A Simple Test of the Affine Class of Term Structure Models

PIERLUIGI BALDUZZI and I-HSUAN ETHAN CHIANG*

This draft: October 21, 2008

*Balduzzi is from Boston College; Chiang is from the University of North Carolina at Charlotte. The authors thank Yacine Aït-Sahalia, Christian Gouriéroux, Bruce Lehmann, Allan Timmermann, and seminar participants at the "First Symposium on Econometric Theory and Applications (SETA)" for useful comments. The authors also thank Dan Waggoner for providing the raw data and the Fortran codes used to generate the zero-coupon yields. Corresponding author: Pierluigi Balduzzi, Carroll School of Management, Boston College, 140 Commonwealth Avenue, Chestnut Hill, Massachusetts, 02467; Tel: (617) 552-3976; Fax: (617) 552-0431; email:balduzzp@bc.edu.

A Simple Test of the Affine Class of Term Structure Models

ABSTRACT

This paper develops a simple test of the affine class of term structure models. Affine term structure models imply an affine relation between yields and factors, and between yields and yields. We test whether a set of yield changes is linearly related to a small set of changes in empirical factors, which, in turn, are linear combinations of other yields. The test leads to only weak rejections of the affine class. An out-of-sample hedging exercise also supports the validity of the affine class: the constant hedge ratios implied by the affine class generally outperform time-varying weights.

JEL # G12

Keywords: affine term structure models, semi-nonparametric specification, bootstrap

Introduction

Most of the existing term structure models belong to the affine class. As shown by Duffie and Kan (1996), if the instantaneous riskless rate (the "short" rate) is an affine function of the underlying factors, and if the drifts and volatilities of the processes followed by the underlying factors, under the risk-adjusted measure, are also affine in the factors, then the zero-coupon yields of all maturities are affine functions of the factors and the converse is also true. Some of the examples of affine models are those developed by Vasicek (1977), Cox, Ingersoll, and Ross (1985) (CIR), Dai and Singleton (2000), Duffee (2002), Ang and Piazzesi (2003), and Duarte (2004). Affine models have computational advantages over other models, since they specify a log-linear dependence of bond prices on the risk variables. The coefficients of this dependence are the solutions to a system of Riccati equations, which can sometimes be calculated in closed-form (e.g. Vasicek, 1977, and CIR); and even when numerical techniques are needed, the computational burden is modest. Affine models also have advantages from the standpoint of estimation, allowing for estimation by method of moments and maximum-likelihood (see Dai and Singleton, 2003, for a review).

This paper proposes a new and simple test of this important class of term structure models. A distinctive feature of the affine class is that it is possible to uniquely invert from the factors to the observable yields, and all yields can be expressed as linear functions of as many other yields (or linear combinations of yields) as the number of underlying factors of uncertainty. This observation suggests a simple regression-based test of the affine class. We define three linear combinations of yields as the "empirical factors:" the level, slope, and curvature of the yield curve. To offset the high persistence of yields and factors, we formulate the regression-based test in first differences, where the first differences of a yield are a linear function of the first differences of the factors and of the first differences of *nonlinear* functions of the factors. If the nonlinear terms are jointly significant in a Wald-style test, then we have evidence against the affine class. We consider four popular semi-nonparametric (SNP) nonlinear specifications: simple polynomials, Legendre polynomials, Fourier transforms, and Hermite polynomials.

We consider weekly data for U.S. Treasury rates for the 1961–2002 sample and for four subsamples, corresponding to different operating procedures of the Federal Reserve. Overall, we document only mild rejections of the affine class. For example, in our baseline test exercise, the simple-polynomial specification leads to rejections at the 5% level in only two out of 40 tests.

Importantly, in the test illustrated above, we correct the critical values of the Wald statistic with a bootstrap. Indeed, a simulation shows that a bootstrap adjustment is crucial to obtain tests with the correct size. The simulation also shows that even with a bootstrap adjustment some of the SNP specifications lead to tests that substantially over-reject the null. Finally, the simulation shows that bootstrap-corrected tests have adequate power when the data is generated by a multi-factor quadratic term structure model (Ahn et al., 2002), particularly when the test is implemented on individual yields.

We further investigate the validity of the affine class through a hedging exercise. The affine class of term structure models implies that hedge ratios are constant over time. We compare the out-of-sample performance of hedged portfolios with constant weights and weights time-varying as a function of the empirical factors. We find that in most cases the constant-weight hedging strategy outperforms the strategies with time-varying weights, particularly if the weights include high-order, nonlinear functions of the empirical factors. Indeed, in our baseline hedging exercise, the constant-weight portfolio beats the time-varying-weight portfolios in 36 out of 40 scenarios.

While our approach is based on SNP specifications, fully nonparametric tests of the affine class of term structure models have been developed by Aït-Sahalia (1996a,b) and Stanton (1997). Aït-Sahalia (1996a) compares the parametric density of the short rate implied by affine one-factor models (Vasicek, 1977, and CIR) to the same density estimated nonparametrically. The comparisons between the implied parametric densities and the estimated nonparametric density leads to a rejection of the affine models. Aït-Sahalia (1996b) estimates the drift and diffusion terms of the short rate based on the density of the short rate, finding strong evidence of nonlinearities in both. Stanton (1996) uses a new technique to estimate the drift and volatility terms of the instantaneous riskless rate process. The drift and volatility are modeled as possibly nonlinear functions of the short rate by kernel regression. He finds evidence of nonlinearities in both the drift and the volatility. Yet, the results of Aït-Sahalia (1996a) have been disputed by Pritsker (1998), who argues that Aït-Sahalia's test rejects the affine null too often, because interest rate data are highly persistent, while the asymptotic distribution of the test treats the data as i.i.d. In addition, Chapman and Pearson (2000) dispute the results of both Aït-Sahalia (1996a) and Stanton (1997). They show in a simulation that the drift estimators of Aït-Sahalia (1996a) and Stanton (1997) would suggest nonlinearities when the process is, in fact, affine.

A related paper that takes an SNP approach similar to that of the present study is Ghysels and Ng (1998). Ghysels and Ng (1998) model the drifts of various yields as nonlinear (quadratic) functions of two yield spreads. Their tests, based on asymptotic critical values, find the quadratic terms to be statistically significant.

Our methodology and results contribute in several ways to the literature above. First, we develop a simple regression-based test of the affine class, and a simulation exercise demonstrates that with a bootstrap correction the test can deliver the correct size and adequate power. Second, since we focus on the contemporaneous relations among yields, we do not have to worry about the issues arising in the estimation of discretely sampled continuous-time processes. Third, we test a linear relation between yield changes, which is consistent with the state variables following affine processes under the risk-adjusted measure, but not necessarily under the actual measure. Hence, our null hypothesis allows for departures from the completely affine (Dai and Singleton, 2000) and essentially affine (Duffee, 2002) paradigms, such as the semiaffine square-root model of Duarte (2004). Fourth, our tests allow for a multi-factor affine null hypothesis, whereas the studies above mainly focus on one-factor models. Fifth, our out-of-sample hedging exercise provides evidence of the performance of the affine class in the context of a practical application. Finally, our test results are supportive of the affine class and, hence, they are broadly consistent with the conclusions of Pritsker (1998) and Chapman and Pearson (2000).

The paper is organized as follows: Section I describes the methodology for modeling the contemporaneous relationship between yields and yields. Section II performs the test schemes on yield data. Section III performs the hedging exercises. Section IV concludes the paper.

I. Testing the Affine Class

This section describes the baseline testing scheme. We define empirical yield curve factors, and describe SNP specifications and asymptotic and bootstrap test statistics. Other alternative testing schemes are discussed in Section II when applicable.

A. Empirical Yield Curve Factors

It is well documented that one needs at least three factors to explain the term structure.[1] We consider the following three "empirical" factors

$$\text{Level:} \quad X_t^{(1)} = Y_{0.25,t}, \tag{1}$$

$$\text{Slope:} \quad X_t^{(2)} = Y_{8,t} - Y_{0.25,t}, \tag{2}$$

$$\text{Curvature:} \quad X_t^{(3)} = (Y_{8,t} - Y_{2,t}) - (Y_{2,t} - Y_{0.25,t}), \tag{3}$$

where $Y_{\tau,t}$ is a τ-maturity zero-coupon yield at t, $t = 1, \ldots, T$, and time is measured in years. In the affine setting, the above three factors are a rotation (affine transformation) of the unobservable underlying factors.

B. SNP Specifications and Estimation

We consider the following general SNP specification,

$$Y_{\tau,t} = \alpha_\tau + \sum_{j=1}^{3} \beta_\tau^{(j)} X_t^{(j)} + \sum_{j=1}^{3} m_\tau(X_t^{(j)}; \gamma_\tau^{(j)}) + u_{\tau,t}, \tag{4}$$

where α_τ is a scalar intercept term, $m_\tau(X_t^{(j)}; \gamma_\tau^{(j)})$ is nonlinear in $X_t^{(j)}$ with an $M^{(j)} \times 1$ parameter vector $\gamma_\tau^{(j)}$, where we restrict $M^{(1)} = M^{(2)} = M^{(3)} = M$, and $u_{\tau,t}$ is the error term. The form of $m_\tau(X_t^{(j)}; \gamma_\tau^{(j)})$ depends on the choice of SNP specification. Our baseline testing scheme considers three test assets: $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$. For each factor $X_t^{(j)}$, we consider M nonlinear terms, where M is chosen using the Akaike information criterion (AIC). Section II.D further addresses the model selection issue.

[1]See, for example, Knez, Litterman, and Scheinkman (1994) and Chapman and Pearson (2001).

We consider four different SNP specifications and describe their generic forms as follows. While we set at 5 the maximum number of terms in each of the generic forms for $m(.;.)$, the AIC may lead to selecting $M < 5$.[2] The first specification is a simple polynomial, where (for notational convenience, in the polynomial specifications we drop the subscript τ and the superscript (j))[3]

$$m(X_t;\gamma) = \gamma_1 X_t^2 + \gamma_2 \frac{1}{X_t} + \gamma_3 X_t^3 + \gamma_4 X_t^4 + \gamma_5 X_t^5. \tag{5}$$

We de-mean and standardize each $X_t^{(j)}$ before forming the polynomials to reduce the chances of singular moment matrices.[4]

The second specification is a linear combination of orthogonal Legendre polynomials, where[5]

$$m(X_t;\gamma) = \gamma_1 P_1(X_t) + \gamma_2 P_2(X_t) + \gamma_3 P_3(X_t) + \gamma_4 P_4(X_t) + \gamma_5 P_5(X_t), \tag{7}$$

with

$$P_1(X_t) = \frac{1}{2}(3X_t^2 - 1), \tag{8}$$
$$P_2(X_t) = \frac{1}{2}(5X_t^3 - 3X_t), \tag{9}$$
$$P_3(X_t) = \frac{1}{8}(35X_t^4 - 30X_t^2 + 3), \tag{10}$$
$$P_4(X_t) = \frac{1}{8}(63X_t^5 - 70X_t^3 + 15X_t), \tag{11}$$
$$P_5(X_t) = \frac{1}{16}(231X_t^6 - 315X_t^4 + 105X_t^2 - 5). \tag{12}$$

As is standard with Legendre polynomials, the original X_t variables are transformed so that the transformed variable falls in the $[-1, 1]$ region, the region over which the polynomials are orthogonal to each other.

The third specification is a Fourier transform, where[6]

$$m(X_t;\gamma) = \gamma_1 X_t^2 + \gamma_2 \sin(X_t) + \gamma_3 \cos(X_t) + \gamma_4 \sin(2X_t) + \gamma_5 \cos(2X_t). \tag{13}$$

Again, the different trigonometric terms are orthogonal to each other.[7] The original X_t variables are transformed so that the transformed variable falls in the $[-\pi, \pi]$ region.

---

[2]We set $M \leq 5$ to avoid numerical issues.

[3]A simple polynomial is used, for example, in modeling the drift of the short rate by Aït-Sahalia (1996a,b).

[4]By de-meaning the X_t variables we reduce the correlation between them. By standardizing the X_t variables, we limit the differences in their order of magnitude.

[5]In this context, orthogonality means

$$\int_X g_1(X) g_2(X) \varpi(X) dX = 0, \tag{6}$$

where $g_1(X)$ and $g_2(X)$ are two different polynomials and $\varpi(X)$ is a weighting function specific to the polynomial. In the case of Legendre polynomials, $\varpi(X) = 1$.

[6]A quadratic term is present in all SNP specifications, because the minimization of the Schwartz information criterion (SIC) leads to a SNP specification with a linear and a quadratic term.

[7]In this case, the weighting function equals one.

The fourth specification is a linear combination of orthogonal Hermite polynomials,[8]

$$m(X_t;\gamma) = \gamma_1 H_1(X_t) + \gamma_2 H_2(X_t) + \gamma_3 H_3(X_t) + \gamma_4 H_4(X_t) + \gamma_5 H_5(X_t), \tag{14}$$

where

$$H_1(X_t) = 4X_t^2 - 2, \tag{15}$$
$$H_2(X_t) = 8X_t^3 - 12X_t, \tag{16}$$
$$H_3(X_t) = 16X_t^4 - 48X_t^2 + 12, \tag{17}$$
$$H_4(X_t) = 32X_t^5 - 160X_t^3 + 120X_t, \tag{18}$$
$$H_5(X_t) = 64X_t^6 - 480X_t^4 + 720X_t^2 - 120. \tag{19}$$

As in the case of the simple polynomial, we de-mean and standardize each X_t before forming the polynomial to reduce the possibility of singular moment matrices.

Empirically, both $Y_{\tau,t}$ and X_t are highly persistent and the strong autocorrelation in $u_{\tau,t}$ is a serious concern when making statistical inference. To mitigate the issue, we consider a regression in the first differences,

$$\Delta Y_{\tau,t} = \sum_{j=1}^{3} \beta_\tau^{(j)} \Delta X_t^{(j)} + \sum_{j=1}^{3} \Delta m_\tau(X_t^{(j)};\gamma_\tau^{(j)}) + \Delta u_{\tau,t}. \tag{20}$$

The model in (20) can be easily generalized to a multi-equation setting in which several yields, $Y_{\tau,t}$ are considered and M, which is the same for all equations, is selected by the AIC.

C. Hypothesis Testing

We test the null hypothesis of linearity by means of a Wald-style test of the joint significance of the γ_τ parameters associated with the nonlinear terms. The covariance matrix of the estimates is either White's (1980) heteroskedasticity-consistent covariance matrix estimator or Newey-West's (1987) heteroskedasticity- and autocorrelation-consistent covariance matrix estimator, where the number of lags is determined by the number of significant autocorrelations in $\Delta u_{\tau,t}$. The test easily generalizes to the case of multiple yields.

Our baseline test examines the following three null hypotheses: $\gamma_1 = 0$, $\gamma_5 = 0$, and $\gamma_{10} = 0$, individually, and also tests the joint hypothesis of $\gamma_1 = \gamma_5 = \gamma_{10} = 0$.

[8] In this case, the weighting function is proportional to the normal density.

D. Tests Based on Bootstrap-corrected Critical Values

Throughout this paper, we use a "residual-based" bootstrap method to attain asymptotic refinements. The steps are as follows. Recognizing that the null hypothesis is linearity, we run the restricted regression,

$$\Delta Y_{\tau,t} = \sum_{j=1}^{3} \beta_{\tau}^{(j)} \Delta X_t^{(j)} + \Delta u_{\tau,t}, \tag{21}$$

and save the estimated coefficients and residuals. We use a block bootstrap to jointly resample regressors and residuals, so that the bootstrap data preserve the possible higher-moment dependence between factors and noise.

Suppose we obtain a bootstrap sample of regressors and residuals with length $T-1$. Using the coefficient estimates, we can form a bootstrap sample for the dependent variable, $\Delta Y_{\tau,t}$. Next, we form a bootstrap sample of the empirical factors $X_t^{(j)}$ in levels: we randomly draw an $X_t^{(j)}$ and treat it as the initial value. We apply it to the bootstrap $\Delta X_t^{(j)}$ to restore a $T \times 1$ time series of $X_t^{(j)}$ in levels, which enables us to obtain nonlinear terms in $m(.;.)$ and estimate the regression model in (20). We then test the null hypotheses $\gamma_1 = 0$, $\gamma_5 = 0$, $\gamma_{10} = 0$, and $\gamma_1 = \gamma_5 = \gamma_{10} = 0$, and we compute the corresponding Wald statistics. We repeat the above procedure B times and save B Wald statistics. Based on the empirical distribution of the B bootstrap Wald statistics, we can compute the bootstrap p-values for given sample Wald statistics.

II. Empirical Results of the Regression Tests

A. Data Sources

We use an extended version of the Bliss (1997) dataset for zero-coupon yields, where the yields estimated by the unsmoothed Fama-Bliss algorithm,[9] using all available issues and assuming no tax effects. The original dataset is at the daily frequency, and we form weekly datasets by extracting Wednesday-to-Wednesday data to avoid possible weekend effects. When Wednesday is not a trading day, we use the data from the next business day. The resulting weekly dataset covers the period from June 14, 1961 to December 26, 2002, for a total of 2,167 observations. Note that observations from September 11, 2001 (Tuesday) to September 20, 2001 (Thursday) are missing,[10] and we use the September 21, 2001 (Friday) observation and then resume our sampling of Wednesday data. Relative to other datasets used in the literature (e.g. the monthly McCulloch and Kwon (1993) dataset, which will be studied in section D.5), the Bliss dataset has the advantage of allowing us to generate higher frequency data, and a large number of observations.

[9]See Fama and Bliss (1987), Bliss (1997), and Waggoner (1997), for more details about the estimation methods for yield curves. Notice that although Bliss (1997) prefers the "smoothed" Fama-Bliss algorithm to the "unsmoothed" Fama-Bliss algorithm, the former algorithm often produces irregular patterns, such as negative yields, zero yields, or sudden spikes, when it is applied to more recent data. The unsmoothed Fama-Bliss algorithm does not suffer from such problems.

[10]U.S. financial markets were closed during the period September 12-September 17, 2001.

Table 1 reports summary statistics for $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$, and for the empirical yield curve factors, Level, Slope, and Curvature, for the whole sample period and four subsamples. The first subsample covers June 1961 through August 1971, during which the Bretton Woods agreement was in place and the operating target of monetary policy was free reserves. The second subsample runs from September 1971 to September 1979, during which the operating target of monetary policy was the growth of money aggregates. The third subsample is from October 1979 to September 1982, during which inflation was high and volatile and the operating target of monetary policy was non-borrowed reserves. The final subsample covers October 1982 through December 2002, during which inflation experienced an overall downward trend, and the operating target was the fed funds rate. Throughout the paper, we report empirical results for the whole sample as well as for the subsamples, to account for possible structural breaks.

Both yields and empirical yield curve factors are highly autocorrelated in the full sample and in all four subsamples. The first-order autocorrelation coefficients are all above 0.94, except for the Curvature factor. Overall, yields and the Level factor are more autocorrelated than Slope and Curvature. In addition, the third subsample features the lowest autocorrelations for both yields and factors, among the four subsamples.

B. Testing the Affine Class: Baseline Case

Table 2 presents results of the test of linearity, using data for the whole sample and for four subsamples. We are interested in testing the null hypothesis of the affine three-factor model against the alternative hypothesis of a nonlinear three-factor model. We consider all four SNP specifications, run regressions in first-differences, and report asymptotic and block-bootstrap p-values of the test statistics. We perform single-equation tests for each of the three yields $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$, and we also conduct joint tests using all three yields.

We first consider a simple-polynomial specification (Panel A of Table 2). The asymptotic tests strongly reject the three-factor affine model, in both single- and multiple-equation settings, with the exception of $Y_{10,t}$ in the first subsample and $Y_{5,t}$ in the four subsamples. Indeed, out of 40 test statistics, the corresponding p-values are below 5% in 30 instances. The inference changes dramatically when we consider bootstrap corrections. The bootstrap-corrected tests, based on 5,000 bootstrap replications of blocks of data of length 2, mainly fail to reject the null hypotheses at conventional significance levels. Indeed, out of 40 test statistics, the corresponding p-values are equal or below 5% only in two instances.

Model diagnostics show that the simple-polynomial specification captures well the time-series variation in yield changes. The adjusted R^2 is between 67% and 88% in the full sample, while the explanatory power is slightly lower in the first subsample, featuring adjusted R^2-s between 42% to 82%. Durbin-Watson statistics indicate that the residuals are only weakly autocorrelated in most instances.

When considering the Legendre polynomial specification, we still reject the null hypothesis based on the asymptotic p-values. Panel B of Table 2 shows that, out of 40 instances, the asymptotic tests reject the null hypothesis 33 times at the 5% significance level. When we turn to bootstrap tests, we find slightly stronger rejections in the full sample, possibly driven by the results in the fourth subsample where the null hypothesis is rejected in all instances. In the first three subsamples, on the other hand, none of the p-values are below 10%. Overall, there are 14 rejections at the 5% level in 40 bootstrap tests.

Panel C of Table 2 uses the Fourier transform specification. Similar to the other specifications, the asymptotic tests reject the null hypothesis of the three-factor affine model in 34 out of 40 instances (at the 5% significance level); whereas, after the bootstrap adjustment, we only observe mild rejections (ten out of 40 tests), driven by the data in the fourth subsample. The SNP specification based on Hermite polynomials results in similar patterns of the test statistics: the asymptotic tests reject the null 33 times in 40 instances, while the bootstrap tests reject 14 times; see Panel D of Table 2 for details.

Our baseline test results can be summarized as follows: First, tests based on asymptotic p-values seem to suffer from small-sample biases and tests based on bootstrap-corrected p-values dramatically change the inference. While asymptotic tests strongly reject the null hypothesis of three-factor affine term structure models, we only observe mild rejections when the bootstrap-corrected critical values are used instead. Second, the simple-polynomial specification leads to weaker rejections of the affine class, particularly in the case of bootstrap-corrected tests.

C. Simulation Evidence

In the following, we describe the results of a simulation where we generate data from realistically-calibrated affine and quadratic term structure models (see Ahn et al., 2002), and where we implement the regression-based test (20) using both asymptotic and bootstrap-corrected critical values. Details of the two models and of the simulation exercises are in the Appendix. When we simulate an affine model, we want to see whether p-values based on asymptotics or based on bootstrap corrections are close to the nominal sizes of the test. When we simulate a quadratic model, we want to see how often the null of linearity is rejected for different nominal sizes of the test.

There are several important reasons for performing this simulation exercise. First, when we bootstrap the regression model under the null of linearity, we are not imposing the full set of restrictions deriving from an affine term structure model. Hence, it is not obvious that, even with a bootstrap correction, one can obtain appropriate p-values under the null. Second, if we want to study the power of our tests, we need to simulate a nonlinear model. If we simply bootstrap data based on the nonlinear regression model (20), there is no assurance that the data that we generate are consistent with any well-defined term structure model. Moreover, we would generate different bootstrap data depending on the SNP specification, making comparisons difficult. Third, we have seen in the previous section that tests based on different SNP specifications lead to somewhat different inference, even with a bootstrap adjustment. Without a simulation exercise, we do not know which results to trust.

We first consider the size of the test. We want to test the null hypothesis for the individual yields $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$, and for the three yields jointly. For a given combination of SNP specification (simple polynomial, Legendre polynomial, Fourier transform, or Hermite polynomial), covariance estimator (White, 1980, or Newey and West, 1987), we obtain an empirical Wald statistic and a p-value for each path simulated from the affine model. By collecting 1,000 Wald statistics and p-values and calculating the rejection probabilities across 1,000 simulations we can assess the size of the SNP test for given nominal significance levels (1%, 5%, and 10%); bootstrap-corrected tests are based on 500 bootstrap samples of a given simulated path.[11]

We find that bootstrap critical values provide a significant improvement, particularly for the simple-polynomial specification (detailed results available upon request). For example, in the case of the joint test and the simple-polynomial specification (Newey-West covariance matrix) the actual rejection rates are 1.5%, 6.3%, and 10.6% (theoretical sizes: 1%, 5%, and 10%, respectively). On the other hand, rejection rates are still high for the other SNP specifications. For example, in the case of the Legendre Polynomial specification, actual sizes are 10.8%, 22.4%, and 32.4%.

Next, we assess the power of the tests. Given the previous results, we focus on the simple-polynomial specifications. We simulate 1,000 paths from a quadratic model and we perform our tests on the yields $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$, for each path, and save 1,000 Wald statistics and p-values. We assess the power of the tests for given significance levels (1%, 5%, and 10%) by calculating the rejection probabilities across 1,000 simulations.

As one would expect, rejection rates are high when asymptotic critical values are used (detailed results available upon request). For example, in the case of the joint test (Newey-West covariance matrix) the rejection rates are 100% for all three critical values (1%, 5%, and 10%). When the critical values are bootstrap-corrected, rejection rates are much lower: 1.6%, 11.6%, and 48.9%. Note, though, that for the bootstrap-corrected tests, rejection rates are much higher on for the tests on individual yields, particularly for the intermediate and shorter maturities. For example, in the tests for the $Y_{5,t}$ yield rejection rates are 33.1%, 99.9%, and 100.0%, for the three critical values, respectively.

In summary, the simulations confirm the need for a bootstrap correction of our tests. Moreover, they show that, even with a bootstrap correction, three of the four SNP specifications lead to over-rejections. Finally, they show that for the simple-polynomial specification the power of the bootstrap-corrected tests is adequate, especially for the intermediate- and shorter-maturity yields.

[11] Note that the convergence of p-values is very quick as the number of bootstrap samples increases. Indeed, we obtain essentially the same results for 100 bootstrap samples, for each simulated sample. Moreover, for the actual sample data set, we increase the number of bootstrap samples from 500, to 1,000, 5,000, and 10,000 and find only marginal changes in the bootstrap-corrected p-values.

D. Robustness

In this section, we consider an array of variations of the SNP tests to assess the robustness of the results in section B. We consider alternative orders in the SNP specifications, definitions of yield curve factors, test assets, bootstrap methods, and a dataset with lower frequency. Given the simulation results from the previous section, the tables only report results obtained from the simple-polynomial specification. Results obtained from the other three SNP specifications are available upon request.

D.1. Model Selection

We examine whether the number of nonlinear terms affects their joint explanatory power. Our baseline models use the AIC to determine the number of nonlinear terms, and the AIC tends to include 4 to 5 nonlinear terms for each empirical factor, with the only exception that x^2 is the sole nonlinear term when $Y_{5,t}$ is the test asset in a regression with simple polynomials.

We use the SIC as an alternative criterion to select the models. We find that only the square terms are chosen. While not reported in the tables (results available upon request), the asymptotic p-values are below 5% in 12 out of 40 tests, and all the rejections are either from the full sample or from the fourth subperiod. Bootstrap-corrected p-values are larger than asymptotic ones in almost all instances, leading to six rejections at the 5% level, all of them in the fourth subperiod.

We also consider an intermediate case, in which only the first M nonlinear terms are included, such that $1 < M < 5$. Specifically, we set $M = 2$ for the simple polynomial, Legendre polynomial, and Hermite polynomial specifications, and $M = 3$ for the Fourier transforms so that both the sine and cosine terms are included. Results for the simple-polynomial specification are reported in Table 3. For the asymptotic tests, p-values are lower than in the most parsimonious case (orders determined by SIC), but higher than in the baseline case (orders determined by AIC). Based on asymptotic tests, except for the first subperiod, the null hypothesis is rejected in most cases. However, with a bootstrap correction, we have only two rejections out of 38 tests[12].

D.2. Alternative Factors

Our baseline test uses the yields $Y_{0.25,t}, Y_{2,t}$ and $Y_{8,t}$, to define the empirical yield curve factors. We can use other yields instead to define the three factors, e.g.,

$$\text{Level:} \quad X_t^{(1')} = Y_{0.5,t}, \tag{22}$$

$$\text{Slope:} \quad X_t^{(2')} = Y_{7,t} - Y_{0.5,t}, \tag{23}$$

$$\text{Curvature:} \quad X_t^{(3')} = (Y_{7,t} - Y_{3,t}) - (Y_{3,t} - Y_{0.5,t}). \tag{24}$$

[12]We are unable to invert the covariance matrices in the multi-equation specification in the third subsample.

The change in the definition of the factors does not change the basic results: We find strong rejections in the asymptotic tests, but only mild rejections using bootstrap-corrected p-values (results available upon request).

Alternatively, we replace the empirical yield curve factors with the first three principal components extracted from a broad set of yields. Specifically, we extract three principal components from $\{\Delta Y_{\tau,t}\}$, for $\tau = 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7, 8, 9$, and 10, and we apply the ordered eigenvectors to $\{Y_{\tau,t}\}$ to form alternative factors, which are also linear combinations of yields. With these factors, we can form the linear and nonlinear terms in the SNP specifications and test the overall explanatory power of the nonlinear terms. Table 4 shows that the asymptotic tests still strongly reject the null of linearity (at the 5% level in 34 out 40 tests) and bootstrap-corrected tests do not reject the null in the full sample and the fourth subsample, although they occasionally reject the null in the first three subsamples (for a total of 11 rejections). Hence, with the principal components as empirical factors, we obtain somewhat stronger rejections than in the baseline case of Table 2.

D.3. Alternative Test Assets

Our baseline model uses the three yields $Y_{1,t}$, $Y_{5,t}$, and $Y_{10,t}$. Alternatively, we can use $Y_{0.5,t}$, $Y_{3,t}$, and $Y_{7,t}$ and perform individual and joint tests (results available upon request). When a simple-polynomial specification is used, the null hypothesis is strongly rejected in the asymptotic tests. With bootstrap-corrected critical values, we reject the null hypotheses under the 5% significance level in seven out of 40 tests, and six of them are from the fourth subsample. When using SNP specifications other than simple polynomials, we observe a few more rejections, mostly from the full sample and the fourth subperiod.

D.4. Alternative Bootstrap Methods

We consider two other methods to construct the bootstrap-corrected p-values. Table 5 uses a sieve bootstrap to resample the vector $\mathcal{X}_t$, which includes the empirical factors and the residuals of the regression model (20). We fit a vector autoregressive system,

$$\mathcal{X}_t = \sum_{i=1}^{I} A_i \mathcal{X}_{t-i} + v_t, \tag{25}$$

where I is determined by the SIC to avoid overfitting. We randomly draw a set of factor realizations $X_t^{(j)}$ and treat them as the initial values, resample v_t, and form restored factors. Table 5 shows again that the bootstrap-corrected tests are much less likely to reject the null hypothesis (13 rejections) than are the asymptotic tests (30 rejections), especially in the subsamples.

Another bootstrap method that we consider is the stationary bootstrap of Politis and Romano (1994). To construct the stationary bootstrap samples, we resample blocks of empirical-factor realizations and residuals, with random block lengths determined by a geometric distribution with parameter 0.5, such that the average length is 2. The bootstrap tests fail to reject the null hypothesis in 39 out of 40 instances, and the bootstrap-corrected p-values resemble in magnitude those in Table 2 (results available upon request).

D.5. McCulloch-Kwon Dataset

To check the robustness of our findings against the choice of dataset, we also perform our tests using the McCulloch and Kwon (1993) data set, covering the 1946–1991 period at the monthly frequency. This is the dataset used by Ahn et al. (2002) in the estimation of affine and quadratic term structure models. Results from this alternative dataset confirm our previous conclusions. Tests based on the asymptotic distribution of the statistics lead to strong rejections of the affine class. On the other hand, when we use the empirical distributions of the statistics, we observe only mild rejections. Table 6 shows 33 rejections in 40 asymptotic tests, and 11 rejections in 40 bootstrap-corrected tests.

III. Hedging

This section uses a hedging exercise to validate our results obtained from the previous section. We relate the hedging strategies to distinct classes of term structure models, and we propose several competing hedging strategies. We compare the in-sample and out-of-sample hedging performance of the different strategies.

A. Hedging Strategies and Term Structure Models

A.1. Setup

We consider a situation where we want to hedge a bullet portfolio with maturity τ_h in a three-factor economy. The hedging portfolio has weights $w_{\tau_1,t}$, $w_{\tau_2,t}$, and $w_{\tau_3,t}$ on three other zero-coupon bonds with maturities τ_1, τ_2, and τ_3, respectively; the fraction $1 - w_{\tau_1,t} - w_{\tau_2,t} - w_{\tau_3,t}$ of the hedging portfolio is invested in the riskless asset.

Consider the representation

$$\begin{aligned} RET_{\tau_h,t+\Delta t} &= w_{\tau_1,t} RET_{\tau_1,t+\Delta t} + w_{\tau_2,t} RET_{\tau_2,t+\Delta t} + w_{\tau_3,t} RET_{\tau_3,t+\Delta t} \\ &\quad + [1 - w_{\tau_1,t} - w_{\tau_2,t} - w_{\tau_3,t}] RF_t + \varepsilon_{\tau_h,t+\Delta t}, \end{aligned} \tag{26}$$

$$RET_{\tau,t+\Delta t} \equiv \frac{p_{\tau-\Delta t,t+\Delta t}}{p_{\tau,t}} - 1, \tag{27}$$

where $RET_{\tau,t+\Delta t}$ denotes the Δt-period return of a τ-year zero-coupon bond invested at t, $RF_t \equiv RET_{\Delta t,t+\Delta t}$ is the risk-free rate for a Δt-period investment, known at t, and $p_{\tau,t}$ is the time-t price of a zero-coupon bond maturing at time $t+\tau$. In order to hedge the bullet portfolio, at time t we have to find the weights of the τ_1-year, τ_2-year and τ_3-year bonds, $w_{\tau_1,t}$, $w_{\tau_2,t}$ and $w_{\tau_3,t}$, respectively, such that the variation in $RET_{\tau_h,t+\Delta t}$ is captured by the linear combination of the right-hand-side returns. The error term $\varepsilon_{\tau_h,t+\Delta t}$ is the "hedging error." The optimal hedging strategy minimizes the sum of squared hedging errors $\sum_{t=1}^{T} \varepsilon^2_{\tau_h,t+\Delta t}$, or equivalently, the mean squared hedging error.

Equation (26) has the following equivalent representation,

$$ER_{\tau_h,t+\Delta t} = w_{\tau_1,t} ER_{\tau_1,t+\Delta t} + w_{\tau_2,t} ER_{\tau_2,t+\Delta t} + w_{\tau_3,t} ER_{\tau_3,t+\Delta t} + \varepsilon_{\tau_h,t+\Delta t}, \tag{28}$$

where $ER_{\tau,t+\Delta t}$ is the excess return over the holding horizon Δt, defined as

$$ER_{\tau,t+\Delta t} = RET_{\tau,t+\Delta t} - RF_t. \tag{29}$$

The next step is to determine the functional forms of the portfolio weights. First, we examine the determinants of the return on a zero coupon bond, approximated by

$$\begin{aligned} RET_{\tau,t+\Delta t} &\approx \ln p_{\tau-\Delta t,t+\Delta t} - \ln p_{\tau,t} \\ &= -(\tau - \Delta t) Y_{\tau-\Delta t,t+\Delta t} + \tau Y_{\tau,t}. \end{aligned} \tag{30}$$

For a very short rebalancing horizon, $\Delta t \to 0$, the change in yields converges to $dY_{\tau,t}$. Furthermore, the infinitesimal change in the factor $F_t^{(j)}$ results in the change in yield of $\left(\partial Y_{\tau,t} / \partial F_t^{(j)}\right) dF_t^{(j)}$, as suggested by existing term structure models. In the affine term structure setting, $\partial Y_{\tau,t} / \partial F_t^{(j)}$ is constant over time, whereas outside of the affine class the derivative is a function of the underlying factors and thus it is time-varying.

Motivated by this fundamental difference between affine and nonlinear term structure models, we consider two types of weight functions. If the term structure follows an affine model, the implied portfolio weights are constant over time. If the term structure follows a nonlinear process, the implied hedging portfolio weights are time-varying.

Suppose at time t the portfolio weight function for the zero coupon bond with maturity τ is modeled as

$$w_{\tau_i,t} = \alpha_{\tau_i} + \sum_{j=1}^{3} \beta_{\tau_i}^{(j)} X_t^{(j)} + \sum_{j=1}^{3} m_{\tau_i}(X_t^{(j)}; \gamma_\tau^{(j)}). \tag{31}$$

The constant portfolio weights are nested in (31) if we set $\beta_{\tau_i} = \gamma_{\tau_i} = 0$.

We consider five types of time-varying weights: if we set $\gamma_\tau = 0$, then the weights are affine in the empirical factors. We call such portfolios the "affine portfolios."[13] Furthermore, we can model higher-order nonlinearities by allowing for four SNP specifications for the nonzero γ_τ: If we define $m_{\tau_i}(X_t; \gamma_{\tau_i})$ according to (5), then we have a "simple-polynomial portfolio." Similarly, if we define $m_{\tau_i}(X_t; \gamma_{\tau_i})$ according to (7), (13), or (14), then we have a "Legendre-polynomial portfolio," a "Fourier-transform portfolio," or a "Hermite-polynomial portfolio," respectively. Collectively, we call these portfolios the "SNP portfolios." Note that the coefficients in (31) are constant over time, and the time variation in $w_{\tau,t}$ is driven by the change in the empirical factors. The coefficients in (31) can be estimated by least squares since the objective function is the mean squared hedging error.

[13]Note, though, that the affine portfolio is inconsistent with the affine class of term structure models, as the affine class implies constant weights.

A.2. Hedging Performance

For each hedging strategy, we examine its in-sample and out-of-sample hedging performance, measured by the root mean squared (hedging) error (RMSE). The in-sample performance is evaluated using all of the observations within the corresponding (sub-)samples. The out-of-sample performance exercise is implemented using four schemes: fixed-window, recursive, rolling-window, and exponentially-weighted moving average (EWMA). For the fixed-window scheme, parameters are estimated only once, using the data in the initial estimation window, and the parameter estimates are applied to the remaining observations out-of-sample. For the recursive scheme, parameters are updated periodically when a new observation is added to the initial window, and the parameter estimates are applied to the one-step-ahead observation. For the rolling-window scheme, parameters are also updated periodically when a new observation is available, but the most distant observation is dropped so that the estimation window size is fixed. For the EWMA scheme, in-sample hedging errors are assigned different weights, exponential in the remoteness of the observations. The time-t optimization problem is

$$\min \sum_{s=1}^{t} \xi^{t-s} \varepsilon_{\tau_h,s}^2, \tag{32}$$

where ξ is called the "decay factor," and is set to 0.95.[14] For all of the out-of-sample evaluations, we preserve the first 50% of observations of the corresponding sample period as the initial estimation window, and compute the RMSEs of the hedging portfolios.

For each of the nonlinear hedging portfolios, the dimension of γ_{τ_i} is determined by the AIC, and it is kept fixed. For the same SNP portfolio, we set M the same across τ_i, while M can be different for various SNP portfolios.

When evaluating the hedging performance, we compute the RMSE of the constant-weight portfolio (implied by affine term structure models) and the five different dynamic portfolios (implied by nonlinear term structure models). We compare these RMSEs to the RMSE of the excess return on the unhedged portfolio. In addition, we consider a benchmark hedging strategy based on a portfolio of the τ_1-year and the τ_3-year bonds, a "barbell portfolio," where $\tau_1 < \tau_h < \tau_3$, with the same modified duration as the bullet portfolio. This hedging strategy is an interesting benchmark as it does not require any estimation.[15,16]

B. Empirical Performance of Hedging Strategies: Baseline Case

In this subsection we consider a baseline hedging exercise to evaluate the out-of-sample performance of different hedging strategies. For the bullet portfolio, we set $\tau_h = 3$ or 7. For the bonds in the hedging portfolio, we consider

[14] RiskMetrics™ chooses 0.94 as the decay factor for daily data and 0.97 for monthly data; our choice of 0.95 falls in this range and tilts toward the value for daily data.

[15] We thank Bruce Lehmann for suggesting this hedging strategy.

[16] While the weights of the barbell portfolio are also time-varying, as modified durations change with yields, we only categorize portfolios with weights explicitly modeled as functions of the empirical factors as time-varying weight portfolios.

$\tau_1 = 1$, $\tau_2 = 5$, and $\tau_3 = 10$. The rebalancing period is one week ($\Delta t = 1/52$ year).[17]

Table 7 reports the hedging performance for the whole sample and for the four subsamples. Not surprisingly, the SNP portfolios perform best in-sample, although the affine portfolio is often close. As an example, consider hedging a 3-year bond over the whole sample. The RMSE for the unhedged portfolio is 0.555%. Hedging with the barbell portfolio reduces the RMSE to 0.311%, while the constant-weight portfolio has a RMSE of 0.202%. The four SNP portfolios all deliver a RMSE of 0.184%. The affine portfolio is close at 0.197%.

Comparing in-sample performance across subsamples, it is easiest to hedge interest rate risk in the first subsample, featuring the lowest RMSEs of 0.121% ($\tau_h = 3$) and 0.091% ($\tau_h = 7$). It is most difficult to hedge interest rate risk in the third subsample, featuring the lowest RMSEs of 0.364% ($\tau_h = 3$) or 0.886% ($\tau_h = 7$). Comparing Panels A and B, we find that, with the exception of the first subsample, it is harder to hedge the risk of a 7-year bond than of a 3-year bond.

We then examine how the hedging strategies fare out of sample. Interestingly, we now obtain the opposite results: the constant-weight portfolio generally performs the best, while the SNP portfolios perform poorly, sometimes even worse than the unhedged asset. Indeed, out of 40 out-of-sample hedging exercises (20 for each of the two bonds), the constant-weight portfolio delivers the smallest RMSEs in 35 scenarios, and beats the time-varying weight portfolios 36 times. When hedging a 7-year bond, the barbell portfolio has the best performance in 3 instances, all of them in the first subperiod.

Out of sample, the largest RMSEs are always delivered by the SNP portfolios. The affine portfolio, on the other hand, performs much better, and in a few instances (four out of 40) it outperforms even the constant-weight portfolio. As an example, consider again hedging a 3-year bond over the whole sample. The RMSE for the unhedged portfolio is 0.505%. Hedging with the barbell portfolio reduces the RMSE to 0.253%.[18] The constant-weight portfolio has RMSEs ranging between 0.157% and 0.162%. The SNP portfolios have RMSEs ranging between 0.248% and 3.634%. The affine portfolio, on the other hand, has RMSEs between 0.164% and 0.206%.

While not reported in the table, we also performed our analysis for the cases where the time-varying weights are functions of one and two factors, and where one and two bonds, respectively, are used to construct a hedging portfolio. These exercises are useful because there are fewer parameters to estimate in the one- and two-factor cases, and this could affect the comparison between constant-weight and time-varying weights strategies. In summary, these additional hedging exercises generally confirm the findings reported for the three-factor case. In sample, the SNP portfolios perform best, while they often perform poorly out-of-sample. Out-of-sample, the strategy with

[17] While not reported in the table, we also computed the percentage of times that each hedging strategy leads to a squared hedging error higher than the squared hedging error of the constant-weight strategy. With few exceptions, these percentages deliver the same conclusion as the comparison of RMSEs, indicating that our results are not driven by a few outliers.

[18] Note that RMSEs of the unhedged portfolio differ from the in-sample figures, because we are now selecting a subsample of the whole sample, covering only the last 50% of the observations. This is to allow for a direct comparison with the hedging strategies, which use the first 50% of the data for estimation.

time-varying weights that performs best is the affine-weight strategy. Relative to the three-factor case, now there are more instances where the affine-weight strategy outperforms the constant-weight strategy. One important observation in performing these tests is that, as the number of factors increases, RMSEs drop substantially, both in sample and out of sample. For example, consider the case of the 3-year bond and the out-of-sample performance of the constant-weight portfolio for the fixed-window case. The RMSE falls from 0.259% (one factor), to 0.186% (two factors), to 0.107% (three factors). This indicates that accounting for the correct number of factors in the term structure is crucial for hedging performance.

We can summarize the baseline-case results by saying that, in most cases, the constant-weight portfolio (implied by the affine class) outperforms the affine-weight portfolios out of sample. SNP portfolios perform poorly out of sample and can do even worse than the unhedged portfolios.

C. Robustness

In this section, we consider several variations of the hedging exercises to assess the robustness of the results in Section B. We consider alternative orders in the SNP weights, definitions of the yield curve factors, assets in the hedging portfolios, and lengths of the initial estimation window.

C.1. Model Selection

The poor out-of-sample performance of the time-varying weights featuring higher order terms is possibly due to overfitting. Similar to the case of model selection when testing nonlinearities, the AIC tends to choose models with more higher order terms when allowing for nonlinearities. In the baseline hedging exercise, the AIC sets $M = 5$ for almost all SNP specifications, with the only exception of $M = 4$ for the Fourier transform when $\tau_h = 3$. To address the likely overfitting problem, we perform a hedging exercise with the most parsimonious SNP specifications (M=1), and another exercise considering an intermediate case. Note that the variations considered here only involve the SNP portfolios, and the RMSEs of other strategies are the same as the results shown in Table 7.

Table 8 presents results for the intermediate case: it sets $M = 3$ for the Fourier-transform portfolios and $M = 2$ for the other hedging portfolios with SNP weights. While the relative rankings of the hedging portfolios' in-sample and out-of-sample RMSEs do not vary much, the out-of-sample performance of the SNP portfolios improves when we reduce the number of terms. Compared with Table 7, the in-sample RMSEs of the SNP portfolios increase several basis points, but the out-of-sample RMSEs of those portfolios are much lower. Taking again the full-sample results for $\tau_h = 3$ as an example, the time-varying weights of the SNP portfolios produce RMSEs ranging from 0.174% to 0.656%. Comparing the 16 corresponding hedging errors, none of the RMSEs in Table 8 is larger than the corresponding one in Table 7.

In the most parsimonious case, the only higher-order terms are the squared terms, and thus all four SNP hedging portfolios are the same (results available upon request). In this case, the in-sample performance worsens slightly, but the out-of-sample performance improves further. For example, when hedging a three-year bond in the full sample, the range of out-of-sample RMSEs for the SNP portfolios is between 0.170% and 0.288%. When we look at subperiods, all SNP portfolio out-of-sample RMSEs are well below 1%, with only one exception in the third subperiod.

In summary, including higher-order terms deteriorates the out-of-sample hedging performance of the SNP portfolios. We can improve their performance by employing more parsimonious models. However, even in the most parsimonious setting, the SNP portfolios are still outperformed by affine- or constant-weight portfolios, with only few exceptions.

C.2. Alternative Factors

In the baseline hedging exercise, we model the time-varying weights as functions of empirical factors, defined as in (1)-(3). As in section II, we now consider two alternative sets of factors: the empirical factors defined in (22)-(24) and the principal components extracted from yield curve data. Note that these changes only affect the performance of strategies with time-varying weights and the RMSEs of the unhedged portfolio. The performance of the barbell portfolio and the constant-weight portfolio are the same as in Table 7.

When we use the factors defined by Equations (22)-(24), the affine portfolio is able to outperform the constant-weight portfolio in more instances out of sample, as compared to Table 7 (results available upon request). For example, in the full sample and when $\tau_h = 3$, the affine portfolio's RMSEs are 0.158% for both the recursive and rolling evaluation schemes, which is 0.1 to 0.4 basis points better than the constant-weight portfolio. The advantage is even stronger in the first subperiod when hedging against the risk of the 7-year bond out of sample, for the recursive, rolling, and EWMA schemes. Yet, in general, the constant-weight portfolio is still preferred out of sample: it has the lowest RMSEs among all strategies in 31 out of 40 out-of-sample hedging exercises.

Table 9 uses the principal components to model time variation in the hedging portfolio weights. The results are qualitatively and quantitatively very similar to the baseline case in table 7.

In summary, changing the definition of the empirical factors driving the time-varying portfolio weights affects only marginally the results, and the superior out-of-sample performance of the constant-weight portfolios still holds.

C.3. Alternative Assets in Hedging Portfolios

In this section, we examine whether including different assets in the hedging portfolios affects the relative performance of the hedging strategies. Specifically, we set $\tau_1 = 0.25$, $\tau_2 = 2$, and $\tau_3 = 8$, as opposed to 1, 5, and 10, respectively, in the baseline case.

When forming hedging portfolios using these alternative assets, we obtain results similar to the baseline case (results available upon request). Out of sample, the constant-weight portfolio beats all other portfolios in 27 out of 40 exercises. While the affine portfolio performs better than the SNP portfolios, it only outperforms the constant-weight portfolio three times, with the largest RMSE difference being 0.5 basis points.

C.4. Changing the Initial Estimation Window

Our last robustness check explores whether changing the length of the initial estimation window affects the out-of-sample performance of the different strategies. Our baseline analysis uses the first 50% of observations in each sample as the initial estimation window. We now expand it to 70% and shorten it to 30%.

Table 10 shows the results with the initial estimation window covering 70% of sample observations. The relative ranking of the hedging performance remains the same: SNP portfolios having the smallest RMSEs in-sample; whereas out of sample the constant-weight portfolios are the preferred strategy. Moreover, it appears that a larger initial estimation window mitigates the issue of estimation error, and leads to enhanced hedging performance. For example, when hedging a three-year bond out of sample in the full period, the constant-weight portfolio has RMSEs between 0.100% and 0.107%, and the range for affine portfolio is 0.111% to 0.127%, and 0.109% to 0.996% for the Fourier transform portfolios. Recall that in Table 7 the corresponding ranges are 0.157% to 0.162%, 0.164% to 0.206%, and 0.248% to 1.100%, respectively.

When including 30% of observations in the initial window (results available upon request), we find that a constant-weight portfolio is still the preferred hedging strategy. For both the constant-weight and time-varying weight portfolios, we find some deterioration in hedging performance.

IV. Conclusions

This paper develops a new and simple test of the affine class of term structure models. The affine class of term structure models implies affine relations between yields and factors, and between yields and yields. Hence, we test whether a set of yield changes is linearly related to a small set of changes in empirical factors, which, in turn, are linear combinations of other yields.

The test leads to only weak rejections of the affine class. Moreover, we consider hedging strategies, where the hedge ratios are constant (as implied by the affine class) or time-varying (as implied by nonlinear models). In most instances, the constant-weight strategies outperform the time-varying-weight strategies.

In summary, we regard our evidence as broadly supportive of the affine class of term structure models.

Appendix: Simulation Exercise

We first consider Chen's (1996) three-factor affine term structure model.[19] Following Dai and Singleton (2000), Chen's (1996) model has the representation

$$d\begin{pmatrix} F_t^{(1)} \\ F_t^{(2)} \\ F_t^{(3)} \end{pmatrix} = \begin{pmatrix} \kappa_{11} & 0 & 0 \\ 0 & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \\ 0 \end{pmatrix} - \begin{pmatrix} F_t^{(1)} \\ F_t^{(2)} \\ F_t^{(3)} \end{pmatrix} \right] dt \tag{33}$$

$$+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \sqrt{\begin{pmatrix} S_t^{(11)} & 0 & 0 \\ 0 & S_t^{(22)} & 0 \\ 0 & 0 & S_t^{(33)} \end{pmatrix}} dW_t \tag{34}$$

$$\equiv \mathcal{K}(\Theta - F_t)dt + \Sigma\sqrt{S_t}dW_t, \tag{35}$$

where F_t is the vector of latent factors at t, $S_t^{(ii)}$ are affine transformations of F_t, and W_t is a 3×1 vector of mutually independent standard Brownian motions. The short riskless rate is an affine function of F_t. As a result, the price at time t of a bond with maturity τ, $p_{\tau,t}$, is exponential affine in the factors, i.e.,

$$p_{\tau,t} = \exp[A(\tau) - B(\tau)^\top F_t], \tag{36}$$

where $A(\tau)$ is a scalar and $B(\tau)$ is a 3×1 vector, satisfying the Riccati equations specific to Chen's (1996) model, with boundary conditions $A(0) = 0$ and $B(0) = 0_3$. The bond yield with maturity τ is given by

$$Y_{\tau,t} = -\frac{1}{\tau}[A(\tau) - B(\tau)^\top F_t], \tag{37}$$

which is an affine function of the factors.

We calibrate the economy using the estimates in Ahn et al. (2002). We simulate the time series of the factors (35) by replacing W_t with i.i.d. standard normal random vectors. The initial values of the factors are set to be their long run means, Θ, and then we impose a warm-up period of one year before the actual simulated sample starts. In the simulation, we set the step size equal to half a day and we then generate weekly observations. In each Monte Carlo path, we simulate 52 (weeks) $\times$ 42 (years) = 2,184 weekly observations to match the frequency and horizon of our data set.

In the nonlinear setting, we assume a quadratic term structure model. Following Ahn et al. (2002), we assume that the stochastic differential equations of the factors are mean-reverting multivariate Gaussian processes,

$$dF_t = (\mu + \xi F_t)dt + \Sigma dW_t, \tag{38}$$

where μ is a 3×1 vector, and ξ and Σ are 3×3 matrices. The short rate is a quadratic function of the factors F_t. As a result, bond prices are exponential quadratic functions of the factors,

$$p_{\tau,t} = \exp[A(\tau) + B(\tau)^\top F_t + F_t^\top C(\tau)F_t], \tag{39}$$

[19] As in Dai and Singleton (2000) and Ahn et al. (2002), we employ Chen's (1996) "benchmark" model.

where $A(\tau)$ is a scalar, $B(\tau)$ is a 3×1 vector, and $C(\tau)$ is a 3×3 matrix, and they are the solutions for the Riccati equations with boundary conditions $A(0) = 0$, $B(0) = 0_3$, and $C(0) = 0_{3\times3}$, respectively. The Riccati equations can be solved by numerical methods as well. The yield is a quadratic function of the factors,

$$Y_{\tau,t} = -\frac{1}{\tau}[A(\tau) + B(\tau)^\top F_t + F_t^\top C(\tau)F_t]. \tag{40}$$

We assume orthogonal factors and no interactions, i.e., we assume Ahn et al.'s (2002) QTSM 3 model. To construct a nonlinear economy, we simulate the processes of the factors by replacing W_t with i.i.d. standard normal random vectors, and we set parameter values equal to the estimates in Ahn et al. (2002). The starting values of the factors are their unconditional means, $-\xi^{-1}\mu$, with a one-year warm-up period. In the simulation, we set the step size equal to one week.[20] With the solutions to the Riccati equations, we can convert the simulated factors into yield curves.

In each of the simulated economies, we can form theoretical yield series of different maturities. We then introduce heteroskedastic and autocorrelated noise around the theoretical yields generated in the simulation:

$$\hat{Y}_{\tau,t} = Y_{\tau,t} + e_{\tau,t}, \tag{41}$$

$$e_{\tau,t} = \rho_\tau e_{\tau,t-1} + \epsilon_{\tau,t}, \tag{42}$$

$$\epsilon_{\tau,t} = \phi_\tau Y_{\tau,t-2} z_{\tau,t}, \tag{43}$$

where $z_{\tau,t}$ is standard normal. We choose the serial-correlation parameters ρ_τ and the volatility parameters ϕ_τ so that the average adjusted R^2 and Durbin-Watson statistics in the simulated regression-based tests roughly match the adjusted R^2 and Durbin-Watson statistics in the tests implemented in the actual data, for the case of the simple polynomial.

[20] We do not have to consider a shorter step size due to the assumption that the factors follow Gaussian processes, and the volatilities of the factors are constant.

References

[1] Ahn, D.-H., R.F. Dittmar, and A.R. Gallant, 2002, Quadratic term structure models: theory and evidence, *Review of Financial Studies* 15, 243-288.

[2] Aït-Sahalia, Y., 1996a, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.

[3] Aït-Sahalia, Y., 1996b, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

[4] 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.

[5] Bliss, R.R., 1997, Testing term structure estimation models, *Advances in futures and options research* 9, 197-231.

[6] Chapman, D.A., and N.D. Pearson, 2000, Is the short rate drift actually nonlinear?, *Journal of Finance* 55, 355-388.

[7] Chapman, D.A., and N.D. Pearson, 2001, Recent advances in estimating term-structure models, *Financial Analysts Journal* 57, 77-95.

[8] Cox, J., J. Ingersoll, and S.A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-406.

[9] Dai, Q., and K. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

[10] Dai, Q., and K. Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16, 631-678.

[11] Duarte, J., 2004, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405-443.

[12] Duffee, G.R., 2002, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 379-404.

[13] Duffie, D., and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379-406.

[14] Fama, E.F., and R.R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680-692.

[15] Ghysels, E., and S. Ng, 1998, A semi-parametric factor model of interest rates and tests of the affine term structure, *Review of Economics and Statistics* 80, 489-502.

[16] Knez, P.J., R. Litterman, and J. Scheinkman, 1994, Explorations into factors explaining money market returns, *Journal of Finance* 49, 1861-1882.

[17] McCulloch, J.H., and H.-C. Kwon, 1993, U.S. term structure data, 1947–1991, working paper, Ohio State University.

[18] Newey, W.K., and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

[19] Pearson, N.D., and T.-S. Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279-1304.

[20] Politis, D.N., and J.P. Romano, 1994, The stationary bootstrap, *Journal of American Statistical Association* 89, 1303-1313.

[21] Pritsker, M., 1998, Nonparametric density estimation and tests of continuous time interest rate models, *Review of Financial Studies* 11, 449-487.

[22] Stanton, R., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

[23] Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-188.

[24] Waggoner, D., 1997, Spline methods for extracting interest rate from curves from coupon bond prices, working paper, Federal Reserve Bank of Atlanta.

[25] White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.

Table 1: Summary Statistics

The table reports the summary statistics of weekly observations for three selected unsmoothed Fama-Bliss zero-coupon Treasury yields and for three empirical yield curve factors, for the sample period June 1961-December 2002, and for the four subsamples June 1961-August 1971, September 1971-September 1979, October 1979-September 1982, and October 1982-December 2002. "Mean" is the sample mean. "Std." is the standard deviation. "AR1" is the first-order autocorrelation coefficient. The kurtosis measure is standardized such that it is zero for a normal random variable. $Y_{\tau,t}$ is the τ-year yield. "Level" is defined as $Y_{0.25,t}$. "Slope" is defined as $Y_{8,t} - Y_{0.25,t}$. "Curvature" is defined as $(Y_{8,t} - Y_{2,t}) - (Y_{2,t} - Y_{0.25,t})$. All yields and factors are reported in percentage points.

	Mean	Std	Skewness	Kurtosis	AR1
Whole Sample: June 1961 - December 2002					
Y_1	6.429	2.683	1.044	1.310	0.995
Y_5	7.016	2.439	0.923	0.588	0.996
Y_{10}	7.221	2.354	0.890	0.539	0.996
Level	6.056	2.744	1.297	2.292	0.994
Slope	1.155	1.351	-0.563	0.840	0.986
Curvature	-0.042	0.639	-0.081	2.467	0.934
Period 1: June 1961 - August 1971					
Y_1	4.768	1.412	0.504	-0.724	0.993
Y_5	4.998	1.250	0.735	-0.567	0.995
Y_{10}	5.102	1.161	0.710	-0.718	0.991
Level	4.503	1.412	0.522	-0.536	0.994
Slope	0.582	0.611	0.266	-0.060	0.978
Curvature	-0.044	0.376	-0.195	0.126	0.914
Period 2: September 1971 - September 1979					
Y_1	6.851	1.568	0.443	-0.622	0.984
Y_5	7.221	0.916	0.114	-0.791	0.985
Y_{10}	7.364	0.812	0.180	-0.717	0.980
Level	6.391	1.785	0.438	-0.768	0.982
Slope	0.933	1.340	-0.490	-1.101	0.979
Curvature	-0.249	0.534	0.062	-0.130	0.879
Period 3: October 1979 - September 1982					
Y_1	12.754	1.988	-0.636	-0.068	0.951
Y_5	12.265	1.572	-0.495	-0.871	0.956
Y_{10}	12.226	1.469	-0.158	-1.019	0.953
Level	12.688	2.575	-0.409	-0.515	0.943
Slope	-0.523	2.097	0.110	-0.855	0.947
Curvature	-0.263	1.208	0.030	0.459	0.876
Period 4: October 1982 - December 2002					
Y_1	6.159	2.171	0.180	-0.178	0.993
Y_5	7.174	2.128	0.657	-0.027	0.994
Y_{10}	7.489	2.010	0.687	-0.082	0.994
Level	5.719	2.069	0.108	-0.354	0.994
Slope	1.782	1.093	-0.138	-1.010	0.989
Curvature	0.074	0.632	0.113	0.436	0.961

Table 2: Testing the Affine Class; Baseline Case

The table reports the p-values of the regression-based tests using the asymptotic distribution and bootstrap distribution of the test statistics. The null hypothesis is the three-factor affine model. The empirical factors are the "level," "slope," and "curvature" of the yield curve. We consider four SNP specifications: simple polynomial (Panel A), Legendre polynomial (Panel B), Fourier transform (Panel C), and Hermite polynomial (Panel D). The order of the SNP specification is determined by the Akaike Information Criterion. The regression models are estimated in first differences; see (20). The null hypothesis is tested separately for the three individual yields Y_1, Y_5, and Y_{10}, and then jointly using all three yields ("All"). Under "White" we report the p-values for the Wald test statistics with heteroskedasticity adjustment due to White (1980); under "NW" we report the p-values for the Wald test statistics with heteroskedasticity and autocorrelation adjustment due to Newey and West (1987). For each bootstrap-corrected test, 5,000 bootstrap replications are produced using a block bootstrap. For goodness-of-fit measures, we report the adjusted coefficients of determination under "Adj-R^2" and the Durbin-Watson statistics under "DW." The yield dataset is the extended "unsmoothed" Fama-Bliss dataset.

Panel A: Simple Polynomial

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.000	0.000	0.075	0.080	0.765	2.357
Y_5	0.030	0.032	0.059	0.058	0.876	2.409
Y_{10}	0.003	0.008	0.129	0.210	0.669	2.736
All	0.000	0.000	0.159	0.146		
Period 1: June 1961 - August 1971						
Y_1	0.034	0.040	0.316	0.408	0.515	2.268
Y_5	0.785	0.789	0.812	0.819	0.822	2.153
Y_{10}	0.961	0.988	0.781	0.890	0.417	2.989
All	0.003	0.000	0.768	0.715		
Period 2: September 1971 - September 1979						
Y_1	0.001	0.001	0.145	0.233	0.685	2.195
Y_5	0.970	0.972	0.979	0.980	0.736	2.057
Y_{10}	0.007	0.006	0.188	0.242	0.569	2.148
All	0.000	0.000	0.655	0.670		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.120	0.086	0.854	2.374
Y_5	0.188	0.180	0.283	0.283	0.877	2.673
Y_{10}	0.000	0.000	0.089	0.168	0.691	2.738
All	0.000	0.000	0.090	0.130		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.035	0.050	0.758	2.467
Y_5	0.418	0.456	0.517	0.542	0.929	2.452
Y_{10}	0.006	0.008	0.162	0.196	0.797	2.510
All	0.000	0.000	0.154	0.157		

Panel B: Legendre Polynomial

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.000	0.000	0.033	0.025	0.766	2.357
Y_5	0.043	0.053	0.354	0.391	0.877	2.393
Y_{10}	0.000	0.000	0.017	0.025	0.673	2.733
All	0.000	0.000	0.006	0.006		
Period 1: June 1961 - August 1971						
Y_1	0.086	0.054	0.470	0.455	0.523	2.262
Y_5	0.018	0.005	0.341	0.277	0.826	2.137
Y_{10}	0.976	0.993	0.779	0.891	0.417	2.984
All	0.018	0.000	0.742	0.494		
Period 2: September 1971 - September 1979						
Y_1	0.000	0.000	0.135	0.153	0.692	2.131
Y_5	0.449	0.270	0.911	0.865	0.736	2.028
Y_{10}	0.002	0.003	0.107	0.153	0.583	2.115
All	0.000	0.000	0.613	0.586		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.128	0.186	0.865	2.351
Y_5	0.000	0.000	0.219	0.348	0.880	2.595
Y_{10}	0.024	0.010	0.632	0.613	0.666	2.773
All	0.000	0.000	0.209	0.267		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.005	0.007	0.758	2.456
Y_5	0.000	0.000	0.016	0.017	0.932	2.395
Y_{10}	0.000	0.000	0.022	0.021	0.801	2.506
All	0.000	0.000	0.007	0.008		

Panel C: Fourier Transform

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.000	0.000	0.017	0.017	0.768	2.352
Y_5	0.030	0.033	0.271	0.280	0.878	2.389
Y_{10}	0.013	0.040	0.158	0.250	0.671	2.730
All	0.000	0.000	0.005	0.003		
Period 1: June 1961 - August 1971						
Y_1	0.154	0.121	0.527	0.508	0.519	2.277
Y_5	0.030	0.010	0.301	0.223	0.826	2.128
Y_{10}	0.951	0.983	0.645	0.769	0.417	2.984
All	0.006	0.000	0.523	0.266		
Period 2: September 1971 - September 1979						
Y_1	0.001	0.000	0.105	0.096	0.691	2.153
Y_5	0.676	0.606	0.957	0.948	0.737	2.041
Y_{10}	0.007	0.010	0.127	0.172	0.577	2.137
All	0.000	0.000	0.579	0.642		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.002	0.268	0.441	0.866	2.323
Y_5	0.000	0.000	0.155	0.200	0.879	2.639
Y_{10}	0.009	0.003	0.434	0.396	0.668	2.769
All	0.000	0.000	0.375	0.512		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.001	0.002	0.761	2.455
Y_5	0.010	0.017	0.112	0.135	0.932	2.410
Y_{10}	0.000	0.000	0.017	0.016	0.801	2.506
All	0.000	0.000	0.000	0.000		

Panel D: Hermite Polynomial

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.000	0.000	0.037	0.029	0.766	2.357
Y_5	0.043	0.053	0.361	0.397	0.877	2.393
Y_{10}	0.000	0.000	0.012	0.021	0.673	2.733
All	0.000	0.000	0.007	0.007		
Period 1: June 1961 - August 1971						
Y_1	0.086	0.054	0.465	0.445	0.523	2.262
Y_5	0.018	0.005	0.343	0.277	0.826	2.137
Y_{10}	0.976	0.993	0.786	0.887	0.417	2.984
All	0.018	0.000	0.734	0.491		
Period 2: September 1971 - September 1979						
Y_1	0.000	0.000	0.134	0.157	0.692	2.131
Y_5	0.449	0.270	0.921	0.877	0.736	2.028
Y_{10}	0.002	0.003	0.114	0.149	0.583	2.115
All	0.000	0.000	0.622	0.602		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.121	0.177	0.865	2.351
Y_5	0.000	0.000	0.226	0.347	0.880	2.595
Y_{10}	0.024	0.010	0.636	0.615	0.666	2.773
All	0.000	0.000	0.211	0.278		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.003	0.006	0.758	2.456
Y_5	0.000	0.000	0.016	0.017	0.932	2.395
Y_{10}	0.000	0.000	0.018	0.018	0.801	2.506
All	0.000	0.000	0.005	0.005		

Table 3: Testing the Affine Class; Intermediate SNP Orders

The table reports the p-values of the regression-based tests using the asymptotic distribution and bootstrap distribution of the test statistics. The null hypothesis is the three-factor affine model. The empirical factors are the "level," "slope," and "curvature" of the yield curve. We report results for the simple-polynomial specification with two nonlinear terms. The regression models are estimated in first differences; see (20). The null hypothesis is tested separately for the three individual yields Y_1, Y_5, and Y_{10}, and then jointly using all three yields ("All"). Under "White" we report the p-values for the Wald test statistics with heteroskedasticity adjustment due to White (1980); under "NW" we report the p-values for the Wald test statistics with heteroskedasticity and autocorrelation adjustment due to Newey and West (1987). For each bootstrap-corrected test, 5,000 bootstrap replications are produced using a block bootstrap. For goodness-of-fit measures, we report the adjusted coefficients of determination under "Adj-R^2" and the Durbin-Watson statistics under "DW." The yield dataset is the extended "unsmoothed" Fama-Bliss dataset.

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.597	0.667	0.811	0.870	0.757	2.356
Y_5	0.013	0.011	0.178	0.206	0.876	2.409
Y_{10}	0.035	0.019	0.167	0.159	0.665	2.760
All	0.000	0.000	0.236	0.252		
Period 1: June 1961 - August 1971						
Y_1	0.220	0.306	0.500	0.637	0.510	2.287
Y_5	0.654	0.524	0.822	0.788	0.821	2.155
Y_{10}	0.472	0.586	0.368	0.534	0.427	2.994
All	0.113	0.199	0.733	0.833		
Period 2: September 1971 - September 1979						
Y_1	0.008	0.017	0.181	0.285	0.671	2.172
Y_5	0.010	0.000	0.248	0.120	0.735	2.057
Y_{10}	0.259	0.252	0.492	0.566	0.533	2.175
All	0.000	0.000	0.543	0.504		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.029	0.019	0.851	2.349
Y_5	0.006	0.000	0.267	0.224	0.877	2.630
Y_{10}	0.000	0.000	0.098	0.117	0.674	2.838
All	N/A					
Period 4: October 1982 - December 2002						
Y_1	0.002	0.001	0.144	0.175	0.757	2.470
Y_5	0.030	0.025	0.276	0.309	0.929	2.449
Y_{10}	0.006	0.021	0.094	0.182	0.790	2.543
All	0.000	0.000	0.400	0.450		

Table 4: Testing the Affine Class; Principal Components

The table reports the p-values of the regression-based tests using the asymptotic distribution and bootstrap distribution of the test statistics. The null hypothesis is the three-factor affine model. The empirical factors are the first three principal components, extracted from 13 yield changes. We report results for the simple-polynomial specification, where the number of terms is chosen based on the Akaike Information Criterion. The regression models are estimated in first differences; see (20). The null hypothesis is tested separately for the three individual yields Y_1, Y_5, and Y_{10}, and then jointly using all three yields ("All"). Under "White" we report the p-values for the Wald test statistics with heteroskedasticity adjustment due to White (1980); under "NW" we report the p-values for the Wald test statistics with heteroskedasticity and autocorrelation adjustment due to Newey and West (1987). For each bootstrap-corrected test, 5,000 bootstrap replications are produced using a block bootstrap. For goodness-of-fit measures, we report the adjusted coefficients of determination under "Adj-R^2" and the Durbin-Watson statistics under "DW." The yield dataset is the extended "unsmoothed" Fama-Bliss dataset.

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.001	0.001	0.096	0.136	0.914	2.470
Y_5	0.064	0.051	0.334	0.350	0.932	2.363
Y_{10}	0.001	0.003	0.107	0.158	0.854	2.678
All	0.000	0.000	0.571	0.653		
Period 1: June 1961 - August 1971						
Y_1	0.000	0.000	0.017	0.025	0.806	2.135
Y_5	0.000	0.000	0.091	0.074	0.916	2.194
Y_{10}	0.000	0.000	0.020	0.012	0.992	2.183
All	0.000	0.000	0.018	0.016		
Period 2: September 1971 - September 1979						
Y_1	0.001	0.000	0.112	0.145	0.873	2.404
Y_5	0.000	0.000	0.021	0.024	0.851	2.033
Y_{10}	0.000	0.000	0.074	0.045	0.799	2.053
All	0.000	0.000	0.101	0.096		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.028	0.020	0.956	2.548
Y_5	0.000	0.000	0.121	0.091	0.932	2.460
Y_{10}	0.000	0.000	0.223	0.215	0.870	2.742
All	0.000	0.000	0.243	0.211		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.074	0.112	0.873	2.539
Y_5	0.799	0.837	0.939	0.959	0.958	2.402
Y_{10}	0.197	0.204	0.639	0.646	0.891	2.713
All	0.000	0.000	0.496	0.511		

Table 5: Testing the Affine Class; Sieve Bootstrap

The table reports the p-values of the regression-based tests using the asymptotic distribution and bootstrap distribution of the test statistics. The null hypothesis is the three-factor affine model. The empirical factors are the "level," "slope," and "curvature" of the yield curve. We report results for the simple-polynomial specification, where the number of terms is chosen based on the Akaike Information Criterion. The regression models are estimated in first differences; see (20). The null hypothesis is tested separately for the three individual yields Y_1, Y_5, and Y_{10}, and then jointly using all three yields ("All"). Under "White" we report the p-values for the Wald test statistics with heteroskedasticity adjustment due to White (1980); under "NW" we report the p-values for the Wald test statistics with heteroskedasticity and autocorrelation adjustment due to Newey and West (1987). For each bootstrap-corrected test, 5,000 bootstrap replications are produced using a sieve bootstrap. For goodness-of-fit measures, we report the adjusted coefficients of determination under "Adj-R^2" and the Durbin-Watson statistics under "DW." The yield dataset is the extended "unsmoothed" Fama-Bliss dataset.

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: June 1961 - December 2002						
Y_1	0.000	0.000	0.021	0.025	0.765	2.357
Y_5	0.030	0.032	0.007	0.009	0.876	2.409
Y_{10}	0.003	0.008	0.611	0.821	0.669	2.736
All	0.000	0.000	0.119	0.113		
Period 1: June 1961 - August 1971						
Y_1	0.034	0.040	0.048	0.088	0.515	2.268
Y_5	0.785	0.789	0.534	0.557	0.822	2.153
Y_{10}	0.961	0.988	0.970	0.995	0.417	2.989
All	0.003	0.000	0.588	0.522		
Period 2: September 1971 - September 1979						
Y_1	0.001	0.001	0.005	0.009	0.685	2.195
Y_5	0.970	0.972	0.564	0.644	0.736	2.057
Y_{10}	0.007	0.006	0.017	0.023	0.569	2.148
All	0.000	0.000	0.036	0.041		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.121	0.084	0.854	2.374
Y_5	0.188	0.180	0.201	0.212	0.877	2.673
Y_{10}	0.000	0.000	0.077	0.171	0.691	2.738
All	0.000	0.000	0.093	0.131		
Period 4: October 1982 - December 2002						
Y_1	0.000	0.000	0.033	0.046	0.758	2.467
Y_5	0.418	0.456	0.812	0.827	0.929	2.452
Y_{10}	0.006	0.008	0.616	0.662	0.797	2.510
All	0.000	0.000	0.348	0.405		

Table 6: Testing the Affine Class; McCulloch-Kwon Dataset

The table reports the p-values of the regression-based tests using the asymptotic distribution and bootstrap distribution of the test statistics. The null hypothesis is the three-factor affine model. The empirical factors are the "level," "slope," and "curvature" of the yield curve. We report results for the simple-polynomial specification, where the number of terms is chosen based on the Akaike Information Criterion. The regression models are estimated in first differences; see (20). The null hypothesis is tested separately for the three individual yields Y_1, Y_5, and Y_{10}, and then jointly using all three yields ("All"). Under "White" we report the p-values for the Wald test statistics with heteroskedasticity adjustment due to White (1980); under "NW" we report the p-values for the Wald test statistics with heteroskedasticity and autocorrelation adjustment due to Newey and West (1987). For each bootstrap-corrected test, 5,000 bootstrap replications are produced using a block bootstrap. For goodness-of-fit measures, we report the adjusted coefficients of determination under "Adj-R^2" and the Durbin-Watson statistics under "DW." The yield dataset is the McCulloch-Kwon dataset.

	Asymptotic Tests		Bootstrap-corrected Tests		Goodness of Fit	
	White	NW	White	NW	Adj-R^2	DW
Whole Sample: December 1946 - February 1991						
Y_1	0.159	0.046	0.644	0.543	0.957	2.623
Y_5	0.000	0.000	0.000	0.000	0.975	2.639
Y_{10}	0.002	0.000	0.367	0.304	0.975	2.574
All	0.000	0.000	0.008	0.004		
Period 1: June 1961 - August 1971						
Y_1	0.000	0.000	0.024	0.018	0.920	2.572
Y_5	0.304	0.229	0.383	0.312	0.953	2.661
Y_{10}	0.000	0.000	0.199	0.173	0.973	2.488
All	0.011	0.011	0.129	0.139		
Period 2: September 1971 - September 1979						
Y_1	0.000	0.000	0.400	0.423	0.926	2.686
Y_5	0.036	0.028	0.125	0.149	0.951	2.515
Y_{10}	0.000	0.000	0.199	0.238	0.963	2.405
All	0.066	0.076	0.543	0.624		
Period 3: October 1979 - September 1982						
Y_1	0.000	0.000	0.777	0.763	0.983	2.034
Y_5	0.349	0.286	0.587	0.567	0.987	2.538
Y_{10}	0.000	0.000	0.046	0.055	0.974	2.233
All	0.013	0.013	0.421	0.546		
Period 4: October 1982 - February 1991						
Y_1	0.000	0.000	0.069	0.070	0.979	2.293
Y_5	0.002	0.002	0.023	0.032	0.984	2.528
Y_{10}	0.000	0.000	0.103	0.091	0.983	2.309
All	0.000	0.000	0.004	0.005		

Table 7: Hedging; Baseline Case

This table reports the in sample and out-of-sample (OOS) root mean square hedging errors (in percentage points) for the whole sample and for four subsamples. We use zero coupon bonds with maturities of 1, 5, and 10 years to hedge a 3- (Panel A) or 7- (Panel B) year zero coupon bond using the following six strategies: constant-weight, affine, simple polynomial, Legendre polynomial, Fourier transform, and Hermite polynomial portfolios. In the time-varying weight strategies, the factors are empirical factors constructed by $Y_{0.25}$, Y_2, and Y_8, and the orders of the polynomials are determined by the Akaike Information Criterion. As benchmarks, we also report the root mean square errors for the unhedged portfolio (Unhedged) and for the portfolio hedged by means of a barbell portfolio of the 1- and 10-year bonds (Barbell). The in-sample evaluation uses all of the observations within the sample. Four out-of-sample schemes are considered: fixed window, recursive scheme, rolling window, and equally-weighted moving average (EWMA), and the first 50% of the observations within the sample are used as the initial estimation window.

Panel A: Hedge a 3 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	0.555	0.311	0.202	0.197	0.184	0.184	0.184	0.184
OOS(Fixed)	0.505	0.253	0.159	0.170	0.473	3.634	0.507	3.634
OOS(Recursive)	0.505	0.253	0.159	0.164	0.290	1.063	0.248	1.063
OOS(Rolling)	0.505	0.253	0.162	0.165	0.348	1.081	0.256	1.081
OOS(EWMA)	0.505	0.253	0.157	0.206	3.093	2.397	1.100	2.397
Period 1: June 1961 - August 1971								
In-Sample	0.352	0.285	0.143	0.139	0.121	0.123	0.124	0.123
OOS(Fixed)	0.478	0.236	0.209	0.258	149.046	832.822	34.382	832.822
OOS(Recursive)	0.478	0.236	0.187	0.198	0.848	2.046	0.448	2.046
OOS(Rolling)	0.478	0.236	0.185	0.200	3.654	6.707	2.260	6.707
OOS(EWMA)	0.478	0.236	0.188	0.246	11.628	31.901	4.716	31.901
Period 2: September 1971 - September 1979								
In-Sample	0.427	0.322	0.176	0.173	0.160	0.160	0.158	0.160
OOS(Fixed)	0.392	0.302	0.195	0.197	0.353	0.444	0.281	0.444
OOS(Recursive)	0.392	0.302	0.195	0.199	0.289	0.316	0.240	0.316
OOS(Rolling)	0.392	0.302	0.196	0.217	0.487	0.913	0.462	0.913
OOS(EWMA)	0.392	0.302	0.221	0.257	0.595	0.763	0.468	0.763
Period 3: October 1979 - September 1982								
In-Sample	1.335	0.662	0.502	0.467	0.390	0.374	0.364	0.374
OOS(Fixed)	1.205	0.720	0.619	1.023	386.016	2375.623	48.856	2375.623
OOS(Recursive)	1.205	0.720	0.607	0.673	5.374	8.531	3.685	8.531
OOS(Rolling)	1.205	0.720	0.612	0.671	15.282	40.550	9.293	40.550
OOS(EWMA)	1.205	0.720	0.618	0.702	8.198	19.088	6.645	19.088
Period 4: October 1982 - December 2002								
In-Sample	0.484	0.224	0.149	0.138	0.129	0.129	0.129	0.129
OOS(Fixed)	0.406	0.183	0.104	0.100	0.372	0.586	0.308	0.586
OOS(Recursive)	0.406	0.183	0.102	0.097	0.117	0.147	0.126	0.147
OOS(Rolling)	0.406	0.183	0.098	0.099	0.147	0.234	0.179	0.234
OOS(EWMA)	0.406	0.183	0.094	0.119	1.227	2.412	1.418	2.412

Panel B: Hedge a 7 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	1.158	0.753	0.508	0.503	0.460	0.458	0.457	0.458
OOS(Fixed)	1.171	0.603	0.425	0.529	1.604	1.065	0.974	1.065
OOS(Recursive)	1.171	0.603	0.411	0.438	1.065	0.851	0.945	0.851
OOS(Rolling)	1.171	0.603	0.407	0.435	1.153	0.910	0.978	0.910
OOS(EWMA)	1.171	0.603	0.408	0.491	25.831	3.662	2.896	3.662
Period 1: June 1961 - August 1971								
In-Sample	0.637	0.678	0.280	0.150	0.091	0.100	0.099	0.100
OOS(Fixed)	0.851	0.120	0.386	1.058	185.576	183.123	44.947	183.123
OOS(Recursive)	0.851	0.120	0.364	0.221	0.574	0.364	0.374	0.364
OOS(Rolling)	0.851	0.120	0.192	0.169	1.105	2.041	1.502	2.041
OOS(EWMA)	0.851	0.120	0.115	0.118	5.099	0.668	0.216	0.668
Period 2: September 1971 - September 1979								
In-Sample	0.807	0.788	0.556	0.545	0.493	0.488	0.492	0.488
OOS(Fixed)	0.745	0.714	0.432	0.641	1.135	0.907	0.973	0.907
OOS(Recursive)	0.745	0.714	0.407	0.443	0.783	0.687	0.505	0.687
OOS(Rolling)	0.745	0.714	0.416	0.464	1.155	1.064	0.921	1.064
OOS(EWMA)	0.745	0.714	0.400	0.412	1.520	0.807	0.912	0.807
Period 3: October 1979 - September 1982								
In-Sample	2.632	1.664	1.266	1.198	0.920	0.910	0.886	0.910
OOS(Fixed)	2.424	1.922	1.459	2.153	851.458	247.853	78.830	247.853
OOS(Recursive)	2.424	1.922	1.412	1.521	10.809	5.308	3.939	5.308
OOS(Rolling)	2.424	1.922	1.417	1.526	34.799	21.336	15.942	21.336
OOS(EWMA)	2.424	1.922	1.415	1.701	15.250	6.530	5.345	6.530
Period 4: October 1982 - December 2002								
In-Sample	1.122	0.512	0.333	0.326	0.310	0.310	0.310	0.310
OOS(Fixed)	0.947	0.332	0.218	0.238	0.914	0.889	0.954	0.889
OOS(Recursive)	0.947	0.332	0.213	0.222	0.282	0.276	0.267	0.276
OOS(Rolling)	0.947	0.332	0.212	0.220	0.468	0.453	0.592	0.453
OOS(EWMA)	0.947	0.332	0.204	0.266	4.349	2.505	12.584	2.505

Table 8: Hedging; Intermediate SNP Orders

This table reports the in sample and out-of-sample (OOS) root mean square hedging errors (in percentage points) for the whole sample and for four subsamples. We use zero coupon bonds with maturities of 1, 5, and 10 years to hedge a 3- (Panel A) or 7- (Panel B) year zero-coupon bond using the following six strategies: constant-weight, affine, simple polynomial, Legendre polynomial, Fourier transform, and Hermite polynomial portfolios. In the time-varying weight strategies, the factors are empirical factors constructed by $Y_{0.25}$, Y_2, and Y_8. For each factor, three nonlinear terms are included in the Fourier transform, and two for the other SNP specifications. As benchmarks, we also report the root mean square errors for the unhedged portfolio (Unhedged) and for the portfolio hedged by means of a barbell portfolio of the 1- and 10-year bonds (Barbell). The in-sample evaluation uses all of the observations within the sample. Four out-of-sample schemes are considered: fixed window, recursive scheme, rolling window, and equally-weighted moving average (EWMA), and the first 50% of the observations within the sample are used as the initial estimation window.

Panel A: Hedge a 3 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	0.555	0.311	0.202	0.197	0.193	0.190	0.188	0.190
OOS(Fixed)	0.505	0.253	0.159	0.170	0.205	0.305	0.295	0.305
OOS(Recursive)	0.505	0.253	0.159	0.164	0.174	0.218	0.252	0.218
OOS(Rolling)	0.505	0.253	0.162	0.165	0.259	0.222	0.252	0.222
OOS(EWMA)	0.505	0.253	0.157	0.206	0.557	0.346	0.656	0.346
Period 1: June 1961 - August 1971								
In-Sample	0.352	0.285	0.143	0.139	0.129	0.130	0.126	0.130
OOS(Fixed)	0.478	0.236	0.209	0.258	0.581	3.862	13.611	3.862
OOS(Recursive)	0.478	0.236	0.187	0.198	0.242	0.254	0.280	0.254
OOS(Rolling)	0.478	0.236	0.185	0.200	0.266	0.311	0.448	0.311
OOS(EWMA)	0.478	0.236	0.188	0.246	0.497	0.936	2.921	0.936
Period 2: September 1971 - September 1979								
In-Sample	0.427	0.322	0.176	0.173	0.168	0.168	0.165	0.168
OOS(Fixed)	0.392	0.302	0.195	0.197	0.227	0.203	0.208	0.203
OOS(Recursive)	0.392	0.302	0.195	0.199	0.209	0.210	0.215	0.210
OOS(Rolling)	0.392	0.302	0.196	0.217	0.240	0.246	0.278	0.246
OOS(EWMA)	0.392	0.302	0.221	0.257	0.318	0.314	0.408	0.314
Period 3: October 1979 - September 1982								
In-Sample	1.335	0.662	0.502	0.467	0.435	0.418	0.405	0.418
OOS(Fixed)	1.205	0.720	0.619	1.023	3.692	7.240	38.150	7.240
OOS(Recursive)	1.205	0.720	0.607	0.673	1.505	1.041	1.461	1.041
OOS(Rolling)	1.205	0.720	0.612	0.671	1.751	1.392	2.499	1.392
OOS(EWMA)	1.205	0.720	0.618	0.702	1.996	1.626	2.108	1.626
Period 4: October 1982 - December 2002								
In-Sample	0.484	0.224	0.149	0.138	0.135	0.133	0.131	0.133
OOS(Fixed)	0.406	0.183	0.104	0.100	0.116	0.164	0.194	0.164
OOS(Recursive)	0.406	0.183	0.102	0.097	0.099	0.104	0.110	0.104
OOS(Rolling)	0.406	0.183	0.098	0.099	0.100	0.103	0.106	0.103
OOS(EWMA)	0.406	0.183	0.094	0.119	0.321	0.368	0.955	0.368

Panel B: Hedge a 7 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	1.158	0.753	0.508	0.503	0.486	0.483	0.473	0.483
OOS(Fixed)	1.171	0.603	0.425	0.529	0.519	0.675	0.897	0.675
OOS(Recursive)	1.171	0.603	0.411	0.438	0.457	0.564	0.517	0.564
OOS(Rolling)	1.171	0.603	0.407	0.435	0.513	0.565	0.520	0.565
OOS(EWMA)	1.171	0.603	0.408	0.491	9.054	1.142	1.832	1.142
Period 1: June 1961 - August 1971								
In-Sample	0.637	0.678	0.280	0.150	0.102	0.122	0.114	0.122
OOS(Fixed)	0.851	0.120	0.386	1.058	1.403	3.516	11.624	3.516
OOS(Recursive)	0.851	0.120	0.364	0.221	0.233	0.226	0.529	0.226
OOS(Rolling)	0.851	0.120	0.192	0.169	0.188	0.281	0.907	0.281
OOS(EWMA)	0.851	0.120	0.115	0.118	0.195	0.241	0.676	0.241
Period 2: September 1971 - September 1979								
In-Sample	0.807	0.788	0.556	0.545	0.521	0.523	0.512	0.523
OOS(Fixed)	0.745	0.714	0.432	0.641	0.684	0.678	0.732	0.678
OOS(Recursive)	0.745	0.714	0.407	0.443	0.462	0.462	0.496	0.462
OOS(Rolling)	0.745	0.714	0.416	0.464	0.576	0.525	0.642	0.525
OOS(EWMA)	0.745	0.714	0.400	0.412	0.704	0.572	0.742	0.572
Period 3: October 1979 - September 1982								
In-Sample	2.632	1.664	1.266	1.198	1.081	1.080	1.022	1.080
OOS(Fixed)	2.424	1.922	1.459	2.153	16.067	11.520	59.628	11.520
OOS(Recursive)	2.424	1.922	1.412	1.521	5.176	2.235	2.866	2.235
OOS(Rolling)	2.424	1.922	1.417	1.526	5.476	2.756	4.788	2.756
OOS(EWMA)	2.424	1.922	1.415	1.701	10.625	2.763	3.203	2.763
Period 4: October 1982 - December 2002								
In-Sample	1.122	0.512	0.333	0.326	0.320	0.318	0.314	0.318
OOS(Fixed)	0.947	0.332	0.218	0.238	0.260	0.250	0.293	0.250
OOS(Recursive)	0.947	0.332	0.213	0.222	0.232	0.232	0.235	0.232
OOS(Rolling)	0.947	0.332	0.212	0.220	0.237	0.253	0.307	0.253
OOS(EWMA)	0.947	0.332	0.204	0.266	0.750	0.948	2.051	0.948

Table 9: Hedging; Principal Components

This table reports the in sample and out-of-sample (OOS) root mean square hedging errors (in percentage points) for the whole sample and for four subsamples. We use zero coupon bonds with maturities of 1, 5, and 10 years to hedge a 3- (Panel A) or 7- (Panel B) year zero-coupon bond using the following six strategies: constant-weight, affine, simple polynomial, Legendre polynomial, Fourier transform, and Hermite polynomial portfolios. In the time-varying weight strategies, the factors are the principal components extracted from 13 yield changes, and the orders of the polynomials are determined by the Akaike Information Criterion. As benchmarks, we also report the root mean square errors for the unhedged portfolio (Unhedged) and for the portfolio hedged by means of a barbell portfolio of the 1- and 10-year bonds (Barbell). The in-sample evaluation uses all of the observations within the sample. Four out-of-sample schemes are considered: fixed window, recursive scheme, rolling window, and equally-weighted moving average (EWMA), and the first 50% of the observations within the sample are used as the initial estimation window.

Panel A: Hedge a 3 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	0.555	0.311	0.202	0.198	0.186	0.184	0.185	0.184
OOS(Fixed)	0.505	0.253	0.159	0.166	0.308	0.790	0.383	0.790
OOS(Recursive)	0.505	0.253	0.159	0.163	0.210	0.281	0.196	0.281
OOS(Rolling)	0.505	0.253	0.162	0.164	0.203	0.291	0.213	0.291
OOS(EWMA)	0.505	0.253	0.157	0.204	23.882	1.461	1.538	1.461
Period 1: June 1961 - August 1971								
In-Sample	0.352	0.285	0.143	0.140	0.123	0.122	0.123	0.122
OOS(Fixed)	0.478	0.236	0.209	0.247	298.818	2420.265	407.175	2420.265
OOS(Recursive)	0.478	0.236	0.187	0.199	0.853	2.021	1.170	2.021
OOS(Rolling)	0.478	0.236	0.185	0.201	0.979	2.235	1.499	2.235
OOS(EWMA)	0.478	0.236	0.188	0.248	5.155	12.312	9.388	12.312
Period 2: September 1971 - September 1979								
In-Sample	0.427	0.322	0.176	0.173	0.161	0.160	0.159	0.160
OOS(Fixed)	0.392	0.302	0.195	0.197	1.064	1.365	0.858	1.365
OOS(Recursive)	0.392	0.302	0.195	0.199	0.352	0.373	0.292	0.373
OOS(Rolling)	0.392	0.302	0.196	0.218	0.468	0.705	0.501	0.705
OOS(EWMA)	0.392	0.302	0.221	0.252	0.859	1.150	0.881	1.150
Period 3: October 1979 - September 1982								
In-Sample	1.335	0.662	0.502	0.471	0.343	0.340	0.338	0.340
OOS(Fixed)	1.205	0.720	0.619	1.031	342.000	703.428	147.517	703.428
OOS(Recursive)	1.205	0.720	0.607	0.690	6.869	9.002	5.015	9.002
OOS(Rolling)	1.205	0.720	0.612	0.699	6.515	38.368	20.872	38.368
OOS(EWMA)	1.205	0.720	0.618	0.725	10.055	14.935	9.770	14.935
Period 4: October 1982 - December 2002								
In-Sample	0.484	0.224	0.149	0.140	0.132	0.131	0.131	0.131
OOS(Fixed)	0.406	0.183	0.104	0.101	0.900	3.209	1.422	3.209
OOS(Recursive)	0.406	0.183	0.102	0.097	0.127	0.195	0.143	0.195
OOS(Rolling)	0.406	0.183	0.098	0.100	0.249	0.279	0.251	0.279
OOS(EWMA)	0.406	0.183	0.094	0.115	3.266	5.629	4.074	5.629

Panel B: Hedge a 7 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	1.158	0.753	0.508	0.502	0.468	0.464	0.464	0.464
OOS(Fixed)	1.171	0.603	0.425	0.478	1.913	2.493	1.444	2.493
OOS(Recursive)	1.171	0.603	0.411	0.435	0.889	0.648	0.569	0.648
OOS(Rolling)	1.171	0.603	0.407	0.430	1.134	0.683	0.592	0.683
OOS(EWMA)	1.171	0.603	0.408	0.492	68.923	6.559	6.174	6.559
Period 1: June 1961 - August 1971								
In-Sample	0.637	0.678	0.280	0.147	0.092	0.089	0.089	0.089
OOS(Fixed)	0.851	0.120	0.386	1.236	1320.105	1519.875	225.554	1519.875
OOS(Recursive)	0.851	0.120	0.364	0.216	0.389	2.277	1.572	2.277
OOS(Rolling)	0.851	0.120	0.192	0.161	1.249	5.787	3.955	5.787
OOS(EWMA)	0.851	0.120	0.115	0.120	1.206	2.150	1.635	2.150
Period 2: September 1971 - September 1979								
In-Sample	0.807	0.788	0.556	0.545	0.495	0.480	0.476	0.480
OOS(Fixed)	0.745	0.714	0.432	0.705	4.261	7.151	3.077	7.151
OOS(Recursive)	0.745	0.714	0.407	0.461	1.472	1.766	0.883	1.766
OOS(Rolling)	0.745	0.714	0.416	0.481	1.714	2.766	1.603	2.766
OOS(EWMA)	0.745	0.714	0.400	0.445	2.543	1.585	1.083	1.585
Period 3: October 1979 - September 1982								
In-Sample	2.632	1.664	1.266	1.196	1.020	0.987	0.982	0.987
OOS(Fixed)	2.424	1.922	1.459	2.169	1090.641	2297.066	860.236	2297.066
OOS(Recursive)	2.424	1.922	1.412	1.511	18.741	64.613	25.924	64.613
OOS(Rolling)	2.424	1.922	1.417	1.502	41.749	148.532	86.220	148.532
OOS(EWMA)	2.424	1.922	1.415	1.739	34.171	63.097	24.375	63.097
Period 4: October 1982 - December 2002								
In-Sample	1.122	0.512	0.333	0.323	0.311	0.310	0.310	0.310
OOS(Fixed)	0.947	0.332	0.218	0.243	4.173	10.649	4.244	10.649
OOS(Recursive)	0.947	0.332	0.213	0.223	0.375	0.397	0.318	0.397
OOS(Rolling)	0.947	0.332	0.212	0.223	0.468	0.710	0.603	0.710
OOS(EWMA)	0.947	0.332	0.204	0.250	17.731	7.063	6.172	7.063

Table 10: Hedging; Longer Initial Estimation Window

This table reports the in sample and out-of-sample (OOS) root mean square hedging errors (in percentage points) for the whole sample and for four subsamples. We use zero-coupon bonds with maturities 1, 5, and 10 years to hedge a 3- (Panel A) or 7- (Panel B) year zero-coupon bond using the following six strategies: constant-weight, affine, simple polynomial, Legendre polynomial, Fourier transform, and Hermite polynomial portfolios. In the time-varying weight strategies, the factors are empirical factors constructed by $Y_{0.25}$, Y_2, and Y_8, and the orders of the polynomials are determined by the Akaike Information Criterion. As benchmarks, we also report the root mean square errors for the unhedged portfolio (Unhedged) and for the portfolio hedged by means of a barbell portfolio of the 1- and 10-year bonds (Barbell). The in-sample evaluation uses all of the observations within the sample. Four out-of-sample schemes are considered: fixed window, recursive scheme, rolling window, and equally-weighted moving average (EWMA), and the first 70% of the observations within the sample are used as the initial estimation window.

Panel A: Hedge a 3 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	0.555	0.311	0.202	0.197	0.184	0.184	0.184	0.184
OOS(Fixed)	0.414	0.195	0.107	0.117	0.183	0.178	0.141	0.178
OOS(Recursive)	0.414	0.195	0.106	0.111	0.118	0.110	0.109	0.110
OOS(Rolling)	0.414	0.195	0.107	0.111	0.129	0.111	0.112	0.111
OOS(EWMA)	0.414	0.195	0.100	0.127	1.556	0.906	0.996	0.906
Period 1: June 1961 - August 1971								
In-Sample	0.352	0.285	0.143	0.139	0.121	0.123	0.124	0.123
OOS(Fixed)	0.545	0.270	0.212	0.259	18.212	21.600	5.565	21.600
OOS(Recursive)	0.545	0.270	0.211	0.225	0.366	0.344	0.331	0.344
OOS(Rolling)	0.545	0.270	0.209	0.225	0.467	0.385	0.386	0.385
OOS(EWMA)	0.545	0.270	0.214	0.278	2.115	2.035	1.666	2.035
Period 2: September 1971 - September 1979								
In-Sample	0.427	0.322	0.176	0.173	0.160	0.160	0.158	0.160
OOS(Fixed)	0.338	0.239	0.190	0.190	0.306	0.303	0.235	0.303
OOS(Recursive)	0.338	0.239	0.190	0.194	0.239	0.226	0.222	0.226
OOS(Rolling)	0.338	0.239	0.191	0.196	0.255	0.238	0.234	0.238
OOS(EWMA)	0.338	0.239	0.230	0.269	0.586	0.545	0.511	0.545
Period 3: October 1979 - September 1982								
In-Sample	1.335	0.662	0.502	0.467	0.390	0.374	0.364	0.374
OOS(Fixed)	1.082	0.693	0.541	0.630	0.978	0.925	0.907	0.925
OOS(Recursive)	1.082	0.693	0.534	0.568	0.866	0.810	0.798	0.810
OOS(Rolling)	1.082	0.693	0.532	0.595	3.506	2.585	2.073	2.585
OOS(EWMA)	1.082	0.693	0.520	0.623	1.433	1.186	1.164	1.186
Period 4: October 1982 - December 2002								
In-Sample	0.484	0.224	0.149	0.138	0.129	0.129	0.129	0.129
OOS(Fixed)	0.416	0.207	0.114	0.103	0.282	0.279	0.210	0.279
OOS(Recursive)	0.416	0.207	0.113	0.102	0.120	0.118	0.119	0.118
OOS(Rolling)	0.416	0.207	0.105	0.106	0.117	0.116	0.112	0.116
OOS(EWMA)	0.416	0.207	0.103	0.133	1.540	1.730	1.810	1.730

Panel B: Hedge a 7 -Year Zero-Coupon Bond Using 3 Factors

	Unhedged	Bar-bell	Const.-Weight	Time-Varying Weights				
				Affine	Simple Poly.	Legendre Poly.	Fourier Trans.	Hermite Poly.
Whole Sample: June 1961 - December 2002								
In-Sample	1.158	0.753	0.508	0.503	0.460	0.458	0.457	0.458
OOS(Fixed)	0.956	0.357	0.263	0.270	0.480	0.614	0.371	0.614
OOS(Recursive)	0.956	0.357	0.255	0.256	0.312	0.278	0.270	0.278
OOS(Rolling)	0.956	0.357	0.249	0.240	0.357	0.271	0.265	0.271
OOS(EWMA)	0.956	0.357	0.235	0.293	4.498	10.019	8.754	10.019
Period 1: June 1961 - August 1971								
In-Sample	0.637	0.678	0.280	0.150	0.091	0.100	0.099	0.100
OOS(Fixed)	0.964	0.152	0.419	0.341	15.923	60.818	14.355	60.818
OOS(Recursive)	0.964	0.152	0.405	0.242	0.301	0.349	0.252	0.349
OOS(Rolling)	0.964	0.152	0.300	0.194	0.284	0.242	0.219	0.242
OOS(EWMA)	0.964	0.152	0.147	0.151	0.229	0.162	0.136	0.162
Period 2: September 1971 - September 1979								
In-Sample	0.807	0.788	0.556	0.545	0.493	0.488	0.492	0.488
OOS(Fixed)	0.619	0.467	0.311	0.339	0.620	0.698	0.755	0.698
OOS(Recursive)	0.619	0.467	0.308	0.299	0.433	0.452	0.443	0.452
OOS(Rolling)	0.619	0.467	0.313	0.289	0.454	0.588	0.555	0.588
OOS(EWMA)	0.619	0.467	0.259	0.284	0.814	0.744	0.724	0.744
Period 3: October 1979 - September 1982								
In-Sample	2.632	1.664	1.266	1.198	0.920	0.910	0.886	0.910
OOS(Fixed)	2.554	1.719	1.257	2.195	3.824	4.104	3.446	4.104
OOS(Recursive)	2.554	1.719	1.231	1.550	3.027	2.247	2.131	2.247
OOS(Rolling)	2.554	1.719	1.231	1.557	8.075	14.308	8.116	14.308
OOS(EWMA)	2.554	1.719	1.376	1.650	3.453	3.526	3.501	3.526
Period 4: October 1982 - December 2002								
In-Sample	1.122	0.512	0.333	0.326	0.310	0.310	0.310	0.310
OOS(Fixed)	0.940	0.374	0.194	0.217	1.023	2.431	1.097	2.431
OOS(Recursive)	0.940	0.374	0.193	0.210	0.304	0.336	0.282	0.336
OOS(Rolling)	0.940	0.374	0.201	0.204	0.348	0.471	0.406	0.471
OOS(EWMA)	0.940	0.374	0.192	0.268	5.478	11.622	9.671	11.622