A NEW APPROACH
FOR EVALUATING ECONOMIC 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

A NEW APPROACH
FOR EVALUATING ECONOMIC 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

Abstract

This paper presents a new approach to evaluating multiple economic forecasts. In the past, evaluations have focused on the forecasts of individual variables. However, many macroeconomic variables are forecast at the same time and are used together to describe the state of the economy. It is, therefore, appropriate to examine those forecasts jointly. This specific approach is based on the Sinclair and Stekler (forthcoming) analysis of data revisions. The main contributions of this paper are (1) the application of this technique to the Survey of Professional Forecasters (SPF) and (2) showing that there is a bias that is associated with the stages of the business cycle.

A NEW APPROACH FOR EVALUATING ECONOMIC FORECASTS

This paper presents a new approach for evaluating macroeconomic forecasts and then applies it to the Survey of Professional Forecasters (SPF). The traditional evaluation approach has been to examine the forecasts of each variable (GDP, inflation, unemployment, etc.) separately. This has been the way that previous analyses of the SPF forecasts have been conducted (Baghestani, 1994; Baghestani, 2006; Clements, 2006; 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. Sinclair, Stekler and Kitzinger (SSK, 2010) introduced a procedure to jointly evaluate forecasts in terms of their direction using contingency tables. SSK justified a joint evaluation of the directional accuracy of the Fed's GDP growth and inflation forecasts because the Federal Reserve considered both variables jointly in making monetary policy decisions. However, their analysis 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.

We first present the methodology for jointly evaluating the quantitative forecasts of several variables that can describe the state of the economy. This is 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.[1] Then this

---

[1] 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.

procedure is 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 for evaluating multivariate forecasts: measuring accuracy within this framework and testing for bias. We also explicitly take into consideration that there may be asymmetries in terms of forecasting performance in recessions as compared to expansions.

I. Multivariate Evaluation

There have been many evaluations of economic forecasts; these evaluations have separately examined the forecasts of select variables such as GDP, inflation, and unemployment. There is a concern that arises from this univariate evaluation approach: namely that these forecasts are produced and/or used jointly and, therefore, should be judged together on whether they are unbiased and provide an accurate comprehensive picture of the entire state of the economy.

Sinclair and Stekler (forthcoming) analyzed a set of estimates of the growth rates of ten GDP sub-components as a vector comprising a particular vintage of data relating to that particular quarter. The revised estimates for that quarter created a different vector. As an accuracy metric, a generalization of Euclidian distance known as Mahalanobis distance was used to test whether there was a difference between the two vectors of estimates. Sinclair, Stekler and Carnow (2012) developed a VAR procedure based on Holden and Peel (1990) for testing whether variables comprising a vector of forecasts that were issued jointly were biased. In this paper we utilize the same methodologies and apply them 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.

II. Data

We examine the SPF's consensus forecasts of three variables: the growth rate of real GDP, the rate of inflation and the level 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.

III. Methodology

A. Single Variable Analysis (bias).

We first analyze the forecast errors of each variable separately and determine whether they were systematically related to the actual data. First, we test this relationship using the Mincer-Zarnowitz (1969) regression:

$$A_t = \beta_0 + \beta_1 F_t + e_t, \tag{1}$$

where A_t and F_t are the actual real-time data and the SPF forecasts, respectively. The null hypothesis is: $\beta_0 = 0 \ and \ \beta_1 = 1$. 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.[2]

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. Similarly, inflation was under-predicted when it was rising and over-predicted when it was declining. In 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 these systematic errors exist.

In order to determine whether the SPF forecasts similarly failed to incorporate

---

[2]An alternative procedure for testing for bias is to use: $A_t - F_t = \beta_0 - e_t$. (Holden and Peel, 1990).

information about the state of the economy, we modified (1) as in Sinclair, Joutz, and Stekler (2010) and it now becomes:

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

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.

B. Multivariate Analysis (Bias)

We next 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 unbiasedness: 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 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:

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

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.

C. Multivariate Analysis (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 are independent 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.[3] 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:

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

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.[4] Under the assumption of normality, 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.[5]

III. 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 two Mincer-Zarnowitz (MZ) equations and from the joint test using the 3-

---

[3]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).

[4] 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.

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

equation VAR. The results depend on the test that is used. The null of no bias is rejected for the current quarter growth and unemployment forecasts regardless of which MZ equation is used. For the one-quarter-ahead forecasts, the traditional equations do not reject the null, but it is rejected when the cyclical dummy is included, indicating the presence of offsetting errors. 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 and in all but the current quarter inflation equation for the univariate evaluations. 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.

V. Conclusions

We adapted a new methodology that enabled us to evaluate the SPF predictions of GDP growth, inflation, and unemployment in a joint framework. 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.

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

	Wald Test		VAR of Forecast Errors			
	MZ	MZ with Dummy	Signif. Constant	Signif. Own Lags	Granger Causality	Signif. Dummy
Real GNP/GDP	**0.043**	**0.003**	**0.076**	0.499	**0.033**	**0.002**
Unemployment	**0.030**	**0.000**	**0.044**	**0.007**	0.859	**0.000**
Inflation	0.230	0.114	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)

	Wald Test		VAR of Forecast Errors			
	MZ	MZ with Dummy	Signif. Constant	Signif. Own Lags	Granger Causality	Signif. Dummy
Real GNP/GDP	0.986	**0.000**	0.751	0.241	**0.000**	**0.000**
Unemployment	0.151	**0.000**	0.151	**0.000**	**0.001**	**0.000**
Inflation	0.537	**0.004**	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**[6]	
Observations	136		135	

---

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

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

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

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

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.

Clements, Michael P. "Evaluating the Survey of Professional Forecasters' Probability Distributions of Expected Inflation Based on Derived Event Probability Forecasts," Empirical Economics 31:1 (2006), pp. 49-64.

Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis. "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: Oxford University Press, 1999.

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 (forthcoming). Review of Economics and Statistics. 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 (forthcoming). Jointly Evaluating the Federal Reserve's Forecasts of GDP Growth and Inflation, International Journal of Forecasting.

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. (forthcoming). Examining the Quality of Early GDP Component Estimates. International Journal of Forecasting.

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