Multivariate Forecast Errors and the Taylor Rule

Tara M. Sinclair[1]
Department of Economics
George Washington University
Washington, DC 20052
tsinc@gwu.edu

Edward N. Gamber
Department of Economics and Business
Lafayette College
Easton, PA 18042
gambere@lafayette.edu

H.O. Stekler
Department of Economics
George Washington University
Washington, DC 20052
hstekler@gwu.edu

Elizabeth Reid
Department of Economics
George Washington University
Washington, DC 20052
ereid@gwu.edu

JEL Codes: C53, E37, E52, E58

Keywords: Forecast Evaluation, Federal Reserve Forecasts, Monetary Policy

March 18, 2009

Abstract

This paper evaluates the potential impact of forecast errors on policy. We use the Taylor (1993) monetary policy rule to jointly evaluate the Federal Reserve staff forecasts of U.S. real output growth and the inflation rate. Our simple methodology generates "policy forecast errors" which have a direct interpretation for the impact of forecast errors on policy. In the case of the Federal Reserve, we find that, on average, Fed policy based on the Taylor rule would have been approximately a full percentage point away from the intended target because of errors in forecasting growth and inflation.

[1] Corresponding author. Phone: 202-994-7988. Fax: 202-994-6147. The authors wish to thank Mike Bradley, Dean Croushore, Fred Joutz, George Monokroussos, Keith Ord, Michael Owyang, Roberto Sameniego, Peter Tinsley, and Harry Watson for helpful discussions. We also thank participants in the GWU Brown Bag Seminar Series on Forecasting, the 16th Federal Forecasters Conference, the 28th Annual International Symposium on Forecasting, West Virginia University Economics Department Seminar, Richmond University Economics Department Seminar, University of Memphis Economics Department Seminar, and Wesleyan University Economics Department Seminar for their comments and suggestions. Alexander Kapinos provided essential research assistance.

Introduction

Individuals, businesses, and government policymakers use economic forecasts as guides to many important economic and financial decisions. Thus, it is imperative to evaluate the accuracy of economic forecasts and determine the effects that they have on policy. There have been innumerable evaluations of these forecasts. While most evaluations focused on the accuracy of a particular variable, there have also been discussions about the appropriate procedures for evaluating multivariate forecasts (See Hymans, 1968; Clements and Hendry and subsequent comments, 1993; Eisenbeis, et al., 2002; and Komunjer and Owyang, 2007). The recent research has examined forecasts of multiple variables taking into consideration the characteristics of these multivariate forecasts such as the relative variances and co-movement of the series. However, these evaluations involving multiple variables were not done in the context of policy loss functions or decision rules.

In this paper we develop a simple methodology to jointly evaluate the forecasts of a number of variables and quantify their potential policy implications. Because we focus on the potential policy impact of the forecast errors, our work is in the spirit of evaluating forecasts by considering the economic costs of prediction errors (see Clements, 2004; Granger and Pesaran, 2000a and 2000b; and Pesaran and Skouras, 2002). We ask the following question: Given a specific policy rule, what is the impact of forecast errors on the policy decision?

We use this methodology to evaluate the Greenbook forecasts that are made by the staff of the Board of Governors of the Federal Reserve. The Greenbook forecasts are produced before each meeting of the Federal Open Market Committee and contain projections on the economy up to eight quarters into the future. Many studies have already examined these predictions of the real output growth and inflation rate (e.g. Clements et al., 2007; Joutz and Stekler, 2000; Romer and Romer, 2000; Sims, 2002; Stekler, 1994). The previous studies have measured the accuracy of these variables individually. However, as these two variables are related and forecasted at the same time, it is necessary to evaluate them jointly.

A joint evaluation could consider either the directional accuracy of the two variables or a quantitative measure of the loss associated with making policy based on forecasts of real output growth and inflation. Sinclair et al. (forthcoming) provide the procedures for jointly evaluating the directional accuracy of these variables. They showed that while the directional forecasts, by themselves, were found to not always be valuable, the joint forecasts were valuable overall. However, they did not consider the procedures for jointly evaluating the *quantitative* predictions, nor did they show the impact of the forecast errors on policy.

This paper focuses on this problem by constructing a quantitative policy forecast error measure that is based on a particular policy decision rule. While the methodology for evaluating policy forecast errors is general, for illustration we apply it to the Federal Reserve's decision rule for setting monetary policy. The next section presents the policy forecast error measure. Subsequent sections present the data, the Fed's implicit policy rule, the evaluation procedure, the results, robustness checks, and our conclusions.

I. Methodology for Combining the Forecast Errors

In order to jointly evaluate forecasts of any two or more variables simultaneously, an appropriate weighted measure is needed to combine the errors of the variables. We suggest that forecasts of multiple series are generally generated for a specific policy purpose, whether the forecasts are used as inputs for monetary policy, fiscal policy, a firm's pricing policy, or for any other use.

Specifically, let $P^f_{t,t+h}$ be a policy decision at time *t* that is a linear function of the h-step ahead forecasts of N $\geq$ 1 variables ($x^f_{i,t+h}$, $i = 1,...N$). The superscript *f* indicates that the policy decision is based on forecasts rather than the actual outcomes of the variables:

$$P^f_{t,t+h} = p(x^f_{1,t+h},...x^f_{N,t+h}) . \qquad (1)$$

If policymakers have perfect foresight, the policy decision would simply be P_t without the superscript *f*:

$$P_{t,t+h} = p(x_{1,t+h},...x_{N,t+h}) . \qquad (2)$$

However, because policy is based on forecasts, rather than on the actual data, policy is subject to errors which are functions of the mistakes made in forecasting the underlying variables $x_{i,t}$, $i = 1,...N$. We call the difference between the actual policy and the policy that would have been pursued under perfect foresight the policy forecast error (PFE):

$$PFE_t = P_{t,t+h} - P^f_{t,t+h} = p(x_{1,t+h},...x_{N,t+h}) - p(x^f_{1,t+h},...x^f_{N,t+h}) = e(e_{1,t+h},...e_{N,t+h}), \qquad (3)$$

where $e_{1,t+h},...e_{N,t+h}$, are the forecast errors associated with the individual series. Thus the PFE is composed of the individual forecast errors weighted by their importance in the policy rule.

As noted in the introduction, the standard practice has been to separately evaluate the forecast errors of each of the individual series which enter the policy decision function. It is common to compute some measure of forecast accuracy (such as root mean squared forecast error-RMSFE, for example) to use in a comparison with other forecasts. These standards of comparison could be either a naïve benchmark or forecasts generated by another set of forecasters. But, in a policy decision environment, such comparisons are difficult to interpret for at least three reasons. First, it is often difficult to make general statements about relative forecast ability because one forecaster may have a lower RMSFE for a subset of the forecasted variables and a higher RMSFE for the other forecasted variables. Second, single variable comparisons ignore the ultimate objective of the forecasts which serve as inputs to the policy decision. By examining the errors of each variable separately, we are implicitly assigning equal weights (in the policy decision function) to all variables being forecasted. Finally, single variable comparisons ignore the possibility that errors in two or more of the series being considered may, partly or completely, offset each other within the policy decision function. It is for these reasons that we undertake this multivariable evaluation.

II. Data

The data are the Federal Reserve's Greenbook forecasts for each quarter from 1965.4 through 2001.4.[1] The projections used in this analysis are the growth rate of real output (GNP from 1965 to 1991 and GDP from 1992 on)[2] and the inflation rate (based on the implicit price deflator through the first quarter of 1996, then the chain-weighted price index from 1996.2 on). We only analyze the projections that are made for the current quarter and one quarter ahead.

We focus on short horizons for the following reason. The Fed 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 varying

[1] 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.
[2] The last forecast in the fourth quarter of 1991 was the first forecast of GDP.

path for monetary policy, the current and one-quarter-ahead forecasts will be unaffected by those assumptions.

In the Greenbook there are at least two forecasts per quarter, each made in various months of the quarter. Consequently the forecasts made for the current and next quarter have leads of zero to five months to the end of the relevant quarter. These leads are the horizons of the forecasts. Thus, horizons 0 through 2 are forecasts for the current quarter; with horizon 0 being the forecast made in the last month of the quarter but before all the data that refer to that quarter have been released. Similarly, horizons 3 through 5 are forecasts made for one quarter ahead (so the horizon 0 and horizon 3 forecasts were made at the same time). Because forecasts were not made at all horizons in every quarter, the number of observations differs between horizons.

The actual figures were the data published approximately 90 days after the end of the quarter to which they refer.[3] Use of the real time data avoids definitional and classification changes.

III. The Fed's Policy Rule

In order to show the effects of the Federal Reserve's forecasts of real output growth and the inflation rate, it is necessary to first establish an appropriate policy rule. We assume that the Fed implicitly follows the Taylor rule (Taylor, 1993) as the monetary policy rule. Although Taylor (1993) originally proposed his rule as an empirical description of past Fed policy actions, Woodford (2001a, 2001b) has shown that the Taylor rule can also be justified based on a firm theoretical foundation. The Taylor rule can be derived from an optimizing macro model in which the monetary authority attempts to minimize a quadratic loss function of the form:

[3] These are thus, in modern terminology, the "final" releases by the Bureau of Economic Analysis. Our results are robust to using instead the data published 45-60 days after the end of the quarter.

$L_t = \pi^2_t + \lambda(y_t - y_t^e)^2$, where π is inflation, y is output, λ is a positive constant and y_t^e is the efficient level of output (Woodford, 2001a). Woodford (2001b) showed that this loss function approximates the loss function in a representative household's optimization problem. Moreover, under the assumptions that the dynamic behavior of the economy is linear and the objective function of the Fed is quadratic, the optimal policy is certainty equivalent (Mishkin, 2008). The application of certainty equivalence requires interest rate smoothing (Giannoni and Woodford, 2002). We will address the impact of smoothing in Section VI.

If the Fed were to follow this policy rule in real time, it would require "now-casts," i.e. forecasts for the current quarter, for both inflation and output, due to the lag in data availability. Further research has also suggested that the Taylor rule should be forward-looking, employing expected future values of both output and inflation (see Clarida et al, 2000).

According to the forward-looking Taylor rule, the Fed, sets a target federal funds rate, i_t^{Tf}, based on equation (4), where, as above, the superscript "f" denotes that the target is based on forecasted variables.[4] The Fed's policy decision $(P^f_{t,t+h})$ is written as:

$$P^f_{t,t+h} = i_t^{Tf} = r^* + \pi^f_{t+h} + 0.5(\pi^f_{t+h} - \pi^*) + 0.5(y^f_{t+h} - y^*)\,, \tag{4}$$

where r* is the equilibrium real interest rate, π^* is the Fed's implicit inflation rate target, and y* is the potential output growth rate.[5] The Fed forecasts both inflation, π^f_{t+h}, and output growth,

[4] Following Orphanides (2001), we assume that the Fed uses the Greenbook forecasts in their decision rule. The members of the FOMC also make their own forecasts, but have access to the staff forecasts of the Greenbook when doing so. For an evaluation of those forecasts, see Romer and Romer (2008).

[5] While the output gap is typically used in the Taylor rule, the growth rate is typically used in forecast evaluation. The growth rate of the actuals is approximately $\ln(Y_t) - \ln(Y_{t-1})$, whereas the growth rate of the forecasts is approximately $\ln(Y^f_t) - \ln(Y_{t-1})$. Thus, when we subtract one from the other for the policy forecast error, we have $\ln(Y_t) - \ln(Y^f_t)$. Approximating the output gaps in the same manner, we have $\ln(Y_t) - \ln(Y^*)$ and $\ln(Y^f_t) - \ln(Y^*)$, so again we have $\ln(Y_t) - \ln(Y^f_t)$. It is this result that permits us to use the growth rate in order to construct the PFEs. This analysis does assume, however, that potential output, Y*, is known rather than a forecast. This assumption is based on the lack of forecasts for this variable in the Greenbook. For a discussion of the role of real time output gap estimates and the Taylor rule, see Orphanides (2001).

y^f_{t+h}, *h* periods ahead. The robustness of our choice of the original Taylor (1993) weights of 0.5 on both the inflation and output gaps will be explored in Section VI.

The actual outcome in period *t+h*, however, may differ from the Fed's forecasts. Therefore, if the members of the FOMC had known the actual values for π_{t+h} and y_{t+h} (i.e. if they had perfect forecasts or perfect foresight), they would have chosen a (potentially different) federal funds rate. Consequently, their policy decision under perfect foresight $(P_{t,t+h})$ would have been:

$$P_{t,t+h} = i^T_t = r^* + \pi^A_{t+h} + 0.5(\pi^A_{t+h} - \pi^*) + 0.5(y^A_{t+h} - y^*)\,, \tag{5}$$

where π^A_{t+h} and y^A_{t+h} represent the actual realizations of π_{t+h} and y_{t+h}. The difference between i^{Tf} and i^T measures the difference in the Fed funds rate that occurs because of inaccurate forecasts of output growth and inflation and thus represents the Federal Reserve's policy forecast error, PFE_t:

$$PFE_t = i^T_t - i^{Tf}_t = 1.5\left(\pi^A_{t+h} - \pi^f_{t+h}\right) + 0.5\left(y^A_{t+h} - y^f_{t+h}\right). \tag{6}$$

The differences, $\left(\pi^A_{t+h} - \pi^f_{t+h}\right)$ and $\left(y^A_{t+h} - y^f_{t+h}\right)$, are the Fed's forecast errors for the inflation rate and real output growth respectively. Given the PFEs, the evaluation procedures are similar to those used in judging individual forecast errors.

Working with the PFE of the Fed allows us to focus on the impact of forecast errors on the policy rule without explicitly estimating the policy rule.[6] This means that we do not use the interest rate, the equilibrium real interest rate, the Fed's implicit inflation rate target, or potential output in our analysis. Considerable research has gone into estimating these variables (e.g. Clark and Kozicki, 2005, for the equilibrium real interest rate and Leigh, 2008, for the Fed's implicit

[6] For a discussion of the impact of using projections on the estimates of the Taylor rule, see Orphanides and Wieland (2008).

inflation rate target). While these variables may be time varying, as long as they are known to the policymaker each quarter, they drop out of our expression.

IV. Evaluation Procedure

We calculate the mean absolute policy forecast errors (MAPFE) and root mean squared policy forecast errors (RMSPFE) for the full sample as well as for the sub-sample where we have alternative forecasts available for comparison. We test for bias and compare these errors with forecasts obtained from two standards of comparison: a naïve model and the Survey of Professional Forecasters (SPF).

A. Testing for Bias

Holden and Peel (1990) showed that a necessary condition for bias is that the mean error be significantly different from zero. We test for bias in the traditional way by regressing the PFE on a constant and use a t-test that the constant is zero. This statistic will be calculated separately for each forecast horizon for the entire period as well as for alternative subsamples detailed below.

B. Standards of Comparison

In order to evaluate the policy losses attributable to the forecasts, we develop two standards of comparison: the PFEs that would have occurred if (1) naïve forecasts had been used to set the policy or (2) if the median forecasts from the Survey of Professional Forecasters had been used. In each case, the comparison uses the same weights and has the same number of observations as were used in constructing the loss functions associated with the Fed forecasts that we are evaluating.

The naïve (same-change) forecast assumes that the same growth rate of output and the same growth rate in the price level (the inflation rate) as were observed in the previous quarter will occur in the future period(s). These forecasts are the real-time data, published in the Greenbook, on GDP growth and inflation rate for the quarter before the Greenbook forecasts were made. This assures that we have the information available to the forecasters at the time the predictions were made.

We also compare the Greenbook forecasts to the median forecasts from the Survey of Professional Forecasters (SPF). These data come from the Federal Reserve Bank of Philadelphia (for a description of this survey, see Croushore, 1993). The individuals who respond to the SPF report their forecasts of real and nominal output (plus other variables not considered here). We use the median forecasts from this survey to construct forecasts comparable to our Greenbook data. Since the SPF forecasts are only reported once per quarter (in the middle of the second month of each quarter), we only have data comparable to horizons one and four. The SPF forecasts are available beginning with the fourth quarter of 1968.

We use the Diebold-Mariano (1995) test to determine whether the Fed's policy forecast errors are significantly different from those generated by each of the two alternative forecasts. We employ the modification recommended by Harvey et al. (1997) which results in an improvement in the behavior of the test statistic for moderately-sized samples:

$$S_1^* = S_1\left(\frac{T+1-2(h+1)+h(h+1)/T}{T}\right)^{-\frac{1}{2}}, \quad S_1 = \frac{\bar{d}}{[\hat{V}(\bar{d})]^{1/2}} \tag{7}$$

where h is the horizon, $\bar{d}$ is the mean absolute difference of the prediction errors, $\hat{V}(\bar{d})$ is the estimated variance, S_1 is the original DM statistic, and S_1* is the modified DM statistic. The modified Diebold-Mariano (DM) test statistic is estimated with Newey-West (1987) corrected standard errors that allow for heteroskedastic autocorrelated errors.

V. Results

Figures 1 and 2 show the policy forecast errors for the zero month horizon and the three month horizon respectively (other horizons are similar). The y-axis is expressed in percentage points and can be interpreted as the amount by which the federal funds rate would have differed by using forecasts in the policy rule instead of the realized data. A positive policy forecast error suggests that the federal funds rate would have been set higher had the actual data rather than the forecasts been used. The mean absolute policy forecast error (MAPFE) based on Fed forecasts at the zero month horizon is 101 basis points for the entire sample 1965:4-2002:4 (Table 1). In general, Fed policy based on the Taylor rule was at least a full percentage point away from the intended target because of errors in forecasting growth and inflation at the zero horizon.

Table 1 also shows the root mean squared policy forecast errors (RMSPFE) for the samples and horizons we analyze in our subsequent tests. As expected, both the MAPFE and RMSPFE in general increase with the length of the forecast horizon and the MAPFE and RMSPFE are both smaller than those of the naïve and SPF forecasts in all cases. This result is consistent with previous findings from evaluations of a single variable. Comparing the results in Panel B and Panel C do suggest a small amount of improvement in terms of smaller MAPFE and RMSPFE in the more recent period. The similarity of these results with the full sample results, however, suggests that our findings for the full-sample are not dependent upon the possibility that the Fed did not follow the Taylor Rule in the earlier part of the sample.[7] Furthermore, the improvement is also present for the naïve forecasts.

Table 2 presents the results from our tests for bias. Panel A shows that the null of no bias (zero mean) in the Fed's PFE is not rejected at the 5% for any horizon in the entire sample or for the Volcker-Greenspan subsample. For the Pre-Volcker sample, however, the null of no

[7] These subsamples are representative of several alternative subsamples we explored.

bias is rejected for the 4 month and 5 month horizons in favor of a positive bias. A positive bias suggests that the Taylor rule would have implied a higher target federal funds rate had the actual data rather than the forecasts been used. Thus, in terms of bias the Fed's forecasts have improved in the recent period.

In addition to testing for bias we compared the policy forecast error based on Fed forecasts with the policy forecast errors based on our two alternative sets of forecasts: the naïve forecast and the median SPF forecast. We used the modified Diebold-Mariano test of equal predictive ability (Table 3) to determine whether or not the Fed's PFEs are statistically significantly different from the alternative forecasts.

The average gap between the alternative forecasts and the Fed's forecasts are positive in all cases, indicating that the Fed's PFEs are smaller than those that would have been obtained if the alternative forecasts had been used to set policy. We further find that the gap between the absolute value of the naïve policy forecast error and the Fed's policy forecast error is statistically significant at the 5% level for all horizons for the full sample and for both sub-samples, except for horizon 5 in the pre-Volcker subsample. Similarly, in both the full sample and the pre-Volcker sample, the Fed's PFEs are significantly smaller than those generated by using the SPF forecasts. The results in Panel C of Table 3 show, however, that for the Volcker-Greenspan sample, the Fed's policy forecast errors are <u>not</u> significantly smaller than the policy forecast errors generated by the SPF at either horizon for which the SPF forecasts are available.[8]

VI. Robustness

In our analysis we have chosen to focus on the policy rule as initially formulated in Taylor (1993). Since then there has been, however, a long literature estimating the policy rule

[8] Similar results were found for inflation alone by Gamber and Smith (2008).

actually used by the Fed (e.g. Clarida et al, 2000; Orphanides, 2001). Two key issues are the role of the policy rule weights and the role of interest rate smoothing. Below we explore the impact of these two issues on our analysis.

VI.a Robustness to the Policy Rule Weights

One focus of this literature, and the part relevant to our analysis, is the weights used on inflation and output. To address the possibility that the Fed uses weights other than the Taylor weights, we performed a simple exercise. While the Fed may choose their weights based on many criteria, we focus on the weights which show the minimum impact of the forecast errors on the policy instrument. We thus determined the weights that minimize the RMSPFE of the Fed's forecasts (similar results were found when we instead minimized the MAPFE). While we calculated different weights for each horizon, the weights were constant for the full sample. These weights are listed in the first two columns of Table 4. These weights are surprisingly similar to the Taylor weights.

Then, following Clarida et al (2000), we divided the sample into two periods and allowed for weights that might differ over the two periods. The first period was for the pre-Volcker Fed, from 1965.4 through 1979.2, and the other period was for the Volcker-Greenspan Fed from 1979.3-20024. These weights are in the last columns of Table 4[9] These weights do not necessarily imply anything about actual Fed behavior, but they allow us to find the minimum MAPFE and RMSPFE for the policy if those weights had been used. Any other weights used by the Fed would increase the impact of forecast errors on the policy instrument, so these results give us a lower bound. The minimum RMSPFE and MAPFE are presented in Table 5. These results are very similar to what we found using the Taylor weights. Based on these tables our

[9] We also estimated the PFEs using the weights found by Clarida et al (2000) in their analysis and found that our results were robust to that choice of weights as well.

conclusion still holds that the impact of forecast errors alone on the monetary policy rule is approximately 100 basis points.

VI.b Robustness to Interest Rate Smoothing

According to Giannoni and Woodford (2002), the application of certainty equivalence requires interest rate smoothing. Therefore, suppose we modify the policy rules as follows to allow for interest rate smoothing:

$$P_{t,t+h}^{f} = i_t^{Tf} = \rho i_{t-1} + (1-\rho)\left[r^* + \pi_{t+h}^{f} + 0.5(\pi_{t+h}^{f} - \pi^*) + 0.5(y_{t+h}^{f} - y^*)\right] \text{ and} \qquad (4a)$$

$$P_{t,t+h} = i_t^{T} = \rho i_{t-1} + (1-\rho)\left[r^* + \pi_{t+h}^{A} + 0.5(\pi_{t+h}^{A} - \pi^*) + 0.5(y_{t+h}^{A} - y^*)\right], \qquad (5a)$$

where $0 < \rho < 1$. Estimates of the smoothing parameter, ρ, have varied widely in the literature. The policy forecast error in this case would be a simple proportion of the policy forecast error without smoothing. Thus, the worse the forecasts are expected to be, the greater the benefit of smoothing.

VII. Conclusions

In this paper we developed a simple methodology in order to evaluate the impact of forecast errors on the Fed's monetary policy as characterized by the Taylor rule. We find that the Fed's policy forecast error is in general unbiased and significantly smaller than the errors that would have resulted from naïve forecasts but not always from the SPF predictions. Nevertheless, the mean absolute policy forecast error of the Fed forecasts is approximately 100 basis points.

Moreover, this methodology can be applied in the context of any policy decision rule. This methodology may be useful for policy decisions in both the public and private sectors. Some policy applications include evaluating future budget plans for government programs, health policies, and environmental policies. For private decision makers, we could evaluate inventory decisions which depend on forecasted sales, costs, and real interest rates, using a decision rule such as the one discussed in Maccini et al. (2004).

References

Clarida, R., J. Gali and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *The Quarterly Journal of Economics* 115(1): 147-180.

Clark, T. E. and S. Kozicki, (2005). "Estimating Equilibrium Real Interest Rates in Real Time," *North American Journal of Economics and Finance* 16(3), pp. 395-413.

Clements, M. P. (2004). "Evaluating The Bank Of England Density Forecasts Of Inflation," *The Economic Journal*, 114 (October), 855–877.

Clements, M. P. and D. F. Hendry (1993). "On the Limitations of Comparing Mean Square Forecast Errors" and subsequent comments, *Journal of Forecasting,* 12(8), 617-676.

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(1), 121-136.

Croushore, D. (1993). "Introducing: The Survey of Professional Forecasters." *Federal Reserve Bank of Philadelphia Business Review,* November/December, 3-13.

Diebold, F.X. and R. Mariano (1995). "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-265

Eisenbeis, R., D. Waggoner, and T. Zha (2002). "Evaluating Wall Street Journal Survey Forecasters: A Multivariate Approach," *Business Economics*, July, 11-21.

Gamber, E. N. and J. K. Smith (2008). "Are the Fed's Inflation Forecasts Still Superior to the Private Sector's?" Research Program on Forecasting Working Paper no. 2007-002.

Giannoni, M. P. and M. Woodford (2002). "Optimal Interest Rate Rules: II. Applications," NBER Working Paper No. 9420.

Granger, C. W. J. and M. H. Pesaran (2000a). "A decision-based approach to forecast evaluation", in (W.S. Chan, W. K. Li, and H. Tong, eds). *Statistics and Finance: An Interface*, London: Imperial College Press.

Granger, C. W. J. and M. H. Pesaran (2000b). "Economic and statistical measures of forecast accuracy," *Journal of Forecasting*, vol. 19, pp. 537–60.

Harvey, D., S. Leybourne and P. Newbold (1997). "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 281-91.

Holden, K. and D. Peel (1990). “On Testing for Unbiasedness and Efficiency of Forecasts,” *The Manchester School*, pp. 121-127.

Hymans, Saul H. (1968) “Simultaneous Confidence Intervals in Econometric Forecasting,” *Econometrica*, 36(1), pp. 18-30.

Joutz, F. and H.O. Stekler, (2000). “An Evaluation of the Predictions of the Federal Reserve,” *International Journal of Forecasting*, 16, 17-38.

Komunjer, I. and M. T. Owyang (2007). “Multivariate Forecast Evaluation and Rationality Testing,” Federal Reserve Bank of St. Louis Working Paper No. 2007-047.

Leigh, D. (2008). “Estimating the Federal Reserve's implicit inflation target: A state space approach,” *Journal of Economic Dynamics and Control*, 32(6), pp. 2013-2030

Maccini, L. J., B. J. Moore, and H. Schaller (2004). “The Interest Rate, Learning, and Inventory Investment,” *The American Economic Review*, 94(5), 1303-1327.

Mishkin, F. S. (2008). “Monetary Policy Flexibility, Risk Management, and Financial Disruptions,” Speech at the Federal Reserve Bank of New York, New York, New York, January 11, 2008.

Newey, W. and K. West (1987). “A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.

Orphanides, A. (2001). “Monetary Policy Rules Based on Real-Time Data,” *The American Economic Review*. 91(4), 964-985.

Orphanides, A., and V. Weiland (2008). “Economic Projections and Rules of Thumb for Monetary Policy,” *Federal Reserve Bank of St. Louis Review*. 90(4), 307-24.

Pesaran, M. H. and S. Skouras (2002). “Decision-based methods for forecast evaluation,” in (M.P. Clements, and D.F. Hendry, eds), *A Companion to Economic Forecasting*, pp. 241–67, Oxford: Blackwell.

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

Romer, C. D. and D. H. Romer. (2008). "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?", *American Economic Review* 98(2), pp. 230-35.

Sims, C. A., (2002). “The Role of Models and Probabilities in the Monetary Policy Process.”

Brookings Papers on Economic Activity 2, 1-40.

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

Stekler, H.O. (1994). “Are economic forecasts valuable?” *Journal of Forecasting* 13, 495-505.

Taylor, J.B. (1993). “Discretion versus Policy Rules in Practice,” *Carnegie-Rochester Conference Series on Public Policy,* 39,195-214.

Woodford, M. (2001a). “The Taylor Rule and Optimal Monetary Policy,” *American Economic Review*, 91 (2), 232-237.

Woodford, M. (2001b). ”Inflation Stabilization and Welfare," National Bureau of Economic Research working paper 8071.

Figure 1: Policy Forecast Error in Percentage Points, Horizon 0
(NBER Recessions in Gray)

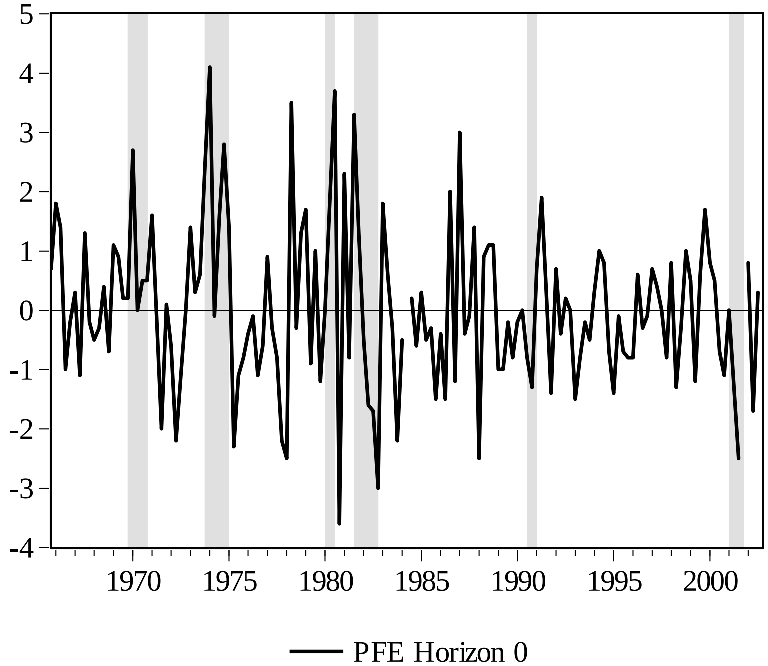

Figure 2: Policy Forecast Error in Percentage Points, Horizon 3
(NBER Recessions in Gray)

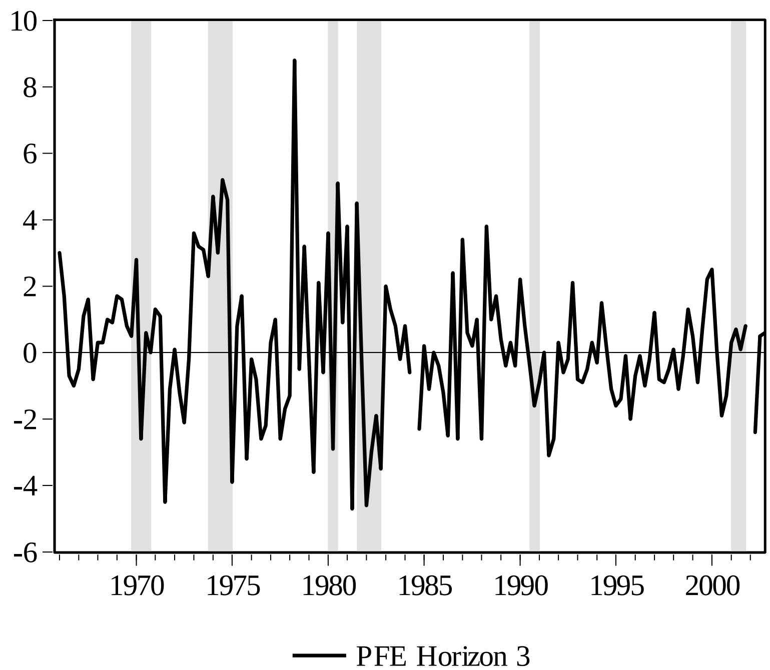

Table 1: Mean Absolute Policy Forecast Error (MAPFE) and Root Mean Squared Policy Forecast Error (RMSPFE)

Panel A: Full Sample (1965.4 – 2002.4)[10]

horizon	$MAPFE_{Fed}$	$MAPFE_{naïve}$	$MAPFE_{SPF}$	$RMSPFE_{Fed}$	$RMSPFE_{naïve}$	$RMSPFE_{SPF}$	N
0	1.01	1.95	---	1.31	2.77	---	159
1	1.25	2.08	1.54	1.67	2.83	1.99	110[11]
2	1.54	1.99	---	1.93	2.68	---	94
3	1.56	2.31	---	2.08	3.03	---	159
4	1.83	2.41	2.21	2.34	3.15	2.77	105[12]
5	1.80	2.38	---	2.43	3.13	---	88

Panel B: Pre-Volcker Sample (1965.4 – 1979.2)[10]

horizon	$MAPFE_{Fed}$	$MAPFE_{naïve}$	$MAPFE_{SPF}$	$RMSPFE_{Fed}$	$RMSPFE_{naïve}$	$RMSPFE_{SPF}$	N
0	1.03	2.06	---	1.35	2.98	---	68
1	1.38	2.26	1.95	1.73	3.18	2.36	54[13]
2	1.71	2.21	---	2.15	2.97	---	51
3	1.84	2.70	---	2.38	3.46	---	67
4	2.09	2.74	2.78	2.69	3.56	3.36	48[14]
5	2.38	2.81	---	3.01	3.58	---	44

Panel C: Volcker-Greenspan Sample Panel C: 1979.3 – 2002.4

horizon	$MAPFE_{Fed}$	$MAPFE_{naïve}$	$MAPFE_{SPF}$	$RMSPFE_{Fed}$	$RMSPFE_{naïve}$	$RMSPFE_{SPF}$	N
0	1.00	1.87	---	1.29	2.60	---	91
1	1.12	1.92	1.23	1.62	2.45	1.66	56
2	1.35	1.73	---	1.65	2.29	---	43
3	1.36	2.02	---	1.82	2.67	---	92
4	1.61	2.14	1.81	2.01	2.75	2.26	57
5	1.23	1.96	---	1.65	2.61	---	44

[10] SPF sample starts with forecasts made in 1968.4.
[11] N = 98 for SPF.
[12] N = 97 for SPF.
[13] N = 42 for SPF.
[14] N = 40 for SPF.

Table 2: Tests for Unbiasedness of Fed's Policy Forecast Errors, Alternative Sample Periods

Panel A: Full Sample (1965.4 – 2002.4)[15]

Horizon (months)	0	1	2	3	4	5
Estimated mean	0.05	0.15	0.12	0.20	0.32	0.49
Std. Error	0.10	0.16	0.20	0.16	0.23	0.25
P-value (H_0: zero mean)	0.61	0.34	0.55	0.23	0.16	0.06
Number of Observations	159	110	94	159	105	88

Panel B: Pre-Volcker Sample (1965.4 – 1979.2)[15]

Horizon (months)	0	1	2	3	4	5
Estimated mean	0.27	0.30	0.35	0.58	0.78	1.10
Std. Error	0.16	0.23	0.30	0.28	0.38	0.43
P-value (H_0: zero mean)	0.10	0.21	0.25	0.05	0.04	0.01
Number of Observations	68	54	51	67	48	44

Panel C: Volcker-Greenspan Sample Panel C: 1979.3 – 2002.4

Horizon (months)	0	1	2	3	4	5
Estimated mean	-0.11	0.01	-0.15	-0.08	-0.06	-0.13
Std. Error	0.14	0.22	0.25	0.19	0.27	0.25
P-value (H_0: zero mean)	0.43	0.95	0.55	0.69	0.83	0.61
Number of Observations	91	56	43	92	57	44

[15] SPF sample starts with forecasts made in 1968.4.

Table 3: Modified Diebold Mariano Test of Equal Predictive Ability (H_0: Gap=0)

Panel A: Full Sample (1965.4 – 2002.4)[16]

Horizon	Average Gap |Naïve PFE|-|Fed PFE| (Modified DM statistic)	Average Gap |SPF PFE|-|Fed PFE| (Modified DM statistic)	N
0	.94** (6.17)	---	159
1	.83** (4.96)	.22* (2.09)	110 (98 for SPF)
2	.45** (2.51)	---	94
3	.75** (4.31)	---	159
4	.58** (2.98)	.32** (3.13)	105 (97 for SPF)
5	.57** (2.54)	---	88

** indicates significant at the 1% level, * indicates significant at the 5% level.

Panel B: Pre-Volcker Sample (1965.4 – 1979.2)[16]

Horizon	Average Gap |Naïve PFE|-|Fed PFE| (Modified DM statistic)	Average Gap |SPF PFE|-|Fed PFE| (Modified DM statistic)	N
0	1.03** (4.05)	---	68
1	0.87** (2.86)	.40* (2.22)	54 (42 for SPF)
2	0.50* (1.87)	---	51
3	0.87** (2.57)	---	67
4	0.65* (1.69)	.50** (3.09)	48 (40 for SPF)
5	0.42 (1.01)	---	44

** indicates significant at the 1% level, * indicates significant at the 5% level.

[16] SPF sample starts with forecasts made in 1968.4.

Table 3: Modified Diebold Mariano Test of Equal Predictive Ability (continued)

Panel C: Volcker-Greenspan Sample Panel C: 1979.3 – 2002.4

Horizon	Average Gap |Naïve PFE|-|Fed PFE| (Modified DM statistic)	Average Gap |SPF PFE|-|Fed PFE| (Modified DM statistic)	N
0	0.87** (4.71)	---	91
1	0.79** (5.46)	.10 (0.92)	56
2	0.39* (1.73)	---	43
3	0.66** (4.10)	---	92
4	0.52** (3.18)	.19 (1.54)	57
5	0.72** (4.26)	---	44

** indicates significant at the 1% level, * indicates significant at the 5% level.

Table 4: Weights that Minimize RMSPFE for the Fed Forecasts

	Full Sample Weights		**Split-Sample Weights**			
Horizon	**GDP Weight**	**Inflation Weight**	**GDP Weight**		**Inflation Weight**	
			1965.4 – 1979.2	**1979.3 – 2002.4**	**1965.4 – 1979.2**	**1979.3 – 2002.4**
0	0.69	1.31	0.77	0.56	1.23	1.44
1	0.57	1.43	0.69	0.37	1.31	1.63
2	0.55	1.45	0.65	0.34	1.35	1.66
3	0.52	1.48	0.69	0.31	1.31	1.69
4	0.49	1.51	0.65	0.21	1.35	1.79
5	0.64	1.36	0.74	0.39	1.26	1.61

Table 5: RMSPFE and MAPFE for Different Weights

horizon	**Taylor $RMSPFE_{Fed}$**	**Full Sample Min Weights $RMSPFE_{Fed}$**	**Split Sample Min Weights $RMSPFE_{Fed}$**	**Taylor $MAPFE_{Fed}$**	**Full Sample Min Weights $MAPFE_{Fed}$**	**Split Sample Min Weights $MAPFE_{Fed}$**
0	1.31	1.24	1.23	1.01	0.97	0.96
1	1.67	1.66	1.60	1.25	1.23	1.19
2	1.93	1.93	1.85	1.54	1.56	1.46
3	2.08	2.07	1.97	1.56	1.56	1.50
4	2.34	2.34	2.18	1.83	1.83	1.67
5	2.43	2.34	2.24	1.80	1.79	1.66