International Journal of Forecasting 31 (2015) 157–164

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Evaluating a vector of the Fed's forecasts

Tara M. Sinclair *, H.O. Stekler, Warren Carnow

Department of Economics, The George Washington University, Monroe Hall #340, 2115 G Street NW, Washington, DC 20052, United States

ARTICLE INFO

JEL classification:
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 evaluate jointly 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.

1. Introduction

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 previous evaluations have used a univariate methodology that examined the Fed's forecasts of select variables such as GDP, inflation, and unemployment separately (Clements, Joutz, & Stekler, 2007; Joutz & Stekler, 2000; Romer & Romer, 2000; Sims, 2002; Stekler, 1994).

However, there have been a small number of studies that have considered some multivariate characteristics of the Fed's forecasts. For example, Sinclair, Stekler, and Kitzinger (2010) examined the joint *directional* forecasts of GDP and inflation using contingency tables; and Sinclair, Gamber, Stekler, and Reid (2012) calculated the costs of forecast errors of GDP and inflation jointly within the context of a Taylor-type rule. These studies, however, did not develop a *general* approach for the joint evaluation of quantitative forecasts.

Focusing on the rationalizability of the forecasts, Caunedo, DiCecio, Komunjer, and Owyang (2013) jointly test the rationality of the Federal Reserve's forecasts of inflation, unemployment, and output growth. To do this, 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. Forecast comparison is relevant because we may want to know whether we can substitute one forecast for another, or use forecasts in place of actual data for policy decisions in real time.

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 with either a univariate or a multivariate approach. 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 forecasts of the ten main components of GDP have never

* Corresponding author. Tel.: +1 202 994 7988.
E-mail address: tsinc@gwu.edu (T.M. Sinclair).

http://dx.doi.org/10.1016/j.ijforecast.2014.02.002
0169-2070/

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 provide an accurate, comprehensive view of the various sectors of the economy. This question is especially important if the forecasts are 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 probably 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 no simple answer, because there are a number of possible ways of doing this analysis, ranging from the simple univariate single-horizon method to more complex methods which aggregate across the various dimensions of the data.

In general, the database of forecasts 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 the 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 most appropriate procedure for such aggregation depends on the question that is being investigated. For example, Eisenbeis, Waggoner, and Zha (2002) 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 et al. (2007) and Davies and Lahiri (1995, 1999) do not pool across variables,[2] but pool across horizons for each variable. One difference between those 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 the 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 the 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 calculate MSEs for each of these additional variables. If the MSE of one variable were "small", while those 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?

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 (2013) utilized for analyzing early GDP component estimates from the Bureau of Economic Analysis. Their 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. In order to determine whether the two sets of estimates were related, it was necessary to compare the difference between the two vectors. Sinclair and Stekler utilized the Mahalanobis measure for estimating the relationship between the 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 for analyzing the Fed's forecasts. One vector will consist of all of the forecasts of the different variables that the Fed made at one time that refer to a particular point in time, while the other vector will comprise the actual outcomes (or alternative forecasts) for those variables. In addition, we will also apply the new vector generalization of the Holden and Peel (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

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.

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 Forecasters for GDP growth, unemployment, and inflation.