Detecting Time-dependent Bias in the Fed’s Greenbook Forecasts of Foreign GDP Growth

Neil R. Ericsson† Emilio J. Fiallos‡ J E. Seymour§
Revised: February 2017

Abstract
Building on Sinclair, Joutz, and Stekler (2010) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2013), this paper examines publicly available Fed Greenbook forecasts of several foreign countries’ GDP growth, focusing on potential biases in the forecasts. While standard tests typically fail to detect biases, recently developed indicator saturation techniques detect economically sizable and highly significant time-varying biases. Estimated biases differ not only over time, but by country and across the forecast horizon.

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

1. Introduction

The Fed’s monetary policy 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 of U.S. economic variables have been extensively analyzed, including by Romer and Romer (2008), Sinclair, Joutz, and Stekler (2010), Nunes (2013), and Ericsson, Hood, Joutz, Sinclair, and Stekler (2015). Surprisingly, Greenbook forecasts of foreign economic variables have not been examined, even though foreign economic activity is often a topic of discussion in the Fed’s deliberations on monetary policy; see Yellen (2015) inter alia. This paper thus examines the properties of Greenbook forecasts of foreign GDP growth—a key measure of foreign economic activity.

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) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2013), the current paper analyzes Greenbook forecasts of output growth in several foreign countries over 1998–2008. Standard tests typically fail to detect any important forecast biases. However, a recently developed technique—impulse indicator saturation—detects economically large and highly statistically significant time-varying biases. Biases differ across

---

1The 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 Shaghil Ahmed, David Hendry, Jun Ma, Ricardo Nunes, Andrea Raffo, John Rogers, and Herman Stekler 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)

‡Department of Statistics, Rutgers, The State University of New Jersey, New Brunswick, NJ (emilio_f1@yahoo.com)

§Division of International Finance, Board of Governors of the Federal Reserve System, Washington, DC 20551 USA (jedwardseymour@gmail.com)
the country being forecast, the horizon, and the date of the forecast. For example, forecasts of Chinese real GDP growth are systematically biased, with a statistically significant and economically large bias of approximately two percent per annum. For all countries examined, there is little observed predictability beyond two quarters ahead.

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

As input to the decision-making process of the Federal Open Market Committee, the staff of the Federal Reserve Board (the “Fed”) produce a document called the Greenbook, which includes forecasts of U.S. and foreign economic activity. This section describes the foreign Greenbook forecasts analyzed in this paper and the data being forecast. See Ericsson, Fiallos, and Seymour (2014) for further details.

The data being forecast are real GDP growth rates for nine countries:

- Brazil (BZ),
- Canada (CA),
- China (CH),
- Germany (GE),
- Japan (JA),
- South Korea (KO),
- Mexico (MX),
- the United Kingdom (UK), and
- (for comparison) the United States (US).

Country abbreviations are in parentheses. The sample period is determined by the presence of the forecasts in the publicly available Greenbooks: 1998Q1–2008Q4 for Canada, Germany, Japan, the United Kingdom, and the United States; and 1999Q4–2008Q4 for Brazil, China, Mexico, and South Korea. The Greenbook forecasts are from the final Greenbook of each quarter so as to allow as much information to be available for the forecasts being made in a given quarter. The forecast horizon $h$ (in quarters) is $h = -1, 0, 1, 2, 3, 4$, where $h = -1$ denotes the one-quarter backcast, $h = 0$ denotes the nowcast, and $h = 1, 2, 3, 4$ denote the one-, two-, three-, and four-quarter-ahead forecasts. Output growth is measured in quarterly rates expressed as percent changes at an annual rate. Measured actual values are the GDP growth rates as reported in the Greenbook with a two-quarter lag.

The Greenbooks are publicly available from the Federal Reserve Bank of Philadelphia:


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.
Several properties of the data, the Greenbook forecasts, and the corresponding forecast errors are apparent upon graphing. The Chinese forecasts systematically underpredict actual growth, albeit by different amounts, varying over time. Forecasts for other countries’ growth likewise under- or over-predict actual growth, and the degree of inaccuracy depends on the country, the horizon, and the date of the forecast. Forecast errors are often persistent, suggestive of systematic biases in the forecasts. See Ericsson, Fiallos, and Seymour (2014) for further details. For some previous analyses of Greenbook 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), Tsuchiya (2013), and Ericsson, Hood, Joutz, Sinclair, and Stekler (2015).

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. This section’s exposition draws on Ericsson (2015) and Ericsson, Hood, Joutz, Sinclair, and Stekler (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 \( \xi_t \) at time \( t \) and its forecast \( \hat{\xi}_t \), estimate the equation:

\[
(x_t - \hat{\xi}_t) = a + e_t \quad t = 1, \ldots, T, \tag{1}
\]

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 \( \hat{\xi}_t \) is unbiased for the variable \( \xi_t \). For current-period and one-step-ahead forecasts, the error \( e_t \) may 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):

\[
(x_t - \hat{\xi}_t) = b_0 + b_1 z_t + e_t \tag{2}
\]

where 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) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2015), the regressor \( z_t \) includes a dummy variable that is indicates the business cycle’s phase—either contraction or expansion. Another choice of \( z_t \) is \( \hat{\xi}_t \), proposed by Mincer and Zarnowitz (1969, p. 11). In Ericsson (2015) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2015), “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 \( z_t = \hat{\xi}_t \). Other choices for \( z_t \) include an alternative forecast \( \tilde{\xi}_t \) or the differential between the two forecasts \( (\hat{\xi}_t - \tilde{\xi}_t) \), 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 \( (\hat{x}_t - \hat{x}_t) \) 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:

\[
(x_t - \hat{x}_t) = a_t + e_t = \sum_{i=1}^{T} c_i I_{it} + e_t \quad t = 1, \ldots, T,
\]

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 non-technical exposition of IIS, and Hendry and Doornik (2014) for extensive analysis in the context of automatic model selection. This section’s exposition draws on Ericsson (2015) and Ericsson, Hood, Joutz, Sinclair, and Stekler (2015).


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

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Variables</th>
<th>Definition of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulse indicator saturation</td>
<td>Zero-one dummies</td>
<td>{I_{it}}</td>
<td>( I_{it} = 1 ) for ( t = i ), zero otherwise</td>
</tr>
<tr>
<td>Super saturation</td>
<td>Step functions</td>
<td>{I_{it}, S_{it}}</td>
<td>( S_{it} = 1 ) for ( t \geq i ), zero otherwise</td>
</tr>
</tbody>
</table>

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 an extension: super 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).

Table 1 is by no means an exhaustive list of extensions to IIS. One direct extension is ultra saturation, with searches across \( \{I_{it}, S_{it}, T_{it}\} \), where the \( \{T_{it}\} \) are broken linear trends. Broken quadratic trends, broken cubic trends, and higher-order broken trends are also feasible. Other extensions include sequential \( (j = 1) \) and non-sequential \( (j > 1) \) pairwise impulse indicator saturation for an indicator \( P_{at} \), defined as \( I_{at} + I_{at+j} \); sequential multiplet indicator saturation for an indicator \( M_{at}^{j+1} \), defined as \( I_{at} + \cdots + I_{at+j} \); zero-sum pairwise IIS for an indicator \( Z_{at} \), defined as \( \Delta I_{at} \); 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 $S_{kt}x_{kt}$. 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, dummies for economic expansions and contractions 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 Greenbook forecasts of output growth for eight foreign countries and for the United States. 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. Forecast biases differ numerically across the forecast horizon, country being forecast, and the date of the forecast, albeit with some qualitative similarities.

Section 5.1 reports a standard summary statistic on forecast performance: root mean squared forecast errors. Section 5.2 reports standard Mincer–Zarnowitz tests of forecast bias. Section 5.3 employs IIS-type procedures to test for and estimate time-varying forecast bias.

5.1 Summary Statistics of Forecast Performance

Figure 1 plots the root mean squared forecast errors (RMSEs) for the nine countries as a function of the forecast horizon $h$. The RMSEs for four developed countries (Germany, Canada, the United Kingdom, and the United States) are considerably smaller at every horizon than the RMSEs for the remaining countries. Notably, the remaining countries include Japan—a developed country—and all of the emerging market economies analyzed (Brazil, Mexico, South Korea, and China). For all
countries, the RMSEs generally increase with the forecast horizon; and the RMSEs increase little beyond a horizon of two quarters ahead.

5.2 Standard Tests of Forecast Bias

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

Table 2 reports estimated intercepts and estimated standard errors for the Mincer–Zarnowitz regression in equation (1). Here and in Table 3, HAC estimated standard errors appear under regression coefficients in square brackets [·]. The symbols +, *, and ** respectively denote significance at the 10%, 5%, and 1% levels. Statistically, there is little evidence of forecast bias, except for Chinese output growth, for which the estimated bias is 1.5%–2.0% per annum for the nowcast and the forecasts.

5.3 Estimated Time-varying Bias Using Indicator Saturation

To assess possible time dependence of the forecast biases, this subsection estimates IIS-type equations in the form of equation (3). Time dependence is detected for all countries at some or all forecast horizons.

Table 3 reports estimated intercepts and estimated standard errors for the Mincer–Zarnowitz regression in equation (3); and Table 4 lists the impulse and step dummies retained from super saturation with a target size of 0.5%. Tables 3 and 4 highlight the dependence of bias on the country and on the forecast hori-
Table 2: Estimated intercepts and HAC standard errors for the Mincer–Zarnowitz test of bias in Greenbook forecasts for GDP growth.

<table>
<thead>
<tr>
<th>Country</th>
<th>Forecast horizon $h$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−1</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>[0.27]</td>
</tr>
<tr>
<td>Canada</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>[0.10]</td>
</tr>
<tr>
<td>China</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>[0.14]</td>
</tr>
<tr>
<td>Germany</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>[0.09]</td>
</tr>
<tr>
<td>Japan</td>
<td>−0.42+</td>
</tr>
<tr>
<td></td>
<td>[0.24]</td>
</tr>
<tr>
<td>Korea</td>
<td>0.66*</td>
</tr>
<tr>
<td></td>
<td>[0.25]</td>
</tr>
</tbody>
</table>
| Mexico        | −0.03| −0.30| −0.64| −1.10| −1.21| −1.36+
|               | [0.16]| [0.48]| [0.56]| [0.67]| [0.77]| [0.75]|
| United Kingdom| 0.10+| 0.01| −0.16| −0.19| −0.23| −0.28|
|               | [0.05]| [0.11]| [0.15]| [0.23]| [0.26]| [0.25]|
| United States | −0.06*| 0.56**| 0.39| 0.14| −0.01| −0.06|
|               | [0.03]| [0.18]| [0.30]| [0.45]| [0.50]| [0.51]|

Table 3: IIS-estimated intercepts and HAC standard errors for the Mincer–Zarnowitz test of bias in the Greenbook forecasts for GDP growth.

<table>
<thead>
<tr>
<th>Country</th>
<th>Forecast horizon $h$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−1</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>[0.11]</td>
</tr>
<tr>
<td>Canada</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>[0.05]</td>
</tr>
<tr>
<td>China</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[0]</td>
</tr>
<tr>
<td>Germany</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>[0.05]</td>
</tr>
<tr>
<td>Japan</td>
<td>−0.11</td>
</tr>
<tr>
<td></td>
<td>[0.16]</td>
</tr>
<tr>
<td>Korea</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>[0.02]</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>[0.13]</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>[0.05]</td>
</tr>
<tr>
<td>United States</td>
<td>−0.00</td>
</tr>
<tr>
<td></td>
<td>[0.03]</td>
</tr>
</tbody>
</table>

*Represents significance at the 10% level.
**Represents significance at the 5% level.
***Represents significance at the 1% level.
Table 4: Dummy variables selected by super saturation at the 0.1% level at at least one forecast horizon, by country.

<table>
<thead>
<tr>
<th>Country</th>
<th>Indicators selected</th>
<th>Impulse</th>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>1999(4)</td>
<td>2000(3), 2001(3)</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>—</td>
<td>2006(2), 2007(4), 2008(1)</td>
<td></td>
</tr>
</tbody>
</table>
zon $h$. The selected dummies in Table 4 indicate the pervasiveness of time-varying bias. Because forecast errors tend to be very small for backcasts ($h = -1$), only non-negative forecast horizons are considered in Table 4.

Graphs directly convey a sense of the magnitude and extent of the biases present. Figures 2–5 thus plot actual values and forecasts and the forecast errors for two of the countries analyzed: China and the United States. Each figure is a panel of $2 \times 3$ for forecast horizons $h (-1, 0, 1 \ldots, 4)$.

Figure 2 plots the actual and forecast values for Chinese growth; and Figure 3 plots the corresponding forecast errors and the bias as estimated from super saturation. Large biases are evident for all horizons except $h = -1$, and the biases appear somewhat different before and after 2003.

Figure 4 plots the actual and forecast values for U.S. growth; and Figure 5 plots the corresponding forecast errors and the bias as estimated from super saturation. The biases are notably time-dependent and persistent at all non-negative horizons. Ericsson, Hood, Joutz, Sinclair, and Stekler (2015) show that those biases depend primarily on the phase of the business cycle.

Forecast biases vary markedly over time, being sometimes positive and other-times 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 the sample, even though this bias would be obvious from (e.g.) graphing the data. Super saturation often detects time-varying bias, and for historically and economically consequential years. The dates of the retained dummies are important and informative, and those dummies often appear to reflect cyclical movements.

6. Conclusions

Building on Sinclair, Joutz, and Stekler (2010), the current paper analyzes Greenbook forecasts of foreign output growth for potential biases over 1999–2008. Standard tests typically fail to detect bias. However, super saturation detects economically large and highly statistically significant time-dependent biases across all countries being forecast. Biases depend on the country, the forecast horizon, and the date of the forecast. Saturation as a technique 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. In particular, such biases imply an opportunity to robustify the forecasts, as with intercept correction; see Clements and Hendry (1999, 2002), Hendry (2006), and Castle, Fawcett, and Hendry (2010).
Figure 2: Chinese GDP growth and its Greenbook forecasts at different forecast horizons \( h \ (h = -1, 0, 1, \ldots, 4) \).

Figure 3: Greenbook forecast errors for Chinese GDP growth at different forecast horizons \( h \ (h = -1, 0, 1, \ldots, 4) \), and estimated forecast biases as calculated using super saturation.
Figure 4: US GDP growth and its Greenbook forecasts at different forecast horizons $h$ ($h = -1, 0, 1 \ldots, 4$).

Figure 5: Greenbook forecast errors for US GDP growth at different forecast horizons $h$ ($h = -1, 0, 1 \ldots, 4$), and estimated forecast biases as calculated using super saturation.
REFERENCES


