International Journal of Forecasting 31 (2015) 312–324

Contents lists available at ScienceDirect

International Journal of Forecasting

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

How good are US government forecasts of the federal debt?

Andrew B. Martinez

Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Evaluating Forecasts
Government Forecasting
Macroeconomic Forecasts
Forecast Encompassing
Deficit

ABSTRACT

This paper compares annual one-year-ahead and five-year-ahead forecasts from government agencies for the US gross federal debt and deficit from 1984 to 2013. Other studies have compared two of these agencies' forecasts, but not for debt. The current paper finds that the forecast from the Analysis of the President's Budget performs best across both horizons but does not encompass the other forecasts. Instead, each of the forecasts lacks information included by the other agencies and therefore a combination of all three outperforms all individual forecasts.

1. Introduction

In the aftermath of the recent financial and economic crisis, the rapidly increasing government debt around the world has heightened worries about economic growth. In 2013, the US total federal government debt outstanding was 99% of GDP, a share that had not been reached since World War II. US government agencies have forecast that government debt will continue to rise. These forecasts have prompted concerns that the US debt burden will become unsustainable; and the actual debt and its forecasts have led to a national debate over the debt, the debt ceiling, and deficits.[1]

The intense focus on the US federal debt makes it important that we understand how well both the debt and the deficit are forecast. The available forecasts vary considerably, highlighting the need to know which forecast tracks the trajectory of the debt most closely. The short-, medium- and long-term forecasts are all of interest, given concerns about the debt ceiling and about debt sustainability.

This paper aims to answer these questions. Using annual data since 1984, this paper compares the three different forecasts of the US federal debt and deficit at both the one- and five-year-ahead horizons. The forecasts are denoted by their sources:

- CBO (Congressional Budget Office) from its *Budget and Economic Outlook*,
- OMB (Office of Management and Budget) from its *Budget of the US Government*, and
- APB (Analysis of the President's Budget).

The Congressional Budget Office and the Office of Management and Budget are different agencies within the US federal government. The *Analysis of the President's Budget* is produced by the Congressional Budget Office, but the policy assumptions embedded in the forecasts from the *Analysis of the President's Budget* differ from those in the forecasts from the *Budget and Economic Outlook*. Thus, these two forecasts are referred to as the "APB forecast" and the "CBO forecast", noting that both are produced by the Congressional Budget Office. Also, for expositional convenience, the three forecasts are referred to as "agency forecasts", even though only two agencies are involved.

The current analysis adds to the previous literature in several ways. First and foremost, it extends comparisons of government agencies' forecasts to include the federal debt. Second, it compares the CBO, OMB, and APB forecasts with each other individually, and with averages of the agencies' forecasts. Finally, the analysis uses both root mean square forecast errors (RMSE) and forecast-encompassing tests to

E-mail address: andrew.martinez@economics.ox.ac.uk.

[1] For some examples, see IMF Survey (2013), and Standard and Poors (2011).

http://dx.doi.org/10.1016/j.ijforecast.2014.08.014
0169-2070/

compare the forecasts. These tools help to assess whether certain forecasts can outperform other forecasts. The analysis finds that all of the short-term agency forecasts have relatively small forecast errors, except during recessions. It also finds value in examining debt forecasts separately from deficit forecasts. Furthermore, it shows that either the APB's debt forecast or a combination of the agencies' forecasts performs better than individual CBO and OMB debt forecasts.

In public discourse, the discussion of the deficit often overshadows any discussion of the federal debt, since the deficit is commonly thought of as equaling the change in debt. However, the deficit excludes certain items that contribute to the change in debt. Focusing on the deficit alone could miss significant components of the debt. Therefore, it is important to examine forecasts of the federal debt, which aggregate across multiple sources of debt, including the deficit.

Comparing forecasts from different government agencies is complicated because they condition on different economic and policy assumptions. The CBO assumes that current law will remain unchanged, whereas the OMB and APB assume that the policy changes proposed in the president's budget will be implemented. However, that being said, analyzing these different forecasts relative to actual outcomes can give a sense of the degree of usefulness of the agencies' forecasts. Also, as a linguistic matter, CBO refers to its forecasts of deficits and debts as "projections", while OMB refers to them as "estimates", and the APB publications typically call them "re-estimates". The current paper uses the term "forecasts" throughout, drawing on the more encompassing usage given by Clements and Hendry (2002, chap. 1, p. 2): "A forecast is any statement about the future".[2]

Poor performance – whether measured as forecast non-encompassing or as a large RMSE – has both economic and statistical significance. However, that being said, the specific meaning of "poor forecast performance" depends in part on whether an agency's forecasts are interpreted as "forecasts" or as "projections", where "projections" are taken in the sense of being policy simulations conditional upon a certain set of assumptions. If the agency's forecasts are interpreted *qua* forecasts, then a poor forecast performance implies potential room for improvement in terms of these performance measures. If the forecasts are interpreted *qua* projections, then a poor forecast performance implies a limited usefulness of the forecasts as representing interesting hypothetical paths for economic policy.

In this analysis, a poor forecast performance is interpreted in the latter sense. Therefore, in that context, the overarching questions that this paper seeks to address are: which forecast represents the most useful path for economic policy, and does any one of the agency forecasts incorporate all of the relevant information such that the other agency forecasts are not useful?

This paper is structured as follows. Section 2 reviews the literature on comparisons of US government agency forecasts. Section 3 provides a background to the forecast-encompassing tests used. Section 4 describes the data and some initial comparisons of the forecasts. Section 5 presents the empirical findings and analysis. Section 6 concludes.

2. Literature review

There is a considerable body of literature on the comparison of US government agency forecasts. These studies can be divided into two types. The first and more popular type typically compares the agencies' forecasts using summary statistics such as RMSE, mean absolute errors (MAE), and mean absolute percent errors (MAPE). The second type uses forecast-encompassing tests to compare the forecasts. Both types of studies provide valuable information about the forecasts.

Studies of the first type have obtained a variety of results. Kamlet, Mowery, and Su (1987) compare one-year-ahead and multi-year-ahead forecasts from CBO, OMB, their own ARIMA model, and the ASA/NBER survey for the real growth rate, inflation rate, and unemployment from 1976 to 1984.[3] They find that neither agency outperforms the other for short-term forecasts. However, for forecasts extending beyond three years, the authors find that OMB forecasts are more optimistic than CBO forecasts, but not less accurate. Plesko (1988) examines the CBO and OMB forecasts of nominal GNP, current receipts, current outlays, and the deficit from 1974 to 1988, and finds similar results for the short-term forecasts.

McNees (1995) compares forecasts from the Federal Reserve Board (FRB), the CBO, the Council of Economic Advisors (CEA), and several private forecasters for inflation, GNP, and unemployment from 1976 to 1994.[4] For long-term forecasts, McNees finds that the CEA is more optimistic than any of the other forecasters. Frendreis and Tatalovich (2000) compare CBO, OMB, and FRB one-year-ahead forecasts of GNP growth, inflation, and unemployment from 1979 to 1997. While all three agencies' forecasts tend to be close, the authors find that the CBO forecasts are the best, followed by those of the FRB, then those of OMB.

The CBO itself conducts a semi-annual comparison of its own forecasts with OMB and Blue Chip forecasts. The most recent update is that of US CBO (2013), which compares two-year-ahead and five-year-ahead forecasts for output, inflation, three-month Treasury rates, long-term interest rates, and wage and salary disbursements from 1980 to 2010. Similarly to previous studies, it finds that the CBO's forecasts are as accurate as the OMB and Blue Chip forecasts.

Studies of the second type use forecast-encompassing tests to compare forecasts, and also obtain mixed results. Howard (1987) compares CBO and OMB forecasts of the

[2] See also Rasche (1985), who discusses the differences between projections and forecasts.

[3] The ASA/NBER survey is a survey of private forecasters that is conducted by the American Statistical Association and the National Bureau of Economic Research.

[4] The CEA and OMB forecasts of the federal debt are identical. Thus, studies use one or the other for comparisons with the CBO. The private forecasters considered are the American Statistical Association (ASA), Data Resources Inc (DRI), and Blue Chip Economic Indicators.

Table 1
Previous studies.

Study	Forecasters	Variables	Horizon	Sample	Summary of findings
Kamlet et al. (1987)	CBO, OMB, ASA/NBER, ARIMA	Real GNP growth rate, inflation, unemployment	Short, long	1976–1984	OMB more optimistic than CBO (long)
Howard (1987)	CBO, OMB	Real GNP growth rate, GNP deflator, CPI, unemployment, Treasury rates	Short	1976–1985	OMB forecasts are optimistic
Plesko (1988)	CBO, OMB	Nominal GDP, revenues, outlays	Short, long	1974–1988	OMB and CBO perform equally well (short and long)
Belongia (1988)	CBO, CEA, ASA/NBER	Real GNP growth rate, GNP deflator, unemployment	Short	1976–1987	ASA/NBER best, CBO and CEA equally poor
McNees (1995)	CBO, CEA, FRB, ASA, Blue Chip, DRI	Inflation, GNP, unemployment	Long	1976–1994	CEA more optimistic than CBO, FRB, ASA, Blue Chip and DRI
Frendreis and Tatalovich (2000)	CBO, OMB, FRB	GNP growth, inflation, unemployment	Short	1979–1997	CBO best, followed by FRB, then OMB
Cohen and Follette (2003)	CBO, OMB, FRB	Budget	Short	1977–2003	CBO encompasses OMB
Douglas and Krause (2005)	CBO, OMB, FRB	Real and nominal GDP, inflation, unemployment, revenues, outlays, budget	Short	1976–2001	FRB best for unemployment, CBO worst in tax revenues, all else indistinguishable
Corder (2005)	CBO, OMB, SSA	GDP, inflation, unemployment, interest rates	Short, long	1976–2003	CBO better for GDP, OMB better for unemployment, neither for interest rates (long)
US CBO (2013)	CBO, OMB, Blue Chip	Output, inflation, Treasury rates, long-term interest rates, wage and salary disbursements	Short, long	1982–2010	CBO and OMB perform equally well (short and long)

real GNP growth rate, the GNP deflator, the consumer price index, the unemployment rate, and the three-month Treasury bill rate from 1976 to 1985. By regressing the OMB's forecast errors on a constant and the CBO's forecast errors, Howard finds that the forecast errors are strongly correlated with each other and that the OMB forecasts are biased optimistically.

Belongia (1988) compares CEA, CBO, and ASA/NBER one-year-ahead forecasts of real GNP growth, the GNP deflator, and unemployment from 1976 to 1987. By regressing the actual growth rate of each variable on a constant and different pairs of predicted growth rates, Belongia finds that, in general, the ASA/NBER forecasts perform better than those of either the CBO or CEA, while neither CBO nor CEA outperforms the other. These results suggest that the CBO and CEA forecasts may be encompassed by the ASA/NBER forecasts, but do not encompass one another.

Cohen and Follette (2003) compare one-year-ahead CBO, OMB, and FRB forecasts of the budget deficit from 1977 to 2003. They regress the actual outcomes on OMB and CBO forecasts over different periods, and find that, for most samples, CBO forecasts encompass OMB forecasts. Douglas and Krause (2005) also compare one-year-ahead CBO, OMB, and FRB forecasts of real and nominal GDP, inflation, unemployment, tax revenues, government outlays, and the budget deficit from 1976 to 2001. They use a variety of encompassing tests and find that, with the exception of unemployment and tax revenues, the forecasts are not statistically distinguishable from one another. They find that the FRB unemployment forecasts perform better than those from either the CBO or the OMB, while the CBO tax-revenue forecasts perform worse than those of either the OMB or the FRB.

Corder (2005) examines forecasts of GDP, inflation, unemployment, and interest rates from the Social Security Administration (SSA), the CBO, and the OMB between 1976 and 2003. Using two different tests to check for bias and efficiency, he finds that the CBO forecasts encompass the OMB forecasts for GDP, the OMB forecasts encompass the CBO forecasts for unemployment and inflation, and neither encompasses the other for interest rates. He concludes that the agencies' forecasts could be improved if they incorporated information from each other.

Overall, both types of studies obtain mixed results. Some studies find that CBO forecasts are significantly better than OMB forecasts, while other studies find that OMB forecasts are on a par with CBO forecasts. In a few instances, studies even find that the OMB forecasts perform better than the CBO forecasts. Table 1 summarizes the findings from previous studies.

In addition to the studies comparing government agencies' forecasts, there is also a considerable body of literature that focuses on the performances of individual agency forecasts. While the focus of these studies is not the same as that of the studies that compare forecasts, they do tend to use similar methods. For instance, Kliesen and Thornton (2012) rely primarily on summary statistics to examine how well CBO forecasts of the deficit have performed relative to a random walk model; and Frankel (2011) examines potential biases in OMB forecasts of the deficit using an approach that is akin to a simple forecast-encompassing test.

Standard forecast summary statistics and forecast-encompassing tests both have limitations. Ericsson (1992) shows that, while a small MSE is a necessary condition for being the better forecast, it is not sufficient to determine whether one forecast can explain another forecast's error (i.e., encompass it). Equally, US CBO (2010) cautions against using statistical tests with such small sample sizes because "particular errors can have an unduly large influence on the measures" (p. 2). Thus, rather than relying on one single test, the current analysis uses both RMSEs and forecast-encompassing tests to compare government forecasts of debt and deficits. This ensures that the risk of choosing a less powerful test is spread over a range of tests, while allowing for a comparison of the results across tests.

3. Forecast-encompassing tests

The analysis in this paper relies heavily on the concept of forecast encompassing, as developed by Chong and Hendry (1986). They propose a procedure for comparing two forecasts and determining whether one of them is "sufficient" in a very specific statistical sense. Their basic framework, re-interpreted in the current context for comparing the three agency forecasts, is:

$$y_t = b_1{}^* x_t + b_2{}^* w_t + b_3{}^* z_t + e_t, \quad (1)$$

where y_t denotes the actual value of the variable being forecast, x_t denotes the forecast from the first agency, w_t denotes the forecast from the second agency, z_t denotes the forecast from the third agency, and e_t is a residual.[5]

Eq. (1) defines a sharp testable hypothesis: $\{b_1 = 1, b_2 = 0, b_3 = 0\}$. This null hypothesis holds if the first agency's forecast provides an explanation of the dependent variable, with the second and third agencies' forecasts offering no additional information to that explanation. The agencies' roles can be reversed by testing $\{b_1 = 0, b_2 = 1, b_3 = 0\}$ or $\{b_1 = 0, b_2 = 0, b_3 = 1\}$. See Bates and Granger (1969) for the development of the related concept of forecast combination.

Ericsson (1993) proposes a modification of Eq. (1), noting that if each forecast is cointegrated (+1:−1) with the variable being forecast, then $b_1 + b_2 + b_3 = 1$. Ericsson (1992) illustrates the desirability of accounting for that cointegration in forecast-encompassing tests. The resulting modified framework is:

$$(y_t - x_t) = b_4{}^*(w_t - x_t) + b_5{}^*(z_t - x_t) + e_t, \quad (2)$$

where $(y_t - x_t)$ is the first agency's forecast error and $(w_t - x_t)$ is the difference between the second and first agencies' forecasts, while $(z_t - x_t)$ is the difference between the third and first agencies' forecasts. That is, $(w_t - x_t)$ and $(z_t - x_t)$ measure the information that is present in the second and third agencies' forecasts but not in the first agency's forecast. Under this approach, the null hypothesis tested is $\{b_4 = 0, b_5 = 0\}$. That is, assuming that the coefficients on the agencies' forecasts in Eq. (1) sum to 1, the first agency's forecast error cannot be explained by its deviations from the other agencies' forecasts. The roles of the three agencies can be reversed by testing the null hypothesis that $\{b_4 = 1, b_5 = 0\}$ or $\{b_4 = 0, b_5 = 1\}$ in Eq. (2).

Ericsson and Marquez (1993) observe that, in general, the framework in Eq. (1) implicitly assumes unbiased forecasts, because Eq. (1) omits an intercept. By introducing an intercept to Eq. (1), additional tests can be performed to examine whether the forecasts are biased.

These observations about Eqs. (1) and (2) generate fifteen variants of the forecast-encompassing test. Specifically, for a triplet of forecasts, five different hypotheses about $\{b_1, b_2, b_3, b_4, b_5\}$ are of particular interest:

(a) $b_1 = 1, b_2 = 0$ and $b_3 = 0$,
(b) $b_2 = 0$ and $b_3 = 0$,
(c) $b_1 \equiv 1, b_2 = 0$ and $b_3 = 0$,
(d) $b_1 + b_2 + b_3 = 1$, and
(e) $b_1 + b_2 + b_3 \equiv 1, b_4 = 0$ and $b_5 = 0$.

In (c), "$b_1 \equiv 1$" indicates that b_1 is constrained to unity, whether b_2 and b_3 are constrained to zero or not; and likewise for "$b_1 + b_2 + b_3 \equiv 1$" in (e). Tests of (a)–(c), (d) and (e) are adaptations of those proposed by Chong and Hendry (1986) and Ericsson (1992, 1993), respectively.

In addition, three different hypotheses about the intercept (denoted b_0) can be tested:

(i) an intercept is included in Eq. (1) and is left unrestricted ("b_0 unrestricted"),
(ii) an intercept is included in Eq. (1) but is set to zero ("$b_0 = 0$") when any of (a)–(e) is imposed, and
(iii) an intercept is not included in Eq. (1) ("$b_0 \equiv 0$").

The hypotheses in (a)–(e), in combination with those in (i)–(iii), generate fifteen variants of the forecast-encompassing test, where the variants differ in their assumptions about $\{b_0, b_1, b_2, b_3, b_4, b_5\}$ under the null hypothesis of forecast encompassing and the alternative hypothesis of a lack of forecast encompassing. Thus, each variant has implications for the size and power of the statistical test. As a group, these variants provide a comprehensive view of whether any one set of forecasts is able to explain the actual outcomes better than any other set of forecasts. Because a non-zero intercept reflects a systematic bias, the analysis below focuses on variants involving (ii).

4. Data sources, description, and summary statistics

This section describes the data and provides some visual comparisons of the forecasts. The primary variable of interest is the total US gross federal debt outstanding, in billions of dollars, from 1984 to 2013 ("DEBT"). The analysis in this paper also examines the US federal deficit ("DEFICIT"). The data for both DEBT and DEFICIT are published by the US Department of Treasury's Financial Management Service in the December issue of the *Treasury Bulletin*, and are measured on the basis of fiscal years ending on September 30th.

In public discourse, discussions of the deficit often overshadow discussions of the federal debt, since the deficit is commonly thought of as equaling the change in debt. However, the deficit does not in fact equal the

[5] Similar approaches were used by Belongia (1988), Cohen and Follette (2003), Corder (2005), and Douglas and Krause (2005).

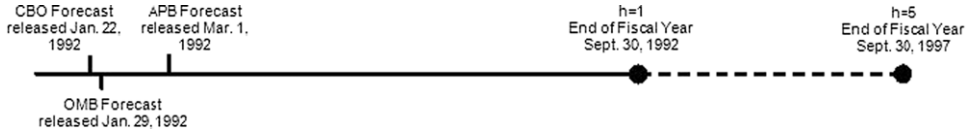

Fig. 1. Illustrative timeline of CBO, OMB, and APB forecasts.

Table 2
Comparisons of actual debt and forecasts of debt from CBO and OMB for 2009 (billions of $).

Budget item	Actual	CBO	OMB
Change in public debt	1741.7	1390.0	2728.7
Deficit	1412.7	1186.0	1841.2
Changes in financial assets and liabilities	329.0	204.0	887.5
Change in cash balances	−96.3	−297.0	−301.6
TARP	105.4	461.0	166.4
Other	319.9	40.0	1022.7
Change in debt held by the government	148.1	153.0	153.1
Change in agency-issued debt	−0.3	1.0	−0.2
Total (i.e., change in the federal debt)	1889.5	1544.0	2881.6

Notes:
1. The change in debt held by the public is the sum of the deficit and "other".
2. Other changes in financial assets and liabilities include direct loans and guaranteed loans, purchases of mortgage-backed securities, and additional financing accounts for "potential additional financial stabilization efforts".
3. The change in federal debt is the sum of changes to public debt, debt held by the government, and agency-issued debt.
4. OMB values come from the OMB's February 2009 publication *Budget of the US Government Fiscal Year 2010.*
5. CBO values come from the CBO's January 2009 publication *Budget and Economic Outlook: Fiscal Years 2009–2019.*
6. Actual values come from the OMB's February 2010 publication *Budget of the US Government Fiscal Year 2011.*

change in debt, as the deficit excludes certain items that contribute to the change in debt, such as the Troubled Asset Relief Program (TARP) and changes in cash balances held by the Treasury. The inclusion or exclusion of such items can alter the measure of debt calculated significantly, and the CBO and OMB debt forecasts and their relative merits depend on which measure of debt is used. For example, the difference between the CBO and OMB debt forecasts in 2009 was due largely to differences in the forecasts of the change in financial assets and liabilities in response to the financial crisis; see Table 2. Furthermore, the gross federal debt is a closer measure of the debt subject to the debt ceiling than is debt held by the public.[6] Therefore, it is important to examine the forecasts of the federal debt, which aggregates across multiple sources of debt, including the deficit. Focusing on the deficit alone could miss significant components of debt.

The forecasts examined are one-year-ahead ($h = 1$) and five-year-ahead ($h = 5$) forecasts of the federal debt and deficit. These forecasts span the period 1984–2013 and appear in annual releases of the CBO's *Budget and Economic Outlook*, the OMB's *Budget of the United States Government*, and the CBO's *Analysis of the President's Budget*. The forecasts are denoted CBO, OMB, and APB respectively. They are released at the beginning of the year, usually about a month apart (see Table 3) between January and March, and are of the level of the US federal debt and deficit at the end of the (then) current fiscal year and of future fiscal years. Thus, they are not precisely one-year-ahead and five-year-ahead forecasts, but will be referred to as such for ease of reference. See Fig. 1 for an illustrative timeline, and Martinez (2011, Table 2) for specific dates.

Table 3
Differences in forecast release dates (in number of days, 1984–2013).

	OMB relative to CBO	APB relative to OMB
Mean	17	36
Median	12	31
Max	89	89
Min	2	12

It is important to emphasize that these agency forecasts are conditioned on different policy assumptions. The CBO forecasts assume that current laws will remain unchanged over the forecast horizon, whereas the OMB and APB forecasts assume that the policy changes proposed in the president's budget will be implemented. In this sense, the forecasts can be viewed more as different policy scenarios than as true unconditional forecasts. However, it is still of interest to determine how useful these different policy scenarios are (both relatively and absolutely), and whether any of these scenarios incorporate all of the relevant information available in the other scenarios, especially with the prominence they are given in policy debates and decisions. For more information on the forecasts and their assumptions, see Martinez (2011).

In the current analysis, the forecasts that are of primary interest are the CBO, OMB, and APB one-year-ahead and five-year-ahead forecasts of the federal debt and deficit. Fig. 2 plots the logs of the debt forecasts and of the actual debt so as to provide an initial indication of how much the debt has grown over the past 30 years, and how the one-year-ahead and five-year-ahead forecasts have performed. Fig. 3 plots the debt forecast errors, which were generated

[6] However, the gross federal debt is not exactly the same as the debt subject to limit, as the debt subject to limit excludes most debt issued by agencies other than the Treasury and the Federal Financing Bank. The difference between the two measures was $17 billion in 2013, and has averaged about $30 billion since 1990.

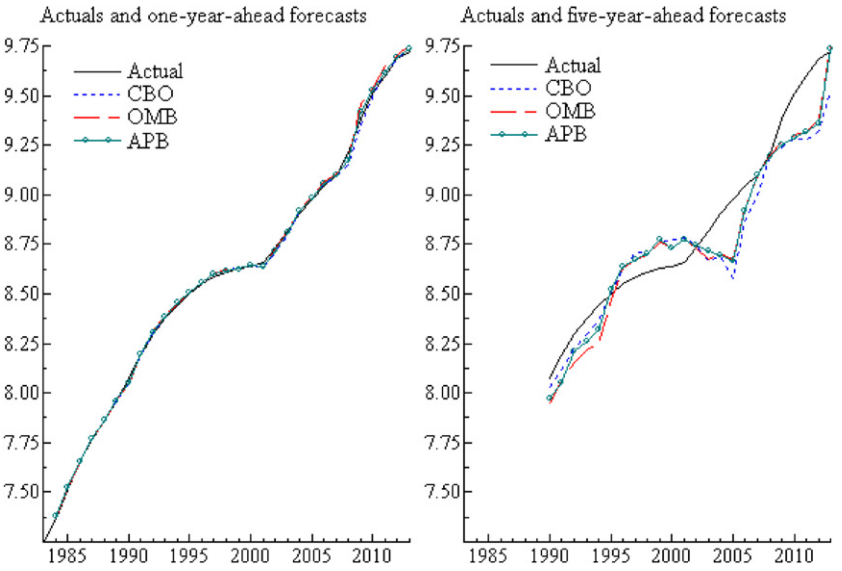

Fig. 2. The logs of actual debt and of the one-year-ahead and five-year-ahead CBO, OMB, and APB forecasts.

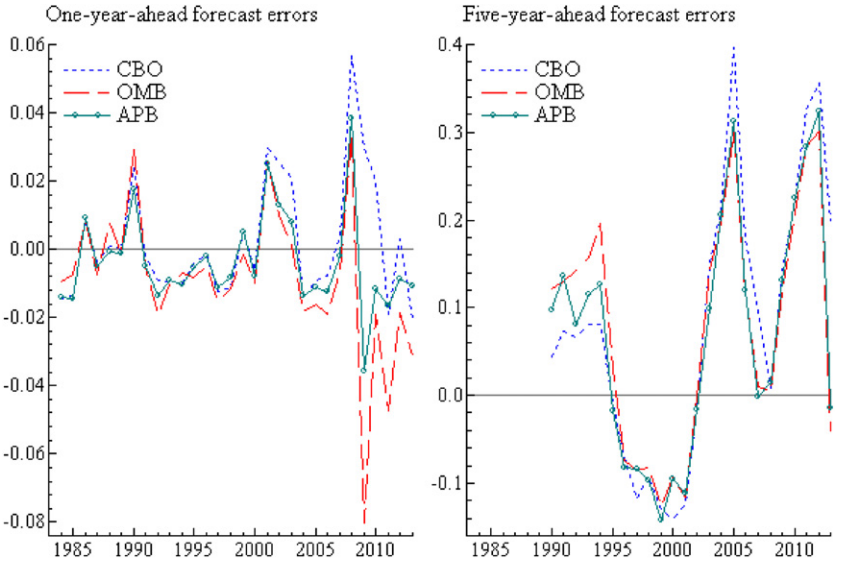

Fig. 3. One-year-ahead and five-year-ahead CBO, OMB, and APB debt forecast errors.

by subtracting the logs of the forecasts from the log of the actual debt. Fig. 4 plots the deficit forecasts and actual values, while Fig. 5 plots the deficit forecast errors. The largest errors for one-year-ahead debt and deficit forecasts were in 1990, 2001, 2002, 2008, 2009, 2011 and 2013. In each of these years, either the US was entering a recession or expansion (according to the NBER), or there were major policy changes. On the other hand, the five-year-ahead debt and deficit forecasts exhibit large and prolonged errors that imply under-prediction, except for a brief period of over-prediction from 1995 to 2001.

For purposes of further comparison, the analysis below includes an average of the logs of the OMB and CBO forecasts (AVE1), an average of the logs of the OMB and APB forecasts (AVE2), an average of the logs of the CBO and APB forecasts (AVE3), and an average of the logs of all three agency forecasts (AVE4). Comparing the individual agency forecasts with their averages provides an additional assessment of whether a simple combination of the forecasts incorporates all of the relevant information in the individual agency forecasts. For details of the data series and their sources, see Table 4.

Estimated forecast biases, the error variances, and the RMSEs provide further insights into the relative performances of the forecasts. Table 5 reports these summary statistics for the three agency one-year-ahead debt forecasts and for the forecasts from a random walk model, including over various subsamples. A negative bias indicates

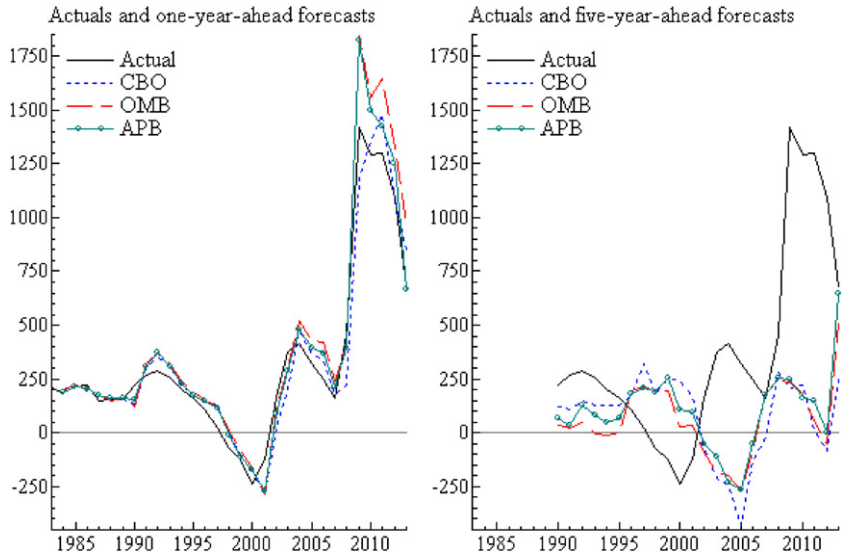

Fig. 4. The actual deficit and the one-year-ahead and five-year-ahead CBO, OMB, and APB forecasts of the deficit ($ billions).

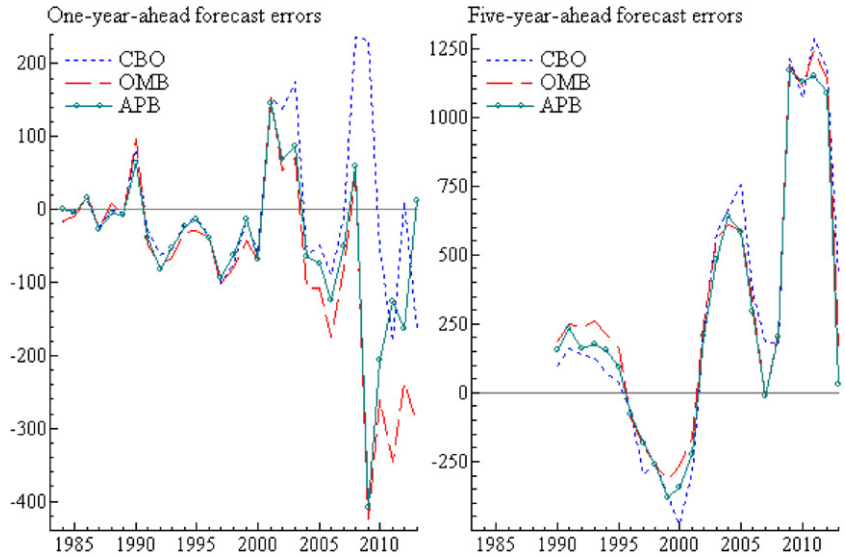

Fig. 5. One-year-ahead and five-year-ahead CBO, OMB, and APB deficit forecast errors ($ billions).

Table 4
Variables (in $ billions, 1984–2013).

Variable	Description	Source
DEBT	Annual value of the gross federal debt outstanding (held by public and intra-governmental holdings).	Financial Management Service
DEFICIT	Annual value of the federal deficit.	Financial Management Service
CBO	Annual one-year-ahead and five-year-ahead forecasts of the debt and deficit from CBO.	Congressional Budget Office
OMB	Annual one-year-ahead and five-year-ahead forecasts of the debt and deficit from OMB.	Office of Management and Budget
APB	Annual one-year-ahead and five-year-ahead forecasts of the debt and deficit from CBO's analysis of the President's Budget.	Congressional Budget Office

systematic over-prediction. For all subsamples, the agencies' forecasts perform better than the forecasts from a random walk model. For the subsample ending in 2008, the RMSE for OMB is smaller than that for CBO, but the converse holds for subsamples ending between 2009 and 2013. However, regardless of which subsample is chosen,

Table 5
Summary statistics for comparing one-year-ahead forecasts of debt (in %).

Forecast	Statistic	1984–2008	1984–2009	1984–2010	1984–2011	1984–2012	1984–2013
CBO	Bias	0.18	0.28	0.35	0.27	0.27	0.19
	Variance	0.03	0.03	0.03	0.03	0.03	0.03
	RMSE	1.73	1.80	1.80	1.81	1.78	1.79
OMB	Bias	−0.25	−0.55	−0.60	−0.75	−0.79	−0.87
	Variance	0.02	0.04	0.04	0.05	0.05	0.05
	RMSE	1.47	2.16	2.15	2.29	2.28	2.31
APB	Bias	−0.12	−0.26	−0.29	−0.34	−0.36	−0.38
	Variance	0.02	0.02	0.02	0.02	0.02	0.02
	RMSE	1.35	1.50	1.49	1.49	1.48	1.46
Random walk model	Bias	−2.01	−1.60	−1.39	−1.35	−1.32	−1.44
	Variance	0.14	0.18	0.18	0.18	0.17	0.17
	RMSE	4.23	4.50	4.49	4.41	4.33	4.35

Notes:
1. Bias is calculated as the mean forecast error over the sample. A negative bias indicates over-prediction.
2. The variance is calculated as the squared standard deviation of the forecast error over the sample.
3. The RMSE is calculated by taking the square root of the sum of the variance and the squared bias.

Table 6
Summary statistics for comparing five-year-ahead forecasts of debt (in %).

Forecast	Statistic	1990–2008	1990–2009	1990–2010	1990–2011	1990–2012	1990–2013
CBO	Bias	3.64	4.17	5.03	6.28	7.56	8.07
	Variance	1.92	1.88	1.94	2.19	2.47	2.42
	RMSE	14.34	14.33	14.81	16.09	17.44	17.54
OMB	Bias	5.10	5.44	6.17	7.18	8.18	7.66
	Variance	1.60	1.54	1.57	1.72	1.88	1.86
	RMSE	13.64	13.54	13.98	14.96	15.95	15.64
APB	Bias	3.47	3.95	4.83	5.90	7.05	6.70
	Variance	1.50	1.47	1.56	1.74	1.96	1.91
	RMSE	12.74	12.75	13.39	14.44	15.68	15.34
Random walk model	Bias	−16.74	−15.69	−14.46	−13.23	−11.98	−11.13
	Variance	1.98	2.09	2.31	2.53	2.77	2.83
	RMSE	21.86	21.34	20.97	20.68	20.51	20.16

Notes:
1. See the notes for Table 5.

the APB forecasts outperform both the CBO and OMB forecasts in terms of biases, error variances, and RMSEs. The same results hold for the one-year-ahead deficit forecasts, except that APB performs worse than CBO for subsamples ending in 2009 or later.

Table 6 summarizes the properties of the five-year-ahead debt forecasts. The agencies' forecasts continue to outperform the random walk forecasts, while OMB performs better than CBO. APB also continues to outperform both OMB and CBO across all subsamples.[7]

None of the agencies' five-year-ahead *deficit* forecasts outperform the random walk forecasts. In fact, Kliesen and Thornton (2012) find that the CBO deficit forecasts fail to outperform random walk forecasts for two-year-ahead forecasts. This preliminary analysis suggests that, while the agency forecasts perform well relative to a random walk forecast, the relative performance depends in general on the variable, the sample, and the forecast horizon.

[7] The performance also depends on which naive forecast is used. A simple robust forecasting device of the double difference of debt is much more in line with the performances of the agency forecasts of debt in terms of the RMSE at both horizons. Furthermore, it outperforms the CBO's five-year-ahead debt forecast for the full sample ending in 2013.

5. Analysis and results

Forecast-encompassing tests provide an alternative method of comparing forecasts. The analysis below initially examines one-year-ahead debt forecasts. The analysis is then extended to five-year-ahead debt forecasts, and to one-year-ahead and five-year-ahead deficit forecasts. The results show that each agency's forecast of the debt is incomplete, and therefore a combination of the agencies' forecasts performs better than the individual forecasts.

Generalizing on Eq. (1) so as to include an intercept, the baseline regression for the forecast-encompassing tests is obtained by regressing the actual debt on an intercept and the CBO, OMB, and APB forecasts:

$$\begin{array}{lllll} \text{debt} = & -0.069 & -0.13^{*}\text{cbo} & -1.01^{*}\text{omb} & +2.15^{*}\text{apb}, \\ & (0.043) & (0.21) & (0.17) & (0.47) \end{array} \quad (3)$$

as appears in Table 7 (regression #1). Lowercase variables denote the logs of uppercase variables, estimated standard errors are in parentheses, and the sample period is 1984–2013. The coefficients suggest that the APB forecast should be given double the weight, whereas the coefficient for the CBO forecast is not statistically significant, and a negative weight should be given to OMB, suggesting that

Table 7
One-year-ahead debt and deficit forecast-encompassing regressions.

#	(1984–2013)	Sigma	$b_1 + b_2 + b_3 = 1$
1	debt = −0.069 −0.131 × cbo−1.012 × omb+2.151 × apb (0.043) (0.208) (0.337) (0.4712)	1.26%	1. 33 [0.281] (2, 26)
2	(debt-apb) = −0.0012 −0.117 × (cbo-apb) − 0.672 × (omb-apb) (0.0026) (0.213) (0.265)	1.30%	–
3	deficit = 11.8 +0.135 × cbo+0.215 × omb+0.499 × apb (14.6) (0.164) (0.290) (0.228)	62.50	5.07* [0.014] (2, 26)
4	(deficit-apb) = −4.9 +0.564 × (cbo-apb)−0.508 × (omb-apb) (15.7) (0.106) (0.206)	72.17	–

Notes:
1. "cbo", "omb", and "apb" in rows 1 and 2 refer to those agencies' debt forecasts, whereas in rows 3 and 4 they refer to those agencies' deficit forecasts.
2. The three entries within a given block of numbers in the last column are: the approximate F-statistic for testing the null hypothesis, the tail probability associated with that value of the F-statistic (in square brackets), and the degrees of freedom for the F-statistic (in parentheses).
* Denotes rejection at the 5% critical value.
** Denotes rejection at the 1% critical value.

one should move in the opposite direction to the OMB forecast given the APB forecast.

Now consider the five null hypotheses in (a)–(e) above, as applied to Eq. (3). Implicitly, in the discussion below, the intercept is also being tested to be zero as in (ii) above (i.e., $b_0 = 0$).

The first null hypothesis is (a), which is $\{b_{CBO} = 1, b_{OMB} = 0, b_{APB} = 0\}$, where b_{CBO}, b_{OMB}, and b_{APB} are the coefficients on cbo, omb and apb, respectively, in Eq. (3). A test of this hypothesis examines whether the CBO forecast on its own explains the debt, with the OMB and the APB forecasts providing no additional explanatory information. If that null hypothesis is not rejected, then the CBO forecast encompasses the OMB and APB forecasts. In Eq. (3), this null hypothesis is rejected strongly, with an F-statistic of 8.04 and a p-value of less than 0.1%. This outcome is reported as the upper left "triplet" of numbers in the first major row of Table 9, with the parenthetical elements in the triplet specifying the degrees of freedom for the F-statistic: (4, 26).

The second null hypothesis is (b), which is $\{b_{OMB} = 0, b_{APB} = 0\}$. This hypothesis is a slightly weaker version of (a), and focuses on whether the APB and OMB forecasts provide any additional information over and above that provided by the CBO forecast. The null hypothesis (b) is also rejected strongly, with an F-statistic of 10.45 and a p-value of less than 0.1%.

The third null hypothesis is (c), which is $\{b_{CBO} \equiv 1, b_{OMB} = 0, b_{APB} = 0\}$. This hypothesis focuses on whether the OMB and APB forecasts help explain the CBO's forecast error. The null hypothesis (c) is not rejected, with an F-statistic of 0.43 and a p-value of 73.6%. The rejection of (a) and (b) is not surprising, given the regression coefficients in Eq. (3), and implies that CBO does not encompass the OMB and APB forecasts. The failure to reject (c), on the other hand, implies that the other agencies' forecasts cannot explain CBO's forecast errors.

The fourth null hypothesis is (d), which is $\{b_{CBO} + b_{OMB} + b_{APB} = 1\}$. If this hypothesis holds, then the sum of the coefficients on the OMB and CBO forecasts in Eq. (3) is equal to unity. The null hypothesis (d) is not rejected (see Table 7, regression #2): the F-statistic is 1.33, with a p-value of 28.1%. The homogeneity restriction in (d) allows Eq. (1) to be rewritten as Eq. (2), which leads to the examination of hypothesis (e).

The fifth null hypothesis is (e), which is $\{b_{CBO} + b_{OMB} + b_{APB} \equiv 1, b_4 = 0, b_5 = 0\}$, and so conditions on the hypothesis tested in (d). This hypothesis can be examined by estimating Eq. (2) with an unrestricted intercept. Letting x_t be the CBO forecast, w_t be the OMB forecast, and z_t be the APB forecast, the estimated equation for (e) is:

$$\underset{}{\text{(debt-cbo)}} = \underset{(0.0026)}{-0.0012} \ \underset{(0.26)}{-0.67^{*}\text{(omb-cbo)}} \ \underset{(0.42)}{+1.79^{*}\text{(apb-cbo)}}. \quad (4)$$

As was discussed in Section 3, testing $\{b_4 = 0, b_5 = 0\}$ in Eq. (2) corresponds to testing whether the coefficients on (omb-cbo) and (apb-cbo) are zero in Eq. (4). That is, do the discrepancies between the OMB and CBO *forecasts* and the discrepancies between the APB and CBO *forecasts* help explain the CBO *forecast errors*? The null hypothesis (e) is rejected strongly, with an F-statistic of 9.39 and a p-value of less than 0.1% (Table 9). Thus, the differences between the OMB and CBO forecasts and the differences between the APB and CBO forecasts do help to explain the CBO forecast errors.

The results with the OMB as the reference forecast are reported in the second major row in Table 9. Specifically, Eq. (3) can be used to test whether the OMB forecasts encompass the CBO forecasts, simply by swapping the coefficients on which the hypotheses are tested. Thus, the new null hypothesis for (a) is $\{b_{CBO} = 0, b_{OMB} = 1, b_{APB} = 0\}$, which examines whether OMB's forecasts explain the debt, with CBO and APB providing no additional explanatory information. The null hypothesis is rejected strongly: the F-statistic is 17.89, with a p-value of less than 0.1%. The new hypotheses for (b) and (c) are $\{b_{CBO} = 0, b_{APB} = 0\}$ and $\{b_{OMB} \equiv 1, b_{CBO} = 0, b_{APB} = 0\}$ respectively: both hypotheses are rejected strongly. The new hypothesis for (d) is identical that reported in Table 7 for equation #1. The new null hypothesis for (e) corresponds to testing

Table 8
Five-year-ahead debt and deficit forecast-encompassing regressions.

#	(1990–2013)	Sigma	$b_1 + b_2 + b_3 = 1$
1	debt = 0.096 −0.143 × cbo+0.721 × omb+0.419 × apb (1.450) (0.974) (1.366) (1.239)	14.6%	1. 00 [0.385] (2, 20)
2	(debt-apb) = 0.072 −0.133 × (cbo-apb)+0.701 × (omb-apb) (0.051) (0.495) (1.044)	14.2%	–
3	deficit = 368.6 −1.411 × cbo+1.313 × omb+0.310 × apb (167.2) (0.715) (2.613) (2.106)	460.67	6.89** [0.005] (2, 20)
4	(deficit-apb) = 332.0 −1.201 × (cbo-apb)+2.301 × (omb-apb) (180.1) (0.707) (2.164)	470.36	–

Notes:
1. See the notes for Table 7.
2. The standard errors for the five-year-ahead forecasts are calculated using the HAC variance–covariance matrix.
3. Five-year-ahead forecasts were first produced by all agencies in 1986.

$\{b_4 = 1, b_5 = 0\}$ in Eq. (2), which means testing that the coefficient on (omb-cbo) is unity and the coefficient on (apb-cbo) is zero in Eq. (4). This hypothesis is rejected strongly. Thus, the OMB forecasts do not encompass the CBO and APB forecasts, as they provide additional information that can be incorporated into the OMB forecasts.

The results for APB as the reference forecast are reported in the third major row in Table 9. The test statistics show that the APB forecast does not encompass the CBO and OMB forecasts: three of the four null hypotheses are rejected at the 5% level. These results stand in contrast to the finding that APB has the lowest RMSE of the three agency forecasts, at 1.46%.

The contrast between the RMSE ranking and the forecast-encompassing tests can be understood by rewriting Eq. (4) with APB as the reference forecast:

$$\underset{(0.0026)\ (0.21)\qquad\qquad(0.27)}{\text{(debt-apb)} = -0.0012\ \ -0.12^{*}\text{(cbo-apb)}\ -0.67^{*}\text{(omb-apb)}}. \quad (5)$$

In Eq. (5), sigma is 1.30%, whereas the RMSE for APB is considerably larger, at 1.46%. Thus, the CBO and OMB forecasts explain about 10% of the APB's RMSE. While APB has the smallest RMSE among the individual agency forecasts, it still lacks some relevant information that is available from the other agencies' forecasts. This empirical result reflects the theoretical finding of Ericsson (1992) that dominance in RMSE is necessary but not sufficient for forecast encompassing.

In Eq. (5), the negative coefficient on (omb-apb) implies that forecast improvements can be achieved by moving away from the OMB forecast, relative to the APB forecast. That result is also reflected in Eq. (3), where the OMB forecast has a negative coefficient and the APB forecast has a coefficient that is greater than unity. Because APB often performs well among the individual agency forecasts, APB is used as the reference forecast in Tables 7 and 8 for regressions of the form of Eq. (2).

The results from using each forecast average as the reference forecast are displayed in the remaining major rows in Table 9. None of these forecast averages encompass the individual agencies' forecasts fully. In other words, simple averages of the agency forecasts fail to incorporate all of the relevant information that is available in the agencies' forecasts themselves. However, that being said, AVE3 has the smallest RMSE of the forecast averages considered, and AVE3's RMSE is equal to that of the APB.

All of the results mentioned above were obtained with the intercept being restricted to equal zero under the null hypothesis. The results are robust to the treatment of the intercept, to changes in the sample, and to the removal of years with large forecast errors; see Martinez (2011).

The results do change when the analysis is extended to longer forecast horizons. The right-hand panel in Table 9 presents the results of the five-year-ahead debt forecasts, and shows that most of the agencies' forecasts and their averages encompass one another, albeit with some evidence to suggest that CBO does not encompass the other agencies' forecasts. APB has the lowest RMSE, at 15.34%, of which the CBO and OMB forecasts can explain 7% (see Table 8, equation #2).

The results also change when they are extended to the deficit. Table 10 shows that none of the agencies' deficit forecasts or their averages encompass any of the others at either forecast horizon. However, when the intercept is left unrestricted at the five-year-ahead forecast horizon (not shown), all of the forecasts encompass one another, though with some evidence suggesting that CBO does not encompass the other agencies' forecasts. AVE3 has the lowest RMSEs for the one-year-ahead forecasts, while APB has the lowest RMSEs for the five-year-ahead forecasts.

The findings in this paper show that, at both short and long horizons, a combination of agencies' forecasts is better at explaining the debt than any of the forecasts individually, suggesting that a combination of the forecasts is more robust to changes in the economy than the individual agency forecasts. This is supported by the theoretical analysis of Clements and Hendry (2004), which shows that pooling forecasts can add value when individual forecasting models are differentially mis-specified. Furthermore, Hendry and Mizon (2001) illustrate that pooling across models may be beneficial in forecasting when there are structural breaks or policy regime shifts. Thus, the weaknesses of individual forecasts may be ameliorated by combining forecasts.

Table 9
Forecast-encompassing test statistics for alternative US federal debt forecasts (log levels), by horizon and null hypothesis.

	One-year-ahead (1984–2013)					Five-year-ahead (1990–2013)				
Reference forecast	RMSE (%)	$b_1 = 1$, $b_2 = 0$, $b_3 = 0$	$b_2 = 0$, $b_3 = 0$	$b_1 \equiv 1$, $b_2 = 0$, $b_3 = 0$	$b_1 + b_2 + b_3 \equiv 1$, $b_4 = 0$, $b_5 = 0$	RMSE (%)	$b_1 = 1$, $b_2 = 0$, $b_3 = 0$	$b_2 = 0$, $b_3 = 0$	$b_1 \equiv 1$, $b_2 = 0$, $b_3 = 0$	$b_1 + b_2 + b_3 \equiv 1$, $b_4 = 0, b_5 = 0$
CBO	1.79	8.04** [0.000] (4, 26)	10.45** [0.000] (3, 26)	0.43 [0.736] (3, 27)	9.39** [0.000] (3, 27)	17.54	4.14* [0.013] (4, 20)	2.39 [0.099] (3, 20)	2.59 [0.080] (3, 21)	5.26** [0.007] (3, 21)
OMB	2.31	17.89** [0.000] (4, 26)	18.37** [0.000] (3, 26)	5.27** [0.005] (3, 27)	21.86** [0.000] (3, 27)	15.64	1.81 [0.166] (4, 20)	0.05 [0.987] (3, 20)	1.46 [0.254] (3, 21)	2.41 [0.095] (3, 21)
APB	1.46	3.26* [0.027] (4, 26)	3.38* [0.033] (3, 26)	2.00 [0.138] (3, 27)	3.35* [0.034] (3, 27)	15.34	0.93 [0.469] (4, 20)	0.20 [0.898] (3, 20)	1.07 [0.382] (3, 21)	1.27 [0.310] (3, 21)
AVE1	1.68	6.38** [0.001] (4, 26)	7.68** [0.001] (3, 26)	4.19* [0.015] (3, 27)	7.30** [0.001] (3, 27)	16.21	2.52 [0.073] (4, 20)	0.61 [0.618] (3, 20)	1.74 [0.190] (3, 21)	3.23* [0.043] (3, 21)
AVE2	1.84	8.95** [0.000] (4, 26)	9.16** [0.000] (3, 26)	3.07* [0.045] (3, 27)	10.55** [0.000] (3, 27)	15.43	1.25 [0.322] (4, 20)	0.03 [0.994] (3, 20)	1.26 [0.314] (3, 21)	1.69 [0.201] (3, 21)
AVE3	1.46	3.19* [0.029] (4, 26)	4.20* [0.015] (3, 26)	0.91 [0.448] (3, 27)	3.26* [0.037] (3, 27)	16.19	2.13 [0.115] (4, 20)	0.88 [0.468] (3, 20)	1.52 [0.238] (3, 21)	2.78 [0.066] (3, 21)
AVE4	1.59	4.99** [0.004] (4, 26)	5.78** [0.004] (3, 26)	4.35* [0.013] (3, 27)	5.53** [0.004] (3, 27)	15.86	1.86 [0.158] (4, 20)	0.38 [0.767] (3, 20)	1.38 [0.275] (3, 21)	2.43 [0.094] (3, 21)

Notes:
1. See the notes for Table 8.
2. For each test statistic, the maintained hypothesis is the regression with all three forecasts. For the test statistics with the average forecasts, the maintained hypothesis is the regression with two agency forecasts and the average forecast.
3. $b_4 \equiv b_2 - b_1$, $b_5 \equiv b_3 - b_1$.
4. AVE1 is the average of the CBO and OMB forecasts. AVE2 is the average of the APB and OMB forecasts. AVE3 is the average of the CBO and APB forecasts. AVE4 is the average of the CBO, OMB and APB forecasts.

There are several potential explanations for why the individual agency debt forecasts may be differentially misspecified. First, different agencies release their forecasts on different dates, meaning that forecasts with later release dates (typically APB) may incorporate newer information. Even small differences in release dates could have significant effects at the onset of a recession or in the midst of large economic and policy changes. Second, as was mentioned earlier, the agencies condition their forecasts on different assumptions. These different assumptions could lead to large differences in the forecasts, especially when policies are changing rapidly. While a determination of the specific characteristics of the agencies' forecasts that lead to their divergence is beyond the scope of this analysis, a combination of the forecasts can reduce their susceptibility to changes in policies or in the economy.

6. Conclusions

This paper compares three decades of one-year-ahead and five-year-ahead forecasts for US federal debt and deficits from the Congressional Budget Office and the Office of Management and Budget. Summary statistics and forecast-encompassing tests lead to consistent and robust conclusions. Overall, the agency forecasts perform better than a simple benchmark forecast model. At the one-year-ahead forecast horizon, the APB forecasts outperform the CBO and OMB forecasts in terms of the RMSEs (for the debt but not the deficit). This is not particularly surprising, given APB's later release date, and therefore potential information advantage. However, the APB forecasts are not able to fully forecast-encompass the CBO and OMB forecasts (both deficit and debt). At the five-year-ahead forecast horizon, the APB performs best in terms of RMSEs, but is still unable to encompass the CBO and OMB forecasts. This suggests that the other agencies' forecasts provide additional useful information at both horizons.

Each of the agency forecasts is relatively successful in forecasting the debt and deficit. However, that being said, each agency forecast of the federal debt remains incomplete in both the short and medium term, and could benefit from information that the other agency forecasts take into account. When only one agency's forecast is used, we have an incomplete and potentially distorted picture of the future levels of and changes in federal debt.

The APB forecasts behave somewhat like a quasi-combination of the OMB and CBO forecasts, and the evidence suggests that the APB forecasts outperform the other agency debt forecasts at both horizons. However, despite being released as much as several months after the CBO and OMB forecasts, the APB forecast is still missing information that the other agency forecasts take into account. The utilization of information from each of the agencies' forecasts provides a more complete picture of the future trajectory of the US federal debt.

Table 10
Forecast-encompassing test statistics for alternative US federal deficit forecasts (levels, in $ billions), by horizon and null hypothesis.

	One-year-ahead (1984–2013)					Five-year-ahead (1990–2013)				
Reference forecast	RMSE	$b_1 = 1$, $b_2 = 0$, $b_3 = 0$	$b_2 = 0$, $b_3 = 0$	$b_1 \equiv 1$, $b_2 = 0$, $b_3 = 0$	$b_1 + b_2 + b_3 \equiv 1$, $b_4 = 0$, $b_5 = 0$	RMSE	$b_1 = 1$, $b_2 = 0$, $b_3 = 0$	$b_2 = 0$, $b_3 = 0$	$b_1 \equiv 1$, $b_2 = 0$, $b_3 = 0$	$b_1 + b_2 + b_3 \equiv 1$, $b_4 = 0$, $b_5 = 0$
CBO	101.3	12.54** [0.000] (4, 26)	15.92** [0.000] (3, 26)	3.71* [0.024] (3, 27)	10.04** [0.000] (3, 27)	591.2	4.47** [0.010] (4, 20)	4.71* [0.012] (3, 20)	3.20* [0.044] (3, 21)	4.77* [0.011] (3, 21)
OMB	149.7	35.43** [0.000] (4, 26)	1.85 [0.164] (3, 26)	36.28** [0.000] (3, 27)	32.92** [0.000] (3, 27)	561.8	3.55* [0.024] (4, 20)	3.97* [0.023] (3, 20)	5.07** [0.009] (3, 21)	2.79 [0.066] (3, 21)
APB	109.9	16.03** [0.000] (4, 26)	1.98 [0.142] (3, 26)	17.29** [0.000] (3, 27)	13.53** [0.000] (3, 27)	544.6	3.46* [0.027] (4, 20)	3.56* [0.033] (3, 20)	4.64* [0.012] (3, 21)	2.22 [0.116] (3, 21)
AVE1	103.7	13.55** [0.000] (4, 26)	4.66** [0.010] (3, 26)	56.13** [0.000] (3, 21)	14.96** [0.000] (3, 21)	573.8	3.81* [0.019] (4, 20)	4.35* [0.016] (3, 20)	5.73** [0.005] (3, 21)	3.61* [0.030] (3, 21)
AVE2	125.9	23.16** [0.000] (4, 26)	0.50 [0.686] (3, 26)	54.44** [0.000] (3, 21)	13.18** [0.000] (3, 21)	552.6	3.46* [0.027] (4, 20)	3.75* [0.028] (3, 20)	5.02** [0.009] (3, 21)	2.46 [0.091] (3, 21)
AVE3	80.5	5.57** [0.002] (4, 26)	1.79 [0.175] (3, 26)	53.90** [0.000] (3, 21)	13.75** [0.000] (3, 21)	565.2	3.66* [0.022] (4, 20)	4.12* [0.020] (3, 20)	5.78** [0.005] (3, 21)	3.31* [0.040] (3, 21)
AVE4	98.7	11.68** [0.000] (4, 26)	1.12 [0.360] (3, 26)	56.63** [0.000] (3, 21)	13.99** [0.000] (3, 21)	563.2	3.58* [0.023] (4, 20)	4.05* [0.021] (3, 20)	7.24** [0.002] (3, 21)	3.08* [0.050] (3, 21)

Notes:
1. See the notes for Table 9.

Acknowledgments

All numerical results were obtained using PcGive Version 13.1 in OxMetrics Version 6.10; see Doornik and Hendry (2009). The author is grateful for helpful suggestions and comments from an associate editor and two referees, as well as Michael Clements, Neil R. Ericsson, David F. Hendry, Frederick L. Joutz, Tara M. Sinclair, Herman O. Stekler, Peter A. Tinsley, Christopher Williams, and participants at the GWU Brown Bag Seminar Series on Forecasting, the 18th Federal Forecasters Conference, the Southern Economic Association 81st Annual Meetings, the 11th OxMetrics User Conference, and the 32nd Annual International Symposium on Forecasting Conference.

Appendix A. Supplementary data

The forecasts and data used in this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2014.1008.014.

References

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Operations Research Quarterly*, *20*, 451–468.
Belongia, M. (1988). Are economic forecasts by government agencies biased? Accurate? *Federal Reserve Bank of St. Louis Review*, *70*, 15–23.
Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, *53*(4), 671–690.
Clements, M. P., & Hendry, D. F. (2002). An overview of economic forecasting. In M. P. Clements & D. F. Hendry (Eds.), *A companion to economic forecasting* (pp. 1–18). Oxford: Blackwell Publishers.
Clements, M. P., & Hendry, D. F. (2004). Pooling of forecasts. *Econometrics Journal*, *7*, 1–31.
Cohen, D., & Follette, G. (2003). *Forecasting exogenous fiscal variables in the United States*. Federal Reserve Board, Finance and Economics discussion series, no. 59.
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.
Doornik, J. A., & Hendry, D. F. (2009). *Empirical econometric modeling using PcGive: Volume I* (6th ed.). London: Timberlake Consultants Press.
Douglas, J. W., & Krause, G. A. (2005). Institutional design versus reputational effects on bureaucratic performance: evidence from US government macroeconomic and fiscal projections. *Journal of Public Administration Research and Theory*, *15*(2), 281–306.
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. (1993). On the limitations of comparing mean square forecast errors: clarifications and extensions. *Journal of Forecasting*, *12*(8), 644–651.
Ericsson, N. R., & Marquez, J. (1993). Encompassing the forecasts of US trade balance models. *The Review of Economics and Statistics*, *75*(1), 19–31.
Frankel, J. A. (2011). Over-optimism in forecasts by official budget agencies and its implications. *Oxford Review of Economic Policy*, *27*(4), 536–562.
Frendreis, J., & Tatalovich, R. (2000). Accuracy and bias in macroeconomic forecasting by the administration, the CBO, and the Federal Reserve Board. *Polity*, *32*(4), 623–632.
Hendry, D. F., & Mizon, G. E. (2001). Forecasting in the presence of structural breaks and policy shifts. In D. W. Andrews, & J. H. Stock (Eds.), *Identification and inference for econometric models: essays in honor of Thomas Rothenberg* (pp. 480–502). Cambridge: Cambridge University Press.
Howard, J. A. (1987). Government economic projections: a comparison between CBO and OMB forecasts. *Public Budgeting and Finance*, *7*(3), 14–25.
IMF Survey (2013). High government debt threatens growth prospects. *IMF Survey Magazine*, January 8, 2013. https://www.imf.org/external/pubs/ft/survey/so/2013/NEW010813A.htm (Accessed on 12 January 2013).
Kamlet, M. S., Mowery, D. C., & Su, T.-T. (1987). Whom do you trust? An analysis of executive and congressional economic forecasts. *Journal of Policy Analysis and Management*, *6*(3), 365–384.

Kliesen, K. L., & Thornton, D. L. (2012). How good are the government's deficit and debt projections and should we care? *Federal Reserve Bank of St. Louis Review*, *94*, 21–39.

Martinez, A. B. (2011). *Comparing government forecasts of the United States' gross federal debt*. GWU Forecasting Program working paper, no. 2011-002.

McNees, S. K. (1995). An assessment of the 'official' economic forecasts. *New England Economic Review*, *July/August*, 13–23.

Plesko, G. A. (1988). The accuracy of government forecasts and budget projections. *National Tax Journal*, *41*(4), 483–501.

Rasche, R. H. (1985). *Deficit projections vs. deficit forecasts*. FRBSF Weekly Letter, 5.

Standard & Poors (2011). United States of America long-term rating lowered to 'AA+' due to political risks, rising debt burden; outlook negative. Press Release, August 5, 2011. http://www.standardandpoors.com/ratings/articles/en/us/?assetID=1245316529563 (Accessed on 5 August 2011).

US Congressional Budget Office (2010). *CBO's economic forecasting record: 2010 update*. www.cbo.gov/doc.cfm?index=11553 (Accessed on 10 July 2010).

US Congressional Budget Office (2013). *CBO's economic forecasting record: 2013 update*. https://www.cbo.gov/publication/43846 (Accessed on 17 January 2013).

Andrew B. Martinez is a research assistant at the Institute for New Economic Thinking at the Oxford Martin School and a graduate student in Economics at the University of Oxford. He received his M.A. in International Trade and Investment Policy from the George Washington University in 2010 and his B.A. in Economics from Guilford College in 2007. His research and professional interests are related to time series econometrics and forecasting, as well as international macroeconomics.