Non-linearities in U.S. Treasury Rates: A Semi-Nonparametric Approach

by

Pierluigi Balduzzi and Young Ho Eom*

First draft, March 1997

This revision, August 1998

Abstract

This paper systematically tests for non-linearities in the relation between U.S. Treasury rates of different maturities. We document statistically and economically significant non-linearities that are robust to i) estimation method, ii) sample period, and iii) the presence of up to three underlying factors of uncertainty. These non-linearities are relevant for both pricing and hedging fixed-income securities. But we also find that the inclusion of extra factors of uncertainty in a linear specification is more important than the modeling of non-linearities.

JEL # G12

* Balduzzi is with the Carroll School of Management, Boston College. Eom is with Yonsei University, Seoul. The authors thank Dave Backus, Kobi Boudoukh, Silverio Foresi, George Martin, Matt Richardson, and seminar participants at the Federal Reserve Bank of New York, New York University, the Sixth Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, University of Massachusetts at Amherst, and the NBER for useful comments. The authors also thank Robert Bliss for providing the raw data and the Fortran codes used to generate the spot rates. This research was initiated when both authors were affiliated with New York University. Research support from New York University and Boston College is gratefully acknowledged. All correspondence should be addressed to Pierluigi Balduzzi, Finance Department, The Carroll School of Management, Fulton Hall, 140 Commonwealth Avenue, Chestnut Hill, Massachusetts 02167-3808; tel. (617) 552-3976, e-mail balduzzp@bc.edu.

Non-linearities in U.S. Treasury Rates: A Semi-Nonparametric Approach

August 1998

Abstract

This paper systematically tests for non-linearities in the relation between U.S. Treasury rates of different maturities. We document statistically and economically significant non-linearities that are robust to i) estimation method, ii) sample period, and iii) the presence of up to three underlying factors of uncertainty. These non-linearities are relevant for both pricing and hedging fixed-income securities. But we also find that the inclusion of extra factors of uncertainty in a linear specification is more important than the modeling of non-linearities.

Introduction

This is the first paper to systematically test for the presence of non-linearities in the relationship among zero-coupon, or "spot," interest rates of different maturities. This is also the first paper to compare the performance of affine term-structure models against non-affine models, in the context most relevant in practical applications: pricing and hedging fixed-income securities. Affine term structure models imply that all spot rates are linear functions of a common set of state variables. Hence, in the context of affine models, it is possible to uniquely invert the relationship between rates and risk variables, and any spot rate can be written as an exact linear function of as many other rates as the number of underlying risk factors. Deviations from linearity in the relation between spot rates of different maturity is thus an indication of the failure of the affine class.

The importance of the affine term structure model stems from its computational appeal, and in fact most of the models known in the literature belong to this class. Some of the many examples of affine models are Vasicek (1977), Cox, Ingersoll, and Ross (1985) (henceforth CIR), Pearson and Sun (1994), Chen and Scott (1993), Longstaff and Schwartz (1992), and Dai and Singleton (1998). 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 solution to a system of Riccati equations which can sometimes be calculated in closed-form (e.g. CIR); and even when numerical techniques are needed, the computational burden is modest.

The main results of our study can be summarized as follows: First, we document statistically-significant non-linearities in the relation between spot rates of maturities one, five, and ten years, and three spot-curve "factors:" the Level, Slope, and Curvature. These non-linearities generate substantial pricing differences relative to linear models, and are robust to: i) correction for serial correlation in the residuals; ii) controlling for simultaneity and errors-in-variables; iii) controlling for regime breaks; and iv) allowing for up to three underlying factors of uncertainty. Second, we find that non-linearities also matter in a hedging context: hedging based on a non-linear model, as opposed to a linear model, allows for a substantial reduction in the volatility of the hedged portfolio. Third, in the context of non-nested tests we find that increasing the number of risk factors is more important than accounting for non-linearities. Hence, the overall message of our analysis is that although non-linearities are present and robust, allowing for the appropriate number of factors (at least three) may be even more important than modeling the non-linearities.

The possibility that spot-curve behavior deviates from the affine paradigm was first suggested by

Aït-Sahalia (1996a,b) and Stanton (1997). These studies find that both the conditional mean and the conditional variance–the "drift" and "diffusion" functions in a continuous-time specification–of a short-term rate are non-linear functions of the level of the short rate. Indeed, Aït-Sahalia (1996b) considers all existing specifications where both drift and diffusion terms of the short-rate process are linear functions of the short rate, and finds that no specification is adequate in explaining the unconditional probability distribution of the short rate. More recently, Boudoukh, Richardson, Stanton, and Whitelaw (1997b) and Ghysels and Ng (1997) extend the analysis to a multi-factor setting, and document the presence of strong non-linearities in the joint process for the short rate and measures of the slope of the term structure. In addition, Stanton (1997) and Boudoukh et al. (1997b) document non-linearities in the dependence of risk premia on the underlying state variables.

While the evidence above suggests that non-linearities in the dynamics of the short rate and the risk premia are important, relatively little work has been done to model the spot curve based on non-linear processes. Among the few examples of non-affine term structure models in analytical closed form we have Constantinides (1992) and Longstaff (1989). On the other hand, Andersen and Lund (1997), Aït-Sahalia (1996b), Boudoukh et al. (1997b), and Stanton (1997) offer numerical solutions for the term structure and interest-rate derivatives.

Hence, both academics and practitioners are reluctant to abandon the affine class for more complicated models whose empirical performance is not yet clear. In addition, a recent study by Pritsker (1997) indicates that some of the evidence against the affine class may not be reliable. Specifically, Pritsker argues that Aït-Sahalia's (1996b) non-parametric test rejects the null hypothesis of linearity too often when using asymptotic critical values. The reason for the high rejection rate is that the asymptotic distribution of the test does not depend on persistence (which is very high for all spot rates), but the finite sample performance of the estimator does. Similar concerns are raised by Chapman and Pearson (1998) about the findings of both Aït-Sahalia (1996b) and Stanton (1997).

This paper aims at clarifying some of the issues by focusing on the contemporaneous relation between spot rates and spot-curve factors. This approach has several advantages over the estimation of the dynamic behavior of interest rates. First, the pricing implications of a non-linear *vs* a linear process for the relevant state variables are not immediate. And the implications for the term structure obtained by Aït-Sahalia (1996b), Andersen and Lund (1997), Boudoukh et al. (1997b), and Stanton (1997) using numerical methods are not compared to actual data. On the other hand, one would like to know the size of the pricing errors if one were to price bonds using the wrong (i.e. linear) model. By estimating linear and non-linear relations between spot rates and factors we directly address the issue of the pricing implications of non-linearities in interest rates. Second, the dynamics of the short rate may depend on other factors than the short rate itself: for example, a stochastic central tendency (Balduzzi, Das, and Foresi, 1998) and/or a stochastic volatility [Balduzzi, Das, Foresi, and Sundaram (1996), Andersen and Lund (1996, 1997), and Dai and Singleton (1998)]. These factors are unobservable and this makes the estimation of the short-rate process problematic.[1] On the other hand, when we consider the contemporaneous relation between spot rates and factors, it is relatively easy to extend the analysis to multiple spot-curve factors, and all of these factors are observable. This allows us to characterize the true relationship between different points and indicators of the term structure. Finally, spot-curve models are typically formulated based on the *continuous-time* process for one or more state variables. Unfortunately, the actual sampling of interest rates is discrete and may lead to a *discretization bias*.[2] By focusing on the contemporaneous relation between spot rates and factors we overcome the discretization-bias problem and we can rely on more standard econometric techniques.

The approach advocated here is based on the analysis of Duffie and Kan (1996). According to Duffie and Kan, a necessary and sufficient condition for zero-coupon bond rates to be affine functions of the underlying factors of uncertainty is that the short rate is linear in the factors and the factors themselves follow linear processes in the risk-neutral measure. Equivalently, a *sufficient* condition for bond yields to be linear is that the short rate and the risk premia are linear in the factors, and the factors follow linear processes in the real measure. One implication of Duffie and Kan's analysis is that any zero-coupon rate can be written as an exact linear function of as many other spot rates as the number of factors in the economy. Hence, if the data show that bond yields relate non-linearly to each other, this is direct evidence of non-linearities in the short rate and the risk-adjusted process for the factors. This is also strong evidence of non-linearities in the process for the risk factors, and of a non-linear dependence of risk premia on the factors.

Our tests compare the null hypothesis of a linear specification to a semi-nonparametric (SNP) non-linear alternative hypothesis. We choose an SNP as opposed to a nonparametric alternative specification for several reasons. First, the SNP approach allows us to draw from well-developed econometric techniques for identification and testing. Second, the semi-nonparametric approach is less prone to the curse-of-dimensionality problem of nonparametric techniques. This allows us

[1] Andersen and Lund (1996), for example, apply an efficient method of moments approach developed by Gallant and Tauchen (1994). Dai and Singleton (1998) use a simulated method of moments technique.

[2] This why Aït-Sahalia (1996a,b) and Stanton (1997) have developed density-based estimation techniques.

to study economies with several underlying factors of uncertainty. Third, the semi-nonparametric approach allows us to explicitly take into account the statistical properties of the residuals of the model, for instance, serial correlation.

Our approach allows us to address several issues of both academic and non-academic relevance. For example, we test whether the presence of persistent pricing errors might be responsible for "spurious" non-linearities in the relation between yields of different maturity.[3] We test whether changes in monetary policy regime may generate non-linear patterns in the data. We also test whether by allowing for several factors of uncertainty the non-linearities disappear. We run a horse-race of sorts between linear and non-linear models to establish whether modeling non-linearities is more or less important than the inclusion of several factors of uncertainty. Finally, we test for the statistical and economic relevance of non-linearities for the purpose of hedging fixed-income securities.

The paper is organized as follows: Section I presents the affine term structure models and discusses the empirical methodology used in the paper. Section II discusses the data. Section III performs some preliminary tests. Section IV performs some further tests where the estimation method explicitly accounts for the serial correlation in the residuals. Section V corrects for possible simultaneity and errors-in-variables biases. Section VI tests for non-linearities within subperiods. Section VII extends the analysis to several factors of uncertainty. Section VIII performs non-nested tests. Section IX performs hedging exercises. Section X concludes.

I. Affine Term Structure Models

In the following, we review three different approaches to the derivation of affine term structure models. The first approach is the "risk-neutral" approach, which is based on assumptions on the behavior of the state variables under the risk-neutral probability measure. This is the approach of Duffie and Kan (1996). The second approach is the "pricing-kernel" approach, which is based on assumptions on the behavior of the stochastic discount factor under the real probability measure. This is the approach of Backus and Zin (1994) in a discrete-time setting, and Bakshi and Chen (1997) in a continuous-time setting. The third approach is the "fundamental-valuation" approach, which is based on assumptions on the behavior of the short rate and the market prices of risk also

[3]The presence of persistent deviations from term structure models is widely documented. See, for example, Duffie and Singleton (1998).

under the real probability measure. This is the standard derivation of term structure models which is found, for example, in CIR. All three approaches are equivalent in terms of their implications for bond prices: bond prices are log-linear functions of the state variables and spot rates are *linear* functions of the same variables.

A. The Risk-Neutral Approach

The discussion that follows mimics closely Duffie and Kan (1996). Consider a given complete probability space (Ω, F, P) and the augmented filtration $\{F_t : t \geq 0\}$ generated by a standard Brownian motion in R^n. We suppose that there is a Markov process X, valued in some open subset D of R^n, such that, for any times t and τ, the market value $p_t(\tau)$ at time t of a zero-coupon bond maturing at time $t+\tau$, with face value \$1, is given by $f(X_t, \tau)$. f is twice continuously differentiable in X and once continuously differentiable in τ, $f \in C^{2,1}(D \times [0, \infty))$. The short-rate process R is defined by continuity, in the sense that there is a measurable function $R : D \to R$ defined as the limit of spot rates as maturity goes to zero, or

$$R_t = R(X_t) = \lim_{\tau \downarrow 0} \frac{-\log[f(X_t, \tau)]}{\tau} \tag{1}$$

As is well understood from Harrison and Kreps (1979), only technical regularity conditions are required for the equivalence between the absence of arbitrage and the existence of an equivalent martingale measure. An equivalent martingale measure is a probability measure Q equivalent to the real probability measure P, under which the price process of any security is a Q-martingale after normalization at each time t by the value $\int_0^t R(X_s)ds$ of continual re-investment of interest from one unit of account held from time zero at the short rate. Namely, we have

$$E_Q\left(\frac{dp_t(\tau)}{p_t(\tau)} \mid X_t\right) = R(X_t)dt \; a.s., \quad 0 \leq t \leq t+\tau \leq \infty \tag{2}$$

Suppose that X satisfies a stochastic differential equation of the form

$$dX_t = \mu^*(X_t)dt + \sigma(X_t)dW_t^* \tag{3}$$

where $\mu^* : D \to R^n$ and $\sigma : D \to R^{n\times n}$ are regular enough for (3) to have a unique solution. Additional regularity conditions imply that there is a standard Brownian motion W in R^n under the Q-measure such that

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{4}$$

where $\mu : D \to R^n$ is a function that can be calculated in terms of μ^*, σ, and f.

Here, we are interested in situations where

$$f(X_t, \tau) = E_Q \left[\exp \left(- \int_t^{t+\tau} R(X_s) ds \right) \mid X_t \right] \; a.s., \quad 0 \leq t \leq t + \tau \leq \infty \tag{5}$$

Equation (5) is merely the definition of Q as an equivalent martingale measure.

We consider a class of models where

$$f(X_t, \tau) = \exp[A(\tau) + B(\tau)X_t] \tag{6}$$

for which, given the assumptions on f, we know that $A(\tau)$ and $B(\tau)$ are continuous functions of τ on $[0, \infty)$. This parametric class of models is called *exponential-affine*, in light of the affine relationship between spot rates and factors. Equation (6) is the starting point of our empirical analysis. Duffie and Kan (1996) show that if μ, $\sigma\sigma^\top$, and R are affine functions of the state variables X, then f is exponential affine. Conversely, if f is exponential affine, then R is affine (by a continuity argument). Moreover, under regularity conditions, μ and $\sigma\sigma^\top$ are also affine.

Duffie and Kan also show that in order for μ and $\sigma\sigma^\top$ to be affine, we can essentially take (4) to be of the form

$$dX_t = (aX_t + b)dt + \Sigma V dW_t, \quad X_0 \in D \tag{7}$$

where $a \in R^{n \times n}$, $b \in R^n$, and $\Sigma \in R^{n \times n}$, and V is a diagonal matrix where

$$V_{ii} = \sqrt{\alpha_i + \beta_i^\top X} \tag{8}$$

with α_i a scalar and $\beta \in R^n$, for each i.

In light of the discussion above, testing the assumptions of the affine model amounts to testing the linearity of R, μ, and $\sigma\sigma^\top$. There are several difficulties with this approach. First, the short rate R is not easily observable, being the rate on a default-free obligation with an infinitesimal time to maturity. In the U.S. financial markets, for example, the shortest-maturity contracts are overnight, and only the fed funds and repo markets are active enough that price quotes are reliable. And both the fed funds and the repo rate have shortcomings as proxies of the short rate, since fed funds loans are unsecured, while repo rates include the convenience yield of holding liquid Treasury securities as collateral.[4] Second, the state variables X are, in general, not observable. This is true both for the "additive" version of the affine model, where the short rate is an affine function of a vector of diffusions [see, for example, Chen and Scott (1993), Pearson and Sun (1994), and Duffie and Singleton, (1998)], and for the "terraced" version of the affine model, in which the short rate itself is a state variable, and one of the state variables is the stochastic long-run mean of R [see Balduzzi et al. (1996), Balduzzi, Das, and Foresi (1998), Chen (1996), and Dai and Singleton (1998)]. Third, even if R and X were observable, the dynamic behavior of X can only be estimated under the real measure, while the assumption of the affine model relates to the dynamics of X under the risk-adjusted measure. Without explicit assumptions on preferences and attitudes towards risk, one cannot make the connection between the real and the risk-adjusted processes.[5] Fourth, the dynamic process for X is formulated in continuous time, while any actual sampling is generally discrete. This leads to the *aliasing* phenomenon, by which different continuous-time processes may appear to be identical when sampled at regular time intervals. Moreover, ignoring the effect of discrete sampling may lead to a discretization bias. Lo (1988), for example, shows that the common approach of estimating parameters of an Ito process by applying maximum likelihood to a discretization of the stochastic differential equation does not yield consistent estimators.

B. The Pricing-Kernel Approach

Sufficient conditions for the affine class can be derived based on assumptions on the stochastic discount factor. Assume that the prices of m bonds follow the Ito process $p = [p(\tau_1), p(\tau_2), \ldots, p(\tau_m)]$ in R^m,

$$dp_t = \mu_p(X_t)dt + \sigma_p(X_t)dW_t^* \tag{9}$$

[4]This effect is especially pronounced in the proximity of Treasury auctions when on-the-run securities go on "special." See Duffie (1996) for a discussion of the price effects induced by specialness.

[5]It is worth noting that, while the drift function differs in the risk-adjusted and real measures, the volatility function σ is the same.

where $\sigma_p \in R^{m\times n}$. Assume the absence of arbitrage opportunities. Then, given technical conditions, there exists a state price deflater π such that πp is a martingale under the real measure P. For any time t and $s > t$,

$$p(X_t, \tau_i) = E_P\left[\frac{\pi(X_s)}{\pi(X_t)}p(X_s, \tau_i - (s-t))\right] \tag{10}$$

The ratio π_s/π_t is the stochastic discount factor or *pricing kernel* for pricing the m securities in the absence of arbitrage. By Ito's lemma, the pricing kernel satisfies

$$\frac{d\pi_t}{\pi_t} = -R(X_t)dt - \Lambda'(X_t)dW_t^* \tag{11}$$

Integrating, we obtain

$$p(X_t, \tau) = E_P\left\{\exp\int_t^{t+\tau}[-R(X_s) + \Lambda'(X_s)\Lambda(X_s)/2]dt - \Lambda'(X_s)dW_t^* \mid X_t\right\} \tag{12}$$

The affine term structure model obtains by imposing the additional assumptions that R is affine in X and

$$\Lambda(X_t) = V(X_t)\lambda \tag{13}$$

where $\lambda \in R^n$, which in turn implies that $\Lambda'(X)\Lambda(X)$ is affine in X.[6]

The pricing-kernel derivation of the affine class suggests a different approach to testing the affine model. Rather than focusing on R and the dynamics of X under the risk-neutral measure Q, we can focus on π and the dynamics of X under the real measure P. In fact, for the affine model to hold, it is sufficient that $\pi(X_s)/\pi(X_t)$ is log-linear in X_s, and X follows an affine process.

Testing the assumptions of the affine model based on this pricing-kernel derivation amounts to testing that the pricing kernel is log-linear and the state variables follow an affine process. So far, few

[6]It is also worth noting that the pricing kernel formulation allows us to establish a link between the real and risk-adjusted process for the state variables X. Namely, we have

$$dX_t = (a^* + b^*X_t)dt + \Sigma V dW_t^*$$

where

$$a = a^* - \Sigma[\lambda_i\alpha_i], \quad b = b^* - \Sigma[\lambda_i\beta_i^\top]$$

authors have attempted to model the pricing kernel non-parametrically or semi-nonparametrically, and we are not aware of any paper which specifically tests the null hypothesis of log-linearity. For example, Bansal and Vishwanathan (1996) use selected asset returns as "factors" driving the pricing kernel. Their paper rejects the linearity of the kernel in the factors, but they do not test for log-linearity. Bansal, Hsieh, and Vishwanathan (1996) extend the analysis to the pricing of international asset returns, and do not reject linearity of a one-factor pricing kernel driven by the world market return. Again, the null of log-linearity is not tested. Chapman (1997) formulates the pricing kernel as a function of per-capita consumption growth and technology shocks, and his tests also do not consider the null of log-linearity.

C. The Fundamental-Valuation Approach

A direct implication of the pricing-kernel approach above is the standard APT-style equation used by CIR and many others:

$$\sigma_{pt}\Lambda_t = \mu_{pt} - R_t p_t \tag{14}$$

Equation (14) above is the counterpart of the fundamental valuation equation (FVE), equation (22), of CIR (1985). If R is affine in X, X follows an affine process under the real measure P, and Λ satisfies (13), then the affine term structure model obtains.

Based on the FVE derivation, one can test the assumptions of the affine class by testing the linearity of R, μ^*, $\sigma\sigma^\top$, and $\Lambda\Lambda^\top$. This approach has been taken by Stanton (1997) and Boudoukh et al. (1997b), who find evidence of non-linearities both in the dynamics of (a proxy for) the short-rate and in the risk premia.

D. Properties of the Affine Model

The main advantage of affine term structure models is their computability.[7] Another important feature of the affine term structure model is that it allows us to uniquely invert from the state variables X to a "basis" of n spot rates $Y^B \in R^n$. We have

$$Y_t(\tau_i)^B = -\frac{A(\tau_i) + B(\tau_i)X_t}{\tau_i} \tag{15}$$

Provided the matrix K, whose (i,j)-element is $-B_j(\tau_i)/\tau_i$, is non-singular, we know that $X_t = K^{-1}(Y_t^B + H)$, where $H_i = A(\tau_i)/\tau_i$, making the change in variables possible. In this case, we can write

$$dY_t^B = \mu^*(Y_t^B)dt + \sigma^*(Y_t^B)dW_t \tag{16}$$

where

$$\mu^*(Y_t^B) = \mu(K^{-1}(Y_t^B + H)), \quad \sigma^*(Y_t^B) = \sigma(K^{-1}(Y_t^B + H)) \tag{17}$$

The invertibility property has also been used in testing the affine model by Ghysels and Ng (1997) who estimate the dynamic process for yields of different maturities. In their specification, the drift and diffusion functions depend on a number of observable spot-curve spreads, and the null hypothesis of linearity in the spreads is nested within a general nonparametric alternative.

The equivalent term structure model is

$$f^E(Y_t^B, \tau) = \exp\left[A^E(\tau) + B^E(\tau)Y_t^B\right], \tag{18}$$

where $A^E(\tau) = A(\tau) + B(\tau)^\top K^{-1}H$ and $B^E(\tau)^\top = B(\tau)^\top K^{-1}$. Hence, in the affine class, any

[7] In fact, Duffie and Kan (1996) show that A and B can be calculated as the solutions of a pair of linear-quadratic Riccati equations:

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= a'B(\tau) + \frac{1}{2}\sum_{i=1}^{n}[\Sigma' B(\tau)]_i^2 \alpha_i \\ \frac{dB(\tau)}{d\tau} &= b'B(\tau) + \frac{1}{2}\sum_{i=1}^{n}[\Sigma' B(\tau)]_i^2 \beta_i + \delta \end{aligned}$$

where $\delta = \partial R/\partial X'$. Solutions exist on the whole time domain for special cases, such as CIR (1985), and, for any given particular case, they exist up to some time $\bar{T} > 0$.

spot rate can be written as a linear combination of as many other spot rates as the number of factors. This property has been extensively exploited in the estimation of term structure models, for example, Pearson and Sun (1994).

Equation (18) above is the implication of the affine model that we test in this paper. Thus, while other studies have tested the *assumptions* of the affine class, we focus on the *pricing implications* of the affine class.

E. Testing the Affine Class

In the empirical analysis we consider linear transformations of the basis Y^B. Specifically, we use variables which describe the position, slope, and curvature of the spot curve. To save on notation, these observable spot-curve factors are denoted with X, where $X \equiv \Gamma Y^B$, and $\Gamma \in R^{n\times n}$. Note that the theory implies that the number of observable spot-curve factors used in the empirical analysis equals the number of underlying risk factors. This consideration motivates the choice of up to three spot-curve factors which will be described in the following data section.

We focus on both the level specification

$$Y_t(\tau) = -A^E(\tau) - B^E(\tau)\Gamma^{-1}X_t \tag{19}$$

and the first-difference specification

$$Y_t(\tau) - Y_{t-\Delta t}(\tau) = -B^E(\tau)\Gamma^{-1}[X(t) - X_\tau(t-\Delta t)] \tag{20}$$

Empirically we test the specifications:

$$\begin{aligned} Y_t(\tau) &= -A^E(\tau) - B^E(\tau)\Gamma^{-1}X_t + g(X_t,\tau) + \text{error} \qquad (21)\\ Y_t(\tau) - Y_{t-\Delta t}(\tau) &= -B^E(\tau)\Gamma^{-1}(X_t - X_{t-\Delta t}) + l(X_t,\tau) + \text{error} \qquad (22)\end{aligned}$$

where g and l are generic smooth functions of X. If g and l enter significantly into the two models (21) and (22) then we can reject the affine family of term structure models in favor of more general non-linear alternatives.

Note that in both specifications the error terms have the interpretation of pricing errors: the component of Y_τ that cannot be explained in terms of the common underlying risk factors X. Although the presence of these pricing errors cannot be justified on theoretical grounds, they have been allowed in several existing tests of term structure models (see, for example, Duffie and Singleton, 1997). Since our approach is regression-based, we can explicitly test the properties of these pricing errors.

II. The Data

The main data set used in the analysis is the continuously-compounded, annualized, zero-coupon or *spot* rates, for maturities of three months, and one, two, eight, and ten years. The frequency of the data is weekly (Wednesday to Wednesday), and the sample covers the period June 14, 1961, through December 27, 1995. These rates were generated using the "Smoothed Fama-Bliss" estimation method described by Bliss (1996), applied to the daily CRSP bond data. This method was chosen because it is the best performing of the five methods reviewed by Bliss,[8] all of which involved fitting a smooth function to the prices of actively traded U.S. Treasury securities.

The ensuing empirical analysis focuses on the one-, five-, and ten-year rates as the dependent variables. These three rates should capture most of the dynamics in the spot curve for maturities of up to ten years. As explanatory variables we use three term-structure factors: the Level, Slope, and Curvature. The Level is proxied by the three-month rate of interest: $Y(0.25)$. The Slope is measured by the spread between the eight-year rate and the three-month rate: $Y(8) - Y(0.25)$. And, the Curvature is measured by the difference between two spreads: the spread between the eight-year and the two-year rate and the spread between the two-year and the three-month rate: $[Y(8) - Y(2)] - [Y(2) - Y(0.25)]$.

Table I presents summary statistics for the rates and the spot-curve factors for the entire 1961-1995 sample and for four separate subperiods, characterized by different monetary policy regimes. The first sub-period runs from June 1961 to August 1971. During this period the Bretton Woods exchange-rate agreement was in place and the operating target of monetary policy was "free reserves" (bank reserves in excess of required reserves and discount loans). The second sub-period within the sample goes from September 1971 to September 1979. During this period shocks to the prices of oil and other raw materials hit the U.S. economy and the focus of monetary policy shifted, now targeting the growth of monetary aggregates, mainly M1. The third sub-period spans from

[8] The five estimation methods tested by Bliss are: "Unsmoothed Fama-Bliss," "McCulloch Cubic Spline," "Fisher-Nychka-Zervos Cubic Spline," "Extended Nelson and Siegel," and "Smoothed Fama-Bliss."

October 1979 to September 1982. This period was characterized by high and volatile inflation and by operating procedures which exercised tight control of "non-borrowed reserves" (bank reserves in excess of discount loans). The last subperiod is from October 1982 to December 1995 period. This period is characterized by a general downward trend in inflation and interest rates and a renewed emphasis on interest rate targeting with a concern for the effects of monetary policy on exchange rates.

Examination of Table I shows that over the entire sample there is evidence of positive skewness and excess kurtosis.[9] Within individual sub-periods, sample skewness is less pronounced, even turning negative in some instances. In addition, kurtosis is systematically negative, both for the short ($Y(0.25)$) and the long ($Y(10)$) rate, and for all sub-periods. This evidence suggests the data may be generated by a mixture of distributions, where the mixing variable changes according to the interest-rate regime. Also, all variables follow highly persistent processes, with autocorrelation coefficients above 0.9, regardless of the sub-period.

Additional features of the data are presented in Figures 1 and 2. Figure 1 displays time-series graphs for the three spot-curve factors: Level, Slope, and Curvature; whereas, Figure 2 displays the average spot curves for the four sub-periods together with one-standard error bands.

In order to determine whether the three spot-curve factors adequately capture the joint behavior of the different spot rates, we also performed principal component analysis. We estimated the principal components from the spot rates for the 1961-1995 period. The correlations between the principal component factors and the Level, Slope, and Curvature is high. Interestingly, the Level factor displays very strong correlation with the first principal component (0.96). The Slope factor displays the strongest correlation with the second principal component (0.97) and the Curvature factor displays the strongest correlation with the third principal component (0.85). Hence, it appears that our three spot-curve factors capture the variability of interest rates similarly to the first three principal components.

III. Preliminary Evidence

This section presents some preliminary evidence on the relationship between spot rates of different maturities and the three spot-curve factors.

[9] Note that the Table reports *standardized* kurtosis, which is positive when the distribution has fatter tails than the standard normal.

Figures 3-5 present scatter plots and contour maps describing the joint behavior of the ten-year rate and the three factors. The contour maps denote the regions where most of the data is concentrated. These regions help provide a first indication of the patterns that will be further documented in the subsequent analysis. For example, the contour map of Figure 3 shows a flattening of the relation between the ten-year and the three-month rate when both rates are high relative to the bulk of the data. This is a first indication of a non-linear pattern in the relation between the two rates.

Table III presents some preliminary evidence based on three semi-nonparametric specifications. We estimate the models

$$Y_t(\tau) = m(X_t) + \epsilon_t \tag{23}$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_t is the "factor:" the Level, Slope, or Curvature of the spot curve.[10] The function $m(.)$ nests the linear specification as a special case. Hence equation (23) above is a restatement of equation (21).

We use three SNP specifications for the function m. The first specification is a simple polynomial:

$$m(X_t) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \eta_3 X_t^3 + \eta_4 X_t^4 + \eta_5 X_t^5 + \eta_6 \frac{1}{X_t} \tag{24}$$

The second specification is a sum of Fourier transforms:

$$m(X_t) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \sum_{j=1}^{2} [\theta_j \sin(jX_t) + \lambda_j \cos(jX_t)] \tag{25}$$

The third specification is a sum of orthogonal Legendre polynomials:

$$m(X_t) = \eta_0 + \eta_1 P_1(X_t) + \eta_2 P_2(X_t) + \eta_3 P_3(X_t) + \eta_4 P_4(X_t) + \eta_5 P_5(X_t) \tag{26}$$

where $P_j(.)$ is the j-th order Legendre polynomial:

$$\begin{aligned} P_1(X_t) &= [X_t/\sqrt{2}]\sqrt{3} \\ P_2(X_t) &= 0.5[(3X_t^2 - 1)/\sqrt{2}]\sqrt{5} \end{aligned}$$

[10] In the semi-nonparametric specifications based on Fourier transforms and Legendre polynomials, we take appropriate linear transformations of the three factors.

$$P_3(X_t) = 0.5[(5X_t^3 - 3X_t)/\sqrt{2}]\sqrt{7}$$
$$P_4(X_t) = [((35X_t^4 - 30X_t^2 + 3)/8)/\sqrt{2}]\sqrt{9}$$
$$P_5(X_t) = [((63X_t^5 - 70X_t^3 + 15X_t)/8)/\sqrt{2}]\sqrt{11}$$

Similar SNP specifications have been previously used in finance applications. Aït-Sahalia (1996b), for example, uses a polynomial specification for the drift function of the short-term rate. Legendre polynomials have been used by Chapman (1997) who argues that orthogonal polynomials have the advantage of reducing collinearity in the expansion terms relative to other SNP techniques.

One advantage of all three SNP specifications is that they are *linear* in the parameters, while they allow for a *non-linear* dependence on the state variables X. Hence, we can use ordinary least squares in the estimation, without resorting to more complicated numerical optimization techniques. The number of terms in all specifications was chosen so that the models are under-parameterized according to the Schwartz information criterion. This was done to avoid the possibility that the estimation picks up spurious non-linearities induced by an over-parameterized model. This approach is justified by the fact that the purpose of the study is not to estimate the "best" SNP specification, but to test the null hypothesis of linearity.

If spot rates are described by a one-factor linear model as in CIR and Vasicek (1977), the relationship between bonds yields and contemporaneous spot-curve factors is linear. Hence, we would expect the coefficient on all non-linear terms to be not significantly different from zero.

Table II reports the p-values for the chi-square statistics of the null hypothesis of linearity based on the heteroskedasticity adjustment due to White (1980), and based on the Newey and West (1987) correction for serial correlation, with 50 lags. The sample period is June 1961-December 1995.

The main results from Table II can be summarized as follows. First, the linearity assumption is strongly rejected (at the 1% level) for all three maturities, and all spot-curve factors. Second, the fit of different one-factor models changes substantially according to the choice of the factor. The Level is the "best" factor, explaining between 75% and 97% of the variability of the three spot rates. The Slope explains between 23% and 30% of the variability of the three spot rates, while the Curvature only explains between 17% and 18% of the variability of the three spot rates. Moreover, the residuals of the regressions exhibit a high degree of serial correlation. These results are remarkably similar across semi-nonparametric specifications, which suggests that this evidence is robust and that the non-linearities are not driven by overfitting the data with a particular SNP specification.

Figure 6 presents the relation between the ten-year rate and the three factors estimated using SNP specifications (left three panels) versus using non-parametric (NP) gaussian kernel regression (right three panels). For the kernel regression we use three different band-widths: the optimal band-width under the assumption of serial independence in the errors, h^*, and then twice and five times h^*. The widening of the band is required to correct for the correlation in the residuals.[11]

All three SNP specifications recover a concave relation between the ten-year rate and the Level of the spot curve. This basic shape is confirmed by the results of the kernel regression with band-width at h^* and $2h^*$.

The relationship between the long rate and the Slope is best described as a "U" shape. This pattern is consistent for both the SNP and NP specifications (h^* and $2h^*$), at least for a range of Slope values between -3% and 3%. When the Slope is positive, the long rate and the Slope correlate positively, which has the interpretation that a further steepening is likely to result in a higher long rate. In contrast, when the Slope is negative, the long rate and Slope correlate negatively since an increase in steepness occurs at the same time as the whole spot curve shifts down.

The relationship between the long rate and the Curvature is likewise described as a "U" shape. Again, this pattern is consistent across SNP and NP specifications (h^* and $2h^*$), at least for Curvature values between -2% and 2%. When the Curvature is negative, which is true for most of the sample, the long rate and the Curvature correlate negatively. This negative correlation indicates that a further increase in concavity (Curvature becomes more negative) is likely to result in a higher long rate. When the Curvature is positive though, the long rate correlates positively with the Curvature; a decrease in concavity (Curvature becomes more positive) occurs at the same time as the long rate increases.

While formal tests of the presence of non-linearities within individual sub-periods will be performed later on in the paper, as an example Figure 7 presents the relation between the long rate and the three factors for the 1961-1971 period. Interestingly, the non-linear patterns persist within this sub-period, and the general shape of the non-linearities is consistent with the evidence discussed above for the entire 1961-1995 period.

[11] Altman (1990) shows that standard band-width selection techniques favor undersmoothing when the errors are positively correlated. See also the discussion in Härdle and Linton (1994).

IV. Serial Correlation

Table III illustrates that the error term ϵ is highly persistent.[12] Persistence in the departures of actual bond prices from the values implied by term structure models is not uncommon. For example, Duffie and Singleton (1997) estimate a two-factor model of the term structure of swap rates and obtain pricing errors with a serial correlation coefficient of about 0.83, using weekly data.

Although the persistence in pricing errors may appear surprising, we argue that there are good reasons for it. First of all, it is unlikely that common risk factors will explain all the variation of each individual spot rate. The residual component, which is specific to each spot rate, may reflect liquidity and tax effects, for example. In order to prevent excess profits, one would expect this maturity-specific component to evolve similarly to a random walk. Hence, the high degree of serial correlation is consistent with an efficient-markets view of bond prices.

Second, the persistence of pricing errors also has to do with the statistical properties of the variables involved in the analysis. In fact, we have

$$m(X_t) \equiv E[Y_t(\tau)|X_t] \tag{27}$$

This is the expectation of the τ-maturity spot rate conditional on the realization of the variables in X. Let $\hat{m}(X)$ denote the estimate of $m(X)$. The residual $\hat{\epsilon}_t \equiv Y_t(\tau) - \hat{m}(X_t)$ can be decomposed as

$$\hat{\epsilon}_t = [Y_t(\tau) - m_t(X_t)] + [m_t(X_t) - \hat{m}_t(X_t)] \tag{28}$$

The term $Y(\tau) - m(X)$ is the "theoretical" component of the residual, while the term $m(X) - \hat{m}(X)$ is the "statistical" component of the estimation error. If the variables in X are very persistent, which is the case of all factors (see Table I), this translates into a very persistent statistical component of the residual.

The high persistence of the errors ϵ means that the least-squares estimates are highly inefficient and that the reported standard errors of the estimates are incorrect (the moving-average adjustment, even with 50 lags is clearly inadequate), which affects the size of the tests. In principle, the small-sample bias could be responsible for the appearance of non-linearities that are "spurious," and could

[12]Recall that an estimator of the serial correlation coefficient of ϵ is $1 - \frac{1}{2}$D.W. This means, for example, that the error terms from the third regression in Table III (ten-year rate) exhibit a serial correlation of $1 - 0.018/2 = 0.991$.

lead to an over-rejection of the null hypothesis of linearity. Hence, it is important to understand the nature of the serial correlation and to adjust for it in the estimation.

In the estimation, we consider two approaches to correct for the serial correlation in the residuals. In the first approach we estimate the model

$$Y_t(\tau) = m(X_t) + \epsilon_t \tag{29}$$

where[13]

$$\epsilon_t = \rho\epsilon_{t-1} + u_t \tag{30}$$

The error u_t is assumed to be serially uncorrelated. It is worth noting that this approach is equivalent to estimating the model

$$Y_t(\tau) - \rho Y_{t-1}(\tau) = [m(x_t) - \rho m(x_{t-1})] + u_t \tag{31}$$

One appealing feature of this approach is that we let the data determine the value of ρ. Also, it is important to note that the feasible GLS estimation above does not simply produce appropriate standard errors, but it also generates more efficient estimates.

An alternative approach is that of estimating the first-differences specification[14]

$$Y_t(\tau) - Y_{t-1}(\tau) = [m(X_t) - m(X_{t-1})] + u_t \tag{32}$$

This second approach is equivalent to the one above when $\rho = 1$.

Table III presents the results of estimating the Fourier-transform, SNP specification in equation (25) with the GLS correction for serial correlation and in first-differences form.

The main findings can be summarized as follows. First, the linearity assumption is rejected in all instances when the test is performed based on the level specification with GLS adjustment. We fail to reject linearity only for the first-differences specification for the relation between the

[13]While the AR(1) specification is arbitrary, it fits the data well. This is corroborated by the lack of serial correlation in the error term u found in all tests.

[14]Note that in the first-differences specification (32) we use the original function $m(X)$ on the r.h.s. Alternatively, we could have on the r.h.s. a flexible function of $X_t - X_{t-1}$. This approach would still allow us to test the affine class, since $Y_t - Y_{t-1}$ should be linear in $X_t - X_{t-1}$, but it would prevent us from recovering the relation between Y and X.

ten-year rate and the Level, and for the relation between the three spot rates and the Slope. These are instances where the Fourier transform specification is probably too parsimonious. In fact, the same tests performed using the Legendre polynomials specification lead to a strong rejection of the model. Second, the estimated values for ρ in the GLS estimation are extremely close to one. Hence, it appears that the first-difference specification is also appropriate. Third, the residuals u display little serial correlation.

In Figure 8 we graph the results of the estimation. For the three spot rates we present the difference between the estimated non-linear and linear relation. The graphs show how the non-linearities that we detect can be economically as well as statistically significant. In the case of the relation with the Slope, for example, the difference between the non-linear and the linear models can be as large as 80 basis points.

V. Instrumental Variables Estimation

One possible concern with the estimations performed so far is that the error u_t and the factor X_t could be correlated. This would lead to a systematic bias in the estimates, and possibly to spurious non-linearities.

We control for this possible source of bias in two ways. First, in the context of the GLS estimation, we instrument the spot-curve factors with their lagged values. This corrects for the possible correlation between the factors and the errors. Second, in the context of the first-differences estimation, we instrument the three factors with alternative spot-curve indicators: $Y(0.5)$ for the Level, $Y(7) - Y(0.5)$ for the Slope, and $[Y(7) - Y(3)] - [Y(3) - Y(0.25)]$ for the Curvature. This corrects for the possible correlation between the realizations of the factors and a component of the error which is specific to the relation between a given maturity-rate and a particular measure of the spot curve.

As in the previous section, we test the general specification in equation (25) against the null hypothesis of a one-factor linear model.

The results of the instrumental variables estimation are reported in Table IV. We find that the instrumental variables correction does not alter the results: non-linearities are significant across spot rates and estimation approaches. Indeed, we now find that linearity is strongly rejected in all instances for the first-differences specification.

VI. Structural Breaks

This section tests whether the non-linearities documented in the previous two sections proxy for structural breaks which take place when the monetary regime is changed. Hence, we use as break points those previously identified in Section III. We test the following three null hypotheses:

1. linear specification without structural breaks: L-NB;
2. linear specification with structural breaks: L-B;
3. non-linear specification without structural breaks: NL-NB;

against the alternative hypothesis: *non-linear* specification *with* structural breaks. We consider five possibilities for the presence of breaks. First, we allow for four different regimes. Then we allow for each of the four subperiods to be a separate regime relative to the rest of the sample. The presence of breaks is tested by allowing the coefficients θ_j and λ_j in the Fourier transform to differ across subsamples.

The results of the tests are summarized in Table V. First, as one would expect, the null of a linear model with no breaks (L-NB) is strongly rejected in most models. Second, the hypothesis of a linear model with breaks (L-B) is also widely rejected, which leads us to conclude that the non-linearities that we had previously detected are not simply proxying for regime changes. Third, the hypothesis of a non-linear model without structural breaks (NL-NB) is also widely rejected, with the exception of the models which use the Level as the spot-curve factor and allow the 1979-1982 period to differ from the rest of the sample. This finding is somewhat surprising since we expected that if one regime were to stand out, it should be the 1979-1982 period. Our interpretation for this finding is that the high volatility of the 1979-1982 period prevents us from estimating the regime-specific parameters with enough accuracy, and hence this period does not stand out simply for lack of power of the test.

The main message of these tests is that the non-linearities estimated in one-factor models do not proxy for structural breaks. Although controlling for breaks is needed, one still needs to model the yield curve non-linearly. In the remainder of the paper we shall extend the analysis to the case where more than one factor of uncertainty is at work in the yield curve. In our analysis of multi-factor models, we shall ignore explicit regime dummies. Still, we hope that the presence of several factors might capture the possible effects of changes in regime.

VII. Multi-Factor Models

In this section we first consider two-factor economies, where the Level and Slope, or the Level and Curvature are both needed to characterize the spot curve. We then consider three-factor economies, where Level, Slope, and Curvature are all needed for a complete characterization of the spot curve. Finally, we consider non-nested tests, where one-factor (Level) and two-factor (Level and Slope) non-linear models "compete" with two-factor (Level and Slope) and three-factor (Level, Slope, and Curvature) linear models, respectively.

A. Two-Factor Models

We estimate the model:

$$Y_t(\tau) = m_1(X_{1t}) + m_2(X_{2t}) + \epsilon_t \tag{33}$$

where the functions $m_1(.)$ and $m_2(.)$ are Fourier transforms and X_2 is made orthogonal to X_1.[15] The model above is estimated in first-differences.

We test the following four null hypotheses:

1. linear two-factor model: L;
2. two-factor model linear in the first factor, non-linear in the second factor: L1;
3. two-factor model non-linear in first factor, linear in the second factor: L2;
4. non-linear one-factor model (Level): NL1.

The alternative hypothesis is a non-linear two-factor model.

Our findings are summarized in Table VI. The null of linearity (L) is strongly rejected for both choices of second factor. This result is not surprising given the evidence from the previous sections. Second, the null of a model linear in the Level and non-linear in the second factor (L1) is not rejected for the Curvature. Third, the null of a model non-linear in the Level and linear in the second factor (L2) is strongly rejected. These last two results confirm the visual evidence of Figure 6: the non-linearities are stronger for the Slope and Curvature factors, than for the Level. Lastly, the null of a non-linear one-factor (NL1) model is strongly rejected. Hence, even when modeling

[15] In other words, X_2 is the residual of a linear regression of the "raw" factor X_2 on X_1 and a constant.

non-linearities, a second factor of uncertainty is needed: non-linearities in the one-factor context do not simply proxy for a missing second factor.

B. Three-Factor Models

This section considers the relation between spot rates of different maturity and all three spot-curve factors: Level, Slope, and Curvature.

We estimate the model:

$$Y_t(\tau) = m_1(X_{1t}) + m_2(X_{2t}) + m_3(X_{3t}) + \epsilon_t \tag{34}$$

where the functions $m_1(.)$, $m_2(.)$, and $m_3(.)$ are Fourier transforms and X_2 is made orthogonal to X_1, while X_3 is made orthogonal to both X_1 and X_2. The model above is estimated in first-differences.

We test the following five null hypotheses:

1. linear three-factor model: L;

2. three-factor model linear in the first and second factor, non-linear in the third factor: L1-L2;

3. three-factor model linear in the first and third factor, non-linear in the second factor: L1-L3;

4. two-factor non-linear model, Level and Slope: NL1-NL2;

5. two-factor non-linear model, Level and Curvature: NL1-NL3.

The alternative hypothesis is a non-linear three-factor model.

Our findings are summarized in Table VII. All null hypotheses are widely rejected. Hence, the non-linearities do not seem to proxy for the absence of a third factor. Moreover, even when modeling non-linearities in two factors, the presence of a third factor is required. This finding motivates the analysis of the next section where we compare the importance of modeling non-linearities to the importance of allowing for several sources of uncertainty.

C. Non-Nested Tests

Although non-linearities seem to play an important role for pricing, even when several spot-curve factors are considered, an open question is whether non-linearities are more or less important than including the “right” number of factors. This question is relevant from a practical standpoint, because most of the modeling efforts to date have focused on multi-factor affine models of interest rates, as opposed to non-linear models.

This question is addressed by running a horse race of sorts between a non-linear model with "few" factors and a linear model with more factors. The econometric approach is that of an encompassing regression where the fitted values from the two competing non-nested models are used as regressors in a third model. Our test is in the same spirit as the J-test of Davidson and MacKinnon (1981), where the fitted values of one model are used as a regressor together with the explanatory variables of the other model. One disadvantage of the J-test is that it may lead to the rejection of both models. Hence, we modify the J-test regression by estimating separately the two non-nested models and then using the fitted values as regressors in the encompassing regression. The slope coefficients and t-statistics of the fitted values from the two models provide an indication of which one of the two non-nested hypotheses is closest to describing the data.

We consider the following two pairs of non-nested hypotheses:

1. non-linear one-factor model, Level: NL1– linear two-factor model, Level and Slope: L2;
2. non-linear two-factor model, Level and Slope: NL2– linear three-factor model: L3.

The tests are performed for all three maturities.

The results of the non-nested tests are reported in Table VIII and these are probably among the most interesting findings of the paper. For both encompassing regressions and for all maturities, the linear model with one extra factor clearly outperforms the non-linear alternative. In fact, the slope coefficients on the fitted values of the linear models are always very close to one, and the t-ratios are high. On the other hand, the slope coefficients on the fitted values of the non-linear models are close to zero and insignificant. These results suggest that in the development of a term-structure model it is more important to account for the right number of factors than to model non-linearities.

VIII. Non-Linearities and Hedging

The previous sections have focused on the *pricing* implications of linear versus non-linear models. This section considers a second, equally important application of a term-structure model: hedging.

Several papers have studied the problem of hedging U.S. Treasury fixed-income securities in a linear context; see, for example, Elton, Gruber, and Nabar (1988), Elton, Gruber, and Michaely (1990), and Nelson and Shafer (1983). Recently, Boudoukh, Richardson, Stanton, and Whitelaw

(1997a) have extended the analysis to a non-linear setting, where non-parametric techniques are used to hedge mortgage-backed securities. The present paper combines these two strands of literature by applying a non-linear approach to the hedging of U.S. Treasury securities.

Consider the problem of hedging a zero-coupon bond with maturity τ_h. In a two-factor economy with continuous trading this can be achieved with the help of two other bonds with maturities τ_1 and τ_2. Namely, we want to construct a portfolio of three bonds whose exposure to the two risk factors is zero. Let $w(\tau_1)$ and $w(\tau_2)$ denote the portfolio weights of the two bonds with maturity τ_1 and τ_2, respectively. We have

$$\begin{bmatrix} w(\tau_1) \\ w(\tau_2) \end{bmatrix} = \begin{bmatrix} \tau_1 \frac{\partial Y(\tau_1)}{\partial X_1} - \tau_h \frac{\partial Y(\tau_h)}{\partial X_1} & \tau_2 \frac{\partial Y(\tau_1)}{\partial X_1} - \tau_h \frac{\partial Y(\tau_h)}{\partial X_1} \\ \tau_1 \frac{\partial Y(\tau_1)}{\partial X_2} - \tau_h \frac{\partial Y(\tau_h)}{\partial X_2} & \tau_2 \frac{\partial Y(\tau_1)}{\partial X_2} - \tau_h \frac{\partial Y(\tau_h)}{\partial X_2} \end{bmatrix}^{-1} \begin{bmatrix} -\tau_h \frac{\partial Y_t(\tau_h)}{\partial X_1} \\ -\tau_h \frac{\partial Y_t(\tau_h)}{\partial X_2} \end{bmatrix} \quad (35)$$

Note that, if the affine model holds, the derivatives of all rates with respect to the factors are constant, and hence the weights of the hedged portfolio are also constant over time. On the other hand, if rates depend non-linearly on the factors, the derivatives of the rates with respect to the factors change with the value of the underlying factors. In turn, this means that the portfolio weights are time-varying. Hence, deviations from the affine class imply that the hedging weights change over time as functions of the underlying risk factors.

The analysis above is empirically implemented as follows. Consider the $\Delta-$period excess return of the portfolio of three bonds: a one-year, a five-year, and a ten-year zero-coupon bond. The five-year bond corresponds to the τ_h-maturity bond that we want to hedge with the help of the two other bonds. The weights of the portfolio are, respectively, $w_t(1)$, $w(5) = 1 - w(1) - w(10)$, and $w(10)$. Let e_t denote the rate of return on the hedged portfolio of the three bonds, in excess of the risk-free rate.[16] We have

$$\begin{aligned} e_{t+\Delta t} &= w_t(1)[RET_{t+\Delta t}(1) - R_t] + [1 - w_t(1) - w_t(10)][RET_{t+\Delta t}(5) - R_t] \\ &\quad + w_t(10)[RET_{t+\Delta t}(10) - R_t] \end{aligned} \quad (36)$$

R_t is the beginning-of-period risk-free rate, and $RET_{t+\Delta t}(\tau)$ is the rate of return on a τ-maturity bond over the t-$t + \Delta t$ period.

Motivated by the analysis in continuous time, we model the optimal portfolio weights as unknown

[16] To prevent arbitrage, a perfectly hedged portfolio should earn the risk-free rate.

smooth functions of the observable factors. We then estimate the unknown weight function by approximating it with a sequence of orthogonal polynomials in the spot-curve factors. Namely, we model the weight functions as

$$w_t(\tau) = \phi_0(\tau) + \sum_{i=1}^{2} \sum_{j=1}^{5} \phi_{ij}(\tau) P_{ij}(X_{it}), \text{ for } \tau = 1, 10 \tag{37}$$

where $P_{ij}(.)$ is a j-th order Legendre polynomial.[17] We choose the parameters of the weight functions in equation (37) to minimize the variance of the hedging error in equation (36) above.

In our hedging exercise, we consider the *linear* case, where the weights are constant, and the *non-linear* case, where the weights are functions of two observable factors, the Level and the Slope. We use one week as the hedging interval (Δt =one week).[18] This approach is consistent, for example, with previous results by Elton, Gruber, and Nabar (1988), who show that:

1. two-index models outperform single-index models–hence we use a portfolio of three bonds to hedge two underlying risk factors;

2. observable indices perform better than factors extracted from the data–hence we use our spot-curve factors as opposed to the principal components.

Table IX reports standard deviations for the excess returns on an "unhedged" portfolio (the five-year bond), the hedged portfolio with constant weights (affine case), and the hedged portfolio with time-varying weights (non-linear case). In addition, we report the p-value of the F-statistic for the following three null hypotheses:

1. weights vary with second factor only, the Slope: L1;

2. weights vary with first factor only, the Level: L2;

3. weights are constant: L.

Table IX shows that the annualized standard deviation for the hedged portfolio with constant weights is 1.06 percent, while the standard deviation of the "unhedged" portfolio (five-year bond)

[17]Note that an alternative way to calculate the optimal weights would be to directly estimate the partial derivatives $\partial Y(\tau)/\partial f_i$ from the data and use these estimated partial derivatives to obtain the weights, as in equation (35). When we implemented this alternative procedure we obtained disappointing results in terms of hedging performance.

[18]Since reliable figures for a weekly rate cannot be derived using the available data, we constructed the weekly rate from the monthly one.

is 6.03 percent. Allowing for time-varying weights further reduces the standard deviation to 0.98 percent. This is approximately a 7.1 percent reduction relative to the constant weights case. Hedging also substantially reduces the average excess return on the portfolio: from 7.5 percent to 0.82 percent. Based on the F-statistics, the null hypothesis of constant weights is strongly rejected (0.000 p-value). The null hypotheses of constant weights either in the Level or in the Slope are also strongly rejected, indicating that nonlinearities in both the Level and the Slope are important in implementing a hedging strategy.

Figure 9 provides a further illustration of the hedging performance. The first two panels of Figure 9 show the behavior over time of the two weights associated with the one-year and the five-year bond. Both time-varying weights exhibit substantial volatility around the time of the Fed experiment, the 1979-1982 period. Over the remaining part of the sample the time-varying weights are quite close to the constant ones obtained under the assumption of linearity. The lower four panels show the behavior of the two weights as functions of the two spot-curve factors, Level and Slope. For values of the Level between 2 and 15 percent, the time-varying weights exhibit little variation and are quite close to the constant ones. When the Level exceeds 15 percent the two weights quickly decrease. The behavior of the two weights as functions of the second factor, the Slope, is similar. For values of the Slope between -1 and 4 percent, the time-varying weights are quite close to the constant ones, but when the Slope goes below -1 percent the two weights quickly increase.

The main messages from the hedging exercises can be summarized as follows. First, hedging with the affine model substantially reduces the variability of excess returns, although a further reduction can be achieved with the time-varying weights motivated by a non-linear model. Second, the time-variation of the hedging weights is statistically significant. Third, hedging with time-varying weights is more important for high values of the Level of the spot curve and for very negative values of the Slope (strongly inverted curve). This means that in "normal" situations, when interest rates are at low to intermediate values and the spot curve is positively sloped, hedging with the affine model gives results similar to hedging with the more complicated non-linear model.

IX. Conclusions

Several recent studies have documented non-linearities in the dependence of the conditional mean and volatility of spot rates on observable spot-curve factors. In addition, some studies have found that the estimated market price of risk also depends non-linearly on observable factors. According to Duffie and Kan (1996), these non-linearities in interest-rate dynamics and risk premia should translate into non-linear relations between contemporaneous zero-coupon rates of different maturity. This is the implication tested in this paper.

This paper finds evidence of statistically-significant non-linearities in the relation between spot rates of maturity one, five, and ten years, and three spot-curve factors (Level, Slope, and Curvature). These non-linearities generate substantial pricing discrepancies relative to linear models and are robust to: i) correction for serial correlation in the pricing errors; ii) controlling for simultaneity and errors-in-variables; iii) controlling for regime breaks; and iv) allowing for up to three factors of uncertainty. But, in the context of non-nested tests, we find that increasing the number of risk factors is more important than accounting for non-linearities. We also find that non-linearities matter in a hedging context: hedging based on a non-linear model, as opposed to a linear model, allows for a reduction in the volatility of the hedged portfolio. Yet, hedging with an affine model is appropriate for "normal" values of the position and shape of the spot curve.

References

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

Aït-Sahalia, Yacine, 1996b, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, 527-560.

Altman, N.S., 1990, Kernel smoothing of data with correlated errors, *Journal of the American Statistical Association* 85, 749-759.

Andersen, Torben G., and Jesper Lund, 1996, The short rate diffusion revisited: an investigation guided by the efficient method of moments, *Working paper*, Northwestern University.

Andersen, Torben G., and Jesper Lund, 1997, Stochastic volatility and mean drift in the short term interest rate diffusion: sources of steepness, Level and Curvature in the yield curve, *Working paper*, Northwestern University.

Backus, Dave K., and Stan E. Zin, 1994, Reverse engineering the yield curve, *Working paper*, New York University.

Bakshi, Gurdip S., and Zhiwu Chen, 1997, Asset pricing without consumption or market portfolio data, *Working paper*, University of Maryland, College Park.

Balduzzi, Pierluigi, Sanjiv R. Das, and Silverio Foresi, 1998, The central tendency: a second factor in bond yields, *Review of Economics and Statistics* 80, 62-72.

Balduzzi, Pierluigi, Sanjiv R. Das, Silverio Foresi, and Rangarajan K. Sundaram, 1996, A simple approach to three-factor affine models of the term structure, *Journal of Fixed Income* 6, 43-52.

Bansal, Ravi, and S. Vishwanathan, 1996, No arbitrage and arbitrage pricing: A new approach, *Journal of Finance* 48, 1231-1262.

Bansal, Ravi, S. Vishwanathan, and David A. Hsieh, 1996, A new approach to international arbitrage pricing, *Journal of Finance* 48, 1719-1747.

Bliss, Robert R., 1996, Testing term structure estimation methods, *Working paper*, Federal Reserve Bank of Atlanta

Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert Whitelaw, 1997a, Pricing mortgage-backed securities in a multifactor interest rate environment: a multivariate density estimation approach, *Review of Financial Studies* 10, 405-446.

Boudoukh, Jacob, Matthew Richardson, Richard Stanton, and Robert Whitelaw, 1997b, The stochastic behavior of interest rates, *Working paper*, New York University.

Chapman, David A., 1997, Approximating the asset pricing kernel, *Journal of Finance* 52, 1383-1410.

Chapman, David A., and Neil D. Pearson, 1998, Is the short rate drift actually nonlinear? *Working paper*, University of Texas, Austin.

Chen, Lin, 1996, Stochastic mean and stochastic volatility-a three-factor model of the term structure of interest rates and its application to the pricing of interest rate derivatives *Blackwel Publishers.*

Chen, Ren-Raw, and Louis Scott, 1993, Maximum likelihood estimation for a multi-factor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14-31.

Constantinides, George M., 1992, A theory of the nominal term structure of interest rates, *Review of Financial Studies* 5, 531-552.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385-407.

Davidson, Robert, and John G. MacKinnon, 1981, Several tests for model specification in the presence of alternative hypotheses, *Econometrica* 49, 781-793.

Dai, Qiang, and Kenneth J. Singleton, 1998, Specification analysis of affine term structure models, *Working paper*, Stanford University.

Duffie, Darrell, 1996, Special repo rates, *Journal of Finance* 51, 493-526.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* October, 379-406.

Duffie, Darrell, and Kenneth J. Singleton, 1997, An econometric model of the term structure of interest-rate swap yields, *Journal of Finance* 52, 1287-1321.

Elton, Edwin J., Martin J. Gruber, and Prafulla G. Nabar, 1988, Bond returns, immunization and the return generating process, *Studies in Banking and Finance* 5, 125-154.

Elton, Edwin J., Martin J. Gruber, and Roni Michaely, 1990, The structure of spot rates and immunization, *Journal of Finance* 45, 629-642.

Gallant, A. Ronald, and George E. Tauchen, 1994, Which moments to match?, *Econometric Theory* 12, 657-681.

Härdle, Wolfgang, and Olivier Linton, 1994, Applied nonparametric models, in *The Handbook of Econometrics*, R.F. Engle and D.F. McFadden eds., North Holland.

Harrison, J. Michael, and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381-408.

Ghysels, Eric, and Serena Ng, 1997, A semi-parametric factor model of interest rates and tests of the affine term structure, *Working paper*, Pennsylvania State University.

Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory* 4, 231-247.

Nelson, John, and Stephen Shaefer, 1983, The dynamics of the term structure and alternative portfolio immunization strategies, in G. Bierwag, G Kaufman, and A. Toevs, eds.: *Innovations in Bond Portfolio Management: Duration Analysis and Immunization*, (JAI Press, Greenwich, CT).

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

Longstaff, Francis A., 1989, A non-linear general equilibrium model of the term structure of interest rates, *Journal of Financial Economics* 23, 195-224.

Longstaff, Francis A., and Eduardo Schwartz, 1992, Interest-rate volatility and the term structure: a two-factor general equilibrium model, *Journal of Finance* 47, 1259-1282.

Pearson, Neil, and Tong-Sheng Sun, 1994, An empirical examination of Cox, Ingersoll, and Ross model of the term structure of interest rates using the method of maximum likelihood, *Journal of Finance* 49, 929-359.

Pritsker, Matt, 1997, Nonparametric density estimation and tests of continuous time interest rate models, *Working paper*, Board of Governors of the Federal Reserve System.

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

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

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

Table I

Summary Statistics

We report summary statistics for six selected zero-coupon Treasury spot rates and for the *Slope* and *Curvature* "factors" for the sample period June 1961-December 1995, and for the four subperiods June 1961-August 1971, September 1971-September 1979, October 1979-September 1982, and October 1982-December 1993. All spot rates and factors are quoted in percentage points per year. Observations are weekly (Wednesday to Wednesday, or next business day if a holiday). "Mean" is the sample mean. "Std." is the standard deviation. "Auto" is the first-order correlation coefficient. The kurtosis measure is standardized: it is zero for a normal random variable. $Y(\tau)$ is the τ-year zero-coupon rate. "Slope" is defined as $Y(8) - Y(0.25)$. "Curvature" is defined as $[Y(8) - Y(2)] - [Y(2) - Y(0.25)]$.

	Mean	Std	Skewness	Kurtosis	Auto
	June 1961-December 1995				
$Y(0.25)$	6.37	2.82	1.22	1.72	0.995
$Y(1)$	6.79	2.74	0.99	0.91	0.996
$Y(5)$	7.36	2.51	0.74	0.20	0.997
$Y(10)$	7.59	2.45	0.62	0.03	0.997
Slope	1.16	1.39	-0.59	0.63	0.988
Curvature	-0.12	0.64	-0.20	2.82	0.944
	June 1961-August 1971				
$Y(0.25)$	4.48	1.39	0.50	-0.56	0.995
$Y(10)$	5.12	1.19	0.75	-0.69	0.994
Slope	0.61	0.62	0.63	0.14	0.983
Curvature	-0.10	0.41	0.07	0.50	0.906
	September 1971-September 1979				
$Y(0.25)$	6.42	1.80	0.42	-0.81	0.998
$Y(10)$	7.38	0.80	0.00	-0.86	0.997
Slope	0.92	1.32	-0.49	-1.13	0.992
Curvature	-0.23	0.50	0.23	-0.09	0.901
	October 1979-September 1982				
$Y(0.25)$	12.63	2.57	-0.43	-0.53	0.968
$Y(10)$	12.20	1.43	-0.30	-0.95	0.959
Slope	-0.42	2.08	0.11	-0.87	0.975
Curvature	-0.27	1.23	-0.15	0.22	0.944
	October 1982-December 1995				
$Y(0.25)$	6.37	2.03	-0.05	-0.85	0.997
$Y(10)$	8.59	1.79	0.66	-0.32	0.996
Slope	2.09	1.01	-0.50	-0.62	0.989
Curvature	-0.02	0.64	-0.06	0.37	0.959

Table II
Preliminary Evidence

We estimate the models

$$Y_t(\tau) = m(X_t) + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_t is the "factor:" the Level, Slope, or Curvature of the spot curve. In the semi-nonparametric specifications based on Fourier transforms and Legendre polynomials, we take appropriate linear transformations of the three factors. We use three semi-nonparametric specifications for the function m. The first specification is a simple polynomial:

$$m(X_t) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \eta_3 X_t^3 + \eta_4 X_t^4 + \eta_5 X_t^5 + \eta_6 \frac{1}{X_t}$$

The second specification is a sum of Legendre polynomials:

$$m(X_t) = \eta_0 + \eta_1 P_1(X_t) + \eta_2 P_2(X_t) + \eta_3 P_3(X_t) + \eta_4 P_4(X_t) + \eta_5 P_5(X_t)$$

where $P_j(.)$ is the j-th order Legendre polynomial The third specification is a sum of Fourier transforms:

$$m(X_t) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \sum_{j=1}^{2} [\theta_j \sin(jX_t) + \lambda_j \cos(jX_t)]$$

Under "χ^2(W)" we report the p-value for the chi-square statistic of the null hypothesis of linearity, based on the heteroskedasticity adjustment due to White (1980). Under "χ^2(NW)" we report the p-value for the chi-square statistic of the null hypothesis of linearity, based on the Newey and West (1987) correction with 50 lags. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. The sample period is June 1961-December 1995.

Simple Polynomial

	χ^2(W)	χ^2(NW)	ADJ. R^2	D.W.
		Level		
$Y(1)$	0.000	0.000	0.975	0.096
$Y(5)$	0.000	0.000	0.829	0.022
$Y(10)$	0.000	0.001	0.747	0.018
		Slope		
$Y(1)$	0.000	0.000	0.294	0.035
$Y(5)$	0.000	0.000	0.212	0.032
$Y(10)$	0.000	0.000	0.235	0.033
		Curvature		
$Y(1)$	0.000	0.000	0.180	0.059
$Y(5)$	0.000	0.000	0.191	0.052
$Y(10)$	0.000	0.000	0.172	0.051

Table II Continued

Legendre Polynomials

	χ^2(W)	χ^2(NW)	ADJ. R^2	D.W.
		Level		
$Y(1)$	0.000	0.000	0.974	0.095
$Y(5)$	0.000	0.004	0.829	0.022
Y(10)	0.000	0.019	0.747	0.018
		Slope		
$Y(1)$	0.000	0.000	0.293	0.030
$Y(5)$	0.000	0.000	0.211	0.027
Y(10)	0.000	0.000	0.234	0.028
		Curvature		
$Y(1)$	0.000	0.000	0.180	0.058
$Y(5)$	0.000	0.000	0.191	0.051
Y(10)	0.000	0.000	0.172	0.049

Fourier Transforms

	χ^2(W)	χ^2(NW)	ADJ. R^2	D.W.
		Level		
$Y(1)$	0.000	0.000	0.975	0.097
$Y(5)$	0.000	0.004	0.829	0.022
Y(10)	0.000	0.018	0.747	0.018
		Slope		
$Y(1)$	0.000	0.000	0.302	0.026
$Y(5)$	0.000	0.000	0.221	0.023
Y(10)	0.000	0.000	0.244	0.024
		Curvature		
$Y(1)$	0.000	0.000	0.187	0.065
$Y(5)$	0.000	0.000	0.194	0.054
Y(10)	0.000	0.000	0.174	0.051

Table III
Correcting for Serial Correlation

We estimate the Fourier transform model

$$Y_t(\tau) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \sum_{j=1}^{2} [\theta_j \sin(jX_t) + \lambda_j \cos(jX_t)] + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_t is the "factor:" the Level, Slope, or Curvature of the spot curve. The model is estimated in levels with serial correlation adjustment, and in first differences. Under "F" we report the p-value for the F statistic of the null hypothesis of linearity. Under "χ^2" we report the p-value for the chi-square statistic of the null hypothesis of linearity. "ρ" is the autocorrelation coefficient for the errors in the level specification. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. In the first-differences specification, the chi-square statistic is calculated based on the White (1980) correction. The sample period is June 1961-December 1995.

Levels Specification, GLS Adjustment

	F	ρ	ADJ. R^2	D.W.
		Level		
$Y(1)$	0.000	0.987	0.998	1.862
$Y(5)$	0.000	0.996	0.997	1.809
$Y(10)$	0.008	0.996	0.997	1.901
		Slope		
$Y(1)$	0.000	0.997	0.995	1.681
$Y(5)$	0.000	0.997	0.996	1.729
$Y(10)$	0.000	0.997	0.996	1.848
		Curvature		
$Y(1)$	0.000	0.996	0.994	1.582
$Y(5)$	0.020	0.997	0.996	1.780
$Y(10)$	0.001	0.997	0.996	1.865

First-differences Specification

	χ^2	ADJ. R^2	D.W.
		Level	
$Y(1)$	0.003	0.687	1.873
$Y(5)$	0.010	0.407	1.812
$Y(10)$	0.392	0.280	1.904
		Slope	
$Y(1)$	0.139	0.197	1.683
$Y(5)$	0.001	0.036	1.731
$Y(10)$	0.000	0.051	1.850
		Curvature	
$Y(1)$	0.322	0.044	1.585
$Y(5)$	0.468	0.055	1.782
$Y(10)$	0.199	0.009	1.866

Table IV
Instrumental Variables Estimation

We estimate the Fourier transform model

$$Y_t(\tau) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \sum_{j=1}^{2} [\theta_j \sin(jX_t) + \lambda_j \cos(jX_t)] + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_t is the "factor:" the Level, Slope, or Curvature of the spot curve. The model is estimated in levels, with serial correlation adjustment, and first differences, using as instruments the functional differences of $Y(0.5)$, $Y(7) - Y(0.5)$, and $[Y(7) - Y(3)] - [Y(3) - Y(0.5)]$. Under "F" we report the p-value for the F statistic of the null hypothesis of linearity. Under "χ^2" we report the p-value for the chi-square statistic of the null hypothesis of linearity. "ρ" is the autocorrelation coefficient for the errors in the level specification. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. In the first-differences specification, the chi-square statistic is calculated based on the White (1980) correction. The sample period is June 1961-December 1995.

Levels Specification, GLS Adjustment

	F	ρ	ADJ. R^2	D.W.
		Level		
$Y(1)$	0.207	0.965	0.998	1.937
$Y(5)$	0.030	0.997	0.997	1.799
$Y(10)$	0.017	0.997	0.996	1.879
		Slope		
$Y(1)$	0.000	0.996	0.994	1.769
$Y(5)$	0.000	0.997	0.995	1.788
$Y(10)$	0.003	0.997	0.996	1.884
		Curvature		
$Y(1)$	0.000	0.996	0.993	1.691
$Y(5)$	0.000	0.997	0.995	1.820
$Y(10)$	0.000	0.997	0.995	1.898

First-differences Specification

	χ^2	ADJ. R^2	D.W.
		Level	
$Y(1)$	0.000	0.656	1.932
$Y(5)$	0.000	0.379	1.802
$Y(10)$	0.025	0.256	1.873
		Slope	
$Y(1)$	0.041	0.136	1.788
$Y(5)$	0.008	-0.017	1.830
$Y(10)$	0.001	-0.002	1.937
		Curvature	
$Y(1)$	0.002	-0.591	1.758
$Y(5)$	0.000	-0.171	1.821
$Y(10)$	0.041	-0.122	1.859

Table V
Testing for Structural Breaks

We estimate the Fourier transform model

$$Y_t(\tau) = \eta_0 + \eta_1 X_t + \eta_2 X_t^2 + \sum_{j=1}^{2}[\theta_j \sin(jX_t) + \lambda_j \cos(jX_t)] + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_t is the "factor:" the Level, Slope, or Curvature of the spot curve. The model is estimated in first differences. Under "Whole Period," June 1961-December 1993, we test whether four different regimes took place. Under "First Period," June 1961-August 1971, "Second Period," September 1971-September 1979, "Third Period," October 1979-September 1982, and "Fourth Period," October 1982-December 1993, we test whether two different regimes took place (the "Period" and the rest of the sample). Under "χ^2:L-NB" we report the p-value for the chi-square statistic of the null hypothesis of linearity with no structural breaks. Under "χ^2:L-B" we report the p-value for the chi-square statistic of the null hypothesis of linearity with structural breaks. Under "χ^2:NL-NB" we report the p-value for the chi-square statistic of the null hypothesis of non-linearity with no structural breaks. The chi-square statistics are calculated based on the White (1980) correction. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic.

Level

	χ^2:L-NB	χ^2:L-B	χ^2:NL-NB	ADJ. R^2	D.W.
			Whole Period		
$Y(1)$	0.000	0.000	0.000	0.702	1.895
$Y(5)$	0.000	0.000	0.000	0.451	1.824
$Y(10)$	0.000	0.000	0.000	0.315	1.908
			First Period		
$Y(1)$	0.000	0.001	0.000	0.687	1.872
$Y(5)$	0.026	0.014	0.078	0.407	1.808
$Y(10)$	0.149	0.181	0.608	0.279	1.899
			Second Period		
$Y(1)$	0.000	0.000	0.000	0.697	1.863
$Y(5)$	0.000	0.000	0.000	0.425	1.799
$Y(10)$	0.000	0.000	0.000	0.296	1.895
			Third Period		
$Y(1)$	0.024	0.009	0.939	0.688	1.886
$Y(5)$	0.080	0.058	0.712	0.409	1.830
$Y(10)$	0.388	0.613	0.501	0.281	1.918
			Fourth Period		
$Y(1)$	0.000	0.000	0.000	0.696	1.907
$Y(5)$	0.000	0.000	0.000	0.443	1.830
$Y(10)$	0.000	0.000	0.000	0.308	1.914

Table V Continued

Slope

	χ^2:L-NB	χ^2:L-B	χ^2:NL-NB	ADJ. R^2	D.W.
			Whole Period		
$Y(1)$	0.000	0.000	0.000	0.264	1.698
$Y(5)$	0.000	0.000	0.000	0.135	1.776
$Y(10)$	0.000	0.000	0.000	0.205	1.867
			First Period		
$Y(1)$	0.000	0.000	0.000	0.203	1.691
$Y(5)$	0.000	0.000	0.000	0.047	1.741
$Y(10)$	0.000	0.000	0.000	0.119	1.835
			Second Period		
$Y(1)$	0.007	0.025	0.005	0.202	1.696
$Y(5)$	0.000	0.003	0.000	0.039	1.732
$Y(10)$	0.000	0.001	0.077	0.053	1.848
			Third Period		
$Y(1)$	0.001	0.002	0.176	0.219	1.681
$Y(5)$	0.000	0.000	0.057	0.074	1.743
$Y(10)$	0.000	0.000	0.009	0.106	1.860
			Fourth Period		
$Y(1)$	0.000	0.000	0.000	0.246	1.689
$Y(5)$	0.000	0.000	0.000	0.109	1.748
$Y(10)$	0.000	0.000	0.000	0.116	1.870

Curvature

	χ^2:L-NB	χ^2:L-B	χ^2:NL-NB	ADJ. R^2	D.W.
			Whole Period		
$Y(1)$	0.000	0.000	0.000	0.106	1.585
$Y(5)$	0.000	0.000	0.000	0.120	1.813
$Y(10)$	0.000	0.000	0.000	0.085	1.870
			First Period		
$Y(1)$	0.000	0.215	0.000	0.042	1.588
$Y(5)$	0.000	0.344	0.000	0.054	1.786
$Y(10)$	0.000	0.000	0.000	0.038	1.849
			Second Period		
$Y(1)$	0.034	0.146	0.083	0.045	1.589
$Y(5)$	0.000	0.134	0.000	0.056	1.784
$Y(10)$	0.000	0.000	0.000	0.017	1.858
			Third Period		
$Y(1)$	0.017	0.032	0.023	0.081	1.584
$Y(5)$	0.097	0.150	0.032	0.084	1.794
$Y(10)$	0.240	0.170	0.363	0.019	1.882
			Fourth Period		
$Y(1)$	0.000	0.000	0.000	0.094	1.587
$Y(5)$	0.000	0.000	0.000	0.115	1.815
$Y(10)$	0.000	0.000	0.000	0.064	1.891

Table VI
Two-Factor Models

We estimate the model

$$Y_t(\tau) = m_1(X_{1t}) + m_2(X_{2t}) + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_{it} is the "factor:" the Level, Slope, or Curvature of the spot curve. $m_i(X_i)$ is the sum of Fourier transforms

$$m_i(X_{it}) = \eta_{i0} + \eta_{i1}X_{it} + \eta_{i2}X_{it}^2 + \sum_{j=1}^{2}[\theta_{ij}\cos(jX_{it}) + \lambda_{ij}\sin(jX_{it})]$$

The model is estimated in first differences. Under "χ^2:L" we report the p-value for the chi-square statistic of the null hypothesis of linearity in both factors. Under "χ^2:L1" we report the p-value for the null hypothesis of linearity in the first, but not the second factor. Under "χ^2:L2" we report the p-value for the null hypothesis of linearity in the second, but not the first factor. Under "χ^2:NL1" we report the p-value for the null hypothesis of non-linear one-factor (the first factor) model. In all tests, the statistics are calculated based on the White (1980) correction. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. The sample period is June 1961-December 1995.

	χ^2:L	χ^2:L1	χ^2:L2	χ^2:NL1	ADJ. R^2	D.W.
Level and Slope						
$Y(1)$	0.000	0.017	0.005	0.000	0.817	2.027
$Y(5)$	0.000	0.000	0.053	0.000	0.929	2.291
$Y(10)$	0.000	0.000	0.018	0.000	0.977	2.367
Level and Curvature						
$Y(1)$	0.025	0.807	0.005	0.000	0.907	2.280
$Y(5)$	0.000	0.407	0.005	0.000	0.610	1.956
$Y(10)$	0.000	0.375	0.000	0.000	0.330	1.972

Table VII
Three-Factor Model

We estimate the model

$$Y_t(\tau) = m_1(X_{1t}) + m_2(X_{2t}) + m_3(X_{3t}) + \epsilon_t$$

where $Y_t(\tau)$ denotes the spot rate of maturity 1, 5, and 10 years; X_{it} is the i-th "factor:" the Level, Slope, or Curvature of the spot curve. $m_i(X_i)$ is the sum of Fourier transforms

$$m_i(X_{it}) = \eta_{i0} + \eta_{i1} X_{it} + \eta_{i2}^2 X_{it}^2 + \sum_{j=1}^{2} [\theta_{ij} \cos(jX_{it}) + \lambda_{ij} \sin(jX_{it})]$$

The model is estimated in first differences. Under "χ^2:L" we report the p-value for the chi-square statistic of the null hypothesis of linearity in all three factors. Under "χ^2:L1-L2" we report the p-value for the null hypothesis of linearity in the first two factors (the Level and the Slope), but not the third factor (the Curvature). Under "χ^2:L1-L3" we report the p-value for the null hypothesis of linearity in the first and third factor (the Level and the Curvature), but not the second factor (the Slope). Under "χ^2:NL1-NL2" we report the p-value for the null hypothesis of non-linear two-factor (Level and Slope) model. Under "χ^2:NL1-NL3" we report the p-value for the null hypothesis of non-linear two-factor (Level and Curvature) model. In all tests, the statistics are calculated based on the White (1980) correction. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. The sample period is June 1961-December 1995.

	χ^2:L	χ^2:L1-L2	χ^2:L1-L3	χ^2:NL1-NL2	χ^2:NL1-NL3	ADJ. R^2	D.W.
$Y(1)$	0.125	0.135	0.316	0.000	0.000	0.951	2.323
$Y(5)$	0.000	0.000	0.000	0.000	0.000	0.978	2.431
$Y(10)$	0.000	0.000	0.007	0.000	0.000	0.978	2.477

Table VIII
Non-Nested Tests

We test the hypothesis of two-factor linear model (L2)

$$Y_t(\tau) = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} + \epsilon_t$$

against the non-nested hypothesis of one-factor non-linear model (NL1)

$$Y_t(\tau) = m_1(X_{1t}) + \epsilon_t$$

$Y_t(\tau)$ denotes the spot rate of maturity 1, 5, or 10 years; X_{1t} is the Level and X_{2t} is the Slope. $m_i(X_i)$ is the sum of Fourier transforms

$$m_i(X_{it}) = \eta_{i0} + \eta_{i1} X_{it} + \eta_{i2} X_{it}^2 + \sum_{j=1}^{2} [\theta_{ij} \cos(jX_{it}) + \lambda_{ij} \sin(jX_{it})]$$

Similarly, we test the hypothesis of three-factor linear model (L3)

$$Y_t(\tau) = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} + \lambda_2 X_{3t} + \epsilon_t$$

against the non-nested hypothesis of two-factor non-linear model (NL2)

$$Y_t(\tau) = m_1(X_{1t}) + m_1(X_{2t}) + \epsilon_t$$

All models are estimated in first differences. We report the coefficients and t-ratios (in parentheses) from a regression on the fitted values of the two competing models. In all tests, the statistics are calculated based on the White (1980) correction. "ADJ. R^2" is the adjusted R-square. "D.W." is the Durbin-Watson statistic. The sample period is June 1961-December 1995.

	constant	NL1	L2	NL2	L3	ADJ. R^2	D.W.
$Y(1)$	-0.000	0.051	0.956			0.809	2.046
	(-0.034)	(0.994)	(19.429)				
	0.000			0.017	0.985	0.949	2.307
	(0.048)			(0.644)	(39.395)		
$Y(5)$	-0.000	0.019	0.992			0.926	2.295
	(-0.131)	(0.760)	(54.270)				
	-0.000			0.042	0.960	0.976	2.461
	(-0.136)			(1.087)	(26.711)		
$Y(10)$	0.000	-0.004	1.001			0.975	2.387
	(0.131)	(-0.287)	(99.063)				
	0.000			0.132	0.868	0.985	2.507
	(0.134)			(2.275)	(14.064)		

Table IX
Hedging

We construct a portfolio of three bonds: a one-year, a five-year, and a ten-year zero-coupon bond. The three weights of the portfolio minimize the variance of the weekly excess return on the portfolio. In the affine case the optimal weights are constant, while in the non-linear case the weights are functions of the Level and Slope factors. In the non-linear case, we model the weight of the τ-maturity bond in the portfolio as

$$w(X_{1,t}, X_{2,t}\tau) = \phi_0(\tau) + \sum_{i=1}^{2}\sum_{j=1}^{5} \phi_{ij}(\tau)P_{ij}(X_{it})$$

where j is a j-th order Legendre polynomial. "RET_5" is the excess return on the five-year bond. "RET_{hL}" is the excess return on the hedged portfolio with constant weights. "RET_{hNL}" is the excess return on the hedged portfolio with time-varying weights. Under "L" we report the p-value of the F-statistic of the two weights being constant in both factors (affine case). Under "L1" we report the p-value of the F-statistic of the two weights being constant in the Level. Under "L2" we report the p-value of the F-statistic of the two weights being constant in the Slope. The sample period is June 1961-December 1995.

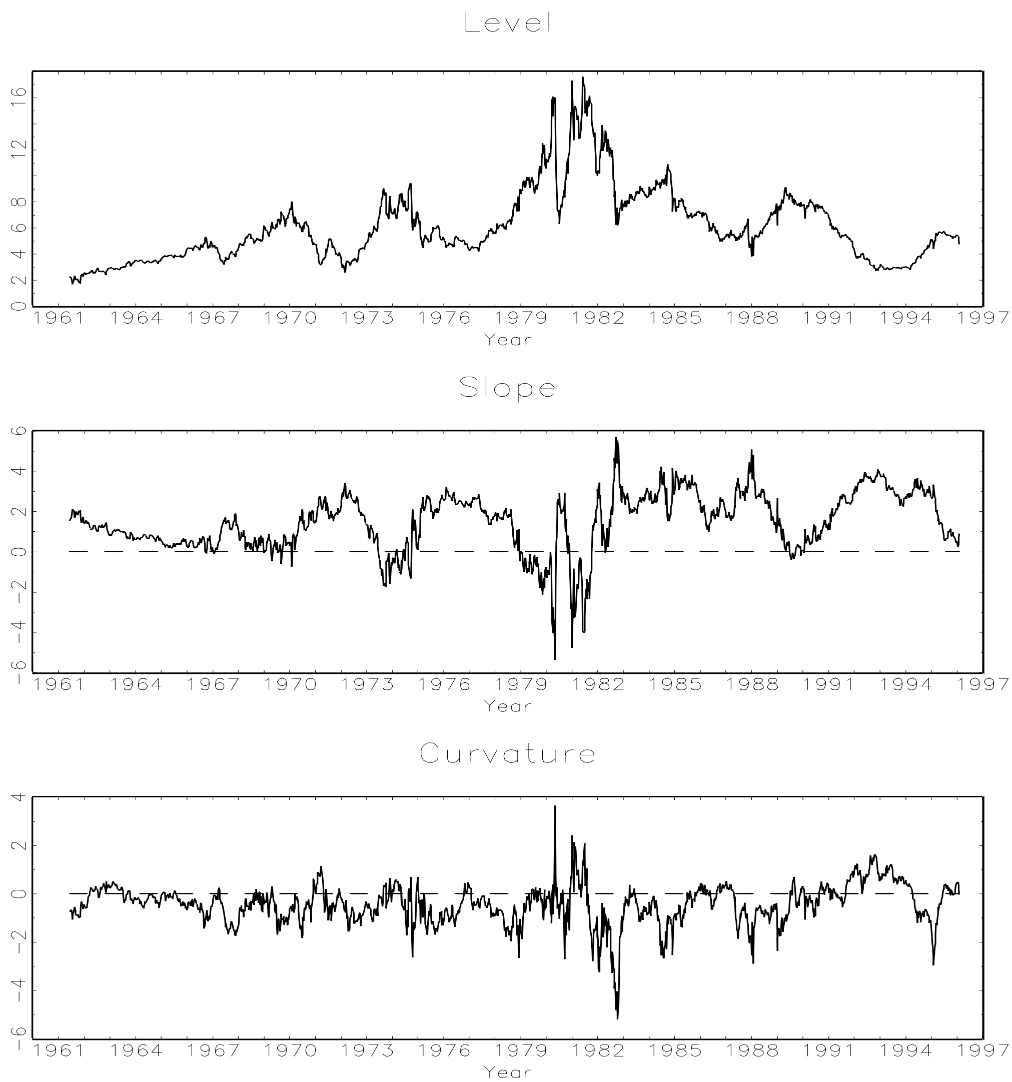
Summary Statistics

	Mean	Std.
RET_5	7.5031	6.0320
RET_{hL}	0.8254	1.0561
RET_{hNL}	0.8166	0.9816

Tests

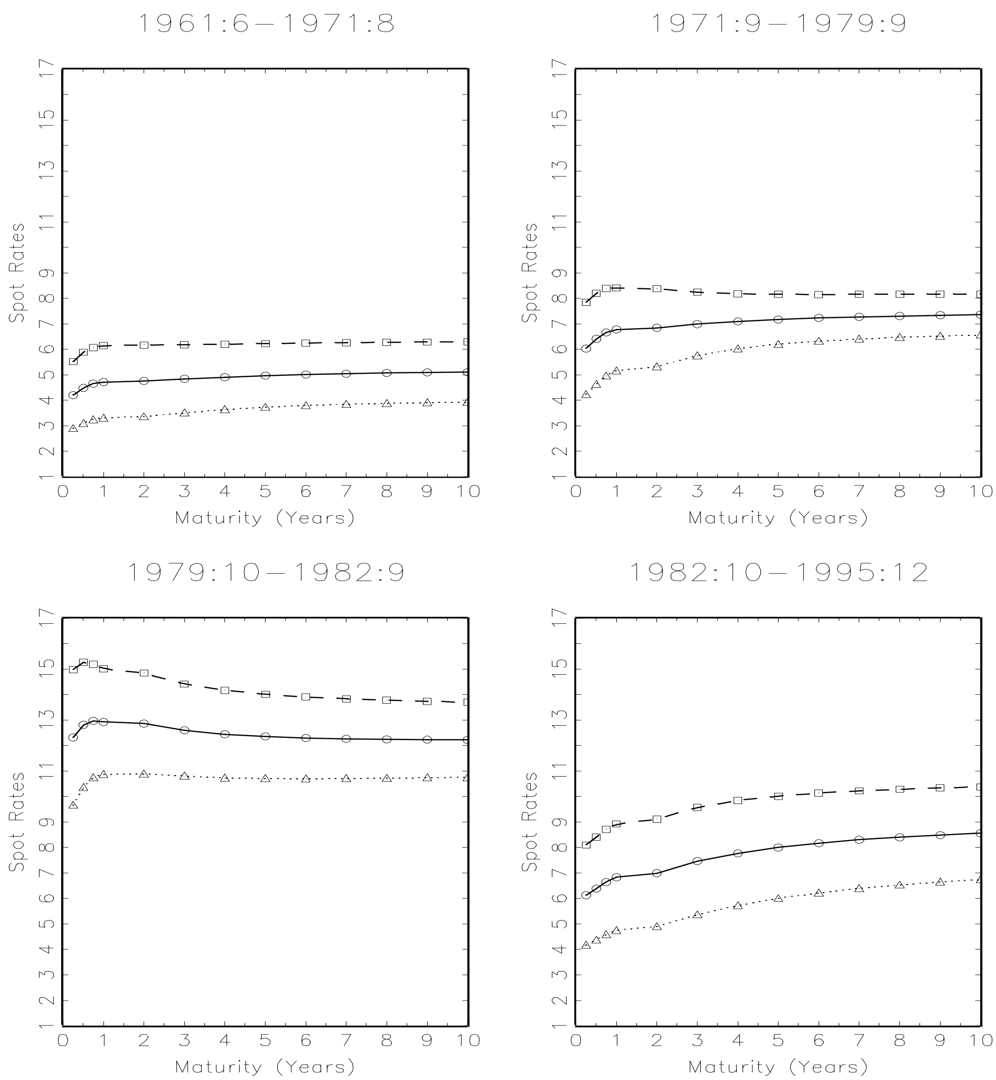
L1	L2	L
0.000	0.000	0.000

Figure 1



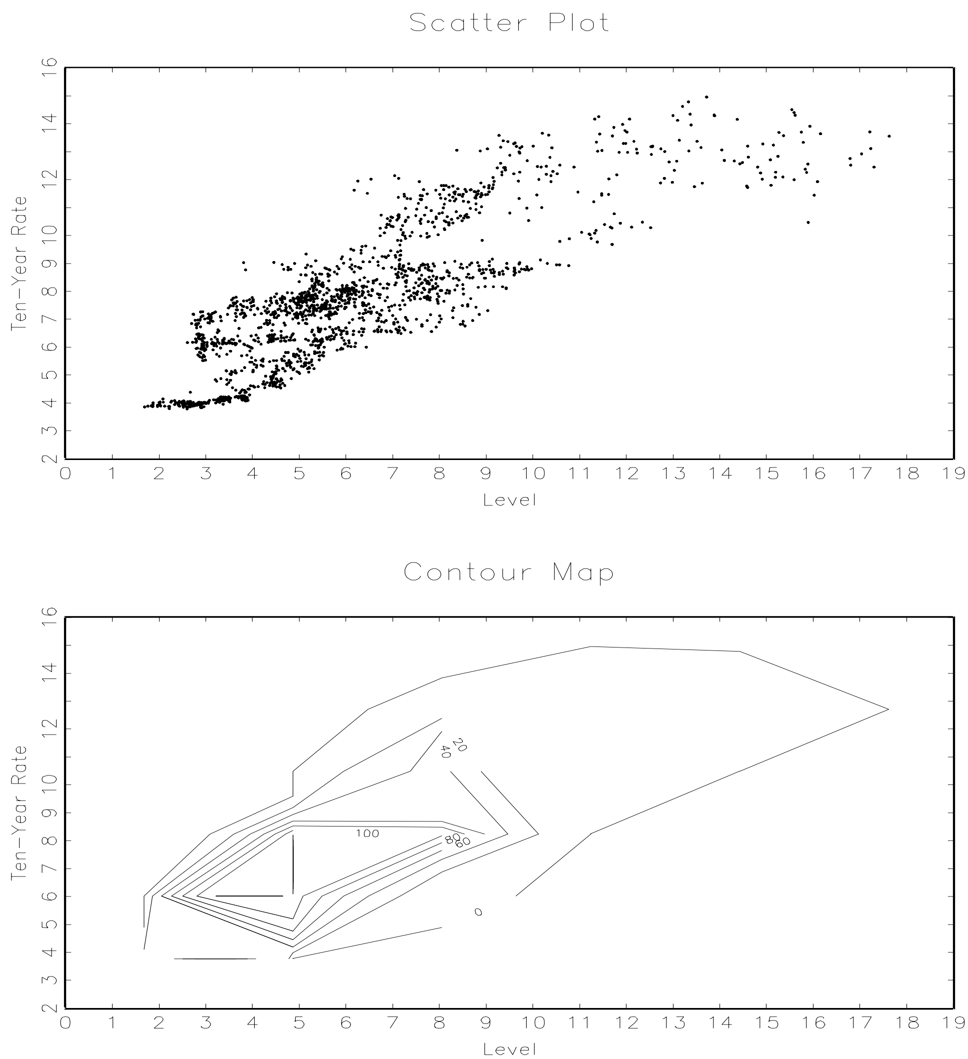
Spot-Curve Time Series. Figure 1 displays the behavior of three spot curve indicators: The Level of the spot curve, as measured by the three-month rate. The Slope of the spot curve, as measured by the spread between the eight-year and the three-month rate. The Curvature of the spot curve, as measured by the difference between two spreads: the spread between the eight- and two-year rate and the spread between the two-year and three-month rate. The sample period is June 1961-December 1995.

Figure 2



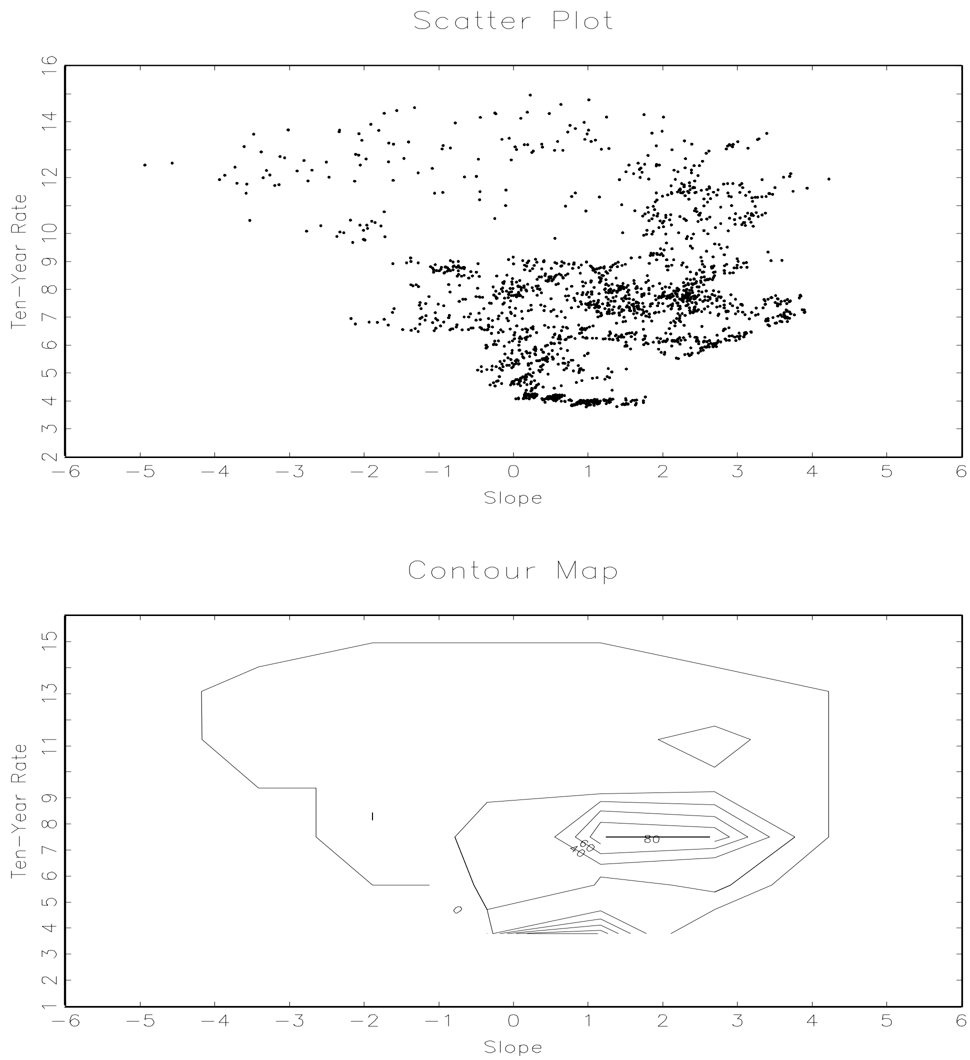
Average Spot Curves. Figure 2 displays the average spot curves for the four subperiods: June 1961-August 1971, September 1971-September 1979, October 1979-September 1982, and October 1982-December 1993. We report the mean rate together with one-standard error bands.

Figure 3



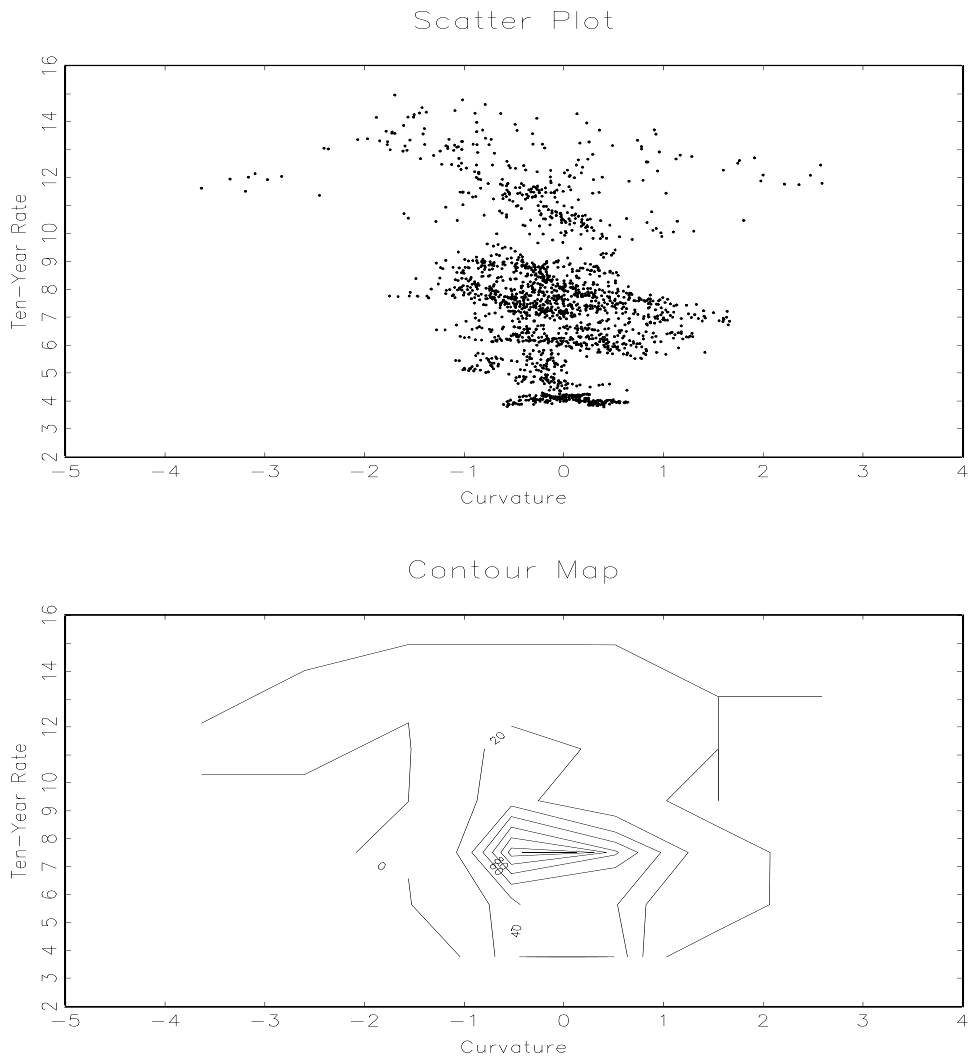
The Ten-Year Rate and the Level. Figure 3 describes the joint behavior of the ten-year rate and the Level (the three-month rate). The figures in the contour map denote the percentage of observations which lie outside that contour. The sample period is June 1961-December 1995.

Figure 4



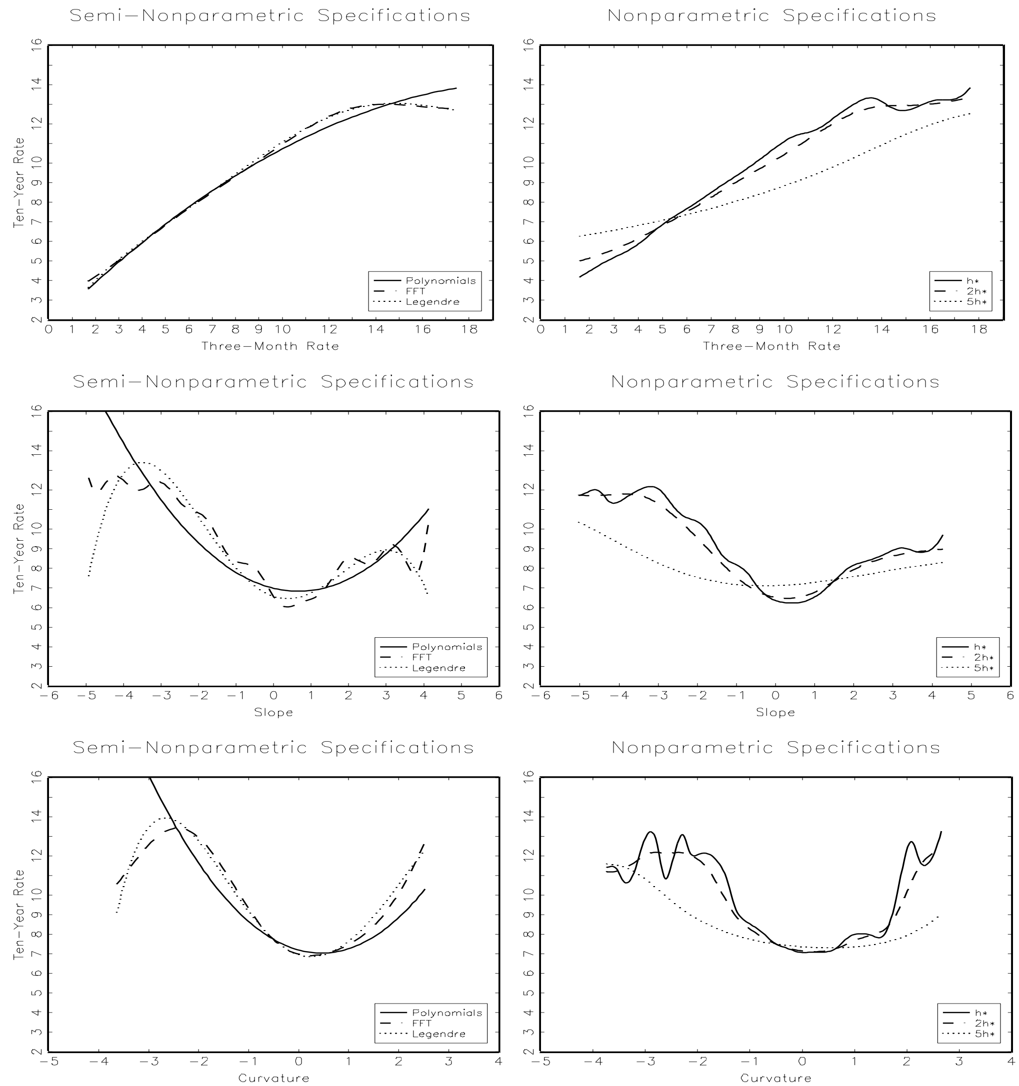
The Ten-Year Rate and the Slope. Figure 4 describes the joint behavior of the ten-year rate and the Slope (the spread between the eight-year and the three-month rate). The figures in the contour map denote the percentage of observations which lie outside that contour. The sample period is June 1961-December 1995.

Figure 5



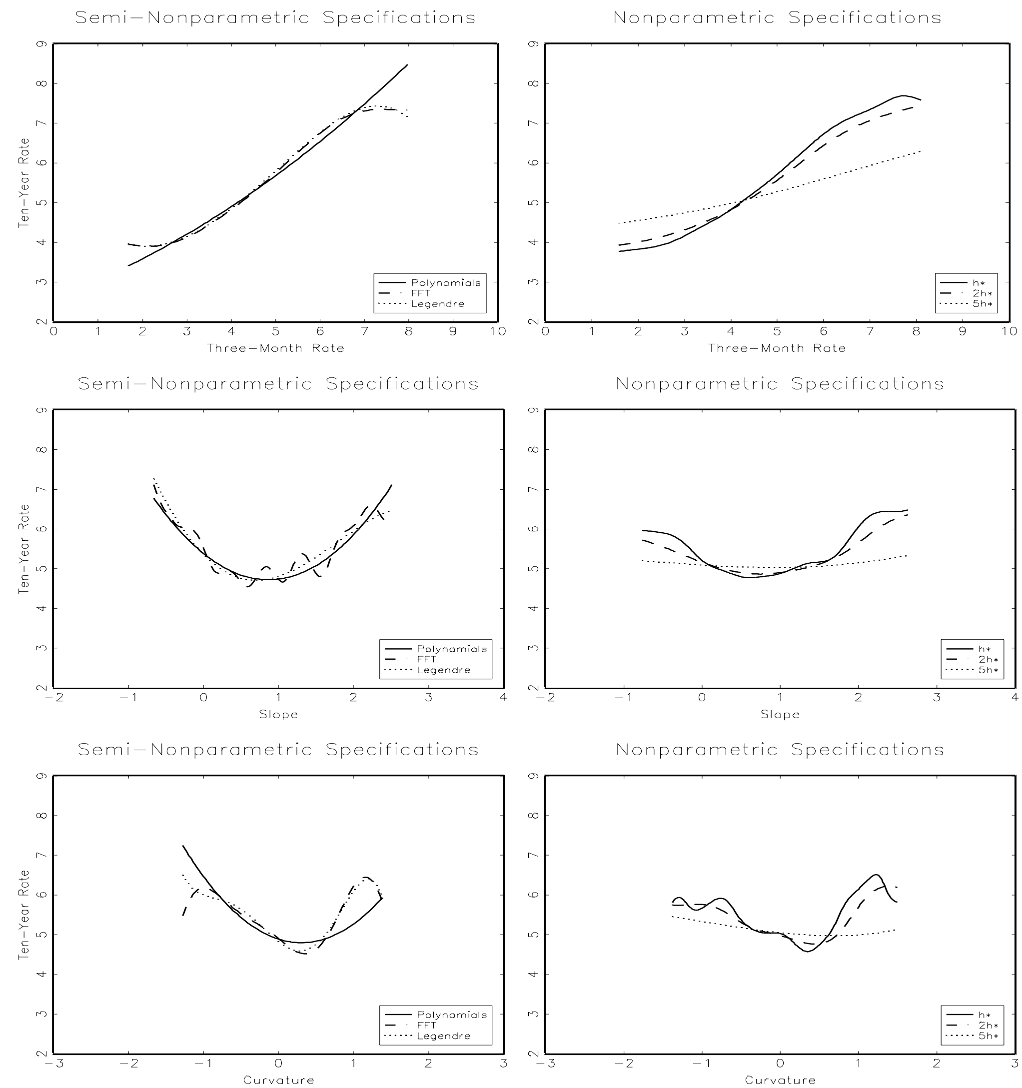
The Ten-Year Rate and the Curvature. Figure 5 describes the joint behavior of the ten-year rate and the Curvature (the difference between the eight-year-two-year and the two-year-three-month spreads). The figures in the contour map denote the percentage of observations which lie outside that contour. The sample period is June 1961-December 1995.

Figure 6



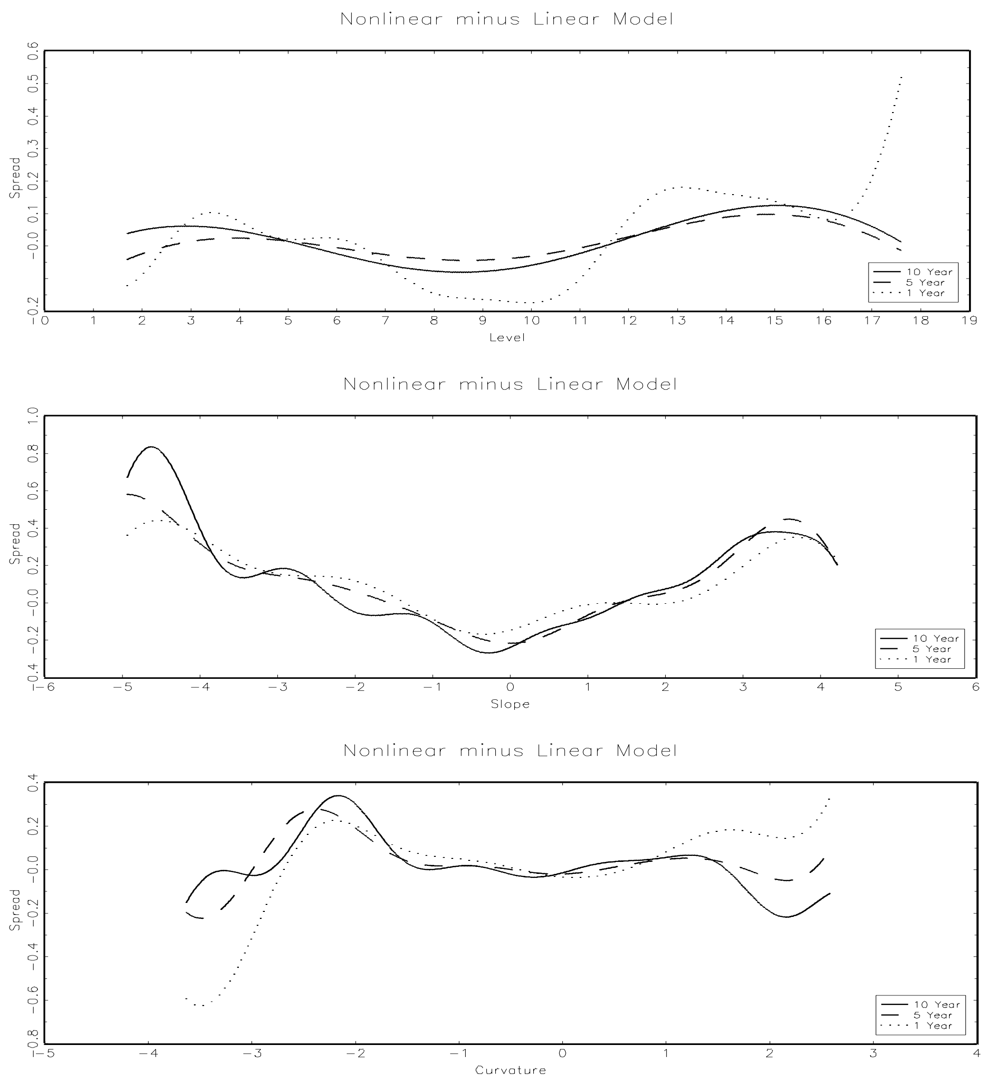
The Ten-Year Rate and the Three Factors: Whole Period. Figure 6 presents the estimated relation between the ten-year rate and the three spot-curve factors. We use semi-nonparametric specifications: i) polynomial, ii) Fourier transforms (FFT), iii) Legendre polynomials. The right-hand panel shows the estimated relations obtained using a kernel regression approach for different values of the band-width. h^* denotes the optimal band-width calculated under the assumption of no serial correlation in the residuals. The sample period is June 1961-December 1995.

Figure 7



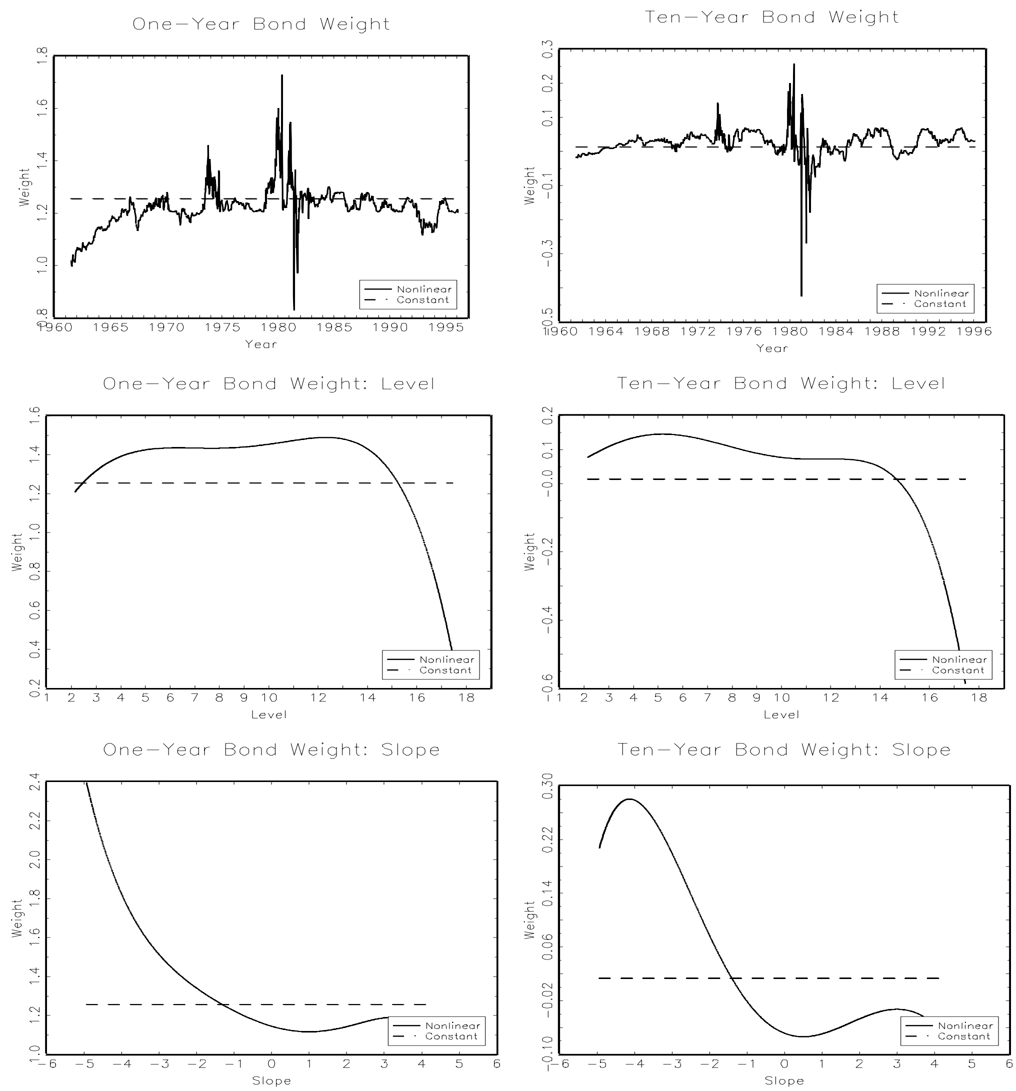
The Ten-Year Rate and the Three Factors: First Period. Figure 7 presents the estimated relation between the ten-year rate and the three spot-curve factors. We use semi-nonparametric specifications: i) polynomial, ii) Fourier transforms (FFT), iii) Legendre polynomials. The right-hand panels shows the estimated relations obtained using a kernel regression approach for different values of the band-width. h^* denotes the optimal band-width calculated under the assumption of no serial correlation in the residuals. The sample period is June 1961-August 1971.

Figure 8



Correcting for Serial Correlation. Figure 8 presents the difference between the non-linear and the linear models for the ten-year, five-year, and one-year rates. The estimation is done in first differences and the semi-nonparametric specification uses sums of Fourier transforms. The sample period is June 1961-December 1995.

Figure 9



Weights of Hedged Portfolio. Figure 9 presents the behavior of the two weights of the minimum-variance portfolio. The three panels on the left present the weight of the one-year bond. The three panels on the right present the weight of the ten-year bond. The top two panels present the time-series behavior of the two weights. The intermediate two panels present the behavior of the two weights against the realizations of the Level. The bottom two panels present the behavior of the two weights against the realizations of the Slope.