A Parametric Nonlinear Model of Term Structure Dynamics

Dong-Hyun Ahn
Bin Gao
University of North Carolina

Recent nonparametric estimation studies pioneered by Aït-Sahalia document that the diffusion of the short rate is similar to the parametric function, $r^{1.5}$, estimated by Chan et al., whereas the drift is substantially nonlinear in the short rate. These empirical properties call into question the efficacy of the existing *affine* term structure models and beg for alternative models which admit the observed behavior. This article presents such a model. Our model delivers closed-form solutions for bond prices and a concave relationship between the interest rate and the yields. We show that in empirical analyses, our model outperforms the one-factor affine models in both time-series as well as cross-sectional tests.

There has been a surge in studies of the term structure of interest rates. Since the introduction of continuous-time mathematics two decades ago, starting with Vasicek (1977), Dothan (1978) and Courtadon (1982), many studies have developed one-factor term structure models using different interest rate specifications.[1] Among term structure models, affine class models, where yields are an affine function of the interest rate, are state of the art. The Vasicek (1977) and Cox, Ingersoll, and Ross (1985) (hereafter CIR) models belong to this class. The idea that the affine model is typically associated with affine drift and squared diffusion is foreshadowed in CIR and is formally proven in Duffie and Kan (1996) (hereafter DK). DK also present the most general affine class model which nests the Vasicek (1985) and CIR models. These affine class models share a number of desirable features. The diffusion processes of the interest rate specified in these models provide closed-form expressions for transition and marginal densities of the interest rate, and also bond prices. As a result, these models are analytically tractable and easy to implement, which is a major strength. However, empirical evidence on the

---

This article has benefited considerably from the comments and suggestions of the editor, Bernard Dumas, and an anonymous referee. We thank Jacob Boudoukh, Jennifer Conrad, Robert Dittmar, Silverio Foresi, Eric Ghysels, Jay Huang, Ravi Jagannathan, Paisan Limratanamongkol, Matthew Richardson, Anil Shivdasani, Steve Slezak, Robert Whitelaw, Marc Zenner, seminar participants at Pennsylvania State University, University of North Carolina and the 1998 Western Finance Association Meetings and discussant, Gregory Duffie, and in particular, Yacine Aït-Sahalia for their helpful comments and suggestions. However, all remaining errors are ours. Address correspondence to Dong-Hyun Ahn, Finance Department, Kenan-Flagler Business School, University of North Carolina, CB 3490, McColl Bldg., Chapel Hill, NC 27599-3490.

[1] Along with one-factor models, there have also been multifactor models. See, for example, Brennan and Schwartz (1979), Schaefer and Schwartz (1987), Chen and Scott (1992), and Duffie and Kan (1996).

The Review of Financial Studies Special 1999 Vol. 12, No. 4, pp. 721–762

diffusion process of the interest rate is at odds with the affine class models. In particular, the nonparametric studies pioneered by Aït-Sahalia (1996a,b) evidence strong nonlinearities in both the drift and diffusion functions. The results call into question the validity of the affine class models, and beg for alternative term structure models which admit the observed dynamics of the interest rate. This article develops a new term structure model that is based on the stochastic process of the interest rate which admits the empirical observed behavior.

A useful first step to investigate the empirical validity of alternative term structure models is to examine the diffusion process of the short rate. Chan et al. (1992) investigate this issue by estimating a Euler approximation of the constant elasticity of volatility model (CEV model) for the short rate of the form

$$dr(t) = [\beta_0 + \beta_1 r(t)]dt + \sigma r(t)^{\beta_2} dz(t). \tag{1}$$

This specification nests the Vasicek (1977) model when $\beta_2 = 0$, the CIR model when $\beta_2 = 0.5$, and Brennan and Schwartz (1979) when $\beta_2 = 1$. They find that even though the estimation of the drift (in terms of β_0 and β_1) is problematic, the estimation of the volatility is quite precise and a result of $\beta_2 = 1.5$ rejects all popular models with $\beta_2 \leq 1$.

Campbell, Lo, and MacKinlay (1996) also support this result, showing that the heteroscedasticity of the short rate can be markedly reduced as one increases β_2 from 0 to 1.5. In the words of Campbell, Lo, and Mackinlay (1996, p. 451):

> *...Second, single-factor affine-yield models require that $\gamma = 0$ or 0.5...*[2]
> *The estimated value of 1.5 takes one outside the tractable class of affine-yield models and forces one to solve the term-structure model numerically.*
> *There is as yet no consensus about how to resolve these problems...*

Previous empirical findings can be summarized as follows. First, the heteroscedastic form of the conditional volatility is precisely estimated, and its functional form is $r^{1.5}$, a convex function of the short rate. Second, the estimation of the drift is very imprecise, and its form is controversial. The problems associated with these studies are twofold. First, the estimation of the diffusion process relies critically on the parametric specification for the process of the unrestricted model, such as Equation (1), against which extant parametric models are tested. Second, the Euler approximation of a diffusion process is subject to discretization errors.

Recently, in order to overcome these undesirable features of parametric estimation of the diffusion process of the interest rate, nonparametric estimation, a model-free approach pioneered by Aït-Sahalia (1996a,b), has

[2] γ is equivalent to β_2 in Equation (1).

been introduced to address these issues and add to our understanding of the time-series evolution of the short rate. Specifically, Aït-Sahalia (1996a) proposes a nonparametric method to estimate the diffusion, given a parametric specification for the drift. Further, to infer the implied diffusion process, he (1996b) nonparametrically estimates the marginal density function of the short rate. Since the stochastic process of the short rate is fully recovered from the data without any parametric restrictions, the specification test of alternative parametric models is more powerful. Conley et al. (1997) also use the constant volatility elasticity parameterization of the diffusion coefficient to interpret interest rate data in the presence of subordination while permitting the drift to be nonlinear. For this they provide two new methods for estimation, which use the marginal distribution of the interest rate. Finally, Stanton (1997) extends this approach to develop a procedure for estimating both the drift and the diffusion nonparametrically from observed data. The benefit of these studies is that they avoid the possible misspecification of a parametric model, and more importantly, do away with discrete approximation errors.

The estimated volatility in these studies is similar to that estimated by Chan et al. (1992).[3] However, the results of these studies show that the nonparametrically determined drift has substantial nonlinearity. Although for low and medium interest rate levels, there is little mean reversion, similar to the finding of Chan et al. (1992), the degree of mean reversion increases dramatically as interest rates climb above the midrange. In particular, Aït-Sahalia (1996b) rejects even the model of Equation (1) as well as all the popular diffusion models nested in it, and states that the linearity of the drift imposed in the literature appears to be the main source of misspecification.

These empirical studies strongly suggest that the dynamics of the interest rate specified in the affine term structure models conflict with the observed interest rate behavior. Thus the interesting question that remains is whether we can develop a parsimonious term structure model that captures both the empirical findings of a nonlinear drift and a diffusion term proportional to $r^{1.5}$. If so, what testable implications does this new model offer to differentiate itself from existing affine-class term structure models?

We address these questions by proposing an alternative single-factor term structure model that is consistent with the dynamics of the interest rate process documented in the nonparametric literature. The specification for the diffusion is $r^{1.5}$, as suggested in empirical studies. The novel feature is the drift, which is specified as a quadratic function instead of a linear one, so that it exhibits substantial nonlinear mean-reverting behavior when the interest rate is above its long-run mean. The model is just as parsimonious

[3] Aït-Sahalia's (1996b) β_2 estimate in his general parametric model is close to 1, while the estimate of β_2 in Stanton's (1997) nonparametric model is about 1.5. Conley et al. (1997) find the estimate of β_2 between 1.5 and 2.

as the models of CIR (1985) and Vasicek (1977) with only three structural parameters. In addition, the model provides a closed-form solution for bond prices, resolving the issue raised by Campbell, Lo, and MacKinlay (1996).

Using the monthly data of McCulloch and Kwon (1993), we empirically test our model and compare it to affine term structure models including the DK model, which is the most general affine-class model. We find that our model performs better than the affine model in explaining the stochastic process of the short rate; the R^2 of our model is substantially greater than that of the DK model for both the drift and the diffusion. Further, our model is shown to capture the nonlinearities in the drift and the diffusion detected in the nonparametric literature even though we do not force the parameters to reflect these shapes. This provides evidence on the superior performance of the model in the time-series dimension or under the true probability measure.

By construction, the model endows a nonaffine relationship between the interest rate and yields. The only extant non-affine-class models are a class of squared autoregressive models such as Beaglehole and Tenney (1992), Constantinides (1992), and Ahn (1997). However, these models do not share the strength of single-factor affine-class models since the interest rate is not a sufficient statistic. This feature makes the implementation of these models quite demanding. In contrast, the interest rate is still a sufficient statistic in our model. Thus its empirical test and implementation does not demand extra effort beyond what is needed in the affine-class models.

We find that the shape of the nonlinearity generated by our model depends on the parameter values. First, in order to investigate the probable nonlinearity contained in the data, we perform generic nonlinear regression analyses such as Taylor expansion series, flexible Fourier form expansion series, and nonparametric kernel regression. We find these alternative tests posit a strong concave nonlinearity in the relationship. Our nonlinear model is able to capture this nonlinear relationship very well, and reduces the pricing errors of the DK model by about 15%. Therefore we conclude that our model performs better than the affine model in a cross-sectional dimension under the risk-neutral probability measure. Overall our model can explain some peculiar nonlinear features of the dynamics of the interest rate and the relationship between the interest rate and yields which are not considered by the existing parametric term structure models.

The article is organized as follows. In Section 1, we suggest an alternative stochastic process of the interest rate and analyze the characteristics of the implied interest rate distribution. Section 2 shows that our model outperforms other single factor models in capturing the dynamics of the interest rate documented in the nonparametric literature. Section 3 proposes a specification for the market price of risk, derives closed-form solutions for bond prices, and analyzes the behavior of yields to maturity. In Section 4 we present empirical evidence of a nonlinear relationship between the short rate and the yields to maturity. In addition, we show that our model implies the types of nonlinearity presented in the data. We also document empirical support for our model specification of the market price of risk. Section 5 concludes and discusses possible extensions.

1. The Stochastic Process of the Interest Rate

The economy is represented by the augmented probability space $(\Omega, \mathcal{F}, F, P)$, where $F = \{\mathcal{F}_t\}_{0 \leq t \leq \mathcal{T}}$. First, we define the stochastic discount factor $M(t, T)$,

$$S(t) = E_t^P\left[M(t, T)S(T)\right], \tag{2}$$

where $S(t)$ is the price of an asset, $S(t, \omega) : [0, \infty) \times \Omega \rightarrow R^+$. As is well known, alternative asset pricing models differ in their description of this stochastic discount factor.

As shown by Harrison and Kreps (1979) and Harrison and Pliska (1981), under the assumption of a complete market, there is a unique probability measure Q under which all money market scaled security prices follow a martingale:

$$\frac{S(t)}{B(t)} = E_t^Q\left[\frac{S(T)}{B(T)}\right] = E_t^P\left[\frac{dQ(T)}{dP(T)}\frac{S(T)}{B(T)}\right], \tag{3}$$

where the money market account is represented as

$$B(t) = \exp\left(\int_0^t r(s)\,ds\right),$$

and $r(t)$ is the *locally* riskless interest rate at time t. $\frac{dQ}{dP}$ is called the Radon–Nikodym derivative, which can be described as

$$\Psi(t, T) \stackrel{\text{def}}{=} \frac{dQ(T)}{dP(T)} = \exp\left(\int_t^T \eta(s)\,dw_q(s) - \frac{1}{2}\int_t^T \eta(s)^2\,ds\right),$$

where $\eta(t)$ is a vector of diffusion parameters in the stochastic discount factor and $w_q(t)$ is a vector of independent standard Brownian motions.

Rewrite Equation (3) as

$$S(t) = E_t^P\left[\frac{dQ}{dP}\frac{B(t)}{B(T)}S(T)\right]. \tag{4}$$

Given the unique martingale measure, the equivalence of Equations (2) and (4) relates the stochastic discount factor and Radon–Nikodym derivative:

$$\begin{aligned} M(t,T) &= \frac{B(t)}{B(T)}\Psi(t,T) \\ &= \exp\left(\int_t^T \eta(s)\,dw_q(s) - \int_t^T \left(r(s) + \frac{1}{2}\eta(s)^2\right) ds\right). \end{aligned}$$

Therefore the stochastic discount factor can be rewritten as the product of the inverse of the value of the money market account and the Radon–Nikodym derivative. Different models vary in their descriptions of $r(t)$ and $\eta(t)$, which result in different stochastic discount factors. From Ito's lemma, we can write the stochastic differential equation of the stochastic discount factor,

$$\frac{dM(0,t)}{M(0,t)} = -r(t)dt + \eta(t)dw_q(t). \tag{5}$$

We first specify the dynamics of the interest rate $r(t)$ in Assumption 1. The specification of $\eta(t)$ is discussed later in Section 4.

Assumption 1. *The development of the interest rate $r(t)$ is given by the stochastic differential equation (SDE)*

$$dr(t) = \kappa\,(\theta - r(t))\,r(t)dt + \sigma r(t)^{1.5}dw(t), \tag{6}$$

where $w(t)$ is a standard Brownian motion defined under measure P. [4]

Several features of the SDE in Equation (6) are noteworthy. First, the diffusion in Equation (6), $\sigma r(t)^{1.5}$, is the same as that estimated parametrically by Chan et al. (1992), and reflects the nonparametric estimation result of Aït-Sahalia (1996b) and Stanton (1997). A rapidly increasing diffusion function would generate a dense population of large values, which can partially explain the slowly decaying unconditional distribution of the interest rate documented in Aït-Sahalia (1996b). Second, of particular note, our drift term is different from that in Chan et al. (1992), but it is consistent with the empirical findings of Aït-Sahalia (1996b) and Stanton (1997). In particular, Aït-Sahalia (1996a,b) broadens our understanding of the dynamics of interest rates by addressing the fact that the linearity of the drift imposed in the affine-class models or Chan et al. (1992) appears to be the main source of misspecification in these models. He argues that given the diffusion, a dramatic decline in the drift at high interest rates is required to prevent the interest rate from exploding toward infinity, which is also con-

[4] Marsh and Rosenfeld (1983) study the time-series property of instantaneous interest rate based on a process $dr = (Ar^{-(1-\beta)} + Br)dt + \sigma r^{\beta/2}dw$, with $0 \le \beta \le 2$. Our process corresponds to $\beta = 3$, which is not studied in their model. Even though it seems innocuous to nest our model in their model, the boundary conditions should be redefined, and the density of the interest rate is somewhat different.

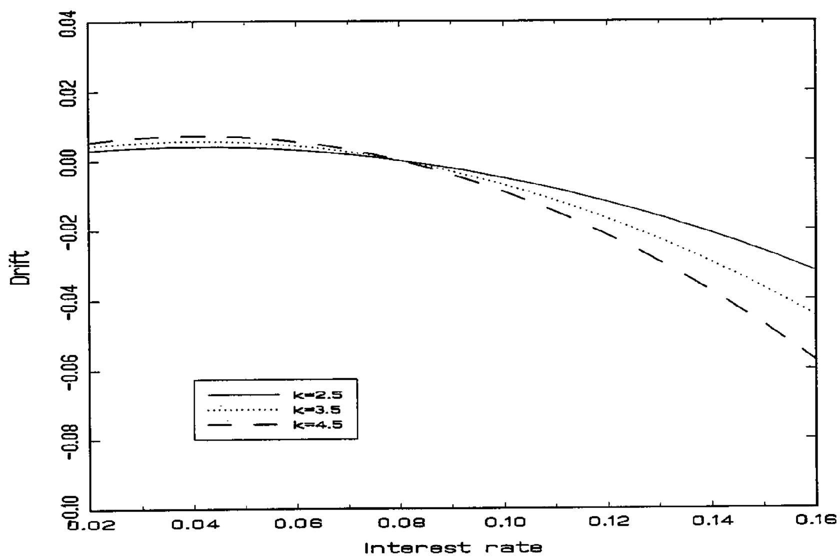


Figure 1
The relationship between κ and the drift of the model where $\theta = 0.08$

firmed by Stanton (1997). In other words, there is a need for balancing the effects of a stronger mean-reverting drift with a higher diffusion in order to make the process "more able" to pull back the interest rate toward its long-term mean from a high level even in the presence of a higher variation [see Aït-Sahalia (1996a, p. 545, and 1996b)].

The drift suggested in Equation (6) reflects the feature documented in the nonparametric studies since it is a quadratic function of the interest rate. To see this, let us investigate the economic implication of θ and κ. First, θ is the *threshold level* of the interest rate at which the drift is zero. Mathematically, θ is the intercept of the x-axis in the interest rate-drift domain. Under the assumption of $\kappa > 0$ (which is a necessary condition for the stationarity of the interest rate process, as demonstrated below), if the interest rate is below θ, the drift is positive, implying that the interest rate reverts to the normal range. Once the interest rate climbs above θ, the drift becomes negative for the same reason. Therefore θ determines the range of the interest rates wherein the drift is positive; the larger (smaller) θ is, the wider (narrower) is the range of positive drift.

κ also has an important meaning in the determination of drift. As illustrated in Figure 1, κ determines the curvature of the drift. For larger values of κ, the curvature of the drift becomes steeper, and vice versa. The curvature, which is equal to the second derivative of the drift with respect to the interest rate, $|-2\kappa|$, measures the degree of nonlinearity in the drift. In addition, a higher curvature means a larger absolute value of the drift at a given level of interest rate, as is shown in Figure 1, so that mean reversion will be stronger. However, unlike the linear drift models, the *mean-reversion speed* in our model is a linear function of the interest rate $\kappa r(t)$, with the mean-reversion speed increasing with the interest rate. This is a novel feature of our model which generates the balancing effects of a stronger mean reversion with a higher interest rate. Initially the drift climbs until the interest rate reaches $\theta/2$. At this point the drift hits its maximum value, $\kappa\theta^2/4$. Once the interest rate passes this point, the drift begins to decline, reaching zero when the interest rate is θ. If the interest rate climbs above θ, then the drift becomes negative, pulling the high interest rate back toward its normal range. The further above θ, the faster the mean reversion. In contrast, linear drift models maintain the same mean-reversion speed, $-\beta_1$, regardless of the level of the interest rate.[5]

In addition, the specification of the interest rate dynamics in Equation (6) is interesting in considering the relationship between inflation and the nominal interest rate. As an illustration, consider a Fisher effect. In its extreme version, the real interest rate remains constant, so the time-varying movement of the nominal interest rate stems from changes in the expected inflation rate. Informal evidence suggests that the relative (percentage) variance of the expected inflation rate increases as its level increases [see CIR (1985)]. Due to the one-to-one correspondence between the expected inflation rate and the nominal interest rate, the nominal interest rate can have the dynamics described above. This is only an illustrative example that can support the above specification of the stochastic differential equation. In a more general setup where the real interest rate and the expected inflation rate are time varying, the variable representing the combination of the two may be assumed to follow the above stochastic differential equation.

1.1 The distribution of the interest rate

In this section we examine the conditions under which the stochastic process in Equation (6) is stationary. We also derive the distributions implied by the stochastic process proposed in Assumption 1.

An examination of the boundary condition shows that r will be stationary and zero is not attainable if κ and $\sigma > 0$ (see Appendix A). Intuitively, if κ is negative, the leading term in the drift, $-\kappa r(t)^2$, is positive, and the drift term is U-shaped rather than an inverted U shape. Under this condition and given diffusion of $\sigma r(t)^{1.5}$, the process explodes.[6]

---

[5] Aït-Sahalia (1996b) and Conley et al. (1997) provide evidence of strong mean reversion when the interest rate is below the normal range (about 2%), which implies a U-shaped function of the mean-reversion speed. Our model fails to generate this feature. In contrast, Stanton (1997) documents no evidence of strong mean reversion at low levels of the interest rate, which is similar to our specification.

[6] We thank the referee for pointing this out and for providing us with an extensive analysis of this point.

The probability density of the interest rate at time s, conditional on its value at the current time t, is given by[7]

$$f(r(s), s|r(t), t) = \alpha^{-1} e^{-u-v} v^{q/2+2} u^{-q/2} I_q\left[2(uv)^{1/2}\right],$$

where

$$\begin{aligned} \alpha &= \frac{2\kappa\theta}{\sigma^2\left(1 - e^{-\kappa\theta(s-t)}\right)} \\ u &= \frac{\alpha}{r(t)} e^{-\kappa\theta(s-t)} \\ v &= \frac{\alpha}{r(s)} \\ q &= \frac{2\left(\kappa + \sigma^2\right)}{\sigma^2} - 1. \end{aligned}$$

$I_q(\cdot)$ is the modified Bessel function of the first kind of order q,

$$I_q\left[2(uv)^{1/2}\right] = (uv)^{q/2} \sum_{p=0}^{\infty} \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)}.$$

Straightforward calculation yields the νth conditional moments for the interest rate following the process of Equation (6):

$$E\left[r(s)^\nu|r(t)\right] = \alpha^\nu e^{-u} \frac{\Gamma(1+q-\nu)}{\Gamma(1+q)} M(1+q-\nu, 1+q, u),$$

where $M(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function (or Kummer's function) and can be represented in the Pochhammer symbols,

$$M(a, b, x) = \sum_{n=0}^{\infty} \frac{(a)_n x^n}{(b)_n n!} = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{xz} z^{a-1} (1-z)^{b-a-1}\, dz,$$

and

$$(a)_n \equiv a(a+1)(a+2)\cdots(a+n-1), \quad (a)_0 \equiv 1.$$

In particular, the expected value and variance of $r(s)$ conditional on $r(t)$ are

$$\begin{aligned} E[r(s)|r(t)] &= \frac{\alpha e^{-u}}{q} M(q, 1+q, u) = \alpha e^{-u} (-u)^{-q} \gamma(q, -u) \\ \mathrm{var}[r(s)|r(t)] &= \frac{\alpha^2 e^{-u}}{q} \left[\frac{M(q-1, 1+q, u)}{q-1} - \frac{e^{-u} M(q, 1+q, u)^2}{q}\right], \end{aligned}$$

[7] The derivation is shown in Appendix B.

where $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function. Since the interest rate displays mean reversion, then as s becomes large, its distribution is well defined. The steady-state density function is[8]

$$\pi(r) = \frac{1}{\Gamma(2+\frac{2\kappa}{\sigma^2})}\left(\frac{2\kappa\theta}{\sigma^2}\right)^{2+\frac{2\kappa}{\sigma^2}} r^{-\left(3+\frac{2\kappa}{\sigma^2}\right)} \exp\left(-\frac{2\kappa\theta}{\sigma^2 r}\right).$$

The unconditional νth moments of the interest rate are

$$E\left[r^{\nu}\right] = \frac{\left(\frac{2\kappa\theta}{\sigma^2}\right)^{\nu}}{\Gamma\left(2+\frac{2\kappa}{\sigma^2}\right)}\Gamma\left(2+\frac{2\kappa}{\sigma^2}-\nu\right).$$

In particular, the steady-state mean and variance are

$$\begin{aligned} E(r) &= \frac{2\kappa\theta}{2\kappa+\sigma^2} \\ \mathrm{var}(r) &= \frac{2\kappa\theta^2\sigma^2}{\left(2\kappa+\sigma^2\right)^2} \end{aligned}$$

Of interest, even the steady-state mean is determined by all three structural parameters. Given $\kappa > 0$, the steady-state mean is smaller than the threshold level of the interest rate, θ because the mean-reversion speed is an increasing function of the interest rate. Since it takes more time to reach θ from the lower interest rate than from the higher interest rate, the frequency of these levels of the interest rate is greater, which shifts the unconditional mean below θ. In summary, the introduction of the additional nonlinearity in the drift in a fashion consistent with the nonparametric evidence results in a closed-form expression for the transition and marginal densities unlike the CEV process in Equation (1), and resolves the issue raised by Campbell, Lo, and MacKinlay (1996).

2. The Empirical Performance of the Stochastic Process

In this section we estimate the diffusion process of our model to evaluate its ability to capture the actual behavior of the interest rate. Since Equation (6) is the stochastic differential equation under the true probability measure, this test provides a measure of the relative performance of our model under the P measure. Further, we compare the performance of our model to linear drift models including the affine-class models. We conjecture that our model should outperform the affine-class models since our specification arises from the extant empirical literature rather than from “thin air.” However, we investigate whether our model can produce statistically significant

[8] The derivation is in Appendix B.

Table 1
Alternative models and parameter restrictions

Model	Specification	Restrictions	j	Boundary conditions
Vasicek	$dr = \kappa(\theta - r)dt + \sigma dw$	$\alpha_3 = \alpha_5 = \alpha_6 = 0$	3	$\kappa > 0$ (0 is attainable)
CIR	$dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dw$	$\alpha_3 = \alpha_4 = \alpha_6 = 0$	3	$\kappa > 0$ and $2\kappa\theta \geq \sigma^2$
DK	$dr = \kappa(\theta - r)dt + \sqrt{\alpha + \beta r}dw$	$\alpha_3 = \alpha_6 = 0$	2	$\kappa > 0$ and $2\kappa(\theta - r^*) \geq \beta$ ($r^* = -\alpha/\beta \geq 0$ is the lower bound)
CKLS	$dr = \kappa(\theta - r)dt + \sigma r^{1.5}dw$	$\alpha_3 = \alpha_4 = \alpha_5 = 0$	3	$\kappa > 0$ and $\theta > 0$
AG	$dr = \kappa(\theta - r)rdt + \sigma r^{1.5}dw$	$\alpha_1 = \alpha_4 = \alpha_5 = 0$	3	$\kappa > 0$ and $\theta > 0$

The table documents the specifications of and the corresponding restrictions imposed on alternative models of the interest rate which are nested within the unrestricted model:

$$dr = (\alpha_1 + \alpha_2 r + \alpha_3 r^2)dt + \sqrt{\alpha_4 + \alpha_5 r + \alpha_6 r^3}dw.$$

The following natural restrictions have to be imposed on the parameter values:

(i) $\alpha_4 \geq 0$ (or $\alpha_5 > 0$ if $\alpha_4 = 0$)
(ii) $\alpha_6 > 0$ (or $\alpha_5 > 0$ if $\alpha_6 = 0$)
(iii) $\alpha_3 \leq 0$ (or $\alpha_2 < 0$ if $\alpha_3 = 0$)
(iv) $\alpha_1 \geq 0$ (or $\alpha_2 > 0$ if $\alpha_1 = 0$), $\alpha_4 = 0$ and $2\alpha_1 \geq \alpha_5 \geq 0$

Equations (i) and (ii) are necessary for the volatility to be positive in the neighborhood of the zero and infinity boundaries, respectively. Equations (iii) and (iv) ensure that infinity and zero are unattainable.[9] Similar boundary conditions for the nested models are described in the last column.[10] The index j denotes the number of overidentification restrictions in the GMM estimation nested within the unrestricted model. AG stands for the model of this article.

improvements, and if it does, in what aspects of the dynamics of the interest rate. For this purpose we consider two benchmarks, one which is parametric and another which is nonparametric.

2.1 The parametric benchmark

As in Marsh and Rosenfeld (1983) and Chan et al. (1992), we set up a parametric model in which a variety of affine term structure models as well as our model are nested and their performance is benchmarked. The time series of the interest rate that we consider is the following stochastic differential equation:

$$dr(t) = (\alpha_1 + \alpha_2 r(t) + \alpha_3 r(t)^2)dt + \sqrt{\alpha_4 + \alpha_5 r(t) + \alpha_6 r(t)^3}dw(t). \quad (7)$$

These dynamics are different from and more general than those in Chan et al. (1992). The stochastic differential equation defines a broad class of interest rate processes which reduces to many well-known affine-class models by placing appropriate restrictions on the parameters. We focus on three different affine class models: Vasicek (1977), CIR (1985), and DK (1996), and a linear-drift nonaffine-class model of Chan et al. (1992) (hereafter CKLS). The specifications of these models and their corresponding parameter restrictions are summarized in Table 1.

In line with CKLS (1992), we first estimate the unrestricted model of Equation (7). Define $\underline{\alpha}$ to be the parameter vector with elements $(\alpha_1, \ldots, \alpha_6)$. Given $\epsilon(t+1) = r(t+1) - r(t) - (\alpha_1 + \alpha_2 r(t) + \alpha_3 r(t)^2)$, let the vector

$g(t, \underline{\alpha})$ be

$$\underline{g}(t, \underline{\alpha}) = \begin{bmatrix} \epsilon(t+1) \otimes \underline{z_1}(t) \\ \epsilon(t+1)^2 \otimes \underline{z_2}(t) \end{bmatrix},$$

where $\underline{z_1}(t)^\top = (1, \ r(t), \ r(t)^2)$ and $\underline{z_2}(t)^\top = (1, \ r(t), \ r(t)^3)$ with corresponding moment conditions

$$\underline{\bar{g}}_T(\underline{\alpha}) = \frac{1}{T} \sum_{t=1}^{T} \underline{g}(t, \underline{\alpha}).$$

Then we choose parameter estimates to minimize the quadratic form,

$$J_T^U(\underline{\alpha}) = T \underline{\bar{g}}_T(\underline{\alpha})^\top W_T(\underline{\alpha}) \underline{\bar{g}}_T(\underline{\alpha}),$$

where $W_T(\underline{\alpha})$ is a weighting matrix, that is, the inverse of the covariance matrix $\Sigma = E[\underline{g}(t, \underline{\alpha}) \underline{g}(t, \underline{\alpha})^\top]$, with the sample estimate adjusted for serial correlation and heteroscedasticity using the method of Newey and West (1987a) with Bartlett weights. Since the system is exactly identified in the unrestricted model, $J_T^U = 0$.

For a null hypothesis of model k of the form $H_0 : \underline{\alpha}_k = \underline{0}$, where $\underline{\alpha}_k$ is a vector of order j, we use the following hypothesis testing method which focuses on evaluating the restrictions imposed by the restricted models on the unrestricted model:

$$\mathrm{LRT}_k(j) = J_T(\underline{\acute{\alpha}})^R - J_T(\underline{\hat{\alpha}})^U = J_T(\underline{\acute{\alpha}})^R \overset{a}{\sim} \chi^2(j),$$

which is the difference of the restricted $J_T(\underline{\acute{\alpha}})^R$ and the unrestricted $J_T(\underline{\hat{\alpha}})^U$ objective function for the efficient GMM estimator where the weighting matrix from the *unrestricted model* is used to calculate both $J_T(\underline{\acute{\alpha}})^R$ and $J_T(\underline{\hat{\alpha}})^U$. The second equation holds from $J_T(\underline{\hat{\alpha}})^U = 0$. This is analogous to the likelihood ratio test. The likelihood ratio test requires a two-stage GMM estimation, estimating the optimal weighting matrix in the unrestricted model, followed by estimating the parameters of the restricted model.

In addition to these statistical tests, we also examine the economic importance of differences between alternative models by investigating the ability of the model to capture the mean and variance of changes in the interest rate. The R^2 metrics of the drift and the diffusion will provide insight into this issue, where the R^2, the coefficient of determination, is the proportion of the total variation in the ex post interest rate changes (R_1^2) and *squared interest*

[9] To be precise, condition (iii) is a sufficient condition for stationarity and unattainability of infinity. The necessary condition is weaker, $2\alpha_3 < \alpha_6$ if α_6 is not zero. However, we impose the negative leading term in the drift to ensure mean reversion at high values of the interest rate.

[10] The above boundary conditions can be easily derived as in Appendix A. For a more detailed reference, please see Aït-Sahalia (1996b).

rate changes (R_2^2) that can be explained by the conditional expected interest rate changes and conditional volatility measures, respectively.[11] This metric provides information about how well each model is able to forecast the future level and volatility of the interest rate. Along with the formal statistical test procedures above, this R^2 metric will provide an intuitive way of evaluating the economic significance of differences between alternative models.

2.2 The nonparametric benchmark

Stanton (1997) presents a nonparametric method of estimating the parameters of a continuous-time diffusion process using discrete data.[12] This methodology is virtually free of discretization error, and in addition, it is model-free so that the functional form of the drift can be estimated without any restrictions. Putting aside the probable estimation errors of nonparametric methods, which is beyond the scope of this article, we visually compare the drift and the diffusion of our model estimated with the above GMM procedure to those of the interest rate estimated by Stanton's (1997) procedure. Even though this procedure does not provide a formal statistical test as in the previous parametric test, it provides us with some intuition regarding how well our model captures the dynamics described in the nonparametric literature.

2.3 Data

We use the monthly data of McCulloch and Kwon (1993) over the period from December 1946 to February 1991.[13] Figure 2 depicts the historical

---

[11] The explained variable in the R_2^2 is $r(t)^2$, not $\epsilon(t+1)^2$ from which the parameters governing the diffusion function are estimated, since $\epsilon(t+1)^2$ varies across the models.

[12] The drift $\mu(r(t))$ is approximated

$$\mu(r(t)) = \frac{1}{6\triangle}\left[18E_t(r(t+\triangle)-r(t))-9E_t(r(t+2\triangle)-r(t))+2E_t(r(t+3\triangle)-r(t))\right],$$

where $\triangle$ is an observation interval of the data, which is a month in our case. The conditional expectation of the interest rate change is estimated as

$$E_t[r(t+\triangle)-r(t)|r(t)=r] \approx \frac{\sum_{s=1}^{T-1}[r(s+\triangle)-r(s)]K\left(\frac{r-r(s)}{h}\right)}{\sum_{s=1}^{T-1}K\left(\frac{r-r(s)}{h}\right)},$$

where $K(z) = 1/\sqrt{2\pi}\exp\left(-.5z^2\right)$. The conditional variance $\sigma(r(t))$ is estimated in a similar way. The critical parameter is the bandwidth, h, which is chosen based on the dispersion of the observations. Following Boudoukh et al. (1997), we choose $h = 1.5\sigma(\hat{r})T^{-(1/5)}$, which will yield similar functions to those in Stanton (1997). Since the purpose of this article is a visual comparison of our model and the nonparametric model, the choice of the bandwidth is not critical. For a determination of optimal bandwidth, see Aït-Sahalia (1996b).

[13] Stanton (1997) documents that the monthly frequency of the data does not adversely affect the estimation. He finds that the discrete approximation errors of the diffusion process are quite small, as long as the sampling frequency is monthly or finer.

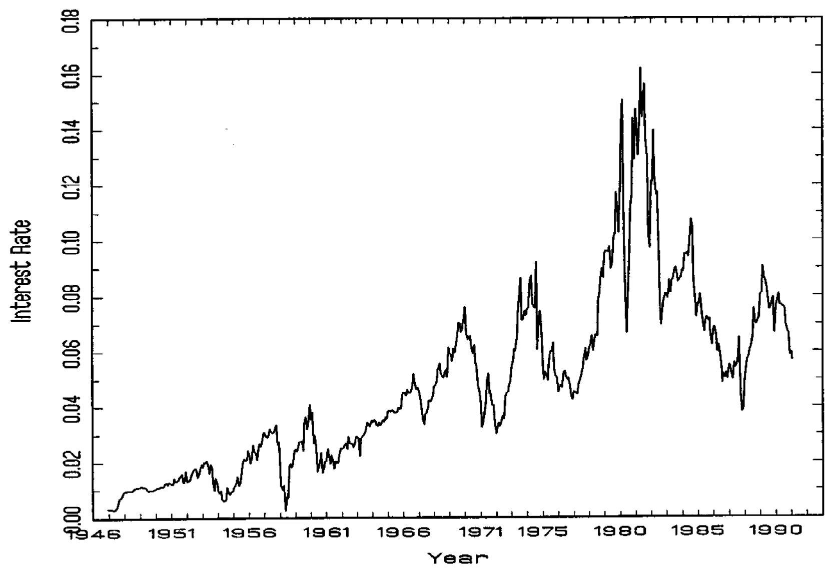


Figure 2
Historical values of the interest rate of McColluch and Kwon (1993) data

levels of the interest rate over the period. It is clear that this data covers virtually the entire range of interest rates from very low to high levels. In addition, in order to compare our results to CKLS (1992), we truncate the sample before January 1960. One impact of this truncation is that it eliminates very low levels of interest rates (say, 2%), which have not been observed since 1960. Further, we are able to check if the parameter values are stable over time.

Table 2 reports the means, standard deviations, and first five autocorrelations of the one-month interest rate. We choose the one-month interest rate as a proxy for the short rate, since the instantaneous interest rate in McCulloch and Kwon (1993) contains a large degree of extrapolation error.[14] The truncation of the earlier period increases the mean of the interest rate and lowers its standard deviation.

2.4 The empirical results

2.4.1 Comparison with linear drift-class models. We begin by estimating the unrestricted model and the five restricted interest rate processes using the LRT test statistic outlined above as a measure of goodness-of-fit

[14] The empirical results with the instantaneous interest rate are available upon request from the authors. The results are similar to those reported in the article, but are much more favorable to our model.

Table 2
Descriptive statistics

Variable	Sample size	Mean	SD	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
1946–1991								
r_t	531	.0482	.0319	.982	.963	.944	.930	.919
$\triangle r_t$	530	.0001	.0061	.022	$-.020$	$-.117$	$-.077$	$-.017$
1960–1991								
r_t	374	.0618	.0282	.969	.936	.905	.881	.863
$\triangle r_t$	373	.0001	.0070	.030	$-.021$	$-.124$	$-.098$	$-.018$

Means, standard deviations, and autocorrelations of monthly interest rates and interest rate changes are computed from December 1946 through February 1991, and from January 1960 through February 1991, respectively. The variable r_t stands for the interest rate and $\triangle r_t = r_{t+1} - r_t$ is the associated monthly interest rate change. ρ_j denotes the jth-order autocorrelation coefficient.

of each nested model. Table 3 reports the parameter estimates, p-values, the LRT test statistic, and the R^2 for the unrestricted model and for each of the five nested models.

As shown, all three affine-class models are strongly rejected at conventional confidence levels in the LRT test for goodness-of-fit, and these results are robust for different samples. The Vasicek (1977) and CIR (1985) models are strongly rejected with χ^2 values in excess of 20 and can be rejected at the 99% confidence level. These are followed by the DK (1996) model, which is rejected at the 90% or 95% confidence levels. Of particular note, the estimate of the intercept term in the diffusion, $\hat{\alpha}$, is negative and significant. Does this negative value of α satisfy the regularity condition of the DK model in Table 1? It is straightforward to see that the lower bound $r^* = -\alpha/\beta$ is 0.0171 (for 1946–1991) and 0.0332 (for 1960–1991). Then, checking the boundary condition in Table 1,

1946–1991	$2\kappa(\theta - r^*) = 2(.0608)(.0877 - .0171) = .0086$	$\not>$	$\beta = .0117$
1960–1991	$2\kappa(\theta - r^*) = 2(.1347)(.0762 - .0332) = .0116$	$\not>$	$\beta = .0181$

which violates the extended Feller condition. Further, it is obvious from Figure 2 that there are some observed interest rates below the theoretical lower bounds, r^*. All these results are challenging to the validity of the DK model.

As we discussed, another way to gauge the relative performance of the alternative models is the R^2, which measures their forecast power for interest rate changes and squared interest rate changes, and provides simple ex post measures of interest rate change and its volatility. The results are presented in the last column of Table 3.

The R_1^2's of the affine models are very low, and the models appear similar in their weak forecasting ability. Specifically the models explain only 0.4% to 1% of the total variation in interest rate changes. The models have a better ability to forecast the volatility of interest rate changes. The proportion of the total variation in volatility captured by the models ranges from 5% to 16%. In particular, the intercept in the diffusion term seems to be very important

Table 3
Parameter estimate for the spot rate process

	κ	θ	σ			$\chi^2(3)$	R^2
Vasicek: $dr = \kappa(\theta - r)dt + \sigma dw$							
1946–1991	.0484	.1132	.0112			39.81	$R_1^2 = .0040$
	(.310)	(.253)	(.000)			(.000)	$R_2^2 = .0000$
1960–1991	.1283	.0827	.0133			34.10	$R_1^2 = .0092$
	(.209)	(.042)	(.000)			(.000)	$R_2^2 = .0000$
CIR: $dr = \kappa(\theta - r)dt + \sigma\sqrt{r}dz$							
1946–1991	.0613	.0902	.0688			19.63	$R_1^2 = .0049$
	(.277)	(.196)	(.000)			(.000)	$R_2^2 = .0685$
1960–1991	.1343	.0781	.0657			22.49	$R_1^2 = .0096$
	(.206)	(.033)	(.000)			(.000)	$R_2^2 = .0517$
DK: $dr = \kappa(\theta - r)dt + \sqrt{\alpha + \beta r}dw$							
1946–1991	.0608	.0877	$-.0002$	.0117		7.41	$R_1^2 = .0049$
	(.286)	(.207)	(.000)	(.000)		(.025)	$R_2^2 = .1382$
1960–1991	.1347	.0762	$-.0006$	.0181		5.51	$R_1^2 = .0096$
	(.221)	(.043)	(.000)	(.000)		(.064)	$R_2^2 = .1626$
CKLS: $dr = \kappa(\theta - r)dt + \sigma r^{1.5}dw$							
1946–1991	$-.0203$	$-.1547$	1.2825			7.51	$R_1^2 = -.002$
	(.425)	(.437)	(.000)			(.057)	$R_2^2 = .2669$
1960–1991	.0472	.1359	1.1964			2.71	$R_1^2 = .0038$
	(.396)	(.348)	(.000)			(.438)	$R_2^2 = .2550$
AG: $dr = \kappa(\theta - r)rdt + \sigma r^{1.5}dw$							
1946–1991	3.6438	.0823	1.2801			5.87	$R_1^2 = .0244$
	(.063)	(.000)	(.000)			(.118)	$R_2^2 = .2671$
1960–1991	3.4387	.0828	1.1920			1.54	$R_1^2 = .0255$
	(.098)	(.000)	(.000)			(.672)	$R_2^2 = .2548$
Unrestricted: $dr = (\alpha_1 + \alpha_2 r + \alpha_3 r^2)dt + \sqrt{\alpha_4 + \alpha_5 r + \alpha_6^3 r^3}dw$							
	α_1	α_2	α_3	α_4	α_5	α_6	R^2
1946–1991	$-.0071$	.6907	-7.476	.0000	.0003	1.3111	$R_1^2 = .0312$
	(.184)	(.066)	(.058)	(.393)	(.479)	(.014)	$R_2^2 = .2644$
1960–1991	$-.0251$	1.199	-10.466	$-.0001$	.0031	1.1926	$R_1^2 = .0373$
	(.125)	(.068)	(.054)	(.378)	(.373)	(.047)	$R_2^2 = .2493$

The estimation horizons for r_t, the annualized one-month U.S. Treasury bill rate, are from 12/1946 to 02/1991 and 01/1960 to 02/1991, respectively. The parameters are estimated by the generalized method of moments with z-statistics in parentheses. The R_i^2 values are computed as the proportion of the total variation of the actual short rate changes ($i = 1$) and their volatility ($i = 2$) explained by the respective predictive values for each model. The $\chi^2(j)$ test statistics evaluate overidentifying restrictions imposed by alternative models on the unrestricted model. All these statistics are reported with p-values in parentheses where j stands for degrees of freedom. All estimates are adjusted for conditional heteroscedasticity and serial correlation using the method of Newey and West (1987a).

in forecasting the volatility, as is evidenced by the difference between the CIR model and the DK model. Therefore α in the DK model is somewhat problematic, since it enhances forecasting power immensely but its value is inconsistent with the regularity condition of the model.

The performance of the CKLS model is not as clear as that of the affine-class models. For the sample from 1946 to 1991, the model is rejected at the 90% confidence level. Further, κ, the mean-reversion parameter, is negative, which violates the regularity condition of the model, and the R_1^2 of the drift equation turns out to be negative.[15] However, the model seems to perform quite well for 1960–1991, such that the model is not rejected at even the 90% confidence interval, and the mean-reversion parameter κ becomes positive. The R_2^2 of the model is about 26%, suggesting that the model can capture the variation in volatility far better than the affine-class model. In summary, except for the CKLS model (for 1960–1991), the linear drift models fail to survive the goodness-of-fit tests and have very poor forecasting power for actual interest rate changes as well as volatility.

In contrast, our model is not rejected at even the 90% confidence level regardless of the sample periods, and in particular, the model works particularly well for 1960–1991, with χ^2 values of about 1.5. Further, all the parameters outside of our model — α_1, α_4, and α_5 — are individually insignificant, which confirms that the χ^2 results do not stem from the insignificance of one or two particular parameters, but from the three parameters together. On the other hand, all the parameters which are relevant in our model — α_2, α_3, and α_6 — are individually significant. In fact, the given parameter estimates of the unrestricted model do not satisfy the boundary conditions stated in Table 1. Only when we eliminate the insignificant parameters in the unrestricted model does the model satisfy the boundary condition, which is, in fact, reduced to our model. In addition, the forecasting ability of our model is better than the linear drift models. The R_1^2 is about 2.5%, which is greater than that of the best linear drift model by 2.5 to 5 times. The R_2^2 is similar to that of the CKLS model since both models share the same specification of diffusion of the interest rate. Both R_2^2 values are quite close to those of the unrestricted model, which is encouraging considering the small number of parameters in our model.[16]

Another point worth noting is that the mean-reversion parameter κ in all four linear drift models including the CKLS model are insignificant regardless of the sample periods. This is consistent with the finding of CKLS (1992) who document that there is only very weak evidence of mean reversion and state further that the additional generality obtained by allowing the interest rate process to be mean reverting may not be justified. However, Aït-Sahalia (1996a,b) questions the latter statement and states that this weak evidence of mean reversion is induced mainly by the linear specification of the drift, and evidences the existence of strong mean reversion in nonparametric estimation. His statement is confirmed with our model in that the mean-reversion coefficient κ of our model is

[15] Even though the model has an intercept, negative R^2 values are possible since the drift function is jointly estimated with the diffusion function, and $\epsilon(t+1)$ should be orthogonal to $r(t+1)^2$ in addition to the usual restrictions of the regression.

[16] Since the R_2^2's are the explained proportion of the total variation in $r(t)^2$, not $\epsilon(t+1)^2$, the R_2^2's of the restricted models can be higher than that of the unrestricted model.

fairly significant. Therefore the effort to allow the interest rate process to be mean reverting can be justified as long as the model specifications are correct.

A more important result is that our model maintains similar parameter values over different samples, whereas the parameter estimates of the linear drift models critically depend on the sample period. Notice that the sample of 1946–1991 contains the low level of interest rates ranging from 0.3% to 3.5%, which are missing in the sample of 1960–1991. Since our model captures well the dynamics of the interest rate over this range, the parameter values remain fairly stable. However, the linear drift models have to adjust their parameter values to account for this range, since they are sensitive to the dense region of the interest rate.

Finally, we compare the CKLS model and our model for the sample period 1960–1991, since none of the models is rejected. First, χ^2 values of the LRT statistic can be used to rank alternative models if the number of restrictions is the same across the models. The χ^2 value of our model is 1.54 whereas that of the CKLS model is 2.71; thus we can conclude our model outperforms the CKLS model. To confirm this result we estimate a parsimonious model which nests the CKLS model and our model such that

$$dr(t) = (\alpha_1 + \alpha_2 r(t) + \alpha_3 r(t)^2)dt + \sigma r^{1.5} dw(t). \tag{8}$$

It is obvious to see that the model will be reduced to the CKLS model if $\alpha_3 = 0$, whereas $\alpha_1 = 0$ results in our model. The GMM estimation results are documented as follows:

Parameters	α_1	α_2	α_3	σ
Estimate	−.0078	.7326	−7.5325	1.2424
p-value	(.1573)	(.0535)	(.0488)	(.0000)

Since α_3 is significant whereas α_1 is not, only our model is supported. In summary, we can conclude that our model outperforms all linear drift models, including the CKLS model, across several measures of performance and captures the mean-reverting behavior of the interest rate.[17]

2.4.2 Comparison with the nonparametric model. Figure 3A,C illustrates the comparison of the nonparametrically determined drift and the fitted drift of the model given the parameter values estimated above. Overall

[17] We perform the Wald test in addition to the LRT test. Unlike the LRT test, which requires two-stage estimation, the Wald test statistic can be computed in one stage. If the weighting matrix in the GMM estimation is the same, the Wald test is asymptotically equivalent to the LRT test, as shown by Newey and West (1987b). However, the results can be somewhat different in finite samples. The Wald test results in our empirical analysis turn out to be similar to the results of the LRT test. We report only the LRT results in Table 3.

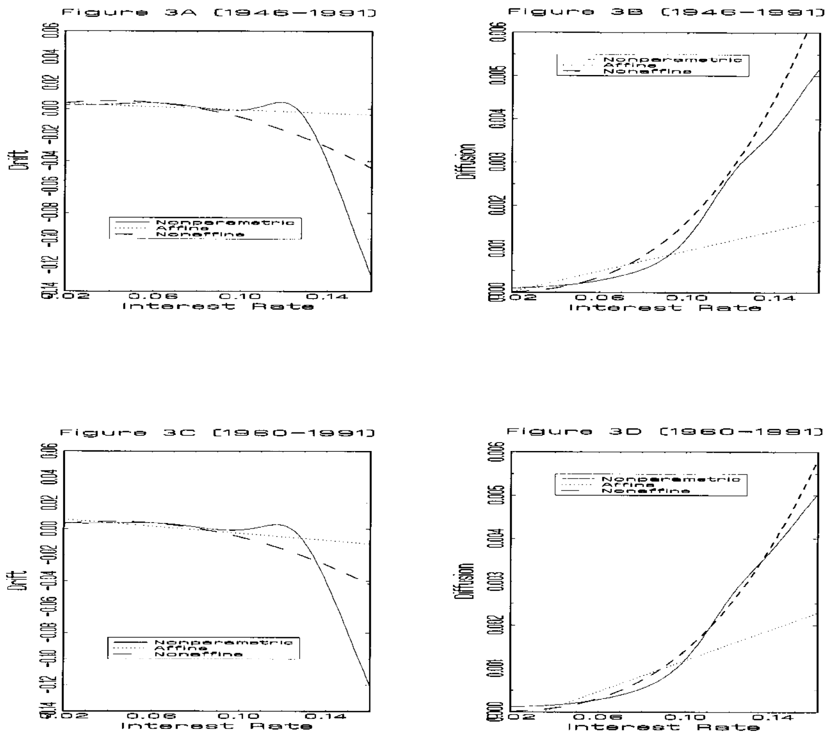


Figure 3
Comparison of nonparametrically estimated and parametric model's drift (A and C) and diffusion (B and D) of the interest rate

our model captures the nonlinearity of the drift fairly well in both samples. In particular, the drift of the model is almost indistinguishable from the nonparametric drift over the low, medium, and reasonably high range of interest rates. In contrast, the mean reversion of our model at the extraordinary high levels of interest rates is smaller than the nonparametric one, but it still performs better than the DK model.[18]

Figure 3A,C also provides insight into why the estimate of the mean-reversion parameter of the linear drift model is so imprecise. The estimation of the drift hinges critically upon the weight of estimation of different ranges of interest rate. If one puts more weight on low and medium interest rates (say, below 12% in Figure 3A), the slope or the mean-reversion speed of the linear drift will be quite small. However, the linear drift cannot explain

[18] The behavior of the nonparametrically estimated drift at this range of the interest rate is very sensitive to the choice of the bandwidth h, since the data is sparse. A larger value of the bandwidth will make the drift function smoother and will lessen the discrepancy between the drift functions in our model and the nonparametric model.

the sharp drop in the drift for high interest rates. On the other hand, if one focuses on the range of high interest rates, the mean-reversion speed should be quite large, which results in poor fitting in the low and medium range of interest rates. This dilemma is a result of the leverage feature of the linear specification. In comparison, the quadratic specification of the drift in our model avoids this problem, allowing a greater degree of freedom in fit over the entire range of interest rates.

Figure 3B,D compares the diffusions (variance) of the parametric models and of the nonparametric estimation. Similar to the findings in CKLS (1992) and Stanton (1997), the nonparametric variance is a convex function of the interest rate. The fitted variance of our model ($\sigma^2 r^3$) approximates that of the nonparametric estimation very well over the entire range of interest rates. For low interest rates, it undershoots by a small magnitude, but as interest rates climb, the model estimate is virtually indistinguishable from the nonparametrically determined function. In contrast, the DK model, which assumes a linear variance function, cannot exhibit the increasingly increasing pattern. Again it is obvious that this poor performance of the DK model results from the leverage feature of the linear specification. Overall the result indicates that the nonlinearities proposed in our model contribute significantly to explaining the dynamics of the interest rate, while maintaining a small number of parameters.

3. The Valuation of Bonds

In this section we will explore the valuation of zero coupon bonds. The prices of zero coupon bonds are determined by two things: the stochastic differential equation of the interest rate, and the diffusion of the stochastic discount factor. The stochastic differential equation of the interest rate is specified in Assumption 1. In this section we will first specify the diffusion of the stochastic discount factor in Assumption 2.

Assumption 2. *The diffusion of the stochastic discount factor is specified as*[19]

$$\begin{aligned} \eta(t) &= \left(\frac{\sigma_{q_1}}{\sqrt{r(t)}} \quad \sigma_{q_2}\sqrt{r(t)} \right) \\ w_q &= (w_{q_1} \;\; w_{q_2})^\top. \end{aligned} \tag{9}$$

Assumption 2 means that the diffusion of the stochastic discount factor

[19] It can be easily checked that the specification of $\eta(t)$ satisfies the Novikov condition $E[\exp(\frac{1}{2}\int_0^T \eta(s)^2\,ds)] < \infty$ under the boundary conditions of $r(t)$.

consists of two components. The first component is linearly proportional to the square root of the interest rate, whereas the second component is inversely related to the square root of the interest rate. This implies that an increase in the interest rate does not necessarily increase the volatility of the stochastic discount factor or the volatility of the time-varying risk premium. In contrast, Vasicek (1977) and CIR (1985) both assume that an increase in the interest rate is either irrelevant for or yields an increase in the volatility of the risk premium of assets.[20] Assumption 2 implies that when interest rates are outside of the normal range (that is, either very low or very high), the volatility of the discount factor will increase. In other words, the volatility of the stochastic discount factor is a U-shaped function of the interest rate.

Given Assumptions 1 and 2 and the Euler equation [Equation (2)], the time t valuation of a bond which pays \$1 at time T is

$$V(r,t,T) = E_t^P\left[M(t,T)V(r(T),T,T)\right] = E_t^P\left[M(t,T)\right], \tag{10}$$

where $r(t)$ evolves according to Equation (6). The *unit* risk premium of interest rate risk is determined by the covariance with the stochastic discount factor such that

$$\begin{aligned}
-\mathrm{cov}\left(\frac{dM(0,t)}{M(0,t)}, dr(t)\right) &= -\mathrm{cov}_t\left(\eta(t)dw_q(t), \sigma r^{1.5}dw(t)\right) \\
&= -\left[\rho_1\sigma_{q_1}\sigma r(t) + \rho_2\sigma_{q_2}\sigma r(t)^2\right]dt \\
&= \left[\lambda_1 r(t) + \lambda_2 r(t)^2\right]dt,
\end{aligned}$$

using Equations (2), (5), and (9). Alternatively, using the Girsanov theorem, we can reexpress the valuation equation [Equation (10)] for the same bond under the equivalent martingale measure Q as[21]

$$V(r,t,T) = E_t^Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\right] \tag{11}$$

[20] The specifications of the diffusion of the Vasicek and CIR models are

$$\begin{array}{lcl}
\text{Vasicek} & \eta = & \sigma_q\sqrt{r(t)} \\
\text{CIR} & \eta = & \sigma_q.
\end{array}$$

[21] The risk-neutral dynamics of the interest rate in Equation (11) are viable under the following technical conditions

$$\lambda_1 < \kappa\theta, \quad \lambda_2 > -\kappa,$$

to procure the equivalence of the two measures.

subject to

$$dr(t) = \left[(\kappa\theta - \lambda_1)r(t) - (\kappa + \lambda_2)r(t)^2\right] dt + \sigma r(t)^{1.5} d\tilde{w}(t),$$

where

$$\tilde{w}(t) = w(t) + \int_0^t \left(\frac{\lambda_1}{\sigma\sqrt{r(s)}} + \frac{\lambda_2\sqrt{r(s)}}{\sigma} \right) ds.$$

Solving either of the valuation equations, Equations (10) and (11), yields the following partial differential equation,

$$\frac{1}{2}\sigma^2 r^3 V_{rr} + r\kappa(\theta - r)V_r + V_t = rV + \left(\lambda_1 r + \lambda_2 r^2\right) V_r, \tag{12}$$

with the terminal condition $V(r, T, T) = 1$. The left-hand side of Equation (12) results from Ito's lemma and the right-hand side derives from the martingale restriction on the stochastic discount factor. The right-hand side can be rewritten as

$$V\left[r + \left(\lambda_1 r + \lambda_2 r^2\right)\frac{V_r}{V}\right].$$

The expression in the square brackets represents the instantaneous return of the bond, which is composed of the riskless interest rate and its risk premium. The risk premium of the bond is the product of the market price of the *unit* risk of the interest rate and the interest rate sensitivity of the bond, a measure of the riskiness of the bond. The solution to the partial differential equation [Equation (12)] is summarized in the following proposition.

Proposition 1. *The time t price of the bond with maturity at T is*

$$V(r, t, T) = \frac{\Gamma(\beta - \gamma)}{\Gamma(\beta)} M\left(\gamma, \beta, -x(r, t, T)\right) x(r, t, T)^{\gamma}, \tag{13}$$

where

$$\begin{aligned} x(r, t, T) &= \frac{2(\kappa\theta - \lambda_1)}{\sigma^2\left(e^{(\kappa\theta - \lambda_1)(T-t)} - 1\right) r(t)} \\ \gamma &= \frac{1}{\sigma^2}\left[\sqrt{\phi^2 + 2\sigma^2} - \phi\right] \\ \beta &= \frac{2}{\sigma^2}\left[\kappa + \lambda_2 + (1+\gamma)\sigma^2\right] \\ \phi &= \kappa + \lambda_2 + \frac{1}{2}\sigma^2 \end{aligned}$$

Proof. See Appendix C. ∎

The corresponding expression for the yield to maturity is defined by

$$y(r, t, T) = -\frac{1}{T-t}\left[\ln\left(\frac{\Gamma(\beta-\gamma)}{\Gamma(\beta)}\right) + \ln M(\gamma, \beta, -x(r, t, T)) + \gamma \ln x(r, t, T)\right]. \tag{14}$$

As maturity increases, the yield to maturity starts from the current interest rate and eventually approaches a limit which is independent of the current interest rate, reflecting the mean-reversion property:

$$y(r, t, \infty) = \gamma(\kappa\theta - \lambda_1).$$

When the current interest rate is below this long-term yield, the term structure is uniformly rising. With the interest rate in excess of $(\kappa\theta-\lambda_1)/(\kappa+\lambda_2)$, the term structure declines. For intermediate values, the yield curve is humped. These properties are similar to those of the CIR (1985) model. Since the bond pricing function is a closed-form expression, all of the other comparative statics for the yield curve are easily obtained. One particularly interesting observation is that the yield to maturity is not an affine or linear function of the interest rate. The sensitivity of the yield with respect to the interest rate is

$$\frac{\partial y}{\partial r} = \frac{1}{(T-t)r}\left[-\frac{\gamma M(\gamma+1, \beta+1, -x(r, t, T))}{\beta M(\gamma, \beta, -x(r, t, T))} x(r, t, T) + \gamma\right],$$

which is not constant but rather a function of the interest rate. Even though the features of the nonlinearity depend on the parameter values, it is a concave function of the interest rate for reasonable values of the parameters. Thus this model belongs to a nonaffine class along with Beaglehole and Tenney (1992), Constantinides (1992), and Ahn (1997). The nonaffine property is not surprising since the power of the interest rate in the diffusion is greater than 0.5 as suggested in Campbell, Lo, and MacKinlay (1996). This is an empirically testable hypothesis. The next section investigates the existence of the probable nonlinearity between the interest rate and the yields.

We investigate the implication of the model for discount bond option prices. Consider a European call option with exercise price K and maturity τ on the bond with maturity $\tau+\tau_p$. The price of this option can be represented as

$$\begin{aligned} C(t, \tau, \tau_p) &= E_t^P\left[M(t, t+\tau)(V(t+\tau, t+\tau+\tau_p) - K)^+\right] \\ &= E_t^Q\left[e^{-\int_t^{t+\tau} r(s)\,ds}\left(V(t+\tau, t+\tau+\tau_p) - K\right)^+\right] \end{aligned}$$

$$
\begin{aligned}
&= E_t^Q\left[e^{-\int_t^{t+\tau} r(s)\,ds} V(t+\tau, t+\tau+\tau_p) | V(t+\tau, t+\tau+\tau_p) \geq K\right] \\
&\quad - K E_t^Q\left[e^{-\int_t^{t+\tau} r(s)\,ds} | V(t+\tau, t+\tau+\tau_p) \geq K\right] \\
&= V(t, t+\tau+\tau_p)\mathrm{Pr}_1(V(t+\tau, t+\tau+\tau_p) \geq K) \\
&\quad - K V(t, t+\tau)\mathrm{Pr}_2(V(t+\tau, t+\tau+\tau_p) \geq K).
\end{aligned}
$$

The probability of being in the money, $\mathrm{Pr}_j(\cdot)$, is defined on the following process,

$$
\begin{aligned}
dr(t) &= \left[(\kappa\theta - \lambda_1)r(t) - (\kappa + \lambda_2)r(t)^2 + \frac{V_r(t, t+\tau_j)}{V(t, t+\tau_j)}\sigma^2 r(t)^3\right] dt \\
&\quad + \sigma r(t)^{1.5} dw_j(t) \\
w_j(t) &= \tilde{w}(t) - \int_0^t \frac{V_r(t, t+\tau_j)}{V(t, t+\tau_j)}\sigma r(t)^{1.5}\,ds,
\end{aligned}
\tag{15}
$$

where $j = 1$ and 2, $\tau_1 = \tau + \tau_p$, $\tau_2 = \tau$, and

$$
\frac{V_r(t, t+\tau_j)}{V(t, t+\tau_j)} = \frac{1}{r}\left[\frac{\gamma M(\gamma+1, \beta+1, -x(r, t, t+\tau_j))}{\beta M(\gamma, \beta, -x(r, t, t+\tau_j))} x(r, t, t+\tau_j) - \gamma\right].
$$

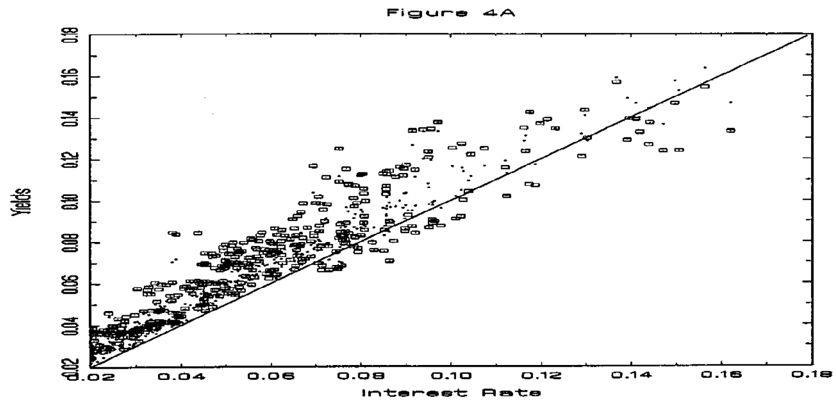
Since the stochastic differential Equation (15) does not yield closed-form expressions for the transition densities of $r(t)$ under the Q_j measure, we solve for $\mathrm{Pr}_j(\cdot)$ by numerical methods such as finite different methods or Monte Carlo simulation methods.

4. The Cross-Sectional Performance of the Model

In this section we empirically investigate the primary testable proposition from the previous section, namely nonlinearity of the relationship between the short-term interest rate and the yields to maturity. However, since the affine-class models are not nested within our model, we cannot directly test this hypothesis. Instead, we first investigate the probable nonlinearity contained in the data using generic nonlinear regression analysis such as Taylor series expansion, flexible Fourier form series expansion, and nonparametric kernel regression.

4.1 The nonlinear relationship: empirical evidence

First, we empirically explore whether there is a nonlinear relationship between the interest rate and yield to maturity, and document the characteristics of the nonlinearity. We use the data of McCulloch and Kwon (1993), and consider five different yields to maturity from one year to five years.



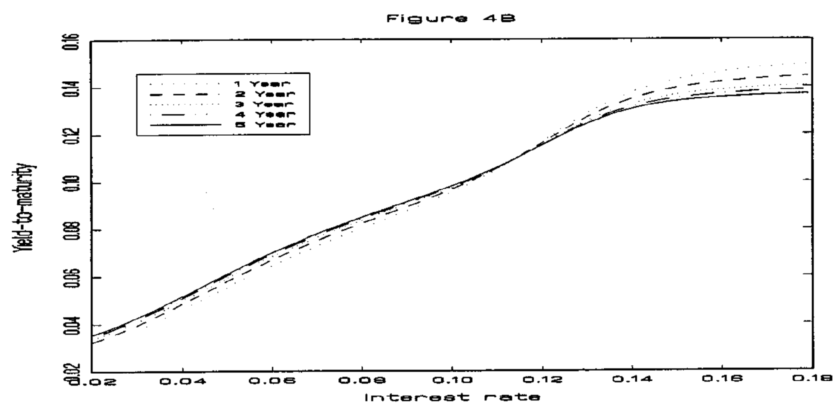


Figure 4
Scatter plot of 1-year (circle) and 5-year (box) yields on the domain of the interest rate (A) and the fitted yields as functions of the interest rate using the nonparametric kernel regression (B)

Figure 4A represents the scatter plots of 1- and 5-year yields for the period 1946–1991.

From Figure 4A, two conjectures may be considered. First, there seems to be a probable nonlinear relationship between the interest rate and the yields. Especially for the higher range of interest rates, the nonlinearity is more prominent, and the pattern is concave. Second, the nonlinearity is stronger for the 5-year yield than for the 1-year yield, so that the magnitude of the nonlinearity may increase as time to maturity increases.

We formally test these conjectures using two alternative tests for nonlinearity. Many series expansions exist in the numerical approximation literature to estimate an unknown functional form. The first one we use is a Taylor series expansion. Because the interest rate has a very small value, we regress the yields on first- and second-order terms of the short-term interest rate:[22]

$$y_{t,\tau} = a_{0,\tau} + a_{1,\tau} r_t + a_{2,\tau} r_t^2 + \epsilon_{t,\tau}, \tag{16}$$

where τ denotes the time to maturity of the yield and $a_{i,\tau}\ \forall\ i = 0, 1,$ and 2 are constants.

The coefficient of the second-order term, $a_{2,\tau}$, measures the extent of nonlinearity. If the shape of the relationship is concave (convex), $\text{sign}(a_{2,\tau})$ is negative (positive). We test our first conjecture, the null of linearity, by testing the restriction that the coefficients, $a_{2,\tau}\ \ \forall\tau = 1, 2, 3, 4,$ and 5 are jointly zero. To test this hypothesis, we construct a Wald test statistic,

$$J = \underline{\hat{a}}_2^\top \text{cov}(\underline{\hat{a}}_2)^{-1} \underline{\hat{a}}_2, \tag{17}$$

where $\underline{\hat{a}}_2$ is a 5×1 vector of estimates of $a_{2,\tau}$'s, and $\text{cov}(\underline{\hat{a}}_2)$ is a covariance matrix of estimates $\underline{\hat{a}}_2$. The asymptotic distribution of this statistic is a $\chi^2(5)$ distribution.

Further, the second derivative of $y_{t,\tau}$ with respect to r_t is $2a_{2,\tau}$, so $|a_{2,\tau}|$ measures the degree of curvature. Hence our second conjecture, that the nonlinearity (or the extent of curvature) is an increasing function of time to maturity, can be expressed as

$$|a_{2,m}|\ >\ |a_{2,n}|\ \text{ if } m\ >\ n.$$

This property can be tested with the Gaussian normal statistic, $z(m, n)$:

$$z(m, n) = \frac{|a_{2,m}| - |a_{2,n}|}{\sqrt{\text{Var}(a_{2,m} - a_{2,n})}}, \tag{18}$$

where $\text{var}(a_{2,m} - a_{2,n}) = \text{var}(a_{2,m}) - 2\text{cov}(a_{2,m}, a_{2,n}) + \text{var}(a_{2,n})$.

Since a Taylor series expansion sometimes provides poor approximations for an unknown function because it holds only locally, we also consider global approximation, the flexible Fourier form (FFF) series expansion suggested by Gallant (1981).[23] Because yields are positive, however, the functions are defined in the interval $(0, U]$ where $U \to \infty$. As a result, there are two alternative specifications: *half-range Fourier sine* or *cosine* series. After applying each series to the data, we find that the half-range Fourier cosine expansion is appropriate for this problem. The functional

[22] We also run the regression with a cubic term, which turns out to be insignificant.

[23] This method is applied in Pagan and Schwert (1989), Pagan and Hong (1991), and Boudoukh, Richardson and Whitelaw (1994).

form we consider is

$$y_{t,\tau} = b_{0,\tau} + b_{1,\tau} r_t + b_{2,\tau} \cos(r_t) + e_{t,\tau}. \tag{19}$$

Similar to the Taylor series expansion, $b_{2,\tau}$ is a measure of nonlinearity. A joint test for nonlinearity of all the five different yields is a $\chi^2(5)$ test statistic defined similarly to Equation (17). Further, the monotonic relationship between the times to maturity and the strength of nonlinearity can be tested by using the $z(m, n)$ test statistic in Equation (18) by replacing $a_{2,\tau}$ with $b_{2,\tau}$.

Table 4A documents the test results using the Taylor series expansion, and the test of the FFF series expansion is summarized in Table 4B. Similar to the empirical analysis in the previous section, we consider two sample periods, 1946–1991 and 1960–1991. The results are striking in that regardless of the chosen estimation method, there is strong evidence of nonlinearity. The χ^2 test for linearity is strongly rejected in both tests. The signs of the coefficients of nonlinearity, $a_{2,\tau}$ in Table 4A and $b_{2,\tau}$ in Table 4B, indicate that the concave relationship exists.[24]

Further, as maturities increase, the absolute values of $a_{2,\tau}$ and $b_{2,\tau}$ increase, as predicted by our second conjecture. The $z(m, n)$ test statics are all significant, so that the nonlinearity is an increasing function of the maturity. This is not surprising since as the maturity approaches zero, the yield becomes close to the interest rate itself, so that we can expect less nonlinearity. Figures 5 and 6 illustrate the fitted curve using the FFF series expansion. The curvature and the monotonic increasing property of the nonlinearity with respect to the maturity is evident. Alternatively, the nonlinear pattern is estimated using the nonparametric kernel regression. Figure 4B represents the relationship between the interest rate and the yields to maturity over the period 1946–1991 and yields similar nonlinear patterns as those in Figure 5. The concavity feature is prominent, and is an increasing function of time to maturity.

4.2 The estimation of the risk premium

In the previous section, the various empirical methods evidenced the strong nonlinearity between the interest rate and yields to maturity. Here we will investigate whether our model can generate similar nonlinear patterns. Since

[24] This result may be interpreted as evidence for the existence of multifactor models. For example, if there are two affine factors, the relationship between the interest rate and the yields is not a function anymore, but a projection of a hyperplane, such that there exist *linear* upper and lower bounds on the yields given at a particular level of the interest rate. If we estimate the functional relationship between the interest rate and the yields, ignoring the existence of multiple factors, the estimation procedure suffers from the problem of omission of a relevant independent variable, which in general makes the estimates biased. However, the omission of a relevant independent variable will not change the functional form by any means. Therefore the nonlinear relationship we document here is robust even in the presence of multiple factors, as long as those factors are affine factors. However, if there are missing data, especially in the high interest rate region, we can find a spurious nonlinear relationship.

Table 4
Test of nonlinearity between the short rate and yields

A: Taylor series expansion

	$a_{0,\tau}$		$a_{1,\tau}$		$a_{2,\tau}$		$z(m,n)$		R^2	
τ	1946–	1960–	1946–	1960–	1946–	1960–	1946–	1960–	1946–	1960–
1 Year	.0021	.0034	1.1907	1.1546	−1.2547	−1.0442	—	—	.9735	.9539
2 Year	.0030	.0048	1.2597	1.2161	−1.9727	−1.7417	7.5145	4.8830	.9525	.9132
3 Year	.0039	.0059	1.2944	1.2493	−2.4090	−2.1854	7.1176	5.0326	.9364	.8817
4 Year	.0049	.0071	1.3029	1.2550	−2.5831	−2.3560	4.0630	2.7741	.9244	.8588
5 Year	.0057	.0079	1.3096	1.2661	−2.7258	−2.5334	5.3313	4.7748	.9144	.8397

1946-: $\chi^2(5) = 142.66$ 1960-: $\chi^2(5) = 67.11$

B: Half-range Fourier cosine series expansion

	$b_{0,\tau}$		$b_{1,\tau}$		$b_{2,\tau}$		$z(m,n)$		R^2	
τ	1946–	1960–	1946–	1960–	1946–	1960–	1946–	1960–	1946–	1960–
1 Year	−2.5144	−2.0912	1.1910	1.1548	2.5164	2.0945	—	—	.9735	.9539
2 Year	−3.9538	−3.4901	1.2601	1.2166	3.9568	3.4948	7.5175	4.8834	.9526	.9132
3 Year	−4.8278	−4.3796	1.2949	1.2499	4.8317	4.3854	7.1183	5.0335	.9364	.8817
4 Year	−5.1759	−4.7202	1.3034	1.2557	5.1808	4.7274	4.0629	2.7733	.9244	.8588
5 Year	−5.4614	−5.0758	1.3101	1.2668	5.4670	5.0837	5.3273	4.7323	.9144	.8397

1946-: $\chi^2(5) = 142.45$ 1960-: $\chi^2(5) = 66.98$

Table 4 provides test results of the regression analysis to investigate the nonlinear relationship between the short rate and the yields by employing a Taylor series expansion (Table 4A) and a half-range Fourier cosine series expansion (Table 4B),

$$y_{t,\tau} = a_{0,\tau} + a_{1,\tau} r_t + a_{2,\tau} r_t^2 + \epsilon_{t,\tau}$$

and

$$y_{t,\tau} = b_{0,\tau} + b_{1,\tau} r_t + b_{2,\tau} \cos(r_t) + e_{t,\tau}$$

for all $\tau = 1, 2, 3, 4$, and 5. $\chi^2(5)$ is a test statistic for testing $a_{2,\tau} = 0,\ \forall\ \tau$ (Table 4A) or $b_{2,\tau} = 0,\ \forall\ \tau$ (Table 4B). $z(m,n)$ is a normal statistic for testing $|a_{2,m}| > |a_{2,n}|$ or $b_{2,m} > b_{2,n}$, where $m > n$. The parameters are heteroscedasticity and serial correlation adjusted, using the method of Newey and West (1987a). All the parameters are significant at the 99% level.

the cross-sectional relationships among asset prices are determined under the risk-neutral Q measure, this test provides an insight into the performance of our model under the Q measure. In particular, we will perform two tests of our model in conjunction with the tests of the DK model for comparison.

Under the Q measure, our short rate has the following process:

$$dr(t) = \left[(\kappa\theta - \lambda_1) r(t) - (\kappa + \lambda_2) r(t)^2\right] dt + \sigma r(t)^{1.5} d\tilde{w}(t),$$

where

$$\tilde{w}(t) = w(t) + \frac{1}{\sigma} \int_0^t \left(\frac{\lambda_1}{\sqrt{r}(s)} + \lambda_2 \sqrt{r}(s) \right) ds.$$

The general DK process follows

$$dr(t) = [\kappa\theta - \lambda_1 - (\kappa + \lambda_2) r(t)] dt + \sqrt{\alpha + \beta r(t)}\, d\tilde{w}(t),$$

where

$$\tilde{w}(t) = w(t) + \int_0^t \left(\frac{\lambda_1 + \lambda_2 r(s)}{\sqrt{\alpha + \beta r(s)}} \right) ds.$$

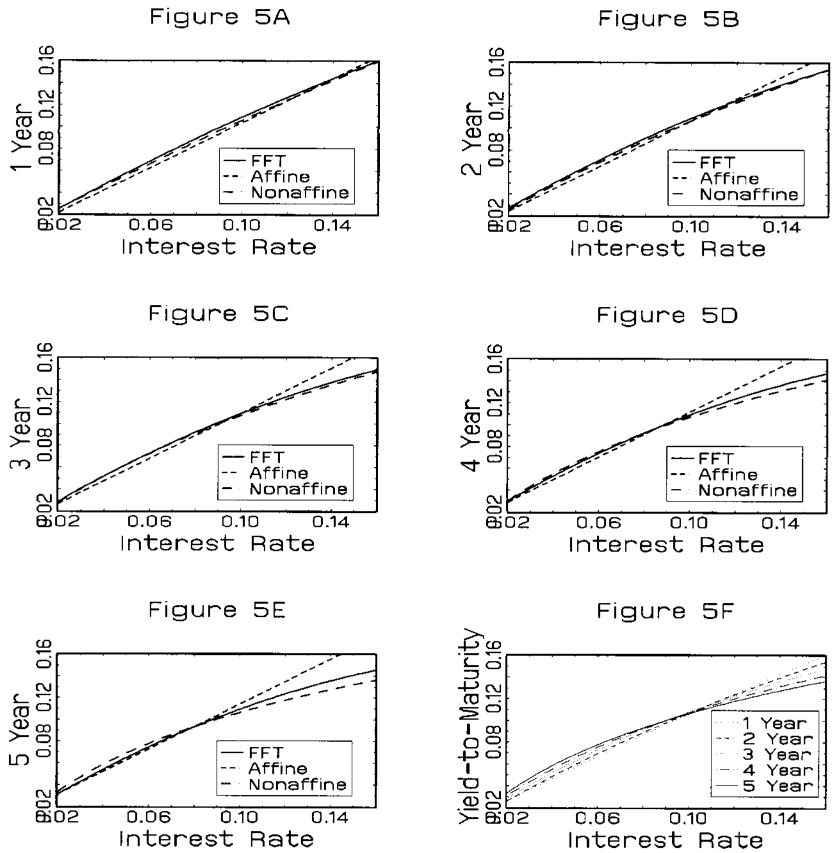


Figure 5
(A–E) Compare the FFF estimation and the calibrated model estimation (based on the two-step method) of the relationship between the interest rate and the yield with various maturities over the period from 1946 to 1991. (F) Represents the estimated yields of the model over different maturities.

One thing to notice is that although λ_1 and λ_2 can be specified freely in our model, this is not true in the DK model. The risk premiums in our model and the DK model are $-(\lambda_1 r(t) + \lambda_2 r(t)^2) V_r / V$ and $-(\lambda_1 + \lambda_2 r(t)) V_r / V$, respectively. As volatility goes to zero ($r(t) \downarrow 0$ in our model and $r(t) \downarrow -\alpha/\beta$ in the DK model) the risk premium should also go to zero; otherwise, there will be an arbitrage opportunity. Clearly this implies $\lambda_1 = \lambda_2 \alpha/\beta$ in the DK model.

Since we have the estimated values for κ, θ, and σ from the interest rate dynamics in Section 3, using McCulloch and Kwon (1993) data, we can estimate the parameters governing the market price of risk, λ_1 and λ_2 such that they minimize the total sum of squared errors (TSSE), a distance measure of the observed yields and the fitted values of the model. Therefore

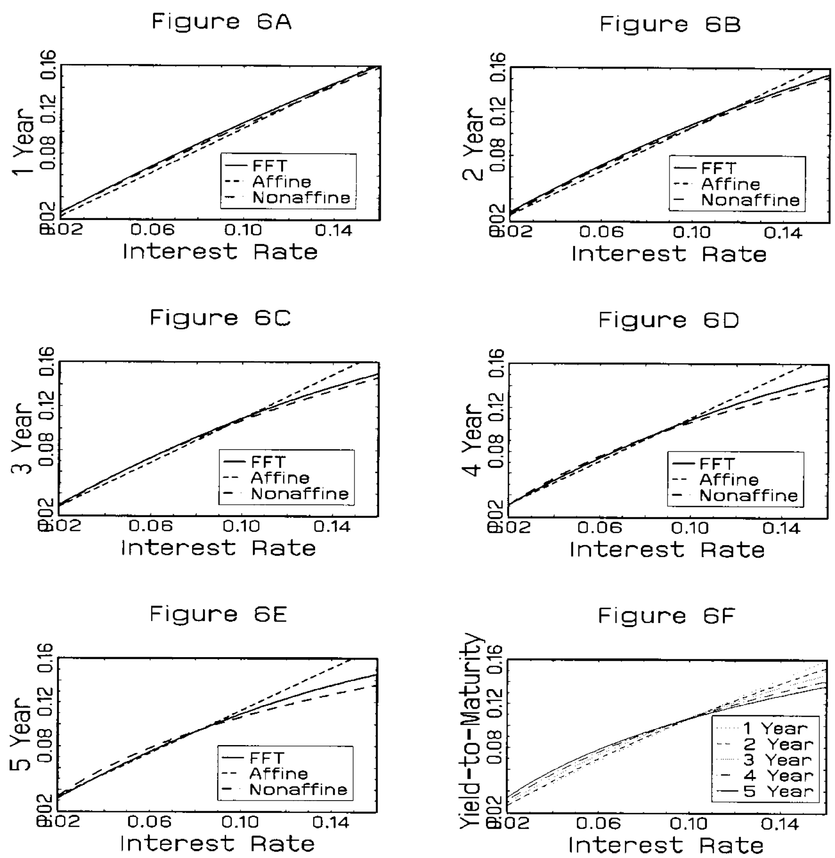


Figure 6
(A–E) Compare the FFF estimation and the calibrated model estimation (based on the two-step method) of the relationship between the interest rate and the yield with various maturities over the period from 1960 to 1991. (F) Represents the estimated yields of the model over different maturities.

the overall estimation procedure is a two-pass scheme wherein the structural parameters of dynamics of the interest rate are estimated in the time-series of the interest rate (in Section 2), and then the market prices of the factor risk are estimated from the cross-sectional data. In this sense, the estimation scheme is similar to that of Fama and MacBeth (1974) and Roll and Ross (1980).

To test the DK model we require the bond price, which can be expressed as

$$V(r, t, T)^{DK} = A(T-t)\exp\left[-B(T-t)r(t)\right], \tag{20}$$

where

$$
\begin{aligned}
A(T-t) &= \exp\left[\frac{\alpha}{\beta}(T-t-B(T-t))\right] \\
&\quad \times \left(\frac{2\psi e^{(\kappa+\lambda_2+\psi)(T-t)/2}}{(\kappa+\lambda_2+\psi)\left(e^{\psi(T-t)}-1\right)+2\psi}\right)^{\frac{2}{\beta}\left(\frac{\alpha}{\beta}(\kappa+\lambda_2)+\kappa\theta-\lambda_1\right)} \\
B(T-t) &= \frac{2\left(e^{\psi(T-t)}-1\right)}{(\kappa+\lambda_2+\psi)\left(e^{\psi(T-t)}-1\right)+2\psi} \\
\psi &= \sqrt{(\kappa+\lambda_2)^2+2\beta} \\
\lambda_1 &= \frac{\alpha}{\beta}\lambda_2.
\end{aligned}
$$

In the estimation, we use λ_2 as the free parameter.

Since these two models are nonnested, it is difficult to statistically test the relative performance of the two alternative models. Therefore our analysis is limited to the comparison of the TSSEs of the models. This is an indirect measure of the goodness-of-fit of the model and provides us with some information regarding the relative performance of the two models.

Table 5 documents the estimation results. The estimated parameter values of our model are $\lambda_1 = 0.0685$ (1946–1991) and 0.0380 (1960–1991), and $\lambda_2 = -2.2109$ (1946–1991) and -1.8073 (1960–1991). Given these parameter values of the market price of risk, we can analyze the risk-neutral distribution. First, $\kappa + \lambda_2 = 1.4330$ (1946–1991) and 1.6314 (1960–1991), so that the degree of the mean reversion or the nonlinearity in the risk-neutral drift is quite small relative to that under the original probability measure P. Second, the risk-neutral threshold rate $(\kappa\theta - \lambda_1)/(\kappa + \lambda_2)$ is 0.1614 (1946–1991) and 0.1512 (1960–1991), which are well above the counterpart under the original measure. We can also see that all these risk-neutral parameter values are quite stable.

In contrast, the estimated λ_2 of the DK model is -0.0957 in the 1946–1991 period and -0.1540 in the 1960–1991 period. These values, combined with the values in Table 3, indicate that the risk-neutral mean reversion parameter of the DK model is -0.0349 (1946–1991) and -0.0193 (1960–1991). Apparently these parameter values are quite unstable over different samples. More importantly, these negative mean-reversion values violate the technical condition required in the DK model for $-\alpha/\beta$ to be inaccessible in both real-measure and risk-neutral measure.

The TSSE criterion also favors our model. Our model has TSSEs of 0.1857 (1946–1991) and 0.1695 (1960–1991), whereas the DK model has TSSEs of 0.2156 (1946–1991) and 0.1841 (1946–1991). Therefore our

Table 5
Estimates of risk premium parameters

Model	λ_1	λ_2	Deviation	1 Year	2 Year	3 Year	4 Year	5 Year	TSSE
1946–1991									
DK		$-.0957$	Mean	$4.00\ 10^{-3}$	$2.92\ 10^{-3}$	$1.27\ 10^{-3}$	$-.53\ 10^{-3}$	$-2.23\ 10^{-3}$	.2156
		(.0022)	Var	$3.40\ 10^{-5}$	$6.03\ 10^{-5}$	$8.05\ 10^{-5}$	$9.43\ 10^{-5}$	$1.06\ 10^{-4}$	
AG	.0685	-2.2109	Mean	$1.08\ 10^{-3}$	$1.88\ 10^{-3}$	$.93\ 10^{-3}$	$-.30\ 10^{-3}$	$-1.56\ 10^{-3}$	.1857
	(.0025)	(.0267)	Var	$3.28\ 10^{-5}$	$5.28\ 10^{-5}$	$6.99\ 10^{-5}$	$8.54\ 10^{-5}$	$1.01\ 10^{-4}$	
FFF			Mean	.00	.00	.00	.00	.00	.1721
			Var	$2.28\ 10^{-5}$	$5.26\ 10^{-5}$	$6.93\ 10^{-5}$	$8.15\ 10^{-5}$	$.91\ 10^{-4}$	
1960–1991									
DK		$-.1540$	Mean	$4.51\ 10^{-3}$	$3.18\ 10^{-3}$	$1.29\ 10^{-3}$	$-.68\ 10^{-3}$	$-2.58\ 10^{-3}$	.1841
		(.0029)	Var	$4.00\ 10^{-5}$	$7.32\ 10^{-5}$	$9.75\ 10^{-5}$	$1.15\ 10^{-4}$	$1.29\ 10^{-4}$	
AG	.0380	-1.8073	Mean	$1.37\ 10^{-3}$	$2.00\ 10^{-3}$	$.83\ 10^{-3}$	$-.58\ 10^{-3}$	$-1.90\ 10^{-3}$	.1695
	(.0034)	(.0404)	Var	$3.89\ 10^{-5}$	$6.93\ 10^{-5}$	$9.22\ 10^{-5}$	$1.12\ 10^{-4}$	$1.31\ 10^{-4}$	
FFF			Mean	.00	.00	.00	.00	.00	.1588
			Var	$3.83\ 10^{-5}$	$6.91\ 10^{-5}$	$9.11\ 10^{-5}$	$1.07\ 10^{-5}$	$1.20\ 10^{-4}$	

The risk premium parameters, λ_1 and λ_2, are estimated by minimizing the squared deviations between the actual yields and the respective predictive values of each model. The standard errors are in parentheses. TSSE stands for the total sum of squared errors, and the mean and var of deviation denote the mean and variance of deviations of each yield to maturity.

model can reduce the pricing error of the DK model by about 8% to 14%, which is economically meaningful, and indicates that our model outperforms the DK model in explaining the cross-sectional variation in bond prices. More specifically, the *mean* pricing error of our model is far smaller than that of the DK model regardless of maturity. For example, the mean deviation of the one-year yield is 1.08×10^{-3} in our model and 4.00×10^{-3} in the DK model for 1946–1991, whereas that of the five-year yield is -1.56×10^{-3} in our model compared to -2.23×10^{-3} in the DK model. Further, the variance of the deviation is smaller in our model than in the DK model over the entire maturity spectrum and for different samples. Therefore we can conclude that the superiority of our model is not induced by bond subsets but by all bonds in our sample. For comparison purposes, Table 5 also reports a TSSE, and the mean and variance of the pricing error of the FFF series expansion estimation of the previous section. The FFF series expansion has a TSSE of 0.1721 (1946–1991) and 0.1588 (1960–1991), which is lower than in our model. However, note that the FFF series expansion method is an unrestricted estimation, which consists of 15 parameters, and our model has only two parameters. By adding 13 more parameters, one can reduce the pricing error by only about 6% to 7%. Since the system of the FFF expansion series estimation is exactly identified, it has zero mean pricing error by construction. On the other hand, the variance of the deviation in the FFF series expansion is quite close to our model regardless of the maturity. Therefore we can conclude that our model can capture the nonlinear function estimated by the FFF method.

One may argue that the above test is unfair to the DK model, because the DK model has only one free parameter λ_2, while our model has two free parameters, λ_1 and λ_2. For the robustness of the result, we perform a cross-sectional test in which we ignore the time-series information altogether. In particular, we define $\widehat{\kappa\theta} = \kappa\theta - \lambda_1$ and $\hat{\kappa} = \kappa + \lambda_2$. Instead of estimating λ's, we estimate $\widehat{\kappa\theta}$, $\hat{\kappa}$, and σ for our model and $\widehat{\kappa\theta}$, $\hat{\kappa}$, α, and β for the DK model. This estimation favors the DK model rather than ours in TSSE because there is one more parameter in the DK model. Table 6 shows the estimation result. The first thing to notice is that both models show improvements with lower TSSE's due to relaxed restrictions. However, the violation of technical condition still exists for the DK model in the 1946–1991 dataset. On the TSSE front, the performance of the DK model lags our model by about 5 to 10%, although it has one more free parameter. It is also worth pointing out that the TSSE of the unrestricted DK model (four parameters) is actually about 1% to 7% higher than our restricted two-step estimation (two parameters) model's TSSE, another indication of our model's superiority. This indicates the limitation of the affine-class model since it cannot capture the nonlinearities in the data by construction no matter how generalized.

Figures 5A–E and 6A–E illustrate the comparison of the nonlinear curves estimated by the FFF method and the fitted curves of our model and the DK model based on the two-step method for different times to maturity for 1946–1991 and 1960–1991, respectively. Consistent with the results in Table 5, our model is quite close to the function estimated by the FFF series expansion method. The DK model, however, is very different from these nonlinear functions, particularly at the high interest rates. One simple way to see this is the levels of yields at an interest rate of 14% in Figure 5A-E. It is clear that the DK model predicts that the higher the maturity, the higher the yield: it predicts an upward sloping yield curve at a 14% interest rate! In contrast, our model predicts that yield curve is downward sloping. For this period, there are nine cases where the interest rate is greater than 14%, and all nine cases show a downward sloping yield curve. A graphical check expanding the maturity to 30 years, shows that the predicted yield curve of the DK model for this range of the interest rate is actually a hump shape, but the shape for 1 to 5 years rests in the upward region. This slow mean-reverting behavior of the yield curve is consistent with the small mean-reversion parameter estimate of the DK model. Since the affine-class model cannot fit the nonlinear relationship between the interest rate and yields with two parameters, it has to sacrifice explanatory power in either the range of low or high levels of the interest rate. Since the data is more heavily populated in the low levels, it will sacrifice explanatory power in the range of the high levels by reducing the mean reversion. This is similar to the problem of the affine-class model in describing the dynamics of the interest rate. In contrast, our model can generate the nonlinear shape of the relationship, and the two parameters are effective in capturing the nonlinearity implied by the data.

Table 6
Cross-sectional estimation

Model	$\widehat{\kappa\theta}$	$\hat{\kappa}$	α	β	Deviation	1 Year	2 Year	3 Year	4 Year	5 Year	TSSE
1946–1991											
DK	.0099	−.0374	.0030	.0327	Mean	$1.19\ 10^{-3}$	$.05\ 10^{-3}$	$-.91\ 10^{-3}$	$-.68\ 10^{-3}$	$.70\ 10^{-3}$	.2012
	(.0008)	(.0139)	(.0006)	(.0100)	Var	$3.34\ 10^{-5}$	$6.01\ 10^{-5}$	$8.06\ 10^{-5}$	$9.43\ 10^{-5}$	$1.06\ 10^{-4}$	
AG	.1963	.9293	1.5915		Mean	$.26\ 10^{-3}$	$1.29\ 10^{-3}$	$.67\ 10^{-3}$	$.02\ 10^{-3}$	$-.05\ 10^{-3}$	.1829
	(.0059)	(.0753)	(.0398)		Var	$3.30\ 10^{-5}$	$5.37\ 10^{-5}$	$7.01\ 10^{-5}$	$8.48\ 10^{-5}$	$1.01\ 10^{-4}$	
1960–1991											
DK	.0167	.0388	.0054	.0118	Mean	$2.02\ 10^{-3}$	$-.02\ 10^{-3}$	$-1.03\ 10^{-3}$	$-.75\ 10^{-3}$	$.74\ 10^{-3}$	.1717
	(.0010)	(.0129)	(.0008)	(.0094)	Var	$4.00\ 10^{-5}$	$7.32\ 10^{-5}$	$9.75\ 10^{-5}$	$1.15\ 10^{-4}$	$1.29\ 10^{-4}$	
AG	.2019	.9801	1.5950		Mean	$.54\ 10^{-3}$	$1.13\ 10^{-3}$	$.30\ 10^{-3}$	$-.30\ 10^{-3}$	$-.57\ 10^{-3}$	.1661
	(.0078)	(.0970)	(.0489)		Var	$3.88\ 10^{-5}$	$7.05\ 10^{-5}$	$9.20\ 10^{-5}$	$1.11\ 10^{-4}$	$1.19\ 10^{-4}$	

The risk-neutral parameters, $\widehat{\kappa\theta}$, $\hat{\kappa}$, and σ for our model and $\widehat{\kappa\theta}$, $\hat{\kappa}$, α, and β for the DK model are estimated by minimizing the squared deviations between the actual yields and the respective predictive values of each model. The standard errors are in parentheses. TSSE stands for the total sum of squared errors, and the mean and var of deviation denote the mean and variance of deviations of each yield to maturity.

5. Conclusion

This article presents a new nonlinear term structure model which can generate dynamics of the interest rate and the market price of risk close to empirical findings documented in Chan et al. (1992), Aït-Sahalia (1996a,b), Conley et al. (1997), and Stanton (1997). By combining nonlinearities in the diffusion process of the interest rate and the market price of risk, we can resolve the problems raised in Campbell, Lo, MacKinlay (1996). The model provides a stationary interest rate process and a closed-form expression for transition and marginal density of the interest rate and bond prices. In addition, the model implies a nonlinear relationship between the interest rate and yields to maturity, which places it beyond the affine-class models. The empirical analysis indicates that our model can dominate the affine-class term structure models in both the time series as well as the cross-sectional dimension. Therefore this model is a better candidate for the hedging or pricing of interest rate contingent claims.

Many extensions are possible. First, the model could be applied toward the valuation of other interest rate sensitive securities such as caps and floors. Given that our model has better performance than affine-class models under the true measure as well as the risk-neutral measure, hedging schemes based on our model are likely to perform better. A horse-race competition of this model to the affine-class models in this dimension seems to be an important task to better our understanding of interest rate behavior.

Second, if one is interested in multifactor models, this model can be extended to a multifactor framework. The extension can be made in a similar manner to the multifactor extensions of the CIR (1985) model as in Chen and Scott (1992), Longstaff and Schwartz (1992), and Pearson and Sun (1992). Also, one could consider a hybrid multifactor model where some factors are affine factors while other factors are nonaffine factors of our model. Such an approach can broaden our understanding regarding the best mixture of factors. These issues represent interesting extensions for future research.

Appendix A: Boundary Conditions

This appendix applies the analysis of Aït-Sahalia (1996b) concerning the constraints on the drift and the diffusion to our model. Therefore, for more extensive analysis regarding the conditions for the general process, refer to the appendix of his study. We first impose constraints on the signs of the parameters to ensure that the volatility $\sigma^2(\cdot) \geq 0$ on $D = (0, \infty)$ and $\lim_{r\downarrow 0} \mu(r) > 0$, $\lim_{r\uparrow\infty} \mu(r) < 0$ for the drift. For our process, the necessary condition is $\kappa > 0$. Further, the following two conditions should be met:

(i) The speed measure M, which is defined as the integral of

$$m(r) = \frac{\exp\left[-\int_r^{\bar{\epsilon}} 2\frac{\kappa\theta x - \kappa x^2}{\sigma^2 x^3}\, dx\right]}{\sigma^2 r^3},$$

converges at both boundaries of D.

(ii) The scale measure S, which is defined as the integral of

$$s(r) = \exp\left[\int_r^{\bar{\epsilon}} 2\frac{\kappa\theta x - \kappa x^2}{\sigma^2 x^3}\, dx\right]$$

diverges at both boundaries of D.

Note that $\bar{\epsilon}$ is a fixed point in D and its particular choice is irrelevant. Condition (i), along with $\sigma^2 r^3 > 0$ on D guarantees the existence of a unique strong solution. We first examine the boundary behavior of the speed and scale measures near zero. We can see immediately, $\mu(r)/\sigma^2(r) \propto \frac{\kappa\theta}{\sigma^2 r^2}$, so $m(r) \propto \exp[-\frac{2\kappa\theta}{\sigma^2 r}]/(\sigma^2 r^3)$ and $s(r) \propto \exp[\frac{2\kappa\theta}{\sigma^2 r}]$. $\theta > 0$ suffices these conditions along with $\kappa > 0$, so that $zero$ is not attainable.

Near infinity, $\mu(r)/\sigma^2(r) \propto -\kappa/(\sigma^2 r)$, so $m(r) \propto (r^{-2\kappa/\sigma^2-3})/\sigma^2$ and $s(r) \propto r^{2\kappa/\sigma^2}$, which requires $\kappa > 0$ again.

Therefore the necessary and sufficient conditions for stationarity and unattainability of 0 and ∞ in finite expected time are the pairs $\kappa > 0$ and $\theta > 0$.

Appendix B: The Analysis of the Distribution of r

The transition density of r

Define $x(t) \equiv 1/r(t)$. Applying Ito's lemma to Equation (6) yields

$$dx(t) = \left[\kappa + \sigma^2 - \kappa\theta x(t)\right] dt - \sigma\sqrt{x(t)}dW,$$

where the boundary condition is that $\frac{2(\kappa+\sigma^2)}{\sigma^2} > 1$, which holds since $\kappa > 0$ as shown in Appendix A. Since $x(t)$ follows the well-known square-root process, the conditional distribution of $x(t)$ is the noncentral chi square, $\chi^2[2\alpha x(s); 2q+2, 2u]$, with $2q+2$ degrees of freedom and parameter of noncentrality $2u$ proportional to the current spot rate. The standard transformation method of the distribution yields

$$\begin{aligned} f(r(s), s|r(t), t) &= f(x(r(s)), s|x(r(t)), t)\frac{\partial x(r(s))}{\partial r(s)} \\ &= \alpha e^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q\left[2(uv)^{1/2}\right]\left(\frac{1}{r(s)^2}\right) \\ &= \alpha e^{-u-v}\left(\frac{v}{u}\right)^{q/2} I_q\left[2(uv)^{1/2}\right]\alpha^{-2}v^2 \\ &= \alpha^{-1}e^{-u-v}v^{q/2+2}u^{-q/2}I_q\left[2(uv)^{1/2}\right] \end{aligned} \tag{21}$$

The conditional moments for r

Here we derive the conditional moments for the interest rate following the process of Equation (6). That is, we must compute

$$\begin{aligned} E\left[r(s)^\nu|r(t)\right] &= \int_0^\infty r(s)^\nu f(r(s), s|r(t), t)\, dr(s) \\ &= \int_0^\infty r(s)^\nu \frac{1}{\alpha}e^{-u-v}v^{q/2+2}u^{-q/2}I_q\left[2(uv)^{1/2}\right] dr(s) \end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \alpha^\nu e^{-u-v} v^{q/2-\nu} u^{-q/2} I_q \left[2(uv)^{1/2}\right] dv \\
&= \alpha^\nu e^{-u} \sum_{p=0}^{\infty} \frac{u^p}{\Gamma(p+1)\Gamma(p+q+1)} \int_0^\infty e^{-v} v^{p+q-\nu}\, dv \\
&= \alpha^\nu e^{-u} \sum_{p=o}^{\infty} \frac{u^p \Gamma(p+q+1-\nu)}{\Gamma(p+1)\Gamma(p+q+1)} \\
&= \alpha^\nu e^{-u} \frac{\Gamma(1+q-\nu)}{\Gamma(1+q)} M(1+q-\nu, 1+q, u).
\end{aligned}
$$

The marginal density of r
The transition density can be written as

$$
\begin{aligned}
f(r(s), s|r(t), t) &= \alpha^{-1} e^{-u-v} v^{q/2+2} u^{-q/2} I_q \left[2(uv)^{1/2}\right] \\
&= \alpha^{-1} e^{-u-v} v^{q/2+2} u^{-q/2} (uv)^{q/2} \sum_{p=0}^{\infty} \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)} \\
&= \alpha^{-1} e^{-u-v} v^{q+2} \sum_{p=0}^{\infty} \frac{(uv)^p}{\Gamma(p+1)\Gamma(p+1+q)}.
\end{aligned}
$$

If $s \to \infty$,

$$
\alpha \to \tfrac{2\kappa\theta}{\sigma^2} \quad u \to 0 \quad v \to \tfrac{\alpha}{r} = \tfrac{2\kappa\theta}{\sigma^2 r}.
$$

Consequentlythe steady-state density function is

$$
\begin{aligned}
\pi(r) &= f(r, \infty|r(t), t) \\
&= \frac{\alpha^{-1} e^{-v} v^{q+2}}{\Gamma(q+1)} \\
&= \frac{\frac{\sigma^2}{2\kappa\theta} \exp\left(-\frac{2\kappa\theta}{\sigma^2 r}\right)\left(\frac{2\kappa\theta}{\sigma^2 r}\right)^{q+2}}{\Gamma\left(\frac{2\kappa}{\sigma^2}+2\right)} \\
&= \frac{1}{\Gamma(2+\frac{2\kappa}{\sigma^2})} \left(\frac{2\kappa\theta}{\sigma^2}\right)^{2+\frac{2\kappa}{\sigma^2}} r^{-\left(3+\frac{2\kappa}{\sigma^2}\right)} \exp\left(-\frac{2\kappa\theta}{\sigma^2 r}\right).
\end{aligned}
$$

The unconditional moments for r
We compute

$$
\begin{aligned}
E\left[r^\nu\right] &= \int_0^\infty r^\nu \pi(r(t))\, dr \\
&= \int_0^\infty r^\nu \frac{1}{\Gamma\left(2+\frac{2\kappa}{\sigma^2}\right)} \left(\frac{2\kappa\theta}{\sigma^2}\right)^{2+\frac{2\kappa}{\sigma^2}} r^{-\left(3+2\frac{\kappa}{\sigma^2}\right)} \exp\left(-\frac{2\kappa\theta}{\sigma^2 r}\right) dr.
\end{aligned}
$$

Define $x(t) \equiv \frac{2\kappa\theta}{\sigma^2 r}$. Then the moment equation can be rewritten as

$$\begin{aligned} E\left[r^{\nu}\right] &= \frac{\left(\frac{2\kappa\theta}{\sigma^2}\right)^{\nu}}{\Gamma\left(2+\frac{2\kappa}{\sigma^2}\right)} \int_0^{\infty} x^{1+\frac{2\kappa}{\sigma^2}-\nu} e^{-x}\, dx \\ &= \frac{\left(\frac{2\kappa\theta}{\sigma^2}\right)^{\nu}}{\Gamma\left(2+\frac{2\kappa}{\sigma^2}\right)} \Gamma\left(2+\frac{2\kappa}{\sigma^2}-\nu\right). \end{aligned}$$

Appendix C: Solution to the PDE

We will prove the solution to athe PDE [Equation (12)] is

$$V(r,t,T) = \frac{\Gamma(\beta-\gamma)}{\Gamma(\beta)} M\left(\gamma, \beta, -x(r,t,T)\right) x(r,t,T)^{\gamma}.$$

Start with PDE [Equation (12)]:

$$\frac{1}{2}\sigma^2 r^3 V_{rr} + r((\kappa\theta-\lambda_1)-(\kappa+\lambda_2)r)V_r - rV + V_t = 0.$$

Define $y = a(t)/r$ as the characteristic function. We want to try the solution

$$V(r,t) = Af(y)y^{\gamma}, \tag{22}$$

where γ and A are constants, $f(y)$ and $a(t)$ are functions, of which the values and functional forms are to be determined.

Given Equation (22), we can calculate the various derivatives,

$$\begin{aligned} \frac{1}{A}V_r &= -\frac{y^{\gamma+2}}{a(t)} f'(y) - \frac{\gamma y^{\gamma+1}}{a(t)} f(y) \\ \frac{1}{A}V_{rr} &= \frac{y^{\gamma+4}}{a^2(t)} f''(y) + \frac{2(\gamma+1)y^{\gamma+3}}{a^2(t)} f'(y) + \frac{\gamma(\gamma+1)y^{\gamma+2}}{a^2(t)} f(y) \\ \frac{1}{A}V_t &= \frac{a'(t)}{a(t)} y^{\gamma+1} f'(y) + \frac{a'(t)}{a(t)} \gamma y^{\gamma} f(y). \end{aligned}$$

Substitute the derivatives into Equation (12), and collect the terms:

$$\begin{aligned} &\frac{1}{2}\sigma^2 a(t) y^{\gamma+1} f''(y) + \Bigg(\frac{a'(t)}{a(t)} y^{\gamma+1} + \sigma^2(\gamma+1)a(t)y^{\gamma} \\ &\qquad - (\kappa\theta-\lambda_1)y^{\gamma+1} + (\kappa+\lambda_2)a(t)y^{\gamma}\Bigg) f'(y) \\ &+ \Bigg(\frac{a'(t)}{a(t)}\gamma y^{\gamma} + \frac{1}{2}\sigma^2\gamma(\gamma+1)a(t)y^{\gamma-1} - (\kappa\theta-\lambda_1)\gamma y^{\gamma} \\ &\qquad + (\kappa+\lambda_2)\gamma a(t) y^{\gamma-1} - a(t)y^{\gamma-1}\Bigg) f(y) = 0. \end{aligned}$$

The above equation has mixed ordinary differentials on both t and y. To reduce the above equation to an ordinary differential equation in y, two conditions need to be satisfied:

$$\frac{a'(t)}{a(t)} - (\kappa\theta - \lambda_1) = a(t)$$

$$\frac{1}{2}\sigma^2\gamma(\gamma + 1) + (\kappa + \lambda_2)\gamma - 1 = 0$$

In the case that the above two equations are satisfied, the PDE is reduced to an ODE, given $y \neq 0$,

$$\frac{1}{2}\sigma^2 y f''(y) + \Big((\kappa + \lambda_2 + (\gamma + 1)\sigma^2) + y\Big) f'(y) + \gamma f(y) = 0.$$

Further define $y = -\frac{1}{2}\sigma^2 z$, and change the variable for the above ODE, then

$$zg''(z) + (\beta - z)g'(z) - \gamma g(z) = 0, \tag{23}$$

where

$$\beta = \frac{2(\kappa + \lambda_2 + (1 + \gamma)\sigma^2)}{\sigma^2}.$$

Equation (23) is a standard confluent hypergeometric function, and the solution is

$$f(y) = g(z) = M(\gamma, \beta; z).$$

We have found the functional form of $f(y)$, we also need to find the functional form of $a(t)$ and the value of γ to complete the proof. To do this we solve the two equations required for thereduction of the PDE to the ODE.

Solving

$$\frac{a'(t)}{a(t)} - (\kappa\theta - \lambda_1) = a(t)$$

yields

$$a(t) = \frac{\kappa\theta - \lambda_1}{e^{(\kappa\theta - \lambda_1)(T-t)} - 1},$$

and solving

$$\frac{1}{2}\sigma^2\gamma(\gamma + 1) + (\kappa + \lambda_2)\gamma - 1 = 0$$

gives

$$\gamma = \frac{\sqrt{\phi^2 + 2\sigma^2} - \phi}{\sigma^2},$$

where

$$\phi = \kappa + \lambda_2 + \frac{1}{2}\sigma^2.$$

So the solution for the PDE is

$$V(r, t, T) = V(r, \tau) \; = \; A f(y) y^{\gamma}$$

$$\begin{aligned}
&= Ag\left(-\frac{2y}{\sigma^2}\right) y^{\gamma} \\
&= A\left(\frac{\sigma^2}{2}\right)^{\gamma} g\left(-\frac{2a(t)}{\sigma^2 r}\right)\left(\frac{2a(t)}{\sigma^2 r}\right)^{\gamma} \\
&= BM(\gamma, \beta; -x(r, t, T))x(r, t, T)^{\gamma},
\end{aligned}$$

where B is a constant to be determined and

$$x(r, t, T) = \frac{2(\kappa\theta - \lambda_1)}{\sigma^2(e^{(\kappa\theta-\lambda_1)(T-t)} - 1)r}.$$

To find B, we use the initial condition of $V(r, T, T) = 1$. By definition,

$$M(\gamma, \beta; -x) = \frac{\Gamma(\beta)}{\Gamma(\gamma)\Gamma(\beta - \gamma)} \int_0^1 e^{-xt} t^{\gamma-1}(1 - t)^{\beta-\gamma-1}\, dt.$$

As $t \to T$,

$$\lim_{t\to T} x(r, t, T) = \infty, \quad \lim_{t\to T} x(\tau)^{\gamma} = \infty \quad \lim_{t\to T} M(\gamma, \beta; -x) = \lim_{x\to\infty} M(\gamma, \beta; -x) = 0.$$

Then

$$\begin{aligned}
\lim_{t\to T} V(r, t, T) &= \frac{B\Gamma(\beta)}{\Gamma(\gamma)\Gamma(\beta - \gamma)} \lim_{x\to\infty} x^{\gamma} \int_0^1 e^{-xt} t^{\gamma-1}(1 - t)^{\beta-\gamma-1}\, dt \\
&= \frac{B\Gamma(\beta)}{\Gamma(\gamma)\Gamma(\beta - \gamma)} \lim_{x\to\infty} \int_0^x e^{-y} y^{\gamma-1}(1 - \frac{y}{x})^{\beta-\gamma-1}\, dy \\
&= \frac{B\Gamma(\beta)}{\Gamma(\gamma)\Gamma(\beta - \gamma)} \int_0^{\infty} e^{-y} y^{\gamma-1}\, dy \\
&= \frac{B\Gamma(\beta)}{\Gamma(\beta - \gamma)} \\
&= 1 \\
B &= \frac{\Gamma(\beta - \gamma)}{\Gamma(\beta)}.
\end{aligned}$$

Finally, we prove that

$$V(r, t, T) = \frac{\Gamma(\beta - \gamma)}{\Gamma(\beta)} M(\gamma, \beta; -x(r, t, T))x(r, t, T)^{\gamma}.$$

References

Ahn, D.-H., 1997, “Generalized Squared Autoregressive Independent Variable Nominal Term Structure Model,” working paper, University of North Carolina, Chapel Hill.

Aït-Sahalia, Y., 1996a, “Nonparametric Pricing of Interest Rate Derivative Securities,” *Econometrica*, 64, 527–560.

Aït-Sahalia, Y., 1996b, “Testing Continuous-Time Models of the Spot Interest Rate,” *Review of Financial Studies*, 9, 385–426.

Aït-Sahalia, Y., 1997, "Do Interest Rates Really Follow Continuous-Time Markov Diffusions?," working paper, University of Chicago.

Andersen, T. G., and J. Lund, 1997, "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rates," *Journal of Econometrics*, 17, 343–377.

Beaglehole, D., and M. Tenney, 1992, "A Nonlinear Equilibrium Model of the Term Structure of Interest Rates: Corrections and Additions," *Journal of Financial Economics*, 32, 345–354.

Boudoukh, J., M. Richardson, R. Stanton, and R. Whitelaw, 1997, "The Stochastic Behavior of Interest Rates: Implications from a Nonlinear Continuous-Time Model," working paper, New York University.

Boudoukh, J., M. Richardson, and R. Whitelaw, 1994, "A Tale of Three Schools: Insights on Autocorrelations of Short-Horizon Stock Returns," *Review of Financial Studies*, 7, 539–573.

Brennan, M. J., and E. Schwartz, 1979, "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 32, 133–156.

Campbell, J. Y., A. W. Lo, A. C. MacKinlay, 1996, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, N.J.

Chan, K. C., G. A. Karolyi, F. A. Longstaff and A. B. Sanders, 1992, "An Empirical Comparison of Alternative Models of Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1228.

Chen, R. R., and L. Scott, 1992, "Pricing Interest Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure," *Review of Financial Studies*, 5, 613–636.

Conley T., L. Hansen, E. Luttmer, and J. Scheinkman, 1997, "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525–577.

Constantinides, G., 1992, "A Theory of the Nominal Structure of Interest Rates," *Review of Financial Studies*, 5, 531–552.

Courtadon, G., 1982, "The Pricing of Options on Default-Free Bonds," *Journal of Financial and Quantitative Analysis*, 17, 75–100.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–406.

Dothan, L. U., 1978, "On the Term Structure of Interest Rates," *Journal of Financial Economics*, 6, 59–69.

Duffie, D., 1992, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, N.J.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

Fama, E., and J. MacBeth, 1974,"Risk, Return and Equilibrium," *Journal of Political Economy*, 81, 607–636.

Feller, W., 1951, "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 173–182.

Gallant, A. R., 1981, "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211–245.

Harrison, M., and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, M., and S. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 215–260.

Litterman, R., and J. Scheinkman, 1991, "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

Longstaff, F. A., and E. S. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Marsh, T. A., and E. R. Rosenfeld, 1983, "Stochastic Processes for Interest Rates and Equilibrium Bond Prices," *Journal of Finance*, 38, 635–646.

McCulloch, J., and H. Kwon, 1993, "US Term Structure Data, 1947–1991," working paper, Ohio State University.

Newey, W., and K. West, 1987a, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W., and K. West, 1987b, "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–788.

Øksendal, B., 1992, *Stochastic Differential Equations* (2nd ed.), Springer-Verlag, New York.

Pagan, A., and Y. S. Hong, 1991, "Non-parametric Estimation and the Risk Premium," in W. Barnett, J. Powell, and G. Tauchen (eds.), *Semiparametric and Non-parametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge, U.K.

Pagan, A., and W. Shwert, 1990, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

Pearson, N. D., and T. S. Sun, 1994, "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model,"*Journal of Finance*, 49, 1279–1304.

Revuz, D., and M. Yor, 1990, *Continuous Martingales and Brownian Motion*, Springer-Verlag, New York.

Roll, R., and S. Ross, 1980, "An Empirical Investigation of the Arbitrage Pricing Model," *Journal of Finance*, 35, 1073–1103.

Schaefer, S. M., and E. S. Schwartz, 1987, "Time Dependent Variance and the Pricing of Bonds," *Journal of Finance*, 42, 113–128.

Stambaugh, R. F., 1988, "The Information in Forward Rates: Implications for Models of the Term Structure," *Journal of Financial Economics*, 21, 41–70.

Stanton, R., 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.