Working Paper Series
Congressional Budget Office
Washington, D.C.

Did Treasury Debt Markets Anticipate the Persistent Decline in Long-Term Interest Rates?

Edward N. Gamber
Congressional Budget Office
ed.gamber@cbo.gov

Working Paper 2017-07

September 2017

To enhance the transparency of the work of the Congressional Budget Office and to encourage external review of that work, CBO's working paper series includes both papers that provide technical descriptions of official CBO analyses and papers that represent original, independent research by CBO analysts. The information in this paper is preliminary and is being circulated to stimulate discussion and critical comment as developmental work for analysis for the Congress. Papers in this series are available at http://go.usa.gov/ULE.

For helpful comments and suggestions, the author thanks Bob Arnold, Wendy Edelberg, Mike Falkenheim, Kim Kowalewski, Damien Moore, Jeff Perry, and Jeff Werling (all of CBO), as well as Jan Hatzius and Brian Sack. The author thanks Christine Bogusz for editing.

Abstract

Private-sector forecasters consistently missed the decline in long-term nominal interest rates over the past three decades, estimating rates that were higher (and, in some cases, much higher) than what actually occurred. This analysis examines whether bond-market participants anticipated with greater accuracy the decline in long-term rates. To explore that issue, the Congressional Budget Office compared the accuracy and bias in forecasts of long-term interest rates from the *Blue Chip* consensus with forecasts based on information derived from the Treasury yield curve as well as several benchmark forecasts and combinations of forecasts. The results indicate that Treasury debt markets did not do a better job than the *Blue Chip* consensus in forecasting the decline in long-term interest rates over the past three decades. Forecasts based on a random walk model of interest rates were more accurate and less biased than those of the *Blue Chip* consensus, especially for the most recent subsample period (1998–2012).

Keywords: long-term interest rates, forecast errors, Treasury yield curve

JEL Classification: E47

Contents

Tables and Figures

Tables

Figures

Introduction

Long-term interest rates in the United States have declined steadily over the past 30 years.[1] Over that same period, private-sector forecasters—as represented by the *Blue Chip* consensus—have generally estimated higher interest rates than what actually occurred.[2] In this analysis, the Congressional Budget Office investigated whether information available from the yield curve at the time the *Blue Chip*'s forecasts were made could have been used to better anticipate the decline in long-term interest rates.[3]

The results of this analysis indicate that compared with the *Blue Chip* consensus, Treasury debt markets did not do a better job forecasting the decline in long-term interest rates over the past three decades. The *Blue Chip* consensus forecast of long-term interest rates was as accurate, or more accurate, than forecasts based on the information contained in the Treasury yield curve. In addition, forecasts based on the Treasury yield curve were not consistently less biased than those of the *Blue Chip* consensus. Combining the two sets of forecasts—those based on the Treasury yield curve and those from the *Blue Chip* consensus—also did not produce significantly more accurate or less biased forecasts.

CBO found that forecasts based on simple benchmark models were more accurate and less biased than those from the *Blue Chip* consensus. That finding was particularly true for forecasts based on the random walk model, especially over the 1998–2012 period. The random walk model naively assumed that interest rates would remain constant, whereas the *Blue Chip* consensus and yield curve–based forecasts assumed that interest rates would move back up to historical levels. By construction, therefore, the random walk–based model was more accurate in an environment of persistently falling interest rates.

---

[1]Real (inflation-adjusted) and nominal long-term interest rates have declined worldwide since the early 1980s. See Rachel and Smith (2015).

[2]The *Blue Chip* consensus is the average of roughly 50 forecasts by private-sector economists (see *Blue Chip Economic Indicators*). It reflects a broader blend of sources and methods than can be expected from a single forecaster. Such combinations of forecasts often provide better estimates than those made by a single forecaster or using a single forecast method. See, for example, Timmerman (2006), Bauer and others (2003), Townsend (1996), and Clemen (1989).

[3]Other researchers have investigated whether the yield curve contains information useful for forecasting interest rates. See, for example, Fama and Bliss (1987), Cochrane and Piazzesi (2005), and Diebold and Rudebusch (2013). Unlike this analysis, those studies have tended to focus on short time horizons (typically one year or less).

The *Blue Chip* Consensus Forecast of Long-Term Interest Rates

To evaluate forecasts of long-term interest rates by the *Blue Chip* consensus, CBO used two measures of forecast quality: statistical bias and accuracy. Statistical bias measures the degree to which the average forecast error differs from zero. Accuracy measures the dispersion of the forecasts around the actual values; it is shown by the root mean square error (RMSE). CBO evaluated the bias and accuracy of the one- through five-year-ahead forecasts of long-term interest rates produced by the *Blue Chip* consensus over the sample period from March 1984 to March 2012, as well as over two subsample periods: March 1984 to October 1997 and March 1998 to March 2012.[4]

The RMSEs of the long-term interest rate forecasts produced by the *Blue Chip* consensus between 1984 and 2012 varied from a low of 0.69 percentage points for the one-year-ahead forecasts to a high of 1.80 percentage points for the five-year-ahead forecasts (see **Table 1**). Over that same period, the forecasts also exhibited significant positive bias (meaning they were too high) for horizons from one through five years (see **Table 2** and **Figure 1**).

Blue Chip's forecasts of long-term interest rates have become more biased and less accurate over time: The RMSEs are substantially larger (in some cases more than twice as large) in the 1998–2012 period as in the 1984–1997 period. Bias also increased from the first half of the sample (1984 through 1997) to the second half (1998 to 2012), which is consistent with the visual impression from Figure 1.

[4]The long-term forecasts by the *Blue Chip* consensus are published in March and October. Long-term interest rates were measured as the AAA corporate bond rate before 1996 and the 10-year Treasury note rate from 1996 onward. The total number of forecasts analyzed in this study is 57. The last forecast in the sample extends through 2016. The two subsamples were chosen by dividing the full sample roughly in half.

Table 1. RMSE of the *Blue Chip* Consensus's 10-Year Treasury Note Forecasts

Horizon (Years)	1984–2012, N=57	1984–1997, N=28	1998–2012, N=29
1	0.69	0.81	0.56
2	1.22	1.16	1.27
3	1.46	1.11	1.74
4	1.59	0.96	2.02
5	1.8	1.11	2.28

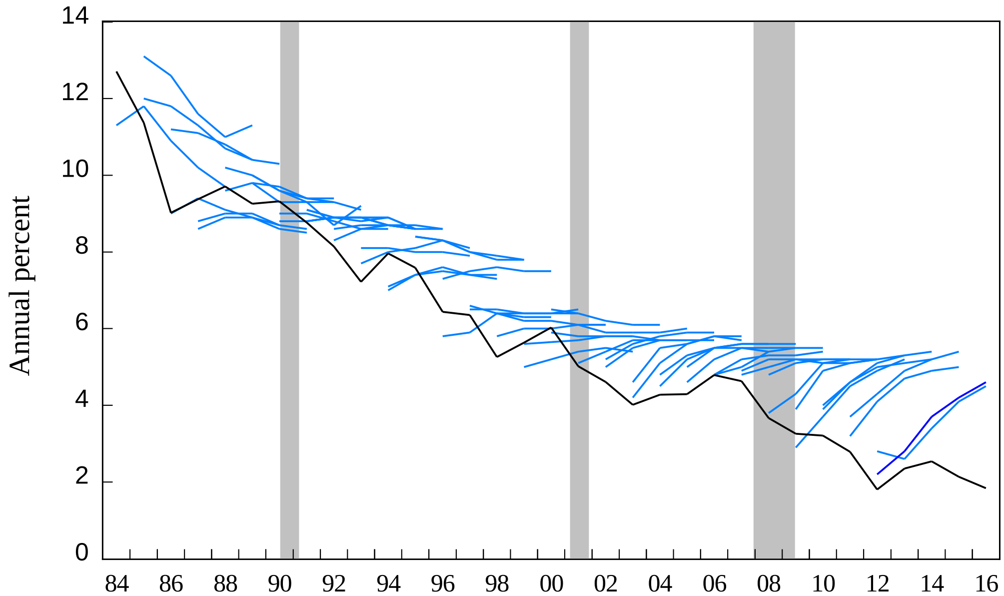
Each cell in the table shows the root mean square error (RMSE) for the forecast horizon listed in the far-left column over the sample shown in the column heading. The RMSE calculations are based on the AAA corporate bond rate from 1984 through 1996 and the 10-year Treasury note rate from 1997 through 2012.

Table 2. Statistical Bias in the *Blue Chip* Consensus's 10-Year Treasury Note Forecasts

Horizon (Years)	1984–2012, N=57	1984–1997, N=28	1998–2012, N=29
1	0.29**	0.23	0.36**
	(0.08)	(0.13)	(0.09)
2	0.85**	0.65**	1.04**
	(0.17)	(0.25)	(0.19)
3	1.18**	0.78**	1.57**
	(0.20)	(0.22)	(0.23)
4	1.33**	0.74**	1.91**
	(0.21)	(0.16)	(0.23)
5	1.55**	0.91**	2.16**
	(0.22)	(0.15)	(0.25)

Each cell in the table shows the average forecast error for the forecast horizon listed in the far-left column calculated over the sample shown in the column heading. Beneath the average error is the Newey-West standard error. Positive numbers indicate that the forecasted value was above the actual value, on average. The asterisks indicate statistical significance corresponding to the null hypothesis that the bias is zero: ** denotes significance at the 0.01 level.

Figure 1. Long-Term Interest Rate Forecasts From the *Blue Chip* Consensus, 1984 to 2012



The figure shows the actual interest rate (the black line) and 57 *Blue Chip* consensus forecasts, published in March and October of each year. Before 1996, the interest rate is the AAA corporate bond rate. After 1996, it is the 10-year Treasury note rate.

The Theoretical Basis for Using the Yield Curve to Forecast Interest Rates

The theoretical basis for using the yield curve to forecast interest rates follows from the expectations hypothesis (EH) of the term structure, which posits that current long-term interest rates equal the average of expected short-term rates (see Cochrane and Piazzesi, 2008). For example, if investors are indifferent between purchasing a 2-period bond with a log yield of $y_t^{(2)}$ and purchasing a sequence of two 1-period bonds with log yields of $y_t^{(1)}$and $y_{t+1}^{(1)}$, then the EH, which assumes an absence of arbitrage opportunities in the bond market, implies:[5]

$$y_t^{(2)} = \frac{y_t^{(1)} + \ {}_t y_{t+1}^{(1)}}{2} \qquad (1)$$

[5]The finance literature typically expresses the expectations hypothesis in terms of log yields. The log yield on an n-period discount bond can be represented as $y_t^n \equiv -\frac{1}{n} p_t^n$ where p_t^n is the log price of an n-period discount bond.

where ${}_{t}y_{t+1}^{(1)}$ is the expected 1-period log yield for a discount bond in time period *t+1* conditional on time *t* information.

Given current yields for 1- and 2-period bonds (the yield curve), equation (1) can be used to back out an estimate of the bond market's expectation of the yield on 1-period bonds one period in the future: ${}_{t}y_{t+1}^{(1)} = 2y_t^{(2)} - y_t^{(1)}$.

Equation (1) can be generalized to *n*-periods:

$$y_t^n = \frac{y_t^{(1)} + {}_{t}y_{t+1}^{(1)} + \cdots + {}_{t}y_{t+n-1}^{(1)}}{n} \qquad (2)$$

Equation (2) implies that the yield curve consisting of interest rates on *1* through *n*-period bonds can be used to back out estimates of the sequence of expected interest rates on *1*-period bonds from *t+1* through *t+n-1* (called forward rates on *1*-period bonds).

If the EH held exactly, it would be straightforward to derive the expected path of future short-term interest rates implied by the current yield curve. But the EH does not hold exactly; there is a persistent (but time-varying) wedge or term premium between long-term rates and the average of expected short-term rates.[6]

$$y_t^n = \frac{y_t^{(1)} + {}_{t}y_{t+1}^{(1)} + \cdots + {}_{t}y_{t+n-1}^{(1)}}{n} + TP_t^{(n)} \qquad (3)$$

The term premium for a bond with a term to maturity *n* (denoted $TP_t^{(n)}$) is unobserved and reflects some combination of the following factors: compensation for the risk of holding long-term bonds, differences in liquidity across bonds of different maturities, investors' preferences for certain maturities over others, and biased expectations of future short-term interest rates. Researchers have developed various methods to estimate the term premium.

For most of the past five decades, the estimated term premium has been positive and increasing in the term to maturity (*n*).[7] The typical positive term premium is explained by the fact that for a

[6]The failure of the expectations hypothesis to explain movements in interest rates is well established. See, for example, Fama and Bliss (1987) and Campbell and Shiller (1991).
[7]A term premium can be defined with respect to any two securities that differ in their term to maturity. For this analysis, the term premium refers to the difference between the yield on the nominal 10-year Treasury note and the

given change in market interest rates, the prices of long-term securities vary by more than the prices of short-term securities, which implies there is more risk associated with holding long-term securities. The typically positive term premium reflects compensation for that risk and can account for the positive slope, on average, of the yield curve.

Since the 2007–2009 recession, the term premium on the 10-year Treasury note relative to the term premium on the one-month Treasury bill has declined from roughly 2 percentage points to negative 50 basis points. (A basis point is one one-hundredth of a percentage point.) Part of the explanation for that decline is that investors have increased their demand for long-term Treasury securities as a hedge against possible negative shocks to inflation and equity prices. Another contributing factor to the low term premium is the Federal Reserve's large-scale purchases of long-term securities through its quantitative easing programs.[8]

The presence of the term premium does not completely eliminate the possibility of extracting information about future expected movements in interest rates from the current yield curve. Given an estimated term premium ($\widehat{TP}_t^{(n)}$), equation (3) can be used to back out the sequence of expected short-term rates that are consistent with that estimated term premium and the current yield curve. Researchers have developed various methods to estimate the *current* term premium and the expected short-term rates contained in the yield curve.[9] For those methods to be usefully applied to forecasts of interest rates, it is necessary to combine the expected short-term rates with a forecast of the *future* term premium. For the forecast comparisons presented below, CBO generated a forecast of the long-term interest rate by combining the average of the expected short-term rates based on the method developed by Adrian, Crump, and Moench (ACM, 2013) with a forecast of the term premium based on a vector autoregression (VAR).[10]

---

average expected yield on the nominal 28-day Treasury bill over the relevant 10-year period. The estimated term premium has been positive for most of the period from 1963 through 2015 (see Figure A-2 in the appendix).

[8]For further discussion of factors that have recently pushed down the term premium, see Congressional Budget Office, *The Budget and Economic Outlook: 2017 to 2027* (January 2017), www.cbo.gov/publication/52370.

[9]See, for example, Duffie and Kan (1996), Vasicek (1977), Cox, Ingersoll, and Ross (1985 a,b), Longstaff and Schwartz (1992 a,b), and Adrian, Crump, and Moench (2013).

[10] See the following section and the appendix for additional details.

Forecasting Methods

CBO compared the accuracy of the *Blue Chip*'s consensus forecasts of long-term interest rates with the accuracy of forecasts generated from several other methods. Three of the comparison methods incorporate information from the yield curve. The other methods serve as benchmarks for the *Blue Chip* consensus forecast and the forecasts generated from the yield curve. Researchers use benchmark forecasts to represent minimum standards for forecast accuracy. For *Blue Chip* forecasts and the forecasts generated from the yield curve to be of practical use, they should be at least as accurate as the benchmark forecasts.

Yield Curve Methods

CBO considered three forecasting methods based on the yield curve. The first method is a reduced-form vector autoregression based on the first three principal components estimated from the Treasury yield curve. The second and third methods are based on the expectations hypothesis.

Reduced-Form Method Based on the Yield Curve. The first three principal components of the Treasury yield curve roughly measure the level, slope, and curvature of the yield curve (see Litterman and Scheinkmann, 1991, and Cochrane and Piazzesi, 2005). CBO estimated a VAR consisting of the nominal long-term interest rate and the first three principal components: [11,12]

$$X_{t+h} = B(L)X_{t-1} + \varepsilon_t \qquad (4)$$

Where $X_t = \begin{bmatrix} i_t \\ PC1i_t \\ PC2i_t \\ PC3i_t \end{bmatrix}$,

$B(L)$ is the matrix of polynomials in the lag operator, $PCi_t, i = 1, \ldots ,3$ are the principal components, $\varepsilon_t$is the regression residual, h=0,… is the forecast horizon, and i_t is the nominal long-term interest rate.[13,14]

---

[11]CBO generated the yields on Treasury securities with terms to maturity ranging from 1 to 120 months using the parameters from Gurkaynak, Sack, and Wright (2006). Changes in the first principal component explain 98 percent of the variation in the nominal 10-year Treasury note rate over the 1961–2015 period.

[12]This method closely follows that used by Bliss (1997). See the appendix for additional details.

[13]The matrix of polynomials in the lag operator is a compact way of representing the relationship between each of the elements of the X vector and their past values.

The VAR measures past correlations between the long-term interest rate and the level, slope, and curvature of the yield curve, but it is only loosely connected to the expectations hypothesis. For example, a positive correlation between the slope of the yield curve and the long-term interest rate could be capturing some combination of expected increases in future rates and the forecastable part of the term premium (the two parts of the right-hand side of equation (3)). But the VAR method does not attempt to sort out that combination and simply relies on the assumption that past correlations will persist in the future.

Methods Based on the Expectations Hypothesis. As alternatives to the reduced-form approaches, CBO estimated two models based on the expectations hypothesis. The first method assumes the expectations hypothesis holds perfectly with a term premium of zero. Under that assumption, the forecast for the nominal 10-year Treasury note rate for horizon *h* is simply the average of the forecast rates from period *h* to *h+9*. For example, the forecast for the nominal 10-year Treasury note rate *h*-years ahead would be based on equation (2) pushed forward in time to begin in period *t+h*:

$$ {}_{t}y_{t+h}^{10} = \frac{{}_{t}y_{t+h}^{(1)} + {}_{t}y_{t+h+1}^{(1)} + \cdots + {}_{t}y_{t+h+9}^{(1)}}{10} \qquad (5)$$

Forecasts based on that method are labeled EH.

CBO also estimated a structural model of the yield curve based on the expectations hypothesis, modified to include an estimated term premium. This method produces separate estimates of the two terms on the right-hand side of equation (3)—the term premium and the sequence of expected future short-term interest rates. Estimates of the term premium are based on the ACM method.

CBO used the ACM method to separate the *current* term premium from the sequence of short-term interest rates. To produce a forecast of long-term interest rates, however, it is necessary to forecast the sequence of *future* term premiums and then combine those term premiums with

[14]Previous researchers have used similar techniques to characterize and forecast the yield curve. Bliss (1997) estimated a VAR using the first three principal components of the yield curve but focused on the estimated impulse responses rather than forecasting the yield curve. Diebold and Li (2006) estimated a VAR with three factors derived from the yield curve. Although their focus was on forecasting, they used a different technique (Nelson-Siegel) to derive the factors underlying the curve.

average expected future short-term rates from the ACM model. CBO used a vector autoregression to forecast the sequence of future term premiums and then added that sequence of forecasted term premiums to the average of the expected future short-term rates derived using the ACM method to arrive at a forecast for the nominal long-term interest rate.[15] The forecasts based on that method are labeled EHTP.

Benchmark Methods

CBO compared the accuracy of the interest rate forecasts from the *Blue Chip* consensus and from the three methods based on the yield curve with three different benchmark methods: a random walk (RW), an autoregressive moving average (ARMA) model of the long-term interest rate, and a simple linear forecast rule using the *Blue Chip*'s forecast of growth rates of nominal gross domestic product (GDP).

Random Walk. The simplest and most naïve forecast is the random walk forecast. The random walk forecast of the nominal interest rate at horizon $t+h$ is simply the value of the interest rate observed at date t. That benchmark serves as a minimum standard for forecast accuracy. To the extent that economic theory has value in forecasting, the forecasts based on theory should outperform the random walk forecast.

Autoregressive Moving Average. A slightly more sophisticated but still naïve method is the autoregressive moving average model. Simple ARMA models are often used as benchmarks in forecast comparisons because they are statistical, not economic, models that are based purely on the historical behavior of the series. The main difference between the random walk forecast and the ARMA forecast is that the ARMA forecast accounts for short-term persistence in the data being forecasted. As with the random walk forecast, ARMA forecasts are considered a minimum standard for accuracy.[16]

[15] CBO uses the curve consisting of yields on Treasury securities with terms to maturity ranging from one month to 10 years to estimate the sequence of short-term interest rates over the next decade. However, forecasts of the 10-year Treasury note rate require estimates of short-term interest rates for years 11 and beyond. For example, the forecast of the nominal 10-year Treasury note rate one year ahead requires a forecast of the short-term interest rates from one year ahead through 11 years ahead. To forecast the short-term interest rates for periods beyond the 10-year window, CBO uses an autoregression.

[16] The specific model chosen for those forecasts was an MA(2) based on in-sample fit over the monthly sample of nominal 10-year Treasury note rates between 1961 and 1984.

Nominal GDP Growth. The final benchmark is based on the *Blue Chip*'s forecast of the growth of nominal GDP. In a standard neoclassical growth model, the real (inflation-adjusted) interest rate roughly corresponds to the growth rate of real GDP.[17] By adding inflation, the standard neoclassical growth model also implies a rough correspondence between nominal interest rates and the growth of nominal GDP. Because the nominal GDP growth forecast is regularly reported as part of the *Blue Chip* consensus's semiannual forecast, it serves as a check on the accuracy of the rough correspondence implied by the neoclassical growth model.

A Comparison of Forecast Accuracy

CBO compared the accuracy of the *Blue Chip's* forecasts of the average annual nominal long-term interest rate, for horizons one through five years, over the full sample period and two subsample periods. The full sample comprised 57 semiannual *Blue Chip* forecasts published between March 1984 and March 2012. The subsample periods were 1984 through 1997 and 1998 through 2012.

The comparison forecasts based on the yield-curve methods were constructed as if in real time using only the data that were available when the *Blue Chip* produced each of its semiannual forecasts. All of the models were estimated using monthly observations from June 1961 through the relevant forecast jump-off date (February of each year for the *Blue Chip* forecast produced in March and September of each year for the *Blue Chip* forecast produced in October). The monthly forecasts were then averaged to produce annual averages. CBO produced the ARMA (and trivially the RW) forecasts in the same manner.

For the full sample of 57 forecasts, the *Blue Chip* consensus was generally as accurate (or more accurate) than the forecasts produced by the yield-curve methods—VAR, EH, and EHTP (see **Table 3**).[18] Compared with the benchmark methods, the *Blue Chip* consensus forecasts were also

[17]The relationship between the real interest rate and real GDP growth in the neoclassical growth model is approximate. In a standard Solow growth model, the real interest rate is equal to $\frac{\alpha}{s}(n+g+\delta)$ where α is capital's share, s is the fixed saving rate, n is the growth of labor input, g is the growth rate of labor productivity, and δ is the depreciation rate. The growth rate of real GDP in that model is simply $n+g.$ In a standard optimal growth model with endogenous saving, $r=\frac{1}{\sigma}g+n+\theta$ where σ is the intertemporal elasticity of substitution in consumption and θ is the rate of time preference. The usefulness of nominal GDP as an input to forecasting nominal interest rates depends on the accuracy of those approximations as well as the accuracy of the forecast of nominal GDP growth.
[18]CBO evaluated relative forecast accuracy using the modified Diebold-Mariano (DM) test statistic. That statistic is the t-test on the difference in the squared errors of the two forecasts (see Diebold and Mariano, 1995). The modified

consistently more accurate than those based on the ARMA model. The RW and ARMA models produced smaller RMSEs than the *Blue Chip* consensus at the longer forecast horizons (years 3 through 5), but the differences were not statistically significant. Using the *Blue Chip* consensus forecast of nominal GDP growth in place of the interest rate forecast would have produced a (significantly) smaller RMSE at the 5-year-ahead horizon.

Table 3. RMSEs of Seminannual Forecasts, 1984 to 2012

		Yield-Curve Models			Benchmark Models		
Forecast Horizon	*Blue Chip*	VAR	EH	EHTP	RW	ARMA	GDP
1	0.69	0.71	0.82	0.87	0.74	0.72	1.74
2	1.22	1.3	1.47	1.51	1.25	1.22	1.44
3	1.46	1.6	1.75	1.74	1.36	1.35	1.48
4	1.59	1.87	1.98	1.96	1.44	1.48	1.49
5	1.8	2.18	2.26	2.20	1.55	1.67	1.56**

Sample size = 57, ** significantly smaller than *Blue Chip*'s RMSE at the 0.05 level based on two-tailed tests using the student's-t distribution (see Harvey, Leybourne, and Newbold, 1997).

Forecasters were largely surprised by the persistent downward movement in long-term interest rates throughout the 1984–2012 period, especially toward the latter half of the period.[19] CBO tested whether the relative accuracy of the various forecasting methods changed over the two halves of the sample. Some of the models based on the yield curve produced more accurate forecasts in the later sample than in the earlier sample (for the EH model at all horizons and the VAR and EHTP models at shorter horizons), CBO found. But even with those improvements, none of those models produced significantly more accurate forecasts than the *Blue Chip*

DM test (see Harvey, Leybourne, and Newbold, 1997) corrects for small sample bias by multiplying the DM test statistic by $\sqrt{\frac{T+1-2h+h(h-1)}{T}}$, where T is the sample size and h is the forecast horizon.

[19]CBO recently updated its analysis of its forecasting record as well as the forecasting records of the *Blue Chip* consensus and the Administration (as published in the annual budget documents prepared by the Office of Management and Budget). That analysis found that across all three forecasters, interest rate forecasts have tended to be more biased and less accurate than forecasts of other economic variables. See Congressional Budget Office, *CBO's Economic Forecasting Record: 2017 Update* (October 2017), www.cbo.gov/publication/53090.

consensus over the 1998–2012 period (see **Table 4** and **Figure 2**). Among the benchmark forecasts, however, the ARMA and RW methods produced significantly more accurate forecasts than the *Blue Chip* consensus at all horizons over the later sample.

The RMSE results indicate that models based on the yield curve would not have produced significantly more accurate forecasts in real time over the full sample and over the two subsamples examined in this analysis. The finding that the benchmark methods produced more accurate forecasts (especially over the later sample) reflects the fact that benchmark forecasts temper the rise in interest rates (or, in the case of the random walk, hold interest rates constant) relative to the yield-curve methods. In an environment with persistently falling interest rates, methods that forecast less of a rise in interest rates or no rise at all will outperform methods that assume that interest rates will rise back toward historical averages.[20]

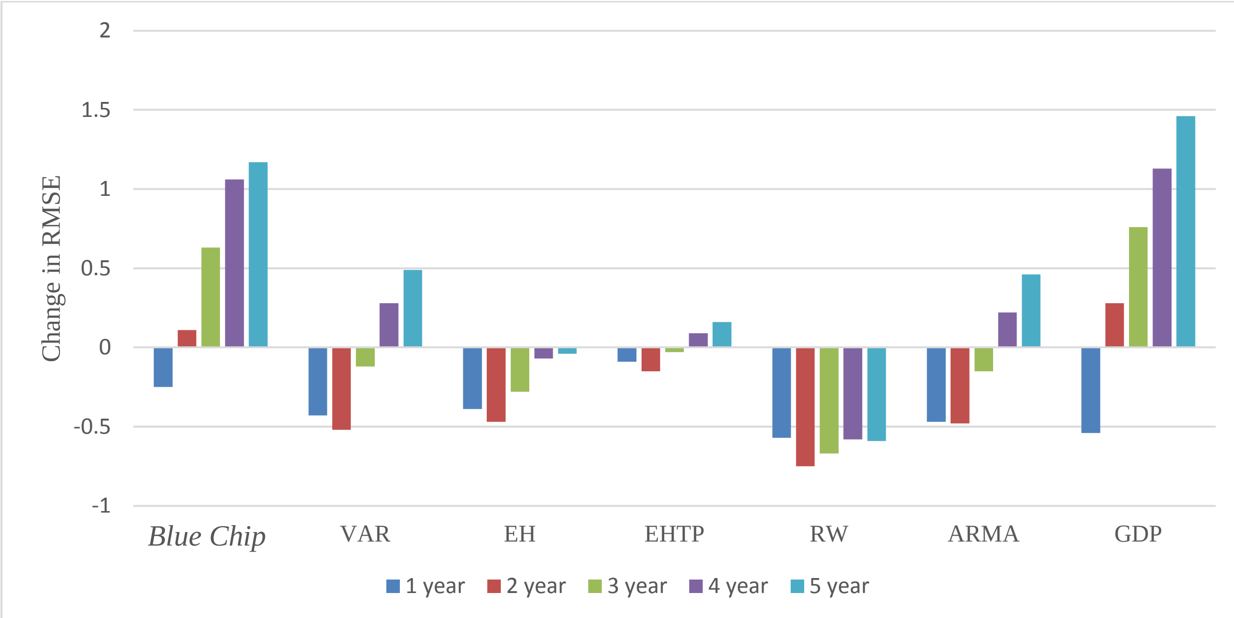
[20]An economic data series is said to be stationary if it exhibits a constant mean and constant covariance structure throughout time. Many of the models examined in this study (such as the VAR, EHTP, and ARMA models) perform poorly when the data they are applied to are nonstationary because those models are constructed to produce forecasts that revert back to the historical mean of the series. The average nominal 10-year Treasury note rate declined steadily throughout the period examined here. Thus, any model that assumes stationarity when used to forecast the nominal 10-year Treasury note rate will fail to anticipate changes in the mean of the series.

Table 4. RMSEs of Seminannual Forecasts for Subsamples 1984–1997 and 1998–2012

			Yield-Curve Models			Benchmark Models		
Forecast Horizon	Sample Period	*Blue Chip*	VAR	EH	EHTP	RW	ARMA	GDP
1	1984–1997	0.81	0.89	1	0.92	0.97	0.92	2
	1998–2012	**0.56**	**0.46**ǂ	**0.61**	**0.83**	**0.40****	**0.45****	**1.46**
2	1984–1997	1.16	1.54	1.69	1.58	1.57	1.44	1.29
	1998–2012	1.27	**1.02**	**1.22**	**1.43**	**0.82***	**0.96***	**1.57**
3	1984–1997	1.11	1.66	1.89	1.75	1.66	1.42	1.05
	1998–2012	1.74	**1.54**	**1.61**	1.72	**0.99****	**1.27***	1.81
4	1984–1997	0.96	1.73	2.02	1.91	1.7	1.36	0.81
	1998–2012	2.02	2.01	**1.95**	2.00	**1.12****	1.58*	1.94
5	1984–1997	1.11	1.92	2.28	2.12	1.83	1.42	0.64**
	1998–2012	2.28	2.41	**2.24**	2.28	**1.24****	1.88*	2.10**

Sample sizes were 28 for the 1984–1997 subsample and 29 for the 1998–2012 subsample. Instances in which the root mean square error (RMSE) was smaller in the 1998–2012 subsample than in the 1984–1997 subsample are shown in bold. Significance levels are reported for the RMSEs that are significantly smaller than the *Blue Chip* consensus's RMSE and are based on two-tailed tests using student's-t distribution (see Harvey, Leybourne, and Newbold, 1997); ǂ denotes significance at the .10 level, * denotes significance at the 0.05 level, ** denotes significance at the 0.01 level.

Figure 2. Change in RMSE, 1984–1997 to 1998–2012



Each bar represents the difference between the root mean square error (RMSE) in the earlier subsample (1984 to 1997) and the later subsample (1998 to 2012) for a particular forecast and horizon.

The bias estimates produced by the various forecasting methods followed a pattern similar to that of the RMSE estimates (see **Table 5**). The VAR, ARMA, and GDP methods produced slightly less bias than the *Blue Chip* consensus at all horizons in the later sample period (1998 to 2012). The EH also produced slightly smaller bias than the *Blue Chip* consensus over the later sample, but only for horizons 2 through 5 years ahead. The RW forecasts were substantially less biased than those of the *Blue Chip* consensus in both the full sample and (especially) the later sample. In some cases, the bias produced by the RW forecasts was less than one-third the size of the bias contained in the *Blue Chip* consensus forecast.

Table 5. Statistical Bias

Horizon	Sample	*Blue Chip*	VAR	EH	EHTP	RW	ARMA	GDP
1	1984–2012	0.29**	0.20*	0.44**	0.54**	0.15	0.19*	-0.75*
		(0.08)	(0.10)	(0.09)	(0.1)	(0.10)	(0.09)	(0.34)
	1984–1997	0.23‡	0.24	0.48**	0.45**	0.25	0.21	-1.76**
		(0.13)	(0.18)	(0.16)	(0.15)	(0.18)	(0.17)	(0.21)
	1998–2012	0.36**	0.16‡	0.4**	0.63**	0.06	0.18‡	0.23
		(0.09)	(0.09)	(0.11)	(0.14)	(0.09)	(0.09)	(0.37)
2	1984–2012	0.85**	0.69**	1**	1.11**	0.48*	.58**	-0.02
		(0.17)	(0.20)	(0.19)	(0.18)	(0.21)	(0.19)	(0.35)
	1984–1997	0.65*	0.72‡	1.1**	1.03**	0.68‡	0.57	-1.07**
		(0.25)	(0.36)	(0.33)	(0.3)	(0.37)	(0.34)	(0.22)
	1998–2012	1.04**	0.66**	0.91**	1.19**	0.27	0.59**	0.99*
		(0.19)	(0.21)	(0.22)	(0.21)	(0.19)	(0.19)	(0.38)
3	1984–2012	1.18**	1.12**	1.41**	1.47**	0.71**	0.87*	0.38
		(0.20)	(0.24)	(0.2)	(0.19)	(0.23)	(0.21)	(0.36)
	1984–1997	0.78**	1.04*	1.46**	1.36**	0.92*	0.74**	-0.71**
		(0.22)	(0.38)	(0.32)	(0.32)	(0.40)	(0.35)	(0.24)
	1998–2012	1.57**	1.20**	1.37**	1.57**	0.5*	1.00**	1.42**
		(0.23)	(0.29)	(0.24)	(0.19)	(0.24)	(0.22)	(0.37)
4	1984–2012	1.33**	1.53**	1.76**	1.75**	0.91**	1.13**	0.66‡
		(0.21)	(0.23)	(0.17)	(0.17)	(0.22)	(0.20)	(0.34)
	1984–1997	0.74**	1.31**	1.73**	1.6**	1.10**	0.86**	-0.39**
		(0.16)	(0.32)	(0.25)	(0.3)	(0.35)	(0.28)	(0.2)
	1998–2012	1.91**	1.74**	1.79**	1.91**	0.74**	1.40**	1.67**
		(0.23)	(0.31)	(0.22)	(0.17)	(0.25)	(0.23)	(0.33)
5	1984–2012	1.55**	1.95**	2.1**	2.05**	1.15**	1.42**	0.89**
		(0.22)	(0.21)	(0.14)	(0.16)	(0.20)	(0.18)	(0.32)
	1984–1997	0.91**	1.67**	2.07**	1.9**	1.37**	1.07**	-0.13
		(0.15)	(0.25)	(0.2)	(0.26)	(0.31)	(0.22)	(0.15)
	1998–2012	2.16**	2.23**	2.13**	2.2**	0.94**	1.75**	1.88**
		(0.25)	(0.30)	(0.2)	(0.18)	(0.24)	(0.22)	(0.32)

Each cell in the table shows the average forecast error for the forecast horizon and sample period listed in the far left columns. Beneath the average error is the Newey-West standard error. Positive numbers indicate that the forecasted value was higher than the actual value, on average. The asterisks indicate statistical significance corresponding to the null hypothesis that the bias is zero: ‡ denotes significance at the .10 level, * denotes significance at the 0.05 level, and ** denotes significance at the 0.01 level.

Does the Yield Curve Contain Information That Would Have Improved the *Blue Chip*'s Forecasts?

In the 1998–2012 subsample, a forecast based on the random walk (that is, assuming a constant interest rate over the entire five-year horizon) would have produced forecasts that were more accurate and less biased than those of the *Blue Chip* consensus. It is still possible, however, that forecasts based on the yield curve would provide information that could be used to improve on the *Blue Chip*. In particular, if the *Blue Chip* consensus forecasts and the forecasts based on the yield curve captured different (complementary rather than redundant) information about future interest rates, then combining those forecasts could lead to improvements in forecast accuracy compared with using a single "best" forecast.[21]

To investigate that possibility, CBO measured the accuracy of combined forecasts that were constructed by taking the simple average of pairs of forecasts analyzed here (see **Table 6**).[22] Over the full sample and the 1984–1997 subsample period, several of the combination forecasts (shown in bold) had lower RMSEs than the *Blue Chip* consensus alone. In two instances—the *Blue Chip* consensus combined with nominal GDP growth at horizons 4 and 5 years—the RMSE was significantly lower than the corresponding RMSE for the *Blue Chip* consensus forecast. When the *Blue Chip* consensus forecast was combined with the ARMA or RW forecast over the second subsample period, the resulting RMSE was lower than the RMSE from the *Blue Chip* forecast. But even in those instances, the RMSEs for the combined forecasts were larger than the RMSEs for the individual (ARMA or RW) forecasts.

---

[21]See Timmerman (2006) for a review of the literature on combining forecasts.

[22]An alternative to averaging is to combine forecasts using weights estimated by regressing actual outcomes on past forecasts. CBO chose the simple average instead of estimated weights, for two reasons. First, using weights estimated from the sample of forecasts would be inconsistent with the real-time nature of the other forecasts compared in this analysis. A forecaster working in real time would not have had those estimated weights when each forecast was made. Second, estimated weights are often imprecise (because of sampling error, especially for small samples), and combining forecasts using a simple average typically results in a forecast that is as good or better than one made using estimated weights [see Clemen (1989), Stock and Watson (2001, 2004), and Timmerman (2006)]. In practice, the gains from estimating the weights are typically not large enough (relative to simple averaging) to outweigh the estimation error. Elliott (2011) explores that trade-off and provides theoretical bounds on the size of the gain from estimating the weights when combining forecasts.

Table 6. RMSEs of Combined Forecasts for the Full Sample and Subsamples

Horizon	Sample Period	*Blue Chip*	VAR	EH	EHTP	RW	ARMA	GDP
1	1984–2012	0.69	**0.66**	**0.68**	0.71	**0.67**	**0.67**	1.15
	1984–1997	0.81	0.81	**0.78**	**0.74**	0.83	0.82	1.36
	1998–2012	0.56	**0.47**	0.57	0.68	**0.45****	**0.49****	0.91
2	1984–2012	1.22	**1.11**	**1.20**	1.23	**1.07**	**1.10**	**1.18**
	1984–1997	1.16	**1.14**	1.17	**1.09**	1.17	**1.12**	**0.92**
	1998–2012	1.27	**1.08**	**1.23**	1.34	**0.97****	**1.09****	1.38
3	1984–2012	1.46	**1.38**	1.46	1.46	**1.22**	**1.30**	**1.37**
	1984–1997	1.11	1.16	1.24	1.14	1.17	**1.08**	**0.80**
	1998–2012	1.74	**1.57**	**1.66**	**1.72**	**1.27****	**1.48****	1.75
4	1984–2012	1.59	1.57	1.62	1.61	**1.29**	1.42	**1.47****
	1984–1997	0.96	1.04	1.14	1.06	1.04	**0.90**	**0.64****
	1998–2012	2.02	**1.96**	**1.97**	**2.00**	**1.48****	**1.78***	**1.97**
5	1984–2012	1.8	1.84	1.85	1.83	**1.45**	**1.62**	**1.61****
	1984–1997	1.11	1.18	1.32	1.22	1.15	**0.98**	**0.62****
	1998–2012	2.28	2.30	**2.25**	**2.26**	**1.69****	**2.06****	**2.18****

Sample sizes were 57 for the 1984–2012 sample, 28 for the 1984–1997 subsample, and 29 for the 1998–2012 subsample. The column labeled *Blue Chip* reports the root mean square error (RMSE) for the *Blue Chip* consensus forecast of the nominal 10-year Treasury note rate. The remaining columns report the RMSE resulting from combining the *Blue Chip* consensus forecast with the forecast listed in the column heading using a simple average. Instances in which the RMSE for the combined forecasts was smaller than the RMSE for the *Blue Chip* consensus forecast are shown in bold. Significance levels are reported for the RMSEs that are significantly smaller than the *Blue Chip* consensus's RMSE and are based on two-tailed tests using the student's-t distribution (see Harvey, Leybourne, and Newbold, 1997); ǂ denotes significance at the 0.10 level, *denotes significance at the 0.05 level, **denotes significance at the 0.01 level.

In several instances, combining forecasts reduced the bias relative to the individual forecast results reported in Table 5 (see **Table 7**). For all horizons and forecasts, bias was positive in the later sample. Except for the forecasts based on the combination of the *Blue Chip* consensus and nominal GDP growth at the one-year horizon, bias was positive and statistically significant in the later sample for all horizons and forecasts. Over the subsample 1997–2012, the biases in the random walk forecasts (Table 5) were smaller than the biases reported for all of the combination forecasts at each horizon.

Table 7. Statistical Bias in Combined Forecasts

Horizon	Sample Period	*Blue Chip*	VAR	EH	EHTP	RW	ARMA	GDP
1	1984–2012	0.29**	0.24**	0.37**	0.42**	0.22**	0.24**	-0.23
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.19)
	1984–1997	0.23ǂ	0.23	0.35*	0.33*	0.23	0.21	-0.77**
		(0.13)	(0.15)	(0.14)	(0.13)	(0.15)	(0.15)	(0.13)
	1998–2012	0.36**	0.26**	0.38**	0.50**	0.21**	0.27**	0.30
		(0.09)	(0.07)	(0.09)	(0.11)	(0.08)	(0.09)	(0.21)
2	1984–2012	0.85**	0.76**	0.92**	0.98**	0.66**	0.71**	0.41ǂ
		(0.17)	(0.16)	(0.17)	(0.16)	(0.17)	(0.17)	(0.24)
	1984–1997	0.65*	0.67*	0.86**	0.83**	0.66*	0.60ǂ	-0.22
		(0.25)	(0.29)	(0.28)	(0.25)	(0.30)	(0.28)	(0.19)
	1998–2012	1.04**	0.85**	0.98**	1.12**	0.66**	0.82**	1.02**
		(0.19)	(0.16)	(0.20)	(0.20)	(0.18)	(0.18)	(0.28)
3	1984–2012	1.18**	1.14**	1.29**	1.31**	0.93**	1.02**	0.77**
		(0.20)	(0.19)	(0.18)	(0.18)	(0.18)	(0.19)	(0.27)
	1984–1997	0.78**	0.89**	1.10**	1.05**	0.83**	0.74**	-0.01
		(0.22)	(0.28)	(0.25)	(0.24)	(0.29)	(0.26)	(0.20)
	1998–2012	1.57**	1.39**	1.47**	1.57**	1.04**	1.28**	1.50**
		(0.23)	(0.21)	(0.23)	(0.20)	(0.21)	(0.21)	(0.29)
4	1984–2012	1.33**	1.42**	1.54**	1.53**	1.11**	1.22**	0.99**
		(0.21)	(0.19)	(0.17)	(0.17)	(0.17)	(0.19)	(0.27)
	1984–1997	0.74**	1.01**	1.22**	1.15**	0.90**	0.78**	0.16
		(0.16)	(0.22)	(0.19)	(0.21)	(0.23)	(0.21)	(0.17)
	1998–2012	1.91**	1.83**	1.85**	1.91**	1.32**	1.65**	1.79**
		(0.23)	(0.22)	(0.21)	(0.19)	(0.21)	(0.21)	(0.28)
5	1984–2012	1.55**	1.74**	1.82**	1.79**	1.34**	1.47**	1.21**
		(0.22)	(0.19)	(0.16)	(0.17)	(0.16)	(0.19)	(0.27)
	1984–1997	0.91**	1.27**	1.47**	1.39**	1.12**	0.98**	0.37*
		(0.15)	(0.18)	(0.16)	(0.18)	(0.21)	(0.17)	(0.14)
	1998–2012	2.16**	2.20**	2.15**	2.18**	1.55**	1.95**	2.02**
		(0.25)	(0.22)	(0.22)	(0.20)	(0.21)	(0.22)	(0.28)

Sample sizes were 57 for the 1984–2012 sample, 28 for the 1984–1997 subsample, and 29 for the 1998–2012 subsample. The column labeled *Blue Chip* reports the average forecast error for the *Blue Chip* consensus forecast of the nominal 10-year Treasury note rate. The remaining columns report the average forecast error resulting from combining the *Blue Chip* consensus forecast with the forecast listed in the column heading using a simple average. Beneath the average error is the Newey-West standard error. Positive numbers indicate that the forecasted value was higher than the actual value, on average. The asterisks indicate statistical significance corresponding to the null hypothesis that the bias is zero: ǂ denotes significance at the 0.10 level, * denotes significance at the 0.05 level, ** denotes significance at the 0.01 level.

Conclusion and Next Steps

This analysis has two main findings. The first is that compared with the *Blue Chip* consensus, Treasury debt markets did not forecast with greater accuracy the decline in long-term interest rates over the past three decades. Over the full sample, none of the methods based on the yield curve and none of the benchmark methods produced more accurate forecasts than the *Blue Chip* consensus. In the context of the yield-curve models examined in this analysis, the answer to the question of whether bond market participants did a better job anticipating the persistent decline in long-term interest rates is "no."

The second main finding is that the forecasts based on the random walk model of interest rates were in general more accurate and less biased than those of the *Blue Chip* consensus—especially over the most recent subsample period of 1998 to 2012. Although forecasters are unlikely to abandon their structural models in favor of the naïve RW forecast, those results suggest that forecasters should reexamine those models to determine which factors have been pushing interest rates back up toward historical levels (and away from the RW forecast) in recent years. One possible source of bias in those models is the term premium, which some analysts have argued is likely to remain suppressed relative to its historical average.[23]

Although the yield-curve-based models examined in this analysis did not outperform the *Blue Chip* consensus, there are other possibilities to explore. In particular, the accuracy of the forecasts produced by the EHTP method depends critically on the accuracy of the forecasts of the term premium. The method employed here relies on past yield curves to forecast the term premium, but there is some debate in the finance literature about whether the information set used to forecast the term premium should contain variables other than the yield curve (such as real GDP growth and inflation).[24] One avenue to explore in future research is whether the forecasts produced from the EHTP model can be improved by including such macroeconomic information.[25]

And, finally, it is important to note that the forecast evaluations in this analysis covered a period when long-term interest rates trended downward, on average. It is not necessarily the case that

[23]That argument is partly based on the increased hedging properties of long-term Treasury securities. Before the early 2000s, the correlation between the price of equities (stocks) and the price of Treasury securities (bonds) tended to be positive. Since that time, the correlation has been negative, which implies that Treasury bonds are now a better hedge against movements in equity prices. The increased desirability to hold Treasury securities as a hedge might be contributing to the low term premium. Moreover, to the extent that well-anchored inflation expectations have contributed to the negative correlation between stock and bond prices (because few inflation surprises would cause both prices to move in the same direction), the low term premium is likely to persist.
[24]See Bauer and Rudebusch (2015) and Bauer and Hamilton (2017).
[25]One downside of such a method is that it would require forecasts of the macroeconomic inputs.

the most accurate forecast models in a period of persistently falling interest rates will also be the most accurate when interest rates are no longer consistently falling.

References

Adrian, Tobias, Richard Crump, and Emanuel Moench, “Pricing the Term Structure With Linear Regressions,” *Journal of Financial Economics*, vol. 110, no. 1 (2013), pp. 11–138.

Bauer, Andy, and others, “Forecast Evaluation With Cross-Sectional Data: The *Blue Chip* Surveys,” *Economic Review*, vol. 88, no. 2 (Federal Reserve Bank of Atlanta, 2003) pp. 17–31.

Bauer, Michael D., and James D. Hamilton, *Robust Bond Risk Premia*, NBER Working Paper 23480 (National Bureau of Economic Research, June 2017).

Bauer, Michael D., and Glenn D. Rudebusch, *Resolving the Spanning Puzzle in Macro-Finance Term Structure Models*, Working Paper 2015-01 (Federal Reserve Bank of San Francisco, 2015).

Bliss, Robert R., “Movements in the Term Structure of Interest Rates,” *Economic Review* (Federal Reserve Bank of Atlanta, Fourth Quarter, 1997), pp. 16–33.

Campbell, John Y., and Robert J. Shiller, “Yield Spreads and Interest Rate Movements: A Bird’s Eye View,” *Review of Economic Studies*, vol. 58 (1991), pp. 495–514.

Clemen, Robert T., “Combining Forecasts: A Review and Annotated Bibliography,” *International Journal of Forecasting*, vol. 5 (1989), pp. 559–583.

Cochrane, John H., and Monika Piazzesi, “Bond Risk Premia,” *American Economic Review*, vol. 95, no. 3 (March 2005), pp. 136–160.

Cochrane, John H., and Monika Piazzesi, “Decomposing the Yield Curve” (draft, University of Chicago, 2008).

Cox, J. C., J. E. Ingersoll, and S. A. Ross, “An Intertemporal General Equilibrium Model of Asset Prices, *Econometrica*, vol. 53 (1985a), pp. 363–384.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, “A Theory of Term Structure of Interest Rates,” *Econometrica*, vol. 53 (1985b), pp. 385–407.

Diebold, Francis X., and Canlin Li, “Forecasting the Term Structure of Government Bond Yields,” *Journal of Econometrics*, vol. 130 (2006), pp. 337–364.

Diebold, Francis X., and Robert S. Mariano, “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, vol. 13 (1995), pp. 253–263.

Diebold, Francis X,. and Glenn D. Rudebusch, *Yield Curve Modeling and Forecasting* (Princeton University Press, 2013).

Duffie, D., and R. Kan, “A Yield-Factor Model of Interest Rates, “*Mathematical Finance*, vol. 6 (1996), pp. 379–406.

Elliott, G., *Averaging and the Optimal Combination of Forecasts* (University of California, San Diego, 2011).

Fama, Eugene F., and Robert R. Bliss, “The Information in Long-Maturity Forward Rates,” *American Economic Review*, vol. 77 (1987), pp. 680–692.

Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright, *The U.S. Treasury Yield Curve: 1961 to the Present*, Working Paper 2006-28, Finance and Economics Discussion Series (Federal Reserve Board, Washington, D.C., 2006).

Harvey, D., S. Leybourne, and P. Newbold, “Testing the Equality of Prediction Mean Squared Errors,” *International Journal of Forecasting*, vol. 13 (1997), pp. 281–291.

Litterman, R., and J. Scheinkmann, “Common Factors Affecting Bond Returns, *Journal of Fixed Income (*June 1991), pp. 54–61.

Longstaff, F. A., and E. S. Schwartz, “Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model,” *Journal of Finance*, vol. 67 (1992a), pp. 1259–1282.

Longstaff, F. A., and E. S. Schwartz, “A Two-Factor Interest Rate Model and Contingent Claims Evaluation,” *Journal of Fixed Income* (1992b), pp. 16–23.

Rachel, Lukasz, and Thomas D. Smith, “Secular Drivers of the Global Real Interest Rate,” Staff Working Paper 571 (Bank of England, December 2015).

Stock, J. H., and M. W. Watson, “A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series,” in R. F. Engle, and H. White, eds., *Festschrift in Honour of Clive Granger* (Cambridge University Press, 2001), pp. 1–44.

Stock, J. H., and M. W. Watson, “Combination Forecasts of Output Growth in a Seven-Country Data Set,” *Journal of Forecasting*, vol. 23 (2004), pp. 405−430.

Timmerman, Allan, “Forecast Combination,” in Graham Elliott, Clive W.J. Granger, and Timmermann, eds., *Handbook of Economic Forecasting* (2006), pp. 136–196.

Townsend, Henry, “A Comparison of Several Consensus Forecasts, 1996,” *Business Economics*, vol. 31, no. 1 (January 1996).

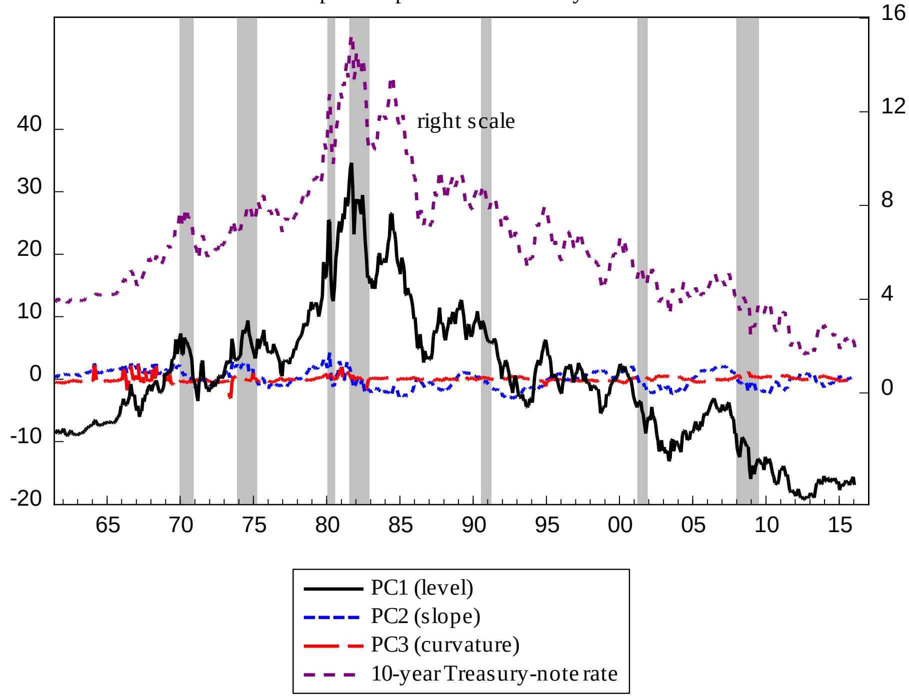
Vasicek, O., “An Equilibrium Characterization of the Term Structure,” *Journal of Financial Economics*, vol. 5 (1977), pp. 177–188.

Appendix: Yield Curve Models

The Congressional Budget Office analyzed the accuracy of forecasts based on three methods that rely on information from the Treasury yield curve. The first method (the vector autoregression, or VAR) is based on the first three principal components estimated from the monthly Treasury yield curves. CBO generated the yields on Treasury securities with terms to maturity ranging from 1 to 120 months using the parameters from Gurkaynak, Sack, and Wright (GSW, 2006).[26]

The macroeconomics finance literature interprets the first three principal components of the yield curve as measures of its level, slope, and curvature at each point in time (see Cochrane and Piazzesi, 2005). Figure A-1 shows CBO's estimates of the first three principal components using the full sample (1961–2015), along with the nominal 10-year Treasury note rate (right scale). Changes in the first principal component explain 98 percent of the variation in the nominal 10-year Treasury note rate over the 1961–2015 period.

Figure A-1. 10-Year Treasury Note Rate
and Principal Components of Monthly Yield Curves



The second method is based on the expectations hypothesis (EH). The expectations hypothesis assumes that investors are indifferent between holding an *n*-period bond to maturity and holding a sequence of one-period bonds in each of the next *n* periods. For example, suppose an investor

[26] For the full set of yield curve parameters, see Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright, *The U.S. Treasury Yield Curve: 1961 to the Present*, Working Paper 2006-28, Finance and Economics Discussion Series (Federal Reserve Board, Washington, D.C., 2006), https://go.usa.gov/xRE94.

wishes to purchase a bond to hold for one month. That investor faces a choice: either purchasing a long-term bond at the beginning of the month and selling it at the end of the month or purchasing a bond with a term to maturity of one month. If investors in general are indifferent between those two options, then buying and selling in the markets for short- and long-term bonds will cause the expected returns for both options to equalize.

CBO constructed the EH forecasts by first calculating the sequence of short-term (one-period) rates implied by the yield curve and then averaging those short-term rates over 10 years as described by equation (2). The EH method assumes that there is no term premium. (The term premium is the additional return that investors require as compensation for holding a long-term bond instead of a one-period bond.)

The third method considered in this analysis is based on the expectations hypothesis of interest rates modified to include an estimated term premium.

As noted in the text, the expected returns on long-term bonds held for one period and one-period bonds tend not to equalize—there is a time-varying term premium. Researchers estimate the time-varying term premium by estimating the expected difference between holding an *n*-period bond for one period and holding a one-period bond. That expected difference, which is called the expected excess holding period return, is a measure of the term premium.

Estimating the term premium involves three steps.

- Step 1 is to estimate the first five principal components of the monthly yield curves generated using the GSW parameters as described above.[27]
- Step 2 is to model the principal components using a VAR (with lag length = 1) and then regress the realized excess holding period returns on the fitted values and the residuals from the VAR to obtain separate estimates of the expected and unexpected excess holding period returns.
- Step 3 is to use the cross-section of excess holding period returns (by term to maturity) to estimate the price of (term) risk and then impose a zero-arbitrage condition on the estimates of the expected excess returns and the price of risk to back out estimates of the expected short-term (risk-free) rates and the term premium.

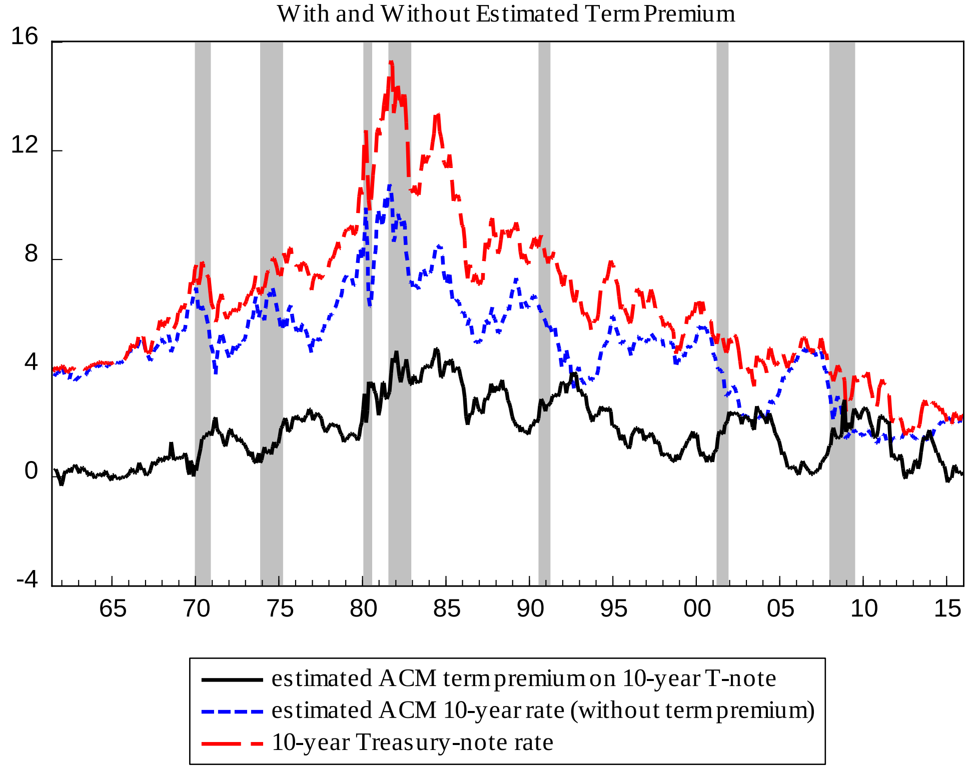
For the forecast comparisons in the main text, CBO using the data available at each date that *Blue Chip* completed its forecasts. Using the entire sample, the resulting estimates of the term premium, the 10-year Treasury note rate without the term premium, and the actual 10-year Treasury note rate are presented in Figure A-2, below.

To produce a forecast based on the Adrian, Crump, and Moench (ACM) model, CBO replaced the *current* term premium estimated at each date with a *forecast* of the term premium and then

[27]CBO chose the five-factor model because that was the benchmark model estimated in Adrian, Crump, and Moench (2013).

combined that forecast with the average of the short-term interest rates implied by the current yield curve. The ACM model produced estimates of the short-term (risk-free) interest rates over the period starting with the forecast jump-off date (February and September of each year in 1984 through 2012) and running through the 10-year period after that jump-off date. Thus, to forecast the 10-year Treasury note rate for periods beyond the current jump-off date requires estimates of both the short-term (risk-free) rates and the term premium beyond the 10-year window. CBO used a simple autoregressive model to forecast the short-term (risk-free) rates and a VAR to forecast the term premium consisting of the estimated term premiums for annual maturities 1 through 10 to forecast the term premium beyond the 10-year window.[28]

Figure A-2. 10-Year Treasury Note Rate
With and Without Estimated Term Premium



[28] An alternative method would be to forecast all of the parameters used in the ACM model to estimate the term premium at each date and construct the forecasted term premium using the forecasted parameters over various horizons.