Has models' forecasting performance for US output growth and inflation changed over time, and when?

Barbara Rossi
Duke University

Tatevik Sekhposyan
University of North Carolina, Chapel Hill

July 2008

ERID Working Paper Number 12

Economic Research Initiatives at Duke
WORKING PAPERS SERIES

Has models' forecasting performance for US output growth and inflation changed over time, and when?

Barbara Rossi and Tatevik Sekhposyan

Duke University *UNC - Chapel Hill*

First draft: July 2007. This revision: August 2008.

Abstract

We evaluate various models' relative performance in forecasting future US output growth and inflation on a monthly basis. Our approach takes into account the possibility that the models' relative performance can be varying over time. We show that the models' relative performance has, in fact, changed dramatically over time, both for revised and real-time data, and investigate possible factors that might explain such changes. In addition, this paper establishes two empirical stylized facts. Namely, most predictors for output growth lost their predictive ability in the mid-1970s, and became essentially useless in the last two decades. When forecasting inflation, instead, fewer predictors are significant (among which, notably, capacity utilization and unemployment), and their predictive ability significantly worsened around the time of the Great Moderation.

Keywords: Output Forecasts, Inflation Forecasts, Model Selection, Structural Change, Forecast Evaluation, Real-time data.

Acknowledgments: We thank seminar participants at the 2007 Midwest Econometrics Group Conference, the Macro-Money seminar at UNC-Chapel Hill and the EMSG at Duke University for comments.

J.E.L. Codes: C22, C52, C53

1 Introduction

This paper investigates whether the relative performance of competing models for forecasting US output growth and inflation has changed over time. While there is widespread empirical evidence on the existence of parameter instability in forecasting GDP growth and inflation (as documented, for example, by Stock and Watson, 2003, and Clark and McCracken, 2005), there is little work on formally testing whether the models' relative performance has actually changed over time. D'Agostino, Giannone, and Surico (2006) undertake a forecast comparison of various models and note a sizeable decline in the relative predictive accuracy of popular forecasting methods based on large data sets of macroeconomic indicators; they associate this decline with the fall in the volatility of most macroeconomic time series (the "Great Moderation"). Interestingly, they also note that the full sample predictability of US macroeconomic series comes from the years before 1985, that constitute a large portion of the full sample. However, their analysis is limited to two sub-samples, and they do not formally test for a change in the relative performance (that is, the difference between the two sub-periods that they document may be just sampling variability rather than a significant change). To fill this gap in the literature, this paper presents a comprehensive analysis of forecast comparisons of various representative models for predicting future output growth and inflation growth and assesses whether their performance has changed over time. Our analysis has the advantage of precisely estimating the time of the reversal in the predictive ability, which provides valuable information for uncovering possible economic causes of the reversals.

In order to assess how the models' relative forecasting performance has changed over time, this paper goes beyond the seminal works of Diebold and Mariano (1995), West (1996), Clark and McCracken (2001), and Clark and West (2006).[1] In fact, these papers only compare the relative forecasting performance of the competing models on average over the forecasting sample. Giacomini and Rossi (2008) notice that this procedure, by focusing on the average performance, involves a loss of information. In particular, it may hide important reversals in the models' relative performance over time. Giacomini and Rossi (2008) propose a Fluctuation test for assessing equal predictive ability that takes into account the possibility that the relative performance might have changed over time, as well as a One-time Reversal procedure to estimate the time of the reversal. We will apply this technique to empirically investigate whether the relative performance of competing models for forecasting US indus-

[1] See also Inoue and Kilian (2006).

trial production growth and consumer price inflation has changed over time. We focus on the same models considered in Stock and Watson (2003) and Clark and McCracken (2005), but use monthly data for industrial production rather than quarterly data for GDP, as well as monthly data for inflation. This ensures a big enough sample for pseudo out-of-sample forecast model comparisons over time. Using both fully revised as well as real-time data, we find substantial reversals in the relative forecasting performance. This analysis, however, is still silent about the economic reasons why such reversals have happened. However, our procedure can estimate the time of the reversal in the relative performance, which allows us to relate such changes to the economic events happening simultaneously.

Our main empirical findings are as follows. First of all, we document that, overall, there is empirical evidence that the economic predictors have forecasting ability in the early part of the sample, but the predictive ability disappears in the later part of the sample. This happens notwithstanding the general result that some explanatory variables help forecasting output growth and inflation beyond a simple autoregression over the full sample. We note that the results that we present in this paper are very robust, and could be made even more striking by a more conservative choice of the bandwidth parameter for the estimate of the variance, or by using a Fluctuation test based on the Clark and West (2006) test statistic. Second, we find empirical evidence in favor of a wide range of instabilities, with sharp reversals in the relative performance of the various models. In particular, when forecasting output growth, we find that interest rates and the spread were useful predictors in the mid-1970s, but their performance worsened at the beginning of the 1980s. Similar results hold for money growth (M2), CPI inflation, stock prices, and the unemployment gap. The results are similar when forecasting inflation, with two notable exceptions. On the one hand, the empirical evidence of models' predictive ability for inflation is much weaker than that of output growth over the full sample, and more evidence of predictive ability can be uncovered only by allowing for changes in the relative performance, unlike the case of output growth. On the other hand, the evidence of predictive ability of most variables breaks down around 1984, which the literature agrees to be the beginning of the Great Moderation. This includes the models with unemployment and other output measures, thus implying that the predictive power of the Phillips curve disappeared around the time of the Great Moderation. Third, we document the robustness of our results to the use of Real-Time data (Croushore and Stark, 2001). Stark and Croushore (2002) and Croushore (2006) show that data revisions matter for forecasting, though the degree to which they matter depends on the case at hand. In particular, they note that in the first half of the 1970s, real-time data forecasts of output growth were significantly better than forecasts based on latest-available data; in other short samples the real-time forecasts were significantly worse than those using latest-available data. Since our analysis allows us to formally analyze changes in the models' relative performance over time, it will shed light on this issue. We show that for some series, such as capacity utilization and M2, the evidence in favor of predictive ability in the early part of the sample is slightly weaker when using real-time as opposed to fully revised data, whereas the opposite happens for other series (such as unemployment). Overall, however, our main qualitative conclusions are strikingly robust to the use of real-time data.

The rest of the paper is organized as follows. Section 2 describes the data and the forecasting models considered in the empirical analysis. Section 3 and 4 present and discuss the empirical results: Section 3 focuses on predicting output growth using both fully revised and data available in real-time, whereas Section 4 focuses on forecasting inflation. Section 5 concludes.

2 A description of the models and data

This paper focuses on the multi-step pseudo out-of-sample forecasting performance of a variety of models for predicting future US output growth and inflation. Our measure of output is the industrial production index (IP), whose data are available on a monthly basis, whereas our measure of inflation is the second difference of the Consumer Price Index (CPI).[2] Following Stock and Watson (2003), the models with explanatory variables (which we will refer to as "economic models") are:

$$y_{t+h} = \beta_0 + \beta_1(L)x_t + \beta_2(L)y_t + \epsilon_{t+h} \tag{1}$$

where y_{t+h} is either the h period ahead output growth at time t defined by $y_{t+h} = 1200\ln(IP_{t+h}/IP_t)/h$ or the h period ahead inflation at time t defined by $y_{t+h} = 1200\ln(CPI_{t+h}/CPI_t)/h - 1200\ln(CPI_t/CPI_{t-1})$, x_t is a possible explanatory variable, y_t is either the period t output growth, that is $y_t = 1200\ln(IP_t/IP_{t-1})$, or the period t change in inflation, that is $y_t = 1200\ln(CPI_t/CPI_{t-1})$, and ϵ_{t+h} is an error term. $\beta_1(L)$ and $\beta_2(L)$ are the lag polynomials, such that $\beta_1(L)x_t = \sum_{j=1}^{p}\beta_{1j}x_{t-j+1}$, $\beta_2(L)y_t = \sum_{j=1}^{q}\beta_{2j}y_{t-j+1}$, and p and q are chosen by the BIC. We consider one year ahead output and inflation growth by setting $h = 12$ months. All models are estimated by OLS.

[2] We chose to work with the second difference of the CPI in order to impose the same I(2) constraint as in Stock and Watson (2003).

The models considered here are bivariate, and they differ in the additional explanatory variable x_t used for forecasting. We consider the Stock and Watson (2003) database when identifying the explanatory variables, omitting housing prices, gold, silver, and the real effective exchange rate, whose sample start much later than the other series, preventing a large out-of-sample size for our forecast comparisons. We will focus mainly on a few representative series, chosen depending on the relevance of the variables for policymaking and economic theory and on how dramatic the changes in the variables' predictive content have been over time, and only succinctly summarize the results for the whole Stock and Watson (2003) database. The sources and exact description of the data are provided in Table 1. The representative series that we consider are:

(i) Short-term interest rates, either the Fed Funds Rate (rovnght level) or the one-year US Treasury rate (rbnds level);

(ii) Interest rate spread (rspread level);

(iii) Unemployment gap (unemp gap) and Capacity utilization (capu level);

(iv) Inflation (cpi Δ ln);

(v) Earnings (earn Δ ln);

(vi) Money growth, either M2 (m2 Δ ln) or M3 (m3 Δ ln).

The Federal Funds Rate is of a special interest as a monetary policy instrument. The economic models that include alternative measures of interest rates (interest rates on short-term and long-term treasury securities) appear to behave similar to the Fed Funds Rate model qualitatively. It is also potentially interesting to consider the role of asset prices by investigating the predictive content of the interest rate spread, as asset prices contain forward looking expectational components. Money is relevant for certain parts of the out-of-sample period as a direct policy instrument. For the rest of the periods, it acts as an intermediate policy target theoretically containing relevant information for future output growth. As Stock and Watson (1999a) discuss, some real variables (such as capacity utilization) lead and others lag the business cycle, making them interesting for forecasting purposes. In addition, when forecasting inflation, unemployment and capacity utilization are relevant from a Phillips curve perspective.

INSERT TABLE 1 HERE

We compare the multi-step pseudo out-of-sample forecasting performance of each of the models above with that of a univariate autoregression. We will refer to the latter as the benchmark model:

$$y_{t+h} = \beta_0 + \beta_2(L)y_t + \eta_{t+h} \tag{2}$$

To capture the time variation in the relative performance, we construct rolling estimates of the relative Mean Square Forecast Errors (MSFE) using a two-sided window of 120 months; the data start in 1959:1, and the first 12-months ahead out-of-sample forecast is made for 1970:3 (we lose two observations because of taking second differences of the data). Let the pseudo out-of-sample forecast errors of models (1) and (2) be denoted, respectively, by $\hat{\epsilon}_{t+h}$ and $\hat{\eta}_{t+h}$. Our object of interest is the rescaled difference between the mean square forecast errors (rMSFE) of the "economic" model (1) and that of the univariate autoregression calculated over these rolling windows ($m = 120$):

$$rMSFE_t = \widehat{\sigma}^{-1}\frac{1}{m}\left(\sum_{j=t-m/2}^{j=t+m/2}(\hat{\epsilon}_{t+h})^2 - \sum_{j=t-m/2}^{j=t+m/2}\left(\hat{\eta}_{t+h}\right)^2\right), \tag{3}$$

where $\widehat{\sigma}$ is an estimate of the standard deviation of the relative MSFE. In order to test whether the relative forecasting performance has changed over time, we utilize the Fluctuation test proposed by Giacomini and Rossi (2008).

3 Forecasting output growth

In this section, we focus on the empirical predictive ability of macroeconomic variables for forecasting US output growth. We begin by considering detailed empirical results for the representative macroeconomic time series, namely the Federal Funds Rate, the interest rate spread, capacity utilization, the unemployment gap, CPI inflation, and the rate of money growth. We then consider a comprehensive survey of all the series in the Stock and Watson (2003) database. We conclude by analyzing the robustness of our results to using real-time data.

3.1 Detailed empirical results using representative series

First, Table 2 reports empirical evidence based on tests of equal predictive ability on average over the full pseudo out-of-sample period, starting in 1970:3 and ending in 2005:12 – except for capacity utilization, oil, and M0, for which the pseudo out-of-sample period stops some time in 2002 and 2003 (consult Table 1 for more details). The first column reports the rescaled MSFE difference calculated over the full out-of-sample period. Negative values mean that the autoregressive model has a higher MSFE than the model with additional explanatory variables. The second column reports the p-values based on the Giacomini and White (2006) test. The table shows that a number of series have predictive content. In fact, we reject the null hypothesis at 10% significance level for the Fed Fund rate, the real Fed Fund rate, the spread, stock prices, capacity utilization, the employment gap, the CPI measures, and money.

INSERT TABLE 2 HERE

When we consider the Fluctuation test, however, we uncover a different picture. Figure 1 reports the Fluctuation test for the representative series. The Fluctuation test consists of rolling estimates of the rMSFE differences *over time*. It is clear from the figure that there is striking empirical evidence of time variation in the relative performance of the economic models relative to a simple autoregression. This is consistent with Stock and Watson's (2003) finding that there is a great deal of instabilities in the ranking of the models in terms of forecast performance. Our analysis, however, gives a better sense on how the relative forecasting performance has evolved over time. Overall, there is ample evidence of reversals in the relative performance, with the economic model losing its predictive ability in the later part of the sample. While this graphical evidence is suggestive of dramatic changes in the relative performance, it is important to econometrically test whether such changes are significant. We test the null hypothesis that the relative performance of the two competing models is the same at each point in time. If this were the case, the paths of rMSFEs depicted in Figure 1 should be inside the two boundary lines also reported in the figure. It is clear that for some variables the paths are outside the bands, thus implying that the relative predictive ability of the two models has not remained the same over time.

Let us focus on each series in more detail. Interest rates such as the Fed Funds rate (labeled "rovnght") or the interest rate spread ("rspread") are considered important predictors for future output growth (see for example, Estrella, 2005, and Kozicki, 1997) although there is widespread evidence of parameter instabilities in such regressions (see Estrella, Rodrigues and Schich, 2003). The first two top panels of Figure 1 suggest that these models performed quite well in the mid- to late Seventies relative to the autoregressive model, whereas their performance has significantly worsened during the Eighties. Similary, the two middle and last panels show that the usefulness of capacity utilization ("capu"), unemployment ("unemp"), and CPI inflation to predict future output growth has worsened in the Eighties and Nineties, relative to the Seventies.

INSERT FIGURE 1 HERE

There is, however, an important difference between the various series. Capacity utilization and unemployment seem to have maintained their predictive ability much longer than the Fed Funds rate, the spread and CPI inflation.

Finally, money deserves special attention in the light of the important debate of whether money predicts future output growth (Stock and Watson (1989), Amato and Swanson (2001), and Inoue and Rossi (2005)). Figure 1 shows that money growth (M2) was a useful predictor for future output growth until the beginning of the 1980s, when its performance became statistically insignificantly different from that of an autoregression.

3.2 Comprehensive overview for all series

We performed a similar analysis to that in the previous sub-section for all the series in the Stock and Watson (2003) database, except for the shorter series mentioned before. Due to space constraints, detailed results are reported in a not-for-publication appendix (Rossi and Sekhposyan, 2008), and we only summarize them here.

Most nominal interest rates behave very similarly to the Fed Funds Rate, although their predictive ability is less significant. The pattern for the real interest rates is similar. The nominal effective exchange rate does not seem to be a good predictor of real activity anywhere in the out-of-sample period. The growth rates of stock prices (both nominal and real) do have significant predictive ability in the late 1970s, but the predictive ability disappears around the 1980s, with a pattern very similar to that of the growth rate of interest rates.

Real activity measures, such as the rate of growth of employment and unemployment, are never significant; however, the employment gap has a pattern very similar to that of the unemployment gap, discussed in the previous section. Variables in the wage and price categories are mostly never significant, although the inflation rate measured by the producer price index difference is significantly better than the autoregressive benchmark in the late 1970s. Finally, considering the money category, we find that, unlike M2, M1 and M3 are never significant, whereas M0 behaves significantly worse than the benchmark in the late 1970s.

Overall, we conclude that there are widespread significant reversals from predictive ability to lack thereof around the late 1970s, and this reversal is stronger for short/medium term interest rates, the employment gap, the producer price index inflation, stock prices, and M2.

3.3 Empirical results for forecasting output growth using real-time data

As it is well-known, using finally revised data in pseudo out-of-sample forecasting exercises has the drawback that the data used in the exercise are not really the same data that the forecasters had available at each point in time. We therefore revisit our analysis in the previous section by using real-time data for industrial production and employment provided by the Philadelphia Fed in the Real-Time Data Set for Macroeconomists (Croushore and Stark, 2001). All data are monthly and seasonally adjusted. For the realized value, we choose the vintage one period (in this case, one month) after the observation date (in monthly vintages).

Many authors, starting from Diebold and Rudebusch (1991a,b), have pointed out that results based on fully revised data are misleading, in that they spuriously find positive empirical evidence in favor of leading indicators. In addition, Stark and Croushore (2002) and Croushore (2006) document that data revisions may matter for forecasting, though how much they matter depends on the case at hand. In particular, they note that in the first half of the 1970s, forecasts of output growth based on real-time data were significantly better than forecasts based on latest-available data, but that in other short samples the real-time forecasts were significantly worse than those using latest-available data. In addition, they found that forecasts of inflation were instead superior when based on latest-available data than when using real-time data in all the sub-samples they considered. Similarly, Orphanides and Van Norden (2005) showed that in real time, out-of-sample forecasts of inflation based on measures of the output gap are not very useful, and Edge, Laubach and Williams (2007) found similar results for forecasting long-run productivity growth. Our methodology allows us to undertake a formal analysis of how the models' relative performance changed over time, and it is well suited to shed further light on this issue.

INSERT TABLE 3 HERE

In this section we focus on the same representative explanatory variables considered in the previous section: the Fed Fund rate, the spread, money (M2), the CPI, the unemployment gap, and capacity utilization. First, we report results based on the full out-of-sample tests in Table 3. Note that some of the predictors are not significant anymore (such as the Fed Funds rate, capacity utilization, CPI, M2 and employment growth). Therefore, the use of real-time data reveals that, at least over the full sample, data revisions actually matter for forecasting, and that real-time data forecasts are significantly worse than those using fully revised data.

The results for the Fluctuation test for real time industrial production data are presented in Figure 3. Overall, for most variables, our results are qualitatively unchanged, although the evidence in favor of predictive ability in the early part of the sample is somewhat weakened when using real-time as opposed to fully revised data. The only notable difference is that the unemployment gap forecasts better with real-time data. As we show in the not-for-publication Appendix, the use of real-time data for employment gap marginally improve the predictive ability of the model in the early 1970s as well.

INSERT FIGURE 2 HERE

4 Forecasting inflation

In this section we focus on forecasting US CPI inflation. We first present detailed empirical results for a few representative time series, that is long term interest rates, the interest rate spread, capacity utilization, unemployment, earnings, and the rate of growth of money (M3). Then we discuss a summary of the results for all the available economic series.

4.1 Detailed empirical results for forecasting future inflation using representative series

The predictive ability of macroeconomic variables for future inflation is much less widespread than that for future output growth. In fact, Table 4 shows that only a very few economic series have predictive content: stock prices, industrial production, the employment gap, some measures of oil prices, and some measures of money. However, there is striking evidence of changes in the relative performance of the models, and once we take that into account, we find much more compelling empirical evidence in favor of economic predictors for capacity utilization, which are perhaps the most important variable for predicting future inflation according to the Phillips curve relationship.[3] For example, Stock and Watson (1999b) found some empirical evidence in favor of the Phillips curve as a forecasting tool, and demonstrated that inflation forecasts produced by the Phillips curve generally are more accurate than forecasts based on other macroeconomic variables, including interest rates, money and

[3]Similar results hold when using employment or unemployment growth rates – see Section 4.2.

commodity prices. Indeed, we find that capacity utilization had significant predictive content in the late 1970s, but that such predictive content disappeared in the late 1980s.

INSERT TABLE 4

Interest rates have been found to be important predictors for future inflation since the works by Mishkin (1990); see also Kozicki (1997). Indeed, we find that short-term (one-year) interest rates had marginal predictive content for inflation in the late 1970s, but that such predictive ability disappeared in the later part of the sample to the point of making the model appear significantly worse. Similarly, the interest rate spread was never significantly better than the autoregressive benchmark over the pseudo out-of-sample period. Interestingly, we find that earnings were a marginally significant predictor throughout the 1980s. Finally, money (M3) had significant predictive content for a long period of time, mostly during the 1980s, then it experienced a sharp reversal towards being insignificant.

4.2 Comprehensive overview for all series and summary of the results

Overall, we find very little predictive content in both nominal and real interest rates for forecasting future inflation. For some interest rates, both real and nominal, however, there have been interesting reversals in their predictive ability during 1980s that resulted in the models becoming significantly worse than the autoregressive benchmark. We also observe interesting reversals in the predictive ability of the nominal effective exchange rate, although such reversals are never significant. The pattern in most activity measures resembles than in capacity utilization, discussed above, except for the employment and unemployment gaps, whose predictive ability is at times significantly worse than the benchmark. There is also very little significance for most wage and price measures. Other definitions of money (M1 and M0) behave similarly to M3 (reported in Figure 2 above), although the predictive ability is somehow smaller in magnitude. M2 is instead a significantly worse predictor than the benchmark throughout the out-of-sample period.[4]

We do not consider real time data for CPI because it is only available in quarterly vintages whereas we focus on monthly vintages to have a sufficiently large pseudo out-of-sample period to obtain meaningful rolling forecast error comparisons.

[4] Again, see the not-for-publication Appendix (Rossi and Sekhposyan, 2008) for detailed results.

5 When did the sharp reversals in the relative forecasting performance happen?

In this section, we analyze more carefully the timing of the sharp reversals in the relative forecasting performance that we documented in the previous sections. In fact, the visual evidence regarding the timing of the break based on Figures 1-3 refers to "smoothed" averages of the relative performance over a window of ten years, and therefore does not allow us to determine the timing of the break exactly. We can estimate the timing of the break precisely by using the "One-time Reversal" procedure in Giacomini and Rossi (2008). Tables 3 and 5 report results for the "One-time Reversal" test (labeled "One-time Reversal"), as well as a test for breaks in the relative predictive ability (labeled "Breaks"). If the latter finds empirical evidence in favor of changes in predictive ability, the table also reports the estimated time of the reversal.

INSERT TABLES 5-7

Table 5 focuses on forecasting output growth. The table shows that the timing of the break for the Fed Fund's rate, and the spread is mid-1976, for M2 is mid-1977, for CPI inflation is mid 1975. From Figure 1, the unemployment gap shows at least two big reversals in the relative performance; the "One-time Reversal" procedure, in this case, estimates the timing of the largest break, which happens to be in early 1976. Also some real interest rates show reversals at the same time. Therefore, interestingly, for all series except employment/unemployment, the most substantial reversal in relative predictive ability happened around mid-1970s. Similar results hold for real-time data for output, see Table 6, with the only exception that the reversal in the predictive ability of unemployment is now dated around 1984. A very different picture emerges when forecasting inflation. Table 7 shows that most reversals happen around 1984 rather than the late 1970s. The reversals in predictive ability happened, therefore, around the time of the Great Moderation, that the literature dates back to 1983-4 (see McConnell and Perez-Quiroz, 2000).

Overall, while our empirical results support the existence of a reversal in the relative predictive ability of a variety of predictors of inflation around the time of the Great Moderation, and therefore support the empirical evidence in D'Agostino, Giannone and Surico (2006), we also find that the reversal in the predictive ability of output happened much earlier than that, around mid-1970s.

6 Conclusion

Our empirical analysis has shown that the predictive ability of a variety of models that aim at predicting future industrial production growth or inflation vary through time. Many predictors have performed considerably well in the beginning of the out-of-sample period that we consider, but worsened relative to the univariate autoregression benchmark during later sample periods. In general, there is more evidence of predictive ability for output than for inflation. The time of the reversal in the relative forecasting ability is very different for the two series: around the mid-1970s for output growth, and around 1983-4 for inflation. We believe that the latter is a new empirical stylized fact that we uncover, and which will be interesting to investigate.

References

[1] Amato, J. and N. Swanson (2001), "The Real-Time Predictive Content of Money for Output", *Journal of Monetary Economics 48*, 3-24.

[2] Clark, T., and M. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", *Journal of Econometrics* 105(1), 85-110.

[3] Clark, T. and M. McCracken (2005), "The Power of Tests of Predictive Ability in the Presence of Structural Breaks", *The Journal of Econometrics* 124, 1-31.

[4] Clark, T., and K. D. West (2006), "Using Out-of-sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis", *Journal of Econometrics* 135, 155-186.

[5] Croushore, D., and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130.

[6] Croushore, D. (2006), "Forecasting with Real-Time Macroeconomic Data", in Elliott, G., C. Granger and A. Timmermann, *Handbook of Economic Forecasting*, Elsevier, Chapter 17, 961-982.

[7] D'Agostino, A., D. Giannone and P. Surico (2006), "(Un)Predictability and Macroeconomic Stability", *ECB Working Paper* 605.

[8] Diebold, F. X., R. S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics,* 13, 253-263.

[9] Diebold, F. X., and G. D. Rudebusch (1991a), "Turning Point Prediction With the Composite Leading Index: An Ex-Ante Analysis," in *Leading Economic Indicators: New Approaches and Forecasting Records*, eds. K. Lahiri and G.H. Moore, Cambridge, U.K.: Cambridge University Press, 231-56.

[10] Diebold, F. X., and G. D. Rudebusch (1991b), "Forecasting Output With the Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association* 86, 603-10.

[11] Edge, R. M., T. Laubach, and J. C. Williams (2007), "Learning and Shifts in Long-Run Productivity Growth," *Journal of Monetary Economics*, 54, 2421-2438.

[12] Estrella, A. (2005), “Why Does the Yield Curve Predict Output and Inflation?”, *Economic Journal* 115, 722-744.

[13] Estrella, A., A.P. Rodrigues, and S. Schich (2003), “How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States”, *Review of Economics and Statistics* 85, 629-644.

[14] Giacomini, R., and B. Rossi (2008), “Forecast Comparisons in Unstable Environments”, Duke University Working Paper 2008-03.

[15] Giacomini, R. and H. White (2006), “Tests of Conditional Predictive Ability”, *Econometrica,* 74, 1545-1578.

[16] Inoue, A., and B. Rossi (2005), “Recursive Predictability Tests for Real-Time Data”, *Journal of Business and Economic Statistics,* 23(3), 336-345.

[17] Inoue, A., and L. Kilian (2006), "On the Selection of Forecasting Models", *Journal of Econometrics* 130(2), 273-306.

[18] Kozichi, S. (1997), “Predicting Real Growth and Inflation with the Yield Spread”, *Federal Reserve Bank of Kansas City Economic Review*, 38-57.

[19] McConnell, M.M., and G. Perez-Quiroz (2000), "Output Fluctuations in the United States: What Has Changed Since the Early 1980’s", *American Economic Review* 90(5), 1464-1476.

[20] Mishkin, F.S. (1990), ”What Does the Term Structure Tell Us About Future Inflation”, *Journal of Monetary Economics* XXV, 77-95.

[21] Orphanides, A., and S. Van Norden (2005), “The Reliability of Inflation Forecasts Based on Output Gaps in Real Time.” *Journal of Money, Credit, and Banking* 37, 583–601.

[22] Rossi, B. and T. Sekhoposyan (2008), “Not-for-publication Appendix to: Has models’ forecasting performance of future US output growth and inflation changed over time, and why?”, http://www.econ.duke.edu/working_papers/working_papers.php.

[23] Stark, T., and D. Croushore (2002), “Forecasting with a Real-Time Data Set for Macroeconomists,” Journal of Macroeconomics 24, 507-31.

[24] Stock, J.H., and M.W. Watson (1989), "Interpreting the Evidence on Money-Income Causality", *Journal of Econometrics* 40, 161-182.

[25] Stock, J.H., and M.W. Watson (1999a), "Business Cycle Fluctuations in US Macroeconomic Time Series", in: *Taylor, J.B. and M. Woodford, eds. Handbook of Macroeconomics. Volume 1A.* Handbooks in Economics, vol. 15. Amsterdam; New York and Oxford: Elsevier Science, North-Holland, 1999; 3-64.

[26] Stock, J.H., and M.W. Watson (1999b), "Forecasting Inflation", *Journal of Monetary Economics* 44, 293-335.

[27] Stock, J.H., and M.W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices", *Journal of Economic Literature.*

[28] West, K. D. (1996), "Asymptotic Inference about Predictive Ability", *Econometrica*, 64, 1067-1084.

7 Tables and Figures

Table 1. Description of Data Series

Label	Freq	Period	Name	Description	S
			Asset Prices		
rovnght	M	1959:1 - 2005:12	FYFF	Int Rate: Federal Funds (Effective)	D
rtbill	M	1959:1 - 2005:12	FYGM3	Int Rate: US Treasury Bills, Sec Mkt, 3-Mo	D
rbnds	M	1959:1 - 2005:12	FYGT1	Int Rate: US Treasury Const Maturities, 1-Yr	D
rbndm	M	1959:1 - 2005:12	FYGT5	Int Rate: US Treasury Const Maturities, 5-Yr	D
rbndl	M	1959:1 - 2005:12	FYGT10	Int Rate: US Treasury Const Maturities, 10-Yr	D
exrate	M	1959:1 - 2005:12	EXRUS	United States; Effective Exchange Rate	D
stockp	M	1959:1 - 2005:12	FSPCOM	S&P's Common Stock Price Index: Composite	D
			Activity		
ip	M	1959:1 - 2005:12	IPN10	Industrial Production Index - Total Index	D
capu	M	1959:1 - 2002:06	IPXMCA	Capacity Utilization Rate: MFG, Total	D
emp	M	1959:1 - 2005:12	LHEM	Civilian Labor Force: Employed, Total	D
unemp	M	1959:1 - 2005:12	LHUR	Unemp Rate: All Workers, 16 Years and Over	D
			Wages & Prices		
cpi	M	1959:1 - 2005:12	PUNEW	CPI-U: All Items	D
ppi	M	1959:1 - 2005:12	PW	Producer Price Index: All Commodities	D
earn	M	1959:1 - 2003:04	LE6GP	Avg Hourly Earnings - Goods - Producing	D
oil	M	1959:1 - 2003:06	WPU0561	Crude Petroleum (Domestic Production)	B
			Money		
m0	M	1959:1 - 2003:06	FMBASE	Monetary Base, Adj For Reserve Req Chgs	D
m1	M	1959:1 - 2005:12	FM1	Money Stock: M1	D
m2	M	1959:1 - 2005:12	FM2	Money Stock: M2	D
m3	M	1959:1 - 2005:12	FM3	Money Stock: M3	D

Note: Sources (S) are abbreviated as follows: B-Bureau of Labor Statistics and D-DRI Basic Economics Database. The same names preceded by an "r" denote the real version of the variable, that is the variable minus CPI inflation. For example, Real Interest Rates (such as rrovnght, rrtbill, rrbnds, rrbndm, rrbndl) are defined as Nominal Interest Rates minus CPI inflation. The spread is defined as the difference between rbndl and rovnght.

Table 2. Forecasting Output Growth: Tests of Average Equal Predictive Ability

Variable		rMSFE	p-value	Variable		rMSFE	p-value
rovnght	level	-1.71	0.09	emp	Δ ln	1.67	0.09
rtbill	level	-1.45	0.15	emp	gap	-2.53	0.01
rbnds	level	-1.34	0.18	unemp	level	-0.50	0.62
rbndm	level	-1.15	0.25	unemp	Δ ln	0.02	0.99
rbndl	level	-1.22	0.22	unemp	gap	-2.55	0.01
rovnght	Δ	-0.29	0.77	cpi	Δ ln	-1.62	0.10
rtbill	Δ	0.92	0.36	cpi	Δ^2 ln	1.97	0.05
rbnds	Δ	0.04	0.97	ppi	Δ ln	-1.28	0.20
rbndm	Δ	-0.74	0.46	ppi	Δ^2 ln	2.39	0.02
rbndl	Δ	-0.74	0.46	earn	Δ ln	-0.16	0.88
rrovnght	level	-1.75	0.08	earn	Δ^2 ln	2.30	0.02
rrtbill	level	-1.39	0.17	oil	Δ ln	0.36	0.72
rrbnds	level	-1.32	0.19	oil	Δ^2 ln	1.49	0.14
rrbndm	level	-1.23	0.22	roil	ln	0.71	0.48
rrbndl	level	-0.98	0.33	roil	Δ ln	0.29	0.77
rrovnght	Δ	-0.18	0.86	m0	Δ ln	2.69	0.01
rrtbill	Δ	0.62	0.54	m0	Δ^2 ln	3.72	0.00
rrbnds	Δ	0.07	0.95	m1	Δ ln	0.88	0.38
rrbndm	Δ	-0.34	0.74	m1	Δ^2 ln	2.19	0.03
rrbndl	Δ	-0.11	0.91	m2	Δ ln	-2.12	0.03
rspread	level	-2.95	0.00	m2	Δ^2 ln	1.74	0.08
exrate	Δ ln	1.29	0.20	m3	Δ ln	0.47	0.64
stockp	Δ ln	-2.54	0.01	m3	Δ^2 ln	1.95	0.05
rstockp	Δ ln	-2.79	0.01	rm0	Δ ln	-1.93	0.05
capu	level	-1.94	0.05	rm1	Δ ln	-1.51	0.13
rm3	Δ ln	-2.33	0.02	rm2	Δ ln	-2.97	0.00

Note: rMSFE denotes the rescaled MSFE difference (negative values indicate that the model with explanatory variables is better than the autoregressive model) and p-value denote the full out-of-sample test p-value.

Table 3. Forecasting Output Growth in Real-Time: Tests of Average Equal Predictive Ability

Variable		rMSFE	p-value	Variable		rMSFE	p-value
rovnght	level	-1.62	0.11	emp	Δ ln	0.76	0.45
rtbill	level	-1.39	0.17	emp	gap	-3.19	0.00
rbnds	level	-1.40	0.16	unemp	level	-0.68	0.50
rbndm	level	-1.41	0.16	unemp	Δ ln	0.86	0.39
rbndl	level	-1.30	0.19	unemp	gap	-3.27	0.00
rovnght	Δ	-0.81	0.42	cpi	Δ ln	-1.39	0.16
rtbill	Δ	0.32	0.75	cpi	Δ^2 ln	1.75	0.08
rbnds	Δ	-0.19	0.85	ppi	Δ ln	-1.25	0.21
rbndm	Δ	-0.51	0.61	ppi	Δ^2 ln	2.64	0.01
rbndl	Δ	-0.27	0.79	earn	Δ ln	0.07	0.95
rrovnght	level	-1.67	0.09	earn	Δ^2 ln	2.13	0.03
rrtbill	level	-1.31	0.19	oil	Δ ln	-0.09	0.93
rrbnds	level	-1.36	0.17	oil	Δ^2 ln	1.52	0.13
rrbndm	level	-1.38	0.17	roil	ln	0.16	0.87
rrbndl	level	-1.01	0.31	roil	Δ ln	-0.20	0.84
rrovnght	Δ	-0.77	0.44	m0	Δ ln	2.94	0.00
rrtbill	Δ	0.17	0.86	m0	Δ^2 ln	4.74	0.00
rrbnds	Δ	-0.10	0.92	m1	Δ ln	1.97	0.05
rrbndm	Δ	-0.21	0.83	m1	Δ^2 ln	3.08	0.00
rrbndl	Δ	0.09	0.93	m2	Δ ln	-1.01	0.31
rspread	level	-2.79	0.01	m2	Δ^2 ln	1.94	0.05
exrate	Δ ln	1.04	0.30	m3	Δ ln	1.02	0.31
stockp	Δ ln	-1.92	0.05	m3	Δ^2 ln	2.42	0.02
rstockp	Δ ln	-2.17	0.03	rm0	Δ ln	-1.50	0.13
capu	level	-1.14	0.25	rm1	Δ ln	-1.05	0.30
				rm2	Δ ln	-2.26	0.02
				rm3	Δ ln	-1.28	0.20

Note: rMSFE denotes the rescaled MSFE difference (negative values indicate that the model with explanatory variables is better than the autoregressive model) and p-value denote the full out-of-sample test p-value.

Table 4. Forecasting Inflation: Tests of Average Equal Predictive Ability

Variable		rMSFE	p-value	Variable		rMSFE	p-value
rovnght	level	0.10	0.92	ip	Δ ln	2.74	0.01
rtbill	level	0.15	0.88	ip	gap	1.39	0.16
rbnds	level	-0.13	0.90	emp	Δ ln	-0.95	0.34
rbndm	level	-0.75	0.45	emp	gap	1.17	0.24
rbndl	level	-0.89	0.37	unemp	level	1.06	0.29
rovnght	Δ	0.19	0.85	unemp	Δ ln	-0.60	0.55
rtbill	Δ	0.17	0.87	unemp	gap	0.66	0.51
rbnds	Δ	0.13	0.90	ppi	Δ ln	0.56	0.58
rbndm	Δ	0.53	0.59	ppi	Δ^2 ln	0.75	0.45
rbndl	Δ	0.69	0.49	earn	Δ ln	-1.47	0.14
rrovnght	level	0.10	0.92	earn	Δ^2 ln	2.72	0.01
rrtbill	level	0.15	0.88	oil	Δ ln	0.63	0.53
rrbnds	level	-0.13	0.90	oil	Δ^2 ln	1.87	0.06
rrbndm	level	-0.75	0.45	roil	ln	0.39	0.69
rrbndl	level	-0.89	0.37	roil	Δ ln	-0.26	0.79
rrtbill	Δ	-0.71	0.48	m0	Δ^2 ln	2.13	0.03
rrbnds	Δ	-0.94	0.35	m1	Δ ln	0.07	0.94
rrbndm	Δ	-1.20	0.23	m1	Δ^2 ln	3.09	0.00
rrbndl	Δ	-1.35	0.18	m2	Δ ln	2.17	0.03
rspread	level	0.40	0.69	m2	Δ^2 ln	2.79	0.01
exrate	Δ ln	0.65	0.52	m3	Δ ln	-1.94	0.05
stockp	Δ ln	2.03	0.04	m3	Δ^2 ln	4.03	0.00
rstockp	Δ ln	2.03	0.04	rm0	Δ ln	-0.26	0.79
capu	level	-0.53	0.60	rm1	Δ ln	0.07	0.94
				rm2	Δ ln	2.17	0.03
				rm3	Δ ln	-1.94	0.05

Note: rMSFE denotes the rescaled MSFE difference (negative values indicate that the model with explanatory variables is better than the autoregressive model) and p-value denote the full out-of-sample test p-value.

Table 5. Forecasting Output Growth:
Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
rovnght	level	0.00	0.00	1976	5	emp	Δ ln	0.84	0.71		
rtbill	level	0.00	0.00	1976	5	emp	gap	0.01	0.02	1976	3
rbnds	level	0.00	0.00	1976	3	unemp	level	0.86	0.82		
rbndm	level	0.03	0.02	1976	2	unemp	Δ ln	1.00	1.00		
rbndl	level	0.07	0.05	1976	2	unemp	gap	0.00	0.00	1976	2
rovnght	Δ	1.00	1.00			cpi	Δ ln	0.00	0.00	1975	10
rtbill	Δ	0.14	0.09	1975	10	cpi	Δ^2 ln	0.58	0.48		
rbnds	Δ	0.81	0.67			ppi	Δ ln	0.19	0.13		
rbndm	Δ	0.85	0.85			ppi	Δ^2 ln	0.06	0.05	1975	9
rbndl	Δ	0.66	0.66			earn	Δ ln	1.00	0.88		
rrovnght	level	0.00	0.00	1976	7	earn	Δ^2 ln	0.71	1.00		
rrtbill	level	0.00	0.00	1976	3	oil	Δ ln	1.00	1.00		
rrbnds	level	0.00	0.00	1976	3	oil	Δ^2 ln	0.20	0.14		
rrbndm	level	0.11	0.08	1976	2	roil	ln	1.00	0.83		
rrbndl	level	0.45	0.33			roil	Δ ln	1.00	1.00		
rrovnght	Δ	1.00	1.00			m0	Δ ln	0.73	1.00		
rrtbill	Δ	0.61	0.47			m0	Δ^2 ln	0.30	0.65		
rrbnds	Δ	0.88	0.77			m1	Δ ln	1.00	1.00		
rrbndm	Δ	0.88	0.84			m1	Δ^2 ln	1.00	1.00		
rrbndl	Δ	1.00	0.84			m2	Δ ln	0.00	0.00	1977	10
rspread	level	0.00	0.00	1976	7	m2	Δ^2 ln	1.00	1.00		
exrate	Δ ln	0.12	0.08	1975	7	m3	Δ ln	0.56	0.40		
stockp	Δ ln	0.00	0.00	1976	7	m3	Δ^2 ln	0.83	0.77		
rstockp	Δ ln	0.00	0.00	1976	7	rm0	Δ ln	0.00	0.00	1975	11
capu	level	0.50	0.61			rm1	Δ ln	0.00	0.00	1975	11
						rm2	Δ ln	0.00	0.00	1976	8
						rm3	Δ ln	0.00	0.00	1975	11

Note: The table reports p-values of Giacomini and Rossi's (2008) test of One-Time Reversal test ("One-time"), the $\sup_t LM_1(t)$ test for a break only ("Break"), as well as the estimate break date when the pvalue is less than 10%.

Table 6. Forecasting Output Growth in Real-Time:
Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
rovnght	level	0.00	0.00	1976	5	emp	Δ ln	1.00	1.00		
rtbill	level	0.00	0.00	1976	5	emp	gap	0.00	0.01	1984	10
rbnds	level	0.00	0.00	1976	3	unemp	level	1.00	1.00		
rbndm	level	0.04	0.03	1976	3	unemp	Δ ln	1.00	1.00		
rbndl	level	0.06	0.04	1976	2	unemp	gap	0.00	0.00	1984	10
rovnght	Δ	1.00	1.00			cpi	Δ ln	0.05	0.03	1976	5
rtbill	Δ	0.44	0.30			cpi	Δ^2 ln	0.86	0.81		
rbnds	Δ	0.81	0.70			ppi	Δ ln	0.15	0.12		
rbndm	Δ	0.81	0.78			ppi	Δ^2 ln	0.00	0.00	1975	10
rbndl	Δ	0.75	0.69			earn	Δ ln	0.76	0.61		
rrovnght	level	0.00	0.00	1976	5	earn	Δ^2 ln	0.40	0.59		
rrtbill	level	0.00	0.00	1976	5	oil	Δ ln	0.82	0.73		
rrbnds	level	0.00	0.00	1976	3	oil	Δ^2 ln	0.29	0.21		
rrbndm	level	0.10	0.08	1976	3	roil	ln	1.00	1.00		
rrbndl	level	0.09	0.11			roil	Δ ln	0.69	0.61		
rrovnght	Δ	0.89	0.89			m0	Δ ln	0.17	0.80		
rrtbill	Δ	1.00	0.86			m0	Δ^2 ln	0.00	0.13		
rrbnds	Δ	0.89	0.80			m1	Δ ln	0.63	0.76		
rrbndm	Δ	0.76	0.70			m1	Δ^2 ln	0.26	0.54		
rrbndl	Δ	0.73	0.63			m2	Δ ln	0.06	0.04	1976	12
rspread	level	0.00	0.00	1976	10	m2	Δ^2 ln	0.74	0.88		
exrate	Δ ln	0.48	0.35			m3	Δ ln	1.00	1.00		
stockp	Δ ln	0.02	0.02	1976	7	m3	Δ^2 ln	0.42	0.68		
rstockp	Δ ln	0.01	0.01	1976	7	rm0	Δ ln	0.13	0.22		
capu	level	1.00	1.00			rm1	Δ ln	0.04	0.03	1976	6
						rm2	Δ ln	0.00	0.00	1976	7
						rm3	Δ ln	0.09	0.06	1976	3

Note: The table reports p-values of Giacomini and Rossi's (2008) test of One-Time Reversal test ("One-time"), the $\sup_t LM_1(t)$ test for a break only ("Break"), as well as the estimate break date when the pvalue is less than 10%.

Table 7. Forecasting Inflation:

Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
rovnght	level	0.30	0.22			ip	Δ ln	0.25	0.58		
rtbill	level	0.17	0.12			ip	gap	0.62	0.68		
rbnds	level	0.09	0.07	1980	9	emp	Δ ln	0.04	0.05	1984	3
rbndm	level	0.62	0.63			emp	gap	0.77	0.80		
rbndl	level	0.47	0.51			unemp	level	0.31	0.18		
rovnght	Δ	0.88	0.79			unemp	Δ ln	0.19	0.18		
rtbill	Δ	1.00	1.00			unemp	gap	0.41	0.25		
rbnds	Δ	1.00	1.00			ppi	Δ ln	1.00	1.00		
rbndm	Δ	1.00	1.00			ppi	Δ^2 ln	1.00	0.89		
rbndl	Δ	1.00	1.00			earn	Δ ln	0.70	0.86		
rrovnght	level	0.30	0.22			earn	Δ^2 ln	0.01	0.10	1986	12
rrtbill	level	0.17	0.12			oil	Δ ln	1.00	1.00		
rrbnds	level	0.09	0.07	1980	9	oil	Δ^2 ln	0.58	0.60		
rrbndm	level	0.62	0.63			roil	ln	0.77	0.64		
rrbndl	level	0.47	0.51			roil	Δ ln	1.00	1.00		
rrovnght	Δ	0.39	0.32			m0	Δ ln	0.22	0.16		
rrtbill	Δ	0.55	0.50			m0	Δ^2 ln	0.55	0.87		
rrbnds	Δ	0.41	0.38			m1	Δ ln	0.60	0.49		
rrbndm	Δ	0.21	0.22			m1	Δ^2 ln	0.17	0.45		
rrbndl	Δ	0.21	0.15			m2	Δ ln	0.03	0.09	1984	12
rspread	level	0.44	0.27			m2	Δ^2 ln	0.06	0.22		
exrate	Δ ln	1.00	1.00			m3	Δ ln	0.15	0.36		
stockp	Δ ln	0.43	0.76			m3	Δ^2 ln	0.00	0.05	1984	1
rstockp	Δ ln	0.43	0.76			rm0	Δ ln	0.22	0.16		
capu	level	0.03	0.03	1984	8	rm1	Δ ln	0.60	0.49		
						rm2	Δ ln	0.03	0.09	1984	12
						rm3	Δ ln	0.15	0.36		

Note: The table reports p-values of Giacomini and Rossi's (2008) test of One-Time Reversal test ("Joint"), the $\sup_t LM_1(t)$ test for a break only ("Break"), as well as the estimate break date when the pvalue is less than 10%.

Figure 1. Forecasting US output growth over time.

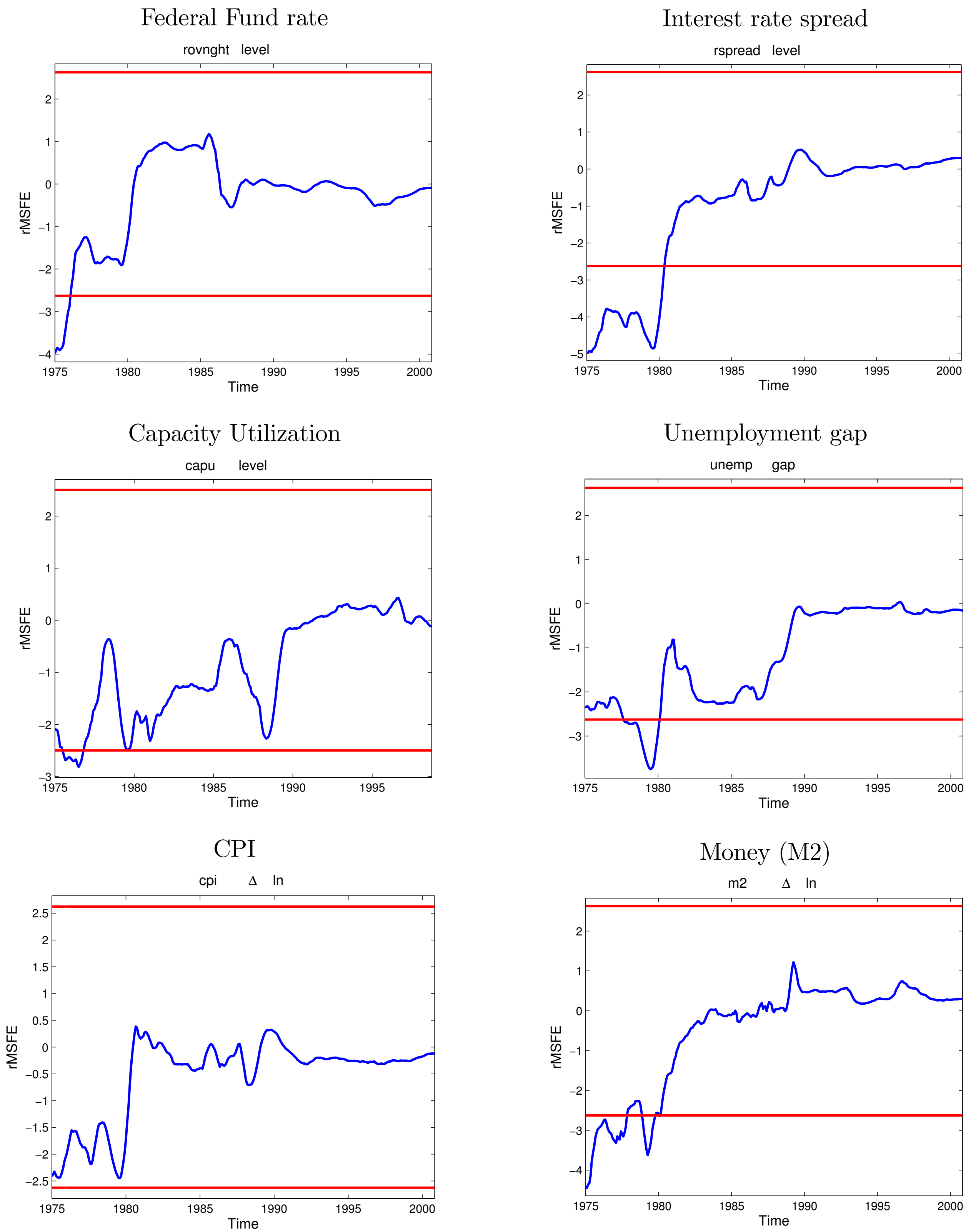

Figure 2. Forecasting output growth using real-time data

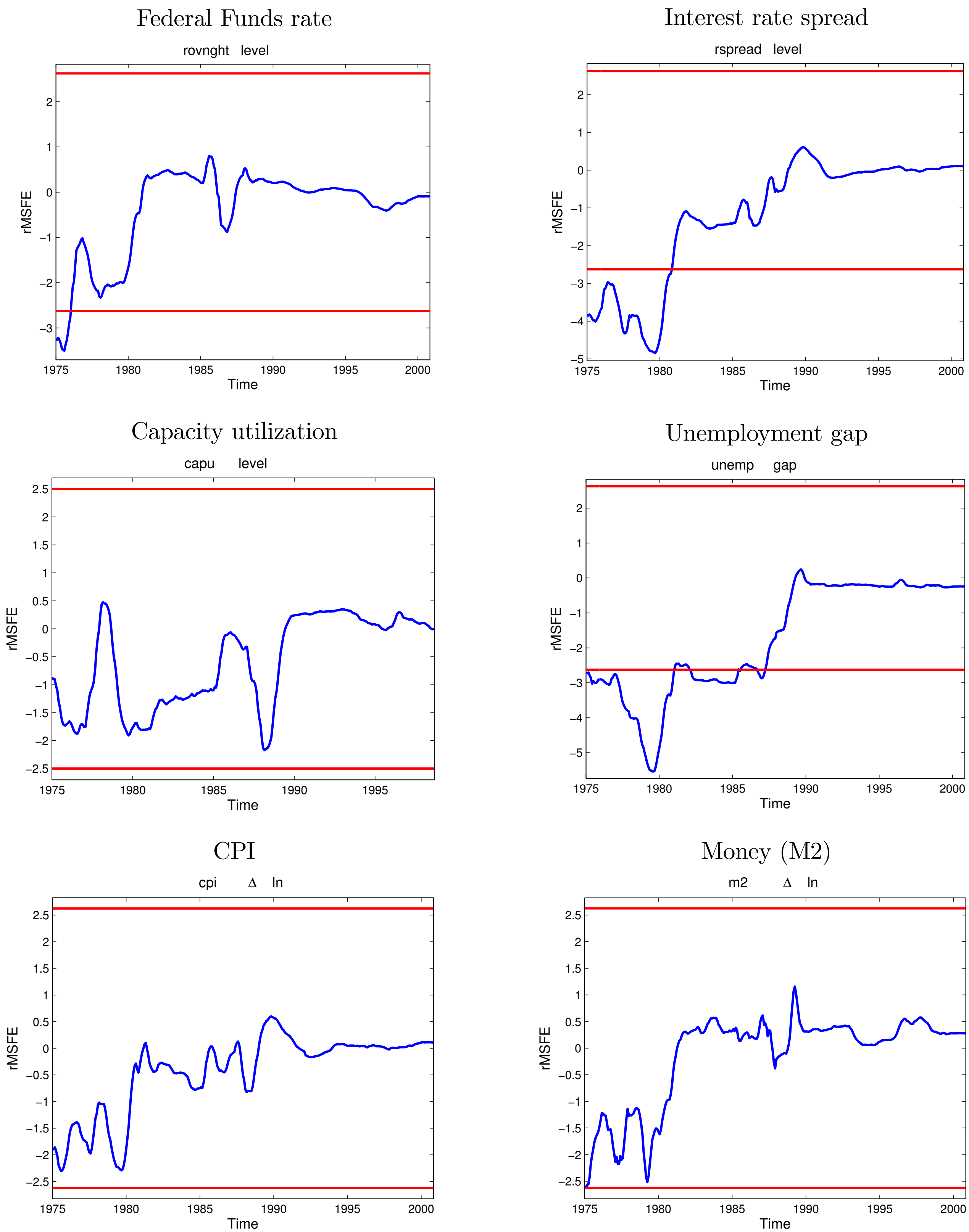

Figure 3. Forecasting US inflation over time