Time-dependent Bias in the Fed's Greenbook Forecasts*

Neil R. Ericsson† Stedman B. Hood‡ Fred Joutz§ Tara M. Sinclair¶
Herman O. Stekler‖

Abstract
Building on Sinclair, Joutz, and Stekler (2010), this paper examines the Federal Reserve Board's Greenbook forecasts of U.S. output growth, inflation, and the unemployment rate for potential biases. Standard tests typically fail to detect biases in one-quarter-ahead forecasts. However, impulse indicator saturation (IIS) detects economically large and highly significant time-varying biases. Biases depend on the variable being forecast and the phase of the business cycle. IIS defines a generic procedure for examining forecast properties, it explains why standard tests fail to detect bias, and it provides a potential mechanism for improving forecasts.

Key Words: Autometrics, bias, Federal Reserve, forecasts, GDP, Greenbook, impulse indicator saturation, inflation, Tealbook, unemployment, United States

1. Introduction

The Fed's monetary policy—including recent large-scale asset purchases and forward guidance—has attracted considerable attention domestically and abroad; see Bernanke (2012) and Yellen (2012) *inter alia* for recent discussions. Monetary policy decisions at the Fed are based in part on the "Greenbook" forecasts, which are economic forecasts produced by the Fed's staff. The Greenbook forecasts have been extensively analyzed, including by Romer and Romer (2008), Sinclair, Joutz, and Stekler (2010), and Nunes (2013).

A central focus in forecast evaluation is forecast bias, especially because forecast bias is systematic, and because ignored forecast biases may have substantive adverse consequences for policy. Building on Sinclair, Joutz, and Stekler (2010), the current paper analyzes Greenbook forecasts of U.S. output growth, inflation, and the unemployment rate for potential biases over 1966Q1–2007Q4. Standard tests typically fail to detect any important biases in one-quarter-ahead forecasts. However, a recently developed technique—impulse indicator saturation—detects economically large and highly statistically significant time-varying biases that depend on the phase of the

*The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The authors are grateful to David Hendry, Ricardo Nunes, Steve Sharpe, and Bob Tetlow for helpful discussions and comments. All numerical results were obtained using PcGive Version 14.0B3, Autometrics Version 1.5e, and Ox Professional Version 7.00 in 64-bit OxMetrics Version 7.00: see Doornik and Hendry (2013) and Doornik(2009).

†Division of International Finance, Board of Governors of the Federal Reserve System, Washington, DC 20551 USA (ericsson@frb.gov), and Research Program on Forecasting, Department of Economics, The George Washington University, Washington, DC 20052 USA (ericsson@gwu.edu)
‡PlateJoy, San Francisco, California USA (stedmanhood@gmail.com)
§Department of Economics, The George Washington University, Washington, DC 20052 USA (bmark@gwu.edu)
¶Department of Economics, The George Washington University, Washington, DC 20052 USA (tsinc@gwu.edu)
‖Department of Economics, The George Washington University, Washington, DC 20052 USA (hstekler@gwu.edu)

business cycle. Biases differ across the variable being forecast and the phase of the business cycle. IIS defines a generic procedure for examining forecast properties; it explains why standard tests fail to detect bias; and it provides a potential mechanism for improving forecasts.

This paper is organized as follows. Section 2 describes the data and the forecasts being analyzed. Section 3 discusses different approaches to testing for potential forecast bias and proposes impulse indicator saturation as a generic test of forecast bias. Section 4 describes indicator saturation techniques, including impulse indicator saturation and several of its extensions. Section 5 presents evidence on forecast bias, using the methods detailed in Sections 3 and 4. Section 6 concludes.

2. The Data and the Forecasts

U.S. output growth, inflation, and the unemployment rate are three key forecasts in the Greenbook produced by the staff of the Federal Reserve Board (the "Fed"). This section describes those Greenbook forecasts and the data being forecast. See Sinclair, Joutz, and Stekler (2010) for further details.

The data being forecast are:

- output growth (Δy),
- inflation (Δp), and
- the unemployment rate (U),

where all are quarterly for the United States over 1966Q1–2007Q4 (168 observations). Output growth and inflation are quarterly rates expressed as percent changes at an annual rate, and the unemployment rate is in percent. Output is GNP initially and GDP from 1991Q4 onwards; inflation is correspondingly calculated from the deflator for GNP or GDP. The values of the data are NIPA, as released approximately 45 days after the beginning of the new quarter. The data are publicly available from the databases FRED (Federal Reserve Economic Data, http://research.stlouisfed.org/fred2/) and ALFRED (ArchivaL Federal Reserve Data, http://alfred.stlouisfed.org/), both maintained by the Federal Reserve Bank of St. Louis.

The Greenbook forecasts are from the final Greenbook of the quarter so as to allow as much information to be available for the forecasts being made in a given quarter. Greenbooks dated within the first 10 days of the subsequent quarter are considered made in the previous quarter because little new information on the data being forecast would have accrued since the end of the previous quarter; see Sinclair, Joutz, and Stekler (2010, footnote 4). The forecasts are for one quarter ahead: Sinclair, Joutz, and Stekler (2010) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2013) analyze the current-quarter forecasts as well. The Greenbook forecasts are publicly available from the Federal Reserve Bank of Philadelphia:

http://www.phil.frb.org/research-and-data/real-time-center/greenbook-data/pdf-data-set.cfm;

These forecasts are made publicly available approximately five years after the fact. The assumptions underlying the Greenbook forecasts, the complex process involved in generating the forecasts, and the goals and objectives of that process are of considerable interest in their own right and merit detailed examination. However, in the spirit of Stekler (1972), Chong and Hendry (1986), and Fildes and Stekler (2002) *inter alia*, the current paper focuses on the properties of the forecasts themselves.

Two constructed variables are also used in the analysis below. The first is a dummy variable ("*NBER*") for the contractionary phases of the business cycle, as dated by the National Bureau of Economic Research (2012). The second is a dummy variable ("*ECRI*") for periods of "inflationary pressure", as determined by the Economic Cycle Research Institute (ECRI; co-founded by Geoffrey H. Moore and Lakshman Achuthan) and obtained from Dr. Anirvan Banerji at the ECRI.

The National Bureau of Economic Research (2012) reports the NBER's turning-point events ("peak" or "trough") relevant to the sample, the date of the event (the "reference date" in the NBER's terminology), the date on which the NBER announced the determination of that event, and the length of time taken to determine that an event had occurred. The corresponding zero-one phase indicator dummies are denoted E_t^{1961}, C_t^{1969}, E_t^{1970}, C_t^{1973}, E_t^{1975}, C_t^{1980}, E_t^{1980}, C_t^{1981}, E_t^{1982}, C_t^{1990}, E_t^{1991}, C_t^{2001}, E_t^{2001}, C_t^{2007}, and E_t^{2009}, where E and C denote expansion and contraction. The phase indicator's superscript designates the calendar year of the beginning of the designated phase, i.e., the calendar year of the reference date; and the phase indicator itself is unity during its phase and zero otherwise. The dummy variable *NBER* is the sum of all C_t^i. The *ECRI* dummy is similarly constructed, but for the periods of inflationary pressure.

Several properties of the data, the Greenbook forecasts, and the corresponding forecast errors are apparent from their graphs. Figure 1 plots actual output growth, inflation, the unemployment rate, and their one-quarter-ahead ("T1") Greenbook forecasts. Figure 2 plots the corresponding forecast errors. For all three variables, the actual and forecast values appear close, reflecting in part the scale of the graphs. One-quarter-ahead forecast errors for output growth and inflation are often small—under 1% per annum in absolute value—but sometimes they are much larger, and with the magnitude and sign of the forecast errors differing over time. Some persistence in the forecast errors is visible, particularly for one-quarter-ahead inflation forecasts during the early 1970s, the late 1990s, and the early 2000s. That persistence is suggestive of systematic biases in the forecasts. For some previous analyses of these and other governmental and institutional forecasts, see Corder (2005), Engstrom and Kernell (1999), Frankel (2011), Joutz and Stekler (2000), Nunes (2013), Sinclair, Joutz, and Stekler (2010), Romer and Romer (2008), and Tsuchiya (2013). Ericsson, Fiallos, and Seymour (2015) analyze the Greenbook forecasts of *foreign* real GDP growth and, using a methodology similar to that below, detect and estimate time-dependent biases.

3. Approaches for Detecting Forecast Bias

This section considers different approaches for assessing potential forecast bias, starting with the standard test of (time-invariant) forecast bias by Mincer and Zarnowitz (1969). This section then considers forms of time-dependent forecast bias, with impulse indicator saturation providing a generic test of potentially time-varying forecast bias. The exposition herein draws on Ericsson (2015).

Mincer and Zarnowitz (1969, pp. 8–11) suggest testing for forecast bias by regressing the forecast error on an intercept and testing whether the intercept is statistically significant. That is, for a variable x_t at time t and its forecast $\hat{x}_t$, estimate the equation:

$$(x_t - \hat{x}_t) \; = \; a \; + \; e_t \qquad t = 1, \ldots, T, \tag{1}$$

where a is the intercept, e_t is the error term at time t, and T is number of observations. A test of $a = 0$ is interpretable as a test that the forecast $\hat{x}_t$ is unbiased for the variable x_t. For current-period and one-step-ahead forecasts, the error e_t may

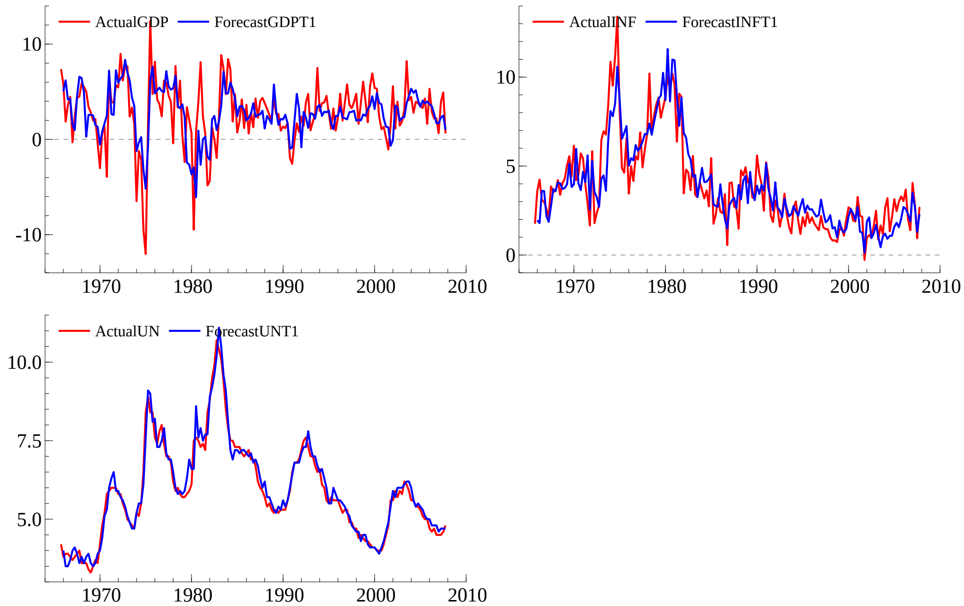

Figure 1: Output growth, inflation, the unemployment rate, and their one-quarter-ahead ("T1") forecasts.

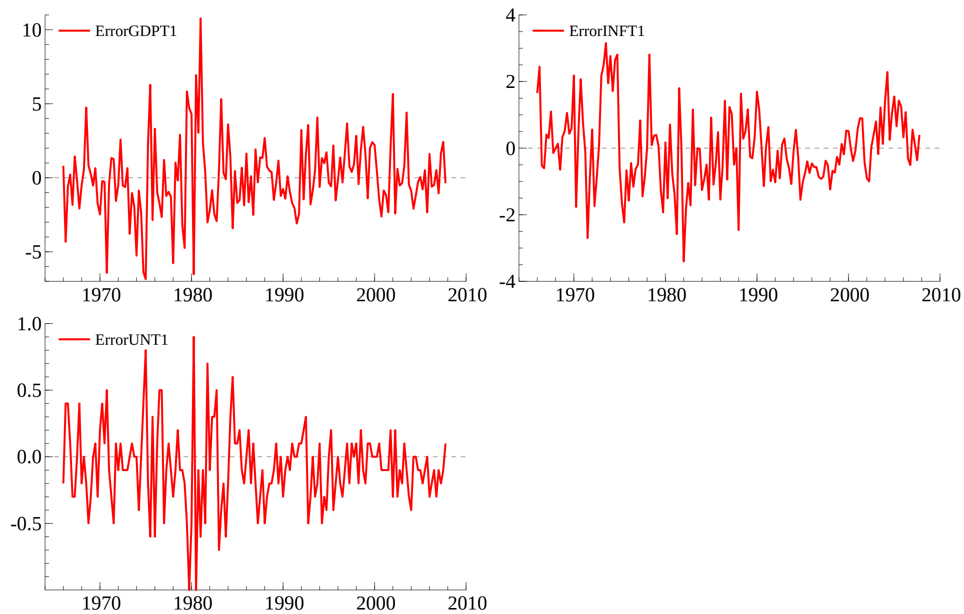

Figure 2: The one-quarter-ahead ("T1") forecast errors for the growth rate, inflation, and the unemployment rate.

be serially uncorrelated, in which case a t- or F-statistic may be appropriate. For multi-step-ahead forecasts, e_t generally will be serially correlated; hence inference about the intercept a may require some accounting for that autocorrelation.

Holden and Peel (1990) and Stekler (2002) discuss a generalization of equation (1):

$$(x_t - \hat{x}_t) \quad = \quad b_0 + b_1' z_t + e_t \qquad t = 1, \ldots, T, \tag{2}$$

in which the right-hand side variables z_t might be any variables; and they interpret a test of $b_1 = 0$ as a test of efficiency. See Holden and Peel (1990) and Stekler (2002) for expositions on these tests as tests of unbiasedness and efficiency, and Sinclair, Stekler, and Carnow (2012) for a recent discussion.

Many forecast tests are interpretable as being based on equation (2). For example, in Sinclair, Joutz, and Stekler (2010), the regressor z_t includes *NBER* or *ECRI*, each indicative of business-cycle phenomena. Another choice of z_t is $\hat{x}_t$, proposed by Mincer and Zarnowitz (1969, p. 11). Below, "Mincer–Zarnowitz A" denotes the regression-based test of $a = 0$ in equation (1), whereas "Mincer–Zarnowitz B" denotes the regression-based test of $\{b_0 = 0, b_1 = 0\}$ in equation (2) with $z_t = \hat{x}_t$. Other choices for z_t include an alternative forecast $\tilde{x}_t$ or the differential between the two forecasts $(\tilde{x}_t - \hat{x}_t)$, generating the forecast-encompassing tests in Chong and Hendry (1986). As Ericsson (1992) discusses, a necessary condition for forecast encompassing is having the smallest mean squared forecast error (MSFE); Granger (1989) and Diebold and Mariano (1995) propose tests of whether one model's MSFE is less than another model's MSFE. Also, the "alternative forecast" could be a forecast made in a different time period, in which case $(\tilde{x}_t - \hat{x}_t)$ is the revision of the forecast. Nordhaus (1987) proposes this test based on forecast revisions across multiple horizons as a test of efficiency. Tversky and Kahneman (1974) earlier described "anchoring" as a phenomenon in which $b_1 > 0$ for forecast revisions; see Campbell and Sharpe (2009) for empirical evidence on anchoring.

In equation (2), the term $(b_0 + b_1' z_t)$ is also interpretable as a specific form of time-dependent forecast bias. That time dependence could be completely general, as follows:

$$\begin{aligned}(x_t - \hat{x}_t) \quad &= \quad a_t + e_t \\ &= \quad \textstyle\sum_{i=1}^{T} c_i I_{it} + e_t \qquad t = 1, \ldots, T,\end{aligned} \tag{3}$$

where the impulse indicator I_{it} is a dummy variable that is unity for $t = i$ and zero otherwise, and c_i is the corresponding coefficient for I_{it}. Because the $\{c_i\}$ may have any values whatsoever, the intercept a_t in (3) may vary arbitrarily over time. In this context, a test that all coefficients c_i are equal to zero is a generic test of forecast unbiasedness. Because equation (3) includes T coefficients, equation (3) cannot be estimated unrestrictedly. However, the question being asked can be answered using impulse indicator saturation, as summarized in Section 4.

4. Indicator Saturation Techniques

Impulse indicator saturation (IIS) uses the zero-one dummies $\{I_{it}\}$ to analyze properties of a model. Unrestricted inclusion of all T dummies in the model (thereby "saturating" the sample) is infeasible. However, blocks of dummies *can* be included, and statistically significant dummies can be retained from those blocks. That insight provides the basis for IIS. See Ericsson and Reisman (2012) for an intuitive nontechnical exposition of IIS, and Hendry and Doornik (2014) for extensive analysis in the context of automatic model selection.

Table 1: Impulse indicator saturation and two extensions, as characterized by the variables involved.

Name	Description	Variables	Definition of variables
Impulse indicator saturation	Zero-one dummies	$\{I_{it}\}$	$I_{it} = 1$ for $t = i$, zero otherwise
Super saturation	Step functions	$\{I_{it}, S_{it}\}$	$S_{it} = 1$ for $t \geq i$, zero otherwise
Ultra saturation	Broken linear trends	$\{I_{it}, S_{it}, T_{it}\}$	$T_{it} = t - i + 1$ for $t \geq i$, zero otherwise

IIS provides a general procedure for robust estimation and for model evaluation—in particular, for testing parameter constancy. IIS is a generic test for an unknown number of structural breaks, occurring at unknown times, with unknown duration and magnitude, anywhere in the sample. IIS is a powerful empirical tool for both evaluating and improving existing empirical models. Hendry (1999) proposes IIS as a procedure for testing parameter constancy. Further discussion, recent developments, and applications appear in Hendry, Johansen, and Santos (2008), Doornik (2009), Johansen and Nielsen (2009, 2013, 2015), Hendry and Santos (2010), Ericsson (2011a, 2011b, 2012), Ericsson and Reisman (2012), Bergamelli and Urga (2013), Hendry and Pretis (2013), Hendry and Doornik (2014), Pretis, Mann, and Kaufmann (2015), and Castle, Doornik, Hendry, and Pretis (2015). Ericsson (2015) proposes a new application for IIS—as a generic test for time-varying forecast bias. Section 5 applies IIS to test for potential bias in the Greenbook forecasts.

Many existing procedures can be interpreted as "special cases" of IIS in that they represent particular algorithmic implementations of IIS. Such special cases include recursive estimation, rolling regression, the Chow (1960) predictive failure statistic (including the 1-step, breakpoint, and forecast versions implemented in OxMetrics), the Andrews (1993) unknown breakpoint test, the Bai and Perron (1998) multiple breakpoint test, tests of extended constancy in Ericsson, Hendry, and Prestwich (1998, pp. 305ff), tests of nonlinearity, intercept correction (in forecasting), and robust estimation. IIS thus provides a general and generic procedure for analyzing a model's constancy. Algorithmically, IIS also solves the problem of having more potential regressors than observations by testing and selecting over blocks of variables.

Table 1 summarizes IIS and two extensions: super saturation and ultra saturation. Throughout, T is the sample size, t is the index for time, i and j are the indexes for indicators, k is the index for economic variables (denoted x_{kt}), and K is the total number of potential regressors considered. A few remarks may be helpful for interpreting the entries in Table 1.

Impulse indicator saturation. This is the standard IIS procedure proposed by Hendry (1999), with selection among the T zero-one impulse indicators $\{I_{it}\}$.

Super saturation. Super saturation searches across all possible one-off step functions $\{S_{it}\}$, in addition to $\{I_{it}\}$. Step functions are of economic interest because they may capture permanent or long-lasting changes that are not otherwise incorporated into a specific empirical model. A step function is a partial sum of impulse indicators; equivalently, it is a parsimonious representation of a sequential subset of impulse indicators that have equal coefficients. Castle, Doornik, Hendry, and Pretis (2015) investigate the statistical properties of a closely related saturation estimator—step indicator saturation (SIS)—which searches among only the step indicator variables $\{S_{it}\}$. Autometrics now includes IIS, SIS, super saturation (IIS+SIS), and zero-sum pairwise IIS (mentioned below); see Doornik and Hendry (2013).

Ultra saturation. Ultra saturation (earlier, sometimes called "super duper" saturation) searches across $\{I_{it}, S_{it}, T_{it}\}$, where the $\{T_{it}\}$ are broken linear trends. Broken linear trends may be of economic interest; mathematically, the $\{T_{it}\}$ are partial sums of the partial sums of impulse indicators. Broken quadratic trends, broken cubic trends, and higher-order broken trends are also feasible.

Table 1 is by no means an exhaustive list of extensions to IIS. Other extensions include sequential ($j = 1$) and non-sequential ($j > 1$) pairwise impulse indicator saturation for an indicator P_{it}, defined as $I_{it} + I_{i+j,t}$; sequential multiplet indicator saturation for an indicator M_{it}^{j+1}, defined as $I_{it} + \cdots + I_{i+j,t}$ for $j \geq 1$; zero-sum pairwise IIS for an indicator Z_{it}, defined as ΔI_{it}; many many variables for a set of K potential regressors $\{x_{kt},\ k = 1, \ldots, K\}$ for $K > T$; factors; principal components; and multiplicative indicator saturation for the set of $S_{it}x_{kt}$. See Castle, Clements, and Hendry (2013) and Ericsson (2011b, 2012) for details, discussion, and examples in the literature. Also, the IIS-type procedure chosen may itself be a combination of extensions; and that choice may affect the power of the procedure to detect specific alternatives. Notably, the *NBER* and *ECRI* dummies are examples of sequential multiplets.

As a more general observation, different types of indicators are adept at characterizing different sorts of bias: impulse dummies $\{I_{it}\}$ for date-specific anomalies, step dummies $\{S_{it}\}$ for level shifts, and broken trends $\{T_{it}\}$ for evolving developments. Transformations of the variable being forecast also may affect the interpretation of the retained indicators. For instance, an impulse dummy for a growth rate implies a level shift for the (log) level of the variable.

IIS-based tests of forecast bias can serve both as diagnostic tools to detect what is wrong with the forecasts, and as developmental tools to suggest how the forecasts can be improved. Clearly, "rejection of the null doesn't imply the alternative". However, for time series data, the date-specific nature of IIS-type procedures can aid in identifying important sources of forecast error. Use of these tests in forecast development is consistent with a progressive modeling approach; see White (1990).

As equation (3) emphasizes, IIS-based tests generalize the Mincer–Zarnowitz tests to allow for arbitrarily time-varying forecast bias. This observation and the observations above highlight the strength of the Mincer–Zarnowitz tests (that they focus on detecting a constant nonzero forecast bias) and also their weakness (that they assume that the forecast bias *is* constant over time). These characteristics of the Mincer–Zarnowitz tests bear directly on the empirical results in Section 5.

5. Evidence on Biases in the Greenbook Forecasts

This section examines the one-quarter-ahead Greenbook forecasts of U.S. output growth, inflation, and the unemployment rate for potential bias. Ericsson, Hood, Joutz, Sinclair, and Stekler (2013) present results on the current-period forecasts as well. Standard (Mincer–Zarnowitz) tests of forecast bias typically fail to detect economically and statistically important biases. By contrast, IIS-type tests detect large time-varying biases that are associated with the phase of the business cycle. Forecast biases differ numerically across the forecast horizon and variable being forecast, albeit with some qualitative similarities.

Section 5.1 reports standard Mincer–Zarnowitz tests of bias for the Greenbook forecasts of output growth, inflation, and the unemployment rate. Section 5.2 replicates results in Sinclair, Joutz, and Stekler (2010) and re-estimates on a longer sample. Section 5.3 employs IIS-type procedures to test for and estimate time-varying forecast bias. Throughout, regressions are on the full sample (1966Q1–2007Q4) unless otherwise indicated; and the dependent variable in a regression is the forecast error.

5.1 Standard Tests of Forecast Bias

This subsection examines the Greenbook forecasts for bias using the standard (Mincer–Zarnowitz) tests, and it finds little evidence of economically and statistically important biases.

Table 2 reports the Mincer–Zarnowitz regressions in equations (1) and (2) for the Greenbook forecasts, with columns alternating between the "A" and "B" versions of the Mincer–Zarnowitz regression. Here and in subsequent tables, OLS and HAC estimated standard errors appear under regression coefficients in parentheses $(\cdot)$ and square brackets $[\cdot]$ respectively, t-ratios appear in curly brackets $\{\cdot\}$, and $\hat{\sigma}$ denotes the residual standard error. For the Mincer–Zarnowitz statistics in Table 2, and for other test statistics in tables below, the entries within a given block of numbers are the F-statistic for testing the null hypothesis against the designated maintained hypothesis, the tail probability associated with that value of the F-statistic (in square brackets), the degrees of freedom for the F-statistic (in parentheses), and (for IIS-type statistics) the retained dummy variables. Superscript asterisks * and ** denote rejections of the null hypothesis at the 5% and 1% levels respectively, and the null hypothesis typically includes setting the coefficient on the intercept to zero. Doornik and Hendry (2013) provide a description of the residual diagnostic statistics. For the IIS-type statistics reported below, K is the number of *potential* regressors for selection, and the target size is chosen much smaller than $1/K$ in order to help ensure that few if any indicators are retained fortuitously.

Table 2 reports Mincer–Zarnowitz regressions for the one-quarter-ahead forecasts. Statistically, there is no evidence of forecast bias for the forecasts of output growth and inflation. Bias is detected for forecasts of the unemployment rate, but that bias appears to be small, e.g., estimated as -0.075% from the Mincer–Zarnowitz A regression.

Table 2: Coefficients, estimated standard errors, t-ratios, and summary statistics for Mincer–Zarnowitz regressions of the one-quarter-ahead forecast errors.

Regressor or statistic	Δy	Δy	Δp	Δp	U	U
Intercept	0.02 (0.20) {0.08}	0.42 (0.30) {1.42}	−0.011 (0.090) {−0.12}	0.21 (0.17) {1.22}	−0.075 (0.023) {−3.32}	0.072 (0.091) {0.79}
Forecast $\hat{x}_t$	–	−0.147 (0.081) {−1.82}	–	−0.057 (0.038) {−1.51}	–	−0.025 (0.015) {−1.66}
$\hat{\sigma}$	2.561%	2.543%	1.170%	1.166%	0.293%	0.291%
RMSE of the forecast	2.553%	2.553%	1.167%	1.167%	0.301%	0.301%
Mincer–Zarnowitz test statistic	0.01 [0.936] $F(1,167)$	1.66 [0.194] $F(2,166)$	0.01 [0.904] $F(1,167)$	1.14 [0.322] $F(2,166)$	11.04** [0.001] $F(1,167)$	6.96** [0.001] $F(2,166)$

Table 3: Coefficients, estimated standard errors, and summary statistics for regressions of one-quarter-ahead forecast errors with NBER or ECRI dummies.

Regressor or statistic	Δy	Δy (SJS)	Δp	Δp (SJS)	U	U (SJS)
Intercept	1.866 [0.570]	1.916 [0.585]	−0.311 [0.287]	−0.446 [0.301]	0.062 [0.093]	0.082 [0.096]
Forecast $\hat{x}_t$	−0.4520 [0.1579]	−0.4511 [0.1601]	−0.0320 [0.0638]	−0.0116 [0.0649]	−0.0285 [0.0168]	−0.0309 [0.0170]
NBER	−3.625 [0.555]	−3.744 [0.578]	–	–	0.198 [0.075]	0.194 [0.079]
ECRI	–	–	0.668 [0.213]	0.665 [0.222]	–	–
$\hat{\sigma}$	2.282%	2.356%	1.125%	1.144%	0.282%	0.294%
F statistic	15.09** [0.000] $F(3,165)$	14.24** [0.000] $F(3,150)$	5.22** [0.002] $F(3,165)$	4.40** [0.005] $F(3,150)$	8.72** [0.000] $F(3,165)$	7.29** [0.000] $F(3,150)$

5.2 Business-cycle Dependence of Forecast Bias

Sinclair, Joutz, and Stekler (2010) generalize the Mincer–Zarnowitz regression to include an indicator of the business-cycle phase, thereby parameterizing a very specific and economically interesting form of time-varying bias. Sinclair, Joutz, and Stekler (2010) find strong evidence for phase-dependent bias in one-quarter-ahead forecasts.

Table 3 replicates those results from Sinclair, Joutz, and Stekler (2010) in the columns labeled "SJS". Each adjacent column to the left re-estimates the same model on a longer sample (1966Q1–2007Q4), noting that the sample in Sinclair, Joutz, and Stekler (2010) ends in 2003Q3. The evidence on bias from Sinclair, Joutz, and Stekler (2010) is virtually unchanged on the longer sample. Forecast bias appears strongly phase-dependent for one-quarter-ahead forecasts. Furthermore, that forecast bias appears to be economically large and highly statistically significant, even while the standard Mincer–Zarnowitz tests failed to detect bias or found biases that were numerically small. These discrepancies in results are explained by the omission of a phase-dependent regressor in the Mincer–Zarnowitz regressions, thereby reducing the power of the corresponding tests to detect forecast bias. By contrast, the tests that focus on cyclicality detect bias in all three one-quarter-ahead forecasts.

5.3 Estimated Time-varying Bias Using Indicator Saturation

To assess more generally the time dependence of the forecast biases, this subsection estimates IIS-type equations of the form in equation (3), conditioning on the inclusion of the *NBER* or *ECRI* dummy. Additional time dependence in the one-quarter-ahead forecast biases is detected, although the dominant effect appears to be the business-cycle phase (or, for inflation, the degree of "inflationary pressure"). Because biases appear more substantial for the one-quarter-ahead forecasts than for the current-period forecasts, this subsection examines the one-quarter-ahead forecasts.

Table 4 reports the "focused" phase-based statistics (from Table 3) and the IIS-type test statistics of forecast bias for the one-quarter-ahead Greenbook forecasts. Table 4 also includes the Mincer–Zarnowitz statistics for comparison. The phase-based statistic detects bias for all three forecasts; see Section 5.2. IIS and its extensions *always* detect additional bias, and they do so for historically and economically consequential years. These estimated biases depend on the selected indicator dummies, in addition to the *NBER* or *ECRI* dummy. Figure 3 plots the implied estimated biases, along with the forecast errors. The graphs highlight the phase-dependent, persistent nature of these forecast biases.

As Figure 3 shows, forecast biases vary markedly over time, being sometimes positive and othertimes negative. The Mincer–Zarnowitz tests have particular difficulty in detecting such biases because the Mincer–Zarnowitz tests average all biases (both negative and positive) over time, and because the Mincer–Zarnowitz tests assign any time variation in bias to the residual rather than to the bias itself. As an extreme example, the Mincer–Zarnowitz A test has no power to detect a forecast bias that is $+10\%$ for the first half of the sample and -10% for the second half of

Table 4: Statistics for testing for bias in the one-quarter-ahead Greenbook forecasts.

Statistic (target size)	K	Δy	Δp	U
Mincer–Zarnowitz A	1	0.01 [0.936] $F(1,167)$	0.01 [0.904] $F(1,167)$	11.04** [0.001] $F(1,167)$
Mincer–Zarnowitz B	2	1.66 [0.194] $F(2,166)$	1.14 [0.322] $F(2,166)$	6.96** [0.001] $F(2,166)$
"Focused" [$NBER$or $ECRI$]	3	15.09** [0.000] $F(3,165)$	5.22** [0.002] $F(3,165)$	8.72** [0.000] $F(3,165)$
IIS (0.3%)	168	15.94** [0.000] $F(3,163)$ $I1980Q1, I1980Q3,$ $I1981Q1$	7.82** [0.000] $F(5,161)$ $I1970Q4, I1973Q3,$ $I1974Q1,$ $I1974Q4, I1982Q1$	12.5** [0.000] $F(4,162)$ $I1975Q1, I1979Q4,$ $I1980Q2, I1980Q3$
Super saturation (0.15%)	334	17.20** [0.000] $F(5,161)$ $I1975Q3, I1980Q3,$ $I1981Q1,$ $S1979Q3, S1980Q2$	24.32** [0.000] $F(2,164)$ $S1973Q1, S1975Q1$	10.8** [0.000] $F(10,156)$ 10 *dummies*
Ultra saturation (0.1%)	501	17.59** [0.000] $F(4,162)$ $I1980Q3, I1981Q1,$ $S1979Q3, S1980Q2$	12.68** [0.000] $F(5,161)$ $I1974Q4, I1978Q2,$ $S1973Q1,$ $T1972Q2, T1975Q1$	9.90** [0.000] $F(8,158)$ 8 *dummies*
$\hat{\sigma}$ (ultra saturation)		2.076%	0.967%	0.238%
RMSE of the forecast		2.553%	1.167%	0.301%

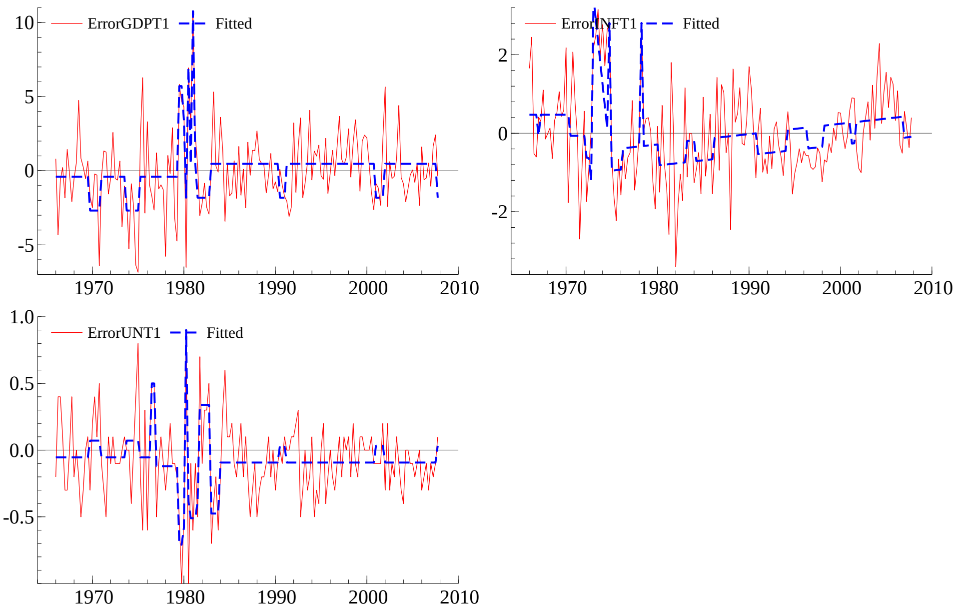

Figure 3: Forecast errors, and the estimates of one-quarter-ahead forecast bias for ultra-saturation with NBER and ECRI dummies.

the sample, even though this bias would be obvious from (e.g.) graphing the data.

The forecast biases are numerically and economically consequential. Specifically, elimination of these time-varying biases would reduce the root mean square forecast error by approximately 20% for each of the three one-quarter-ahead forecasts, as implied by the forecasts' RMSEs and the values of $\hat{\sigma}$ for ultra saturation in Table 4.

For post-1980 output growth, the estimated forecast bias is approximately $+0.5\%$ during expansions and -1.5% during contractions. These biases are economically substantial, noting that output growth averages approximately $+2.8\%$ over the entire sample. One possible explanation for such biases is an asymmetric loss function for forecasting, as discussed in Sinclair, Joutz, and Stekler (2010). Another possible explanation arises from constructing unconditional forecasts in the face of possible regime switches, as between expansions and contractions. The resulting forecast errors could then be bimodal, in which case minimizing the mean square forecast error is no longer an obvious choice for constructing forecasts. In fact, unconditionally unbiased forecasts may generate "unlikely" point forecasts, e.g., ones that lie between two modes of the density. This implication reflects the weighting of two potential disparate outcomes. Equally, this highlights the informational losses in a point forecast, relative to an entire forecast distribution.

In summary, one-quarter-ahead Greenbook forecasts of U.S. output growth, inflation, and the unemployment rate exhibit time-varying biases that are primarily associated with the phase of the business cycle. Biases are not the same across variables, nor are they the same in magnitude for expansions and contractions. Biases appear little affected by other factors. Current-quarter forecasts of output growth exhibit some anchoring bias.

The presence of forecast bias implies the potential for improved forecasts. The feasibility of improvement may depend on the information available when making the forecasts. Because the estimated forecast biases depend primarily on the phase of the business cycle, and because expansions have historically been very long, the prospect is promising for some improvement. Furthermore, the potential improvement could be substantial: ultra saturation in Table 4 achieves approximate 20% reductions in RMSEs for all three one-quarter-ahead forecasts. From an institutional perspective, it may be useful to isolate the causes of the forecast errors according to the various assumptions made about the paths of other economic variables. Such an analysis could lead to improved forecasts, or at least provide a deeper understanding of the sources of forecast error.

6. Conclusions

Building on Sinclair, Joutz, and Stekler (2010), the current paper analyzes Greenbook forecasts of U.S. output growth, inflation, and the unemployment rate for potential biases over 1966Q1–2007Q4. Standard tests typically fail to detect biases in one-quarter-ahead forecasts. However, impulse indicator saturation detects economically large and highly statistically significant time-varying biases. Biases depend on the variable being forecast and the phase of the business cycle.

REFERENCES

Andrews, D. W. K. (1993) "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61, 4, 821–856.

Bai, J., and P. Perron (1998) "Estimating and Testing Linear Models with Multiple Structural Changes", *Econometrica*, 66, 1, 47–78.

Bergamelli, M., and G. Urga (2013) "Detecting Multiple Structural Breaks: A Monte Carlo Study and an Application to the Fisher Equation for the US", draft, Cass Business School, London, March.

Bernanke, B. S. (2012) "U.S. Monetary Policy and International Implications", remarks at the seminar "Challenges of the Global Financial System: Risks and Governance under Evolving Globalization", Bank of Japan, Tokyo, Japan, October 14.

Campbell, S. D., and S. A. Sharpe (2009) "Anchoring Bias in Consensus Forecasts and Its Effect on Market Prices", *Journal of Financial and Quantitative Analysis*, 44, 2, 369–390.

Castle, J. L., M. P. Clements, and D. F. Hendry (2013) "Forecasting by Factors, by Variables, by Both or Neither?", *Journal of Econometrics*, 177, 2, 305–319.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015) "Detecting Location Shifts During Model Selection by Step-indicator Saturation", *Econometrics*, 3, 2, 240–264.

Chong, Y. Y., and D. F. Hendry (1986) "Econometric Evaluation of Linear Macro-economic Models", *Review of Economic Studies*, 53, 4, 671–690.

Chow, G. C. (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 3, 591–605.

Corder, J. K. (2005) "Managing Uncertainty: The Bias and Efficiency of Federal Macroeconomic Forecasts", *Journal of Public Administration Research and Theory*, 15, 1, 55–70.

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

Doornik, J. A. (2009) "Autometrics", Chapter 4 in J. L. Castle and N. Shephard (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford, 88–121.

Doornik, J. A., and D. F. Hendry (2013) *PcGive 14*, Timberlake Consultants Press, London (3 volumes).

Engstrom, E. J., and S. Kernell (1999) "Serving Competing Principals: The Budget Estimates of OMB and CBO in an Era of Divided Government", *Presidential Studies Quarterly*, 29, 4, 820–829.

Ericsson, N. R. (1992) "Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration", *Journal of Policy Modeling*, 14, 4, 465–495.

Ericsson, N. R. (2011a) "Improving Global Vector Autoregressions", draft, Board of Governors of the Federal Reserve System, Washington, D.C., June.

Ericsson, N. R. (2011b) "Justifying Empirical Macro-econometric Evidence in Practice", invited presentation, online conference *Communications with Economists: Current and Future Trends* commemorating the 25th anniversary of the *Journal of Economic Surveys*, November.

Ericsson, N. R. (2012) "Detecting Crises, Jumps, and Changes in Regime", draft, Board of Governors of the Federal Reserve System, Washington, D.C., November.

Ericsson, N. R. (2015) "How Biased Are U.S. Government Forecasts of the Federal Debt?", *International Journal of Forecasting*, forthcoming.

Ericsson, N. R., E. J. Fiallos, and J. E. Seymour (2015) "Detecting Time-dependent Bias in the Fed's Greenbook Forecasts of Foreign GDP Growth", in *JSM Proceedings,* Business and Economic Statistics Section, American Statistical Association, Alexandria, VA, forthcoming.

Ericsson, N. R., D. F. Hendry, and K. M. Prestwich (1998) "The Demand for Broad Money in the United Kingdom, 1878–1993", *Scandinavian Journal of Economics*, 100, 1, 289–324 (with discussion).

Ericsson, N. R., S. B. Hood, F. Joutz, T. M. Sinclair, and H. O. Stekler (2013) "Greenbook Forecasts and the Business Cycle", draft, Board of Governors of the Federal Reserve System, Washington, D.C., December.

Ericsson, N. R., and E. L. Reisman (2012) "Evaluating a Global Vector Autoregression for Forecasting", *International Advances in Economic Research*, 18, 3, 247–258.

Fildes, R., and H. O. Stekler (2002) "The State of Macroeconomic Forecasting", *Journal of Macroeconomics*, 24, 4, 435–468.

Frankel, J. (2011) "Over-optimism in Forecasts by Official Budget Agencies and Its Implications", *Oxford Review of Economic Policy*, 27, 4, 536–562.

Granger, C. W. J. (1989) *Forecasting in Business and Economics*, Academic Press, Boston, Massachusetts, Second Edition.

Hendry, D. F. (1999) "An Econometric Analysis of US Food Expenditure, 1931–1989", Chapter 17 in J. R. Magnus and M. S. Morgan (eds.) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, John Wiley and Sons, Chichester, 341–361.

Hendry, D. F., and J. A. Doornik (2014) *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*, MIT Press, Cambridge, Massachusetts.

Hendry, D. F., S. Johansen, and C. Santos (2008) "Automatic Selection of Indicators in a Fully Saturated Regression", *Computational Statistics*, 23, 2, 317–335, 337–339.

Hendry, D. F., and F. Pretis (2013) "Anthropogenic Influences on Atmospheric CO_2", Chapter 12 in R. Fouquet (ed.) *Handbook on Energy and Climate Change*, Edward Elgar, Cheltenham, 287–326.

Hendry, D. F., and C. Santos (2010) "An Automatic Test of Super Exogeneity", Chapter 12 in T. Bollerslev, J. R. Russell, and M. W. Watson (eds.) *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, Oxford University Press, Oxford, 164–193.

Holden, K., and D. A. Peel (1990) "On Testing for Unbiasedness and Efficiency of Forecasts", *The Manchester School*, 58, 2, 120–127.

Johansen, S., and B. Nielsen (2009) "An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator", Chapter 1 in J. L. Castle and N. Shephard (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, Oxford, 1–36.

Johansen, S., and B. Nielsen (2013) "Outlier Detection in Regression Using an Iterated One-step Approximation to the Huber-skip Estimator", *Econometrics*, 1, 1, 53–70.

Johansen, S., and B. Nielsen (2015) "Asymptotic Theory of Outlier Detection Algorithms for Linear Time Series Regression Models", *Scandinavian Journal of Statistics*, in press.

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

Mincer, J., and V. Zarnowitz (1969) "The Evaluation of Economic Forecasts", Chapter 1 in J. Mincer (ed.) *Economic Forecasts and Expectations: Analyes of Forecasting Behavior and Performance*, National Bureau of Economic Research, New York, 3–46.

National Bureau of Economic Research (2012) "US Business Cycle Expansions and Contractions", webpage, National Bureau of Economic Research, Cambridge, MA, April (www.nber.org/cycles.html).

Nordhaus, W. D. (1987) "Forecasting Efficiency: Concepts and Applications", *Review of Economics and Statistics*, 69, 4, 667–674.

Nunes, R. (2013) "Do Central Banks' Forecasts Take Into Account Public Opinion and Views?", International Finance Discussion Paper No. 1080, Board of Governors of the Federal Reserve System, Washington, D.C., May.

Pretis, F., M. L. Mann, and R. K. Kaufmann (2015) "Testing Competing Models of the Temperature Hiatus: Assessing the Effects of Conditioning Variables and Temporal Uncertainties Through Sample-wide Break Detection", *Climatic Change*, 131, 4, 705–718.

Romer, C. D., and D. H. Romer (2008) "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?", *American Economic Review*, 98, 2, 230–235.

Sinclair, T. M., F. Joutz, and H. O. Stekler (2010) "Can the Fed Predict the State of the Economy?", *Economics Letters*, 108, 1, 28–32.

Sinclair, T. M., H. O. Stekler, and W. Carnow (2012) "A New Approach for Evaluating Economic Forecasts", *Economics Bulletin*, 32, 3, 2332–2342.

Stekler, H. O. (1972) "An Analysis of Turning Point Forecasts", *American Economic Review*, 62, 4, 724–729.

Stekler, H. O. (2002) "The Rationality and Efficiency of Individuals' Forecasts", Chapter 10 in M. P. Clements and D. F. Hendry (eds.) *A Companion to Economic Forecasting*, Blackwell Publishers, Oxford, 222–240.

Tsuchiya, Y. (2013) "Are Government and IMF Forecasts Useful? An Application of a New Market-timing Test", *Economics Letters*, 118, 1, 118–120.

Tversky, A., and D. Kahneman (1974) "Judgment under Uncertainty: Heuristics and Biases", *Science*, 185, 4157, 1124–1131.

White, H. (1990) "A Consistent Model Selection Procedure Based on m-testing", Chapter 16 in C. W. J. Granger (ed.) *Modelling Economic Series: Readings in Econometric Methodology*, Oxford University Press, Oxford, 369–383.

Yellen, J. L. (2012) "Perspectives on Monetary Policy", remarks at the Boston Economic Club Dinner, Federal Reserve Bank of Boston, Boston, Massachusetts, June 6.