



Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Comment on “How Biased are US Government Forecasts of the Federal Debt?”

Edward N. Gamber^{a,*}, Jeffrey P. Liebner^{b,1}^a Department of Economics, Lafayette College, Easton, PA 18042, United States^b Department of Mathematics, Lafayette College, Easton, PA 18042, United States

ARTICLE INFO

Keywords:

Impulse indicator saturation
Forecast bias
Model misspecification

ABSTRACT

In this comment on “How Biased are US Government Forecasts of the Federal Debt?” by Neil R. Ericsson, we investigate the sensitivity of the “bare-bones” application of the impulse indicator saturation technique. We offer an alternative but complementary interpretation of Ericsson’s findings of bias in government debt forecasts. Our findings reinforce his interpretation of the role of the IIS technique as a general diagnostic tool for detecting model misspecification.

© 2014 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Both the Congressional Budget Office (CBO) and the Office of Management and Budget (OMB) are required by law to make medium-term projections of the federal budget.² These projections serve as benchmarks for proposed changes in taxes and expenditures. To the extent that policymakers base current tax and expenditure choices on budget forecasts, errors in making such forecasts translate directly into errors in policy. Thus, it is important to understand the properties of the forecasts, and most importantly whether they are biased.

Neil Ericsson’s paper (Ericsson, 2015) builds on an earlier, unpublished study by Martinez (2011), which also looked at the one-year-ahead debt forecasts produced by CBO and OMB. Martinez’s analysis included (traditional) tests of forecast bias. Ericsson looks at a slightly longer

sample of the same forecasts. The main innovation in Ericsson’s paper is the use of the impulse indicator saturation (IIS) technique to detect *time varying* biases in these forecasts.

While Ericsson focuses on detecting bias using the IIS technique, we explore an alternative interpretation of the technique and its uses. In particular, we interpret the IIS technique as a general diagnostic tool for detecting model specification errors. In the context of Ericsson’s paper, the simplest model is the Mincer–Zarnowitz test for bias (Ericsson’s equation 1), which assumes a constant mean and variance. In our application of the IIS technique to the debt forecast data, we find evidence against the assumption of a constant variance in that simple model.

2. The forecasting process as context

A first step in understanding forecast errors is to understand the context in which the forecasts are produced. The debt forecasts analyzed by Ericsson are the one (fiscal) year ahead forecasts produced by CBO and OMB, typically in late January or early February of each year. Ericsson also analyzes CBO’s debt forecast based on the President’s budget. That forecast, which Ericsson refers to as the APB forecast, is typically produced in March of each year.

* Corresponding author. Tel.: +1 610 330 5310.

E-mail addresses: gambere@lafayette.edu (E.N. Gamber), liebnerj@lafayette.edu (J.P. Liebner).

¹ Tel.: +1 610 330 5910.

² This requirement is stipulated by the Congressional Budget Act of 1974. See Rasche (1985).

All three forecasts of the budget begin with economic forecasts, and it is important to note that, despite the different release dates, all three forecasts are based on the GDP data available through the third quarter of the previous year.³ Thus, any differences in forecast performances across these three forecasts are apparently not due to differences in the availability of macroeconomic data.

As was noted by Penner (2008), the economic forecasts that underlie the budget forecasts are based on traditional large-scale macroeconomic models. Despite the well-known shortcomings of these models (the Lucas critique), they are well suited for the incorporation of an analysis that Penner says “is almost always leavened by a large dose of judgment”.

Judgment is a valuable tool for incorporating information into a forecast that does not come neatly packaged in the form of a consistent time series of published data, but it certainly has its limits as well. The dates identified by Ericsson's application of the IIS technique to the debt forecast data illustrate those limits. The impulse and step dummies that are identified as being significant occur at dates that correspond to large events that might not have been anticipated at the time when the economic and budget forecasts were made. This points to another advantage of the IIS technique as a tool for identifying specific points in time when the model does a particularly poor job of capturing movements in the data.

3. The IIS methodology

We investigate the sensitivity of the IIS technique by applying it to the one-year-ahead forecast errors for the CBO debt forecasts using data from 1984 to 2012. We imitate the bare-bones approach that Ericsson provided as an illustrative example by dividing this 29 year period into two parts, the first of 14 years and the second of 15 years. Note that Autometrics applies the IIS technique with many possibly unequally sized blocks. Because this simpler bare-bones approach uses only one partition, the number of points detected as being significant should be less than or equal to the number detected by the Autometrics software, but will provide a baseline for comparison. Using this bare bones set-up, we use the IIS model to detect years where the observed value was significantly different from the other years using a significance level of 0.01. We find only 2001, 2008, and 2009 to be significant at the 0.01 level when dummy variables are applied to the separate parts of the dataset, and only 2008 remains significant at this level when these three years are used as dummy variables with the entire dataset. In the Autometrics software output that was provided to us, we note that the significance level used appears to increase when performing an extended block search for omitted variables. We also note that the dummy variables that Autometrics originally selected (2001, 2002, 2003, 2008, and 2009) are not all significant at the 0.01

level until this adjustment in significance levels is made. By adjusting our significance level to 0.05, we are able to include 2002, 2003, and 2010 in our potential dummy variables as well. When these are included, all six dummy variables are found to be significant at the 0.02 level when the dummy variables are applied with the entire dataset, with all but 2003 and 2010 being significant at the 0.005 level. Using this bare-bones approach, we are unable to detect 1990 as a significant dummy variable, but we attribute this to our choice of blocks. Thus, we are led to conclude that the choices of blocks and significance levels used are essential for determining the final model when using the IIS technique.

Reflecting on the results themselves, one is led to wonder why particular years are detected as having errors that are significantly different from the mean. Ericsson suggests that these deviations are the result of bias in the forecasts. However, we also consider an alternative scenario. Looking at the errors from the CBO data, the summary statistics reveal that the standard deviation of the first 14 data points for the logarithm of the CBO errors is 0.01007, while the standard deviation of the last 15 data points is 0.02122. We next explore whether the results from the IIS test can be explained by a change in the standard deviation of the data, rather than by the presence of bias.^{4,5} We perform the following simulation. We create one data set consisting of fourteen data points drawn from a normal distribution with mean 0 and standard deviation 0.01007, and another data set with fifteen data points drawn from a normal distribution with mean 0 and standard deviation 0.02122. These represent the two parts of the dataset. By construction, there is no bias in either section of the data, with the only difference between the two parts being the standard deviation. We create 10,000 simulated datasets in this manner, then apply a simple version of the IIS method to the data by mimicking the illustrative example provided by Ericsson. This is the technique that Ericsson describes as the “bare-bones” implementation of IIS, where the data are only blocked into two parts, as opposed to the many, possibly unequally sized, blocks that the Autometrics econometrics software employs. As was noted before, the fact that this simpler technique only uses one partition means that the number of points detected as being significant should be less than or equal to the number detected by the Autometrics software; however, it provides a baseline for comparison. The results are reported in Table 1.

At a significance level of 0.05, we would expect 5% of the twenty-nine years, or 1.45 years on average, to be detected as significant on average if the change in standard deviation had no effect. However, the results show that, using the different standard deviations for the two parts, 4.37 years were detected as significant on average, even when no bias was present. Furthermore, we can use these

³ Although the OMB forecast is usually released shortly after the CBO forecast, it is typically finalized prior to the release of the preliminary GDP data for the fourth quarter. CBO typically does not revise its forecast between the January *Budget and Economic Outlook* and the March *Analysis of the President's Budget*.

⁴ The increased standard deviation of the forecast error is somewhat of a puzzle, given the fact that the second half of the sample includes the Great Moderation, which is characterized by a steep decline in the standard deviation of real GDP growth.

⁵ Hendry and Santos (2010, chap. 12) note the possibility of using the IIS technique to detect general misspecification, including the presence of heteroskedasticity.

Table 1
Bare-bones IIS results.

Number of years	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Percentage	1.94%	6.42%	13.33%	16.47%	17.38%	15.11%	11.88%	8.12%	5.02%	2.71%	0.99%	0.46%	0.13%	0.04%

Notes: The table reports the percentage of iterations in which a particular number of data points were detected as significant using a bare-bones implementation of IIS with one partition on data points y_t , using a significance level of 0.05 simulated from $N(0, \sigma_t)$ with $\sigma_t = 0.01007$ for $t = 1, \dots, 14$ and $\sigma_t = 0.02122$ for $t = 15, \dots, 29$. The simulations are based on 10 000 iterations of a parametric bootstrap.

bootstrap results to obtain a p -value for the original results that were obtained using the IIS model. In this scenario, our null hypothesis is that there is heteroskedasticity but no bias; specifically, that the standard deviation of part 1 of the data is 0.01007 and that of part 2 is 0.02122. The alternative hypothesis is that the model is misspecified, including the possibility that there is bias present. Under the conditions of the null hypothesis, we would detect the six or more significant dummy variables that we found in our original analysis 29.35% of the time. This suggests that the results that we detected using the bare-bones approach are not unusual in the presence of heteroskedasticity, even in the absence of bias. This does not preclude the possibility that the significant forecast errors are the result of forecast bias, but presents an alternative explanation for the observations.⁶

Ericsson acknowledges the issues with interpreting IIS-based tests as tests of the forecast bias because IIS-based tests can detect other forms of misspecification, including outliers that are a result of heteroskedasticity. However, he also points to the significance of the intercept as an indication of bias, noting that the presence of heteroskedasticity fails to explain this observation.⁷ In this, we agree to a certain extent. While the Mincer–Zarnowitz test fails to detect a significant difference in the intercept relative to a null belief that the value is zero, the IIS technique does find a significant difference at the 0.01 level. However, when we examine our simulation to consider the effects of heteroskedasticity, we find that the intercept is significantly different from zero a total of 4.16% of the time when we use a significance level of 0.01. When we rerun the simulation while only retaining dummy variables at a significance level of 0.01, the intercept is still significantly different from zero a total of 2.67% of the time when using a significance level of 0.01. While this does not preclude the intercept being significantly different from zero at a small level of significance, it indicates that the possibility of a type I error may be understated. The principal cause of this discrepancy lies in the estimation of the variance of the error terms. Hendry, Johansen, and Santos (2008) found that the estimation of this variance is downward biased when impulses are introduced. As a consequence, the standard errors of the other parameters that are being estimated are also downward biased. If one then uses the standard t -statistic to test for significance, the p -values are underestimated.

⁶ We test for heteroskedasticity in the CBO debt forecast errors using the Breusch–Pagan test, and reject the null of homoscedasticity at the 0.01 level.

⁷ Another important aspect of Ericsson's argument is that the dummies retained are associated with important economic events (see Ericsson's Section 6).

Note that our simulation study only uses the bare-bones approach to implementing the IIS technique, as opposed to the Autometrics implementation, which uses multiple blocks of unequal sizes. This application of multiple blocks can result in the retention of even more dummy variables, which can have the effect of decreasing the estimation of the variance of the error terms even further.

4. Conclusion

Ex-post analyses of forecast errors are important for both model builders and policymakers. Neil Ericsson's paper explores the biases in the debt forecasts produced by the OMB and the CBO. We (and Ericsson and others) interpret the IIS technique to include a more general method for detecting model misspecification. In our analysis of the debt data, we find a violation of the assumption of homoskedastic errors. In other applications, the IIS technique has the potential to detect a wide variety of model misspecifications.

It is clear that the IIS technique is useful as an ex-post diagnostic tool for detecting points in time when the model is biased. However, the challenge in future is to complete the feedback loop; that is, to use the diagnostic information to improve model building. The traditional definition of bias is that forecasts are systemically too high or too low. If the average forecast error over the entire sample in the case of the debt forecasts had been positive and significant, then in future one could simply adjust the forecasts downward by the amount of the positive bias. However, the type of forecast bias identified here calls for a more nuanced fix. The bias in the debt forecasts appears mostly at business cycle turning points. A more explicit consideration of regime-shifting models, such as those proposed by Hamilton (1989) and more recently Chauvet and Piger (2003), might be useful in addressing the bias at business cycle turning points.

References

- Chauvet, M., & Piger, J. M. (2003). Identifying business cycle turning points in real time. *Econ. Rev. Federal Reserve Bank of St. Louis*, 85(2), 47–61.
- Ericsson, N. R. (2015). How biased are US government forecasts of the federal debt? *International Journal of Forecasting*, in press.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.
- Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23(2), 317–335.
- Hendry, D. F., & Santos, C. (2010). An automatic test of super exogeneity. In T. Bollerslev, J. R. Russell, & M. W. Watson (Eds.), *Volatility and time series econometrics: essays in honor of Robert F. Engle* (pp. 164–193). Oxford: Oxford University Press.
- Martinez, A. B. (2011). *Comparing government forecasts of the United States' gross federal debt*. The George Washington University Research Program on Forecasting, Working paper no. 2011-002.
- Penner, R. G. (2008). Federal revenue forecasting. In J. Sun & T. D. Lynch (Eds.), *Government budget forecasting: theory and practice*. CRC Press, Taylor and Francis Group.
- Rasche, R. H. (1985). *Deficit projections vs. deficit forecasts*. Federal Reserve Bank of San Francisco, Weekly Letter.