Have Economic 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: July 2009.

Abstract

We evaluate various economic 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, and their predictive ability significantly worsened around the time of the Great Moderation.

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

Acknowledgments: The first author gratefully acknowledges financial support from a SAS Forecasting Research Grant. The Grant supported research on new methods for forecasting and model evaluation. This paper shows the empirical relevance of such techniques in the presence of time-varying relative forecasting performance. We thank Sharon Kozicki for information, and seminar participants at the 2007 Midwest Econometrics Group Conference, the 2008 "Nowcasting with Model Combination Workshop" at the Reserve Bank of New Zealand, the 2008 "Workshop on Inflation Forecasting" at the Central Bank of Chile, the Macro-Money seminar at UNC-Chapel Hill, and the EMSG at Duke University for comments. The earlier version of the paper has been greatly improved by suggestions from Todd Clark, Michael McCracken and three anonymous referees.

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), nor they formally study the evolution of the relative performance over time. 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 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). 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 apply these techniques to empirically investigate whether the relative performance of competing models for forecasting US industrial 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. Following the practice of Stock and Watson (2003, Section 4), throughout the paper we will refer to the growth rate of industrial production as output growth.

In particular, we focus on predicting the h-periods ahead output growth and inflation by using autoregressive terms as well as lagged values of important economic explanatory variables, one at a time. In particular, we use interest rates, interest spreads, money supply, unemployment, as well as indices of leading indicators among others. These series have been found to have predictive content for output growth and inflation at different periods in time. Using both fully revised and real-time data, we find substantial reversals in the relative forecasting performance. This analysis, however, is still silent about the economic reasons of why such reversals have happened. However, using the Giacomini and Rossi's (2008) procedure, we 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), the index of supplier deliveries, and the index of leading indicators. 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 weaker than that of output growth over the full sample. 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 models with predictors such as employment and unemployment measures, among others, 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 examine changes in the models' relative performance over time, it will shed light on this issue. We show that for some series 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. 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, and Section 3 discusses the statistical methods used in the paper. Section 4 and 5 present and discuss the empirical results for the Fluctuation test: Section 4 focuses on predicting output growth using both fully revised and data available in real-time, whereas Section 5 focuses on forecasting inflation. Section 6 instead focuses on the empirical results for the One-time Reversal test. Section 7 discusses robustness analysis, and Section 8 concludes.

2 A description of the models and the 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).[1] Following Stock and Watson (2003), the models with explanatory variables (which we refer to as "economic models") are:

$$y_{t+h}^{h} = \beta_0 + \beta_1(L)x_t + \beta_2(L)y_t + \epsilon_{t+h}, \quad t = 1, 2, ...T, \tag{1}$$

where y_{t+h}^{h} is either the h period ahead output growth at time t defined by $y_{t+h}^{h} = 1200 \ln(IP_{t+h}/IP_t)/h$ or the h period ahead inflation at time t defined by $y_{t+h}^{h} = 1200 \ln(CPI_{t+h}/CPI_t)/h - 1200 \ln(CPI_t/CPI_{t-1})$, 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 $y_t = 1200 \ln(CPI_t/CPI_{t-1}) - 1200 \ln(CPI_{t-1}/CPI_{t-2})$, 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 BIC.[2] We consider one year ahead output and inflation growth by setting $h = 12$ months.

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 samples start much later than the other series, preventing a large out-of-sample size for our

[1] 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).

[2] Lag orders are selected once and for all in order to minimize the effect of the lags on the forecasting performance of the models. q is selected based on full sample estimation of the benchmark model, eq. 2 below. After choosing the "best" benchmark specification, p is chosen based on full sample estimation of model 1. The maximum lag length considered in both cases is 12. For robustness we consider recursive lag length selection as well. The results for the recursive lag length selection are discussed in Section 7.

forecast comparisons. In addition, we consider a set of leading indicators such as the Conference Board's index of ten leading indicators, average weekly manufacturing hours and the index of supplier deliveries in order to have a comprehensive coverage of the series that are commonly used by applied forecasters. As recommended by Kozicki and Hoffman (2004), we consider the CPI series with a base year of 1967 in order to avoid distortions in the variability of the implied inflation due to re-basing.[3] Furthermore, we use the consumption deflator as an alternative (housing-consistent) measure of prices.[4] The sources and the exact description of the data are provided in Table 1. Following Stock and Watson (2003), we consider several transformations of the data series, namely levels, differences, second differences and "gaps," where the gaps are estimated by HP (Hodrick and Prescott, 1997) filter.

INSERT TABLE 1 HERE

The predictors that we mainly focus on include a few representative series that are commonly thought of as leading indicators for output growth or inflation. The representative series that we consider for forecasting output growth are the Federal Funds rate, the interest rate spread, the growth rate of money (M2), the index of ten leading indicators, average weekly manufacturing hours and the index of supplier deliveries. We consider the one-year Treasury bond rate, the interest rate spread, the growth rate of money (M3), capacity utilization, the unemployment gap and the growth rate of output as representative series for inflation forecasting. In addition, we succinctly summarize the results for the whole database.

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

$$y^h_{t+h} = \beta_0 + \beta_2(L)y_t + \eta_{t+h}, \quad t = 1, 2, ...T. \tag{2}$$

Models (1) and (2) are both estimated by OLS in rolling samples of 120 observations ($R = 120$). Accordingly, the first 12-months ahead out-of-sample forecast is made for 1970:3 (our data starts in 1959:1, and we lose two observations as we take second differences of some data series).

Let the pseudo out-of-sample forecast errors of models (1) and (2) be denoted, respectively, by $\hat{\epsilon}_{t+h}$ and $\hat{\eta}_{t+h}$.[5] To capture the time variation in the relative performance, we construct rolling estimates of the relative Mean Square Forecast Errors (rMSFE) using a two-sided window of 120

[3] The 1967 base year monthly CPI series provided by the BLS are not seasonally adjusted. We seasonally adjust the series by X-11 filtering.

[4] As referees carefully pointed out, part of the instability in the CPI series observed in the early 1980s can be attributed to the treatment of housing effective January 1983. In order to avoid the distortions introduced by the measurement changes, we consider this alternative price measure.

[5] $\hat{\epsilon}_{t+h}$ is the difference between the realization of y^h_{t+h} and the forecasted value $\hat{y}^h_{t+h|(1)}$ based on model (1). $\hat{\eta}_{t+h}$ is the difference between the realization of y^h_{t+h} and the forecasted value $\hat{y}^h_{t+h|(2)}$ based on model (2).

months. Ultimately, our object of interest is the difference between the mean square forecast errors (rMSFE) of the "economic" model (1) and that of the univariate autoregression (2) calculated over these rolling windows ($m = 120$):

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

We choose $m = R = 120$ to strike a balance between obtaining good estimates of each of the relative MSFE differences (which require m sufficiently large) and obtaining a large enough sample of rolling MSFEs that allows us to follow the evolution of the relative forecast performance over time (which require a large value for $T - m$). We verified the robustness of our results to different choices of the forecast evaluation window size.

3 A description of the statistical methods

In order to test whether the relative forecasting performance has changed over time, we utilize both tests proposed by Giacomini and Rossi (2008): the Fluctuation and the One-time Reversal tests. In what follows, we briefly describe each of these tests and their implementation.

The Fluctuation test relies on a measure of the local relative forecasting performance of the models estimated over rolling windows of data. It is implemented by plotting the sample path of the relative measure of local performance, together with critical values which, if crossed, signal that one of the models outperformed its competitor at some point in time.[6] More in detail, the Fluctuation test is a re-scaled version of $rMSFE_t$, and it is constructed as follows:

$$F_{t,m}^{OOS} = \widehat{\sigma}^{-1} m^{-1/2}\left(\sum_{j=t-m/2}^{t+m/2} \hat{\epsilon}_{t+h}^2 - \sum_{j=t-m/2}^{t+m/2} \hat{\eta}_{t+h}^2\right), \tag{4}$$

for $t = R+h+m/2, ..., T-m/2+1$, where $\widehat{\sigma}^2$ is a Heteroskedasticy and Autocorrelation Consistent (HAC) estimator of the asymptotic variance $\sigma^2 = var\left(P^{-1/2}\sum_{j=R+h}^{T}\left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\right)$, where $P = T - R$. For example,

$$\widehat{\sigma}^2 = \sum_{i=-q(P)+1}^{q(P)-1}(1 - |i/q(P)|)P^{-1}\sum_{j=R+h}^{T}\left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\left(\hat{\epsilon}_{j-i}^2 - \hat{\eta}_{j-i}^2\right), \tag{5}$$

[6] We could implement the Giacomini and Rossi (2008) Fluctuation test either in the Giacomini and White's (2006) or the Clark and West's (2006) frameworks. The fundamental difference in the two frameworks is that they test two different null hypotheses: the null hypothesis in Clark and West (2006) concerns forecast losses that are evaluated at the population parameters, whereas in Giacomini and White (2006) the losses depend on estimated in-sample parameters. Thus, while the former needs a correction for parameter estimation error, the latter does not.

and $q(P)$ is a bandwidth that grows with P (e.g., Newey and West, 1987). In practice, we choose $q(P) = P^{1/4}$.

The null hypothesis of the test is that the models' forecasting performance is the same at each point in time, that is

$$H_0 : E\left(\hat{\epsilon}_t^2 - \hat{\eta}_t^2\right) = 0, \quad t = R+h, ..., T. \tag{6}$$

The asymptotic distribution of the Fluctuation test under the null hypothesis can be approximated by functionals of Brownian motions. Critical values for various significance levels and various window and sample sizes are provided in Giacomini and Rossi (2008). In particular, for the window and sample sizes considered in this paper, for which $m/P \simeq 0.3$, the null hypothesis is rejected at the 10% significance level against the two-sided alternative $E\left(\hat{\epsilon}_t^2 - \hat{\eta}_t^2\right) \neq 0$ when $max_t \left|F_{t,m}^{OOS}\right| > 2.766$. Furthermore, the time path of $F_{t,m}^{OOS}$ contains valuable information. If the path crosses the lower bound then we conclude that the largest model (the "economic" model) forecasts best, whereas if the path crosses the upper bound then we conclude that the small model (the autoregressive benchmark) forecasts best.

The second test that we consider is the One-time Reversal test, which instead is designed for a specific alternative hypothesis. It tests the null hypothesis that the two models perform equally well at each point in time against the alternative that there is a one-time break in the relative performance. One of the advantages of this procedure is that it can be used to estimate the time of the reversal in the relative performance. The One-time Reversal test is implemented as follows. First, we test the null hypothesis of equal performance at each point in time (6) by using the statistic

$$QLR_P^* = \sup_t \left[LM_1 + LM_2(t)\right], \quad t \in \{[0.15P], ... [0.85P]\}, \tag{7}$$

where

$$\begin{aligned} LM_1 &= \hat{\sigma}^{-2} P^{-1} \left[\sum_{j=R+h}^{T} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\right]^2 \\ LM_2(t) &= \hat{\sigma}^{-2} P^{-1} (t/P)^{-1} (1-t/P)^{-1} \left[\sum_{j=R+h}^{t} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right) - (t/P) \sum_{j=R+h}^{T} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\right]^2, \end{aligned}$$

and $\hat{\sigma}^2$ is a HAC estimator of the asymptotic variance σ^2, for example (5). The null hypothesis is rejected at the 10% significance level when $QLR_P^* > 8.1379$. If the test rejects, we analyze whether the rejection is due to instabilities in the relative performance or to a model being constantly better than its competitor. The rejection is attributed to instabilities in the relative forecasting performance if $\sup_t LM_2(t) > 2.71$. The point in time associated with the largest value of $LM_2(t)$ identifies the time of the break: $t^* = \arg\max_{t \in \{0.15P, ..., 0.85P\}} LM_2(t)$. The rejection is instead attributed to a model being constantly better if $LM_1 > 7.17$.[7]

Note that the One-time Reversal test might not reject the null hypothesis even if the test of average equal predictive ability does. In fact, in order to obtain power against reversals in the predictive ability, the One-time Reversal test estimates the predictive ability separately in subsamples of the data, thus 'losing observations' (and therefore power) relative to the average equal predictive ability test when one of the models is constantly better than its competitor over the full sample. On the other hand, however, the average equal predictive ability test has no power to detect situations in which the forecasting ability of the models is changing over time and the changes cancel out on average.

The applicability of the Fluctuation and One-time Reversal tests relies in general on stationarity assumptions (see Giacomini and Rossi, 2008). In particular, note that the assumptions in Giacomini and Rossi (2008) rule out high persistence in the loss function differences, such as unit roots; in order to take care of non-stationarities due to unit roots, in the implementation of the test we rely on appropriately first-differenced or second-differenced data. In addition, these tests rely on the assumption of global covariance stationarity, which rules out breaks in the variance of the MSFEs, and which may or may not be satisfied in the present application. However, unlike the Fluctuation test, the Wald-test version of the One-time Reversal test is robust to one-time changes in the volatility of the relative MSFE at the time of the reversals. This is an important feature of the latter test, as such changes in volatility are typically associated with the Great Moderation, which we find to have an important role in our paper. In this context, the Wald-test version of the One-time Reversal test becomes appropriate, since the variance of the relative MSFE is estimated separately before and after the break. This test can be implemented as follows:

$$QLR_P^* = \sup_t W(t), \quad t \in \{[0.15P], ... [0.85P]\}, \tag{8}$$

where $W(t) = (\Delta L_{1,t} - \Delta L_{2,t})^2 / \left(\frac{\widehat{\sigma}_1^2}{t} + \frac{\widehat{\sigma}_2^2}{P-t}\right)$, $\Delta L_{1,t} = t^{-1} \sum_{j=R+h}^{R+h+t} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)$, while $\Delta L_{2,t} = (P-t)^{-1} \sum_{j=R+h+t+1}^{T} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)$; $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are HAC estimators of the asymptotic variances $\sigma_1^2 = var\left(t^{-1/2} \sum_{j=R+h}^{R+h+t} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\right)$ and $\sigma_2^2 = var\left((P-t)^{-1/2} \sum_{j=R+h+t+1}^{T} \left(\hat{\epsilon}_j^2 - \hat{\eta}_j^2\right)\right)$, respectively. Since Wald and LM-type tests have the same asymptotic distribution under the null hypothesis, we can use the same critical values as originally proposed by Giacomini and Rossi (2008) and reject at the 10% significance level if $QLR_P^* > 8.1379$.

[7] This procedure is justified by the fact that the two components LM_1 and LM_2 are asymptotically independent – see Rossi (2005).

4 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, the hours worked, the indices of leading indicators and of supplier deliveries, and the rate of money growth. We then consider a comprehensive survey of all the series in our database. We conclude by analyzing the robustness of our results to using real-time data.

4.1 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 available sample is shorter and thus the pseudo out-of-sample period stops some time in 2002 and 2003 (consult Table 1 for more details). Panel A focuses on predictors that are commonly considered leading indicators for output. In particular, 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). We also consider a series of leading indicators from the Conference Board's dataset. In particular, we focus on their index of ten leading indicators ("lead"), index of supplier deliveries ("deliveries"), as well as hours worked ("hours"). Money supply ("m2") 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).

The first column reports the re-scaled MSFE difference calculated over the full out-of-sample period.[8] A negative value indicates that the autoregressive model has a higher MSFE than the model with an additional explanatory variable. The second column reports the p-values based on the unconditional 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 most of the series in Panel A, such as the Fed Funds rate, the difference between the short and the long term interest rates, the index of supplier deliveries, the index of ten leading indicators and money supply. Only average weekly hours are not significant.

INSERT TABLE 2 HERE

[8] That is, $\widehat{\sigma}^{-1}P^{-1}\left[\sum_{j=R+h}^{T}\left(\hat{\epsilon}_j^2-\hat{\eta}_j^2\right)\right]^2$ where $\widehat{\sigma}$ is a HAC estimator of the variance of the out-of-sample relative squared forecast error differences.

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 the re-scaled rMSFE differences *over time* (eq. 4). 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 statistically 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 the rMSFEs depicted in Figure 1 would be inside the two boundary lines reported in the figure. It is clear that for some variables the paths are outside the bands, 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. The top panels of Figure 1 suggest that models that use interest rates as predictors perform quite well in the mid-late 1970s relative to the autoregressive model, whereas their performance significantly worsens during the 1980s. Similarly, the middle and last panels show that the performance of traditional leading indicators (such as the Conference Board's index of supplier deliveries and the index of ten leading indicators) worsens in the 1980s and 1990s, relative to the 1970s.

INSERT FIGURE 1 HERE

There is, however, an important difference between the various series. The spread and the Conference Board's ten leading indicators index seem to have maintained their predictive ability much longer than the other variables. Figure 1 also shows that money growth is a useful predictor for future output growth until the beginning of the 1980s, when its performance becomes statistically insignificantly different from that of an autoregression. Finally, hours worked do not have significant predictive content throughout the out-of-sample period.

4.2 Comprehensive overview for all series

We perform a similar analysis to that in the previous sub-section for all the series in our database. Panel B in Table 2 reports such results for the full sample. However, due to space constraints, detailed results for the evolution of the models' relative predictive ability over time for these series are reported in a Not-For-Publication Appendix (Rossi and Sekhposyan, 2009), 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 is not a significant predictor of output growth 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 Fed Funds rate.

Real activity measures, such as the growth rates of employment and unemployment, are never significant; however, employment and unemployment gaps have significant predictive content in the late 1970s but not in the 1980s and 1990s. In general, variables in the wage and price inflation categories are never significant, although the first difference of the inflation rate measured by any of the price indices (the producer price index, the consumer price index and the personal consumption deflator) is significantly worse than the autoregressive benchmark in the late 1970s. The growth rate of oil prices is a significant predictor only at very specific points in time (such as the mid 1970s). Finally, considering the money category we find that, unlike M2, M1 and M3 are never significant. However, the second difference of 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/unemployment gaps, stock prices, and M2.

4.3 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. 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 revised data, but that in other short samples the real-time forecasts were significantly worse than those using revised data. In addition, they found that forecasts of inflation were instead superior when based on revised data rather than 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.

In what follows, we revisit our analysis of the previous section by using real-time data for industrial production. In order to make the economic models suitable for forecasting in real-time, we make the following modifications to the set of variables that we consider. Financial data do not get revised; on the contrary, measures of real activity, money, wages and prices do. For such measures, the forecasting exercise is conducted only if real-time data (vintages) covering our out-of-sample period (1970:3-2005:12) is available. Accordingly, in addition to the asset prices, the series that we use to forecast industrial production in real-time are the non-farm payroll employment and civilian unemployment rate. The data for the non-farm payroll employment and total industrial production index are provided by the Philadelphia Fed in the Real-Time Data Set for Macroeconomists (Croushore and Stark, 2001). The monthly vintages of the civilian unemployment rate comes from the ArchivaL Federal Reserve Economic Data (ALFRED) database of the St. Louis Fed. Since the revisions to the CPI occur primarily due to occasional re-basing, by choosing CPI series with a 1967 base year we obtain a measure of real-time prices (see Clark and McCracken, 2008, 2009).[9]

INSERT TABLE 3 HERE

In this sub-section we concentrate on analyzing predictors such as the Fed Funds rate, the spread, the growth rate of employment ("emp"), the change in unemployment ("unemp Δ), the unemployment gap ("unemp gap") and the second difference of prices ("cpi Δ^2 ln") as explanatory variables. The Fed Funds rate and the spread are among the traditional leading indicators for output growth as discussed in the previous section. We include employment and unemployment measures since they undergo systematic revisions and it is of interest to consider to which degree the revisions matter.

First, we report results based on the full out-of-sample tests in Table 3. The results for the representative variables (reported in Panel A) are very robust, but in general the p-values for the additional variables (see Panel B) increase, to the point that some of the predictors are not significant anymore (such as the second difference of CPI) or they become insignificant at 5% level (such as the real stock price). 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 have less predictive content for real-time data than those using fully revised data.

The results for the Fluctuation test for real-time industrial production data are presented in Figure 2. Overall, for most variables, our results are qualitatively unchanged. Real-time data

[9] For some of the series we consider, namely industrial production and employment, comparable vintages exist in both of the databases. In this case we use the data from the Philadelphia Fed, since it is compiled from many sources, which include the sources already used by ALFRED. Both the Philadelphia Fed dataset and ALFRED contain additional series which we do not consider as their vintages start later than the out-of-sample period we consider.

show larger reversals in the forecasting performance of the models relative to the revised data, notably for employment growth and changes in unemployment rate; nonetheless, the reversals are not significant.

INSERT FIGURE 2 HERE

5 Forecasting inflation

In this section we focus on forecasting US CPI inflation. The measure of inflation that we consider throughout the paper is CPI with a base year of 1967.[10] We first present detailed empirical results for a few representative time series that are commonly considered leading indicators for inflation. Then we discuss a summary of the results for all the available economic series.

5.1 Detailed empirical results using representative series

As leading indicators for inflation, we consider the following series. Interest rates have been found to be important predictors for future inflation since the works by Mishkin (1990) and Kozicki (1997). Thus, we include short/medium interest rates, such as the one year Treasury bonds rate ("rbnds"), and the interest rate spread among our representative series. Perhaps the most important variables for predicting future inflation according to the Phillips curve relationship are real activity measures, such as industrial production and capacity utilization.[11] For example, Stock and Watson (1999a,b, 2008) 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 commodity prices. We therefore include capacity utilization ("capu") among our representative series, as well as the unemployment gap ("unemp gap") and the growth rate of output in light of the results in Orphanides and Van Norden (2005). Finally, we include the growth rate of money ("m3"), which is an important predictor for inflation according to the quantity theory of money.

Overall, the predictive ability of macroeconomic variables for future inflation is less widespread than that for future output growth. In fact, Panel A in Table 4 shows that fewer economic series have predictive content: only industrial production and capacity utilization are significant. However, Figure 3 shows striking evidence of changes in the relative performance of the models. Once we take that into account, we find more compelling empirical evidence in favor of economic predictors such as the Treasury bonds rate and the spread.[12] Indeed, we find that short-term (one-year) interest

[10] Our results do not change if we consider an inflation measure based on CPI with a base year of 1984.
[11] Similar results hold when using employment or unemployment growth rates – see Section 5.2.
[12] That is, while the full out-of-sample tests do not find significant predictive content for these variables, the Fluctuation test does, at least in some portions of the sample.

rates have marginal predictive content for inflation in the late 1970s, but that such predictive ability completely disappears in the mid-1980s. Similarly, the interest rate spread is significantly better than the autoregressive benchmark only sporadically, in the mid-1970s.

INSERT FIGURE 3 AND TABLE 4

Interestingly, we find that industrial production is a significant predictor in the late 1970s, but that, similarly to capacity utilization, its predictive content disappears in the 1980s. Finally, money (M3) never has significant predictive content.

5.2 Comprehensive overview for all series

Overall, we find very little significant predictive content in both nominal and real interest rates for forecasting future inflation when using tests of average out-of-sample predictive performance. For some interest rates, both real and nominal, however, there are interesting reversals in their predictive ability during 1980s, usually indicating that interest rates lost their predictive content in the mid-1980s. 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 that of capacity utilization, discussed above, showing that the predictive content disappears around the early 1980s, except for the industrial production gap, the employment and unemployment gaps, whose predictive ability is never significantly better than the benchmark. There is also very little significance for most wage and price measures, although some measures (such as the second difference of PCE, earnings and producer price inflation) are significantly worse than the benchmark in the late 1970s and/or early 1980s. Other definitions of money (M1 and M0) behave similarly to M3 (reported in Figure 3 above), although the predictive ability is somewhat smaller in magnitude. First differences of money growth are instead significantly worse predictors than the benchmark in the 1980s.[13]

We do not separately consider the CPI inflation forecasting exercise with real-time data. Given the real-time nature of the CPI series we use, the results will generally coincide with the ones presented in this section.

6 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

[13] Again, see the Not-For-Publication Appendix (Rossi and Sekhposyan, 2009) for detailed results.

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 5-7 report results for the "One-time Reversal" test ($QLR_P^*(t)$, labeled "One-time", eq. 7), as well as the test for breaks in the relative predictive ability ($\sup_t LM_2(t)$, labeled "Break"). If the latter finds empirical evidence in favor of changes in predictive ability, the table also reports the estimated time of the reversal "Break Date").

INSERT TABLES 5 AND 6

Table 5 focuses on forecasting output growth using revised data. Panel A in Table 5 shows that the timing of the break for the Fed Funds rate, the spread and the index of supplier deliveries is mid-1976, and for M2 is late 1977. The timing is slightly different for the index of leading indicators (end of 1975). When considering all the remaining series in our database (Panel B), we find that also some nominal interest rates as well as real interest rates show reversals at the same time (mid 1976). Therefore, interestingly, for almost all series, the most substantial reversal in relative predictive ability happened around mid-1970s. By comparing the results in the column labeled "One-time" with those in the column labeled "Break," we note that in most cases both tests reject. Thus, the majority of the rejections of the hypothesis that the two model's predictive ability is the same are linked to reversals in the predictive ability and not just to one model being significantly better than the competitor over the full sample. Similar results hold when we forecast output growth with real-time data – see Table 6. The One-time Reversal tests suggest breaks in real-time data, with timing comparable to the break dates of the revised data. A notable exception is the behavior of the interest rate spread and the unemployment gap, for which the real-time data suggest a reversal in 1984, while the revised data suggest that the break instead occurred in 1976.

A very different picture emerges when forecasting inflation. Table 7 shows that most reversals happen around 1984 rather than the mid 1970s. The reversals in predictive ability happened, therefore, around the time of the Great Moderation, which the literature dates back to 1983-4 (see McConnell and Perez-Quiros, 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.

INSERT TABLE 7

7 Robustness analysis

The empirical results for revised data in Sections 4, 5 and 6 focus on models where the lag length is selected via a BIC criterion over the full sample and then kept constant to minimize the effect of the lag selection on the forecasting performance of the models. In addition, the window size for forecast evaluation is 120 months. One might be concerned that changing the window size and/or the lag length selection criterion might affect our results. One might also wonder whether some of the instabilities in inflation in the early 1980s could be attributed to a measurement change in the CPI due to a shift in the treatment of housing effective in January 1983. Finally, given the discussion of the Great Moderation, it is important to consider the robustness of our results to changes in the variance of the MSFEs, as the tests reported previously rely on the assumption of global covariance stationarity. In this section, we investigate the robustness of our empirical results to these issues.

Tables 8 and 9 are similar to Tables 5 and 7 except that the lag length is selected via a recursive BIC. That is, the lag length is re-optimized each time the models are re-estimated, thus mimicking the behavior of a forecaster as new data become available. Similarly, Figures 4 and 5 are the counterpart of Figures 1 and 3 for the recursive BIC case. Table 8 and Figure 4 show that the results remain basically the same when forecasting output growth based on revised data. Table 9 and Figure 5 show that, for inflation, the results are qualitatively similar when we allow the lag length to be chosen recursively except that some of the predictive ability of interest rates becomes insignificant.

INSERT TABLES 8,9 AND FIGURES 4,5

By comparing Figures 6 and 7 with Figures 1 and 3, respectively, it is clear that the main qualitative results also remain unchanged if we use windows for forecast evaluation with 100 observations. In unreported results, we also verified that the main results are robust to other window sizes including for example 140 and 160.

INSERT FIGURES 6 AND 7

To address the concerns about measurement changes in the CPI, we consider an alternative series for inflation based on the personal consumption expenditure deflator, for which the measurement change is not an issue. Figure 8 reports the empirical results. By comparing Figure 8 and Figure 3, it is clear that the main conclusions of Section 5 are overall robust to using this alternative definition of inflation.

INSERT FIGURE 8

Finally, to address concerns about possible breaks in the variance of the relative MSFEs, Tables 10 and 11 report empirical results for the One-time Reversal test implemented with a Wald test procedure as in eq. (8). The tables show that our main empirical conclusions do not change.

INSERT TABLES 10 AND 11

8 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 periods of the sample. 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, which will be interesting to investigate further.

References

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

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

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

[4] Clark, T. E. and M. W. McCracken (2008), "Forecasting with Small Macroeconomic VARs in the Presence of Instabilities," *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, in D. Rapach and M. Wohar, ed., Emerald Publishing, 2008.

[5] Clark, T. E. and M. W. McCracken (2009), "Averaging Forecasts from VARs with Uncertain Instabilities," *Journal of Applied Econometrics*, forthcoming.

[6] Clark, T. E. and K. D. West (2006), "Using Out-of-sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135(1-2), 155-186.

[7] Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105(1), 111-130.

[8] 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.

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

[10] Diebold, F. X. and R. S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13(3), 253-263.

[11] 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.

[12] 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.

[13] Edge, R. M., T. Laubach, and J. C. Williams (2007), "Learning and Shifts in Long-Run Productivity Growth," *Journal of Monetary Economics* 54(8), 2421-2438.

[14] Estrella, A. (2005), "Why Does the Yield Curve Predict Output and Inflation?" *Economic Journal* 115(505), 722-744.

[15] 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(3), 629-644.

[16] Giacomini, R. and B. Rossi (2008), "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics*, forthcoming.

[17] Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74(6), 1545-1578.

[18] Hodrick, R. J. and E. C. Prescott (1997), "Post-war U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking* 29(1), 1-16.

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

[20] Kozicki, S. (1997), "Predicting Real Growth and Inflation with the Yield Spread," *Federal Reserve Bank of Kansas City Economic Review* 82(QIV), 39-57.

[21] Kozicki, S. and B. Hoffman (2004), "Rounding Error: A Distorting Influence on Index Data," *Journal of Money, Credit and Banking* 36(3), 319-38.

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

[23] Mishkin, F. S. (1990), "What Does the Term Structure Tell Us About Future Inflation," *Journal of Monetary Economics* 25(1), 77-95.

[24] Newey, W. K. and K. D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55(3), 703-708.

[25] Orphanides, A. and S. Van Norden (2005), "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time," *Journal of Money, Credit, and Banking* 37(3), 583-601.

[26] Rossi, B. (2005), "Optimal Tests for Nested Model Selection with underlying Parameter Instability," *Econometric Theory* 21(5), 962-990.

[27] Rossi, B. and T. Sekhoposyan (2009), "Not-For-Publication Appendix to: Has Economic Models' Forecasting Performance for US Output Growth and Inflation Changed Over Time, and When?" *Duke University Working Paper* 2009-06, http://www.econ.duke.edu/working_papers/working_papers.php.

[28] Stark, T. and D. Croushore (2002), "Forecasting with a Real-Time Data Set for Macroeconomists," *Journal of Macroeconomics* 24(4), 507-31.

[29] Stock, J. H. and M. W. Watson (1989), "Interpreting the Evidence on Money-Income Causality," *Journal of Econometrics* 40(1), 161-182.

[30] 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.

[31] Stock, J. H. and M. W. Watson (1999b), "Forecasting Inflation," *Journal of Monetary Economics* 44(2), 293-335.

[32] Stock, J. H. and M. W. Watson (2003), "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature* 41(3), 788-829.

[33] Stock, J. H. and M. W. Watson (2008), "Phillips Curve Inflation Forecasts," *NBER Working Paper* 14322.

[34] West, K. D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica* 64(5), 1067-1084.

9 Tables and Figures

Table 1: Description of Data Series

Label	Period	Name	Description	S
			Asset Prices	
rovnght	1959:1 - 2005:12	FYFF	Int Rate: Federal Funds (Effective)	D
rtbill	1959:1 - 2005:12	FYGM3	Int Rate: US Treasury Bills, Sec Mkt, 3-Mo	D
rbnds	1959:1 - 2005:12	FYGT1	Int Rate: US Treasury Const Maturities, 1-Yr	D
rbndm	1959:1 - 2005:12	FYGT5	Int Rate: US Treasury Const Maturities, 5-Yr	D
rbndl	1959:1 - 2005:12	FYGT10	Int Rate: US Treasury Const Maturities, 10-Yr	D
exrate	1959:1 - 2005:12	EXRUS	United States; Effective Exchange Rate	D
stockp	1959:1 - 2005:12	FSPCOM	S&P's Common Stock Price Index: Composite	D
			Activity	
ip	1959:1 - 2005:12	B5001	Industrial Production Total (sa)	F
capu	1959:1 - 2002:06	IPXMCA	Capacity Utilization Rate: MFG, Total	D
emp	1959:1 - 2005:12	LHEM	Civilian Labor Force: Employed, Total	D
unemp	1959:1 - 2005:12	LHUR	Unemp Rate: All Workers, 16 Years and Over	D
hours	1959:1 - 2005:12	A0M001	Average weekly hours, mfg. (hours)	C
deliveries	1959:1 - 2005:12	A0M032	Index of supplier deliveries - vendor perf. (pct.)	C
			Wages and Prices	
cpi	1959:1 - 2005:12	CUUR0000AA0	CPI - All Urban Consumers (nsa)	B
pce	1959:1 - 2005:12	PCE deflator	Price Indexes for Personal Cons. Expenditures	B
ppi	1959:1 - 2005:12	PW	Producer Price Index: All Commodities	D
earn	1959:1 - 2003:04	LE6GP	Avg Hourly Earnings - Goods - Producing	D
oil	1959:1 - 2003:06	WPU0561	Crude Petroleum (Domestic Production)	B
			Money	
m0	1959:1 - 2003:06	FMBASE	Monetary Base, Adj For Reserve Req Chgs	D
m1	1959:1 - 2005:12	FM1	Money Stock: M1	D
m2	1959:1 - 2005:12	FM2	Money Stock: M2	D
m3	1959:1 - 2005:12	FM3	Money Stock: M3	D
			Miscellaneous	
lead	1959:1 - 2005:12	G0M910	Composite index of 10 leading indicators	C

Note: Sources (S) are abbreviated as follows: B - Bureau of Labor Statistics, C - Conference Board, D - DRI Basic Economics Database, F - Federal Reserve Board of Governors. S pread is defined as the difference between rbndl and rovnght. The same names preceded by an "r" denote the real version of the variable. For example, Real Interest Rates (such as rrovnght, rrtbill, rrbnds, rrbndm, rrbndl) are defined as Nominal Interest Rates minus CPI inflation. Real stock variables such as Real Money Balances (rm0, rm1, rm2, rm3) are defined as the ratio of the Nominal Money Balances and CPI.

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

Variable		rMSFE	p-value	Variable		rMSFE	p-value
Panel A. Representative Series							
rovnght	level	-2.28	0.02	rspread	level	-3.12	0.00
hours	level	-0.69	0.49	deliveries	level	-1.72	0.09
lead	level	-3.34	0.00	m2	Δln	-2.28	0.02
Panel B. Additional Series							
rtbill	level	-1.81	0.07	unemp	Δln	-0.75	0.45
rbnds	level	-1.53	0.13	unemp	gap	-2.65	0.01
rbndm	level	-0.97	0.33	hours	Δ	1.62	0.10
rbndl	level	-0.96	0.33	deliveries	Δ	-1.85	0.06
rovnght	Δ	-0.38	0.70	cpi	Δln	-1.53	0.13
rtbill	Δ	0.28	0.78	cpi	$\Delta^2 ln$	3.40	0.00
rbnds	Δ	-0.45	0.65	pce	Δln	-1.39	0.17
rbndm	Δ	-1.13	0.26	pce	$\Delta^2 ln$	3.00	0.00
rbndl	Δ	-0.92	0.36	ppi	Δln	-0.29	0.77
rrovnght	level	-2.22	0.03	ppi	$\Delta^2 ln$	3.14	0.00
rrtbill	level	-1.81	0.07	earn	Δln	-0.64	0.52
rrbnds	level	-1.55	0.12	earn	$\Delta^2 ln$	3.48	0.00
rrbndm	level	-1.15	0.25	oil	Δln	-0.24	0.81
rrbndl	level	-0.97	0.33	oil	$\Delta^2 ln$	1.63	0.10
rrovnght	Δ	-0.21	0.83	roil	ln	0.39	0.70
rrtbill	Δ	0.13	0.89	roil	Δln	-0.05	0.96
rrbnds	Δ	-0.16	0.87	m0	Δln	1.91	0.06
rrbndm	Δ	-0.34	0.74	m0	$\Delta^2 ln$	2.88	0.00
rrbndl	Δ	0.13	0.89	m1	Δln	-0.16	0.87
exrate	Δln	0.76	0.45	m1	$\Delta^2 ln$	1.23	0.22
stockp	Δln	-2.03	0.04	m2	$\Delta^2 ln$	1.61	0.11
rstockp	Δln	-2.32	0.02	m3	Δln	-0.15	0.88
capu	level	-1.64	0.10	m3	$\Delta^2 ln$	2.62	0.01
emp	Δln	0.94	0.35	rm0	Δln	-1.64	0.10
emp	gap	-2.54	0.01	rm1	Δln	-1.49	0.14
unemp	level	-0.57	0.57	rm2	Δln	-2.80	0.01
unemp	Δ	-0.03	0.97	rm3	Δln	-2.20	0.03

Note: rMSFE denotes the re-scaled average MSFE difference over the full out-of-sample period. A negative value indicate that the model with an explanatory variable outperforms the autoregressive model); p-value is 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
Panel A. Representative Series							
rovnght	level	-2.14	0.03	rspread	level	-2.60	0.01
emp	Δln	0.50	0.62	unemp	Δ	0.85	0.39
unemp	gap	-2.23	0.03	cpi	$\Delta^2 ln$	0.98	0.33
Panel B. Additional Series							
rtbill	level	-1.72	0.08	rrbndl	level	-0.22	0.83
rbnds	level	-1.30	0.19	rrovnght	Δ	0.51	0.61
rbndm	level	-0.66	0.51	rrtbill	Δ	1.30	0.19
rbndl	level	-0.44	0.66	rrbnds	Δ	0.36	0.72
rovnght	Δ	0.26	0.79	rrbndm	Δ	0.55	0.58
rtbill	Δ	1.67	0.10	rrbndl	Δ	0.83	0.41
rbnds	Δ	0.30	0.76	exrate	Δln	1.04	0.30
rbndm	Δ	0.35	0.73	stockp	Δln	-1.66	0.10
rbndl	Δ	0.39	0.70	rstockp	Δln	-1.91	0.06
rrovnght	level	-2.03	0.04	emp	gap	-2.92	0.00
rrtbill	level	-1.73	0.08	unemp	level	-0.50	0.62
rrbnds	level	-1.28	0.20	unemp	Δln	0.23	0.82
rrbndm	level	-0.64	0.52	cpi	Δln	-1.38	0.17

Note: rMSFE denotes the re-scaled average MSFE difference over the full out-of-sample period. A negative value indicate that the model with an explanatory variable outperforms the autoregressive model); p-value is 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
Panel A. Representative Series							
rbnds	level	-1.30	0.19	rspread	level	-0.66	0.51
ip	Δln	-2.02	0.04	capu	level	-2.56	0.01
unemp	gap	-0.55	0.58	m3	Δln	0.37	0.71
Panel B. Additional Series							
rovnght	level	-1.59	0.11	unemp	Δln	-2.38	0.02
rtbill	level	-1.30	0.19	hours	level	-1.36	0.17
rbndm	level	-1.04	0.30	hours	Δ	1.17	0.24
rbndl	level	-1.06	0.29	deliveries	level	-2.35	0.02
rovnght	Δ	-0.09	0.92	deliveries	Δ	-0.92	0.36
rtbill	Δ	-0.08	0.94	pce	Δln	-0.45	0.66
rbnds	Δ	0.03	0.98	pce	$\Delta^2 ln$	2.19	0.03
rbndm	Δ	0.91	0.36	ppi	Δln	1.19	0.24
rbndl	Δ	0.29	0.77	ppi	$\Delta^2 ln$	2.39	0.02
rrovnght	level	-1.49	0.13	earn	Δln	0.84	0.40
rrtbill	level	-1.15	0.25	earn	$\Delta^2 ln$	4.30	0.00
rrbnds	level	-1.16	0.25	oil	Δln	0.67	0.50
rrbndm	level	-0.99	0.32	oil	$\Delta^2 ln$	2.39	0.02
rrbndl	level	-1.08	0.28	roil	ln	-0.96	0.34
rrovnght	Δ	-0.09	0.92	roil	Δln	1.18	0.24
rrtbill	Δ	-0.08	0.94	m0	Δln	0.06	0.95
rrbnds	Δ	0.03	0.98	m0	$\Delta^2 ln$	2.15	0.03
rrbndm	Δ	0.91	0.36	m1	Δln	-0.21	0.83
rrbndl	Δ	0.29	0.77	m1	$\Delta^2 ln$	2.07	0.04
exrate	Δln	0.18	0.86	m2	Δln	0.64	0.52
stockp	Δln	0.61	0.54	m2	$\Delta^2 ln$	1.91	0.06
rstockp	Δln	-0.10	0.92	m3	$\Delta^2 ln$	3.15	0.00
ip	gap	-0.52	0.60	rm0	Δln	-0.90	0.37
emp	Δln	-2.05	0.04	rm1	Δln	-1.07	0.28
emp	gap	-0.43	0.66	rm2	Δln	-1.30	0.19
unemp	level	-2.09	0.04	rm3	Δln	-1.36	0.18
unemp	Δ	-2.12	0.03	lead	level	-1.30	0.20

Note: rMSFE denotes the re-scaled average MSFE difference over the full out-of-sample period. A negative value indicate that the model with an explanatory variable outperforms the autoregressive model); p-value is 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	
Panel A. Representative Series											
rovnght	level	0.00	0.00	1976	5	rspread	level	0.00	0.00	1976	6
hours	level	1.00	1.00			deliveries	level	0.00	0.00	1976	5
lead	level	0.00	0.00	1975	10	m2	Δln	0.00	0.00	1977	10
Panel B. Additional Series											
rtbill	level	0.00	0.00	1976	3	unemp	Δln	1.00	0.87		
rbnds	level	0.00	0.00	1976	3	unemp	gap	0.00	0.00	1976	3
rbndm	level	0.02	0.01	1975	12	hours	Δ	0.65	0.54		
rbndl	level	0.11	0.07	1975	12	deliveries	Δ	0.19	0.37		
rovnght	Δ	0.53	0.37			cpi	Δln	0.00	0.00	1975	10
rtbill	Δ	1.00	1.00			cpi	$\Delta^2 ln$	0.01	0.02	1976	3
rbnds	Δ	1.00	1.00			pce	Δln	0.00	0.00	1975	10
rbndm	Δ	0.15	0.19			pce	$\Delta^2 ln$	0.15	0.15		
rbndl	Δ	0.14	0.17			ppi	Δln	0.73	0.66		
rrovnght	level	0.00	0.00	1976	5	ppi	$\Delta^2 ln$	0.18	0.16		
rrtbill	level	0.00	0.00	1976	3	earn	Δln	0.62	0.49		
rrbnds	level	0.00	0.00	1976	3	earn	$\Delta^2 ln$	0.05	0.28		
rrbndm	level	0.09	0.06	1975	12	oil	Δln	0.62	0.54		
rrbndl	level	0.45	0.33			oil	$\Delta^2 ln$	0.19	0.13		
rrovnght	Δ	0.79	0.65			roil	ln	0.81	0.70		
rrtbill	Δ	1.00	1.00			roil	Δln	1.00	0.87		
rrbnds	Δ	1.00	1.00			m0	Δln	1.00	1.00		
rrbndm	Δ	0.75	0.70			m0	$\Delta^2 ln$	0.71	1.00		
rrbndl	Δ	1.00	1.00			m1	Δln	0.67	0.60		
exrate	Δln	0.58	0.43			m1	$\Delta^2 ln$	1.00	1.00		
stockp	Δln	0.00	0.00	1976	7	m2	$\Delta^2 ln$	1.00	1.00		
rstockp	Δln	0.00	0.00	1976	7	m3	Δln	0.68	0.57		
capu	level	0.53	0.60			m3	$\Delta^2 ln$	0.38	0.67		
emp	Δln	1.00	1.00			rm0	Δln	0.00	0.00	1975	10
emp	gap	0.02	0.02	1976	2	rm1	Δln	0.00	0.00	1975	10
unemp	level	0.90	0.86			rm2	Δln	0.00	0.00	1975	10
unemp	Δ	1.00	1.00			rm3	Δln	0.00	0.00	1975	10

Note: The table reports p-values of Giacomini and Rossi's (2008) One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

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	
Panel A. Representative Series											
rovnght	level	0.00	0.00	1976	5	rspread	level	0.00	0.02	1984	9
emp	Δln	1.00	1.00			unemp	Δ	0.66	0.54		
unemp	gap	0.02	0.06	1984	8	cpi	$\Delta^2 ln$	0.66	0.59		
Panel B. Additional Series											
rtbill	level	0.00	0.00	1976	3	rrbndl	level	0.13	0.11		
rbnds	level	0.00	0.00	1976	3	rrovnght	Δ	0.86	0.82		
rbndm	level	0.04	0.02	1976	2	rrtbill	Δ	0.73	0.62		
rbndl	level	0.08	0.05	1976	1	rrbnds	Δ	0.51	0.38		
rovnght	Δ	1.00	1.00			rrbndm	Δ	0.78	0.64		
rtbill	Δ	0.53	0.43			rrbndl	Δ	1.00	0.86		
rbnds	Δ	0.85	0.75			exrate	Δln	1.00	1.00		
rbndm	Δ	1.00	0.89			stockp	Δln	0.08	0.06	1976	7
rbndl	Δ	0.82	0.70			rstockp	Δln	0.07	0.06	1976	7
rrovnght	level	0.00	0.00	1976	5	emp	gap	0.00	0.02	1984	8
rrtbill	level	0.00	0.00	1976	3	unemp	level	1.00	0.85		
rrbnds	level	0.00	0.00	1976	3	unemp	Δln	1.00	1.00		
rrbndm	level	0.10	0.07	1976	1	cpi	Δln	0.01	0.00	1975	9

Note: The table reports p-values of Giacomini and Rossi's (2008) One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Table 7: Forecasting Inflation: Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
Panel A. Representative Series											
rbnds	level	0.42	0.55			rspread	level	0.28	0.19		
ip	Δln	0.10	0.20			capu	level	0.00	0.00	1983	10
unemp	gap	0.79	0.67			m3	Δln	0.40	0.28		
Panel B. Additional Series											
rovnght	level	0.39	0.57			unemp	Δln	0.00	0.03	1984	8
rtbill	level	0.50	0.63			hours	level	0.14	0.19		
rbndm	level	0.73	0.80			hours	Δ	1.00	1.00		
rbndl	level	0.76	0.83			deliveries	level	0.00	0.02	1983	8
rovnght	Δ	0.80	0.70			deliveries	Δ	0.16	0.11		
rtbill	Δ	1.00	1.00			pce	Δln	0.83	0.78		
rbnds	Δ	1.00	1.00			pce	$\Delta^2 ln$	0.58	0.78		
rbndm	Δ	1.00	1.00			ppi	Δln	1.00	1.00		
rbndl	Δ	1.00	1.00			ppi	$\Delta^2 ln$	0.31	0.26		
rrovnght	level	0.45	0.62			earn	Δln	1.00	0.83		
rrtbill	level	0.60	0.70			earn	$\Delta^2 ln$	0.00	0.07	1983	9
rrbnds	level	0.51	0.61			oil	Δln	1.00	0.83		
rrbndm	level	0.79	0.84			oil	$\Delta^2 ln$	0.18	0.60		
rrbndl	level	0.76	0.83			oil	ln	0.47	0.48		
rrovnght	Δ	0.80	0.70			roil	Δln	0.51	0.38		
rrtbill	Δ	1.00	1.00			m0	Δln	0.80	0.70		
rrbnds	Δ	1.00	1.00			m0	$\Delta^2 ln$	0.65	0.80		
rrbndm	Δ	1.00	1.00			m1	Δln	1.00	0.88		
rrbndl	Δ	1.00	1.00			m1	$\Delta^2 ln$	0.37	0.57		
exrate	Δln	1.00	0.83			m2	Δln	1.00	1.00		
stockp	Δln	1.00	1.00			m2	$\Delta^2 ln$	0.48	0.64		
rstockp	Δln	0.89	0.79			m3	$\Delta^2 ln$	0.06	0.26		
ip	gap	0.82	0.69			rm0	Δln	0.32	0.34		
emp	Δln	0.00	0.03	1983	10	rm1	Δln	0.46	0.47		
emp	gap	0.82	0.70			rm2	Δln	0.70	0.86		
unemp	level	0.00	0.04	1983	10	rm3	Δln	0.29	0.31		
unemp	Δ	0.03	0.08	1984	7	lead	level	0.04	0.06	1984	6

Note: The table reports p-values of Giacomini and Rossi's (2008) One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Table 8: Forecasting Output Growth (Recursive Lag Selection): Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
Panel A. Representative Series											
rovnght	level	0.00	0.00	1976	5	rspread	level	0.00	0.00	1976	6
hours	level	0.59	0.42			deliveries	level	0.00	0.00	1976	6
lead	level	0.00	0.00	1975	10	m2	Δln	0.00	0.00	1977	10
Panel B. Additional Series											
rtbill	level	0.00	0.00	1976	3	unemp	Δln	0.38	0.27		
rbnds	level	0.00	0.00	1976	3	unemp	gap	0.00	0.00	1976	3
rbndm	level	0.02	0.01	1975	12	hours	Δ	0.74	0.63		
rbndl	level	0.02	0.01	1975	12	deliveries	Δ	0.20	0.31		
rovnght	Δ	0.70	0.54			cpi	Δln	0.00	0.00	1975	10
rtbill	Δ	0.86	0.78			cpi	$\Delta^2 ln$	0.68	0.63		
rbnds	Δ	1.00	1.00			pce	Δln	0.02	0.01	1975	10
rbndm	Δ	1.00	0.90			pce	$\Delta^2 ln$	0.46	0.36		
rbndl	Δ	0.57	0.40			ppi	Δln	1.00	0.89		
rrovnght	level	0.00	0.00	1976	5	ppi	$\Delta^2 ln$	0.17	0.15		
rrtbill	level	0.00	0.00	1976	3	earn	Δln	1.00	0.85		
rrbnds	level	0.00	0.00	1976	3	earn	$\Delta^2 ln$	0.07	0.35		
rrbndm	level	0.09	0.06	1975	12	oil	Δln	1.00	1.00		
rrbndl	level	0.35	0.23			oil	$\Delta^2 ln$	0.00	0.00	1975	12
rrovnght	Δ	1.00	1.00			roil	ln	1.00	0.87		
rrtbill	Δ	0.87	0.85			roil	Δln	0.85	0.74		
rrbnds	Δ	1.00	1.00			m0	Δln	0.07	0.02	1999	4
rrbndm	Δ	1.00	1.00			m0	$\Delta^2 ln$	0.11	0.03	1999	4
rrbndl	Δ	1.00	1.00			m1	Δln	0.54	0.48		
exrate	Δln	0.23	0.15			m1	$\Delta^2 ln$	1.00	1.00		
stockp	Δln	0.00	0.00	1976	7	m2	$\Delta^2 ln$	1.00	1.00		
rstockp	Δln	0.00	0.00	1976	7	m3	Δln	0.69	0.57		
capu	level	0.23	0.31			m3	$\Delta^2 ln$	0.25	0.18		
emp	Δln	1.00	0.83			rm0	Δln	0.01	0.00	1975	12
emp	gap	0.01	0.01	1976	5	rm1	Δln	0.00	0.00	1975	10
unemp	level	1.00	1.00			rm2	Δln	0.00	0.00	1975	10
unemp	Δ	0.31	0.21			rm3	Δln	0.00	0.00	1975	10

Note: The table reports p-values of Giacomini and Rossi's (2008) One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Table 9: Forecasting Inflation (Recursive Lag Selection): Tests of Equal Predictive Ability Over Time

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
Panel A. Representative Series											
rbnds	level	0.66	0.72			rspread	level	1.00	0.89		
ip	Δln	0.13	0.17			capu	level	0.02	0.03	1983	10
unemp	gap	1.00	1.00			m3	Δln	0.45	0.34		
Panel B. Additional Series											
rovnght	level	1.00	1.00			unemp	Δln	0.04	0.08	1984	3
rtbill	level	0.75	0.80			hours	level	0.74	0.70		
rbndm	level	1.00	1.00			hours	Δ	1.00	1.00		
rbndl	level	1.00	1.00			deliveries	level	0.11	0.16		
ovnght	Δ	1.00	1.00			deliveries	Δ	0.33	0.34		
rtbill	Δ	1.00	1.00			pce	Δln	1.00	1.00		
rbnds	Δ	1.00	1.00			pce	$\Delta^2 ln$	0.40	0.63		
rbndm	Δ	0.27	0.50			ppi	Δln	1.00	0.87		
rbndl	Δ	1.00	1.00			ppi	$\Delta^2 ln$	0.46	0.40		
rrovnght	level	1.00	1.00			earn	Δln	1.00	1.00		
rrtbill	level	1.00	1.00			earn	$\Delta^2 ln$	0.00	0.14		
rrbnds	level	0.81	0.82			oil	Δln	0.65	0.51		
rrbndm	level	1.00	1.00			oil	$\Delta^2 ln$	0.77	0.89		
rrbndl	level	1.00	1.00			roil	ln	1.00	1.00		
rrovnght	Δ	1.00	1.00			roil	Δln	0.56	0.44		
rrtbill	Δ	1.00	1.00			m0	Δln	1.00	0.80		
rrbnds	Δ	1.00	1.00			m0	$\Delta^2 ln$	0.76	0.56		
rrbndm	Δ	0.74	0.82			m1	Δln	0.86	0.82		
rrbndl	Δ	1.00	1.00			m1	$\Delta^2 ln$	0.29	0.49		
exrate	Δln	0.73	0.62			m2	Δln	1.00	1.00		
stockp	Δln	0.83	1.00			m2	$\Delta^2 ln$	0.17	0.32		
rstockp	Δln	1.00	1.00			m3	$\Delta^2 ln$	0.65	0.82		
ip	gap	0.49	0.34			rm0	Δln	0.24	0.26		
emp	Δln	0.12	0.14			rm1	Δln	0.44	0.45		
emp	gap	1.00	0.88			rm2	Δln	0.47	0.69		
unemp	level	0.21	0.23			rm3	Δln	0.11	0.25		
unemp	Δ	0.08	0.10	1975	12	lead	level	0.11	0.13		

Note: The table reports p-values of Giacomini and Rossi's (2008) One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Table 10: Forecasting Output Growth: Tests of Equal Predictive Ability Over Time (Wald test)

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
Panel A. Representative Series											
rovnght	level	0.02	0.00	1976	5	rspread	level	0.00	0.00	1976	6
hours	level	1.00	1.00			deliveries	level	0.07	0.00	1976	5
lead	level	0.02	0.00	1975	10	m2	Δln	0.05	0.00	1977	10
Panel B. Additional Series											
rtbill	level	0.00	0.00	1976	3	unemp	Δln	1.00	0.87		
rbnds	level	0.01	0.00	1976	3	unemp	gap	0.07	0.00	1976	3
rbndm	level	0.22	0.01	1975	12	hours	Δ	1.00	0.54		
rbndl	level	0.52	0.07	1975	12	deliveries	Δ	0.54	0.37		
rovnght	Δ	0.84	0.37			cpi	Δln	0.23	0.00	1975	10
rtbill	Δ	1.00	1.00			cpi	$\Delta^2 ln$	0.81	0.02	1976	3
rbnds	Δ	1.00	1.00			pce	Δln	0.29	0.00	1975	10
rbndm	Δ	0.45	0.19			pce	$\Delta^2 ln$	1.00	0.15		
rbndl	Δ	0.40	0.17			ppi	Δln	1.00	0.66		
rrovnght	level	0.02	0.00	1976	5	ppi	$\Delta^2 ln$	0.81	0.16		
rrtbill	level	0.00	0.00	1976	3	earn	Δln	0.88	0.49		
rrbnds	level	0.02	0.00	1976	3	earn	$\Delta^2 ln$	0.58	0.28		
rrbndm	level	0.53	0.06	1975	12	oil	Δln	1.00	0.54		
rrbndl	level	0.89	0.33			oil	$\Delta^2 ln$	1.00	0.13		
rrovnght	Δ	0.34	0.65			roil	ln	1.00	0.70		
rrtbill	Δ	0.81	1.00			roil	Δln	1.00	0.87		
rrbnds	Δ	1.00	1.00			m0	Δln	1.00	1.00		
rrbndm	Δ	0.90	0.70			m0	$\Delta^2 ln$	1.00	1.00		
rrbndl	Δ	1.00	1.00			m1	Δln	0.61	0.60		
exrate	Δln	1.00	0.43			m1	$\Delta^2 ln$	1.00	1.00		
stockp	Δln	0.05	0.00	1976	7	m2	$\Delta^2 ln$	1.00	1.00		
rstockp	Δln	0.07	0.00	1976	7	m3	Δln	1.00	0.57		
capu	level	0.61	0.60			m3	$\Delta^2 ln$	1.00	0.67		
emp	Δln	1.00	1.00			rm0	Δln	0.05	0.00	1975	10
emp	gap	0.25	0.02	1976	2	rm1	Δln	0.10	0.00	1975	10
unemp	level	1.00	0.86			rm2	Δln	0.00	0.00	1975	10
unemp	Δ	1.00	1.00			rm3	Δln	0.02	0.00	1975	10

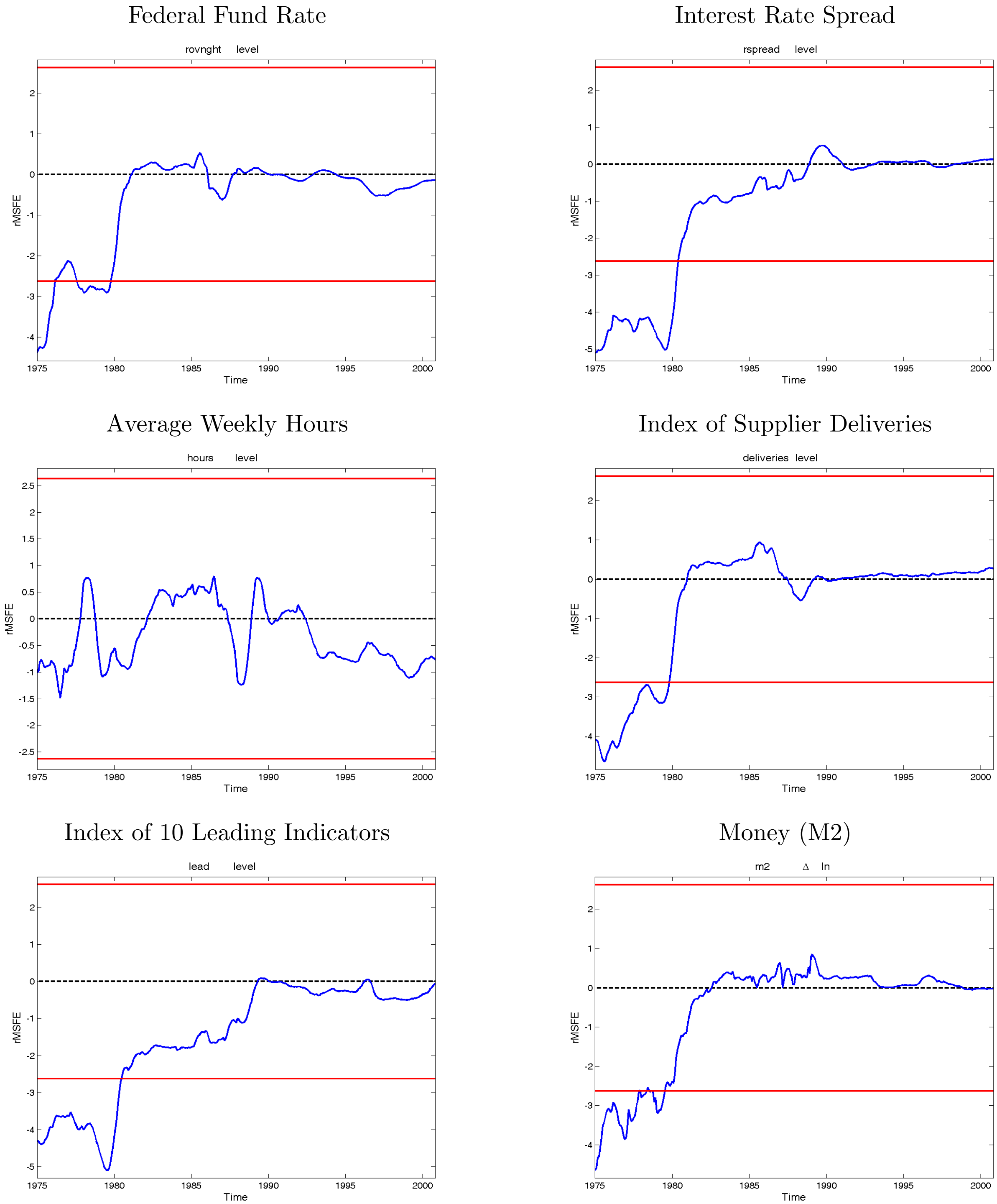
Note: The table reports p-values of Giacomini and Rossi's (2008) Wald version of the One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Table 11: Forecasting Inflation: Tests of Equal Predictive Ability Over Time (Wald test)

Variable		One-time	Break	Break Date		Variable		One-time	Break	Break Date	
Panel A. Representative Series											
rbnds	level	0.67	0.55			rspread	level	0.46	0.19		
ip	Δln	0.55	0.20			capu	level	0.02	0.00	1983	10
unemp	gap	1.00	0.67			m3	Δln	0.47	0.28		
Panel B. Additional Series											
rovnght	level	0.68	0.57			unemp	Δln	0.13	0.03	1984	8
rtbill	level	0.76	0.63			hours	level	0.59	0.19		
rbndm	level	0.89	0.80			hours	Δ	1.00	1.00		
rbndl	level	0.84	0.83			deliveries	level	0.08	0.02	1983	8
rovnght	Δ	0.82	0.70			deliveries	Δ	0.70	0.11		
rtbill	Δ	1.00	1.00			pce	Δln	1.00	0.78		
rbnds	Δ	1.00	1.00			pce	$\Delta^2 ln$	1.00	0.78		
rbndm	Δ	1.00	1.00			ppi	Δln	1.00	1.00		
rbndl	Δ	1.00	1.00			ppi	$\Delta^2 ln$	1.00	0.26		
rrovnght	level	0.73	0.62			earn	Δln	1.00	0.83		
rrtbill	level	0.82	0.70			earn	$\Delta^2 ln$	0.03	0.07	1983	9
rrbnds	level	0.74	0.61			oil	Δln	0.67	0.83		
rrbndm	level	1.00	0.84			oil	$\Delta^2 ln$	0.55	0.60		
rrbndl	level	0.84	0.83			roil	ln	0.77	0.48		
rrovnght	Δ	0.82	0.70			roil	Δln	0.70	0.38		
rrtbill	Δ	1.00	1.00			m0	Δln	0.90	0.70		
rrbnds	Δ	1.00	1.00			m0	$\Delta^2 ln$	1.00	0.80		
rrbndm	Δ	1.00	1.00			m1	Δln	1.00	0.88		
rrbndl	Δ	1.00	1.00			m1	$\Delta^2 ln$	1.00	0.57		
exrate	Δln	0.84	0.83			m2	Δln	1.00	1.00		
stockp	Δln	1.00	1.00			m2	$\Delta^2 ln$	1.00	0.64		
rstockp	Δln	1.00	0.79			m3	$\Delta^2 ln$	0.65	0.26		
ip	gap	0.56	0.69			rm0	Δln	0.58	0.34		
emp	Δln	0.20	0.03	1983	10	rm1	Δln	0.82	0.47		
emp	gap	1.00	0.70			rm2	Δln	0.81	0.86		
unemp	level	0.23	0.04	1983	10	rm3	Δln	0.48	0.31		
unemp	Δ	0.29	0.08	1984	7	lead	level	0.15	0.06	1984	6

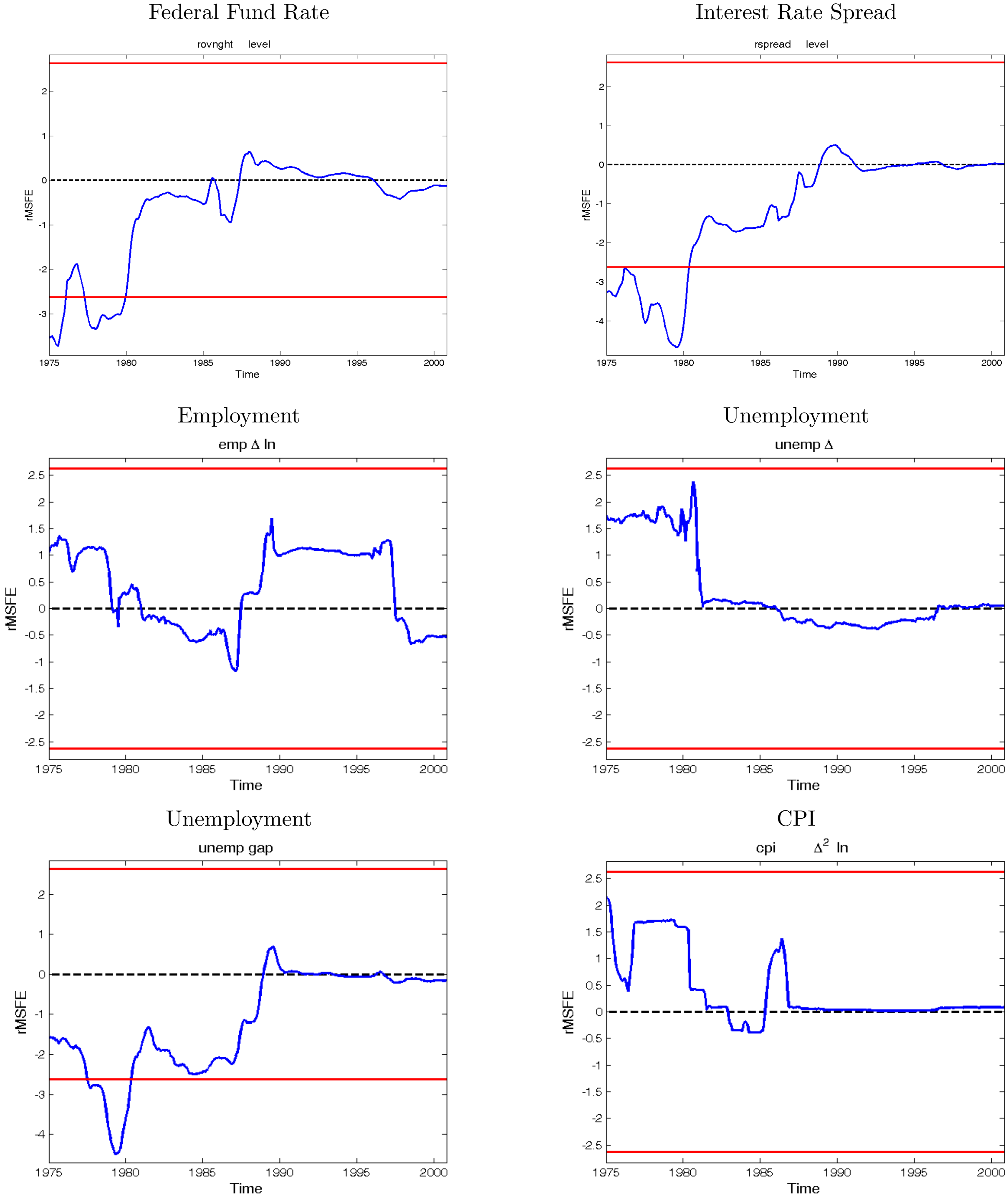
Note: The table reports p-values of Giacomini and Rossi's (2008) Wald version of the One-time Reversal ("One-time") and $\sup_t LM_2(t)$ ("Break") tests. The estimated break dates are reported if significant at least at 10% level.

Figure 1: Forecasting US output growth over time



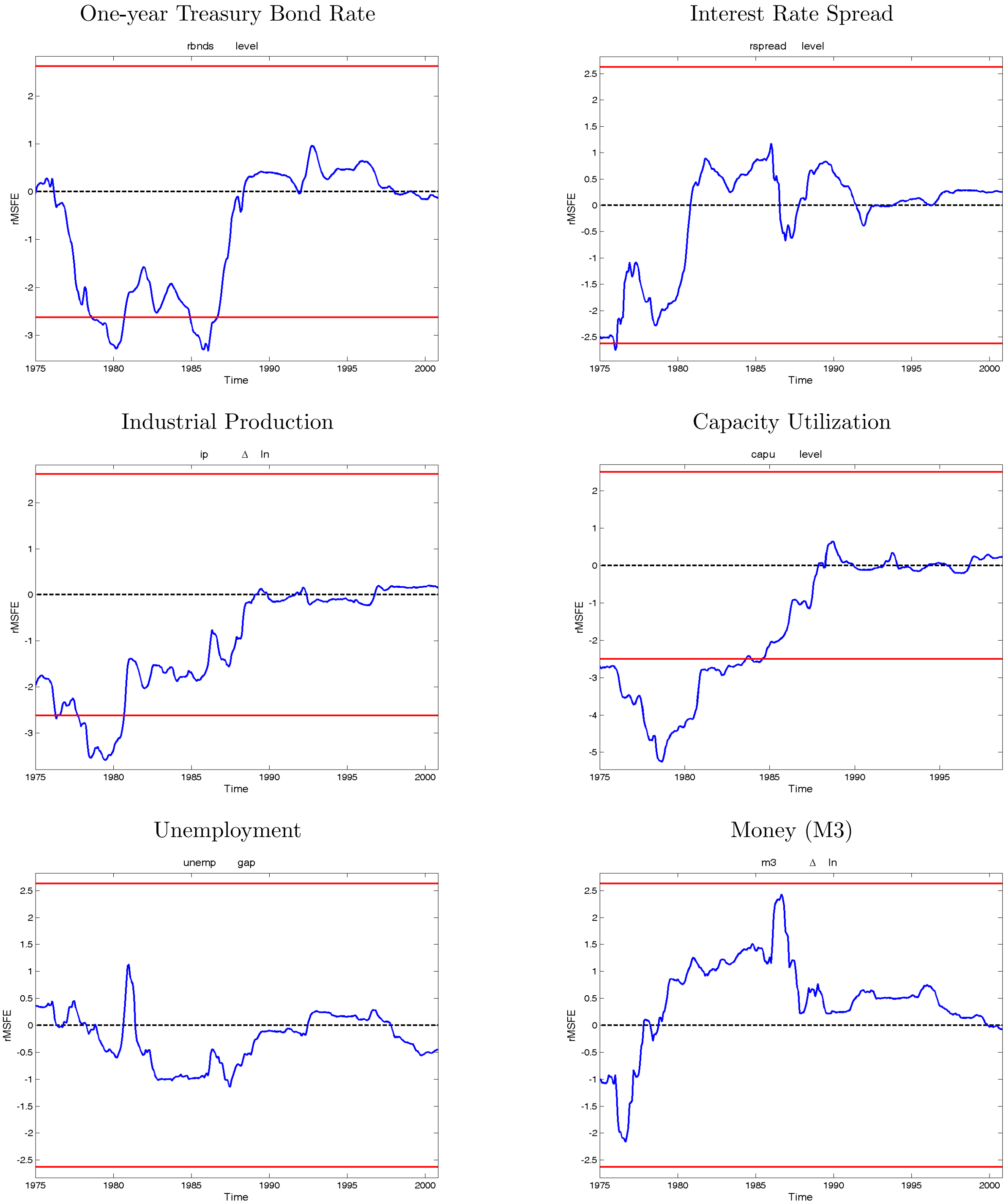
Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 2: Forecasting US output growth over time using real-time data



Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F^{OOS}_{t,m}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

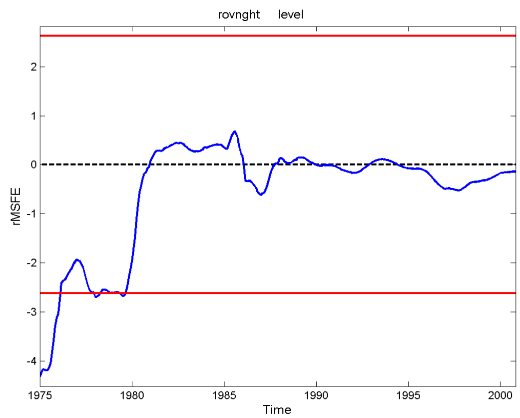
Figure 3: Forecasting US inflation over time



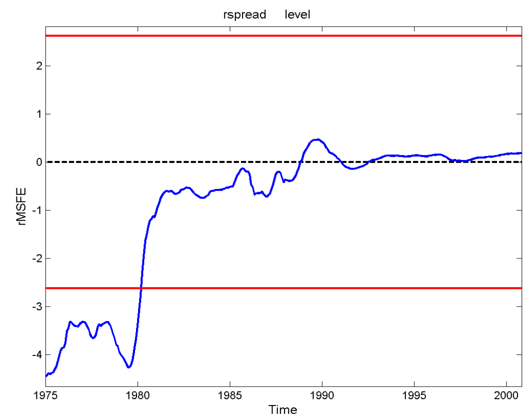
Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 4: Forecasting US output growth over time (recursive lag selection)

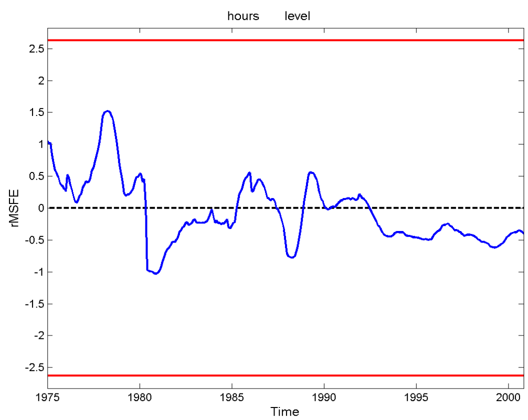
Federal Fund Rate



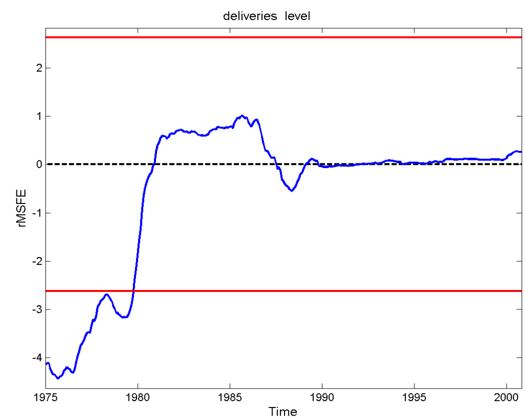
Interest Rate Spread



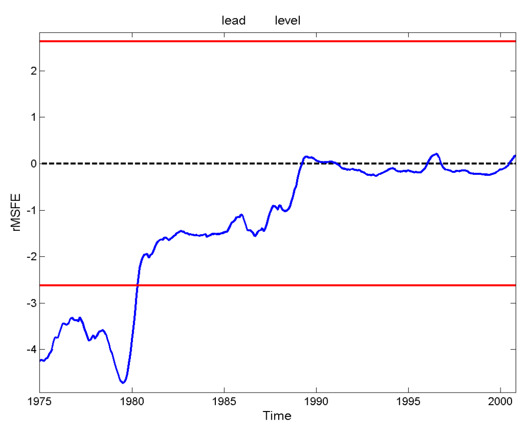
Average Weekly Hours



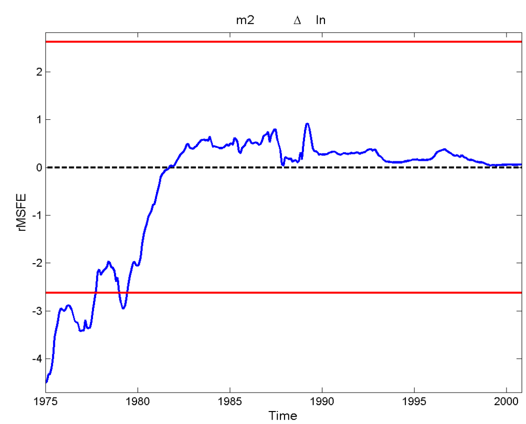
Index of Supplier Deliveries



Index of 10 Leading Indicators

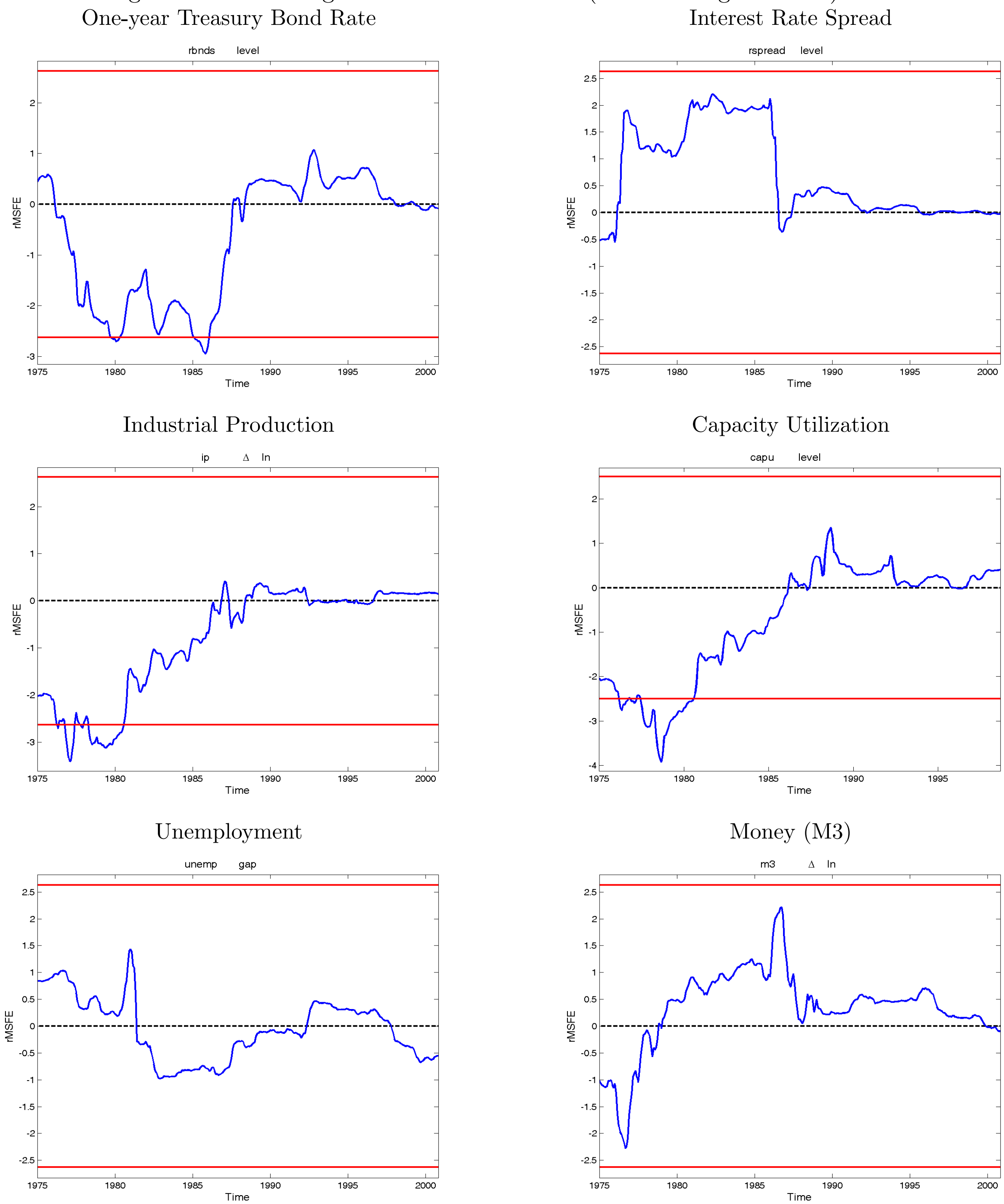


Money (M2)



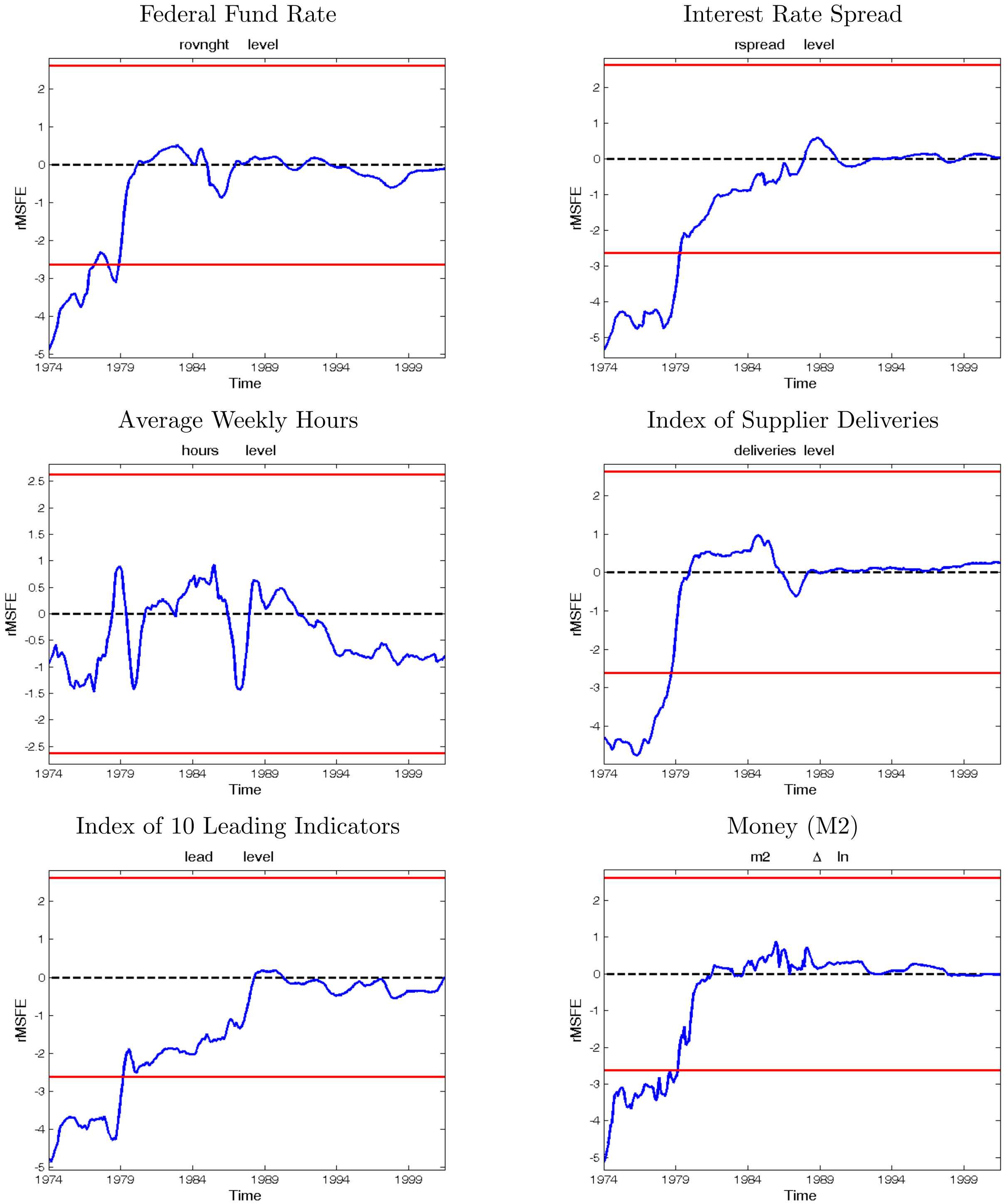
Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 5: Forecasting US inflation over time (recursive lag selection)



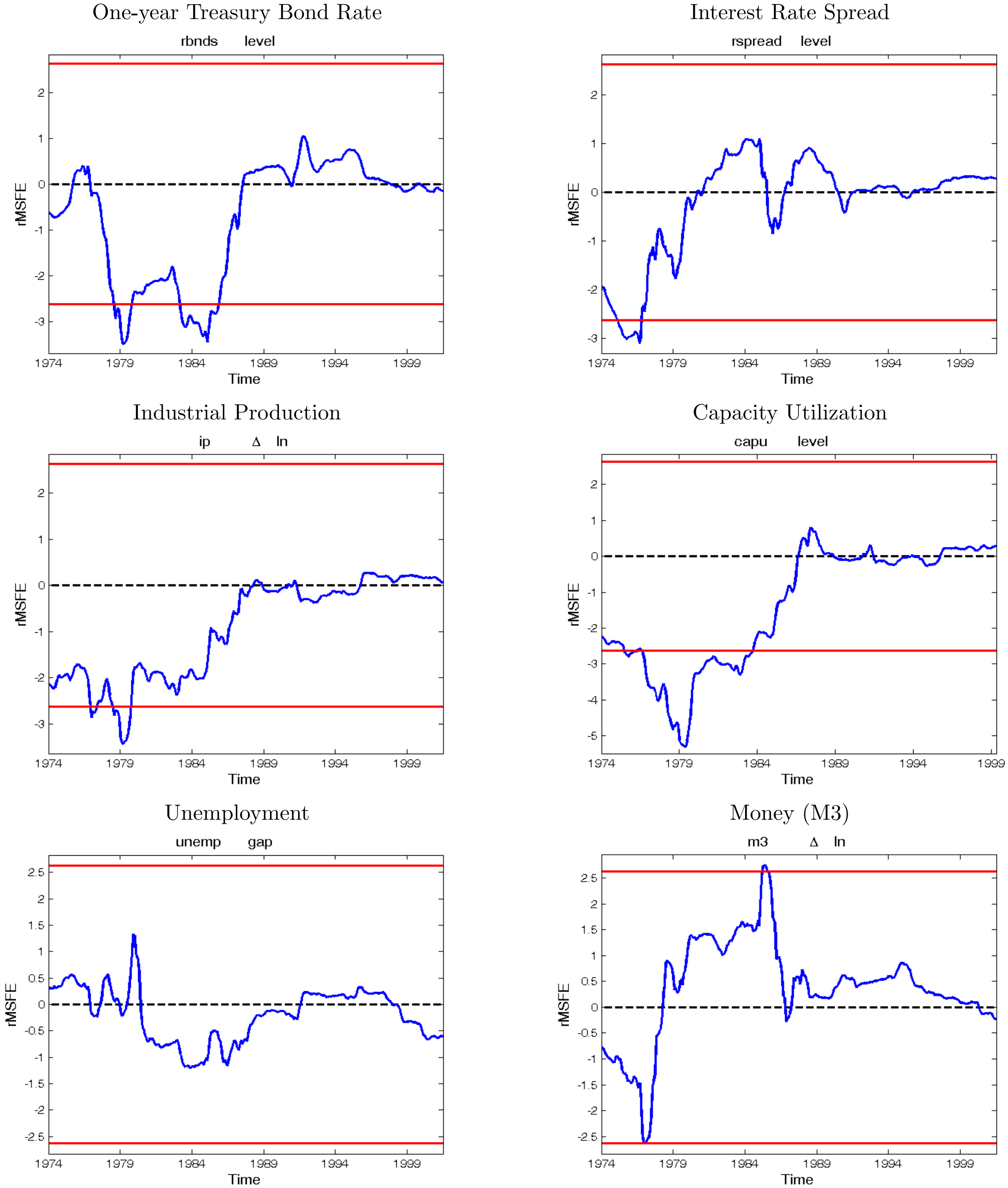
Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F^{OOS}_{t,m}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 6: Forecasting Output Growth ($m = 100$, full sample lag selection)



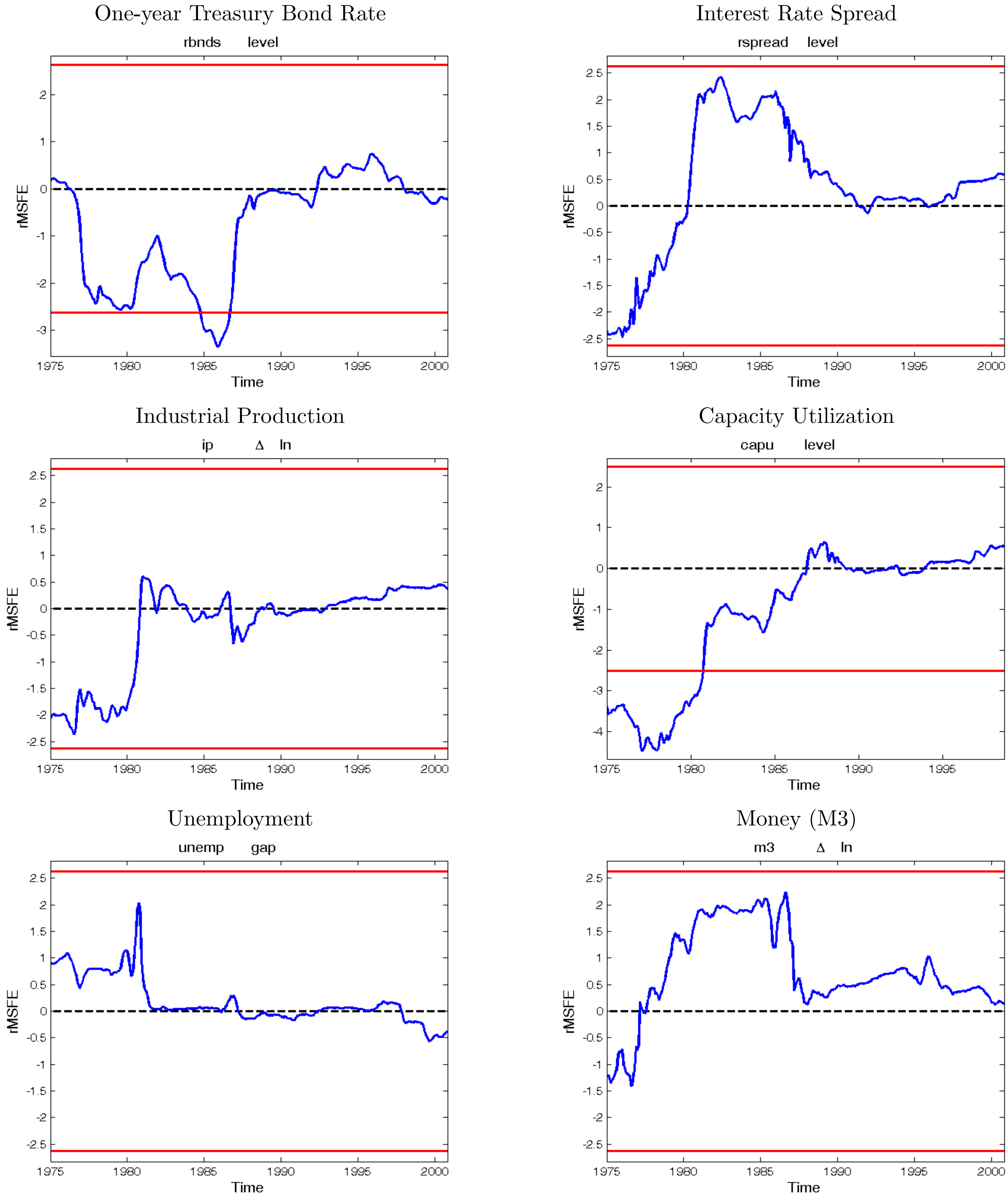
Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1} m^{-1/2} rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 7: Forecasting Inflation ($m = 100$, full sample lag selection)



Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.

Figure 8: Forecasting Consumption Deflator Based Inflation



Note: The dark solid line in the figure reports the re-scaled rMSFEs, $\widehat{\sigma}^{-1}m^{-1/2}rMSFE_t$, that is $F_{t,m}^{OOS}$. The light solid lines report 90% bands for testing the null hypothesis that the models' relative forecasting performance is equal (the test rejects when the dark solid line is outside the bands). Negative values of the re-scaled rMSFE denote situations in which the economic model forecasts better than its competitor.