Economics Bulletin

Volume 32, Issue 3

A new approach for evaluating economic forecasts

Tara M. Sinclair
The George Washington University

H. O. Stekler
The George Washington University

Warren Carnow
The George Washington University

Abstract

This paper presents a recently developed approach for evaluating economic forecasts. Previously, univariate methods were used to evaluate the forecasts of individual variables. However, many macroeconomic variables are forecast at the same time to describe the state of the economy. It is, therefore, appropriate to use a multivariate methodology in evaluating these forecasts. Our approach uses VARs and distance measures. It is applied to the Survey of Professional Forecasters (SPF). Our contributions are the application of the methodology for evaluating multivariate forecasts to the SPF, measuring accuracy, and testing for bias within this framework. We also consider whether there are forecasting performance asymmetries over the business cycle.

The authors gratefully acknowledge generous support from the Institute for International Economic Policy and helpful comments from and discussions with Neil Ericcson, Fred Joutz, Kajal Lahiri, and other participants at the ISF 2012.
Citation: Tara M. Sinclair and H. O. Stekler and Warren Carnow, (2012) "A new approach for evaluating economic forecasts ", *Economics Bulletin*, Vol. 32 No. 3 pp. 2332-2342.
Contact: Tara M. Sinclair - tsinc@gwu.edu, H. O. Stekler - hstekler@gwu.edu, Warren Carnow - wcarnow@gwmail.gwu.edu.
Submitted: April 30, 2012. **Published:** August 23, 2012.

1. Introduction

This paper presents a recently introduced approach for evaluating forecasts of multiple macroeconomic variables which we apply for the first time to the Survey of Professional Forecasters (SPF). The traditional evaluation approach has been to assess the forecasts of each variable (GDP, inflation, unemployment, etc.) separately. This has been the way that previous evaluations of the SPF forecasts have been conducted (Baghestani 1994, Baghestani 2006, Clements 2006 and 2009, and Diebold, Tay, and Wallis 1999).

However, forecasts of multiple macroeconomic variables are often relied upon to provide a holistic picture of the state of the economy. In that case the forecasts of all important variables should be evaluated **jointly** in a multivariate framework. Using contingency tables, Sinclair, Stekler and Kitzinger (SSK, 2010) introduced a procedure to jointly evaluate directional forecasts. SSK justified a joint evaluation of the directional accuracy of the Fed's GDP growth and inflation forecasts because the Federal Reserve considers both variables jointly in making monetary policy decisions. However, SSK only examined the joint directional accuracy of the forecasts. They did not determine whether the quantitative predictions together correctly described the state of the economy or if the forecasts were biased. These are the issues that we examine.[1]

We first present the methodology for jointly evaluating the quantitative forecasts of several variables that can describe the state of the economy. The methodology is based on the approach that Sinclair and Stekler (forthcoming) utilized to determine whether the earliest vintage of estimates of the set of major GDP sub-components was similar to a later vintage of estimates.[2] This procedure is then applied to the SPF's forecast of real growth, inflation, and unemployment that together describe the future state of the economy. This note makes two contributions to the methodology for evaluating multivariate forecasts in the SPF: measuring accuracy within this framework and testing for bias. We also explicitly consider whether there may be asymmetries in terms of forecasting performance in recessions as compared to expansions.

2. Multivariate Evaluation

There have been many evaluations of economic forecasts; these univariate evaluations have separately examined the forecasts of individual variables such as GDP, inflation, and unemployment. Because these forecasts are produced and/or used jointly, they therefore, should be judged using a multivariate methodology to determine whether they are unbiased and provide an accurate picture of the entire state of the economy.

The methodology for evaluating economic forecasts is similar to that used by Sinclair and Stekler (forthcoming). They analyzed the data revisions of the growth rates of ten components of GDP. The first set of estimates relating to a particular quarter comprised one vector. The revised

[1] As an alternative multivariate forecast evaluation, Komunjer and Owyang (forthcoming) use forecast errors in a multivariate framework to derive the weights of a utility function. Another approach, that of Sinclair et al. (2012), is to evaluate the forecasts within the context of a specific loss function.

[2] This methodology has also been used by Sinclair, Stekler, and Carnow (2012) in their analysis of the Fed's Greenbook forecasts of ten major GDP sub-components.

estimates for that quarter constituted a different vector. A generalization of Euclidian distance known as Mahalanobis distance was used as an accuracy metric to test whether there was a difference between the two vectors of estimates. In this paper we utilize the same methodology and apply it to the SPF forecasts. One vector will be the SPF forecasts that refer to a particular point in time; the other will be the actual outcomes for those variables. We also develop a VAR procedure based on Holden and Peel (1990) for testing whether forecasts of a vector of variables were biased.

This note makes a number of contributions to the literature on evaluating macroeconomic forecasts. First, we use a multivariate framework to analyze a set of forecasts. The SPF forecasts have never previously been analyzed in this framework. While these forecasts of growth and inflation had been examined separately, we also include the unemployment rate in the new framework. We thus are able to consider whether those forecasts, taken together, accurately described the state of the economy. Moreover, our approach for testing for bias includes more information that was known and should have been included in the forecast. Finally, we also explicitly consider whether there may be asymmetries in terms of forecasting performance in recessions as compared to expansions.

3. Data

We examine the SPF's consensus forecasts of three variables: the growth rate of real GDP, the rate of inflation and the rate of unemployment. These are the current quarter and the one quarter-ahead predictions made between 1968.4 and 2011.1. The actual data are the estimates released 90 days after the quarter to which they refer.

4. Multivariate Analysis Methodology

4.1 Bias

We first use a *joint* framework to investigate the properties of the forecasts errors of these three variables. We construct a first-order vector autoregression (VAR(1)) of the errors made in forecasting each of the three variables. This is a generalization of a Holden-Peel (1990) test for bias: if the forecasts are unbiased estimates of the outcomes, none of the coefficients in the VAR should be significant: the constant estimates should be zero; the coefficients on the own lags should be zero; and none of the past errors made in forecasting the other variables should Granger-cause any of the other errors. The VAR (1) consisting of the forecast errors of GDP, inflation, and unemployment is:

$$FE_t = \beta_0 + FE_{t-1}\beta_1 + e_t, \tag{1}$$

where FE_t is a vector of the forecast errors for time t, β_0 is a vector of the constant terms, and β_1 is a matrix of coefficients on the lags of the forecast errors. The null hypothesis is that all of the elements of both β_0 and β_1 are zero.

Forecasts sometimes contain systematic errors (Joutz and Stekler 2000, Hanson and Whitehorn 2006) with the rate of growth overestimated during slowdowns and recessions and underestimated during recoveries and booms.[3] In some cases, these systematic errors, associated

[3] Similarly, inflation was under-predicted when it was rising and over-predicted when it was declining.

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 no bias when in fact these systematic errors exist.

In order to determine whether the SPF forecasts similarly failed to incorporate information about the state of the economy, we modified (1) as in Sinclair, Joutz, and Stekler (2010):

$$FE_t = \beta_0 + FE_{t-1}\beta_1 + D_t\beta_2 + e_t, \tag{2}$$

where D_t is a dummy reflecting the state of the economy, taking on the value 1 if during one month of a particular quarter the economy was in an NBER-dated recession. Otherwise, the value of the dummy is zero. 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, they contain information that can explain the forecast errors. That would indicate that the SPF forecasts did not fully incorporate information about the state of the economy.

4.2 Accuracy

We use a distance measure to determine the accuracy or difference of the vectors. There are two common measures of distance, Euclidean and Mahalanobis, but they differ in the assumptions made about the statistical independence of the vectors. Euclidean distance is only applicable to vectors that have independent elements and that are scaled so that they have unit variances. These assumptions do not apply in this analysis. Thus, we will use Mahalanobis Distance, D^2, a generalization of the Euclidian distance, which allows for the scale to differ across the different variables and for nonzero correlation between the variables.[4] 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:

$$D^2 = (\bar{F} - \bar{A})'W(\bar{F} - \bar{A}), \tag{3}$$

where W is the inverse of the pooled sample variance-covariance matrix, and $\bar{F}$ and $\bar{A}$ are the mean vectors of the forecasts and outcomes, respectively.[5] Under the assumption of normality,[6] we can construct an F-statistic based on this measure to test the null hypothesis that the forecasts and outcomes have the same population means.[7]

[4] 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).

[5] 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.

[6] Evidence has been provided to question the assumption of normality for forecast errors in the SPF by both Lahiri and Teigland (1987) and Harvey and Newbold (2003). Our hypothesis test, however, is based on the sampling distribution of the forecasts and the actuals. The test is analogous to a difference-in-means t-test in a univariate framework (see Banternghansa and McCracken 2009). Therefore, the assumption of normality is appropriate with a large sample. Our full sample includes 170 observations which should be sufficiently large.

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

5. Multivariate Results

Tables 1 and 2 present the results from the tests used to determine whether the current and one-quarter-ahead forecasts of the three variables were biased. We show the p-values obtained from the joint test using the 3-equation VAR. The joint test rejects the null in at least one dimension for each of the variables and time periods. The coefficient on the NBER dummy is significant for all of the variables in the VAR. These results suggest that forecasters for the SPF do not know the state of the business cycle when making their forecasts.

Despite the evidence of these biases in the forecasts, we needed to determine whether the forecasts of the three variables, taken together, provided an overall view of the state 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 three forecasts. The null was that the SPF forecasts provided an overall view of the state of the economy that was consistent with the observed data. (Table 3). We did not reject the null for either the current or one-quarter- ahead. These results indicate that the consensus SPF predictions provided a good understanding of the state of the economy. However, when we split the sample into recession observations and expansion observations (Tables 4 and 5), we find that we can reject the null for the one-quarter ahead forecast for both recessions and expansions at the 10% level. This suggests that there are offsetting errors in the one-quarter ahead forecasts.

6. Comparison with Univariate Results

We have presented a multivariate methodology for evaluating the forecasts. Are the results relating to bias different from those obtained using a univariate analysis? The Mincer-Zarnowitz equation (4) is used to test for bias and efficiency in the univariate procedure:

$$A_t = \beta_0 + \beta_1 F_t + e_t, \tag{4}$$

where A_t and F_t are the actual real-time data and the SPF forecasts, respectively. The null hypothesis is: $\beta_0 = 0 \; and \; \beta_1 = 1$. If a recession dummy is included, the equation becomes:

$$A_t = \beta_0 + \beta_1 F_t + \beta_2 D_t + e_t, \tag{5}$$

The joint null hypothesis now is: $\beta_0 = 0, \beta_1 = 1, and \; \beta_2 = 0.$

The results obtained from the two univariate Mincer-Zarnowitz equations are in Table 6. In many cases, the results are identical, but there are differences in the one-quarter-ahead forecasts. Without the recession dummy, the null of no bias is not rejected for any of the three variables using the univariate equations, but the null is rejected for all the variables using the VAR.

Even if the results of the two approaches had been identical, there are several reasons for using the multivariate approach. First, it permits one to determine whether a particular forecast accurately described “general economic conditions” and whether the forecasts of all the variables were consistent with each other.[8] Second, in testing for bias, the new approach includes more information that was known and should have been included in the forecast.

7. Conclusions

This paper presented a recently introduced methodology for evaluating economic forecasts. This multivariate framework enabled us to jointly evaluate the SPF predictions of GDP growth, inflation, and unemployment. We found that both the current quarter and one-quarter-ahead forecasts were generally consistent with the observed data. However, we also found that the forecasts contained biases and offsetting errors, especially during recessions. Many of these results were similar to those obtained from the univariate approach. We argued, however, that there were valid reasons for preferring the multivariate framework.

[8] In another context, Clements (2009) questioned whether two probability forecasts within the SPF were consistent with each other.

Table 1
P-Values of Tests of the Null of No Bias Current Quarter SPF Forecasts
(Sample 1968Q4 – 2011Q1)

	VAR of Forecast Errors			
	Signif. Constant	**Signif. Own Lags**	**Granger Causality**	**Signif. Dummy**
Real GNP/GDP	**0.076**	0.499	**0.033**	**0.002**
Unemployment	**0.044**	**0.007**	0.859	**0.000**
Inflation	0.955	0.326	0.147	**0.026**

Table 2
P-Values of Tests of the Null of No Bias One-Quarter Ahead SPF Forecasts
(Sample 1969Q1 – 2011Q1)

	VAR of Forecast Errors			
	Signif. Constant	**Signif. Own Lags**	**Granger Causality**	**Signif. Dummy**
Real GNP/GDP	0.751	0.241	**0.000**	**0.000**
Unemployment	0.151	**0.000**	**0.001**	**0.000**
Inflation	0.991	**0.000**	0.104	**0.012**

Table 3
Mahalanobis Distance between the SPF and the 90-Day Estimates

	Current Quarter Forecast 1968Q4 – 2011Q1		One Quarter Ahead Forecast 1969Q1 – 2011Q1	
	Mean SPF Forecast	**Mean Actuals**	**Mean SPF Forecast**	**Mean Actuals**
Real GDP Growth	2.326	2.634	2.656	2.629
Unemployment Rate	6.238	6.204	6.259	6.221
Inflation	3.833	3.849	3.775	3.849
Mahalanobis Distance (D^2)	0.013		0.002	
F-statistic	0.363		0.051	
p-value	0.780		0.985	
Observations	170		169	

Table 4
Recessionary Periods
Mahalanobis Distance between the SPF and the 90-Day Estimates

	Current Quarter Forecast Recessions		One Quarter Ahead Forecast Recessions	
	Mean SPF Forecast	**Mean Actuals**	**Mean SPF Forecast**	**Mean Actuals**
Real GDP Growth	-0.815	-1.621	0.871	-1.621
Unemployment Rate	6.306	6.391	6.032	6.391
Inflation	5.168	5.594	4.771	5.594
Mahalanobis Distance (D^2)	0.074		0.751	
F-statistic	0.406		4.125	
p-value	0.749		**0.010**	
Observations	34		34	

Table 5
Expansionary Periods
Mahalanobis Distance between the SPF and the 90-Day Estimates

	Current Quarter Forecast Expansions		**One Quarter Ahead Forecast Expansions**	
	Mean SPF Forecast	**Mean Actuals**	**Mean SPF Forecast**	**Mean Actuals**
Real GDP Growth	3.112	3.698	3.105	3.699
Unemployment Rate	6.221	6.157	6.316	6.178
Inflation	3.499	3.413	3.524	3.409
Mahalanobis Distance (D^2)	0.085		0.096	
F-statistic	1.914		2.134	
p-value	0.128		**0.096**[9]	
Observations	136		135	

Table 6
P-Values of Univariate Tests of the Null of No Bias

	Wald Test Current Quarter 1968Q4 – 2011Q1		**Wald Test One Quarter Ahead 1969Q1 – 2011Q1**	
	MZ	**MZ with Dummy**	**MZ**	**MZ with Dummy**
Real GNP/GDP	**0.043**	**0.003**	0.986	**0.000**
Unemployment	**0.030**	**0.000**	0.151	**0.000**
Inflation	0.230	0.114	0.537	**0.004**

[9] It is simply a coincidence that the p-value of the F-test and the Mahalanobis distance are the same to three decimal places in this case. To four decimal places the Mahalanobis distance is 0.0956 and the p-value is 0.0962.

References

Abdi, H. (2007) "Distance" in *Encyclopedia of Measurement and Statistics* by N.J. Salkind, Ed., Sage: Thousand Oaks CA, 280-284.

Baghestani, H. (2006) "An Evaluation of the Professional Forecasts of U.S. Long-term Interest Rates" *Review of Financial Economics* **15**, 177-191.

Baghestani, H. (1994) "Evaluating Multiperiod Survey Forecasts of Real Net Exports" *Economic Letters* **44**, 267-72.

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

Clements, M. P., F. Joutz and H. O. Stekler (2007) "An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach" *Journal of Applied Econometrics* **22**, 121-136.

Clements, M. P. (2006) "Evaluating the Survey of Professional Forecasters' Probability Distributions of Expected Inflation Based on Derived Event Probability Forecasts" *Empirical Economics* **31**, 49-64.

Clements, M. P. (2009) "Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters" in *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry* by J. L. Castle and N. Shephard, Eds., Oxford University Press: Oxford, 206 - 226.

Diebold, F. X., A. S. Tay, and K. F. Wallis (1999) "Evaluating Density Forecasts: The Survey of Professional Forecasters" in Robert F. Engle and Halbert White, Eds., *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger* Oxford University Press: Oxford, 76-90.

Groen, J. J. J., G. Kapetanios and S. 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 available at http://mshanson.web.wesleyan.edu/research/HWReconsid-Draft.pdf.

Harvey, D. I. and P. Newbold (2003) "The non-normality of some macroeconomic forecast errors" *International Journal of Forecasting* **19**, 635-653.

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

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

Jordá, Ò., M. Knüppel, and M. Marcellino (2010) “Empirical Simultaneous Confidence Regions for Path-Forecasts” European University Institute Economics Working Papers No. ECO2010/18.

Komunjer, I. and M. T. Owyang (forthcoming). “Multivariate Forecast Evaluation and Rationality Testing” *Review of Economics and Statistics*.

Lahiri, K. and Teigland, C. (1987) “On the Normality of Probability Distributions of Inflation and GNP Forecasts” *International Journal of Forecasting* **3**, 269-279.

McLachlan, G. J. (1999) “Mahalanobis Distance” *Resonance*, 20 -26.

Mincer, J. and V. Zarnowitz (1969) “The Evaluation of Economic Forecasts” in *Economic Forecasts and Expectations* by J. Mincer, Ed. National Bureau of Economic Research: New York, 1-46.

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

Romer, C. D. and D. Romer (2000) “Federal Reserve Information and the Behavior of Interest Rates” *American Economic Review* **90**, 429-457.

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

Sinclair, T. M., E. N. Gamber, H. O. Stekler, and E. Reid (2012) “Jointly Evaluating the Federal Reserve’s Forecasts of GDP Growth and Inflation” *International Journal of Forecasting* **28**, 309–314.

Sinclair, T. M., F. Joutz, and H. O. Stekler (2010) “Can the Fed Predict the State of the Economy?” *Economics Letters* **108**, 28 – 32.

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

Sinclair, T. M. and H. O.Stekler (forthcoming) “Examining the Quality of Early GDP Component Estimates” *International Journal of Forecasting*.

Sinclair, T. M., H.O. Stekler, and W. Carnow (2012) “Evaluating a Vector of the Fed's Forecasts.” George Washington University Research Program on Forecasting Working Paper No. 2012-002.