Sinclair, Joutz, and Stekler (2010). We therefore do not replicate them here.

[8] An alternative procedure for testing for bias has been to use equation suggested by Holden and Peel (1990): $A_t - F_t = \beta_0 + e_t$. In this case, the slope is imposed to be one and the test examines whether or not the forecast error has a zero mean, i.e. a simple test of statistical significance of the constant in this equation.

some cases, these systematic errors, associated with the stages of the business cycle, may offset each other. Consequently, the use of (1) in the presence of these offsetting errors may yield regression estimates that do not reject the null of bias when in fact there are systematic errors that are associated with the state of the economy.

In order to determine whether the Fed's forecasts similarly failed to incorporate information about the state of the economy, we modified (1) as in Sinclair, Joutz, and Stekler (2010). The modified Mincer-Zarnowitz regression (2) now becomes:

$$A_t = \beta_0 + \beta_1 F_t + \beta_2 D_t + e_t, \tag{2}$$

where D_t is a dummy that reflects the state of the economy. It takes on the value 1 if during one month of a particular quarter the economy was in a recession. Otherwise, the value of the dummy is zero. The quarters that constituted the recession were those defined by the NBER.[9] The joint null hypothesis now is: $\beta_0 = 0$, $\beta_1 = 1$, *and* $\beta_2 = 0$. If any of the coefficients associated with the dummies are non-zero, the dummies contain information that can explain the forecast errors. If this were the case, it would indicate that the Fed did not fully incorporate information about the state of the economy into the forecasts.

B. Multivariate Analysis

1. Bias

We have described the procedures that are used to test for the existence of systematic errors in the forecasts of each variable. We next investigate the properties of the errors of the forecasts of the same ten variables, but use a new joint framework. To do this we construct a first-order vector autoregression (VAR(1)) of the errors made in forecasting each of the ten components. If the forecasts are unbiased estimates of the outcomes, then none of the

[9] The NBER dates are available at: http://www.nber.org/cycles/cyclesmain.html. Even though the NBER data are not known in real time, there is ample justification for using them, because this is an ex post analysis to determine whether data during recessions were fully incorporated into the forecasts.

coefficients in the VAR should be significant. In other words, the constant estimates should be zero; the coefficients on the own lags should be zero; and none of the errors made in forecasting the other variables should Granger-cause any of the other errors. We, therefore, construct an eleven equation VAR (1) consisting of the forecast errors of GDP and its ten major components (where $\boldsymbol{FE_t}$ is a vector of the forecast errors for time *t*).

$$\boldsymbol{FE_t} = \boldsymbol{\beta_0} + \boldsymbol{FE_{t-1}}\boldsymbol{\beta_1} + \boldsymbol{e_t}. \tag{3}$$

In this case, $\boldsymbol{\beta_0}$ is then a vector of the constant terms and $\boldsymbol{\beta_1}$ is a matrix of coefficients on the lags of the forecast errors. Under the null hypothesis, all of the elements of both $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are zero. In section IV below we also undertake a multivariate analysis of the forecasts of growth, inflation and unemployment to determine whether these predictions accurately capture the overall view of the economy.

2. Accuracy

As mentioned above, we use a distance measure to determine the accuracy or difference of the vectors. There are two common measures of distance, Euclidean and Mahalanobis, that differ in the assumptions made about the statistical independence of the vectors Assume that we have two independent vectors, $\boldsymbol{F_t}$ and $\boldsymbol{A_t}$, representing the forecasts and outcomes consisting of *n* variables in each vector. The difference between the two vectors can be measured by the Euclidean distance between them:

$$d(\boldsymbol{F_t}, \boldsymbol{A_t}) = \sqrt{(\boldsymbol{F_t} - \boldsymbol{A_t})'(\boldsymbol{F_t} - \boldsymbol{A_t})} \tag{4}$$

This procedure is only applicable to vectors that are independent and that are scaled so that they have unit variances. These assumptions do not apply in this analysis. Thus, we will use a generalization of the Euclidian distance that allows for the scale to differ across the different variables and for nonzero correlation between the variables. In order to test if there is a difference between the forecasts and the outcomes, we will focus on the difference between the mean vectors of each set of data relative to the common within-group variation. This measure is called the Mahalanobis Distance, D^2:[10]

$$D^2 = (\bar{\boldsymbol{F}} - \bar{\boldsymbol{A}})' \boldsymbol{W} (\bar{\boldsymbol{F}} - \bar{\boldsymbol{A}}), \tag{5}$$

where $\boldsymbol{W}$ is the inverse of the pooled sample variance-covariance matrix, and $\bar{\boldsymbol{F}}$ and $\bar{\boldsymbol{A}}$ are the mean vectors of the forecasts and outcomes, respectively.[11] We can then construct an F-statistic based on this measure to test the null hypothesis that the forecasts and outcomes have the same population means.[12]

In addition, we will split the sample into periods when the economy was expanding and when the economy was in recession. From this we can see if the difference between the forecasts of the ten GDP components and their outcomes is significant during expansions, during recessions, or in both cyclical phases.

III. Results

A. Directional Accuracy

Among the 10 major components, the sign of the change in consumption services was always positive in both the forecasts and the actual data. The accuracy of the current quarter forecasts of the direction of change of the remaining nine GDP components ranged from 79-89%;

---

[10]Mahalanobis distance is also associated with discriminant analysis. For other economic forecast applications of this measure, see Banternghansa and McCracken (2009) and Jordá et al (2010).

[11] We estimate the sample covariance matrix as the weighted average of the two (bias-corrected) sample covariance matrices from the two sets of data. It is assumed that the two sets of data have a common covariance matrix in the population.

[12] $F = \frac{(n-1-p)n_1 n_2}{p(n-2)(n_1+n_2)} D^2$, with p and n-p-1 degrees of freedom (McLachlan, 1999).

it was 68-83% for the t+1 predictions.[13](See Table 1). The quarter-ahead forecasts displayed a clustering of incorrect signs during the recession of 2001; otherwise there was no obvious clustering.

B. Bias

Tables 2A and 2B present the results from the tests that we used to determine whether the forecasts of the ten components were biased. We show the p-values obtained from the two Mincer-Zarnowitz equations and from the joint test using the 11-equation VAR.[14] We found that at least one of the tests found evidence of bias in six of the component current-quarter forecasts and in nine of the one quarter-ahead projections. Thus in the 1986-2004 period, not only is the headline GDP forecast biased but so are many of the estimated components.

C. Accuracy

Despite the evidence of these biases in the forecasts, we needed to determine whether the forecasts of the ten components, taken together, provided an overall view of the growth of the economy that was consistent with the condition that actually occurred. For this analysis, we used the Mahalanobis Distance measure to jointly evaluate the component forecasts. The null was that the Fed forecasts failed to provide an overall view of the growth of the various sectors of the economy that was consistent with the observed data. (Tables 3A-3C). We did not reject the null for the current quarter forecasts in either the entire sample or in the recession/expansion subsamples. However, the null was rejected for the quarter-ahead forecasts for both the entire sample and for expansionary periods. The p values of the F statistics were 0.05 and 0.02, respectively. These results indicate that the Fed had a good understanding of the composition of

[13] There were thirty quarters in which the signs of the forecasts of at least one component differed from the signs of the actual changes. In seven of those quarters the signs of three or more components disagreed.

[14] In many cases, the null of unbiasedness was rejected when the state of the economy dummy was included in the Mincer-Zarnowitz equation. Newey-West corrected test statistics were used in all cases for the one quarter-ahead regressions.

the GDP changes that were occurring in the current quarter but not how these sectors were likely to change next period.[15]

D. Comparison with BEA Results

It is possible to make a benchmark comparison of these component forecasts. The Bureau of Economic Analysis publishes estimates of all of these components 30 days after the quarter to which they refer. Since the Fed's current quarter forecasts were made during the quarter, they were available at least 30 days before the BEA's first estimates. Sinclair and Stekler (2011) examined the 30 day vintage of BEA's estimates of the GDP components. They found that at least one test rejected the null of no bias in the BEA estimates for every single variable, be it headline GDP or one of the components. While none of the tests rejected the null for the Fed's forecast of four components, the earlier results are basically comparable to the ones that we now present.[16] Thus the Fed's forecasts made in the current quarter seem to be as good as the BEA's estimates released at least 30 days later.

IV. Overall View of the Economy

A good forecast should provide an accurate picture or *overall view* of the state of the economy at a particular point in time. We have, so far, focused on the GDP and component forecasts, but the Fed also projects the inflation rate and the unemployment rate at the same time that it predicts headline GDP. A combination of these three individual forecasts can be viewed as a vector representing an *overall view* of the future condition of the economy. The actual outcomes of the three variables comprise a different vector. Thus, if we are concerned with how well the forecasts reflect the actual changes that have occurred in the economy, we must compare

[15] The null was not rejected in the recessionary quarters. This may be due either to a better understanding of what can occur in recessions or to the fact that the number of observations was too small to be able to reject the hypothesis.
[16] The components were consumption services, fixed non-residential investment, and the two government sectors. (See Table 2A.) It should be noted that the time periods for the two studies were different.

the difference in the two vectors, representing the forecasts and the actual outcomes of the three variables. Our methodology permits us to analyze this issue, and we again used the Mahalanobis Distance measure, this time to jointly evaluate the growth, unemployment, and inflation forecasts.

The null was that the Fed forecasts provide an overall view of the economy that was consistent with the observed data. We did not reject this null for either the current or the quarter-ahead forecasts. We obtained the same results when we divided the sample into expansionary and recessionary periods. The p-values of the F statistic associated with the Mahalanobis Distance measure were always greater than 0.40. (Tables 4A-4C)

From these results we can conclude that the Greenbook forecasts are consistent with the *overall* conditions that actually occurred. However, given the results obtained from the ten component projections, the quarter-ahead estimates made for *particular* sectors may not always reflect actuality.

V. Benchmark Comparison

Our analysis to this point has been exclusively on the Fed's forecasts. This raises an obvious question: In terms of an overall view of the economy, how do these forecasts compare with other predictions? As the benchmark for this comparison, we used the median forecasts of growth, inflation and unemployment obtained from the Survey of Professional Forecasters (SPF). Beginning in 1968.4, these forecasts span a period nearly as long as the entire sample of the Fed's predictions.

The comparison is based on data for forecasts made from 1968.4 through 2005.4. As before we construct two vectors from the predictions of these three variables and use the Mahalanobis Distance measure to determine whether there was a significant difference in the average overall views of these two sets of forecasts. The results are presented in Table 5 and show that there is little difference between the two sets of forecasts.

V. Conclusions

In this paper we evaluated two key sets of the Fed's Greenbook forecasts. We argued that a macro forecast is intended to provide an overall view of the economy and it is, therefore, necessary that the forecasts of all important variables should be evaluated jointly in a multivariate framework. We then showed how a new approach for evaluating economic forecasts permitted us to evaluate the predictions of several variables jointly. We first applied this approach to the Fed's forecasts of the growth rates of ten components of real GDP. We showed that the Fed had a good understanding of the composition of the GDP changes that were occurring in the current quarter but not how these sectors were likely to change next period. Moreover, the Fed's forecasts for the current quarter were comparable in quality to the BEA estimates published at least 30 days after the quarter ended.

We then used the same method to examine the Fed's Greenbook forecasts of growth, inflation and unemployment to determine whether together they presented a substantially correct view of the state of the economy. We found that the Fed's forecasts were generally consistent with the *overall* conditions that actually occurred. We also compared the Fed's forecasts to those of the Survey of Professional Forecasters and found that there was not a significant difference in the quality of the forecasts.

Table 1

Percentage of Time the Signs of the Forecasts of GDP and Components Agreed with the Actuals

	Current Quarter	**1 Quarter Ahead**
GNP/GDP	95%	93%
Consumption Durable Goods	80%	69%
Consumption Non-Durable Goods	82%	72%
Consumption Services	100%	100%
Fixed Investment Nonresidential	83%	77%
Fixed Investment Residential	79%	68%
Private Inventories	86%	72%
Exports	83%	83%
Imports	84%	71%
Government Spending Federal	87%	83%
Government Spending State and Local	89%	80%

Table 2A
P-Values of the Test of No Bias
Real GDP and Components of Current Quarter Forecasts (Sample 1986Q3 – 2004Q4)

	Wald Test		**VAR of Forecast Errors**			
	MZ	**MZ with Dummy**	**Signif. Constant**	**Signif. Own Lags**	**Granger Causality**	**Signif. Dummy**
Real GNP/GDP	**0.005**	**0.014**	**0.000**	0.073	0.053	0.813
Real Consumption Durable Goods	**0.035**	**0.045**	**0.045**	0.346	0.821	0.340
Real Consumption Non-Durable Goods	**0.004**	**0.011**	**0.011**	0.679	0.637	0.624
Real Consumption Services	0.299	0.446	0.979	0.156	0.264	0.996
Real Fixed Investment Nonresidential	0.141	0.261	0.158	0.954	0.353	0.816
Real Fixed Investment Residential	**0.068**	**0.049**	0.259	0.516	0.321	0.327
Real Private Inventories	0.156	**0.003**	0.161	**0.077**	**0.096**	**0.020**
Real Exports	**0.084**	0.177	0.395	0.530	0.643	0.514
Real Imports	**0.004**	**0.000**	**0.064**	0.719	0.764	0.190
Real Government Federal Spending	0.878	0.964	0.518	0.588	0.248	0.892
Real Government State and Local Spending	0.359	0.440	0.584	0.303	0.187	0.197

Rejections of the null hypothesis of no bias at the 10% level in bold.

Table 2B
P-Values of the Test of No Bias
Real GDP and Components of One-quarter Ahead Forecasts (Sample 1986Q3 – 2004Q4)

	Wald Test		VAR of Forecast Errors			
	MZ	**MZ with Dummy**	**Signif. Constant**	**Signif. Own Lags**	**Granger Causality**	**Signif. Dummy**
Real GNP/GDP	**0.014**	**0.000**	0.433	0.175	0.265	**0.003**
Real Consumption Durable Goods	**0.006**	**0.012**	**0.004**	0.796	0.213	0.751
Real Consumption Non-Durable Goods	0.317	0.461	0.677	0.228	0.216	0.315
Real Consumption Services	**0.006**	**0.002**	**0.070**	0.427	**0.071**	0.298
Real Fixed Investment Nonresidential	0.301	**0.068**	0.232	0.540	**0.050**	0.195
Real Fixed Investment Residential	**0.024**	**0.010**	0.328	0.278	0.576	**0.048**
Real Private Inventories	0.936	**0.000**	0.368	0.637	**0.022**	**0.000**
Real Exports	0.933	0.113	0.451	**0.084**	0.009	**0.034**
Real Imports	**0.086**	**0.000**	0.124	**0.034**	0.377	**0.000**
Real Government Federal Spending	**0.096**	0.188	0.116	**0.005**	0.893	0.710
Real Government State and Local Spending	0.210	0.279	0.184	0.152	**0.079**	0.964

Rejections of the null hypothesis of no bias at the 10% level in bold.

Table 3A
Mahalanobis Distance – Entire Sample, 1986Q3 – 2004Q4

	Current Quarter Forecast		One Quarter Ahead Forecast	
	Mean Forecast	**Mean Actual**	**Mean Forecast**	**Mean Actual**
Durable Goods Consumption	4.525	6.329	2.573	6.423
Non-Durable Consumption	1.632	2.295	2.376	2.237
Services Consumption	3.167	3.168	2.856	3.169
Nonresidential Fixed Investment	5.087	6.271	5.219	6.532
Residential Fixed Investment	2.545	3.670	1.209	3.589
Private Inventories	18.080	21.695	21.333	21.513
Exports	5.622	6.762	6.705	6.764
Imports	5.975	8.093	6.271	8.169
Federal Gov. Spending	0.449	0.636	-0.623	1.045
State & Local Gov. Spending	2.122	2.321	1.896	2.315
Mahalanobis Distance (D^2)	0.129		0.539	
F-statistic	0.462		1.898	
p-value	0.912		**0.050**	

Table 3B
Mahalanobis Distance for Recessions – Entire Sample, 1986Q3 – 2004Q4

	Current Quarter Forecast		One Quarter Ahead Forecast	
	Mean Forecast	Mean Actual	Mean Forecast	Mean Actual
Durable Goods Consumption	1.157	5.514	-6.714	1.771
Non-Durable Consumption	-0.743	0.057	-0.129	0.800
Services Consumption	2.100	2.100	2.314	2.514
Nonresidential Fixed Investment	-7.214	-6.043	-4.600	-8.271
Residential Fixed Investment	-7.629	-8.371	-4.929	-3.729
Private Inventories	-29.271	-40.771	-15.457	-47.414
Exports	-3.100	-3.443	2.900	-3.257
Imports	-4.343	-6.757	2.200	-3.357
Federal Gov. Spending	2.543	3.000	1.071	5.414
State & Local Gov. Spending	2.800	3.514	1.886	3.057
Mahalanobis Distance (D^2)	3.678		6.622	
F-statistic	0.322		0.579	
p-value	0.924		0.775	

Table 3C
Mahalanobis Distance for Expansions – Entire Sample, 1986Q3 – 2004Q4

	Current Quarter Forecast		One Quarter Ahead Forecast	
	Mean Forecast	**Mean Actual**	**Mean Forecast**	**Mean Actual**
Durable Goods Consumption	5.216	6.997	3.529	6.901
Non-Durable Consumption	1.796	2.545	2.634	2.385
Services Consumption	3.174	3.191	2.912	3.237
Nonresidential Fixed Investment	5.825	7.154	6.229	8.056
Residential Fixed Investment	3.413	4.672	1.841	4.343
Private Inventories	19.904	24.996	25.121	28.609
Exports	6.299	7.812	7.097	7.796
Imports	6.825	9.371	6.690	9.356
Federal Gov. Spending	0.423	0.261	-0.797	0.596
State & Local Gov. Spending	2.088	2.286	1.897	2.238
Mahalanobis Distance (D^2)	0.253		0.695	
F-statistic	0.815		2.203	
p-value	0.615		**0.022**	

Table 4A
Mahalanobis Distance of the Fed Forecasts of Growth, Unemployment and Inflation – Entire Sample, 1965Q4-2005Q4

	Current Quarter Forecast 1965.4 – 2005.4		One Quarter Ahead Forecast 1966.1 – 2006.1	
	Mean Forecast	**Mean Actual**	**Mean Forecast**	**Mean Actual**
Real GDP Growth	2.595	2.904	2.783	2.891
Unemployment Rate	5.949	5.937	6.013	5.940
Inflation	4.068	4.065	3.973	4.073
Mahalanobis Distance (D^2)	0.010		0.008	
F-statistic	0.271		0.224	
p-value	0.846		0.879	

Table 4B
Mahalanobis Distance for Recessions of the Fed Forecasts of Growth, Unemployment and Inflation – Entire Sample, 1965Q4 – 2005Q4

	Current Quarter Forecast 1965.4 – 2005.4		One Quarter Ahead Forecast 1966.1 – 2006.1	
	Mean Forecast	**Mean Actual**	**Mean Forecast**	**Mean Actual**
Real GDP Growth	-1.878	-1.722	-0.274	-1.722
Unemployment Rate	6.363	6.363	6.281	6.363
Inflation	6.204	6.548	5.907	6.548
Mahalanobis Distance (D^2)	0.023		0.220	
F-statistic	0.102		0.954	
p-value	0.959		0.422	

Table 4C
Mahalanobis Distance for Expansions of the Fed Forecasts of Growth, Unemployment and Inflation – Entire Sample, 1965Q4 – 2005Q4

	Current Quarter Forecast 1965.4 – 2005.4		**One Quarter Ahead Forecast 1966.1 – 2006.1**	
	Mean Forecast	**Mean Actual**	**Mean Forecast**	**Mean Actual**
Real GDP Growth	3.496	3.836	3.399	3.821
Unemployment Rate	5.866	5.851	5.959	5.855
Inflation	3.637	3.564	3.584	3.574
Mahalanobis Distance (D^2)	0.024		0.042	
F-statistic	0.534		0.933	
p-value	0.659		0.425	

Table 5
Mahalanobis Distance Between the SPF and the Fed Forecasts of Growth, Unemployment and Inflation

	Current Quarter Forecast 1968.4 – 2005.4		**One Quarter Ahead Forecast 1969.1 – 2006.1**	
	Mean SPF Forecast	**Mean Fed Forecast**	**Mean SPF Forecast**	**Mean Fed Forecast**
Real GDP Growth	2.428	2.453	2.749	2.687
Unemployment Rate	6.147	6.121	6.175	6.193
Inflation	4.097	4.146	4.031	4.046
Mahalanobis Distance (D^2)	0.001		0.001	
F-statistic	0.035		0.026	
p-value	0.991		0.994	

References

Abdi, Herve (2007). Distance. In N.J. Salkind (Ed.) Encyclopedia of Measurement and Statistics. Thousand Oaks CA: Sage, pp. 280-284.

Banternghansa, Chanont and Michael W. McCracken (2009). Forecast Disagreement Among FOMC Members. Federal Reserve Bank of St. Louis Working Paper No. 2009-059.

Caunedo, Julieta, Riccardo DiCecio, Ivana Komunjer, and Michael T. Owyang (2013) "Federal Reserve Forecasts: Asymmetry and State-Dependence" Federal Reserve Bank of St. Louis Working Paper 2013-012.

Clements, Michael P., Fred Joutz and Herman O. Stekler, (2007). An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach. Journal of Applied Econometrics, 22(1), 121-136.

Davies, Antony and Kajal Lahiri, (1995). "A new framework for testing rationality and measuring aggregate shocks using panel data." *Journal of Econometrics* 68: 205-227.

Davies, Antony and Kajal Lahiri, (1999). "Re-examining the rational expectations hypothesis using panel data on multi-period forecasts." In Hsaio, C., K. Lahiri, L.F. Lee, and M.H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variable Models*, 226-254. Cambridge University Press.

Davies, Antony, Kajal Lahiri, and Xuguang Sheng (2011). "Analyzing Three-Dimensional Panel Data of Forecasts." *Oxford Handbook on Economic Forecasting*, 473-495: 1-13

Groen, Jan J. J., George Kapetanios and Simon Price (2009).Real Time Evaluation of Bank of England Forecasts for Inflation and Growth. International Journal of Forecasting, 25, 74-80.

Hanson, M. S. and J. Whitehorn (2006).Reconsidering the Optimality of Federal Reserve Forecasts. Manuscript.

Holden, K. and D.A. Peel (1990).On testing for unbiasedness and efficiency of forecasts. Manchester School, 48, 120–127.

Joutz, F. and H. Stekler (2000).An evaluation of the predictions of the Federal Reserve. International Journal of Forecasting, 16, 17-38.

Jordá, Òscar, Malte Knüppel, and Massimiliano Marcellino (2010).Empirical Simultaneous Confidence Regions for Path-Forecasts. European University Institute Economics Working Papers No.ECO2010/18.

Komunjer, Ivana and Michael T. Owyang (2012). *Review of Economics and Statistics*. Vol. 94, No. 4: 1066-1080.

McLachlan, G. J. (1999) "Mahalanobis Distance" *Resonance*, pp. 20 -26.

Mincer, J. and V. Zarnowitz (1969).The Evaluation of Economic Forecasts. In J. Mincer (Eds.) Economic Forecasts and Expectations. National Bureau of Economic Research, New York.

Reifschneider, D. and Tulip, P. (2007). Gauging the Uncertainty of the Economic Outlook from Historical Forecasting Errors. Federal Reserve Board Finance and Economics Discussion Series, 2007-60.

Romer, Christina D. and David Romer (2000).Federal Reserve Information and the Behavior of Interest Rates. American Economic Review, 90(3), 429-457.

Sims, Christopher A. (2002). The Role of Models and Probabilities in the Monetary Policy Process. Brookings Papers on Economic Activity, 33(2), 1-62.

Sinclair, Tara M., Edward N. Gamber, Herman Stekler, Elizabeth Reid (2012). Jointly Evaluating the Federal Reserve's Forecasts of GDP Growth and Inflation, International Journal of Forecasting. Volume 28, Issue 2, Pages 309–314.

Sinclair, Tara M., Fred Joutz, and H. O. Stekler (2010). Can the Fed Predict the State of the Economy? Economics Letters, 108, 1, July, 28 – 32.

Sinclair, T. M., Stekler, H.O. and Kitzinger, L. (2010). Directional Forecasts of GDP and Inflation: A Joint Evaluation with an Application to Federal Reserve Predictions. Applied Economics 42(18), 2289 - 2297.

Sinclair, T. M., and Stekler, H.O. (2011). Examining the Quality of Early GDP Component Estimates. George Washington University Research Program on Forecasting Working Paper No. 2011-001.

Sinclair, Tara M., H. O. Stekler and Warren Carnow (2012). "A New Approach for Evaluating Economic Forecasts" *Economics Bulletin,* Vol. 32 No. 3 pp. 2332-2342.