PRELIMINARY AND INCOMPLETE DRAFT
COMMENTS WELCOME

Reconsidering the Optimality of Federal Reserve Forecasts

Michael S. Hanson*

Wesleyan University

and

Jayson Whitehorn†

Epic Systems Corporation

Current version: August 5, 2006

Abstract

We investigate the econometric properties of the Federal Reserve Greenbook forecasts with an integrated real-time database of U.S. macroeconomic data that more precisely characterizes the information sets available to Fed policy makers in advance of the Open Market Committee meetings. Our data set associates historical vintages of NIPA and labor market data with the exact dates of the Greenbook forecasts, allowing us to examine multivariate relationships among the real-time forecast errors for inflation, output growth and the unemployment rate. For most forecast horizons and subsamples we examine, the hypotheses of unbiasedness and efficiency can be individually and jointly rejected. Our results suggest that shifts in lower-frequency components of the forecasted series are an important source for the rejection of forecast rationality.

JEL Categories: C53, C82, E58.

Keywords: Greenbook forecasts, real-time data analysis, monetary policy.

* Corresponding author. Department of Economics, 238 Church Street, Middletown, CT 06459. Phone: (860) 685-4634. FAX: (860) 685-2301. E-mail: `mshanson@wesleyan.edu`

† Wesleyan graduate, class of 2005. E-mail: `jwhitehorn@wesleyan.edu`

Thanks to Manolis Kaparakis and Michael Lovell, and seminar participants at the 2006 Summer Econometric Society Annual Meetings, for helpful comments. Thanks also to Tom Stark at the Philadelphia Fed, Virginia Mannering at the BLS, James Walker at the BEA, and Yu Hung Lin at the New York Fed for assistance with data collection. Stephen Morris provided able research assistance. Any remaining errors are the authors' sole responsibility. Please do not cite this preliminary draft without the authors' prior consent.

1 Introduction

> "The success of monetary policy depends importantly on the quality of forecasting."
>
> — Alan Greenspan (2004)

As the above quote attests, questions about the efficacy of monetary policy should be conditioned on the accuracy of the forecasts upon which policy is based. In this paper we investigate the properties of the Greenbook forecasts made by the staff economists at the Federal Reserve Board. We collect a unique data set that allows us to determine the precise date that real-time macroeconomic information became known to the Fed forecasters. Regression analysis with these data reveals widespread evidence against rationality: the forecasts tend not to make efficient use of information available prior to the Greenbook publication dates, and the forecasts exhibit a systematic pattern of over- and under-prediction in sub-samples. The specific nature of these rejections suggests that unanticipated lower-frequency shifts in the data generating processes for U.S. macroeconomic data may account for many of our findings.

Existing research has not reached a consensus on whether the Greenbook forecasts are rational forecasts, in part because there does not appear to be much consensus on how to approach this issue.[1] In an early contribution, Jansen and Kishen (1996) examine the Greenbook forecasts for inflation, real output growth, and the unemployment rate, made between December 1965 and December 1986. They find evidence against the rationality of the unemployment rate and output growth forecasts, but not the inflation forecasts. Romer and Romer (2000) also do not reject the null hypothesis of rationality for the Greenbook inflation forecasts between November 1965 and November 1991.[2] Additionally, they find that the Greenbook forecasts encompass commercial inflation forecasts.

Romer and Romer (2002) extend their sample through the end of 1996 and also investigate the rationality of the forecasts of the unemployment rate. Again they conclude that the Greenbook forecasts are rational. Subsequent researchers, including Sims (2002), Gavin and Mandal (2003), and Clements et al. (2004), also tend to find evidence supporting the rationality of the Greenbook forecasts for one or more of the series examined by Jansen and Kishen (1996).

Joutz and Stekler (2000), on the other hand, conclude that while the Greenbook forecasts of inflation

[1] A fair degree of variation exists in this literature with regard to the forecasted series that are investigated, the sample period, the forecast horizon(s), the frequency of the data, the specification of the Fed's information set, and the econometric methods to be used.

[2] Romer and Romer (2000) report results for equations of the form of equation (2) below. In a footnote they claim that they find similar conclusions when lagged values and lagged forecast errors were added to their equations.

and real output growth are on average unbiased, the staff has systematically over- or under-predicted inflation around business cycle turning points, and has not forecasted better than the private sector in this dimension.[3] Moreover, they find evidence that the inflation forecasts did not make efficient use of lagged information known at the time of the Greenbook publications.

More recently, Capistrán-Carmona (2005) notes that quarterly forecasts of inflation tend to be serially correlated, and that the Greenbook forecasts appear biased if one focuses on specific sub-samples: the inflation rate tends to have been under-predicted from 1968 to 1979, and over-predicted from 1979 to 1998. Capistrán-Carmona interprets his findings as evidence of an asymmetric loss function for inflation at the Federal Reserve. As we discuss later, this particular explanation does not appear consistent with either the forecasting process or the observed data.

In this study we make several contributions to the literature we have just summarized. First, as we discuss in section 2, our data set offers several improvements over those used in the above-mentioned studies. Second, we undertake a comprehensive set of tests of both the unbiasedness and efficiency of the Greenbook forecasts for inflation, real output growth, and the unemployment rate in section 3. We give special consideration to sub-sample variation in forecast performance. Third, in section 4 we examine several potential explanations for the observed rejection of rationality of the forecasts. We find that systematic deviations from rationality are associated with changes in low-frequency trends in the data. Other candidate explanations are not as plausible. Section 5 concludes with a discussion of the implications of our analysis for understanding the historical performance of U.S. monetary policy.

2 Data Description

For our analysis in this paper we have constructed a database of real-time U.S. macroeconomic data releases and Federal Reserve Greenbook forecasts. One novel aspect of our database is that we have carefully catalogued (from original sources) the exact release date of each piece of real-time information, including all revisions to historical data at the time, and integrated this information with the published data in the Greenbook forecasts. Thus, we have a fairly accurate representation of the information set of the Fed staff economists on the date that each Greenbook was published. This timing information plays a critical role in assuring that our tests of forecast rationality in section 3 are as accurate as possible given

[3] Joutz and Stekler (2000) use a sample from November 1965 through June 1989, and also investigate 5-year sub-samples within this range.

the historical record.

This data set yields a number of advantages for our analysis relative to similar, previously-published studies. First, we effectively have data at a daily frequency for the Greenbook forecasts, which themselves are published between a semi-quarterly and monthly frequency at different times in our sample. Some other studies (for example, Capistrán-Carmona, 2005) have constructed quarterly measures that require either discarding nearly one-half of the forecasts or creating average measures, both of which would likely compromise tests of forecast rationality. Second, because of the exact timing information, we need not make guesses or assumptions about whether a specific release was available to Fed staff economists prior to any given Greenbook forecast. In particular, we can determine when the staff has access to the current month's NIPA revisions, as well as instances in which the staff's information set is a full quarter behind the one-quarter release lag.

Third, we have data on a cross-section of series being forecasted: inflation, real output growth, and the unemployment rate. Thus we can account for the relationships among current and lagged forecast values. Additionally, we have a "term structure" of forecasts, as most series are forecasted several quarters in advance.[4] Finally, we have a longer sample period than most of the above-cited studies — over three decades — which allows for a more accurate characterization of the forecasting relationships over time.

2.1 Real-Time Macroeconomic Data

As macroeconomists have begun to recognize the benefits of working with real-time data, especially when evaluating the historical performance of policy makers, there has been an explosion of research utilizing such data sets.[5] As our interest is the evaluation of the Greenbook forecasts, we collected all of our data from original sources to construct a higher-frequency real-time database with exact release dates. Back issues of the *Survey of Current Business* served as our primary source for both NIPA data produced by the BEA and unemployment data produced by the BLS, supplemented by issues of the *Business Conditions Digest* and *Economic Indicators* (both Commerce Department publications). We recorded a complete vintage of available data for each release date of NIPA or employment data.

For further processing the data was then divided into three releases: advance (the initial release of

[4]This structure is not constant throughout our sample; in particular, longer horizon forecasts are much more prevalent in later years.

[5]Much of this research uses the quarterly-vintage real-time data set for macroeconomists (RTDSM) developed at the Philadelphia Fed. See, e.g., Croushore and Stark (2001) for a detailed discussion of these data.

NIPA data for the preceding quarter), preliminary (the first revision, made in the middle month of a quarter), and final (the second revision, made in the final month of a quarter).[6] Annual revisions of the previous few years of data occur each July, and benchmark revisions occur roughly every five years.

We arranged the unemployment data in a similar pattern, although we note that the release schedule is different: the final release follows one month after the preliminary release, and there is a subsequent revision the following January. Throughout most of our sample the Greenbook forecasts report the quarterly average unemployment rate, which we construct from our database in an analogous manner.

Real output growth and the inflation rate are computed as annualized one-quarter changes in the real-time data from the BEA. Real output is measured as GNP until the end of 1991, and GDP afterwards. The inflation rate is computed similarly as the annualized one-quarter change in the GNP deflator through 1991, and the growth in the GDP deflator afterwards.[7] In 1996, the BEA switched to a chain-weighted method for computing real GDP and its deflator. We construct our database such that the NIPA concept represented by the h-period ahead forecast is always compared with the appropriate realization h-periods ahead.

2.2 Greenbook Forecasts

Scanned copies of the Greenbook forecasts were graciously provided to us by Tom Stark of the Philadelphia Fed. Due to the uneven quality of the digital images, the data were transcribed by hand and checked both against other Federal Reserve publications (such as the publicly available minutes of the FOMC) as well as for internal consistency.[8] We constructed two databases of Greenbook data: a "verbatim" one and a "corrected" one, the latter of which we utilize in this study. The adjustments made to the "corrected" database relative to the "verbatim" one include correcting typographical errors and entering current or lagged data that was known to be available to the Fed from our release-dated real-time database.

Greenbook forecasts were originally produced at a monthly frequency, with a few exceptions, starting in November 1965. The shift to the current system of eight Greenbook publications per year, released roughly one week prior to FOMC meetings, occurred by 1981. Because the frequency of most Greenbook publications is neither monthly (as the unemployment rate data) nor quarterly (as the NIPA data), we

[6] The release schedule was different prior to the middle of 1974; we accommodate these differences in order to maintain the dating relevant for determining the Greenbook-dated information sets.

[7] All growth rates are computed as compounded quarterly changes, expressed as an annual rate to one decimal.

[8] Additionally, we verified our data against the data appendix of Romer and Romer (2000) and the Philadelphia Fed's quarterly Greenbook Excel file for those observation dates that coincide.

constructed our database around the individual dates of the Greenbook publications, associating each publication with a specific vintage of macroeconomic variables known at that date.

The forecast horizon of the Greenbooks has gradually lengthened over time, with four-quarter ahead forecasts available fairly regularly since the mid-1970s and six-quarter ahead forecast available on a regular basis since late 1988. As a consequence, we place most credence on the forecast horizons of four quarters or less, particularly for the earlier sub-sample periods, when relatively few observations are available at the longer forecast horizons. In addition, these shorter horizon forecasts are less likely to have been influenced by subsequent changes in monetary policy, given the "long and variable" lags for the effects of policy upon the economy.

2.3 Forecast Error and Sub-sample Definitions

In this paper we define the forecast error as the difference between the Greenbook forecast made at date t of a variable h-periods into the future, and the realized value of that variable at date $t+h$ from our above-described real-time database: $v_{t+h|t} \equiv x^f_{t+h|t} - x_{t+h}$. Thus, a positive forecast error represents an "over-prediction" — that is, $x^f_{t+h|t} > x_{t+h}$ — while a negative value is an "under-prediction." We presume the object to be forecasted, x_{t+h}, is the final release of the BEA data for inflation and output growth, and the quarterly average unemployment rate.[9]

Figures 1 through 3 plot the actual value of the variable of interest (the thin blue line) against the one-quarter-ahead forecast of that variable (the thick red line) in the top panel, and the corresponding forecast errors for the contemporaneous quarter forecast in the bottom panel (in green). In each panel we have shaded in grey the "Volcker disinflation," from October 1979 though December 1982. A large amount of academic research has investigated whether monetary policy may have changed before and after this disinflation period, and the potential impacts such changes may have had upon the variables in our study. Below we investigate how the performance of the Greenbook forecasts varies between these two sub-samples. Notice that the forecast errors are particularly large during the disinflation period.

The vertical dashed lines denote changes in the chairmanship of the Federal Reserve Board. Our sample begins in late 1965, with William McChesney Martin as chair. Arthur Burns assumed the chair-

[9]We have also evaluated the Greenbook forecasts with respect to the advance and preliminary releases, with qualitative similar results. We choose the final release as it is generally the most complete measure of the actual state of the economy for the forecast horizons we consider. It also avoids the complication of whether the Fed should be expected to forecast much later benchmark revisions. Note that Romer and Romer (2000) use the final release in their study as well.

manship in February 1970, followed by G. William Miller in March 1978. Paul Volcker was appointed chair in August 1979, and Alan Greenspan was chair from August 1987 through the end of our sample. Below we also conduct our analysis on sub-sample periods that correspond with the tenure of individual Fed chairs; as there are not enough observations for either the Martin or Miller sub-periods to be analyzed separately, we group Martin, Burns and Miller into a "pre-Volcker" sub-sample. In practice this definition implies relatively small differences between the "pre-Volcker" and "pre-disinflation" results.

2.4 Summary Statistics

Table 1 lists the average value of the forecast errors for inflation, real output growth, and the unemployment rate, respectively, for each of the sub-samples described above. For inflation, the forecast errors computed over the full sample generally are not statistically discernible from zero. However, prior to 1980, the average forecast error for inflation is negative and statistically significant, while after 1980 the average forecast error is positive and statistically significant. The converse pattern is observed for the forecast errors for real output growth: during the pre-disinflation sub-sample the Greenbook forecasts over-predict real output growth, while they under-predict in the post-disinflation sub-sample. However, not all the underpredictions are statistically discernible from zero in the post-disinflation (i.e. Volcker and Greenspan) sub-sample. On the other hand, the unemployment rate is systematically over-predicted during the latter parts of the sample, while the forecast errors are not discernible from zero for the pre-Volcker, pre-disinflation periods.

The standard deviations of the final revisions for inflation and real output growth are about one-half those of the pre-disinflation period, consistent with evidence on the "Great Moderation." The standard deviation of the unemployment rate also declines in the latter sub-sample, but the difference is not statistically discernible at conventional levels of significance. Due to volatility early in Volcker's tenure, the standard deviation of inflation is comparable under the Volcker and pre-Volcker sub-samples, whereas it is substantially lower during the Greenspan period. Real output growth and the unemployment rate, on the other hand, are less volatile during the Volcker (and Greenspan) sub-sample than before 1980.

The volatility of the Greenbook forecasts also tend to decline with the volatility of the underlying forecasted series, with the standard deviations of the post-disinflation forecasts generally less than one-half as large as their pre-disinflation values. That said, the standard deviations of the inflation forecasts are significantly higher during the Volcker chairmanship than the pre-Volcker sub-sample at all forecast horizons, and comparable if slightly lower for the real output growth forecasts.

These findings are summarized in table 2, which reports the ratio of the standard deviations of the Greenbook forecasts to the standard deviations of the series being forecasted. There is no appreciable decline in relative volatility for any of the three series reported in table 2 between the pre-disinflation and post-disinflation periods. Indeed, the relative variability of inflation and real output growth is always largest during the Volcker sub-sample, and for inflation it is larger in the Greenspan sub-sample than the pre-Volcker one. Despite the evidence on the "Great Moderation" in U.S. real output,[10] the variability of the Greenbook forecasts relative to the variability of the underlying series does not decline over time. Tulip (2005) interprets a similar finding for output as evidence against a reduction in uncertainty concomitant with the reduction in volatility.[11] These findings appear difficult to reconcile with a "good policy" interpretation of the "Great Moderation."

These results also can serve as a simple test of efficiency: under rational expectations the variability in the forecast should be less than the variability in the actual data. By this metric the forecasts of the NIPA data do not appear to be inefficient, as the null hypothesis that the relative volatility ratio is one can be rejected in most cases in favor of the one-sided alternative that the ratio is less than one. The one exception is the Volcker sub-period. The unemployment rate forecasts, on the other hand, generally fail this test of efficiency, particularly during the post-disinflation period.

3 Tests of Forecast Rationality

Under the rational expectations hypothesis, the h-period ahead value of a time series variable can be decomposed into the expected value conditional on information available at time t, and any "news" that arrives after time t that affects the value h periods into the future. If Ω_t represents the collection of all relevant available information at time t to predict x_{t+h}, then by the rational expectations hypothesis,

$$x_{t+h} = \mathrm{E}[x_{t+h}|\Omega_t] + \eta_{t+h}, \tag{1}$$

where η_{t+h}, the rational expectations error, satisfies two statistical properties: $\mathrm{E}[\eta_{t+h}] = 0$ and $\mathrm{E}[\eta_{t+h} z_t] = 0$ for all $z_t \in \Omega_t$. The first condition implies the expectations are not biased, while the second implies that

[10]See, e.g., McConnell and Perez-Quiros (2000), Blanchard and Simon (2001) and Stock and Watson (2002).
[11]See also D'Agostino et al. (2006) for a similar conclusion for a wider collecction of series.

the expectations contain all relevant information for forecasting x_{t+h}.

Of course, an econometrician is unable to observe the complete information set Ω_t, and thus unable to determine exactly what "news" should be contained in η_{t+h}. For some information set $I_t \subset \Omega_t$ of observable data, an econometrician can test the efficiency of the forecast by investigating whether some variable $z_t \in I_t$ is correlated with the forecast error ν_{t+h}. Evidence of a statistically-discernible relationship is sufficient to reject rationality. Information known only after time t cannot be used to accurately determine the rationality of a forecast; hence the attraction of using real-time data: a current vintage data set contains historical revisions that could not have been known when the forecast was made at time t. Our exact timing information provides a further refinement to the specification of the appropriate information set, and thus yields more accurate tests of rationality.

In this section we employ several regression-based tests of the rationality of the Greenbook forecasts of inflation, real output growth and the unemployment rate. Results are reported for all the quarterly forecasts horizons available in the Greenbook data set starting with the contemporaneous quarter ($h = 0$), which is unknown due to lags in the release of data.

3.1 Weak-Form Univariate Rationality Tests

A common way to test whether an observed forecast series satisfies the properties of a rational predictor is to estimate a regression of the form[12]

$$x_{t+h} = \alpha + \beta x^f_{t+h|t} + \varepsilon_{t+h}. \tag{2}$$

The null hypothesis of a rational forecast is a joint hypothesis that $\alpha = 0$ and $\beta = 1$. Due to the fact that our forecasts occur at a higher frequency than the actual data being forecasted, we use a Newey and West (1987) procedure to correct the reported test statistics for the overlapping forecasts.

Table 3 reports the p-values of tests of the joint hypothesis that $\alpha = 0$ and $\beta = 1$ in equation (2). Consistent with the findings of Romer and Romer (2000), albeit for a longer sample period, we find that the null hypothesis of rationality cannot be rejected for the inflation forecast when examined over the entire November 1965 – December 2000 sample period. The same is true for real output growth and the average unemployment rate.

[12] Such regressions are often called "Mincer-Zarnowitz regressions;" see, for example, Diebold (2001).

The results are different when we divide the sample, however. For inflation, the rejections of rationality occur fairly strongly in each sub-sample period for nearly all forecast horizons. For real output growth, on the other hand, there is relatively little evidence against the null hypothesis. The most likely violations occur during the Greenspan sub-sample, although these p-values are not below conventional significance levels for most forecast horizons during this period. For the unemployment rate, rationality can be strongly rejected for the post-disinflation sub-sample but not the pre-disinflation one. However, when split between the Volcker and Greenspan sub-samples, the evidence against rationality of the forecast of the unemployment rate fades.

3.2 Semi-Strong-Form Multivariate Rationality Tests

We next conduct tests with additional lagged information of both the series being forecasted as well as other forecasted series that should have been known at the time the forecast was made. If the forecasts are formed rationally, then any information contained in the lagged values should already have been incorporated into the forecast. One can formulate this concept by augmenting equation (2) to include lagged observations:

$$x_{t+h} = \alpha + \beta x^{f}_{t+h|t} + \sum_{i=1}^{n} \sum_{j=1}^{q} \gamma_{ij} z_{i,t-j} + \varepsilon_{t+h}. \tag{3}$$

That is, one through $q \geq 1$ lags of the variables z_i, $i = 1, \ldots, n$, are added to the regression represented by equation (2). One can then test the joint hypothesis that $\alpha = 0$, $\beta = 1$, and $\gamma_{ij} = 0$ for all i, j.[13]

In implementing this test, we report results for $q = 1$ and $q = 4$ lags of the inflation rate, the growth rate of real output, and the unemployment rate to each forecast regression. Given the construction of our data set discussed in section 2, the subscripts do not necessarily reflect calendar dating. Rather, the dating of information as known by the staff forecasters is determined by the publication date of the Greenbooks relative to the release dates of the data series of interest. These lagged observations are taken directly from the specific Greenbook published at the date in question and thus are unaffected by other specification choices, such as the release used for the forecasted object.

Tables 4 and 5 report the p-values for the null hypothesis corresponding to rationality in the presence of one quarter and four quarters of lagged information, respectively. These multivariate regressions constitute more extensive tests of rationality, and we find fairly consistent evidence against the null hy-

[13]Note that in this particular study, $n = 3$: inflation, real output growth and the unemployment rate are the observed series.

pothesis for all three series at almost all horizons, regardless of the particular sub-sample considered. In particular, whereas the null hypothesis of rationality tended to not be rejected when estimation was undertaken with the full sample period in table 3, in table 4 the null can be rejected in this case more often than not.[14] The most pronounced change from table 3 to table 4 occurs for real output growth, indicating that the forecasts of output growth do not make efficient use of lagged information.

Table 5 also tests the null hypothesis of rationality as expressed in equation (3), but for $q = 4$ lags of inflation, real output growth, and the unemployment rate in each regression. With only a few exceptions, the p-values are lower in table 5 than in table 4. The observation that more distant lagged information increases the likelihood of rejecting rationality is interesting. While this result may be a statistical quirk, it also is consistent with the possibility that information made available in subsequent data revisions is not fully incorporated into the Greenbook forecasts. It also is suggestive of the importance of lower-frequency components of the forecasted series, as we explore in section 4.

3.3 Multivariate Rationality Tests with Forecast Errors

An alternative way to introduce lagged information into a regression test of rationality is to regress the h-period ahead forecast error on lagged forecast errors. If forecasters make efficient use of all available information at the time the forecasts are produced, then lagged forecast errors should be uncorrelated with the difference between the h-period ahead forecast and the h-period ahead realization, and thus have no explanatory power in the following regression:

$$v_{t+h|t} = \sum_{i=1}^{n} \sum_{j=h+1}^{q} \theta_{ij} v_{i,t+h-j|t-j} + \varepsilon_{t+h}, \tag{4}$$

where $i = 1, \ldots, n$ represents the series whose forecast errors are included in the regression; the left-hand side forecast error is one of the n series used in the lags. As written, equation (4) only uses lagged forecast errors for the same horizon; for example, the 4-quarter ahead forecast error is regressed on the last observable 4-quarter ahead forecast error, so that the information set contains only those observations that actually can be known at time t. Thus, given the timing of releases and forecasts, the lagged 4-quarter ahead forecast error may be based on a forecast made five or six quarters ago. Of course, any lagged forecast error that could be computed at time t could have been included in equation (4). For reasons of

[14]Notice that the univariate weak-form tests reported in table 3 can be viewed as the limiting case for which $q = 0$ in equation (3).

parsimony we restrict our investigation to the forecast errors for the same forecast horizon and include $q = 1$ lags for the results shown in table 6.

Table 6 reports the p-values corresponding to the test that $\alpha = 0$ and $\theta_{ij} = 0$ for all values of $i = 1,\ldots,3$ and $j = 1$. While the rejections are not as extensive as those in table 5, there are more instances in which the null hypothesis is rejected than not. Both inflation and real output growth pass this test of rationality when estimated over the full sample, but less so for any of the sub-samples. Furthermore, the p-values indicating rejection tend to be lower the larger is the forecast horizon, h. At conventional levels of significance, the test statistics for the unemployment rate suggest a less-than-rational forecast for most sub-periods and forecast horizons.

Thus the tests reported in this section reveal relatively consistent evidence against the null hypothesis that the Greenbook forecasts are rational, particularly once lagged information is included in the regressions. What is perhaps striking about the results in table 6 is that the gap between the forecast error and the lagged information included in the regression is even longer than for the previous tests reported in table 5, yet the p-values continue to indicate rejection of the null of rationality. Overall, our results indicate that the Greenbook forecasts are both biased and inefficient — particularly once attention is placed on the individual sub-sample periods.

4 Interpretation

In the introduction we noted that much of the existing literature has found the Greenbook forecasts to be rational. When we look at simple measures of rationality, such as the tests of unbiasedness of the average forecast errors in table 1, the tests of efficiency using the ratio of forecast to actual standard deviations in table 2, or the univariate Mincer-Zarnowitz regressions in table 3, we find similar results — provided estimation occurs over the full November 1965 to December 2000 sample period.

Once we examine sub-sample periods, however, a distinct pattern emerges: the Greenbook forecasts tend to under-predict inflation and over-predict real output growth to a statistically discernible degree in the pre-Volcker sub-sample, while in the Volcker and Greenspan sub-samples the forecast errors are largely the opposite: inflation and unemployment are over-predicted while real output growth is under-predicted (although the latter is not significant at conventional levels of statistical significance). Furthermore, semi-strong tests of rationality in tables 5 and 6, which utilize the lagged information known to the Fed staff economists at the time the Greenbook forecasts were prepared, reject the null hypothesis for most forecast horizons. What accounts for these differences among the sub-sample periods?

We explore three possible interpretations. First, our sub-sample results might give a misleading impression of the rationality of forecasters in real time. By restricting our estimation to particular sub-samples, we are both using less information and more information than was available to the forecasters. The real-time history of forecasts and data realizations are available to forecasters in the latter part of our sample, although they may correspond with a different sub-sample in our tests. At the same time, our regressions are *ex post* in nature, as we use all the data available in the sub-sample period to construct our test statistics. As an alternative approach, below we conduct tests based on rolling regressions that use all available data up to the date the forecast is made, but not after.

Second, we investigate how assumptions about the loss function of policy makers might be related to our findings. An extensive literature has argued that the preferences of U.S. monetary policy makers have changed dramatically over time: see, e.g., Taylor (1999) or Clarida et al. (2000). Others have criticized this view based on an analysis of real-time data: Orphanides (2001, 2002) is a prominent example. A recent variation on the theme of time-variation in preferences has been posited by Capistrán-Carmona (2005), who argues that the Fed in the post-disinflation period has an asymmetric loss function — inflation above the desired rate is much more undesirable than below — which could explain the observed bias. We discuss this perspective in more detail below.

Third, there are several significant differences in the attributes of the observed data, as well as the forecasts, between the pre-disinflation and post-disinflation periods. Note initially that the volatility of both the data and the forecasts is noticeably greater in the pre-disinflation period — although table 2 illustrates that there has not been much of a change in the relative volatility over time. In other words, while the forecast errors are generally larger in magnitude in the earlier period considered, their sizes appear to be a constant proportion of the (time-varying) degree of uncertainty inherent in the economy throughout the entire sample.

Additionally, there are two pronounced long-run trends in the inflation and unemployment rate data: a positive trend in the pre-disinflation period, and a negative trend in the post-disinflation period. The series are not monotonic within each sub-sample, however, and there are stretches of time in which the data are locally trending in a different direction. As the graphs of the 4-quarter ahead forecasts against the final released data in figures 4 through 6 indicate, at least some aspects of these trends were not predicted by the Fed staff: witness the large negative errors for inflation and unemployment during the early 1970s — the onset of the "productivity slowdown" — or the persistent positive errors in the same series during the "New Economy" period of the last half of the 1990s.[15] Later in this section we investigate the relationship between these trend shifts and the forecast errors in more detail.

4.1 Rolling Regression Analysis of Forecast Rationality

Figures 7 through 9 report the p-values corresponding to the joint tests of rationality of equation (3), conducted from a rolling regression procedure, for the $h = 1$ period-ahead forecasts of inflation, real output growth, and the unemployment rate, respectively. Estimation begins in January 1971 (starting the estimates with just over five years of available Greenbook forecast data) and continues to add one observation per regression until December 2000. The horizontal dashed lines represent the conventional significance levels of 1%, 5%, and 10%.

In all three cases, the p-values never are greater than 10%. Indeed, for the inflation rate shown in figure 7, only for a few observations at the end of 1994 does the p-value barely cross the 1% threshold. The same holds true for the average unemployment rate, shown in figure 9, in which the p-value exceeds 1% only during the first half of 1980, and never reaches even 4%. Only for real output growth in figure 8 do the p-values from the rationality tests ever exceed 5% — in the mid-1980s and in the early 1990s. Moreover, the p-values slip back towards zero in the latter part of the 1990s.[16]

Thus, our results do not appear to be an artifact of *ex post* estimation over individual sub-sample periods, as even the cumulative rolling regression results shown in figures 7 through 9 support our conclusions that the Greenbook forecasts are not rational.[17]

4.2 Can an Asymmetric Loss Function Explain the Results?

Capistrán-Carmona (2005) uses a quarterly representation of the Greenbooks between 1968Q4 and 1998Q4 to study the properties of the inflation forecasts. Consistent with our results above, he documents that

[15] Schuh (2001) documents sizable and persistent under-prediction errors for real output growth and over-prediction errors for the unemployment rate between 1996 and 2000 by private sector forecasters.

[16] Similar result are found at other forecasts horizons, which are not reported for brevity. The one exception is the contemporaneous forecasts ($h = 0$) for real output growth, in which the p-values range between 15% and 30% from 1983 through the end of our sample.

[17] Rolling estimation of the Mincer-Zarnowitz regressions of equation (2) (not shown for brevity) generally yield a time series of p-values well above 10% for all series and forecast horizons. In other words, rejection of the rationality hypothesis appears to be due less to bias than to inefficient use of available information, as the forecast errors are correlated with lagged data known at the time the forecast was made.

his quarterly inflation forecast errors are negatively biased prior to 1979 and positively biased in subsequent years. He attributes this finding of systematic bias to an asymmetric loss function for the FOMC, which he estimates to have a four-times greater preference for below-target inflation than above.

As a description of the potential preferences of policy makers, Capistrán-Carmona's perspective has a certain appeal. Many analysts have characterized the post-disinflation Fed as "conservative" in the sense of Rogoff (1985). To maintain a credible reputation of conservatism, a central bank may behave in ways that bear a resemblance to, if not formally described as, an asymmetric loss function for inflation per Capistrán-Carmona (2005). The Volcker chairmanship in particular is often characterized in this way. Greenspan has publicly advocated a "risk-management" approach to monetary policy that also might be viewed as asymmetric. Specifically, one might interpret Greenspan's stated preference for "price stability" as the primary feasible goal of monetary policy, in place during a decade of uninterrupted economic growth, as evidence of a strong aversion to above-target inflation rates.[18]

However, Capistrán-Carmona's story is less compelling for the staff economists who construct the Greenbook forecasts. Even if the policy makers on the FOMC have an asymmetric loss function — which is open to question — why would these policy makers want to be provided a biased forecast? Based on interviews with those responsible for creating the forecasts, Sims (2002) documents the separation of policy decisions from forecast construction:

> "The Federal Reserve has an explicit policy of maintaining the forecast as purely a staff forecast, not allowing any policy board participation in the meetings that go into forecast preparation." (p. 4)

Given the five-year publication lag and lack of individual accountability for specific forecasts, the staff has few if any incentives to manipulate the forecasts. Rather, if the staff's objective is to deliver the best possible forecast given available information, that may be well approximated by a symmetric quadratic loss function. In short, forecasters may not share the preferences of policy makers.

Furthermore, an asymmetric loss view is difficult to reconcile with aspects of the data. Consider that Capistrán-Carmona's sample period ends in 1998, whereas the final two years of our sample indicate a shift in the trend inflation rate from negative to positive — other than the near zero rate of growth in the deflator at the end of 2000. With the exception of this final observation, figures 1 and 4 show that

[18] Note that the Fed's concern about the possibility of deflation in the U.S. economy, and the risks associated with too *low* of an inflation rate, did not arise until 2000 — outside of Capistrán-Carmona's sample.

the inflation forecast errors have tended to be negative, not positive, during these few years. A similar pattern can be observed in the late 1980s, as inflation trended upwards prior to the 1991 recession. If the Fed's loss function truly penalized positive forecast errors (that is, under-predicting the inflation rate) to the extent Capistrán-Carmona (2005) reports, such instances simply should not occur.

Since the inflation rate was generally trending downwards throughout Capistrán-Carmona's sample period, it is difficult to distinguish an asymmetric loss function from alternative explanations of the observed data — that is, Capistrán-Carmona's approach may have low power. There are relatively few instances of rising inflation in his sample that provide appropriate opportunities to test his hypothesis.

As a further counter-example, consider the 4-quarter ahead forecasts of the unemployment rate in figure 6: like the inflation rate, the unemployment rate experienced a similar shift of its low frequency trend from positive in the 1970s to negative in the 1990s. However, unlike the inflation rate, there also are sustained periods of generally rising unemployment in the early years of each of the last three decades, due in each case to a recession. Under an asymmetric loss function, these large negative forecast errors would be extremely costly and presumable would not persist for the multiple consecutive years exhibited by the data.

Finally, while an asymmetric loss function might be consistent with the observed inflation data for the post-1979 sub-sample, it is not apparent how an asymmetric loss function can explain the significant residual forecasting power of lagged data that was revealed in tables 5 and 6. An explanation that could account for both the inefficiency and the biasedness of the Greenbook forecasts, across multiple sub-sample periods and forecasted series, might be preferable.

4.3 The Relationship between Forecast Errors and Underlying Trend Shifts

As mentioned above, the forecast errors in figures 4 through 6 appear to have a systematic relationship with the lower-frequency trends in the data being forecasted: when either the inflation rate or the unemployment rate is rising, the Greenbook forecasts tend to under-predict the variable in question; when either is falling the forecasts tend to over-predict. A somewhat similar pattern can be seen in real output growth despite the fact that it is relatively trendless over the full sample (although the mean level is lower during the Volcker sub-sample period than either sub-sample before or after); in this series the graphical evidence points more to missed business cycle turning points as one significant source of forecast error.

To determine whether this conjecture has quantitative content, we regress the h-period ahead forecast error, $v_{t+h|t}$, on a measure of the change in the underlying trend in the variable being forecasted. For the results reported in table 7, the trend is computed by a Hodrick-Prescott filter. We use data from 1960Q1 though 2006Q1 to construct each filtered series.[19] We then compute the one-period rate of change in the filtered series to represent the change in the trend at the date the forecast was formed.

Recall that the graphs suggest a tendency for negative forecast errors ("under-prediction") when the series has an increasing trend, and positive errors ("over-prediction") when the trend is declining. Thus, one would expect a negative sign on the coefficient in these regressions if the trend is systematically related to the forecast error. Additionally, one might conjecture that the consequences of missing a change in the trend are quantitatively greater at longer forecast horizons. In effect, the further out the forecast horizon, the more important is the contribution of the trend vis-à-vis higher frequency aspects (temporary shocks and noise) to the actual realized value. This analysis would imply a larger impact (in absolute value) of changes in the trend for forecasts made over longer horizons.

Table 7 reports the results of this regression exercise. In general, the estimated coefficient on the change in the trend is strongly statistically significant (especially for inflation and the unemployment rate), has a negative sign, and within any particular sub-sample has a greater absolute magnitude for larger values of h.[20] The cases in which the coefficient is not statistical significant are also cases for which the hypothesis of rationality could not be rejected. In particular, for the contemporary forecast horizon ($h = 0$), shifts in low-frequency trends intuitively play a relatively minor role in the current quarter's observed releases.

Recall that the rejection of the rationality of the Greenbook forecasts, as determined by equation (3), tends to increase with the number of periods into the past, q, for which lags are included. This observed behavior is also consistent with the above analysis, as longer lags are more likely to carry a signal about the trend than are more recent observations, which may be correlated in part due to common transitory innovations ("noise") to the series in question. Thus, the observed bias and inefficiency attributes of the forecasts are associated with statistical measures of the change in the underlying trending behavior of the forecasted series. Qualitatively similar regression results are found when either a multi-year centered

[19] Such filters are known to be less accurate at the very beginning and end of a time series. As our estimation occurs for Greenbook publication dates between November 1965 and December 2000, we therefore have at least five years on either side of the filtered series that are not included in the regression results reported in table 7.

[20] As mentioned above, the results for real output growth should be interpreted more as indicative of the effects of turning points at business cycle frequencies than as actual trends in the data.

moving average[21] or simply the quarterly change in the forecasted series used as a measure of the change in the underlying trend, although both of these computed series are far more noisy than the Hodrick-Prescott smoothed trend series.

Of course, a trend is an *ex post* concept that cannot readily be determined in real time. Thus, to the extent that rejection of rationality is strongly associated with a series that is not in the information set of the forecasters, the above rejections of rationality are less straight-forward to interpret. It remains possible that some other data contained in the information set of the staff economists are responsible for the observed rejections. But the statistical association between the forecast errors and the lower-frequency trending behavior appears to be statistically and economically significant, and better accounts for the full range of observed empirical relationships than alternative hypotheses that have been proposed.

5 Conclusion

This paper introduces an integrated database of Federal Reserve Greenbook forecasts and real-time U.S. macroeconomic data for inflation, real output growth and the employment rate. The novel contribution of our data set is the tracking of the exact dates of data releases and forecast publications, which yields a compelling environment for us to examine the statistical properties of the Greenbook forecasts.

Our results have demonstrated pronounced evidence against the rationality of the Greenbook forecasts for all three macroeconomic series whose forecasts we evaluate, particularly once sub-sample behavior is considered. The pre-1979 forecast errors suffer from an "optimistic" bias, in the sense that the Fed staff regularly over-predicted the rate of real growth and under-predicted the rate of inflation. Conversely, the post-1979 forecast errors display a "pessimistic" bias: systematic over-prediction of inflation and the unemployment rate, and under-prediction of the growth rate of real output. One interpretation of these findings infers an "expansionary" bias during the pre-disinflation period. Symmetry might suggest a "contractionary" bias during the Volcker and Greenspan eras as well.

More significantly — and less apparently consistent with the above-mentioned interpretation — regression tests of forecast efficiency are broadly rejected throughout both the full sample and each subsample considered. Moreover, as the lag length of known real-time data included in the regression increases, the rejection of forecast efficiency generally grows stronger. Our results suggest a new interpre-

[21] A 3-year centered moving average for inflation is surprisingly similar to the Hodrick-Prescott filtered series in frequency, albeit with less pronounced amplitude.

tation for these sources of irrationality in the forecasts: the inability to forecast changes in the lower-frequency trending behavior of the series being forecasted.[22] The systematic forecast errors during the "New Economy" period are just one example of this more general pattern of forecast behavior observed throughout our sample. When we regress the forecast errors on a measure of the change in the underlying trend for a given series, we strongly reject the null hypothesis of no relationship in favor of a sizable negative correlation that generally increases with the forecast horizon.

Our preferred interpretation of these results is that the errors at longer horizons can be attributed to changes in the trend rather than to higher-frequency innovation or noise. Moreover, to the extent that observations further away in time are more informative of trend changes than nearer ones (which might share common short-run stochastic components or noise), our interpretation can encompass the evidence that rejections of forecast efficiency increase as the span of historical data included in the information set lengthens.

As such trend shifts cannot necessarily be recognized in real time, these results may exonerate the forecasters and instead call into question the appropriateness of standard tests of rationality. Such questions are the topic of future research with these data. However, as we argue in section 4, the nature of the forecasting problem, as well as the results, are difficult to reconcile with the asymmetric loss approach of Capistrán-Carmona (2005). Our interpretation does not need to appeal to a particular shift in the monetary policy regime in order to account for the forecast failures throughout our sample of available Greenbook forecasts.

Finally, in light of the opening quote of Alan Greenspan, the preponderance of evidence against the unbiasedness and efficiency of the Greenbook forecasts appears to throw some doubt upon claims that "good policy" can explain the generally better U.S. macroeconomic performance over the past two decades or so. Further research into this issue is warranted as well.

[22] Additionally, our results are consistent with the Greenbook forecasters missing turning points at a business-cycle frequency.

References

Blanchard, Olivier J. and John Simon, "The Long and Large Decline in U.S. Output Volatility," in William C. Brainard and George L. Perry, eds., *Brookings Papers on Economic Activity*, Vol. 1 2001, pp. 135–174.

Capistrán-Carmona, Carlos, "Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?," January 2005. manuscript, University of California, San Diego.

Clarida, Richard, Jordi Galí and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, February 2000, *115* (1), 147–180.

Clements, Michael P., Fred Joutz, and Herman O. Stekler, "An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach," September 2004. George Washington University Manuscript.

Croushore, Dean and Tom Stark, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, November 2001, *105*, 111–130.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico, "(Un)Predictability and Macroeconomic Stability," Working Paper 605, European Central Bank 2006.

Diebold, Francis X., *Elements of Forecasting*, Cincinnati, Ohio: South-Western, 2001.

Gavin, William and Rachel Mandal, "Evaluating FOMC Forecasts," *International Journal of Forecasting*, October–December 2003, *19* (4), 655–667.

Greenspan, Alan, "Risk and Uncertainty in Monetary Policy," *American Economic Review Papers and Proceedings*, May 2004, *94* (2), 33–40.

Jansen, Dennis W. and Ruby Pandey Kishen, "An Evaluation of Federal Reserve Forecasting," *Journal of Macroeconomics*, Winter 1996, *18* (1), 89–109.

Joutz, Fred and Herman O. Stekler, "An Evaluation of the Predictions of the Federal Reserve," *International Journal of Forecasting*, January–March 2000, *16* (1), 17–38.

McConnell, Margaret M. and Gabriel Perez-Quiros, "Output Fluctuations in the United States: What Has Changed since the Early 1980s?," *American Economic Review*, 2000, *90* (5), 1464–1476.

Newey, Whitney K. and Kenneth D. West, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, May 1987, *55* (3), 703–708.

Orphanides, Athanasios, "Monetary Policy Rules Based on Real-Time Data," *American Economic Review*, September 2001, *91* (4), 964–985.

__ , "Monetary-Policy Rules and the Great Inflation," *American Economic Review*, May 2002, *92* (2), 115–120.

Rogoff, Kenneth, "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics*, November 1985, *100* (4), 1169–1189.

Romer, Christina and David Romer, "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, June 2000, *90* (3), 429–457.

__ **and** __ , "Changing Views about Stabilization Policy: A Historical Perspective," in "Rethinking Stabilization Policy," Vol. Proceedings of the Federal Reserve Bank of Kansas City 2002 Economic Policy Symposium 2002, pp. 11–78.

Schuh, Scott, "An Evaluation of Recent Macroeconomic Forecast Errors," *New England Economic Review*, January/February 2001, pp. 35–56.

Sims, Christopher A., "The Role of Models and Probabilities in the Monetary Policy Process," in William C. Brainard and George L. Perry, eds., *Brookings Papers on Economic Activity*, number 2 Brookings Institution 2002, pp. 1–62.

Stock, James H. and Mark W. Watson, "Has the Business Cycle Changed and Why?," in Mark Gertler and Kenneth Rogoff, eds., *NBER Macroeconomics Annual*, Vol. 17 MIT Press March 2002, pp. 159 –218.

Taylor, John B., "A Historical Analysis of Monetary Policy Rules," in John B. Taylor, ed., *Monetary Policy Rules*, Chicago: University of Chicago Press, 1999, pp. 319–341.

Tulip, Peter, "Has Output Become More Predictable? Changes in Greenbook Forecast Accuracy," Finance and Economics Discussion Series 2005-31, Board of Governors of the Federal Reserve System August 2005.

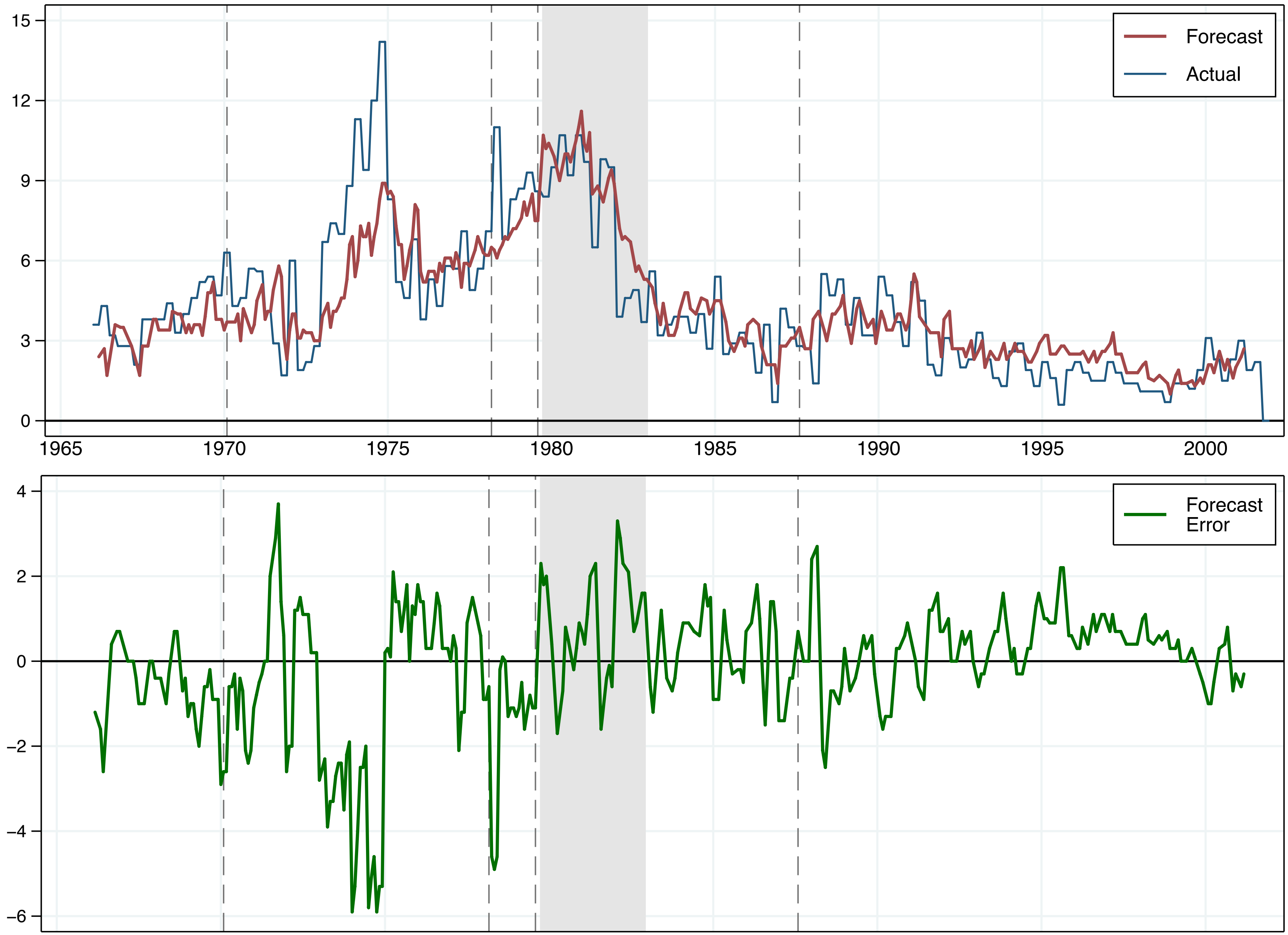

Figure 1: One-Quarter Ahead Greenbook Forecasts of Inflation

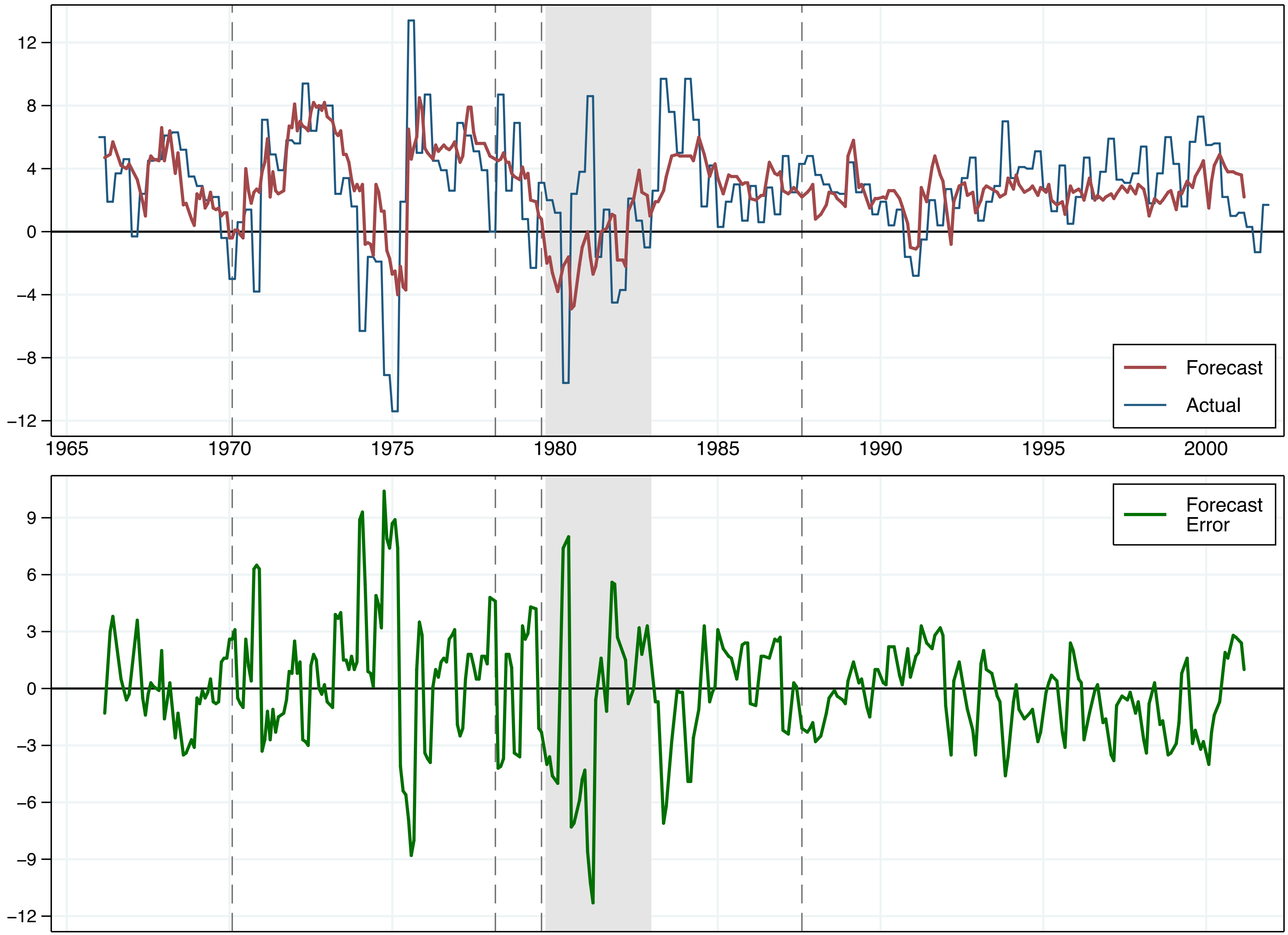

Figure 2: One-Quarter Ahead Greenbook Forecasts of Real Output Growth

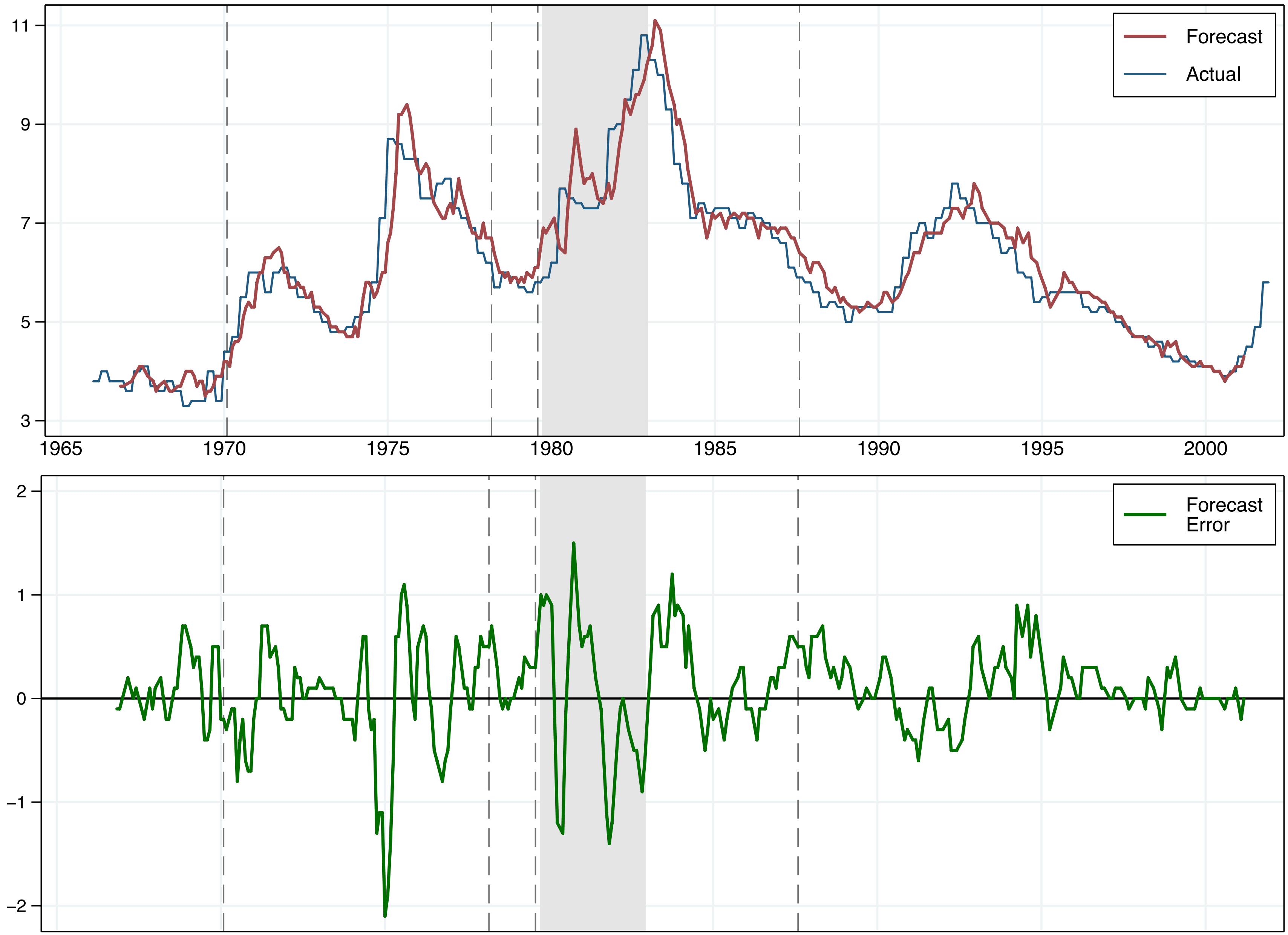

Figure 3: One-Quarter Ahead Greenbook Forecasts of the Unemployment Rate

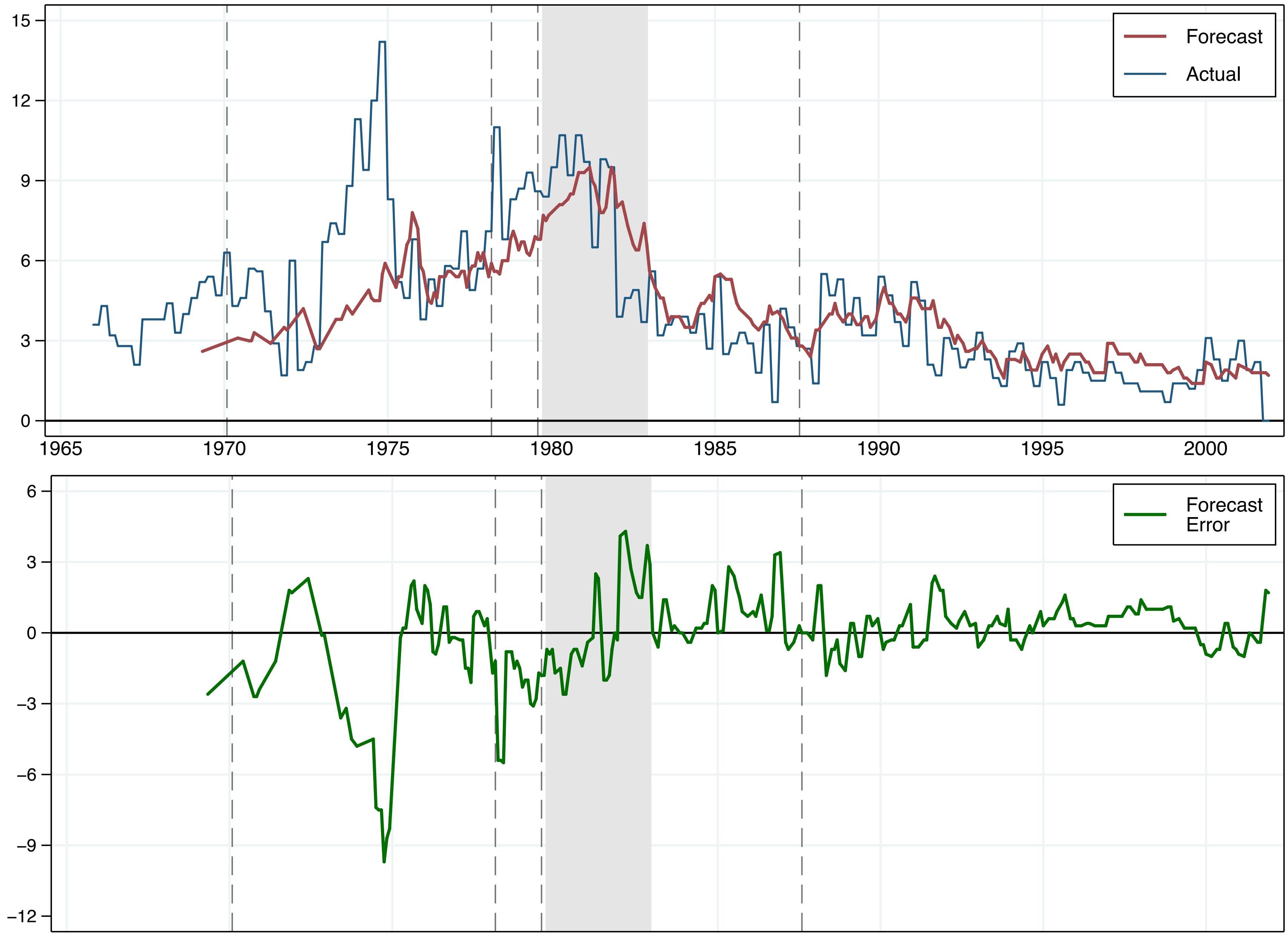

Figure 4: Four-Quarter Ahead Greenbook Forecasts of Inflation

Figure 5: Four-Quarter Ahead Greenbook Forecasts of Real Output Growth

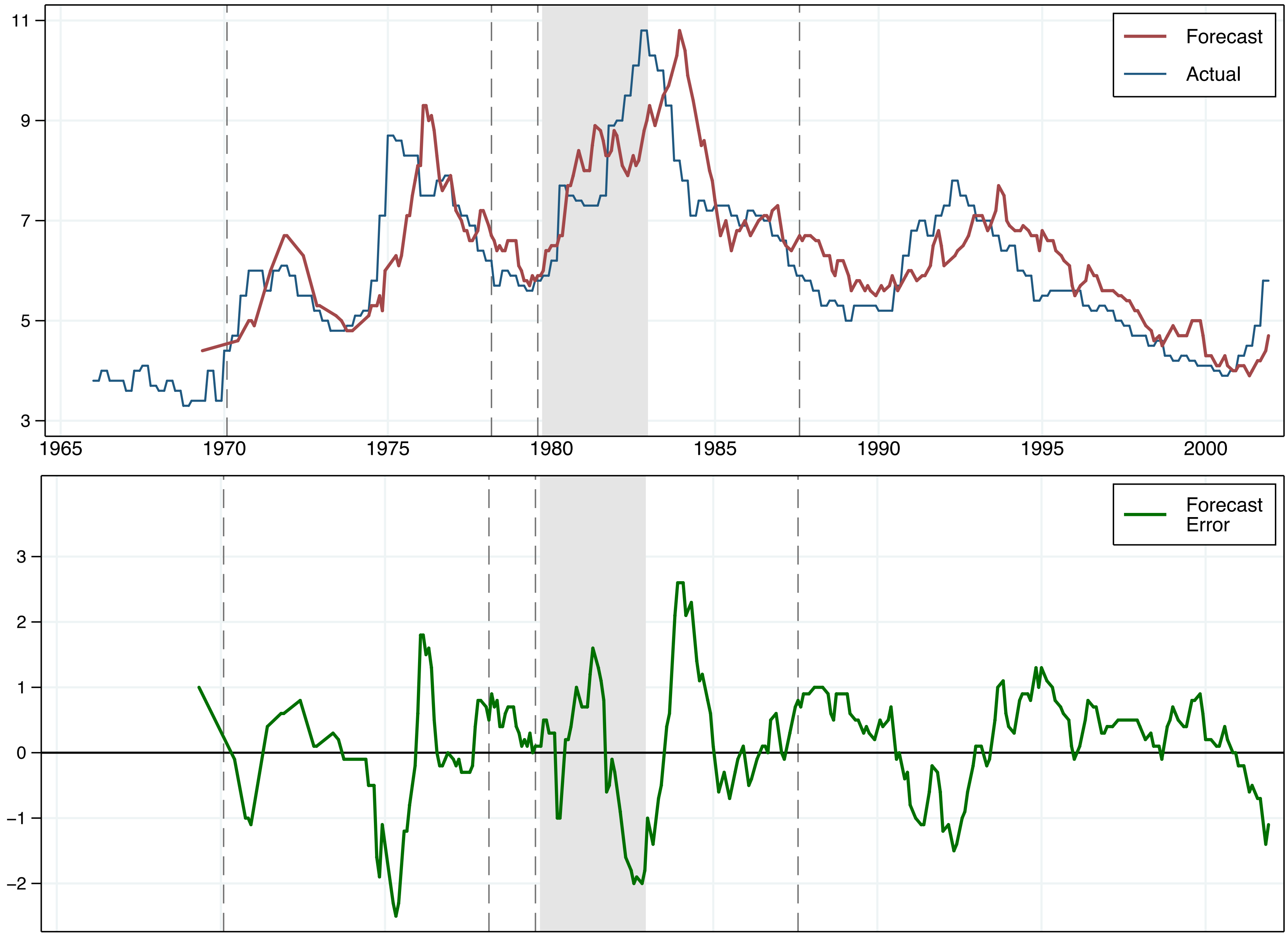

Figure 6: Four-Quarter Ahead Greenbook Forecasts of the Unemployment Rate

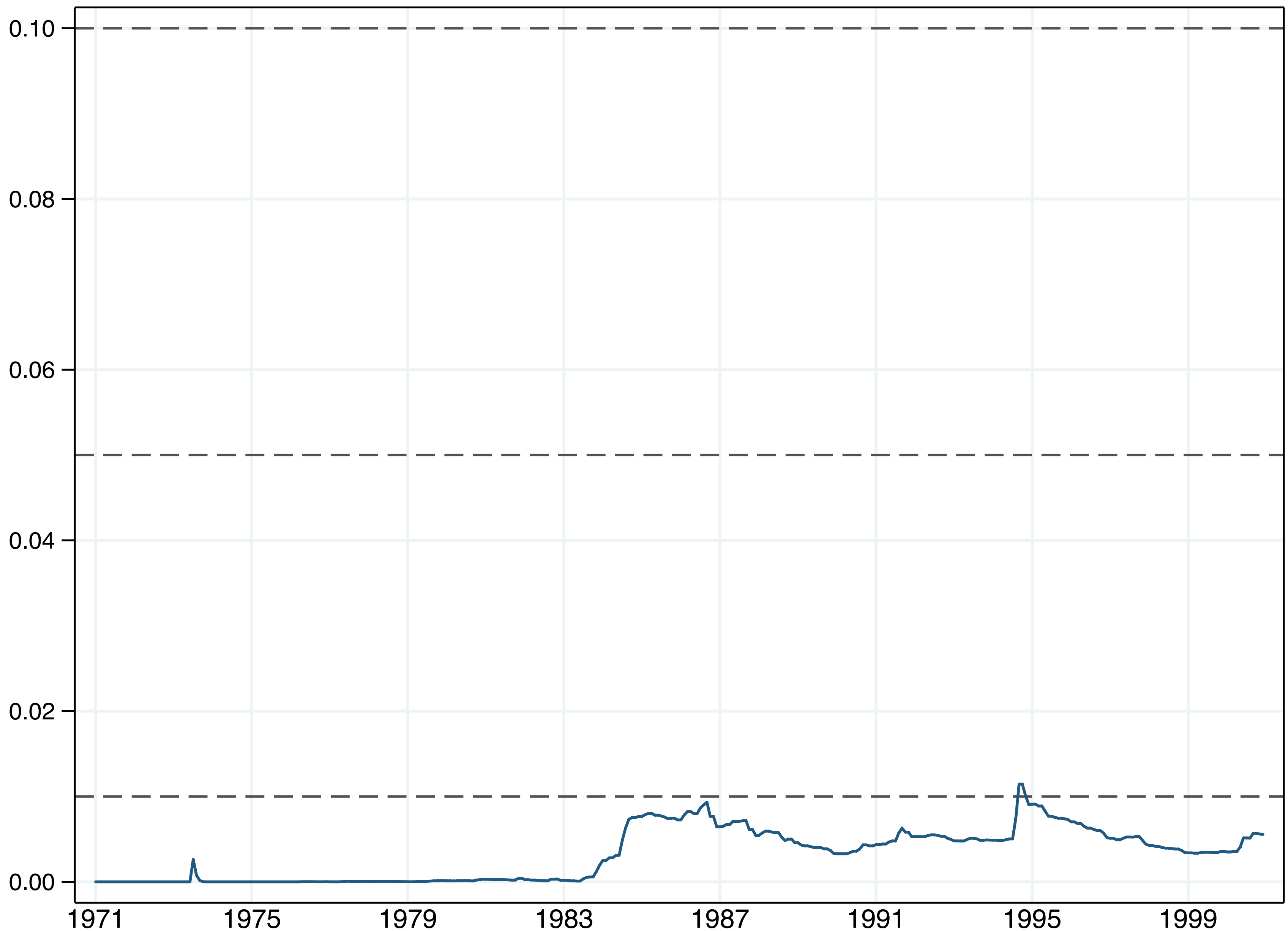

Figure 7: Rolling *p*-values for Semi-Strong-Form Multivariate Rationality Tests, Inflation Rate

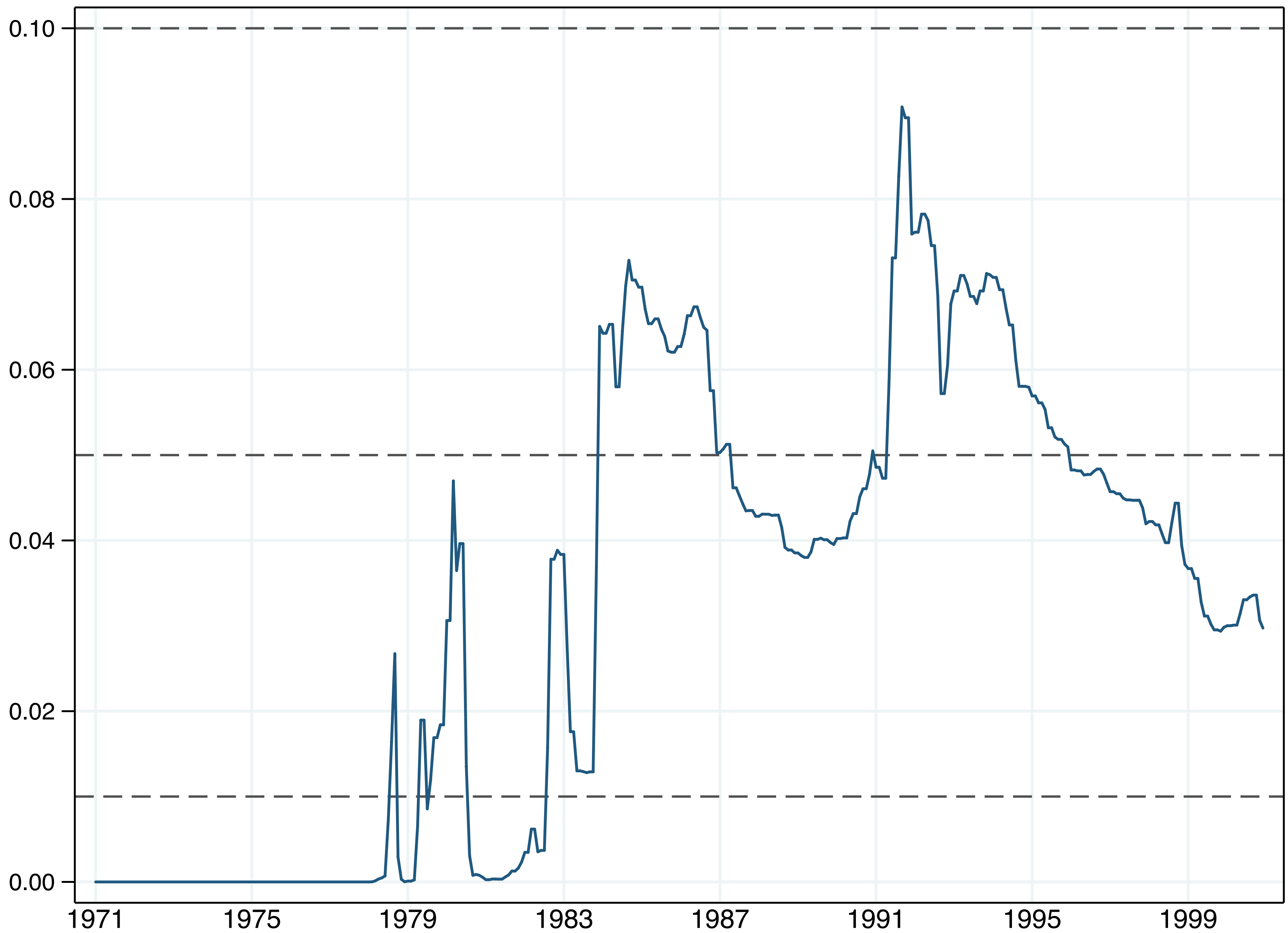

Figure 8: Rolling *p*-values for Semi-Strong-Form Multivariate Rationality Tests, Real Output Growth

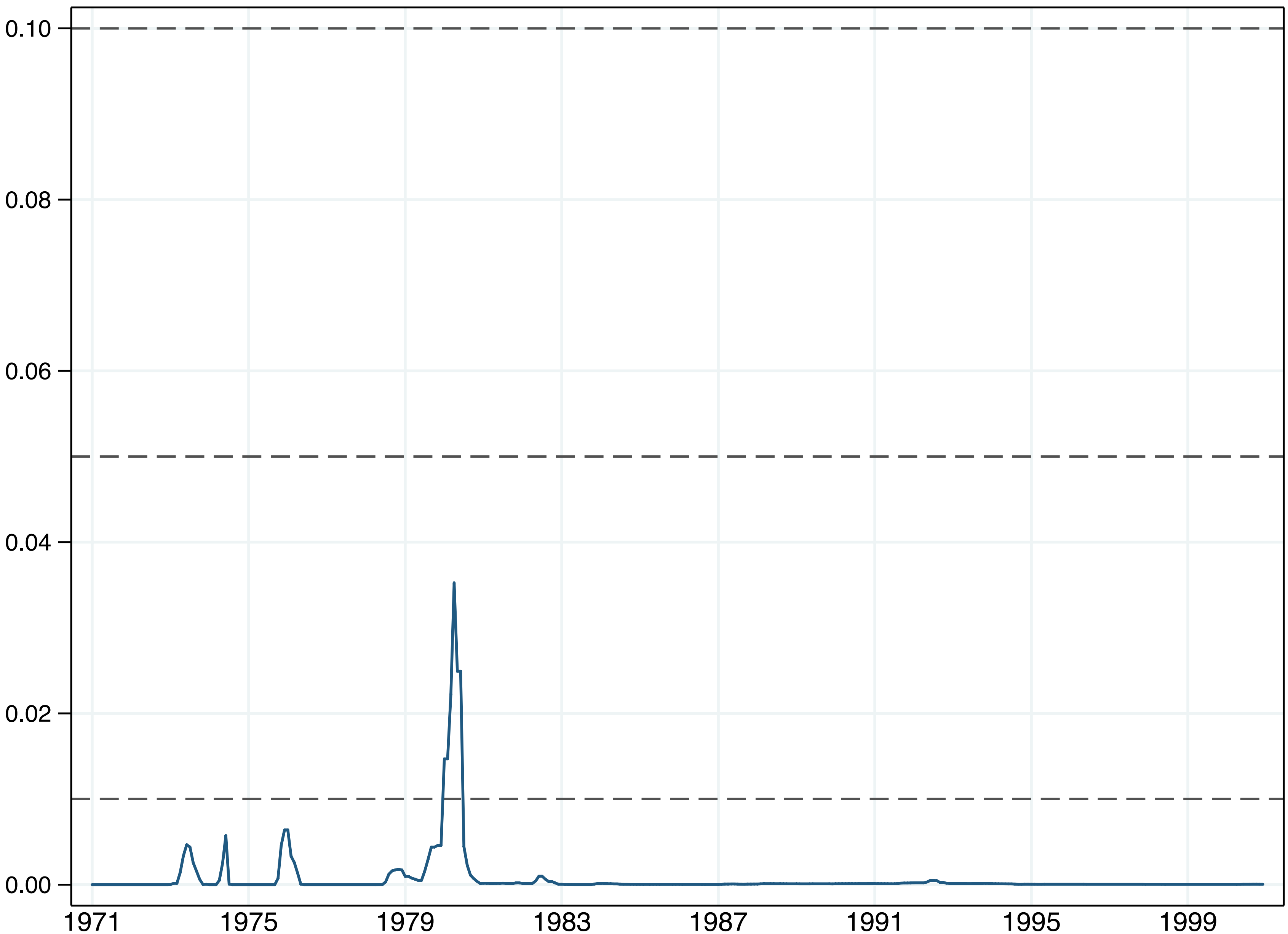

Figure 9: Rolling *p*-values for Semi-Strong-Form Multivariate Rationality Tests, Unemployment Rate

Table 1: Average Forecast Errors and Tests of Forecast Unbiasedness

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	−0.05	-0.17^c	-0.25^b	−0.28	−0.19	0.04	0.36^b	0.74^a
Pre-disinflation	-0.30^a	-0.75^a	-1.10^a	-1.32^a	-1.53^a	-2.25^a	-2.96^b	—
Post-disinflation	0.14^b	0.29^a	0.29^a	0.33^a	0.41^a	0.52^a	0.56^a	0.76^a
Pre-Volcker	-0.31^a	-0.79^a	-1.12^a	-1.31^a	-1.55^a	-2.26^a	-2.96^b	—
Volcker	0.15	0.50^a	0.50^a	0.61^a	0.69^a	1.04^a	0.77^b	1.44^b
Greenspan	0.22^a	0.28^a	0.31^a	0.23^a	0.31^a	0.38^a	0.53^a	0.68^a

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	−0.15	−0.01	0.08	0.28	0.09	0.18	−0.05	−0.08
Pre-disinflation	0.24	0.58^b	0.81^b	1.23^a	0.84^c	1.39^c	0.91	—
Post-disinflation	−0.37	-0.42^b	-0.41^b	−0.26	−0.27	−0.22	−0.15	−0.19
Pre-Volcker	0.29	0.64^b	0.85^b	1.08^b	0.90^c	1.60^b	0.91	—
Volcker	-0.59^c	−0.63	−0.60	−0.05	−0.21	0.25	0.34	1.04
Greenspan	-0.56^a	-0.48^b	−0.41	−0.36	−0.33	−0.36	−0.25	−0.30

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.02^c	0.07^a	0.10^a	0.10^b	0.14^a	0.16^a	0.15^b	0.15
Pre-disinflation	0.02	0.06	0.04	−0.01	0.01	0.05	−0.19	—
Post-disinflation	0.03^b	0.10^a	0.16^a	0.22^a	0.25^a	0.26^a	0.21^a	0.23^b
Pre-Volcker	0.02	0.04	0.02	−0.02	0.01	0.03	−0.19	—
Volcker	0.03	0.11^c	0.19^b	0.21^c	0.23^c	0.19	0.26	0.18
Greenspan	0.01	0.09^a	0.13^a	0.17^a	0.19^a	0.19^a	0.15^c	0.17

Reported value is $\frac{1}{T}\sum_{t=1}^{T}\left(x^f_{t+h|t} - x_{t+h}\right)$. Superscripts ^{a}, ^{b}and ^{c} represent rejection at the 1%, 5% and 10% levels of significance, respectively, of the null hypothesis that the average forecast error equals zero.

Table 2: Relative Volatility and Simple Forecast Efficiency Tests

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.89^b	0.82^a	0.76^a	0.74^a	0.73^a	0.72^a	0.65^a	0.64^a
Pre-disinflation	0.80^a	0.70^a	0.62^a	0.57^a	0.54^a	0.45^a	0.57	—
Post-disinflation	0.85^b	0.74^a	0.79^a	0.80^a	0.82^a	0.81^b	0.78^a	0.79^b
Pre-Volcker	0.79^a	0.67^a	0.60^a	0.56^a	0.52^a	0.41^a	0.57	—
Volcker	1.05	1.03	0.92	0.89	0.85	0.85	0.91	0.99
Greenspan	0.84^c	0.70^a	0.76^a	0.76^a	0.76^a	0.74^a	0.75^a	0.73^a

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.84^a	0.69^a	0.56^a	0.48^a	0.40^a	0.32^a	0.26^a	0.22^a
Pre-disinflation	0.77^a	0.65^a	0.54^a	0.44^a	0.37^a	0.36^a	0.34^a	—
Post-disinflation	0.73^a	0.51^a	0.44^a	0.39^a	0.38^a	0.37^a	0.37^a	0.36^a
Pre-Volcker	0.77^a	0.63^a	0.53^a	0.45^a	0.36^a	0.35^a	0.34^a	—
Volcker	0.86	0.68^a	0.55^a	0.48^a	0.42^a	0.37^a	0.40^a	0.38^a
Greenspan	0.74^a	0.55^a	0.44^a	0.38^a	0.37^a	0.34^a	0.31^a	0.31^a

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	1.00	0.98	0.94	0.90^c	0.88^b	0.87^b	0.86^b	0.81^b
Pre-disinflation	1.00	0.97	0.90	0.81^b	0.74^a	0.74^b	0.81	—
Post-disinflation	1.01	1.03	1.07	1.08	1.06	1.04	1.00	0.96
Pre-Volcker	1.00	0.97	0.89	0.81^b	0.75^a	0.73^b	0.81	—
Volcker	0.99	0.98	0.95	0.91	0.86	0.80^c	0.83	0.81
Greenspan	1.01	1.03	1.00	0.96	0.92	0.89	0.87	0.82^c

Reported value is the ratio of the standard deviation of the h-period ahead forecast to the standard deviation of the forecasted series for the specified sub-sample period. Superscripts ^{a}, ^{b}and ^{c} represent rejection at the 1%, 5% and 10% levels of significance, respectively, of the null hypothesis that the ratio equals one against the one-sided alternative that the ratio is less than one.

Table 3: Univariate Weak-Form Rationality Tests of Greenbook Forecast Values

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.68	0.62	0.60	0.50	0.53	0.48	0.07	0.00
Pre-disinflation	0.16	0.02	0.03	0.06	0.06	0.06	—	—
Post-disinflation	0.22	0.02	0.02	0.05	0.01	0.02	0.00	0.00
Pre-Volcker	0.11	0.02	0.03	0.07	0.07	0.08	—	—
Volcker	0.15	0.01	0.06	0.02	0.01	0.06	—	—
Greenspan	0.02	0.03	0.04	0.24	0.10	0.05	0.00	0.00

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.62	0.53	0.69	0.15	0.22	0.82	0.74	0.70
Pre-disinflation	0.36	0.33	0.36	0.18	0.31	0.28	—	—
Post-disinflation	0.03	0.19	0.30	0.22	0.34	0.39	0.17	0.29
Pre-Volcker	0.17	0.31	0.28	0.25	0.31	0.14	—	—
Volcker	0.30	0.29	0.37	0.46	0.85	0.69	—	—
Greenspan	0.00	0.20	0.18	0.03	0.15	0.31	0.36	0.33

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.26	0.13	0.25	0.40	0.33	0.44	0.49	0.75
Pre-disinflation	0.65	0.66	0.38	0.30	0.02	0.01	—	—
Post-disinflation	0.00	0.01	0.04	0.05	0.03	0.05	0.15	0.16
Pre-Volcker	0.63	0.77	0.50	0.31	0.02	0.03	—	—
Volcker	0.37	0.39	0.55	0.61	0.64	0.79	—	—
Greenspan	0.36	0.01	0.07	0.24	0.43	0.54	0.70	0.69

Reported value is the p-value of the joint null hypothesis that $\alpha = 0$ and $\beta = 1$ in the regression equation: $x_{t+h} = \alpha + \beta x^f_{t+h|t} + \varepsilon_{t+h}$.

Table 4: Multivariate Semi-Strong-Form Rationality Tests of Greenbook Forecast Values, 1 lag

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.00	0.00	0.01	0.03	0.24	0.08	0.08	0.00
Pre-disinflation	0.00	0.00	0.01	0.05	0.00	0.00	—	—
Post-disinflation	0.07	0.02	0.10	0.01	0.00	0.00	0.00	0.00
Pre-Volcker	0.00	0.00	0.01	0.06	0.00	0.00	—	—
Volcker	0.01	0.04	0.09	0.18	0.00	0.00	—	—
Greenspan	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.04	0.12	0.07	0.01	0.04	0.32	0.02	0.00
Pre-disinflation	0.15	0.35	0.01	0.11	0.17	0.04	—	—
Post-disinflation	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pre-Volcker	0.06	0.39	0.01	0.08	0.16	0.01	—	—
Volcker	0.00	0.02	0.00	0.00	0.00	0.80	—	—
Greenspan	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.13	0.26	0.18	0.10	0.04	0.15	0.00	0.00
Pre-disinflation	0.20	0.90	0.27	0.53	0.22	0.01	—	—
Post-disinflation	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Pre-Volcker	0.23	0.92	0.37	0.50	0.22	0.01	—	—
Volcker	0.46	0.70	0.66	0.68	0.12	0.00	—	—
Greenspan	0.20	0.02	0.00	0.00	0.00	0.00	0.00	0.00

Reported value is the p-value of the joint null hypothesis that $\alpha = 0$, $\beta = 1$, and $\gamma_{ij} = 0$ for all i, j in the regression equation: $x_{t+h} = \alpha + \beta x^f_{t+h|t} + \sum_{i=1}^n \sum_{j=1}^q \gamma_{ij} z_{i,t-j} + \varepsilon_{t+h}$, where $q = 1$.

Table 5: Multivariate Semi-Strong-Form Rationality Tests of Greenbook Forecast Values, 4 lags

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00
Pre-disinflation	0.00	0.00	0.00	0.00	0.00	0.00	—	—
Post-disinflation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pre-Volcker	0.00	0.00	0.00	0.00	0.00	0.00	—	—
Volcker	0.00	0.00	0.00	0.00	0.00	0.00	—	—
Greenspan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.18	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Pre-disinflation	0.01	0.02	0.00	0.00	0.00	0.00	—	—
Post-disinflation	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pre-Volcker	0.00	0.01	0.00	0.00	0.00	0.00	—	—
Volcker	0.00	0.00	0.00	0.00	0.00	0.00	—	—
Greenspan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pre-disinflation	0.13	0.00	0.00	0.00	0.00	0.00	—	—
Post-disinflation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pre-Volcker	0.13	0.00	0.00	0.00	0.00	0.00	—	—
Volcker	0.11	0.00	0.00	0.00	0.00	0.00	—	—
Greenspan	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Reported value is the p-value of the joint null hypothesis that $\alpha = 0$, $\beta = 1$, and $\gamma_{ij} = 0$ for all i, j in the regression equation: $x_{t+h} = \alpha + \beta x^f_{t+h|t} + \sum_{i=1}^{n} \sum_{j=1}^{q} \gamma_{ij} z_{i,t-j} + \varepsilon_{t+h}$, where $q = 4$.

Table 6: Multivariate Rationality Tests of Greenbook Forecast Errors

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.51	0.55	0.47	0.93	0.69	0.42	0.00	0.00
Pre-disinflation	0.22	0.08	0.05	0.04	0.00	—	—	—
Post-disinflation	0.03	0.02	0.01	0.21	0.01	0.00	0.00	0.00
Pre-Volcker	0.22	0.05	0.05	0.04	0.00	—	—	—
Volcker	0.56	0.00	0.02	0.00	0.00	0.00	—	—
Greenspan	0.00	0.01	0.01	0.27	0.01	0.00	0.00	0.00

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.25	0.60	0.39	0.73	0.66	0.14	0.27	0.00
Pre-disinflation	0.32	0.61	0.55	0.03	0.01	—	—	—
Post-disinflation	0.00	0.24	0.02	0.04	0.35	0.48	0.27	0.00
Pre-Volcker	0.28	0.50	0.50	0.02	0.00	—	—	—
Volcker	0.28	0.90	0.00	0.13	0.13	0.00	—	—
Greenspan	0.00	0.04	0.39	0.28	0.03	0.08	0.00	0.00

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	0.01	0.00	0.03	0.00	0.34	0.60	0.17	0.63
Pre-disinflation	0.00	0.06	0.00	0.07	0.04	—	—	—
Post-disinflation	0.27	0.00	0.08	0.00	0.00	0.10	0.17	0.63
Pre-Volcker	0.00	0.07	0.00	0.08	0.06	—	—	—
Volcker	0.17	0.05	0.00	0.08	0.07	0.01	—	—
Greenspan	0.26	0.00	0.08	0.00	0.00	0.00	0.00	0.63

Reported value is the p-value of the joint null hypothesis that $\alpha = 0$, and $\theta_{ij} = 0$ for all i, j in the regression equation: $v_{t+h|t} = \sum_{i=1}^{n} \sum_{j=h+1}^{q} \theta_{ij} v_{i,t+h-j|t-j} + \varepsilon_{t+h}$, where $v_{t+h|t} \equiv x^{f}_{t+h|t} - x_{t+h}$.

Table 7: Regression of Forecast Errors on Change in Trend of Forecasted Series

Inflation

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	−2.93^a	−6.64^a	−9.48^a	−10.29^a	−10.40^a	−11.81^a	−10.28^b	−10.61^a
Pre-disinflation	−2.86^b	−6.11^a	−9.19^a	−10.18^a	−10.18^a	−10.89^a	—	—
Post-disinflation	0.33	−6.20^a	−6.54^a	−9.10^a	−7.94^a	−11.10^a	−5.50^c	−6.66^b
Pre-Volcker	−2.87^b	−6.17^a	−9.22^a	−10.18^a	−10.22^a	−10.94^a	—	—
Volcker	−2.19	−3.49	−5.49^a	−4.21^b	−4.91^b	−8.52^b	—	—
Greenspan	−0.14	−7.17^a	−8.67^a	−9.91^a	−8.84^a	−8.57^a	−4.34	−2.70

Real Output Growth

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	−2.10^b	−4.41^a	−6.68^a	−7.13^a	−6.07^a	−6.09^b	−3.87	−5.63^b
Pre-disinflation	−2.03^c	−3.90^a	−6.04^a	−6.80^a	−5.49^a	−5.32	—	—
Post-disinflation	−1.93	−6.67^a	−8.37^a	−8.12^a	−7.93^a	−7.07^a	−4.03	−5.96
Pre-Volcker	−2.12^b	−4.02^a	−6.12^a	−6.60^a	−5.60^a	−5.68^c	—	—
Volcker	−1.23	−4.95^c	−7.97^a	−7.37^a	−4.90^c	−4.98^b	—	—
Greenspan	−2.01	−5.67^b	−8.18^b	−10.78^b	−11.24^b	−9.83^b	−8.90^a	−9.42^a

Unemployment Rate

	Forecast Horizon, h (quarters ahead)							
	0	1	2	3	4	5	6	7
Full Sample	−0.67^a	−1.28^c	−2.78^b	−4.24^a	−4.60^a	−4.82^b	−6.78^a	−9.28^a
Pre-disinflation	−0.72^b	−1.10	−2.47	−3.57^c	−2.84	0.56	—	—
Post-disinflation	−1.00^a	−1.96^b	−4.14^a	−6.03^a	−7.41^a	−9.06^a	−11.01^a	−11.32^a
Pre-Volcker	−0.72^b	−1.19	−2.58	−3.59^c	−2.89	0.42	—	—
Volcker	−0.40	−0.86	−1.88	−3.33	−3.95	−4.51	—	—
Greenspan	−1.26^a	−2.53^a	−5.45^a	−8.19^a	−10.72^a	−12.55^a	−13.96^a	−14.42^a

Reported value is the estimated coefficient of the regression of the h-period ahead forecast error on the change in the Hodrick-Prescott filtered trend of the variable being forecasted over the specified sub-sample period. Superscripts ^{a}, ^{b}and ^{c} represent rejection at the 1%, 5% and 10% levels of significance, respectively, of the null hypothesis that the estimated coefficient is zero.