ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Predictions of Short-Term Rates and the Expectations Hypothesis

Authors	Massimo Guidolin, and Daniel L. Thornton
Working Paper Number	2010-013B
Revision Date	January 2011
Citable Link	https://doi.org/10.20955/wp.2010.013
Suggested Citation	Guidolin, M., Thornton, D.L., 2011; Predictions of Short-Term Rates and the Expectations Hypothesis, Federal Reserve Bank of St. Louis Working Paper 2010-013. URL https://doi.org/10.20955/wp.2010.013

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Predictions of Short-Term Rates and the Expectations Hypothesis*

Massimo Guidolin

Federal Reserve Bank of St. Louis and Manchester Business School

Daniel L. Thornton

Federal Reserve Bank of St. Louis

JEL Classification: E40, E52.

Key Words: expectations hypothesis, random walk, time-varying risk premium, predictability.

January 2011

Abstract

Despite its role in monetary policy and finance, the expectations hypothesis (EH) of the term structure of interest rates has received virtually no empirical support. The empirical failure of the EH has been attributed to a variety of econometric biases associated with the single-equation models most often used to test it; however, none of these explanations appears to account for the massives failure reported in the literature. We note that traditional tests of the EH are based on two assumptions–the EH per se and an assumption about the expectations generating process (EGP) for the short-term rate. Arguing that convential tests of the EH could reject it because the EGP embedded in these tests is significantly at odds with the true EGP, we investigate this possibility by analyzing the out-of-sample predictive prefromance of several models for predicting interest rates and a model that assumes the EH holds. Using standard methods that take into account parameter uncertainty, the null hypothesis of equal predictive accuracy of each models relative to the random walk alternative is never rejected.

.

*The views expressed here are the authors' and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or the Federal Reserve Bank of St. Louis. We would like to thank Michael Dueker, Victor Gaspar, Jeremy Piger, Lucio Sarno, Martin Sola, Jacky So, and Tao Wu for helpful comments on an earlier draft of this paper and John Zhu and John McAdams for valuable research assistance.

> "The forecasting of short term interest rates by long term interest is, in general, so bad that the student may well begin to wonder whether, in fact, there really is any attempt to forecast."—Macaulay (1938, p. 33)

1. **Introduction**

The expectations hypothesis (EH) of the term structure of interest rates—the proposition that the long-term rate is determined by the market's *expectation* of the short-term rate over the holding period of the long-term bond plus a (constant) risk premium—is one of the key economic principles that is at the core of the monetary policy transmission mechanism. Indeed, virtually every central bank conducts monetary policy by targeting a short-term rate. However, as noted by Woodford (1999, 2003) and others, the effectiveness of monetary policy depends critically on a central bank's ability to affect longer-term rates that matter most for aggregate demand. This has prompted at least four central banks—the Reserve Bank of New Zealand (since 1997), the Norges Bank (since 2005), the Swedish Riksbank (since 2007), and the Czech National Bank (since 2008)—to adopt formal policies of providing forward guidance about the path of the relevant short-term interest rate in an attempt to have a larger effect on longer-term interest rates via a typical EH-like mechanism (e.g., see Andersson and Hofmann, 2010). Moreover, the Fed appears to have used forward guidance beginning 2003 and more explicitly since December 2008.[1] Indeed, Kocherlakota (2010) has recently suggested the Fed's quantitative easing program might represent "(...) another form of forward guidance about the path of the fed funds rate."

The recent trend among central banks to increase the effect of their interest rate policy on longer-term rate via the EH stands in stark contrast with the vast empirical evidence against it. The EH has been tested and rejected using a wide variety of interest rate series, over a variety of sample periods, alternative monetary policy regimes, etc. (e.g., Fama, 1984; Mankiw and Miron, 1986; Campbell and Shiller, 1991; Roberds et al., 1996; Kool and Thornton, 2004; Thornton 2005; Sarno, et al., 2007; and Della Corte, et al., 2008). The most common explanation for the EH's failure is that the single-equation models that have been most often used to test it are subject to spurious rejections because of time-varying risk premia, non-rational expectations, peso problems, measurement errors, etc. However, none of numerous attempts to rescue the EH from such problems (e.g., Simon, 1990; Driffill et al., 1997; Tzavalis and Wickens, 1997; Balduzzi, et al. 1997; Roberds and Whiteman, 1999; Bekaert et al., 2001; Dai and Singleton, 2002; Bansal and Zhou, 2002; Hess and Kamara, 2005) has adequately accounted for the EH's failure. The evidence against the EH is strengthened by Bekaert, Hodrick, and Marshall.'s (1997) demonstration that estimates from these models are even less favorable to the EH because of a positive small-sample bias in parameter estimates. The usefulness of these tests is further complicated by Thornton's (2006) demonstration that

[1] At its August 2003 meeting the FOMC stated that "(...) policy accomodation can be maintained for a considerable period." This or very similar language remained in the policy statement until December 2005. In a somewhat more obvious attempt to increase the effect on longer-term rates, following its December 2008 meeting the FOMC stated that "(...) the federal funds rate is likely to remain exceptionally low [zero to 25 basis points] for an extended period."

single-equation models can yield estimates favorable to the EH when the EH is false.

As noted by Froot (1989) and others, conventional tests of the EH are based on two assumptions: a specific (and simple) linear functional relationship linking changes in long-term rates to expected future changes in short term rates (the EH narrowly defined); an assumption about the data generating process for the market's expectations of the future short rates, i.e., the expectations generating process (EGP). Hence, conventional tests of the EH are really a joint tests of the EH and the EGP. Empirical rejections of the EH can occur either because (a) the EH linkages between long and short-term rates are inconsistent with the data, or (b) because the EGP that is assumed is significantly at odds with the true, but unknown, EGP.[2]

Our paper investigates the possibility that the well documented empirical failure of the EH may be due to the inability to forecast future short-term interest rates in the manner assumed by the EGP that is used to derive conventional tests of the EH. To the extent that the empirical failure of the EH stems from (b) rather than (a), our research provides hope that central banks' recent efforts toward forward guidance may be somewhat effective.[3] Our research is further motivated by the fact that while the validity of the EH is independent of the market's ability to predict future short term interest rates, its practical usefulness is. For example, if the market was unable to predict changes in the short-term rate beyond its current level, the EH could still be valid but would be of little practical usefulness: The term spread would provide no useful information about the future path of interest rates and central bankers would have no need to provide forward guidance about their policy rate. Indeed, investors would avoid any temptation to forecast future changes in short-term rates.

We are not the first to recognize the joint hypotheses problem associated with conventional tests of the EH. For instance, Froot (1989) overcame this problem by using survey data in order to test the EH independently of conventional assumption of the expectations generating mechanism. Noting that when coupled with the standard EGP the EH fails miserably to explain existing data on riskless yields, several researchers have investigated the EH using an alternative EGP for the short-term rate. For example, Fuhrer (1996) compares the observed long-term rate with the that implied by the pure EH based on rational expectations of the federal funds rate obtained from a Taylor-style reaction function with interest rate smoothing (i.e., the lagged funds rate) that allows for shifts in the Fed's reaction function. He finds that his EH-implied long-term rate more nearly matches the observed long-term than that implied by a five-variable VAR. Kozicki and Tinsely (2001) perform a similar analysis allowing for historical shifts in

[2]For instance, considering the simple example of a 3-month T-Bill, the (pure, for simplicity) EH only restricts the 2-month rate to obey $r_t^{2m} = 0.5r_t^{1m} + 0.5E_t[r_{t+1}^{1m}]$. Of course, market expectations of future 1-month T-Bills are hard to measure and reliably collect. However, when one assumes $E_t[r_{t+1}^{1m}] = r_{t+1}^{1m} + \epsilon_{t+1}$ (with ϵ_{t+1} white noise) which amounts to imposing rational expectations on forecasts of future short-term yields, and then rejects $r_t^{2m} = 0.5r_t^{1m} + 0.5r_{t+1}^{1m} + 0.5\epsilon_{t+1}$, it remains unclear whether it is the EH or the assumption that $E_t[r_{t+1}^{1m}] = r_{t+1}^{1m} + \epsilon_{t+1}$ that is rejected, because the test is a joint test of the EH and of a very specific EGP. In practice, a process for the market expectations, $\{E_t[r_{t+1}^{1m}]\}_{t=1}^{T}$, may exist such that the EH holds and yet the model $r_t^{2m} = 0.5r_t^{1m} + 0.5r_{t+1}^{1m} + 0.5\epsilon_{t+1}$ is inconsistent with the data.

[3]Of course, it remains true that shoud the EH-implied connections between long and short rates be false, such efforts could have no effect and be essentially mis-directed. However, this aspect is beyond the scope of our research design.

market perceptions about shifts in the Fed's goal for inflation.[4] Based on their analysis they conclude that "(...) empirical rejections might reflect incorrect assumptions about expectations formation rather than incorrect assumptions about the theoretical link between long rates and short rates."[5] Carriero, Favero, and Kaminska (2006) have suggested that the common practice of using the actual short-term rate as a proxy for the h-period ahead expectation of the short-term rate may be grossly inappropriate and report that evidence against the EH is reduced by using an alternative model of the market's expectation of the short-term rate.

Because the EH places no restrictions on how the market participants' expectations of the short-term rate are formed, imposing auxiliary econometric models to capture the dynamics expectations is interesting but arbitrary. Hence, rather than proposing yet another model-specific process of expectations formation, we follow a growing empirical literature on forecasting interest rates (e.g., Chen and Scott, 1993; Dai and Singleton, 2000; Duffee, 2002; Diebold and Li, 2006; Guidolin and Timmermann, 2009, Bali et al., 2009) to investigate the extent to which the inability to predict future short-term rates might account for the vast empirical failure of the EH. Specifically, we produce real-time, out-of-sample forecasts of short-term rates using a variety of models, some of which have been shown elsewhere (e.g., Diebold and Li, 2006; and Duffee, 2002) to have predictive power for forecasting future interest rates, but are not necessarily consistent with the EH itself. We also make forecasts under the assumption that the EH holds, as these have been popular in the empirical and policy literatures. Our model requires only that the EH is true, i.e., that the long-term rate is determined by the expectation of the short-term rate and that risk premiums are constant on average over the sample period. These forecasts are based on observed long-term yields and, as such, must reflect market's actual expectations of future short-term rates. Moreover, by varying the identifying restriction, EH-implied forecasts can be made by allowing considerable variation in the risk premia over time–a standard explanation for the empirical failure of the EH.[6] Some of the models considered impose little or no structure on the term structure of rates, while others impose considerable structure. For instance, Duffee's (2002) family of "essentially affine" term structure models nest standard linear affine term structure models. Affine term structure models allow for variation in the risk premia and impose no-arbitrage; however, they also impose considerable structure on the shape of the yield curve. Finally, we generate forecasts from two naive benchmarks models: the random walk model and a simple regression model that forecasts the short-term rate by using the slope of the yield curve, as suggested by Duffee (2002). The forecasts are made over a range of maturities over the period 1982-2003, using data on

[4] Kozicki and Tinsley (2005) perform a similar analysis but emphasize the fit of long-term yields based on the convention test of the EH rather than on a comparison with the observed long-term yield as Fuhrer (1996) and Kozicki and Tinsely (2001).

[5] Kozicki and Tinsley (2005), p. 444.

[6] To echo the example in our earlier foonote, one can easily invert the (pure) EH restriction $r_t^{3m} = 0.5r_t^{1m} + 0.5E_t[r_{t+1}^{1m}]$ to find that under the EH it should be $E_t[r_{t+1}^{1m}] = 2r_t^{2m} - r_t^{1m}$. This is the forecast that should prevail under the EH when the risk premium is zero. Section 2.1 shows how this simple intuition can be generalized to longer-term yields and to account for non-zero risk premia. The EH-implied forecasts are conceptually similar but not identical to the implied forward rates used in a portion of the fixed income literature. Section 2.1 further discusses the relationships with this literature.

U.S. riskless, zero-coupon rates.

We report negative results on the ability of all the models examined to predict future short-term rates. Specifically, none of these models are able to generate out-of-sample forecasts that are statistically superior to those obtained from a random walk model. Particularly noteworthy is our finding that our EH-consistent forecasting model frequently yields forecasting performance measures that are smaller than models that require considerable structure and that are much more difficult to estimate. However, there were only a few instances where the EH-consistent forecasts dominated the model-based forecasts based on standard statistical tests of differences in predictive accuracy. There were no instances, however, where any of the models considered were statistically significantly superior to the random walk model, and there were no instances where any forecasting model consistently dominated any other model. While the logic of the EH may be fundamentally correct, markets participants appear to forecast future short-term rates in ways that are systematically different than the EGP assumed in conventional tests of the EH. Our evidence suggests that the violation of this assumption alone is sufficient to account for the massive rejections of the EH found in the literature. Consequently, our results provide some hope that central banks may be able to influence yields further out on term structure, but only if they can succeed in making future short-term rates more predictable.

The outline of the paper is as follows. Section 2 introduces the EH, demonstrates the restrictive nature of the standard assumptions on the underlying EGP, and presents our methodology for generating EH-consistent forecasts of the short-term rate under the assumption that risk premia are constant or smoothly time-varying. Section 3 presents Diebold and Li's (2006) three-factor model and Duffee's (2002) affine and essentially affine models. The time series properties of the data and parameter estimates of the affine term structure models are presented in Section 4. Forecasts from all of the models are compared and analyzed in Section 5. Section 6 presents the results of tests of differences in the out-of-sample predictive accuracy of all of the models considered relative to each other. Section 7 concludes.

2. **The EH and the Predictability of the Short-term Rate**

The EH asserts that for $k > 1$,

$$r_t^n = \frac{1}{k}\sum_{i=0}^{k-1} E_t[r_{t+mi}^m] + \pi^{n,m} \qquad n = km > m, \tag{1}$$

where r_t^n denotes the current n-period rate, $E_t[\cdot]$ denotes the time t conditional expectation operator, and $\pi^{n,m}$ denotes a term-specific but constant risk premium.[7] By construction, k is an integer and is defined as $k = n/m$. The most widely used test of the EH is obtained by subtracting r_t^m from both sides of (1)

[7] Shiller, Campbell, and Schoenholtz (1983) remind us that (1) is exact in some special cases and that it can be derived as a linear approximation to a number of nonlinear expectations theories of the term structure.

and rearranging terms to yield

$$(r_t^n - r_t^m) - \pi^{n,m} = \frac{1}{k}\sum_{i=0}^{k-1} E_t[r_{t+mi}^m] - r_t^m = \frac{1}{k}\sum_{i=1}^{k-1} E_t[\Delta r_{t+mi}^m], \tag{2}$$

where $E_t[\Delta r_{t+mi}^m] \equiv E_t[r_{t+mi}^m] - r_t^m = E_t[r_{t+mi}^m - r_t^m]$ is the expected change in the m-period rate between time t and $t + mi$. (2) states that—apart from a (constant) term risk premium—the spread between long- and short-term rates equals the scaled sum of expected future changes of short-term rates. Single-equation tests of the EH are derived by assuming that market participants' expectations are rational in the sense that

$$E_t[r_{t+mi}^m] = r_{t+mi}^m + \epsilon_{t+mi}^m \qquad i = 1, 2, ..., k-1, \tag{3}$$

where ϵ_{t+mi}^m is distributed i.i.d. $(0, \sigma_{m,i}^2)$ and orthogonal to r_{t+mi}^m, i.e., $E[r_{t+mi}^m \epsilon_{t+mi}^m] = 0$. Substituting (3) into (2) and parameterizing the resulting expression yields:

$$\frac{1}{k}\sum_{i=0}^{k-1} r_{t+mi}^m - r_t^m = \varsigma_0 + \varsigma_1(r_t^n - r_t^m) + \eta_t, \tag{4}$$

where $\eta_t = -\frac{1}{k}\sum_{i=1}^{k-1} \epsilon_{t+mi}^m$. Under the EH, $\varsigma_0 = -\pi^{n,m}$ and $\varsigma_1 = 1$.

The EH has been routinely investigated by testing the null hypothesis that $\varsigma_1 = 1$. Estimates of ς_1 are frequently positive and statistically significant from zero; however, the null hypothesis $\varsigma_1 = 1$ is nearly always rejected with very low p-values. Moreover, estimates of the adjusted R-square are typically very small (frequently less than 10 percent), suggesting that the spread between the longer-term and the short-term rates provides relatively little information about future changes in the short-term rate.

Note that (4) is based on two assumptions: (1) and (3), either of which could be false. The (3) constitutes a strong assumption about the predictability of the future short-term rate. To see why substitute (3) into (1) which yields

$$r_t^n = \frac{1}{k}\sum_{i=0}^{k-1} r_{t+mi}^m + \pi^{n,m} + \epsilon_{t+mi}^m. \tag{5}$$

It is clear from (5) that (3) implies that if the EH holds long-term rate would be equal to the average of the realized short-term rate over the holding period of the long-term asset. If the actual EGP is significantly different from that assumed by (3), test of the EH could reject the null hypothesis that $\varsigma_1 = 1$ even if long-term rates were determined in accordance with the EH, i.e., in accordance with (1).

2.1. *Estimating the Theoretical Expected Future Short-Term Rate*

The EH *per se* places no restrictions on how the market participants' form expectations of the future short-term rate. If market's are forward looking, the current long-term yield must simply embody the market's EGP for the future short-term rate. In fact, the EH can be imposed on riskless yield data to retrieve (risk-adjusted, up to a Jensen inequality term) expectations on the future path of short-term rates, which we call EH-consistent forecast of the short-term rate. To see how the expected short-term rate can be computed under the assumption that the EH holds it is convenient to consider the case where $n = 2$ and $m = 1$, so that (2) is rewritten as:[8]

$$2r_t^2 - 2r_t^1 = E_t[r_{t+1}^1] - r_t^1 + 2\pi^{2,1}. \tag{6}$$

Since both r_t^2 and r_t^1 are observable from time t data, $E_t[r_{t+1}^1]$ can be estimated up to a constant term premium under the assumption that the EH holds:

$$E_t[r_{t+1}^1] = 2r_t^2 - r_t^1 - 2\pi^{2,1}. \tag{7}$$

Indeed, this is procedure is commonly used to estimate the so-called forward rate by assuming that the risk premium is zero (i.e., $\pi^{2,1} = 0$).[9] In fact, academic researchers have built a long tradition in which the (risk-neutral, i.e., under an assumption of zero risk premium) forward rate has been used to predict short-term rates (e.g., Hamburger and Platt, 1975; Fama, 1976; Shiller, Campbell and Schoenholtz, 1983; Fama and Bliss, 1987; Deaves, 1996; Park and Switzer, 1997; and Cochrane and Piazzesi, 2005). With the exception of Deaves (1996), however, all prediction tests originally discussed in the academic literature have been of an in-sample type, which clearly limits their usefulness and informativeness.

It is well known that in reality investors are not risk-neutral and so the risk premia reflected in interest rates are usually positive. In fact, failure of the EH is often attributed to the non-constancy of the risk premia (see, e.g., a simple proof in Engle and Ng, 1993). To reflect this basic empirical fact, we explicitly consider risk premia in calculating the expected future short-term rate. For instance, in the simple case above, we would first proceed to obtain an estimate of $\pi^{2,1}$, to be called $\hat{\pi}^{2,1}$, and then proceed to identify the expectation of the future interest rate as $E_t[r_{t+1}^1] = 2r_t^2 - r_t^1 - 2\hat{\pi}^{2,1}$. In general, (7) can be easily generalized to a recursive set of the so-called Fisher-Hicks formulae:

$$E_t[r_{t+n-1}^1] = hr_t^n - (n-1)r_t^{n-1} - n\pi^{n,1} + (n-1)\pi^{(n-1),1}, \tag{8}$$

for all $n \geq 2$, where $\pi^{1,1} = 0$. (8) shows that the expected future short-term rate is a function of current long-term yields and the corresponding risk premia. Consequently, in order to estimate $E_t[r_{t+n-1}^1]$ an identifying assumption is required to estimate the risk premia. Note that the mean forecast error for $E_t[r_{t+n-1}^1]$ is given by:

$$\frac{1}{T}\sum_{t=1}^{T}\left\{r_{t+n-1}^1 - [hr_t^n - (n-1)r_t^{n-1}]\right\} = \frac{1}{T}\sum_{t=1}^{T}\left\{r_{t+n-1}^1 - E_t[r_{t+n-1}^1]\right\} - n\pi^{n,1} + (n-1)\pi^{(n-1),1}. \tag{9}$$

[8] We set $r_t \equiv r_t^1$, the one-month short-term (T-bill) rate, so that occasionally r_t^1 is simply referred to as r_t.

[9] The implications of the EH are sometimes investigated by regressing changes in the short-term rate on the spread between the forward rate and the current short-term rate (e.g., see Fama and Bliss, 1987). MacDonald and Hein (1989) have found that Treasury bill futures rates are significantly more accurate predictors of future spot rates than are the implied forward rates because the latter would contain a possibly time-varying default risk premium due to the short positions needed to replicate a synthetic forward. Our rolling-window risk premium estimation in (11) may in principle also take these components into account.

If it is assumed that expectations are unbiased on average over the sample period, i.e.,

$$\frac{1}{T}\sum_{t=1}^{T}\left\{r_{t+n-1}^{1}-E_t[r_{t+n-1}^{1}]\right\}=0, \tag{10}$$

the constant risk premium $\hat{\pi}^{n,1}$ can be recursively estimated as:[10]

$$\hat{\pi}^{n,1}=-\frac{1}{Th}\sum_{t=1}^{T}\left\{r_{t+n-1}^{1}-[hr_t^n-(n-1)r_t^{n-1}]\right\}+\frac{n-1}{n}\hat{\pi}^{(n-1),1}. \tag{11}$$

Given estimates of the risk premiums, the expected future short-term rates $E_t[r_{t+n}^1]$ can be estimated by

$$E_t[r_{t+n-1}^{1}]=hr_t^n-(n-1)r_t^{n-1}-n\hat{\pi}^{n,1}+(n-1)\hat{\pi}^{(n-1),1}. \tag{12}$$

We call these constant-risk-premium/EH-consistent forecasts. (1) assumes that risk premia are constant. The empirical failure of the EH is often attributed to time-variation in the risk premia (which is more of a tautology than an explanation). Hence, it is interesting to note that these forecasts can be calculated by assuming that expectations are unbiased over any time horizon, P. Consequently, we also make forecast of the future short rate with $P<T$. Estimates of the risk premia vary considerably over time so we call these time-varying-risk-premium/EH-consistent forecasts. Finding that these forecasts were relatively unaffected by the choice of P, we only report the results for a relatively small value of P.

3. **Alternative Forecasting Models**

As noted in the Introduction, in addition to the EH-implied expectations of future rates, we also consider several term-structure econometric models for forecasting future short-term rates. Specifically, we forecast the short-term rate with the three-factor term structure model of Diebold and Li (2006), with a number of alternative affine or essentially affine models, and with a naïve OLS forecasting model suggested by Duffee (2002) and often used by practitioners.

3.1. *Diebold and Li's Model*

Diebold and Li (2006) use the following modified version of the Nelson and Siegel (1987, 1988) three-factor forward rate curve to approximate the yield curve:

$$r_t^n=\xi_{1t}+\xi_{2t}\left[\frac{1-\exp(-\lambda_t n)}{\lambda_t n}\right]+\xi_{3t}\left[\frac{1-\exp(-\lambda_t n)}{\lambda_t n}-\exp(-\lambda_t n)\right]. \tag{13}$$

The parameter λ_t governs the exponential decay rate. Small values produce slow decay and a better fit at longer maturities, while large values tend to provide a better fit at short maturities. λ_t also governs where the loading on ξ_{3t} achieves it maximum. Because the loading on ξ_{1t} is 1 and, hence, its effect does not decay with the horizon parameter n, Diebold and Li interpret it to be the long-term factor corresponding

[10] The starting condition is given by: $\hat{\pi}^{2,1}=-\frac{1}{2T}\sum_{t=1}^{T}\left\{r_{t+1}^1-[2r_t^2-r_t^1]\right\}$.

to the level of the term structure. Because the factor loading on ξ_{2t} decays monotonically from 1 to zero as $n \to \infty$, ξ_{2t} is viewed as a short-term factor, corresponding to the slope of the yield curve. In contrast, the factor loading on ξ_{3t} rises from zero and then decays back to zero as $n \to \infty$. Hence, Diebold and Li suggest that this factor corresponds to the curvature of the yield curve.

Rather than estimating (13) by nonlinear least squares, Diebold and Li fix the value of λ. They argue that this not only greatly simplifies the estimation of the factors, but likely yields more trustworthy estimates as well. Diebold and Li set $\lambda = 0.0609$, precisely the value where the loading on the curvature factor reaches it maximum under the assumption that the curvature of the yield curve attains its maximum at 30 months.

This framework is then used to generate out-of-sample forecasts of rates at all maturities along the yield curve by making $h-$period ahead forecasts of ξ_i, i.e., $\hat{\xi}_{1t+h}$, $\hat{\xi}_{2t+h}$, and $\hat{\xi}_{3t+h}$. This is done by estimating (13) for rates with maturities 1, 2, 3, 6, 9, 12, 15, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months for each of the first N monthly observations. Out-of-sample forecasts of $\hat{\xi}_{1t+h}$, $\hat{\xi}_{2t+h}$, and $\hat{\xi}_{3t+h}$ are obtained by assuming that the factors follow a simple AR(1) process

$$\hat{\xi}_{it} = c + d\hat{\xi}_{it-1} + \upsilon_{it} \qquad i = 1, 2, 3 \tag{14}$$

and by updating the estimates of c and d recursively. Forecasts of the h-period ahead, n-period rate are then obtained from:[11]

$$\hat{r}^n_{t+h} = \hat{\xi}_{1t+h} + \hat{\xi}_{2t+h}\left[\frac{1 - \exp(-0.0609n)}{0.0609n}\right] + \hat{\xi}_{3t+h}\left[\frac{1 - \exp(-0.0609n)}{0.0609n} - \exp(-0.0609n)\right]. \tag{15}$$

Diebold and Li (2006) report an improvement over random walk forecasts at longer forecast horizons. Additionally, Carriero, et al. (2006) also report some (limited) outperformance of Diebold and Li's model over the random walk for short horizons, even though they provide no formal statistical analysis of the improvement.[12] The mounting evidence of the predictive accuracy of Diebold and Li's framework makes it an important benchmark in our recursive forecasting exercise.

3.2. *Affine and Essentially Affine Term Structure Models*

Duffee (2002) shows that some specific members of the class of "essentially affine" models also can beat random walk forecasts according to a simple Mean Square Forecast Error criterion, where the improvement generally increases with the length of the forecast horizon. Even though Duffee (2002) does not test whether the differences in forecasts are statistically significant, this is an important finding because it suggests that "structural" asset pricing models of the yield curve may be able to pin down the dynamics of risk premia

[11] While not shown here, the time series of the Diebold-Li factors are very similar to the level, slope, and curvature factors obtained from the first three principal components of these 18 zero-coupon bond yields.

[12] Carriero, et al. (2006) also find essentially no improvement in the forecasts when the model it is augmented with additional economic variables, specifically, the CPI-inflation and unemployment rates.

to such an extent that they produce useful and accurate predictions of future rates.[13] We therefore briefly review the structure and properties of affine dynamic term structure models, of which the essentially affine class represents a special case. An Appendix presents additional details and provides technical details on the estimation algorithms.

Given an $M \times 1$ vector $\mathbf{x}_t$ collecting all relevant state variables (risk factors), an affine process for the yield curve is one for which the conditional mean and variance of bond yields are linear affine functions of $\mathbf{x}_t$ and for which also the short-term rate follows an affine process, $r(t) = \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t$. An affine term structure model is then just a special diffusion Markov process,[14]

$$\begin{aligned} d\mathbf{x}_t &= \underset{M\times M}{\mathcal{K}}(\underset{M\times 1}{\boldsymbol{\theta}} - \underset{M\times 1}{\mathbf{x}_t})dt + \underset{M\times M}{\mathbf{S}_t^{1/2}}\underset{M\times 1}{d\mathbf{W}_t} \\ [\mathbf{S}_t^{1/2}]_{ii} &= \sqrt{\alpha_i + \boldsymbol{\xi}_i'\mathbf{x}_t} \quad i = 1, ..., M, \end{aligned} \tag{16}$$

where $\mathbf{W}_t$ is a $M \times 1$ vector of independent Brownian motions and $[\mathbf{S}_t^{1/2}]_{ii}$ is the $i-$th element on the main diagonal of $\mathbf{S}_t^{1/2}$. The state variables are mean reverting as long as the elements of $\mathcal{K}$ are positive. Also, the larger they are, the faster is the rate of mean reversion. To price bonds in this framework, assume next that the pricing kernel $\mathcal{M}$ has structure

$$d\mathcal{M}_t = -r_t\mathcal{M}_t dt - \mathcal{M}_t \boldsymbol{\Lambda}(\mathbf{x}_t)' d\mathbf{W}_t, \tag{17}$$

where $\boldsymbol{\Lambda}_t \equiv \boldsymbol{\Lambda}(\mathbf{x}_t)$ is the $M \times 1$ vector of prices of risk associated with each of the M risk factors. By Ito's lemma, we know that $d\ln\mathcal{M}_t = \left(-r_t - \frac{1}{2}\boldsymbol{\Lambda}_t'\boldsymbol{\Lambda}_t\right)dt\ -\boldsymbol{\Lambda}_t' d\mathbf{W}_t$, so that it easy to prove that $d\mathbf{W}_t^{\mathbb{Q}} = d\mathbf{W}_t^{\mathbb{P}} + \boldsymbol{\Lambda}_t dt$, and the physical representation of the stochastic process for the state vector is:

$$\begin{aligned} d\mathbf{x}_t &= \mathcal{K}(\boldsymbol{\theta} - \mathbf{x}_t)dt - \mathbf{S}_t^{1/2}\boldsymbol{\Lambda}(\mathbf{x}_t)dt + \mathbf{S}_t^{1/2}d\mathbf{W}_t^{\mathbb{P}} \\ [\mathbf{S}_t^{1/2}]_{ii} &= \sqrt{\alpha_i + \boldsymbol{\xi}_i'\mathbf{x}_t} \quad i = 1, ..., M. \end{aligned} \tag{18}$$

The representation (18) is important to compute the moments implied by any parameter configuration, including the vector of risk premia $\boldsymbol{\Lambda}(\mathbf{x}_t)$ and is featured in a quasi-maximum likelihood estimation (QMLE) approach. At this point, the derivation of bond prices in the affine case under a discrete time representation is straightforward (see the Appendix).

Within the general affine class, we can distinguish a number of cases that have received attention in the asset pricing literature. In particular, two important cases are obtained depending on whether $\boldsymbol{\Lambda}(\mathbf{x}_t)$ is parameterized as either

$$\Lambda_n(\mathbf{x}_t) = \lambda_n(\alpha_n + \mathcal{B}_n'\mathbf{x}_t)^{1/2} \tag{19}$$

[13]Here "structural" means that we fit to the data specific models of the relationship between the quantity of risk and the risk premium. Although affine models are also useful to forecast the second moments of interest rates (and interesting trade-offs exists between this goal and fitting the level and shape of the yield curve, see Duffee, 2002, for details), in this paper we focus on their ability map predictions of future risk premia into prediction of future interest rates.

[14](16) represents the stochastic process for the state vector under the risk-neutral measure, i.e., without any correction for the price of risk factors. Implicitly, we assume all the necessary restrictions to ensure that the linear affine dynamics is well defined, which requires that $\alpha_i + \boldsymbol{\xi}_i'\mathbf{x}_t$ is nonnegative for all i and all possible values of $\mathbf{x}_t$, see e.g., Dai and Singleton (2000).

$(n = 1, ..., M)$ or

$$\Lambda_n(\mathbf{x}_t) = \lambda_{1n}(\alpha_n + \mathcal{B}_n'\mathbf{x}_t)^{1/2} + \boldsymbol{\lambda}_{2n}' \left[\begin{cases} \frac{1}{(\alpha_n + \mathcal{B}_n'\mathbf{x}_t)^{1/2}} & \inf_i(\alpha_i + \mathcal{B}_i'\mathbf{x}_t) > 0 \\ 0 & \text{otherwise} \end{cases} \right] \mathbf{x}_t, \tag{20}$$

where $\mathcal{B}_n'$ is the n-th row of the matrix $\mathcal{B}$. For instance, Vasicek's (1977) model is obtained when $\mathcal{B}_n = \mathbf{0}$ $(n = 1, ..., M)$ so that $\boldsymbol{\Lambda}(\mathbf{x}_t)$ becomes a vector of constant prices of risk. Likewise, Cox, et al., (1985) is obtained when $\alpha_n = 0$ and $\mathcal{B}_n = \boldsymbol{\iota}_n$ $(n = 1, ..., M)$, so that the prices of risk are time-varying and simply proportional to the risk factors. The *completely* affine case occurs when in (19) $\alpha_n = 0$ $(n = 1, ..., M)$. The models by Vasicek, Cox et al., and Duffie and Kan are all completely affine models, with Vasicek and Cox et al., being particularly restrictive versions of the completely affine family. The *essentially* affine case of Duffee (2002) consists of (20).

An important limitation of completely affine specification of $\boldsymbol{\Lambda}_t$ is that the temporal variation in the instantaneous expected excess returns on h-period zero coupon bonds $(\eta_{D,t}^h)$ is determined entirely by the volatilities of the state variables:

$$\eta_{D,t}^h = -\mathbf{B}_h'\mathbf{S}_t^{1/2}(\mathbf{x}_t)\boldsymbol{\Lambda}(\mathbf{x}_t),$$

where $\mathbf{B}_h$ is an appropriate pricing vector determined in the absence of arbitrage opportunities (see the Appendix for details). Moreover, the sign of each $\boldsymbol{\Lambda}_n(\mathbf{x}_t)$ is fixed over time and determined by the sign of the coefficients in $\mathcal{B}_n$. Although this does not precluded $\eta_{D,t}^h$ from changing sign over time—the sign of $\eta_{D,t}^h$ depends also on the sign and magnitude of the elements of $\mathbf{B}_h$ and $\mathbf{S}_t^{1/2}(\mathbf{x}_t)$—this represents a strong limitation to the flexibility of excess returns to display the patterns that are typical of the data. The essentially affine set up in (20) allows for variation in prices of risk independent of volatilities, the kind of flexibility needed to fit the empirical behavior of excess bond returns.

In this paper we estimate and forecast interest rates using a few alternative *canonical* affine models. A canonical model is one that is admissible (this means that all even moments are guaranteed to be positive), econometrically identified, and maximally flexible within the affine family. Consider the case where there are $L \geq 0$ state variables (without loss of generality, the first L elements of $\mathbf{x}_t$) driving the instantaneous conditional variances of $\mathbf{x}_t$. Then a benchmark $A_L(M)$ affine model in canonical form may be written (under the risk neutral measure) as:[15]

$$\begin{aligned}
d\mathbf{x}_t &= \begin{bmatrix} \underset{L\times L}{\mathcal{K}^{VV}} & \underset{L\times(M-L)}{\mathbf{O}} \\ \underset{(M-L)\times L}{\mathcal{K}^{DV}} & \underset{(M-L)\times(M-L)}{\mathcal{K}^{DD}} \end{bmatrix} \left(\begin{bmatrix} \underset{L\times 1}{\boldsymbol{\theta}^V} \\ \underset{(M-L)\times 1}{\mathbf{0}} \end{bmatrix} - \mathbf{x}_t \right) dt + (\mathbf{S}(\mathbf{x}_t))^{1/2} d\mathbf{W}_t \quad \mathcal{K}_i^{VV}\boldsymbol{\theta}_i^V > 0 \quad 1 \leq i \leq L \\
\mathbf{S}_t &= \begin{bmatrix} \underset{L\times 1}{\mathbf{0}} \\ \underset{(M-L)\times 1}{\mathbf{1}} \end{bmatrix} + \begin{bmatrix} \underset{L\times L}{\mathbf{I}} & \underset{L\times(M-L)}{\mathbf{O}} \\ \underset{(M-L)\times L}{\div^{DV}} & \underset{(M-L)\times(M-L)}{\mathbf{O}} \end{bmatrix} \mathbf{x}_t \quad \div^{DV} \geq \mathbf{O} \\
r_t &= \delta_0 + \boldsymbol{\delta}'\mathbf{x}_t \quad \delta_i \geq 0 \quad L+1 \leq i \leq M,
\end{aligned} \tag{21}$$

[15]When $L = 0$ (a purely Gaussian model with constant second moments) $\mathcal{K}$ is simply required to be either upper or lower triangular. When $L > 0$, additional conditions have to be added to achieve econometric identification.

where $\boldsymbol{\theta}_i^V \geq 0$, $\mathcal{K}_{ij} \leq 0$, $1 \leq i \leq L$, $1 \leq j \leq L$ $(j \neq i)$. The block structure nature of $\mathbf{S}_t$ implies that only the first L state variables impact the conditional variance of the entire vector $\mathbf{x}_t$:

$$[\mathbf{S}_t]_{ii} = \begin{cases} x_{it} & 1 \leq i \leq L \\ 1 + \sum_{l=L+1}^{M} \Xi_{il} x_{lt} & L+1 \leq i \leq M \end{cases}. \tag{22}$$

In terms of estimation, we adopt Duffee's (2002) QMLE method, which can be seen as a special case of an under-identified GMM estimator when only two moment conditions are imposed. Assume that at each month-end t, $t = 1, ..., T$, *yields* on M bonds are measured without error. These bonds have fixed times to maturity h_1, ..., h_M. Yields on $N - M$ other bonds are assumed to be measured with serially uncorrelated, mean-zero measurement errors. As common in the literature, we impose structure on the joint distribution of measurement errors and yields in order to derive the likelihood function of the data: measurement errors collected in the $(N - M) \times 1$ vector $\boldsymbol{\epsilon}_t$ are jointly normally distributed with constant covariance matrix and density $\phi_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t)$. At this point, stack the perfectly observed yields in the vector $\mathbf{Y}_t$ and the imperfectly observed yields in the vector $\check{\mathbf{Y}}_t$. Denote the parameter vector by $\boldsymbol{\theta}$. Because the distribution of $\mathbf{Y}_{t+1}$ conditional on $\mathbf{Y}_t$ is

$$f_Y(\mathbf{Y}_{t+1}|\mathbf{Y}_t) = \frac{1}{\left|\det(\ddot{\mathbf{B}}'_{h_i})\right|} f_X(\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t) \tag{23}$$

($\ddot{\mathbf{B}}_{h_i}$ is a matrix defined in the Appendix and it will depend on the complete or essentially affine nature of the model), the log-likelihood of observation t for $\check{\mathbf{Y}}_t$ is $\ell_t(\boldsymbol{\theta}) = \ln f_Y(\mathbf{Y}_t|\mathbf{Y}_{t-1}) + \ln \phi_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t)$. The estimated parameter vector $\hat{\boldsymbol{\theta}}_T^{QMLE}$ is chosen to solve

$$\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left[\frac{1}{|\det(\ddot{\mathbf{B}}'_{\tau_i})|} f_X(\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t) + \ln \phi_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t) \right], \tag{24}$$

where $f_X(\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t)$ follows a multivariate Gaussian distribution for which it is tedious but possible to derive closed-form representations for the first and second conditional moments (see the Appendix).[16]

3.3. *Naive Benchmarks*

Finally, we also forecast the short-term rate using two models that are frequently used in the financial forecasting literature. The simplest benchmark model is a random walk, where the month t yield on a n-maturity bond is used as the forecast of the month $t + h$ yield on a n-maturity bond. We also consider what Duffee (2002) calls "a more sophisticated benchmark," where the forecast of the future short-term rate is based on the slope of the yield curve. These forecasts are based on OLS regressions of

$$r_{t+h}^m - r_t^m = \varsigma_0 + \varsigma_1 (r_t^{5Y} - r_t^{3m}) + \zeta_{t+h}^m, \tag{25}$$

where r_t^{5Y} is the 5-year Treasury yield and r_t^{3m} is the 3-month T-bill rate. The parameters of (25) are recursively estimated with monthly updating to produce out-of-sample forecasts and forecast errors of the short-term rate at the horizons considered here.

[16]Estimation has been performed by updating the Fortran code kindly made available by Greg Duffie.

4. **Data and Estimation Results for Econometric Benchmarks**

The data are end-of-period monthly observations on continuously compounded yields on riskless pure discount bonds for the U.S. The raw data are from Bloomberg. The riskless pure discount bond yields were obtained using FORTRAN codes provided by Robert Bliss and Dan Waggoner based on Bliss (1997) and Waggoner (1997). The yields were calculated for bonds with maturities of 1, 2, 3, 6, 9, 12, 15, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months for the period January 1970 through December 2003 for maturities between 1- and 72 months, and for slightly shorter samples in the case of maturities between 84 and 120 months. Table 1 reports summary statistics for our implicit zero coupon yields. As one would expect, on average the term structure of US riskless rates has maintained a moderately positive slope, with average nominal yields ranging from 6% at the short end to 7.9% at the back end. This finding also holds with reference to median rates or if one uses a balanced sample common to all series. All the yield series are clearly non-Gaussian and—even after first-differencing—appear to contain strong heteroskedasticity patterns (square changes in interest rates are strongly serially correlated) and robust serial correlation, especially at the shortest end of the yield curve.

Forecasts from the Diebold-Li model are obtained by estimating the three factors using all of the available rates along the term structure, i.e., rates with maturities 1, 2, 3, 6, 9, 12, 15, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months, of each month over the period January, 1972 - December, 1981. h-period ahead forecasts of each of the three factors, $\xi_{it}, i = 1, 2, 3$ are then obtained from (14) using estimates of the factors over this initial sample period. These forecasts are then used to obtain predictions of the 1- and 3-month Treasury rates using (15). The process is updated recursively to generate out-of-sample forecasts of the 1-month T-bill rate for horizon of 1 and 2 months, and of the 3-month T-bill rate at horizons of 3, 6, 9, 12, and 15 months. While not shown here, the estimated factors correspond very closely to estimates of the level, slope, and curvature factors obtained from the first three principal components obtained from the yield data and are comparable to the results reported by Diebold and Li (2006).[17]

Similarly to Litterman and Scheinkman (1991) and Duffee (2002), all of the affine models also assume three underlying factors ($M = 3$). We estimate four different three factor models: a completely affine, mean-reverting purely Gaussian model ($L = 0 < M = 3$); one completely affine model with $L = 2$ and $M = 3$; one essentially affine model that is designed to capture volatility dynamics with high accuracy ($L = 1 < M = 3$); and one essentially affine Gaussian model that trades-off the ability to fit volatility dynamics with the ability to induce rich time variation in bond yields ($L = 0 < M = 3$).[18] As in Duffee (2002), we assume that the bonds with no measurement error are those with maturities of 3 months, 2 years, and 5 years. The remaining maturities fill in the gaps in the term structure and are assumed to

[17]Detailed results are available from the Authors upon request.

[18]The completely affine model with $L = 2$ and $M = 3$ is selected over the case of $L = 3$ because, similarly to what reported by Duffee (2002), this model fails to be rejected (using a standard overidentifying test) when compared to the corresponding essentially affine model. The $L = 0 < M = 3$ essentially affine model is selected because of its good forecasting performance documented in Duffee (2002).

be measured with error. For all the models investigated, we follow Duffee (2002) and also entertain more parsimonious, scaled-down specifications based on the following algorithm:

- first compute the (Wald) t-statistics for the unrestricted parameter estimates;
- set to zero all parameters for which the (robust) p-value exceeds 0.10;
- re-estimate the model under the second-state restrictions.

In the following, we report both in- and out-sample results for both unrestricted and restricted affine models. Finally, we perform a recursive pseudo out-of-sample exercise with a block structure, in the sense that parameter estimates are updated with bi-annual frequency, i.e., starting with 1972:01-1981:12, followed by 1972:01-1983:12, etc., up to 1972:01-2001:12.[19]

Table 2 reports full-sample estimates for the simple $L = 0$ completely affine, mean-reverting three-factor Gaussian model in which volatility fails to depend on $\mathbf{x}_t$.[20] The table reports parameter estimates for both an unrestricted model (apart from the restrictions implied by the canonical form) and for a model in which all unrestricted coefficients with a first-round p-values (approximately) in excess of 0.1 have been forced to zero, and the resulting, restricted model re-estimated by QMLE. For instance, in Table 2 initial unrestricted estimation of the 19-parameter completely affine Gaussian model yields two parameter estimates (the [3,2] element of $\mathcal{K}$ and the risk premium on the first factor, λ_{11}) with pseudo t-ratios below 1.66. Therefore, columns 4-8 of Table 2 re-estimate the completely affine model after setting the two parameters to zero, with the result of obtaining a more parsimonious, 17-parameter model.[21] However, a standard likelihood ratio test of the two restrictions imposed in columns 4-8 rejects the null that the restrictions are not penalized by the resulting optimal likelihood function (the LR statistic is approximately 6.9, which yields a rather small p-value of 0.03 under a $\chi^2_{(2)}$). This is an indication against imposing the constraints, even though the three standard information criteria reported at the bottom of Table 2 signal that the restrictions may in principle improve the out-of-sample forecasting performance of the model, as all the criteria substantially decline when the two restrictions are imposed.[22] Both the restricted and unrestricted models, with saturation ratios (the number of observations available to estimate each of the parameters) of 386 and 431, seem to be based on a sufficient number of observations to deliver reliable inferences.

Table 3 reports estimates for unrestricted and restricted versions of a richer, non-Gaussian completely affine model with $L = 2$. Although the models in Tables 2 and 3 are non-nested and testing the case

[19]The full-sample 1972:01-2003:12 estimates are presented in what follows but actually never used in the recursive predictions.

[20]The canonical form for the completely affine $A_0(3)$ model implies $\mathcal{K}\boldsymbol{\theta} = \mathbf{0}$, $\mathcal{K}$ lower triangular, $\div^{DV} = \mathbf{O}$, and $\boldsymbol{\Lambda}_2 = \mathbf{O}$. Also notice that the canonical model is written in a form that makes the $\boldsymbol{\lambda}$ coefficients *the negative* of the unit prices of risk.

[21]Incidentally, when going from the unrestricted to the resticted ($L = 0$) affine model, we notice that two additional parameters yield pseudo t-stats below 1.66. This suggests further simplications that have not been pursued here. However, in general Tables 3-5 concerning the other affine-class models tend to be free of these problems.

[22]The three statistics are the Bayes-Schwartz, the Akaike, and the Hannan-Quinn information criteria. One should bear in mind that these criteria trade-off in-sample fit with parsimony (hence, potential for out-of-sample predictive accuracy) and that a *declining* information criterion is an indication of a better performing model.

$L = 0$ vs. the case $L = 2$ remains difficult, Table 3 makes it obvious that a richer dependence of the volatility matrix from the state vector does produce a superior in-sample fit, as shown by the fact that the maximized log-likelihood function climbs up from -1386 in Table 1 to -1359 in Table 2 (from -1390 to -1365 when p-value related restrictions are imposed).[23] In fact, in the case of $L = 2$, a first-pass estimation reveals that as many as 4 parameters of the unrestricted model generate p-values in excess of 0.1, leading to the estimation of relatively parsimonious 20-parameter model (only one parameter in excess of the unrestricted $L = 0$ Gaussian completely affine model in Table 2) that gives a rather impressive fit to the data. Also in Table 3 the restrictions led by a pseudo p-value threshold of 0.1 are rejected by a likelihood ratio test (the p-value is 0.02), although the two most parsimonious information criteria (Bayes-Schwartz and Hannan-Quinn) record substantial declines when the 4 restrictions are imposed. Interestingly, three restrictions have clear economic interpretation: the short term rate should not depend on the third risk factor, the volatility of the third factor should not depend on the level of the term structure of interest rates, and the first (level) factor fails to command a significant price of risk.

Tables 4 and 5 report the estimates of the two essentially affine specifications.[24] In both cases the saturation ratios remain well in excess of 200, meaning that there are always at least 200 observations on Treasury yields to estimate each of the parameters implied by the affine models. In Table 4 the model is Gaussian with $L = 0$ which means that the ability to fit the volatility dynamics of the state vector is rather limited. As a result, the in-sample fit provided by an essentially affine, purely Gaussian model with $L = 0$ is only slightly superior to the fit of a completely affine model with $L = 2$. Because the essentially affine model generally has more parameters to be estimated, this translates in higher (i.e., worse) values for the information criteria (e.g., the Schwartz criterion goes from 1.29 in the unrestricted completely affine case to 1.36 in the unrestricted essentially affine Gaussian case; the matching values under restricted estimation are 1.27 and 1.33). Section 6 to follow checks whether this higher information criteria actually translate into an inferior out-of-sample forecasting performance. The in-sample fit obtained is considerably better in Table 5, where results for an essentially affine model with $L = 1$ (i.e., the first state variable is also allowed to drive time variation in volatility for the entire state vector) are displayed. As a result, the log-likelihood is now considerably higher than in Table 3 (e.g., from -1359 to -1258 in the unrestricted case), even though this superior in-sample fit is only partially reflected by the information criteria: while the Akaike and Hannan-Quinn criteria improve when going from the completely affine model with $L = 2$ to the essentially affine model with $L = 1$, this is not the case for the Schwartz criterion. As in Table 4, the pseudo p-value-driven restrictions (5 restrictions in both cases) are rejected by the likelihood ratio test, even though they systematically lead to lower (better) information criteria.

[23] Here the canonical form and absence of arbitrage imply the restrictions $\mathcal{K}\theta_1 = 0$, $\mathcal{K}_{13} = \mathcal{K}_{23} = 0$, $\div_{33} = 0$; $\mathcal{K}\theta_3$ is non-zero with no standard error, because $\theta_3 = 0$.

[24] Also here some restrictions are implied by the absence of arbitrage and the canonical form. For instance, in Table 5 we impose $\mathcal{K}_{12} = \mathcal{K}_{13} = 0$, $\mathbf{\Lambda}_{2,11} = \mathbf{\Lambda}_{2,12} = \mathbf{\Lambda}_{2,13} = 0$, $\theta_2 = \theta_3 = 0$, while $\mathcal{K}\theta_2$ and $\mathcal{K}\theta_3$ can be computed from the implied estimates for θ_1 and $\mathcal{K}$, but they have no standard errors.

Our qualitative findings confirm Duffee's (2002) finding that models that are better able to produce time-varying volatilities have higher maximized log-likelihood (QML) values than models with time-invariant yield volatilities—as the number of factors that affect volatilities increases from zero through three, QML values increase monotonically. Based on the in-sample maximized log-likelihood values, the additional flexibility offered by essentially affine models over completely affine models is important. This is also partially confirmed by the performance of information criteria.

5. **Forecasting Performance**

In this Section, we systematically investigate the recursive (pseudo) out-of-sample forecasting performance of the alternative models presented in Sections 2 and 3. For each model, we estimate/forecast the expected 1-month rate for 1-, 2-month horizons and the expected 3-month rates at 3-, 6-, 9-, 12-, and 15-month horizons. Hence, in what follows all 1- and 2-month ahead forecasts refer to 1-month rates, while the 3, 6, 9, 12, and 15 month ahead forecasts are for 3-month rates. Of course, we focus on both short- and medium-horizon forecasts of short-term rates, given our conjecture in Section 2 that the widespread rejections of the EH reported in the literature may derive from the pervasive difficulty that market participants face when they are called to form rational forecasts of future yields. For affine and essentially affine models the 1-month rate is assumed to be measured with error and it seems sensible to investigate the recursive predictive accuracy of linear affine models both with reference to rates that are assumed to be subject to noise and those which are not. The Diebold-Li, affine, and OLS sloped-based forecasts are initialized using monthly data for the period January 1972, though December 1981. Out-of-sample forecasts are generated recursively for the period 1982:01-2003:12.[25]

5.1. *Theoretical EH Forecasts*

We first generate theoretical, EH-implied forecasts assuming that the risk premium is constant over the entire sample period, 1972:01-2003:12. The estimates of the constant risk premia are $\hat{\pi}^{2,1} = 0.149$, $\hat{\pi}^{3,1} = 0.282$, $\hat{\pi}^{6,3} = 0.238$, $\hat{\pi}^{9,3} = 0.353$, $\hat{\pi}^{12,3} = 0.469$, $\hat{\pi}^{15,3} = 0.601$. These estimates are reasonable and, as one might expect, increase at a decreasing rate as the term to maturity lengthens. The forecast errors under the assumption that the risk premia are constant are very similar to those obtained from the random walk model. This is illustrated in Figure 1, which shows the theoretical (solid lines) and random walk (dashed lines) forecast errors for the 3-month T-bill rate at the 3-month and 15-month horizons, respectively (the figures for the 1-month rate and all other horizons are qualitatively similar and therefore not shown). The forecast errors are sometimes large in absolute value and, not surprisingly, the forecast errors are largest in the early 1980s. Moreover, the absolute size of the forecast errors and their standard deviation tend to increase monotonically as the forecast horizon lengthens. Of course these forecasts are identical —

[25]When $h > 1$ months, the pseudo out of sample evaluation period is 1982:01 - 2003:12$-h$ months.

apart from a constant re-scaling — to those typical of the (implied) forward rate literature, where the risk premium has been typically set to zero. Their visual similarity to random walk forecasts makes us speculate that predicting future short-term rates in the US Treasury market may require investors substantial more effort than the calculation of implied forward rates.

5.2. *Time-Varying Risk Premiums*

The assumption that the risk premium is constant is at odds with the massive rejections of the EH found in the literature. However, with the exception of Dai and Singleton (2002) and Tzavalis and Wickens (1997), whose approaches are flexible enough to account for nearly all of the time variation in the observed risk premiums, time-varying-risk-premium explanations of the lack of empirical success of the EH have been relatively unsuccessful (e.g., Hardouvelis, 1994; Rudebusch, 1995; Bekaert, et al., 1997; and Roberds and Whiteman, 1999). We therefore proceed to compute forecasts by allowing for time variation in risk premia using the methodology outlined in Section 2.

To investigate the effect of time variation in risk premia on the forecast errors, the EH-implied risk premia are alternatively computed by assuming that the forecast errors average to zero over a rolling window of P observations. It is obvious from (11) that the estimated risk premia are likely to vary considerably when estimated over short samples. Several values of P were considered. While the degree of time variation in the estimated risk premia was sensitive to the choice of P, the estimated forecast errors were not. Consequently, the results are presented for P equal to ten months. Estimates of the time-varying risk premiums for the 1-month rate at the 1- and 2-month horizons are presented in Figure 2 along with the corresponding estimate of the constant risk premiums over the entire sample period (again, the figures for the other horizons and for the 3-month rate are very similar to those shown here and all of the estimated time-varying risk premiums are stationary). Interestingly, the risk premia decline below their full-sample average during the period of the so called "great moderation".

Figure 3 compares the forecast errors under the constant and time-varying risk premium assumptions for the 3-month T-bill rate for the 3- and 15-month investment horizons, respectively. Figure 3 shows that differences in the forecast errors are small at the 3-month horizon. The differences are larger at the 15-month horizon; however, as we discuss below, there is relatively little difference in their average forecast performance. Hence, allowing for considerable variation in risk premia, i.e., the failure of the EH, appears to have relatively little effect for the predictive power of the long-term rate.

Table 6 presents summary statistics for the monthly theoretical forecast errors — $e^{n,k}_{t,h} \equiv r^n_{t+h} - \hat{r}^{n,k}_{t,t+h}$ where n is the maturity, k denotes a model, and h is forecast horizon — for all investment horizons and for all of the models considered here. For comparability, these statistics are calculated using forecast errors over the common out-of-sample period, 1982:01-2003:12. Panels A, B, and C report the forecasting performance of the theoretical forecast with a constant risk premium and a time-varying risk premium ($P = 10$), and the forecasts from the random walk model. Not surprisingly, the theoretical forecasts have practically zero means. The medians are small and positive at shorter horizons and small and negative at longer horizons. The summary statistics for the forecast errors from the random walk model show that the mean forecast errors are slightly negative at nearly all horizons, indicating a tendency of the random walk model to under-predict the corresponding short-term rates (the exception is at the 3-month horizon where the 3-month rate is over-predicted by the random walk model). Moreover, the under-prediction increases monotonically as the investment horizon lengthens beyond three months. The similarity in summary statistics suggests a high degree of correspondence between the theoretical and random walk forecasts.

Table 6 also presents standard summary measures of forecasting accuracy, i.e., the Mean Squared Forecast Error (MSFE, and its square root, the RMSFE which is directly comparable to the scale and mean of the predicted series), and the Mean Absolute Forecast Error (MAFE). A comparison of Panels A and C show that theoretical and random walk forecasts are very similar. At $h = 1$, the random walk model performs slightly better, with a RMSFE of 47 b.p. against 51 b.p. for the theoretical model and with a MAFE of 30 b.p. vs. 31 b.p. for the theoretical model. The results are mixed for longer horizons. At some horizons the theoretical models perform marginally better than the random walk model. At $h = 6$, the theoretical model out-performs the random walk model by the RMSFE metric, but is out-performed by the random walk model by MAFE. However, for horizons longer than six months, the theoretical model with constant risk premium is superior to the random walk by both the RMSFE and MAFE metrics.

The impression from Figure 3, that there is little difference in the forecast from the theoretical models based on the constant and time-varying risk premiums assumptions is confirmed by a comparison of forecasting performance of the theoretical model with constant risk premia with those with time-varying risk premia reported in Panels A and B, respectively. In qualitative terms, the predictive accuracy of the constant and time-varying risk premium models are similar in both the RMSFE and MAFE metrics. However, small differences in forecasting performance can be detected at different horizons. For $h = 1$ and 2, modeling time-varying risk premiums model has a slight edge in forecasting accuracy by all three metrics. At horizons of 3 and 6 months, the predictive performances of the two specification are virtually indistinguishable. For horizons of 9 months and longer, forecasts based on the constant risk premium assumption have a modest predictive advantage. The relatively small differences in the forecasting performance of the constant and time-varying risk premia models suggests that the effect of variation in the risk premium on the forecast errors is modest relative to the effect of new information. That is, the forecast errors appear to be dominated by news, which the market participants are unable to forecast.

5.3. *Diebold and Li's Forecasts*

The forecast errors for the Diebold-Li model and the benchmark random walk model at the 3-month and 15-month ahead horizons, respectively, are presented in Figure 4. Also in this case, similar plots for 1-month T-bill rates and/or for alternative forecast horizons gave qualitatively identical indications. The random walk forecast errors closely track the Diebold-Li forecast errors at the 3-month horizon. The differences increase with the forecast horizon, however, they appear to be relatively modest even at the 15-month horizon. This impression is confirmed by the statistics on the forecasting performance of the Diebold-Li model presented in Panel D of Table 6. The Diebold-Li model performs relatively worse than the random walk benchmark by all three forecast metrics at the 1- and 2-month horizons, but somewhat better at the 3-month horizon. The results are mixed for horizons beyond three months, with the Diebold-Li model doing somewhat better than the random walk alternative by some metrics and worse by others. In general, however, differences in the forecasting performance by RMSFE and MAFE metrics are small at horizons of six months and longer.

A comparison of Diebold-Li model forecasts with the theoretical forecasts (either constant or time-varying) yields a similar conclusion. Specifically, Diebold-Li forecasts are somewhat worse than the theoretical forecasts by the RMSFE and MAFE metrics at the 1- and 2-month horizons, somewhat better than the theoretical forecasts at the 3-month horizon, and generally mixed at longer horizons, with the difference typically being very small.

5.4. *The OLS, Slope-Based Naive Model*

Panel E of Table 6 presents the forecasting performance for the naive slope-based benchmark model suggested by Duffee (2002) and popular with fixed income practitioners. The forecasting performance of this model, presented in Panel E of Table 6, indicates that this model performed considerably worse than that of any of the preceding models at all possible horizons. The relative performance of this model is particularly poor at short horizons, but improves as the forecast horizon lengthens. The improved performance of the slope of the yield curve for forecasting the future short-term rate is consistent with Cochrane and Piazzesi's (2005) findings on the in-sample predictive power of forward rates for bond risk premia. The poor performance of the slope of the yield curve relative to the Diebold-Li alternative is not surprising in view of the fact that the Diebold-Li model contains considerably more information about the yield curve. However, that neither of these models performs markedly better than the random walk benchmark (which contain no yield curve information) suggests that information about the yield curve may not be particularly useful for predicting the future short-term rate.

5.5. *Completely Affine Models*

Panels F-G of Table 6 report on the forecasting performance of completely affine models, when risk premiums are simply a linear function of the variance of the price risk factors (here, three). To save space, the results are reported only for a mixture of restricted and unrestricted affine models. Panels F and G report the results for the unrestricted completely affine purely Gaussian model with $L = 0$ and for the restricted (more parsimonious) completely affine model with $L = 2$. Similarly to Duffee's (2002) results, the out-of-sample forecasting performance of completely affine models is extremely disappointing. They systematically outperform only the weak OLS, slope-based benchmark of panel E. Their predictive accuracy is inferior to that of the random walk and, like the slope-based benchmark, very poor at short horizons. For instance, at $h = 1$ (for 1-month rates), an unrestricted Gaussian affine model yields a RMSFE of 62 b.p. and a MAFE of 36 b.p. compared with 47 b.p. and 30 b.p., respectively, for the random walk specification. When restrictions are imposed that make the model more parsimonious, the results (unreported) are essentially identical.

The results for the restricted completely affine model with $L = 2$, presented in Panel G, are qualitatively similar those presented in Panel F. This model performs much worse than the random walk at short horizons and only slightly better for investment horizons longer than 9 months.[26] These results suggest that the completely affine frameworks cannot yield a predictive performance that is comparable to either the random walk or to the best EH-type forecasts. Figure 5 reports a visual impression for both sets of forecasts and plots completely affine forecast errors in comparison to random walk forecast errors. Although the differences never appear major—both when the forecast errors of the two completely affine models are compared and when the comparison is extended to the random walk—it is clear that differences vs. the random walk are modest and tend to favor the random walk benchmark over the structural models.

5.6. *Essentially Affine Models*

The forecasting results for the two essentially affine models estimated in this paper are presented in Panels H and I of Table 6. The unrestricted essentially affine model forecasts are uniformly superior to those of the restricted essentially affine alternative and the completely affine alternatives.[27] The relative improvement in forecasting performance is dramatic at short horizons, but appears to be only marginal at longer horizons. It is interesting to note that the unrestricted essentially affine model does not fit particularly well in-sample because the proposed structure for the dynamics in second moments remains rather rudimental. Nevertheless, as Duffee (2002) has noted, this models is capable of fitting many types of shapes in the term structure. This greater flexibility appears to be rewarded by a competitive out-of-sample forecasting performance. Unlike the other models that utilize information about the shape of the term structure, this model outperforms the random walk by both the RMSFE and MAFE metrics at $h = 1$; however, it does not improve on the random walk benchmark's forecasting performance at longer horizons. Particularly interesting is the fact that there is little or no improvement in model's performance over that of the Diebold-Li model, which allows for considerably less flexibility in shape of the yield curve. Again, this suggests that information about the shape of the yield curve is relatively uninformative for predicting future short-term interest rates. Figure 6 provides a pictorial representation of forecast errors for the two essentially affine models in Table 6. Because the unrestricted essentially affine model performs better than any of the other models considered at the 1-month horizon, Figure 6 presents the forecast errors for 1-month rates at a

[26] Complete results for both restricted and unrestricted models are available from the Author(s) upon request.

[27] The performance of a restricted version of this model is largely similar and omitted to save space.

1-month horizon and as well as those for the 3-month rate at a 15-month horizon.

Overall, the results presented in Table 6 suggest that for the purposes of forecasting, completely affine models are essentially useless.[28] Even the simplest, most naive rule—a random walk—dominates the explanatory power of completely affine models. A corollary is that we should not use completely affine models to attempt to understand why the expectations hypothesis fails, because the models cannot reproduce this failure. By contrast, forecasts from a purely Gaussian essentially affine model dominate naive forecasts at least at short forecast horizons. However, for longer forecast horizons the superior performance of this model over either the random walk or the theoretical EH-implied models is less clear.

6. **Tests of Differences in Out-of-Sample Predictive Accuracy**

There are at least two issues that need to be taken into account to determine whether any of the differences in the forecasting performance noted above can be reliably exploited. The first problem is sampling variation: in the presence of rather small differences between random walk vs. other model performances, it is possible that our finding in favor of either EH-based or of affine-type, no-arbitrage frameworks may be mostly due to pure chance. We therefore test for the statistical significance of differences between the theoretical, the random walk, and the econometric model forecasts using the Diebold and Mariano (1995) test. The second problem stems from the fact that the Diebold-Li and the affine econometric require parameters to be estimated. Indeed, the affine models are richly parameterized. In contrast, the random walk benchmark has no parameters to be estimated, while our EH-based forecasts are based on estimates of a just a few moments (the risk premiums). A sensible procedure to test for the existence of statistically significant differences in predictive accuracy should to take these differences into account. We accomplish by implementing McCracken's (2004) nonparametric test for non-nested models, which takes into account the incremental variation in forecast errors due to parameter uncertainty.[29] We also use the more familiar Diebold and Mariano (1995) test.

The Diebold and Mariano (1995) (henceforth, DM) test is centered around the use of the statistic (for a pair of models indexed as k_1 and k_2)

$$DM^{k_1,k_2} \equiv \frac{\overline{d}}{\sqrt{\widehat{Var}(\overline{d})}}, \tag{26}$$

where $\overline{d}$ is an average over P observations of the values taken by some *differential* loss function, $d_t \equiv$

[28] The empirical rejection and poor forecasting performance of completely affine models is relatively unsurprising in the light of the literature (e.g., see Singleton, 2006).

[29] When parameters are not known but must instead be estimated, West (1996) provides analytical tools that can be used to construct tests of equal forecast accuracy between non-nested models. His results are similar to those in Diebold and Mariano (1995) but require that the loss function used to measure forecast accuracy must be continuously differentiable. McCracken (2000) extends the results of West (1996) to situations where the loss function need not be continuously differentiable (but the expected loss is continuously differentiable). McCracken (2004) provides a method of accounting for the effects of estimation error using numerical methods and without making strong assumptions about the observables or without having to derive the functional form of certain derivatives analytically.

$L(e_{t,h}^{n,k_1}) - L(e_{t,h}^{n,k_2})$ where $L(\cdot)$ is a generic loss function, and $Var(\overline{d})$ is the sample variance of $\overline{d}$. The DM statistic has an asymptotic standard normal distribution under the null hypothesis that $E[d] = 0$, which corresponds to a null of no differential of predictive accuracy between a pair of models/forecasting frameworks. Following standard practice, the variance of $\overline{d}$ is estimated using a heteroskedastic-autocorrelation consistent estimator,

$$\widehat{Var}(\overline{d}) = P^{-1}\left[\hat{\varphi}_0 + 2P^{-1}\sum\nolimits_{j=1}^{K-1}(P-K)\hat{\varphi}_j\right], \tag{27}$$

where $\hat{\varphi}_j \equiv (P-j)^{-1}\sum_{t=j+1}^{P}(d_t - \bar{d})(d_{t-j} - \bar{d})$. Based on the findings of Harvey, et al. (1997) the modified Diebold-Mariano test,

$$MDM^{k_1,k_2} \equiv \left[\frac{P+1-2K+P^{-1}K(K-1)}{P}\right]^{-1/2} DM^{k_1,k_2} \tag{28}$$

is used. The MDM statistic corrects for size distortions associated with the DM statistic.[30]

West (1996) has shown that in general, when loss functions depend on estimated parameters, (27) provides a valid estimate of the asymptotic variance of $\overline{d}$ only is special circumstances, e.g., when the models are estimated consistently by OLS and the loss function is a squared function (i.e., when we evaluate forecast accuracy using MSFE). In general, however, the structure of $Var(d)$ is

$$Var(d) = Var(\hat{d}) + 2\lambda_{dm}(\mathbf{FB}Cov'(d,m)) + \lambda_{mm}\mathbf{FB}Var(m)\mathbf{B}'\mathbf{F}', \tag{29}$$

where in our case of a recursive forecasting exercise, $\lambda_{dm} = 1 - R/P\ln(1+P/R)$, $\lambda_{mm} = 2[1 - R/P\ln(1+P/R)]$, $P = 264 - m$ (the number of recursive pseudo out-of-sample forecasts), and $R = 120$ (the training sample used in estimation). **F** and **B** are matrices that depend on the data used in estimation as well as on derivatives of the loss functions with respect to unknown parameters to be computed in correspondence to the true but unknown population parameters.[31] Finally, m denotes the time series of the scores generated by each model, when estimation occurs by QML. McCracken (2004) proposes to estimate **F** without deriving the functional form for the derivative of the loss function or making strong assumptions about the joint distribution of the observables. The idea is that unknown derivatives can be approximated numerically by using the finite difference method.

We consider two differential loss functions, the absolute forecast error and the squared forecast error. Tables 7 and 8 present the test results for squared forecast error and absolute forecast error loss functions, respectively. The numbers above the diagonal report the standard MDM test statistics and the corresponding significance level in parentheses. The numbers below the diagonal report the West-McCracken test statistics, again with the corresponding significance level in parentheses. The tests are reported for 1-, 6-, and 15-month horizon (the first exercise refers to 1-month T-bill rates, the latter two exercises to 3-month T-bill rates). Finally, to save space, we limit the exercises in this Section to 7 models, the two

[30] Harvey et al. (1997, 1998) also recommend using the critical values from the Student's t distribution rather than those from the normal distribution. The sample sizes used here are large enough, however, that the distinction is trivial.

[31] Exact definitions can be found in McCracken (2000, 2004).

EH-implied forecast models, the random walk, Diebold and Li's, and three representative affine models, including the best performing essentially affine Gaussian with $L = 0$. For convenience, all instances where the null hypothesis of equal predictive accuracy is reject with a p-value below 0.1 are in bold typeface. Finally, for numbers above the main diagonal, a negative (positive) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column. For numbers below the main diagonal, the interpretation is reverse: a positive (negative) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column. For example, the MDM test for the theoretical-time-varying-risk-premium row and the random-walk column of Table 7, Panel A is 0.439, indicating that the random walk model produced a less accurate forecast. Correspondingly, the value of the West-McCracken test statistic in the random-walk row and the theoretical-time-varying-risk-premium column of Panel A is also positive, 0.377, indicating that the theoretical time-varying risk premium model produced the superior forecast. Neither test statistic is statistically significant at any reasonable significance level, however. Hence, both tests indicate that null of equal forecasting power cannot be rejected at the 1-month horizon using the squared forecast error loss function.

While the tables report the test statistics for all pair-wise model comparisons, we focus on comparisons of the other model with the random walk benchmark. The results in Table 7 indicate that none of the models produce statistically significantly better forecast at the 1-, 6-, and 15-month horizons than the random walk benchmark using the mean squared error metric. There are six instances when the null hypothesis is rejected by the MDM test at one of the horizons considered using the squared forecast error metric. Not surprisingly, the corresponding West-McCracken test statistics are uniformly smaller. Moreover, there was no instance where the null hypothesis was rejected. Hence, this tests suggests that all of the models had equal forecasting ability at all horizons using the square error forecast metric.

Table 8 reports the test results using the absolute forecast error metric. As was the case with the squared forecast error metric, there was no instance where any of the other models produced a forecast that was statistically significantly different from the forecasts from the random walk benchmark. There were, however, three instances when one of the other models was statistically superior to another at the 0.10 percent significance level using the West-McCracken test. All of these occurred at the 1-month horizon. The test results indicate that constant-risk-premium theoretical model was statistically superior to the unrestricted essentially affine model, but inferior to the restricted essentially affine model. The test also indicated that the Diebold-Li model was statistically superior to the restricted essentially affine model. Not surprisingly, given the results in Table 6, the West-McCracken test indicates that the forecasting ability of the unrestricted essentially affine model was superior to either the completely affine model or the restricted essentially affine model, both at a very low significance level.

Of particular interest is the fact that the theoretical forecasts allowing for significant time variation in the risk premiums were not statistically significantly different from those based on a constant risk premium at any horizon, using either metric, or either test. This result suggests that the forecast errors from these theoretical models are dominated by response of rates to new information (i.e., "news"), which is essentially unpredictable. From this perspective, the fact that none of the models are able to generate forecast which dominate those from a random walk model is not surprising in that it suggests that the empirical failure of the EH likely stems from the fact that the short-term rates are largely unpredictable beyond their current level. As a result, single equation tests of the EH that are derived using the ex-post short-term rate as proxy for the markets' ex-ante expectation of the future short-term rate (2), could fail because this assumption is greatly at odds with the market's ability to forecast the future short-term rate. The results presented here suggest that the best proxy for the markets' ex-ante expectation of the future short-term rate is the current short-term rate, not the short-term rate that actually materialized $h-$periods in the future as (2) counterfactually assumes. Hence, the empirical failure of the EH seems likely due to fact that interest rates are essentially unpredictable rather than to massive time variation in risk premia.

7. **Summary, Conclusions and Implications**

This paper notes that conventional tests of the EH are based on two assumptions: that long-term rates are equal to the average of the market's expectation of the short-term rate over the holding period of the long-term assets plus a constant risk premium and an assumption about the market's expectation of the future short-term rate. We investigate the possibility that the massive empirical rejections of the EH found in the literature are due the latter assumption being inconsistent with the market's true expectation generating process, rather than to a failure of the EH per se. We do this by comparing the out-of-sample forecasting performance of several interest rate forecasting models. We find that differences in forecasting performance of the models considered appear to be relatively small. This is especially true at longer horizons. Moreover, none of the models outperforms the simple random walk benchmark by all of the metrics and at all forecast horizons. This finding is consistent with the forecasting performance of survey forecasts (e.g., Stark, 2010; Mitchell and Pearce, 2007; and Greer, 2003).

It is also the case that our EH-consistent forecast, which incorporate no information about the term structure of interest rates performs better than models which incorporate significant term structure information, suggesting that information about the term structure is relatively unimportant for forecasting the future interest rate. However, only in a few cases was the improvement in the predictive performance of the EH-consistent forecasts statistically significant. Also, models that were flexible enough to fit yield curves with a wide variety of shapes frequently performed worse than models allowed considerably less flexibility in the shape of the yield curve; however, in no instance was the difference in performance statistically significant. Finally, models that impose the no-arbitrage condition did not forecast significantly better than models that did not.

It is interesting to note that performance differences between our EH-consistent forecast under the assumption that the risk premium is constant over the sample period and the one that allowed for considerable variation in the risk premium were very small, with neither model consistently performing better than the other. In no instance was the difference in the performance of these models statistically significant. Hence, our analysis supports the findings of Kozicki and Tinsley (2005) that the ubiquitous empirical failure of the EH may not due to time-variation in risk premia. Rather, the failure appears to be the consequence of the failure of market participants to forecast short-term rates in the manner assumed in conventional tests of the EH. The future behavior of short-term rates is determined by new information (i.e., news) that appears to be essentially unpredictable. This not only explains why the spread between the long-term and short-term rate is a relatively poor predictor of the future short-term rate, but why conventional tests of the EH consistently reject it.

We tested for statistical differences in forecasting performance of all of the models using both the modified Diebold-Mariano test and a West-McCracken test. The latter test allows for parameter uncertainty, but is computationally more burdensome. While relatively small in number, there were instances where the modified Diebold-Mariano test and the West-McCracken tests yielded different qualitative conclusions. Consistent with expectations, these differences involved comparisons of models with estimated parameters and the results always went in the direction of making the rejection of the null hypothesis of equal predictive accuracy more difficult using the West-McCracken test. Consequently, there is a benefit to bearing the additional costs associated with implementing the West-McCracken test.

Our conclusion that the empirical failure of the EH is likely due to the auxiliary assumption used to derive the conventional test than to the EH gives some hope that the forward guidance policies of some central banks may be successful. Indeed, there is evidence that the federal funds rate has recently become more predictable, at least at short horizons (e.g., Lange, et al., 2003 and Poole, et al., 2002) since the Fed began announcing its funds rate target in 1994. However, evidence by Andersson and Hofmann (2010), Goodhart and Lim (2008), and Rudebusch (2007), suggest that forward guidance has not increased the predictability of the policy rate beyond a month or two.

The finding here and elsewhere that the reaction to unpredictable news is dominant in determining future short-term rates has deep implications for policymakers and financial analysts. If the EH is true, the inability to predict the future short-term rate significantly beyond its current level would imply that the long-term rate is equal to the short-term rate plus a constant risk premium. Such a relationship appears to be inconsistent with the behavior of interest rates, however. Hence, this finding threatens the conventional theory of the term structure of interest rates. The problem, of course, is that theorists have yet to come up with a more appealing alternative. As Fuhrer (1996) has noted "The tendency to fall back on this paradigm [the EH] is so strong because candidates to replace it are so weak."[32] The problem is the profession has a theoretically acceptable theory of the term structure that is at odds with both empirical tests of it and with extensive evidence that interest rates are extremely difficult to predict beyond their current level, but no acceptable theory that can adequately account for the observed behavior of interest rates.

[32] Fuhrer (1996), p. 1183.

Finally, we should note that despite efforts to implement a robust research design, there a number of extensions that could be explored in the attempt to look for cases in which the evidence of predictability in short-term rates may be stronger than what we have uncovered here. For instance, although our econometric benchmarks are inherently multivariate and have employed information from the entire term structure of interest rates, we have not explored the possibility that estimating and exploiting the existence of cointegrating relationships may improve forecast accuracy (see e.g., Hall, et al., 1992). Additionally, our econometric models have linked forecasts to (unobservable) features of the term structure. There is voluminous work in finance on the presence of non-linear dynamics in the latent factors that characterize the term structure (see, e.g., Ang and Bekaert, 2002; Bansal and Zhou, 2002; Engle and Ng, 1993; Hess and Kamara, 2005) with applications to forecasting (e.g., Guidolin and Timmermann, 2009). There is also an expanding literature on the possibility that macroeconomic factors may be suitable drivers for modeling and forecasting riskless yields in addition to standard latent factors (see, e.g., Ang and Piazzesi, 2003; Bekaert, Cho, and Moreno, 2010; Diebold et al. 2006; Favero et al., 2007; Wu, 2006; and Spencer, 2008). Although our linear affine models may in principle capture the idea that US yield curves may contain multiple, complex multi-factor structures that may lead to the possibility that long-term rates may be useful in allowing the investors to extract the dynamics of the latent factors to forecast future short-term rates, a latent factor strategy may offer additional payoffs also in terms of prediction accuracy.

References

[1] Andersson, M. and B., Hoffman. (2010) "Gauging the Effectiveness of Central Bank Forward Guidance," forthcoming in *Inflation Targeting Twenty Years On: Past Lessons and Future Prospects*, edited by D. Cobham, A. Eitrheim, S. Gerlach and J. Qvigstad, Cambridge University Press.

[2] Ang, A. and G., Bekaert. (2002) "Short Rate Nonlinearities and Regime Switches," *Journal of Economic Dynamics and Control* 26, 1243-1274.

[3] Ang., A. and M., Piazzesi. (2003) "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics* 50, 745-787.

[4] Balduzzi, P., G., Bertola and S., Foresi. (1997) "A Model of Target Changes and the Term Structure of Interest Rates," *Journal of Monetary Economics* 39, 223-249.

[5] Bali, T., M., Heidari, and L., Wu. (2009) "Predictability of Interest Rates and Interest-Rate Portfolios," *Journal of Business and Economic Statistics* 27, 517-527.

[6] Bansal, R., and H., Zhou. (2002) "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance* 57, 1997-2043.

[7] Bekaert, G., S., Cho and A., Moreno. (2010) "New Keynesian Macroeconomics and the Term Structure," *Journal of Money, Credit and Banking* 42, 33-62.

[8] Bekaert, G., R., Hodrick and D., Marshall. (1997) "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics* 44, 309-348.

[9] Bekaert, G., R., Hodrick and D., Marshall. (2001) "Peso Problem Explanations for Term Structure Anomalies, *Journal of Monetary Economics* 48, 241-270.

[10] Bliss, R. (1997) "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research* 9, 197-231.

[11] Campbell, J. and R., Shiller. (1991) "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies* 58, 495-514.

[12] Carriero, A., C., Favero and I., Kaminska. (2006) "Financial Factors, Macroeconomic Information and the Expectations Theory of the Term Structure of Interest Rates," *Journal of Econometrics* 131, 339-358.

[13] Chen, R., and L., Scott. (1993) "Maximum Likelihood Estimation for a Multi-Factor Model of the Term Structure of Interest Rates," *Journal of Fixed Income* 4, 14-31.

[14] Cochrane, J., and M., Piazzesi. (2005) "Bond Risk Premia," *American Economic Review* 95, 138-60.

[15] Cox, J., J., Ingersoll and S., Ross. (1985). "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385-407.

[16] Dai, Q., and K., Singleton. (2002) "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics* 63, 415-441.

[17] Dai, Q., and K., Singleton. (2000) "Specification Analysis of Affine Term Structure Models," *Journal of Finance* 55, 1943-1978.

[18] Deaves, R. (1996). "Forecasting Canadian Short-Term Interest Rates," *Canadian Journal of Economics* 29, 615-634.

[19] Della Corte, P., L., Sarno and D., Thornton. (2008) "The Expectations Hypothesis of the Term Structure of Very Short-Term Rates: Statistical Tests and Economic Value," *Journal of Financial Economics* 89, 158-74.

[20] Diebold, F., and C., Li. (2006) "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics* 130, 337-364.

[21] Diebold, F., and R., Mariano. (1995) "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

[22] Diebold, F., G., Rudebusch and S., Aruoba. (2006) "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics* 131, 309-338.

[23] Driffill, J., Z., Psaradakis, and M., Sola. (1997) "A Reconciliation of Some Paradoxical Empirical Results on the Expectations Model of the Term Structure," *Oxford Bulletin of Economics and Statistics* 59, 29-42.

[24] Duffee, G. (2002) "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance* 57, 405-443.

[25] Engle, R., and V., Ng (1993) "Time Varying Volatility and the Dynamic Behavior of the Term Structure," *Journal of Money, Credit and Banking* 25, 336-349.

[26] Fama, E. (1976) "Forward Rates as Predictors of Future Spot Rates," *Journal of Financial Economics* 3, 361-377.

[27] Fama, E. (1984) "The Information in the Term Structure," *Journal of Financial Economics* 13, 509-528

[28] Fama, E., and R., Bliss. (1987) "The Information in Long Maturity Forward Rates," *American Economic Review* 77, 680-692.

[29] Favero C., L., Niu and L., Sala. (2007) "Term Structure Forecasting: No-arbitrage Restrictions vs. Large Information Sets," mimeo, IGIER Bocconi University.

[30] Froot, K. (1989) "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Finance* 44, 283-305.

[31] Fuhrer, J. (1996) "Monetary Policy Shifts and Long-Term Interest Rates," *The Quarterly Journal of Economics* 111, 1183-1209.

[32] Goodhart, C., and W., Lim. (2008) "Interest Rate Forecasts: A Pathology," mimeo, Bank of England.

[33] Greer, M. (2003). "Directional Accuracy Tests of Long-Term Interest Rate Forecasts," *International Journal of Forecasting*, 19, 291-98.

[34] Guidolin, M., and A., Timmermann. (2009) "Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach," *Journal of Econometrics* 150, 297-311.

[35] Hall, A., H., Anderson and C., Granger. (1992) "A Cointegration Analysis of Treasury Bill Yields," *The Review of Economics and Statistics* 74, 116-26.

[36] Hamburger, M., and E., Platt. (1975) "The Expectations Hypothesis and the Efficiency of the Treasury Bill Market," *Review of Economics and Statistics* 57, 190-199.

[37] Hardouvelis, G. (1994) "The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries: Is There a Puzzle?" *Journal of Monetary Economics* 33, 255-283.

[38] Harvey, D., S., Leybourne and P., Newbold. (1997) "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting* 13, 281-291.

[39] Hess, A., A., Kamara. (2005) "Conditional Time-Varying Interest Rate Risk Premium: Evidence from the Treasury Bill Futures Market," *Journal of Money Credit and Banking* 37, 679-698.

[40] Kocherlakota, N. (2010) "Economic Outlook and the Current Tools of Monetary Policy," a speech presented at the European Economics and Financial Centre, London, September 29, 2010.

[41] Kool, C., and D., Thornton. (2004). "A Note on the Expectations Hypothesis at the Founding of the Fed," *Journal of Banking and Finance* 28, 3055-3068.

[42] Kozicki, S. and P., Tinsley. (2001). "Shifting endpoints in the term structure of interest rates," *Journal of Monetary Economics*, 47 613-52.

[43] Kozicki, S. and P., Tinsley. (2005). "What Do You Expect? Imperfect Policy Credibility and Tests of the Expectations Hypothesis," *Journal of Monetary Economics*, 52, 421-47.

[44] Lange, J., B., Sack and W., Whitesell. (2003) "Anticipations of Monetary Policy in Financial Markets," *Journal of Money, Credit and Banking* 35, 889-910.

[45] Littermann, R., and J., Scheinkman. (1991) "Common Factors Affecting Bond Returns," *Journal of Fixed Income* 1, 54-61.

[46] MacDonald, S., and S., Hein. (1989). "Futures Rates and Forward Rates as Predictors of Near-Term Treasury Bill Rates," *Journal of Futures Markets* 9, 249-262.

[47] Macaulay, F. (1938) Some Theoretical Problems Suggested by Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856, National Bureau of Economic Research, New York.

[48] Mankiw, N., and J., Miron. (1986) "The Changing Behavior of the Term Structure of Interest Rates," *The Quarterly Journal of Economics* 101, 211-28.

[49] McCracken, M. (2000) "Robust Out-of-Sample Inference," *Journal of Econometrics* 99, 195-223.

[50] McCracken, M. (2004) "Parameter Estimation and Tests of Equal Forecast Accuracy Between Non-nested Models," *International Journal of Forecasting* 20, 503-514.

[51] Mitchell, K., and D.K. Pearce. (2007). "Professional forecasts of interest rates and exchange rates: Evidence from the Wall Street Journal's panel of economists," *Journal of Macroeconomics*, 29, 840-54.

[52] Nelson, C., and A., Siegel. (1988) "Long-Term Behavior of Yield Curves," *Journal of Financial and Quantitative Analysis* 23, 105-10.

[53] Nelson, C., and A., Siegel. (1987) "Parsimonious Modeling of Yield Curves," *Journal of Business* 60, 473-489.

[54] Park, T., and L., Switzer. (1997) "Forecasting Interest Rates and Yield Spreads: The Informational Content of Implied Futures Yields and Best-Fitting Forward Rate Models," *Journal of Forecasting* 16, 209-224.

[55] Poole, W., R., Rasche and D., Thornton. (2002) "Market Anticipations of Monetary Policy Actions," *Federal Reserve Bank of St. Louis Review* 84, 65-93.

[56] Roberds, W., D., Runkle and C., Whiteman. (1996) "A Daily View of Yield Spreads and Short-Term Interest Rate Movements," *Journal of Money, Credit and Banking* 28, 34-53.

[57] Roberds, W., and C., Whiteman. (1999) "Endogenous Term Premia and Anomalies in the Term Structure of Interest Rates: Explaining the Predictability Smile," *Journal of Monetary Economics* 44, 555-580.

[58] Rudebusch, G. (1995) "Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure," *Journal of Monetary Economics* 35, 245-274.

[59] Rudebusch, G. (2007) "Monetary Policy Inertia and Recent Fed Actions," *Economic Letter 26*, Federal Reserve Bank of San Francisco, Issue Jan 26.

[60] Sarno, L., D., Thornton and G., Valente. (2007) "The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields," *Journal of Financial and Quantitative Analysis* 42, 81-100.

[61] Shiller, R., J., Campbell and K., Schoenholtz. (1983) "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates," *Brookings Papers on Economic Activity* 1, 173-217.

[62] Simon, D. (1990) "Expectations and the Treasury Bill-Federal Funds Rate Spread over Recent Monetary Policy Regimes," *Journal of Finance* 45, 567-577.

[63] Singleton, K. (2006) *Empirical Dynamic Asset Pricing*, Princeton and Oxford: Princeton University Press.

[64] Spencer, P. (2008). "Stochastic Volatility in a Macro-Finance Model of the U.S. Term Structure of Interest Rates 1961-2004," *Journal of Money, Credit, and Banking* 40, 1177-1215.

[65] Stark, T. (2010). "Realistic Evaluation of Real-Time Forecasts in the Survey of Professional Forecasters," Federal Reserve Bank of Philadelphia *Research Rap, Special Report*, May 28, 2010.

[66] Thornton, D. (2006) "Tests of the Expectations Hypothesis: Resolving the Campbell-Shiller Paradox," *Journal of Money, Credit, and Banking* 38, 2039-2071.

[67] Thornton, D. (2005). "Tests of the Expectations Hypothesis: Resolving the Anomalies when the Short-Term Rate Is the Federal Funds Rate," *Journal of Banking and Finance* 29, 2541-2556.

[68] Tzavalis, E., and M., Wickens. (1997) "Explaining the Failures of the Term Spread Models of the Rational Expectations Hypothesis of the Term Structure," *Journal of Money, Credit and Banking* 29, 364-380.

[69] Vasicek, O. (1977) "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics* 5, 177-188.

[70] Waggoner, D. (1997) "Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices," Federal Reserve Bank of Atlanta Working Paper 97-10.

[71] West, K. (1996) "Asymptotic Inference about Predictive Ability," *Econometrica* 64, 1067-84.

[72] Woodford, M. (1999) "Optimal Monetary Policy Inertia," *The Manchester School*, 67 (special supplement), 1-35.

[73] Woodford, M. (2003) "Central Bank Communication and Policy Effectiveness," in *The Greenspan Era: Lessons for the Future*, A Symposium Sponsored by The Federal Resreve Bank of Kansas City, 339-474.

[74] Wu, T. (2006). "Macro Factors and the Affine Term Structure of Interest Rates," *Journal of Money, Credit and Banking* 38, 1848-1873.

Appendix: Details on the Estimation of Models in the Essentially Affine Class (Not for Publication)

Consider the affine term structure model

$$d\mathbf{x}_t = \underset{M\times M}{\mathcal{K}}(\underset{M\times 1}{\boldsymbol{\theta}} - \underset{M\times 1}{\mathbf{x}_t})dt + \underset{M\times M}{\mathbf{S}_t^{1/2}}\underset{M\times 1}{d\mathbf{W}_t} \qquad [\mathbf{S}_t^{1/2}]_{ii} = \sqrt{\alpha_i + \boldsymbol{\xi}_i'\mathbf{x}_t} \quad i = 1, ..., M, \tag{30}$$

where $\mathbf{W}_t$ is a $M \times 1$ vector of independent Brownian motions. When the pricing kernel $\mathcal{M}$ has structure

$$d\mathcal{M}_t = -r_t\mathcal{M}_t dt - \mathcal{M}_t\boldsymbol{\Lambda}_t' d\mathbf{W}_t \tag{31}$$

(where $\boldsymbol{\Lambda}_t$ is the $M \times 1$ vector of prices of risk), the physical measure representation of the stochastic process for the state vector is

$$\begin{gathered} d\mathbf{x}_t = \boldsymbol{\mu}^{\mathbb{P}}(\mathbf{x}_t, t)dt + \mathbf{S}_t^{1/2} d\mathbf{W}_t^{\mathbb{P}} \\ \boldsymbol{\mu}^{\mathbb{P}}(\mathbf{x}_t, t) = \mathcal{K}(\boldsymbol{\theta} - \mathbf{x}_t) - \mathbf{S}_t^{1/2}\boldsymbol{\Lambda}_t \qquad [\mathbf{S}_t^{1/2}]_{ii} = \sqrt{\alpha_i + \boldsymbol{\xi}_i'\mathbf{x}_t} \quad i = 1, ..., M. \end{gathered} \tag{32}$$

in completely affine case where

$$\boldsymbol{\Lambda}_n(\mathbf{x}_t) = \lambda_n(\alpha_n + \mathcal{B}_n'\mathbf{x}_t)^{1/2} \tag{33}$$

and

$$\begin{aligned} d\mathbf{x}_t &= \mathcal{K}(\boldsymbol{\theta} - \mathbf{x}_t)dt - \{\mathbf{S}_t\boldsymbol{\lambda}_1 + \mathbf{I}^-\boldsymbol{\Lambda}_2(\mathbf{x}_t)\}dt + \mathbf{S}_t^{1/2} d\mathbf{W}_t^{\mathbb{P}} \\ \mathbf{I}_{ii}^- &\equiv \begin{cases} 1 & \inf_i(\alpha_i + \mathcal{B}_i'\mathbf{x}_t)) > 0 \\ 0 & \text{otherwise} \end{cases} \qquad i = L+1, ..., M \end{aligned} \tag{34}$$

in the essentially affine case, where:

$$\Lambda_n(\mathbf{x}_t) = \lambda_{1n}(\alpha_n + \mathcal{B}_n'\mathbf{x}_t)^{1/2} + \boldsymbol{\lambda}_{2n}' \left[\begin{cases} \frac{1}{(\alpha_n+\mathcal{B}_n'\mathbf{x}_t)^{1/2}} & \inf_i(\alpha_i + \mathcal{B}_i'\mathbf{x}_t) > 0 \\ 0 & \text{otherwise} \end{cases} \right] \mathbf{x}_t. \tag{35}$$

Here $diag(\cdot)$ denotes the operator that turns an $L \times 1$ vector into an $L \times L$ diagonal matrix. As customary, in our work we use a normalization that sets the first L rows of $\mathbf{\Lambda}_2(\mathbf{x}_t)$ to zero vectors, i.e., for the first L factors the risk premia are indeed completely linear affine in $\mathbf{x}_t$. As a result, when $L = M$, an essentially affine model reduces to a completely affine one. However, when $L < M$, the essentially affine model introduces the possibility that $\mathbf{x}_t$ affects expected excess returns both non-linearly through the terms $\mathbf{I}^{-}\mathbf{\Lambda}_2(\mathbf{x}_t)$ and linearly through the non-zero elements of $\boldsymbol{\lambda}_1$. Crucially, the signs of the premia corresponding to the $M-L$ non-volatility factors may now switch over time, adding additionally variability to the signs of the excess returns.

From basic principles, in the absence of arbitrage opportunities, it must be that price of a bond $h+1$ periods to maturity is determined as $D_t^{h+1} = E_t[\mathcal{M}_{t+1}D_{t+1}^h]$. Let's conjecture that bond prices are log-linear functions of the state:

$$-\ln(D_t^h) = A_h + \mathbf{B}_h'\mathbf{x}_t \Longleftrightarrow D_t^h = \exp\left(-A_h - \mathbf{B}_h'\mathbf{x}_t\right). \tag{36}$$

This means that since a zero-coupon bond with zero-to-maturity left has a value of one ($D_t^0 = 1$), $A_0 = 0$, $\mathbf{B}_0 = \mathbf{0}$, $D_t^1 = E_t[\mathcal{M}_{t+1}]$, and

$$\exp\left(-A_1 - \mathbf{B}_1'\mathbf{x}_t\right) = E_t\left\{\exp\left[-\left(\delta_0 + \sum_{n=1}^{M}\delta_n x_{nt}\right) + \mathbf{\Lambda}(\mathbf{x}_t)\boldsymbol{\varepsilon}_{t+1}\right]\right\} = \exp\left[-\delta_0 - \sum_{n=1}^{M}\delta_n x_{nt} + \frac{1}{2}\sum_{n=1}^{M}\mathbf{\Lambda}_n^2(\mathbf{x}_t)\right], \tag{37}$$

which implies (by matching the coefficients) that

$$A_1 = \delta_0 - \frac{1}{2}\sum_{n=1}^{M}\lambda_{1n}^2\alpha_n \quad \mathbf{B}_1 = \boldsymbol{\delta} - \frac{1}{2}\sum_{n=1}^{M}\tilde{\Lambda}_n^2(\mathbf{x}_t), \tag{38}$$

where $\tilde{\Lambda}_n^2(\mathbf{x}_t)$ is the part of $\Lambda_n^2(\mathbf{x}_t)$ that only depends on $\mathbf{x}_t$; obviously, $\tilde{\Lambda}_n(\mathbf{x}_t) = \lambda_n\mathcal{B}_n'\mathbf{x}_t$ in the completely affine case. In general, it is easy to recognize that a general set of recursive conditions obtains,

$$A_{h+1} = A_h + \delta_0 + \mathbf{B}_h'\mathcal{K}\boldsymbol{\theta} - \frac{1}{2}\sum_{n=1}^{M}(\Lambda_n^2 + B_{hn}^2)\alpha_n \qquad \mathbf{B}_{h+1} = \boldsymbol{\delta} + \mathbf{B}_h'(\mathbf{I}_M - \mathcal{K}) - \frac{1}{2}\sum_{n=1}^{M}(\tilde{\Lambda}_n^2(\mathbf{x}_t) + \mathbf{B}_{hn}^2\mathbf{s}_n), \tag{39}$$

where where $\mathbf{s}_n \equiv [\mathbf{e}_n' \ \boldsymbol{\beta}_n']$. Clearly, higher compensations for risk $\Lambda_n(\mathbf{x}_t)$ and $\tilde{\Lambda}_n(\mathbf{x}_t)$ $(n = 1, ..., M)$ imply lower values of A_h and $\mathbf{B}_h$ and—provided that some conditions on state variables and parameters are satisfied—lower bond prices and higher bond yields. The bond yields for a given maturity h turn out to be themselves linear affine in $\mathbf{x}_t$ as

$$y_t^h = \frac{1}{h}\ln\left(\frac{1}{D_t^h}\right) = -\frac{1}{h}\ln D_t^h = \frac{1}{h}A_h + \frac{1}{h}\mathbf{B}_h'\mathbf{x}_t. \tag{40}$$

One implication is that—assuming the matrix with rows $\mathbf{B}_h'$ is invertible for some set of L bond maturities $(h = 1, ..., L)$—then $\mathbf{x}_t$ can be expressed as a function of L bond yields measured in the market, so that these L bond yields span the risks underlying the variation in the term structure.

As far the estimation is concerned, we adopt Duffee's (2002) QMLE method. Assume that at each month-end t, $t = 1, ..., T$, *yields* on M bonds are measured without error. These bonds have fixed times to maturity h_1, ..., h_M. Yields on $N - M$ other bonds are assumed to be measured with serially uncorrelated, mean-zero measurement errors.[33] Stack the perfectly observed yields in the vector $\mathbf{Y}_t$ and the imperfectly observed yields in the vector $\check{\mathbf{Y}}_t$. Denote the parameter vector by $\boldsymbol{\theta}$. Given $\boldsymbol{\theta}$, $\mathbf{Y}_t$ can be inverted using

$$y_t^{h_i} = \frac{1}{h_i} A_{h_i} + \frac{1}{h_i} \mathbf{B}'_{h_i} \mathbf{x}_t \quad i = 1, ..., M \tag{41}$$

to form an implied state vector series $\hat{\mathbf{x}}_t$,

$$\mathbf{Y}_t = \ddot{\mathbf{a}}_{h_i} + \ddot{\mathbf{B}}'_{h_i} \hat{\mathbf{x}}_t \Longrightarrow \hat{\mathbf{x}}_t = \left(\ddot{\mathbf{B}}'_{h_i}\right)^{-1} (\mathbf{Y}_t - \ddot{\mathbf{a}}_{h_i}) \quad i = 1, ..., M, \tag{42}$$

where the generic elements of the vector $\ddot{\mathbf{a}}_{h_i}$ are $h_i^{-1} A_{h_i}$, and in the case of $\ddot{\mathbf{B}}_{h_i}$ they are $h_i^{-1} \mathbf{B}_{h_i}$. The candidate parameter vector is required to be consistent with $\mathbf{Y}_t$. This is enforced by requiring $\hat{\mathbf{x}}_t$ to be in the admissible space for $\mathbf{x}_t$, which is equivalent to requiring that the diagonal elements of $\mathbf{S}(\mathbf{x}_t)$ be real. Given $\hat{\mathbf{x}}_t$, implied yields for the other $N - M$ bonds can be calculated. Stack them in $\check{\mathbf{Y}}_t$. The measurement error is $\boldsymbol{\epsilon}_t = \hat{\mathbf{Y}}_t - \check{\mathbf{Y}}_t$, where $\hat{\mathbf{Y}}_t$ collects the fitted values from

$$\hat{y}_t^{h_i} = \frac{1}{h_i} A_{h_i} + \frac{1}{h_i} \mathbf{B}'_{h_i} \left(\ddot{\mathbf{B}}'_{h_i}\right)^{-1} (\mathbf{Y}_t - \ddot{\mathbf{a}}_{h_i}) \quad i = M + 1, ..., N. \tag{43}$$

The variance-covariance matrix of the measurement error is assumed to have the following time-invariant Cholesky decomposition: $E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t] = \mathbf{C}\mathbf{C}'$. To compute the quasi-likelihood value, we assume that the one-period-ahead conditional distribution of the state variables $f_X(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is multivariate normal. The mean and variance-covariance matrix of $\mathbf{x}_{t+1}$ are known; thus, $f_X(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is known. Then the distribution of $\mathbf{Y}_{t+1}$ conditional on $\mathbf{Y}_t$ is

$$f_Y(\mathbf{Y}_{t+1}|\mathbf{Y}_t) = \frac{1}{\left|\det(\ddot{\mathbf{B}}'_{h_i})\right|} f_X(\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t). \tag{44}$$

Also, we have assumed in the main text that the measurement error is jointly normally distributed with distribution $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t)$. The log-likelihood of observation t for $\check{\mathbf{Y}}_t$ is then $\ell_t(\boldsymbol{\theta}) = \ln f_Y(\mathbf{Y}_t|\mathbf{Y}_{t-1}) + \ln f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t)$. Stationarity is imposed by requiring that the eigenvalues of a characteristic matrix $\mathcal{K} = \mathbf{N}\mathbf{D}\mathbf{N}^{-1}$ (see below for the definition of $\mathbf{N}$ and $\mathbf{D}$) are positive, allowing $f_Y(\mathbf{Y}_1|\mathbf{Y}_0)$ to be set equal to the unconditional distribution of $\mathbf{Y}_t$. The estimated parameter vector $\hat{\boldsymbol{\theta}}_T^{QMLE}$ is chosen to solve

$$\max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \sum_{t=1}^{T} \left[\frac{1}{|\det(\ddot{\mathbf{B}}'_{\tau_i})|} f_X(\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t) + \ln f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_t) \right], \tag{45}$$

where $\boldsymbol{\epsilon}_t \sim IID\ N(\mathbf{0}, \mathbf{C}\mathbf{C}')$ and

$$\hat{\mathbf{x}}_{t+1}|\hat{\mathbf{x}}_t \sim N\left(\mathbf{N}(\mathbf{I}_N - e^{-\mathbf{D}(T-t)})\boldsymbol{\theta}^* + \mathbf{N}e^{-\mathbf{D}(T-t)}\mathbf{N}^{-1}\mathbf{x}_t,\ \mathbf{N}\boldsymbol{\Upsilon}_{\mathbf{0}}\mathbf{N}' + \mathbf{N}\left(\sum_{i=1}^{N} \boldsymbol{\Upsilon}_i \mathbf{N}^{-1} \mathbf{x}_{it}\right)\mathbf{N}'\right). \tag{46}$$

[33] The recorded prices in the data sets may not be actual market transaction prices or the prices of bonds along the yield curve may not have been recorded at precisely the same time. Alternatively, some have included measurement errors as a result of the explicit recognition of the fact that the pricing model is an approximation and cannot literally fit all market prices.

The yields free of measurement error are picked to span as much of the term structure as possible. At this point, it is tedious but possible to derive closed-form representations for the first and second conditional moments of a state vector that follows the process (16). Assume that $\mathcal{K}$ can be diagonalized, or $\mathcal{K}= \mathbf{N}\mathbf{D}\mathbf{N}^{-1}$, where $\mathbf{D}$ is diagonal. The diagonal elements of $\mathbf{D}$ are denoted $d_1 ..., d_N$. We follow Duffee's (2002) approach and compute the first and second conditional moments of a linear transformation of $\mathbf{x}_t$. The transformation is chosen so that the feedback matrix $\mathcal{K}$ is diagonal under the transformation. The linear transformation is then reversed to calculate the conditional moments of $\mathbf{x}_t$. Defining $\mathbf{x}_t^* \equiv \mathbf{N}^{-1}\mathbf{x}_t$, the dynamics of $\mathbf{x}_t^*$ is

$$d\mathbf{x}_t^* = (\mathbf{N}^{-1}\mathcal{K}\boldsymbol{\theta} - \mathbf{N}^{-1}\mathcal{K}\mathbf{x}_t)dt + \mathbf{N}^{-1}\mathbf{S}(\mathbf{x}_t)^{1/2}d\mathbf{W}_t = \mathbf{D}(\boldsymbol{\theta}^* - \mathbf{x}_t^*)dt + \mathbf{S}^*(\mathbf{x}_t)^{1/2}d\mathbf{W}_t, \tag{47}$$

where $\boldsymbol{\theta}^* \equiv \mathbf{N}^{-1}\boldsymbol{\theta}$, $\mathbf{S}^*(\mathbf{x}_t)^{1/2} \equiv \mathbf{N}^{-1}\mathbf{S}(\mathbf{x}_t)$. We now calculate the first and second moments of $\mathbf{x}_t^*$. The expectation of $\mathbf{x}_T^*$ conditional on $\mathbf{x}_t^*$ is given by:

$$E[\mathbf{x}_T^*|\mathbf{x}_t^*] = \boldsymbol{\theta}^* + e^{-\mathbf{D}(T-t)}(\mathbf{x}_t^* - \boldsymbol{\theta}^*) = (\mathbf{I}_N - e^{-\mathbf{D}(T-t)})\boldsymbol{\theta}^* + e^{-\mathbf{D}(T-t)}\mathbf{x}_t^*. \tag{48}$$

Here, if $\mathbf{Z}$ is a diagonal matrix, the diagonal matrix in which element (i,i) equals $e^{z_{ii}}$ is denoted $e^{\mathbf{z}}$; the N-vector $\mathcal{B}_{\cdot i}$ is column i of $\mathcal{B}$. Given this conditional mean of $\mathbf{x}_T^*$, we reverse the transformation to express the conditional mean of $\mathbf{x}_T$:

$$E[\mathbf{x}_T|\mathbf{x}_t] = \mathbf{N}E[\mathbf{x}_T^*|\mathbf{x}_t^*] = \mathbf{N}(\mathbf{I}_N - e^{-\mathbf{D}(T-t)})\boldsymbol{\theta}^* + \mathbf{N}e^{-\mathbf{D}(T-t)}\mathbf{N}^{-1}\mathbf{x}_t.$$

As for the conditional variance of $\mathbf{x}_T$, the matrix $\mathbf{S}^*(\mathbf{x}_t)\mathbf{S}^*(\mathbf{x}_t)'$ is the instantaneous variance-covariance matrix of the transformed state vector. We can write this as

$$\mathbf{S}^*(\mathbf{x}_t)\mathbf{S}^*(\mathbf{x}_t)' = diag(\boldsymbol{\alpha}^*) + \sum_{i=1}^{N} diag(\mathcal{B}_{\cdot i}^*)\mathbf{x}_{it}^* = \mathbf{G}_0 + \sum_{i=1}^{N} \mathbf{G}_i\mathbf{x}_{it}^*, \tag{49}$$

where $\mathbf{G}_0 \equiv diag(\boldsymbol{\alpha}^*)$ and the $n \times n$ matrices $\mathbf{G}_i$ $(i = 1, ..., n)$ are defined as $diag(\mathbf{B}_{\cdot i}^*)$. Define the $N \times N$ matrix $\mathbf{F}(t,s) \equiv \mathbf{G}_0 + \sum_{i=1}^{n} \mathbf{G}_i [E(\mathbf{x}_s^*|\mathbf{x}_t^*)]_i$. This matrix is the instantaneous variance-covariance matrix of $\mathbf{x}_s^*$, but evaluated at the expectation of $\mathbf{x}_s^*$ (conditional on time t information) instead of the true value of $\mathbf{x}_s^*$. Fisher and Gilles (1996) show the conditional variance of $\mathbf{x}_T^*$ can be written as

$$Var[\mathbf{x}_T^*|\mathbf{x}_t^*] = \int_t^T e^{-\mathbf{D}(T-s)}\mathbf{F}(t,s)e^{-\mathbf{D}(T-s)}ds. \tag{50}$$

Substituting the expression for $\mathbf{F}(t,s)$ into this equation, integrating the resulting expression, and going from $Var[\mathbf{x}_T^*|\mathbf{x}_t^*]$ to the conditional variance of $\mathbf{x}_T$, we obtain:

$$Var[\mathbf{x}_T|\mathbf{x}_t] = \mathbf{N}Var[\mathbf{x}_\mathbf{T}^*|\mathbf{x}_\mathbf{t}^*]\mathbf{N}' = \mathbf{N}\boldsymbol{\Upsilon}_\mathbf{0}\mathbf{N}' + \mathbf{N}\left(\sum_{i=1}^{n}\boldsymbol{\Upsilon}_i\mathbf{x}_{it}^*\right)\mathbf{N}' = \mathbf{N}\boldsymbol{\Upsilon}_\mathbf{0}\mathbf{N}' + \mathbf{N}\left(\sum_{i=1}^{n}\boldsymbol{\Upsilon}_i\mathbf{N}^{-1}\mathbf{x}_{it}\right)\mathbf{N}', \tag{51}$$

where the $n \times n$ matrices $\boldsymbol{\Upsilon}_i$ $(i = 0, ..., n)$ depend on the horizon $T - t$ and their expression can be found in Duffee (2002).

The quasi-likelihood functions implied by the typical affine models tend to have a large number of local maxima: similar quasi-likelihood values can be produced by very different interactions among the elements of the state vector. The most important reason for this is the lack of structure placed on the feedback matrix $\mathcal{K}\boldsymbol{\theta}$. Another difficulty is that any feasible parameter vector must satisfy the requirement that the diagonal elements of $\mathbf{S}_t^{1/2}$ are real for all t, which requires that $\hat{x}_{it} \geq 0$ for all t and $i = 1, ..., L$; this implies LT restrictions on the parameter vector which are very difficult to handle through standard methods. These problems led to implement the following maximization algorithm, similar to Duffee's (2002):

- **Step 1.** Randomly generate parameters from a multivariate normal distribution with a diagonal variance-covariance matrix. The means and variances are set to plausible values equal to the sample estimates of means and variances of 3-month, 2-year, and 5-year Treasury yields.

- **Step 2.** Use $\hat{\mathbf{x}}_t = (\ddot{\mathbf{B}}'_{h_i})^{-1} (\mathbf{Y}_t - \ddot{\mathbf{a}}_{h_i})$ to calculate $\hat{\mathbf{x}}_t$ for all t.

- **Step 3.** If the parameter vector is not feasible, return to step 1; otherwise proceed.

- **Step 4.** Use the simplex method to determine the parameter vector that maximizes the QML value.

- **Step 5.** Using the final parameter vector from Step 4 as a starting point, use numerical optimization (Broyden, Fletcher, Goldfarb, Shanno's algorithm) to make any final improvements to the QML value.

This procedure is repeated until Step 5 is reached 2,000 times. For most of the models, there was little improvement in the QML value after the first 50 to 60 iterations. For each of the models investigated, we have also analyzed the effects of estimating more parsimonious specifications by first computing the (Wald) t-statistics for the unrestricted parameter estimates, then setting to zero all parameters for which the absolute t-statistics did not exceed a 1.66 approximate threshold, before finally re-estimating the restricted model.

As far as forecasts are concerned, given an estimated parameter vector $\hat{\boldsymbol{\theta}}$ associated with a particular model, the implied state vector $\hat{\mathbf{x}}_t$ is given by inverting yields observed at time t. The h-period ahead conditional mean $E[\hat{\mathbf{x}}_{t+h}|\hat{\mathbf{x}}_t]$ can then be constructed. Given this expected state vector, expected h-period ahead bond yields and associated forecast errors can also be constructed. However, differently from Duffee (2002), we forecast the entire term structure and not only the 3 maturities (3 months, 2 years, and 5 years) that we have assumed not to imply measurement errors in yields. Additionally, we perform a recursive pseudo out-of-sample exercise with a block structure, in the sense that parameter estimates are updated with bi-annual frequency, i.e., starting with 1972:01-1981:12, followed by 1972:01-1983:12, etc. up to 1972:01-2001:12.

Table 1

Summary Statistics for Implicit Zero-Coupon Yields in the US Term Structure

The table reports summary statistics for end-of-period monthly observations on continuously compounded yields on riskless pure discount bonds for the U.S. The raw data are from Bloomberg. The riskless pure discount bond yields were obtained using FORTRAN codes provided by Robert Bliss and Dan Waggoner based on Bliss (1997) and Waggoner (1997). The Sharpe ratio is computed with reference to the 1-month T-Bill. The Jarque Bera statistic is used to test the null that the level of bond yields is normally distributed. The Ljung-Box statistics (at 12 lags) test the presence of serial correlation in the first difference of yields and their squares.

Horizon in Months	Sample	Mean	Median	Standard Dev.	Term Premium	Sharpe Ratio (monhly)	Jarque-Bera	Ljung-Box(12) First Diff.	Ljung-Box(12) Squares
1	Jan. 1970 - Dec. 2003	6.042	5.430	2.790	__	__	88.44**	40.54**	189.7**
2	Jan. 1970 - Dec. 2003	6.189	5.538	2.844	0.147	0.015	82.33**	43.87**	140.9**
3	Jan. 1970 - Dec. 2003	6.318	5.690	2.890	0.276	0.028	76.67**	46.85**	142.0**
6	Jan. 1970 - Dec. 2003	6.544	6.046	2.932	0.503	0.049	60.20**	54.74**	176.0**
9	Jan. 1970 - Dec. 2003	6.648	6.152	2.902	0.606	0.060	48.59**	51.46**	166.2**
12	Jan. 1970 - Dec. 2003	6.750	6.273	2.861	0.708	0.071	41.02**	49.27**	136.2**
15	Jan. 1970 - Dec. 2003	6.864	6.394	2.823	0.822	0.084	36.76**	48.88**	119.3**
18	Jan. 1970 - Dec. 2003	6.957	6.537	2.794	0.916	0.095	35.55**	47.06**	115.8**
24	Jan. 1970 - Dec. 2003	7.052	6.668	2.730	1.010	0.107	37.74**	38.80**	129.2**
30	Jan. 1970 - Dec. 2003	7.146	6.752	2.658	1.104	0.120	38.44**	31.65**	147.6**
36	Jan. 1970 - Dec. 2003	7.244	6.814	2.604	1.202	0.133	39.73**	27.70**	152.1**
48	Jan. 1970 - Dec. 2003	7.400	6.943	2.521	1.358	0.155	46.00**	23.25*	144.4**
60	Jan. 1970 - Dec. 2003	7.508	7.040	2.459	1.466	0.172	48.83**	20.12	140.1**
72	Jan. 1970 - Dec. 2003	7.607	7.184	2.420	1.565	0.187	47.75**	16.75	140.6**
84	Aug. 1971 - Dec. 2003	7.726	7.311	2.431	1.684	0.200	39.09**	14.12	141.0**
96	Aug. 1971 - Dec. 2003	7.783	7.386	2.386	1.741	0.211	41.12**	13.14	137.6**
108	Aug. 1971 - Dec. 2003	7.826	7.496	2.333	1.784	0.221	43.39**	13.85	136.0**
120	Nov. 1971 - Dec. 2003	7.879	7.522	2.282	1.837	0.232	43.77**	16.16	130.6**

** indicates significance at size 1% or lower, * significance between 1 and 5%.

Table 2

Full-Sample Estimates of a Purely Affine (Gaussian) Model

The table reports QMLE estimates for a completely affine term structure model obtained from a common data sample that spans the interval 1972:01 – 2003:12. The estimation procedure assumes that nominal yields on 3-month, 2-year, and 5-year Treasury bonds are measured without error, while for all other maturities yields are measured with errors that have a joint i.i.d. Gaussian distribution with constant covariance matrix. Robust (sandwich-style) standard errors are reported in parenthesis. In the table, the restricted model is obtained from the first-stage restricted model after restricting to zero all parameters that have a first-stage p-value in excess of 0.10. Boldfaced coefficients are significant at a size of 5% or lower.

Unrestricted Model				**Restricted Model**			
Parameter	Factor 1	Factor 2	Factor 3	Parameter	Factor 1	Factor 2	Factor 3
δ_0		**0.0776**		δ_0		0.0555	
		(0.0346)				(0.0314)	
δ	**0.0806**	**0.0359**	**0.0115**	δ	**0.0536**	0.0045	**0.0091**
	(0.0113)	(0.0151)	(0.0056)		(0.0033)	(0.0058)	(0.0049)
$\kappa\theta$	0	0	0	$\kappa\theta$	0	0	0
	—	—	—		—	—	—
$\mathcal{K}'_1$	**0.2853**	0	0	$\mathcal{K}'_1$	**0.2847**	0	0
	(0.0013)	—	—		(0.0010)	—	—
$\mathcal{K}'_2$	**-0.7837**	**0.8771**	0	$\mathcal{K}'_2$	**-0.7883**	**0.8711**	0
	(0.1304)	(0.0180)	—		(0.1109)	(0.0133)	—
$\mathcal{K}'_3$	**-0.0202**	0.0735	**0.0360**	$\mathcal{K}'_3$	**-0.0287**	0	**0.0432**
	(0.0051)	(0.0850)	(0.0009)		(0.0043)	—	(0.0006)
α	1	1	1	α	1	1	1
	—	—	—		—	—	—
Ξ'_1	0	0	0	Ξ'_1	0	0	0
	—	—	—		—	—	—
Ξ'_2	0	0	0	Ξ'_2	0	0	0
	—	—	—		—	—	—
Ξ'_3	0	0	0	Ξ'_3	0	0	0
	—	—	—		—	—	—
λ_1	-0.0409	**-0.4307**	**-0.3112**	λ_1	0	**-0.2194**	**-0.4295**
	(0.0331)	(0.1518)	(0.0797)		—	(0.0890)	(0.1162)
C'_1	0.00605	0	0	C'_1	0.00604	0	0
C'_2	0.00400	0.00311	0	C'_2	0.00399	0.00311	0
C'_3	0.00272	0.00287	0.00133	C'_3	0.00270	0.00289	0.00133
Number of free parameters:			19	Number of free parameters:			17
Number of observations:			7326	Number of observations:			7326
Saturation ratio:			385.6	Saturation ratio:			430.9
Log-likelihood function:			-1386.25	Log-likelihood function:			-1389.69
				LR Test of restrictions:			6.888
						p-value:	(0.032)
Schwartz information criterion:			1.2682	Schwartz information criterion:			1.2415
Akaike information criterion:			1.1796	Akaike information criterion:			1.1726
Hannan-Quinn information crit.:			1.2146	Hannan-Quinn information crit.:			1.1998

Table 3

Full-Sample Estimates of a Completely Affine Heteroskedastic Model

The table reports QMLE estimates for a completely affine term structure model obtained from a common data sample that spans the interval 1972:01 – 2003:12 and setting $L = 2$ in the canonical representation. The estimation procedure assumes that nominal yields on 3-month, 2-year, and 5-year Treasury bonds are measured without error, while for all other maturities yields are measured with errors that have a joint i.i.d. Gaussian distribution with constant covariance matrix. Robust (sandwich-style) standard errors are reported in parenthesis. In the table, the restricted model is obtained from the first-stage restricted model after restricting to zero all parameters that have a first-stage p-value in excess of 0.10. Boldfaced coefficients are significant at a size of 5% or lower.

Unrestricted Model				**Restricted Model**			
Parameter	Factor 1	Factor 2	Factor 3	Parameter	Factor 1	Factor 2	Factor 3
δ_0		**0.0238**		δ_0		**0.0225**	
		(0.0095)				(0.0092)	
δ	**0.0576**	**0.0944**	0.0138	δ	**0.0579**	**0.0879**	0
	(0.0029)	(0.0449)	(0.0111)		(0.0019)	(0.0229)	—
$\kappa\theta$	0	**0.1811**	**-1.5328**	$\kappa\theta$	0	**0.1745**	**-1.5395**
	—	(0.0772)	—		—	(0.0814)	—
$\mathcal{K}'_1$	**0.1132**	**-0.1153**	0	$\mathcal{K}'_1$	**0.1325**	**-0.0965**	0
	(0.0041)	(0.0495)	—		(0.0030)	(0.0166)	—
$\mathcal{K}'_2$	-0.0593	**0.2714**	0	$\mathcal{K}'_2$	0	**0.2489**	0
	(0.0360)	(0.0084)	—		—	(0.0017)	—
$\mathcal{K}'_3$	**0.4560**	**-2.6808**	**0.9302**	$\mathcal{K}'_3$	**0.4375**	**-2.6621**	**0.9111**
	(0.2328)	(1.3224)	(0.0246)		(0.1821)	(1.2456)	(0.0238)
α	1	1	1	α	1	1	1
	—	—	—		—	—	—
Ξ'_1	1	0	0	Ξ'_1	1	0	0
	—	—	—		—	—	—
Ξ'_2	0	1	0	Ξ'_2	0	1	0
	—	—	—		—	—	—
Ξ'_3	0.0048	**3.1872**	0	Ξ'_3	0	**2.1797**	0
	(0.0061)	(0.8632)	—		—	(0.0485)	—
λ_1	-0.0305	**-0.0435**	**-0.1544**	λ_1	0	**-0.1314**	**-0.1513**
	(0.0216)	(0.0182)	(0.0638)		—	(0.0715)	(0.0639)
C'_1	0.00601	0	0	C'_1	0.00602	0	0
C'_2	0.00397	0.00311	0	C'_2	0.00398	0.00311	0
C'_3	0.00269	0.00289	0.00132	C'_3	0.00270	0.00289	0.00132
Number of free parameters:			24	Number of free parameters:			20
Number of observations:			7326	Number of observations:			7326
Saturation ratio:			305.3	Saturation ratio:			366.3
Log-likelihood function:			-1359.23	Log-likelihood function:			-1365.28
				LR Test of restrictions:			12.10
				p-value:			(0.017)
Schwartz information criterion:			1.2904	Schwartz information criterion			1.2658
Akaike information criterion:			1.1722	Akaike information criterion:			1.2063
Hannan-Quinn information crit.:			1.2190	Hannan-Quinn information cri			1.1673

Table 4

Full-Sample Estimates of an Essentially Affine (Gaussian) Model

The table reports QMLE estimates for a completely affine term structure model obtained from a common data sample that spans the interval 1972:01 – 2003:12 and setting $L = 0$ in the canonical essentially affine representation. The estimation procedure assumes that nominal yields on 3-month, 2-year, and 5-year Treasury bonds are measured without error, while for all other maturities yields are measured with errors that have a joint i.i.d. Gaussian distribution with constant covariance matrix. Robust (sandwich-style) standard errors are reported in parenthesis. In the table, the restricted model is obtained from the first-stage restricted model after restricting to zero all parameters that have a first-stage p-value in excess of 0.10. Boldfaced coefficients are significant at a size of 5% or lower.

Unrestricted Model

Parameter	Factor 1	Factor 2	Factor 3
δ_0		**0.1220**	
		(0.0337)	
δ	**0.0501**	**0.0415**	0.0573
	(0.0029)	(0.0167)	(0.0464)
$\kappa\theta$	0	0	0
	—	—	—
$\mathcal{K}'_1$	**0.3831**	0	0
	(0.0061)	—	—
$\mathcal{K}'_2$	**-0.4631**	**2.7282**	0
	(0.1269)	(0.0667)	—
$\mathcal{K}'_3$	**-0.1854**	-0.4056	**0.0899**
	(0.0747)	(1.0612)	(0.0025)
α	1	1	1
	—	—	—
Ξ'_1	0	0	0
	—	—	—
Ξ'_2	0	0	0
	—	—	—
Ξ'_3	0	0	0
	—	—	—
λ_1	**-0.2914**	**-0.3583**	**-0.3478**
	(0.1506)	(0.1284)	(0.0885)
Λ'_1	**-0.3753**	**0.6533**	-0.0332
	(0.1800)	(0.2887)	(0.0197)
Λ'_2	**0.3870**	**-0.9091**	-0.0348
	(0.1343)	(0.3701)	(0.0277)
Λ'_3	**0.3103**	**0.5415**	-0.0466
	(0.1163)	(0.2494)	(0.0368)
C'_1	0.00604	0	0
C'_2	0.00400	0.00304	0
C'_3	0.00271	0.00288	0.00131
Number of free parameters:			28
Number of observations:			7326
Saturation ratio:			261.6
Log-likelihood function:			-1338.4
Schwartz information criterion:			1.3619
Akaike information criterion:			1.1864
Hannan-Quinn information crit.:			1.2548

Restricted Model

Parameter	Factor 1	Factor 2	Factor 3
δ_0		**0.0163**	
		(0.0037)	
δ	**0.0529**	**0.0524**	0
	(0.0017)	(0.0191)	—
$\kappa\theta$	0	0	0
	—	—	—
$\mathcal{K}'_1$	**0.3787**	0	0
	(0.0017)	—	—
$\mathcal{K}'_2$	**-0.4459**	**2.7285**	0
	(0.0659)	(0.0430)	—
$\mathcal{K}'_3$	**-0.1555**	0	**0.0925**
	(0.0247)	—	(0.0012)
α	1	1	1
	—	—	—
Ξ'_1	0	0	0
	—	—	—
Ξ'_2	0	0	0
	—	—	—
Ξ'_3	0	0	0
	—	—	—
λ_1	**-0.2678**	**-0.3257**	**-0.3114**
	(0.1479)	(0.1376)	(0.0851)
Λ'_1	-0.2432	**0.6715**	0
	(0.1918)	(0.2417)	—
Λ'_2	**0.2942**	**-0.7947**	0
	(0.1197)	(0.4089)	—
Λ'_3	**0.2001**	**0.4524**	0
	(0.1153)	(0.1766)	—
C'_1	0.00605	0	0
C'_2	0.00400	0.00311	0
C'_3	0.00271	0.00289	0.00132
Number of free parameters:			23
Number of observations:			7326
Saturation ratio:			318.5
Log-likelihood function:			-1374.31
LR Test of restrictions:			71.82
p-value:			(0.000)
Schwartz information criterion:			1.3322
Akaike information criterion:			1.1944
Hannan-Quinn information crit.:			1.2489

Table 5

Full-Sample Estimates of an Essentially Affine Model

The table reports QMLE estimates for a completely affine term structure model obtained from a common data sample that spans the interval 1972:01 – 2003:12 and setting $L = 1$ in the canonical essentially affine representation. The estimation procedure assumes that nominal yields on 3-month, 2-year, and 5-year Treasury bonds are measured without error, while for all other maturities yields are measured with errors that have a joint i.i.d. Gaussian distribution with constant covariance matrix. Robust (sandwich-style) standard errors are reported in parenthesis. In the table, the restricted model is obtained from the first-stage restricted model after restricting to zero all parameters that have a first-stage p-value in excess of 0.10. Boldfaced coefficients are significant at a size of 5% or lower.

Unrestricted Model				Restricted Model			
Parameter	Factor 1	Factor 2	Factor 3	Parameter	Factor 1	Factor 2	Factor 3
δ_0		**0.0280**		δ_0		**0.0213**	
		(0.0149)				(0.0112)	
δ	**0.0579**	0.0083	0.0018	δ	**0.0500**	0	0
	(0.0035)	(0.0069)	(0.0014)		(0.0008)	—	—
$\kappa\theta$	**0.1672**	**-1.4098**	**0.6548**	$\kappa\theta$	**0.1648**	**-1.4026**	**0.6515**
	(0.0765)	—	—		(0.0758)	—	—
$\mathcal{K}'_1$	**0.0322**	0	0	$\mathcal{K}'_1$	**0.0319**	0	0
	(0.0003)	—	—		(0.0003)	—	—
$\mathcal{K}'_2$	**-0.2715**	**0.4939**	**3.4962**	$\mathcal{K}'_2$	**-0.0833**	**0.4990**	**3.4932**
	(0.0343)	(0.0534)	(1.4856)		(0.0104)	(0.0506)	(0.6633)
$\mathcal{K}'_3$	**0.1261**	**-0.1571**	**1.5956**	$\mathcal{K}'_3$	**0.1272**	**-0.1606**	**1.5927**
	(0.0217)	(0.0356)	(0.0676)		(0.0209)	(0.0340)	(0.0648)
α	1	1	1	α	1	1	1
	—	—	—		—	—	—
Ξ'_1	0	0	0	Ξ'_1	0	0	0
	—	—	—		—	—	—
Ξ'_2	**5.5344**	0	0	Ξ'_2	**5.3229**	0	0
	(0.2951)	—	—		(2.7017)	—	—
Ξ'_3	**0.1566**	0	0	Ξ'_3	**0.2394**	0	0
	(0.0062)	—	—		(0.1162)	—	—
λ_1	**-0.0654**	**-4.0354**	**-0.1124**	λ_1	**-0.0647**	**-3.9742**	**-0.1067**
	(0.0362)	(2.2802)	(0.0518)		(0.0365)	(2.1835)	(0.0518)
Λ'_1	0	0	0	Λ'_1	0	0	0
	—	—	—		—	—	—
Λ'_2	**41.7737**	-0.0530	**3.9253**	Λ'_2	**5.5606**	0	**0.1135**
	(23.6237)	(0.0737)	(2.3251)		(1.2671)	—	(0.0796)
Λ'_3	**-0.0320**	0.0072	-0.9415	Λ'_3	**-0.0132**	0	0
	(0.0150)	(0.1127)	(3.1850)		(0.0004)	—	—
C'_1	0.00605	0	0	C'_1	0.00604	0	0
C'_2	0.00407	0.00298	0	C'_2	0.00406	0.00302	0
C'_3	0.00277	0.00279	0.00131	C'_3	0.00276	0.00282	0.00133
Number of free parameters:			31	Number of free parameters:			26
Number of observations:			7326	Number of observations:			7326
Saturation ratio:			236.3	Saturation ratio:			281.8
Log-likelihood function:			-1257.68	Log-likelihood function:			-1292.31
				LR Test of restrictions:			69.26
						p-value:	(0.000)
Schwartz information criterion:			1.2958	Schwartz information criterion:			1.2651
Akaike information criterion:			1.1283	Akaike information criterion:			1.1272
Hannan-Quinn information crit.:			1.1946	Hannan-Quinn information crit.:			1.1818

Table 6

Summary Statistics for Monthly Forecast Errors

The table shows summary statistics for monthly theoretical forecast errors for a range of models and forecast horizons. T-Bills at the 1- and 3-month maturities are considered. For comparability, these statistics are calculated using forecast errors over the common out-of-sample period, 1982:01-2003:12. Panels A and B report the forecasting performance of the theoretical forecast with a constant risk premium and a time-varying risk premium (computed using a rolling window of P = 10), respectively. MSFE and RMSFE are the mean and root-mean squared forecast errors, respectively. MAFE is the mean absolute forecast error.

Statistic	h=1	h=2	h=3	h=6	h=9	h=12	h=15
	1-month rate		3-month rate				
	Panel A - Theoretical Forecast Errors (constant risk premium)						
Mean	-0.001	0.000	0.009	0.002	-0.003	-0.008	-0.012
Median	0.111	0.194	0.337	0.231	0.178	0.007	-0.166
Max.	0.996	0.890	1.613	2.038	2.789	3.678	3.340
Min.	-3.817	-6.766	-5.796	-5.657	-5.369	-5.423	-4.800
S.D.	0.510	0.764	0.758	1.136	1.367	1.556	1.729
MSFE	0.259	0.581	0.605	1.287	1.866	2.435	3.037
RMSFE	0.509	0.762	0.778	1.134	1.366	1.561	1.743
MAFE	0.307	0.448	0.559	0.802	1.016	1.185	1.360
	Panel B - Theoretical Forecast Errors (time-varying risk premium)						
Mean	0.023	0.045	0.236	0.099	0.005	-0.081	-0.169
Median	0.057	0.034	0.210	0.069	0.037	-0.124	-0.269
Max.	2.038	2.729	3.463	3.797	3.483	4.331	4.573
Min.	-2.819	-5.007	-4.037	-4.445	-4.288	-4.622	-4.206
S.D.	0.484	0.735	0.778	1.133	1.389	1.603	1.799
MSFE	0.234	0.540	0.659	1.288	1.921	2.566	3.252
RMSFE	0.484	0.735	0.812	1.135	1.386	1.602	1.803
MAFE	0.306	0.452	0.548	0.801	1.028	1.214	1.395
	Panel C - Random Walk Errors						
Mean	-0.040	-0.078	0.113	-0.025	-0.120	-0.206	-0.294
Median	-0.019	-0.038	0.120	-0.034	-0.143	-0.275	-0.374
Max.	2.406	1.996	1.993	2.800	3.630	3.998	4.790
Min.	-2.745	-5.012	-4.042	-5.479	-5.191	-4.441	-4.576
S.D.	0.471	0.691	0.783	1.134	1.385	1.580	1.777
MSFE	0.223	0.481	0.623	1.282	1.924	2.529	3.233
RMSFE	0.472	0.694	0.789	1.132	1.387	1.590	1.798
MAFE	0.302	0.440	0.566	0.824	1.046	1.241	1.433
	Panel D - Diebold and Li's Errors						
Mean	-0.227	-0.362	-0.170	-0.307	-0.400	-0.485	-0.572
Median	-0.182	-0.257	-0.137	-0.287	-0.252	-0.449	-0.497
Max.	1.481	1.124	1.556	1.826	2.441	3.020	3.779
Min.	-3.414	-5.729	-4.759	-5.710	-5.422	-4.649	-4.541
S.D.	0.471	0.684	0.736	1.095	1.334	1.522	1.707
MSFE	0.273	0.596	0.568	1.289	1.933	2.541	3.229
RMSFE	0.523	0.772	0.754	1.135	1.390	1.594	1.797
MAFE	0.334	0.491	0.537	0.844	1.054	1.249	1.442

Table 6 [continued]

Summary Statistics for Monthly Forecast Errors

Statistic	h=1	h=2	h=3	h=6	h=9	h=12	h=15
	Panel E - Slope-Based Benchmark Errors						
Mean	-0.074	-0.153	-0.201	-0.332	-0.411	-0.495	-0.578
Median	-0.023	-0.112	-0.144	-0.237	-0.384	-0.582	-0.580
Max.	2.013	3.084	1.862	2.751	3.128	3.822	4.513
Min.	-4.984	-5.078	-6.289	-5.490	-5.905	-4.646	-4.471
S.D.	0.687	0.954	1.046	1.259	1.455	1.628	1.823
MSFE	0.475	0.930	1.130	1.689	2.276	2.886	3.643
RMSFE	0.689	0.964	1.063	1.300	1.509	1.699	1.909
MAFE	0.436	0.637	0.730	0.959	1.180	1.374	1.561
	Panel F - Unrestricted Completely Affine Gaussian Model ($A_0(3)$)						
Mean	-0.051	-0.113	-0.160	-0.290	-0.367	-0.449	-0.530
Median	0.010	-0.068	-0.086	-0.205	-0.384	-0.481	-0.574
Max.	1.525	3.151	1.850	2.285	3.269	3.694	4.045
Min.	-5.211	-5.020	-5.761	-5.415	-5.377	-4.470	-4.628
S.D.	0.617	0.861	0.958	1.208	1.389	1.553	1.735
MSFE	0.382	0.751	0.941	1.537	2.056	2.603	3.277
RMSFE	0.618	0.866	0.970	1.240	1.434	1.613	1.810
MAFE	0.364	0.555	0.638	0.888	1.106	1.288	1.471
	Panel G - Restricted Completely Affine Model ($A_2(3)$)						
Mean	-0.048	-0.089	-0.135	-0.262	-0.335	-0.413	-0.490
Median	0.014	0.047	-0.025	-0.135	-0.321	-0.388	-0.426
Max.	1.510	3.311	2.010	2.204	3.032	3.388	3.713
Min.	-5.351	-4.987	-5.880	-5.378	-5.496	-4.376	-4.497
S.D.	0.626	0.889	0.981	1.221	1.394	1.555	1.740
MSFE	0.392	0.795	0.978	1.554	2.048	2.578	3.256
RMSFE	0.626	0.892	0.989	1.246	1.431	1.606	1.804
MAFE	0.367	0.582	0.666	0.907	1.114	1.285	1.473
	Panel H - Unrestricted Essentially Affine Gaussian Model ($A_0(3)$)						
Mean	-0.021	-0.094	-0.141	-0.272	-0.349	-0.432	-0.515
Median	0.026	-0.026	-0.015	-0.170	-0.306	-0.467	-0.525
Max.	1.606	2.625	1.519	2.347	2.601	3.580	3.898
Min.	-3.246	-5.921	-5.122	-5.366	-5.384	-4.890	-4.485
S.D.	0.458	0.783	0.880	1.172	1.359	1.521	1.703
MSFE	0.209	0.619	0.792	1.443	1.961	2.492	3.152
RMSFE	0.458	0.787	0.890	1.201	1.400	1.579	1.776
MAFE	0.282	0.480	0.558	0.848	1.060	1.249	1.431
	Panel I - Restricted Essentially Affine Gaussian Model ($A_1(3)$)						
Mean	-0.058	-0.137	-0.184	-0.314	-0.391	-0.473	-0.554
Median	0.012	-0.043	-0.090	-0.219	-0.388	-0.508	-0.566
Max.	1.535	3.537	2.236	2.225	3.099	3.423	3.736
Min.	-5.452	-4.958	-5.882	-5.243	-5.498	-4.366	-4.486
S.D.	0.635	0.890	0.982	1.215	1.391	1.552	1.739
MSFE	0.405	0.809	0.994	1.569	2.081	2.624	3.319
RMSFE	0.637	0.899	0.997	1.253	1.442	1.620	1.822
MAFE	0.371	0.567	0.661	0.908	1.119	1.301	1.489

Table 7

Tests of Equal Predictive Accuracy: Squared Forecast Error Loss

The table presents test statistics for two types of equal predictive accuracy tests. The numbers above the main diagonal report the standard modified Diebold-Mariano test and the corresponding significance level in parentheses. The numbers below the diagonal report the West-McCracken test statistic, again with the corresponding significance level in parentheses. The tests are performed for 1-, 6-, and 15-month horizon (the first exercise refers to 1-month T-bill rates, the latter two exercises to 3-month T-bill rates). All instances where the null hypothesis of equal predictive accuracy is reject with a p-value below 0.1 are in bold typeface. For numbers above the main diagonal, a negative (positive) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column. For numbers below the main diagonal, the interpretation is that a positive (negative) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column.

Panel A - horizon: 1 month	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian $A_0(3)$	Restricted Essentially Affine $A_1(3)$
Theoretical (constant risk premium)		0.421	0.567	-0.545	**-1.930**	1.338	**-1.877**
		(0.674)	(0.571)	(0.586)	(0.054)	(0.181)	(0.061)
Theoretical (time-varying risk premium)	0.341		0.439	-0.974	-1.298	0.830	-1.335
	(0.733)		(0.660)	(0.330)	(0.194)	(0.407)	(0.182)
Random Walk	0.406	0.377		-1.216	-1.364	0.443	-1.392
	(0.685)	(0.706)		(0.224)	(0.173)	(0.658)	(0.164)
Diebold and Li	-0.401	-0.513	-0.428		-1.341	**3.338**	-1.381
	(0.688)	(0.608)	(0.669)		(0.180)	(0.001)	(0.167)
Unrestricted Completely Affine Gaussian $A_0(3)$	-0.739	-0.370	-0.319	-0.580		1.845	-1.494
	(0.460)	(0.712)	(0.750)	(0.562)		(0.065)	(0.135)
Unrestricted Essentially Affine Gaussian $A_0(3)$	0.290	0.476	0.376	1.278	1.333		**-1.807**
	(0.772)	(0.634)	(0.707)	(0.201)	(0.183)		(0.071)
Restricted Essentially Affine $A_1(3)$	-0.689	-0.355	-0.304	0.546	-1.317	-1.312	
	(0.492)	(0.722)	(0.761)	(0.585)	(0.188)	(0.190)	

Table 7 [continued]

Tests of Equal Predictive Accuracy: Squared Forecast Error Loss

Panel B - horizon: 6 months	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian A0(3)	Restricted Essentially Affine A1(3)
Theoretical (constant risk premium)		-0.001 (0.999)	0.018 (0.986)	-0.009 (0.993)	-1.363 (0.173)	-1.167 (0.243)	-1.600 (0.110)
Theoretical (time-varying risk premium)	-0.002 (0.999)		0.021 (0.983)	-0.004 (0.997)	-0.811 (0.418)	-0.451 (0.652)	-0.932 (0.351)
Random Walk	0.018 (0.986)	0.023 (0.982)		-0.042 (0.966)	-1.144 (0.253)	-0.623 (0.533)	-1.243 (0.213)
Diebold and Li	-0.010 (0.992)	-0.004 (0.997)	-0.041 (0.967)		**-1.655** (0.098)	-0.956 (0.339)	**-1.982** (0.048)
Unrestricted Completely Affine Gaussian $A_0(3)$	-0.516 (0.606)	-0.510 (0.610)	-0.469 (0.639)	-0.511 (0.609)		1.170 (0.242)	-0.783 (0.434)
Unrestricted Essentially Affine Gaussian $A_0(3)$	-0.530 (0.596)	-0.356 (0.722)	-0.400 (0.689)	-0.407 (0.684)	0.565 (0.572)		**-1.793** (0.073)
Restricted Essentially Affine $A_1(3)$	-0.603 (0.547)	-0.553 (0.581)	-0.468 (0.640)	-0.659 (0.510)	-0.589 (0.556)	-0.762 (0.446)	

Panel C - horizon: 15 months	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian A0(3)	Restricted Essentially Affine A1(3)
Theoretical (constant risk premium)		-0.726 (0.468)	-0.504 (0.615)	-0.450 (0.653)	-0.470 (0.639)	-0.272 (0.785)	-0.563 (0.573)
Theoretical (time-varying risk premium)	-0.559 (0.576)		0.038 (0.970)	0.042 (0.967)	-0.043 (0.966)	0.184 (0.854)	-0.108 (0.914)
Random Walk	-0.405 (0.686)	0.036 (0.972)		0.012 (0.990)	-0.210 (0.833)	0.386 (0.700)	-0.317 (0.751)
Diebold and Li	-0.373 (0.709)	0.042 (0.967)	0.010 (0.992)		-0.219 (0.826)	0.503 (0.615)	-0.810 (0.418)
Unrestricted Completely Affine Gaussian $A_0(3)$	-0.390 (0.697)	-0.040 (0.968)	-0.203 (0.839)	-0.205 (0.837)		1.134 (0.257)	-0.395 (0.693)
Unrestricted Essentially Affine Gaussian $A_0(3)$	-0.252 (0.801)	0.163 (0.871)	0.337 (0.736)	0.435 (0.663)	0.616 (0.538)		**-1.930** (0.054)
Restricted Essentially Affine $A_1(3)$	-0.435 (0.664)	-0.104 (0.917)	-0.292 (0.770)	-0.662 (0.508)	-0.323 (0.745)	-0.772 (0.440)	

Table 8

Tests of Equal Predictive Accuracy: Absolute Forecast Error Loss

The table presents test statistics for two types of equal predictive accuracy tests. The numbers above the main diagonal report the standard modified Diebold-Mariano test and the corresponding significance level in parentheses. The numbers below the diagonal report the West-McCracken test statistic, again with the corresponding significance level in parentheses. The tests are performed for 1-, 6-, and 15-month horizon (the first exercise refers to 1-month T-bill rates, the latter two exercises to 3-month T-bill rates). All instances where the null hypothesis of equal predictive accuracy is reject with a p-value below 0.1 are in bold typeface. For numbers above the main diagonal, a negative (positive) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column. For numbers below the main diagonal, the interpretation is that a positive (negative) value of the test statistic implies that the model in the row produces more (less) accurate prediction than the model in the column.

Panel A - horizon: 1 month	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian $A_0(3)$	Restricted Essentially Affine $A_1(3)$
Theoretical (constant risk premium)		0.060	0.287	**-1.847**	**-2.667**	**2.079**	**-2.947**
		(0.952)	(0.774)	(0.065)	(0.008)	(0.038)	(0.003)
Theoretical (time-varying risk premium)	0.062		0.225	-1.319	**-2.114**	1.491	**-2.265**
	(0.950)		(0.822)	(0.187)	(0.035)	(0.136)	(0.023)
Random Walk	0.274	0.216		**-2.164**	**-2.496**	**1.886**	**-2.636**
	(0.784)	(0.829)		(0.031)	(0.013)	(0.059)	(0.008)
Diebold and Li	-1.540	-0.554	-0.716		-1.434	**3.763**	**-1.715**
	(0.124)	(0.580)	(0.474)		(0.151)	(0.000)	(0.086)
Unrestricted Completely Affine Gaussian $A_0(3)$	1.569	-1.477	-1.534	-0.636		**4.080**	-1.448
	(0.112)	(0.140)	(0.125)	(0.525)		(0.000)	(0.148)
Unrestricted Essentially Affine Gaussian $A_0(3)$	**1.689**	0.439	1.204	1.593	**2.467**		**-4.118**
	(0.091)	(0.661)	(0.229)	(0.111)	(0.014)		(0.000)
Restricted Essentially Affine $A_1(3)$	**-1.739**	-1.466	-1.540	**-1.651**	-1.492	**-2.460**	
	(0.082)	(0.143)	(0.124)	(0.099)	(0.136)	(0.014)	

Table 8 [continued]

Tests of Equal Predictive Accuracy: Absolute Forecast Error Loss

Panel B - horizon: 6 months	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian A0(3)	Restricted Essentially Affine A1(3)
Theoretical (constant risk premium)		0.018 (0.985)	-0.350 (0.727)	-0.567 (0.570)	-1.137 (0.256)	-0.745 (0.456)	-1.430 (0.153)
Theoretical (time-varying risk premium)	0.019 (0.985)		-0.321 (0.749)	-0.412 (0.681)	-1.011 (0.312)	-0.557 (0.577)	-1.241 (0.215)
Random Walk	-0.317 (0.751)	-0.288 (0.774)		-0.326 (0.744)	-1.106 (0.269)	-0.438 (0.662)	-1.407 (0.159)
Diebold and Li	-0.418 (0.676)	-0.351 (0.726)	-0.300 (0.764)		-0.960 (0.337)	-0.105 (0.916)	-1.609 (0.108)
Unrestricted Completely Affine Gaussian $A_0(3)$	-0.509 (0.611)	-0.533 (0.594)	-0.573 (0.567)	-0.612 (0.541)		1.609 (0.108)	-1.338 (0.181)
Unrestricted Essentially Affine Gaussian $A_0(3)$	-0.503 (0.615)	-0.439 (0.661)	-0.374 (0.709)	-0.107 (0.915)	1.331 (0.183)		**-2.549** (0.011)
Restricted Essentially Affine $A_1(3)$	-0.498 (0.618)	-0.540 (0.589)	-0.554 (0.580)	-0.713 (0.476)	-0.501 (0.616)	-1.564 (0.118)	

Panel C - horizon: 15 months	Theoretical (constant risk premium)	Theoretical (time-varying risk premium)	Random Walk	Diebold and Li	Unrestricted Completely Affine Gaussian $A_0(3)$	Unrestricted Essentially Affine Gaussian A0(3)	Restricted Essentially Affine A1(3)
Theoretical (constant risk premium)		-0.428 (0.669)	-0.880 (0.379)	-0.733 (0.463)	-0.899 (0.369)	-0.682 (0.496)	-1.036 (0.300)
Theoretical (time-varying risk premium)	-0.362 (0.717)		-0.406 (0.685)	-0.352 (0.725)	-0.561 (0.575)	-0.304 (0.761)	-0.648 (0.517)
Random Walk	-0.483 (0.629)	-0.349 (0.727)		-0.102 (0.919)	-0.698 (0.485)	0.033 (0.974)	-0.849 (0.396)
Diebold and Li	-0.458 (0.647)	-0.317 (0.752)	-0.101 (0.919)		-0.509 (0.611)	0.236 (0.813)	-1.238 (0.216)
Unrestricted Completely Affine Gaussian $A_0(3)$	-0.463 (0.644)	-0.438 (0.662)	-0.530 (0.596)	-0.442 (0.659)		1.552 (0.121)	-0.824 (0.410)
Unrestricted Essentially Affine Gaussian $A_0(3)$	-0.451 (0.652)	-0.277 (0.782)	0.033 (0.974)	0.231 (0.817)	0.991 (0.322)		**-2.593** (0.010)
Restricted Essentially Affine $A_1(3)$	-0.456 (0.648)	-0.449 (0.654)	-0.540 (0.589)	-0.809 (0.419)	-0.503 (0.615)	-1.575 (0.115)	

Figure 1

Forecast Error Series: Comparing EH-Implied (with Constant Risk Premia) and Random Walk Models

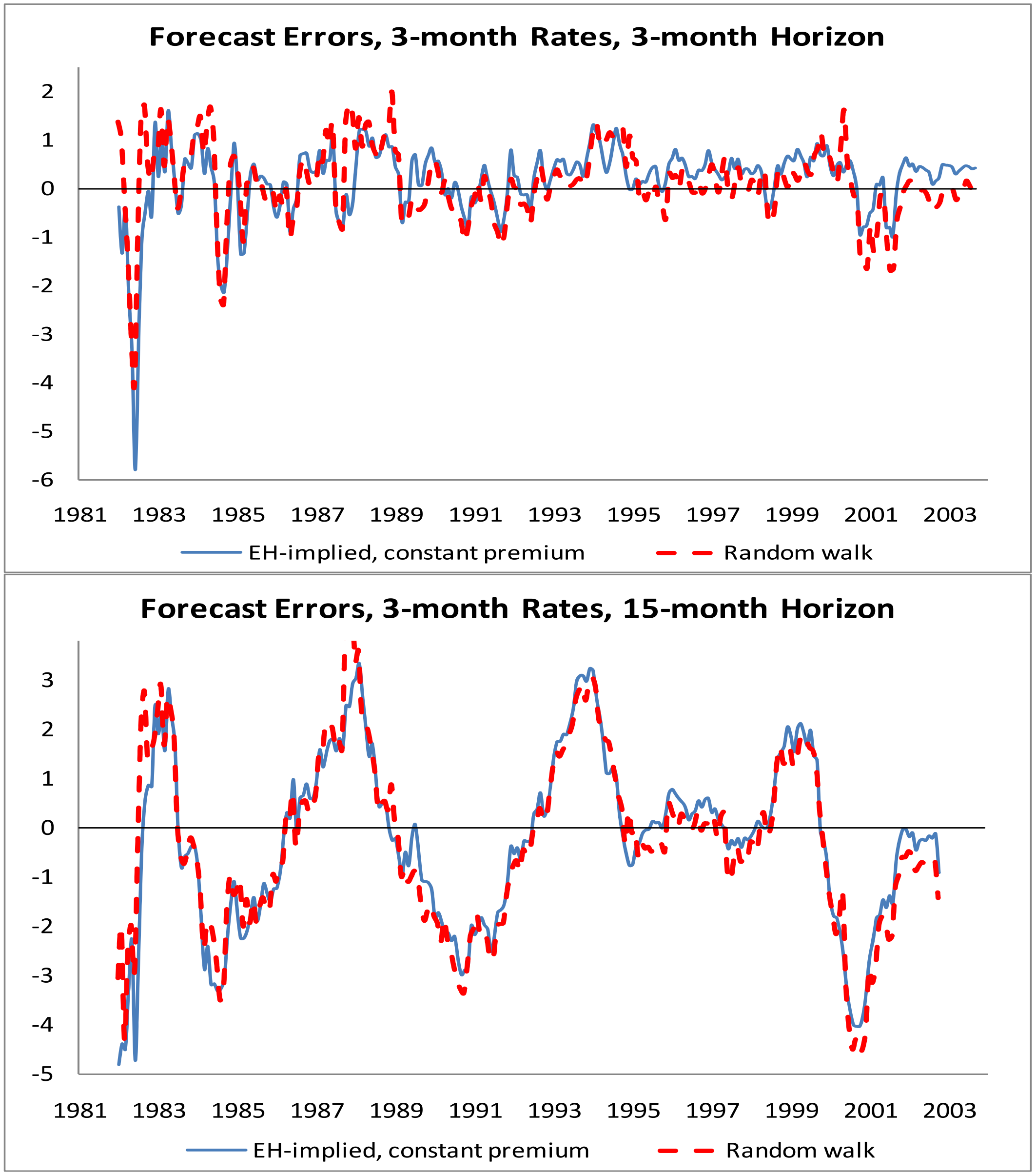


Figure 2

Implied Constant and Rolling-Window (10-month) EH Estimates of the Risk Premium, 1- and 2-month T-Bills

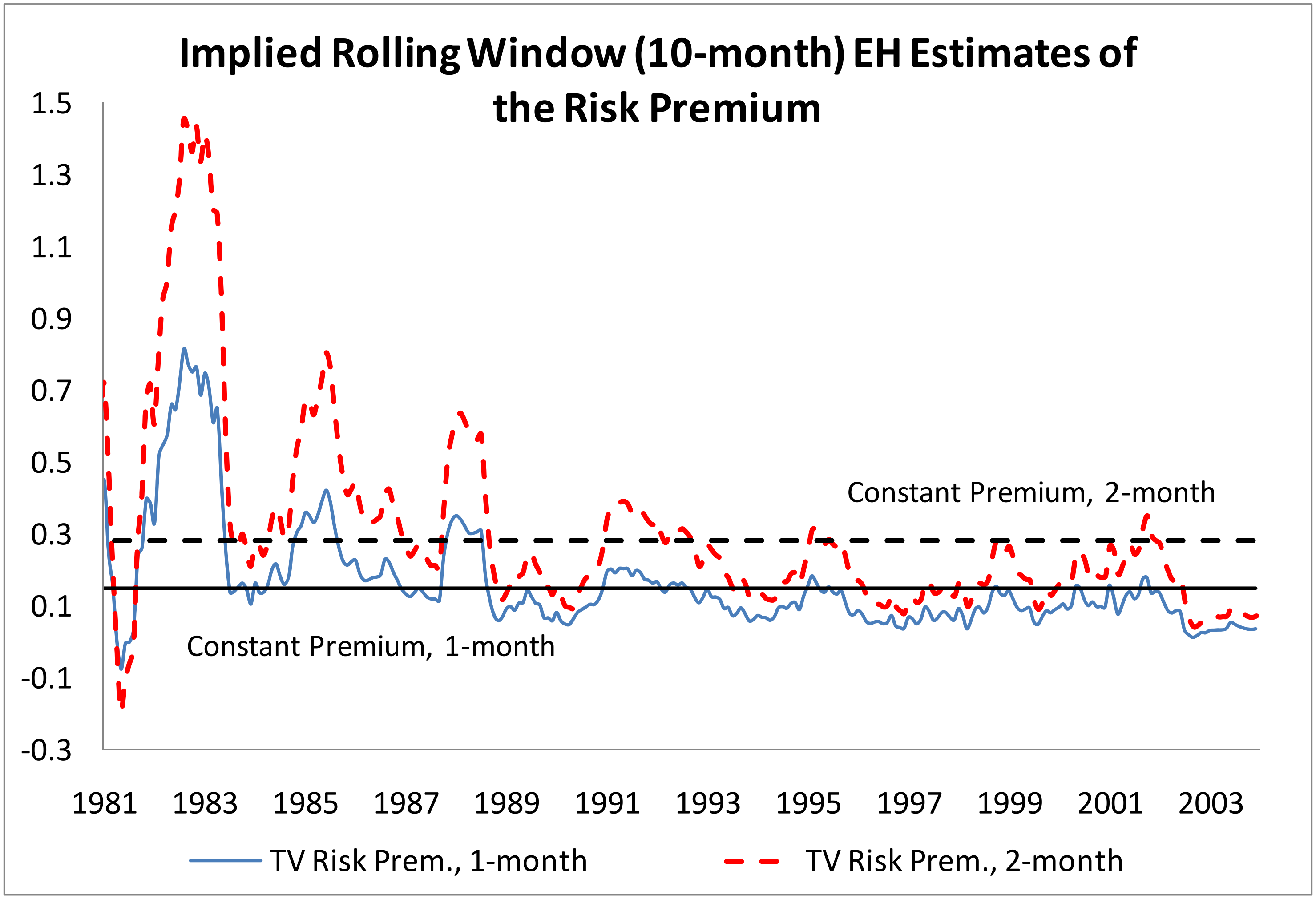
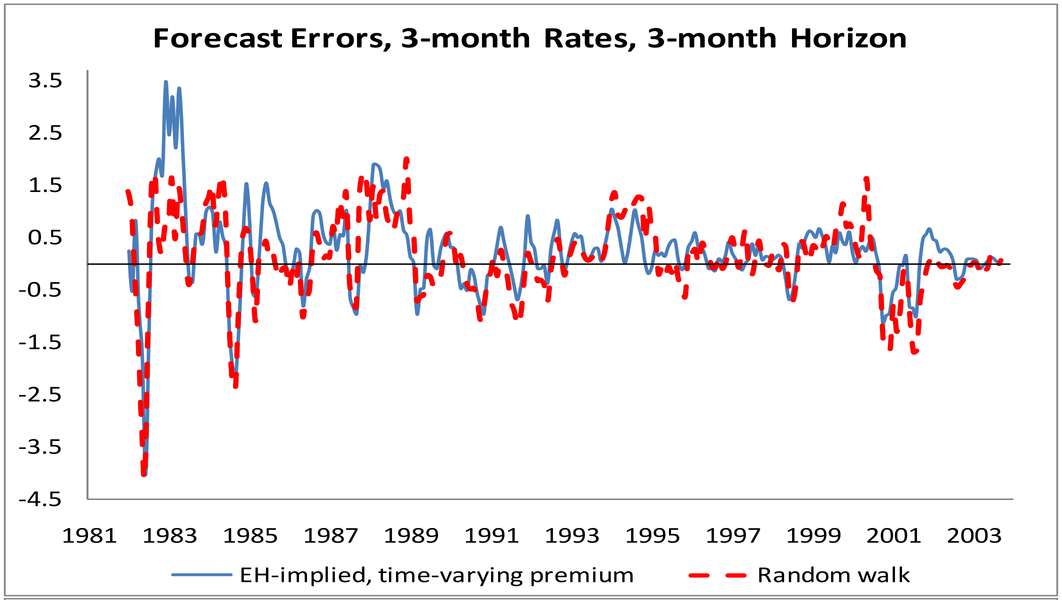


Figure 3

Forecast Error Series: Comparing EH-Implied (with Time-Varying Risk Premia) and Random Walk Models



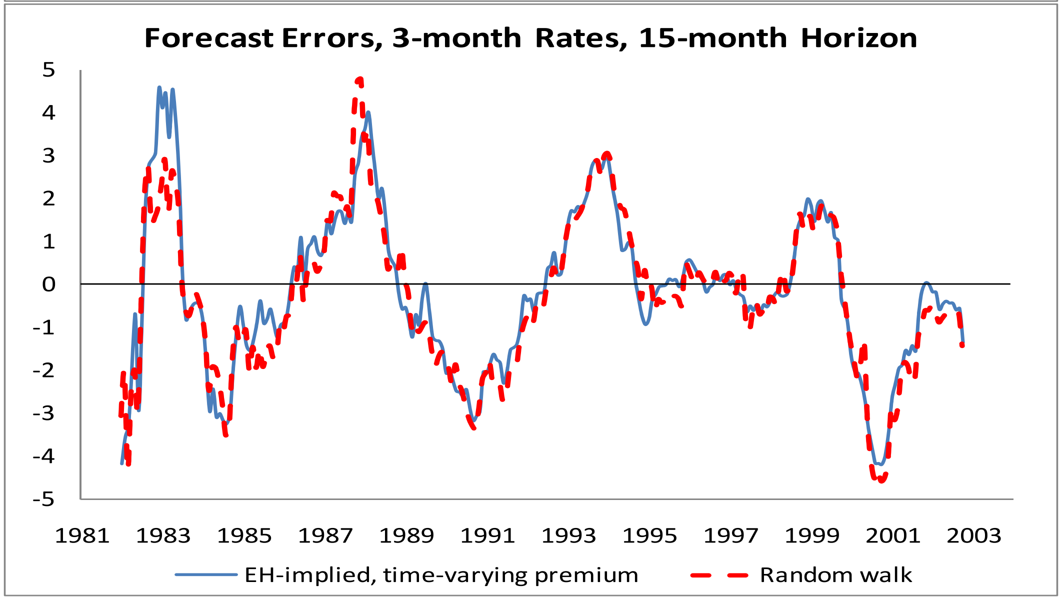
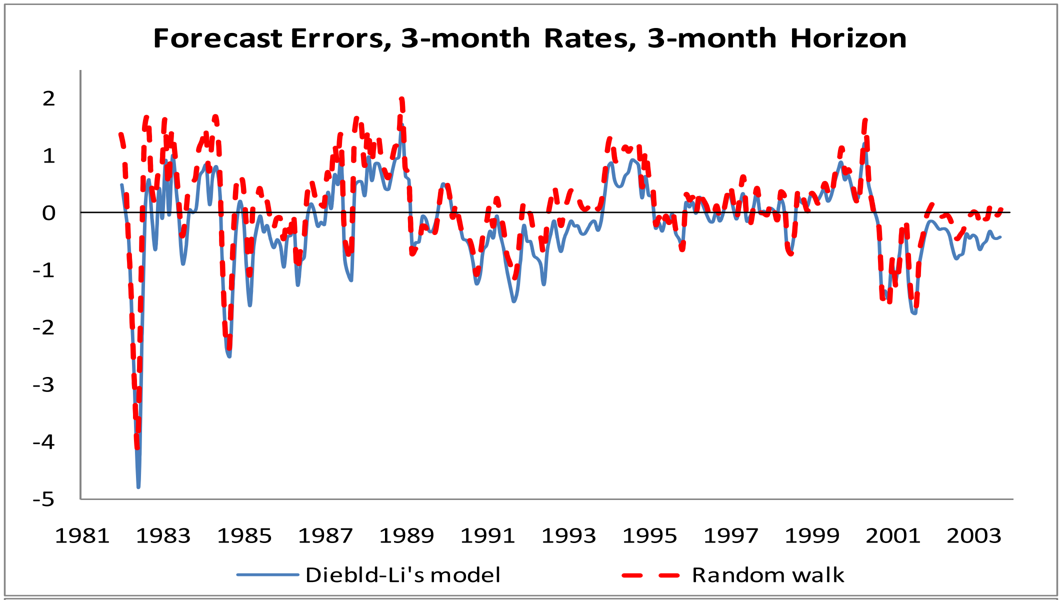


Figure 4

Forecast Error Series: Comparing EH-Implied (with Time-Varying Risk Premia) and Random Walk Models



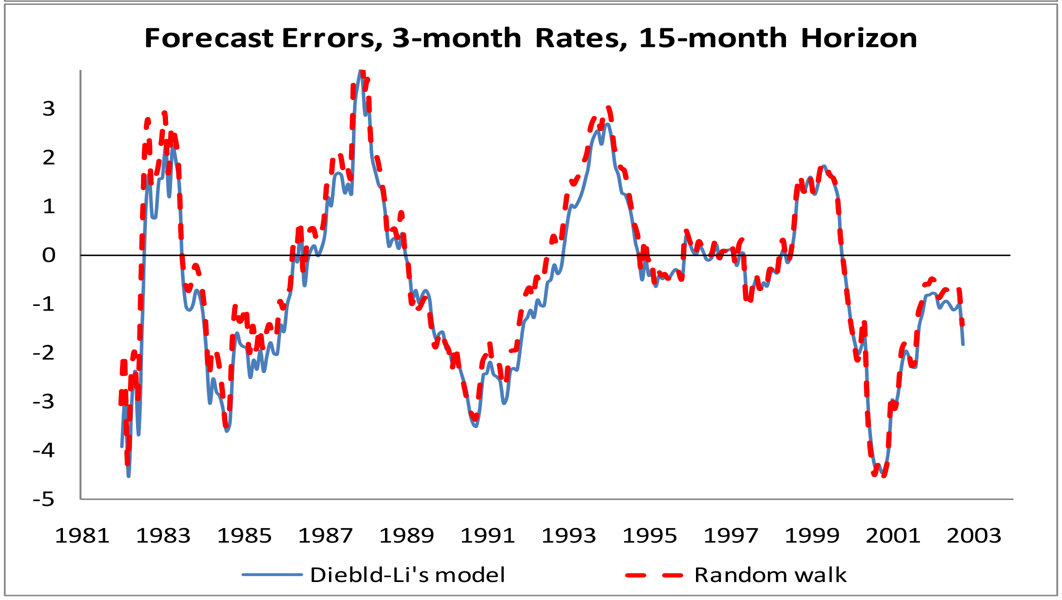
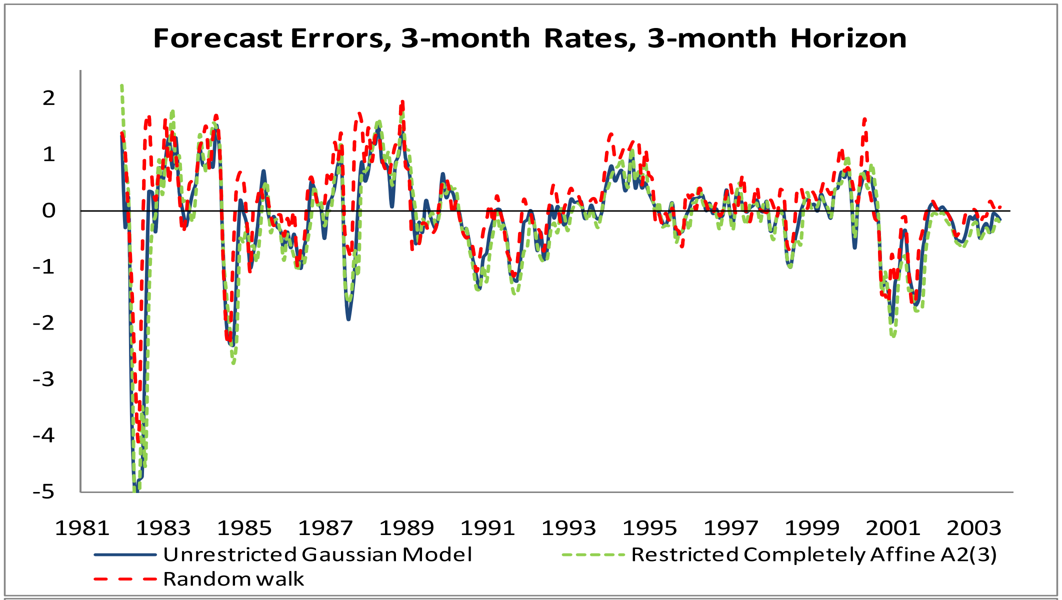


Figure 5

Forecast Error Series: Comparing Completely Affine Term Structure and Random Walk Models



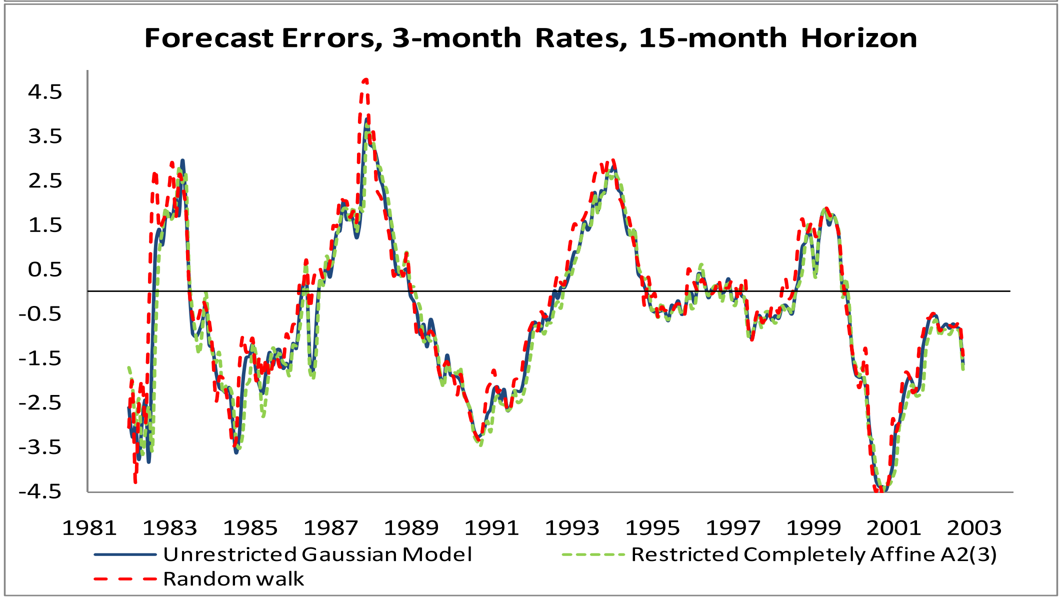


Figure 6

Forecast Error Series: Comparing Essentially Affine Term Structure and Random Walk Models