JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 44, No. 5, Oct. 2009, pp. 1197–1230

doi:10.1017/S0022109009990330

Nonparametric Estimation of the Short Rate Diffusion Process from a Panel of Yields

Abdoul G. Sam and George J. Jiang*

Abstract

In this paper, we propose a nonparametric estimator of the short rate diffusion process using observations of a panel of yields. The proposed estimator can greatly reduce the bias of the nonparametric estimator proposed in Stanton (1997) that uses a single time series of short rate observations. Simulations confirm that the new method significantly attenuates the spurious nonlinearity of the drift function as documented in Chapman and Pearson (2000). We apply the method to estimate the U.S. short rate process using a panel of six Treasury yields. With 42 years' daily observations of the panel of yields, the proposed drift function estimator achieves the same efficiency as the Stanton (1997) estimator based on 145 years of daily short rate observations. Finally, we show that the proposed estimator also has significant economic implications on the pricing of bonds and interest rate derivatives.

I. Introduction

The finance literature has devoted substantial effort to modeling the short-term interest rate, which is generally believed to be the most important state variable driving the dynamics of interest rate term structure. Continuous-time univariate models of the short rate, r_t, are typically specified as the following time-homogenous Itô diffusion process:

$$dr_t \quad = \quad \mu(r_t)dt + \sigma(r_t)dw_t,$$

where w_t is the standard Brownian motion with $t \in [0, T]$, and $\mu(r_t)$ and $\sigma(r_t)$ are, respectively, the drift and diffusion functions. Most existing interest rate models are nested in the parametric specification of Aït-Sahalia (1996b):

$$dr_t \quad = \quad (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1})dt + \sigma r_t^{\gamma} dw_t,$$

*Sam, sam.7@cfaes.osu.edu, Department of Agricultural, Environmental, and Development Economics, Ohio State University, 2120 Fyffe Rd., Columbus, OH 43210; Jiang, gjiang@email.arizona.edu, Eller College of Management, University of Arizona, PO Box 210108, Tucson, AZ 85721. We thank Stephen Brown (the editor), Tyler Brough, Stephen Cosslett, Jianqing Fan, Alfonso Flores-Lagunes, Kei Hirano, Raymond Kan, Alan Ker, Chris Lamoureux, Amilcar Menichini, Neil Pearson (associate editor and referee), Matt Pritsker, Kenneth Roskelley, and seminar participants at the University of Arizona, the Ohio State University, the World Congress of Econometrics Society in London, U.K., and the Conference of Financial Econometrics in Montreal for helpful comments and suggestions. The usual disclaimer applies.

where both the drift and diffusion functions are specified to capture potential nonlinearities. Restrictions on parameter values of the above model lead to the Vasicek (1977) model ($\alpha_2 = \alpha_3 = 0, \gamma = 0$), the Brennan and Schwartz (1979) and the Courtadon (1982) model ($\alpha_2 = \alpha_3 = 0, \gamma = 1$), the Cox, Ingersoll, and Ross (CIR) (1985) model ($\alpha_2 = \alpha_3 = 0, \gamma = 1/2$) or the translated CIR (1985) model in Pearson and Sun (1994), the constant elasticity of volatility (CEV) model of Cox (1975) and Cox, Ingersoll, and Ross (1980) ($\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0, \gamma = 3/2$), and the CEV models of Chan, Karolyi, Longstaff, and Sanders (1992) ($\alpha_2 = \alpha_3 = 0$), among others.

While the main appeal of parametric models is the tractability and potential closed-form pricing formula for bonds and interest rate derivative securities, the downside is the risk of model misspecification. Empirical studies have provided strong evidence rejecting most of the popular parametric diffusion models for the short rate (see empirical tests in, e.g., Chan et al. (1992), Aït-Sahalia (1996b), Andersen and Lund (1997), and Hong and Li (2005)). In addition, some studies have found that misspecified models can have significant economic implications on the pricing of interest rate derivative securities (see, e.g., Backus, Foresi, and Zin (1998), Canabarro (1995)).

For the above reasons, nonparametric modeling of short rate dynamics has received considerable attention in recent years. Building on developments in the statistics literature, various nonparametric estimators of the drift and diffusion functions have been proposed in the finance literature. Aït-Sahalia (1996a) proposes a nonparametric estimator of the diffusion function from discretely observed data with a parametric drift function. Jiang and Knight (1997) extend Florens-Zmirou (1993) and Banon (1978) and propose nonparametric kernel estimators of the drift and diffusion functions of a stationary process via the nonparametric estimator of the marginal density. Using the infinitesimal generator and Taylor series expansion, Stanton (1997) proposes nonparametric estimators of the drift and diffusion functions based on various orders of approximation of the Itô process. Bandi and Phillips (2003) further generalize the nonparametric approach to recurrent diffusion processes, circumventing the assumption of stationarity of the short rate process.

Empirical evidence from model specification tests in Aït-Sahalia (1996b) and the nonparametric drift function estimates in Stanton (1997) and Jiang (1998) suggest that the drift function of the short rate is nonlinear. In particular, the nonparametric test in Aït-Sahalia (1996b) provides evidence that the assumption of a linear drift function is the principal source of model misspecification. The nonparametric drift function estimates in Stanton (1997) and Jiang (1998) share the feature that the short rate exhibits very little mean reversion or behaves like a random walk below the 14% level but has a dramatic mean reversion beyond that. Conley, Hansen, Luttmer, and Scheinkman (1997) report similar results where the estimated drift function is nonzero only for rates below 3% or above 11%.

The findings of a nonlinear drift have, however, been challenged by Pritsker (1998) and Chapman and Pearson (2000). Simulations in Chapman and Pearson (2000) show that the nonparametric drift function estimator proposed in Stanton (1997) can produce spurious nonlinearities even when the underlying drift function is truly linear. More troublesome is the fact that spurious nonlinearity has a pattern similar to the empirical estimate of drift function in Stanton (1997) and Jiang (1998). Chapman and Pearson (2000) argue that a combination of the "truncation" of the observed short rates and a finite sample creates artificial nonlinearity near the boundaries of the support. Abhyankar and Basu (2001) and Li, Pearson, and Poteshman (2004) provide further support to the argument in Chapman and Pearson (2000). They show that if the truncation of the observed short rate process is accounted for, the resulting drift is nonlinear, even if the drift of the unrestricted process is linear.

The literature shows that while increasing the sampling frequency is helpful for the identification and estimation of the diffusion function, it is the increase in the sampling period that is crucial for the identification and estimation of the drift function. Evidence in Pritsker (1998), Chapman and Pearson (2000), and Jones (2003) suggests that as a result of the strong persistence in interest rates, identifying the drift function requires a large number of observations for a given sampling frequency or, equivalently, a long sample period. The ongoing debate about the linearity or nonlinearity of the drift function simply underscores the difficulty of identifying and estimating the drift function.

In this paper, we propose a new nonparametric estimation method for the short rate process. The key difference between the proposed estimator and existing parametric and nonparametric estimators of the short rate diffusion is that instead of using a single time series of short rate observations, the new method uses observations of a panel of yields with different maturities. Our proposed method is implemented in two steps. In the first step, we pool all the yields together and obtain a nonparametric pooled estimator of the drift function. Ideally, if the drift functions of yields with different maturities are identical, then pooling the data results in the optimal drift function estimator. In reality, while the drift functions of yields with different maturities may be similar in functional shape, they are unlikely to be identical. Therefore, in the second step we correct the bias of the pooled estimator using a nonparametric correction factor. The advantage of the two-step procedure is that when the pooled drift function estimator is similar to the drift function of the short rate in functional shape, the correction factor is a flat and smooth function, and thus much easier to estimate nonparametrically. We show that as long as the correction factor is smoother than the true short rate drift function, the proposed approach can reduce the bias of the drift function estimator proposed in Stanton (1997). Simulations confirm that using a panel of yields with different maturities, the proposed estimator can lead to significant efficiency gain relative to using a single time series of short rate observations. In particular, spurious nonlinearities or biases toward the boundaries are substantially reduced.

We apply the proposed method to estimate the U.S. short rate process. In our analysis, we use daily observations of the 3-month T-bill yield over more than 50 years from January 1954 to November 2004, as opposed to the 20 to 30 years of data used in most existing studies. In addition, we use five extra series of bond yields with maturities ranging from 6 months to 10 years. Each of these additional series has 42 years of daily observations. Through simulations and using the mean integrated squared error (MISE) as a yardstick of estimation efficiency, we find that as a result of using the panel of yields, the new drift function estimator achieves the same efficiency as the nonparametric estimator, proposed in Stanton (1997), based on 145 years of daily short rate observations. Our empirical results suggest that the short rate drift function is nonlinear at high interest rate levels. However, the mean reversion is significantly weaker than that documented by Stanton (1997) and Jiang (1998). More importantly, the difference in the drift and diffusion function estimates has significant economic implications on the pricing of interest rate derivatives.

The remainder of the paper is structured as follows. In Section II, we propose a new nonparametric estimator of the short rate diffusion process and derive its statistical properties. Monte Carlo simulations are performed in Section III to assess the finite sample bias of the proposed estimator. In Section IV, an empirical application is undertaken using U.S. interest rate data. Section V concludes.

II. Nonparametric Estimation of the Short Rate Diffusion from a Panel of Yields

We consider the following diffusion process for the short rate $\{r_t^{(1)}, t \geq 0\}$:

$$(1) \qquad dr_t^{(1)} = \mu_1(r_t^{(1)})dt + \sigma_1(r_t^{(1)})dw_t^{(1)} \qquad \text{(the model)},$$

where $w_t^{(1)}$ is a standard Brownian motion and $\mu_1(\cdot)$ and $\sigma_1^2(\cdot)$ are, respectively, the drift and diffusion functions. In addition to the above short rate process, we assume that there are $J-1$ additional interest rates $\{r_t^{(j)}, t \geq 0\}$, $j = 2, \ldots, J$, which follow the diffusion processes:

$$(2) \qquad dr_t^{(j)} = \mu_j(r_t^{(j)})dt + \sigma_j(r_t^{(j)})dw_t^{(j)} \qquad \text{(the auxiliary models)},$$

where $\mu_j(\cdot)$ and $\sigma_j^2(\cdot)$ are, respectively, the drift and diffusion functions of $r_t^{(j)}$, and the standard Brownian motions $w_t^{(j)}$, $j = 1, \ldots, J$, are potentially correlated. We term the short rate model in equation (1) as "the model," since it is the model of interest to us and the one we intend to estimate, and the models in equation (2) as "the auxiliary models." These models, as illustrated later in the paper, are used only to improve the estimation of "the model." If the short rate ($r_t^{(1)}$) is taken to be, say, the yield of the 3-month Treasury bill, then the auxiliary rates could be the yields with maturities longer than 3 months, including the yields on Treasury bills, notes, and bonds. It is noted that the only restriction on the auxiliary model is that the state variable $r_t^{(j)}$ also follows a diffusion process. Without loss of generality, in what follows, all the realized rates are assumed to be equispaced over the time period $[0, T]$, with $\delta = T/n$ being the sampling interval.

Various nonparametric estimators of the drift and diffusion functions have been proposed in the finance literature. In particular, using the infinitesimal generator and Taylor series expansion, Stanton (1997) proposes nonparametric estimators of the drift and diffusion functions based on various orders of approximation of the Itô process. With the discretization interval $\delta > 0$, Stanton (1997) proposes the following nonparametric estimators of the drift and diffusion functions of equation (1) based on the first-order approximation of the discretized process:

$$(3)\qquad \tilde{\mu_1}(r) = \frac{1}{\delta}\frac{\sum_{t=0}^{n-1}\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)K_h\left(r_{t\delta}^{(1)} - r\right)}{\sum_{t=0}^{n-1}K_h\left(r_{t\delta}^{(1)} - r\right)} \qquad \text{and}$$

$$(4)\qquad \tilde{\sigma_1^2}(r) = \frac{1}{\delta}\frac{\sum_{t=0}^{n-1}\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)^2 K_h\left(r_{t\delta}^{(1)} - r\right)}{\sum_{t=0}^{n-1}K_h\left(r_{t\delta}^{(1)} - r\right)},$$

where $K_h(u) = (1/h)K(u/h)$ and $K(\cdot)$ is a kernel function that satisfies common regularity conditions listed in Appendix A. The statistical properties, in particular the asymptotic bias and variance, of the drift function estimator are given in the following proposition:

Proposition 1. Suppose that $\delta = \delta(T,n) \to 0$ and $h = h(T,n) \to 0$ as $T, n \to \infty$, and $Th \to \infty$, given the regularity conditions (see Appendix A) for the short rate process in equation (1) and the kernel function $K(u)$, the basic statistical properties of $\tilde{\mu_1}(r)$ are:

$$(5)\quad \mathrm{E}\left[\tilde{\mu_1}(r) - \mu_1(r)\right] \simeq \frac{h^2}{2}\left(\mu_1''(r) + 2\mu_1'(r)\frac{p_1'(r)}{p_1(r)}\right)\int z^2K(z)dz + \frac{\delta}{2}\left(\mu_1'(r)\mu_1(r) + \frac{1}{2}\sigma_1^2(r)\mu_1''(r)\right) + o(h^2+\delta) \quad \text{and}$$

$$(6)\qquad \mathrm{var}\left[\tilde{\mu_1}(r)\right] \simeq \frac{\sigma_1^2(r)\int K^2(z)dz}{(Th)p_1(r)} + o((Th)^{-1}),$$

where $p_1(r)$ is the marginal density of the short rate.

Equation (5) shows the leading terms of the asymptotic bias of the nonparametric drift function estimator proposed by Stanton (1997). The property of the drift function estimator suggests that even when the discretization interval δ is sufficiently small, bias remains in the nonparametric estimator proposed in Stanton (1997). Thus, it has the potential of generating spurious nonlinearities as illustrated by simulations in Chapman and Pearson (2000). Their simulation is based on the CIR (1985) model, which by specification has a linear drift function. Chapman and Pearson (2000) attribute artificial patterns of nonlinearity near the boundaries of the support to the "truncation" of the observed short rates as a result of a small sample. This is consistent with the bias expression, which shows that the bias is related not only to the slope and/or curvature of the drift function but also to the marginal density function of the short rate. In particular, the term $((\tilde{p_1}(r) - p_1(r))/(\tilde{p_1}(r)))$ omitted from the bias expression explicitly illustrates that severe biases can occur toward boundaries where $p_1(r)$ is poorly estimated.[1]

[1] We thank the referee for pointing this out.

Chapman and Pearson (2000) also account for the "boundary effect" using the jackknife kernel proposed in Rice (1984) and show that it leads to no reduction in the spurious nonlinearity of the estimated drift function toward the boundaries. While certain bandwidth choices may reduce the boundary bias, they do so at the cost of increasing the overall bias, because the boundary bias partially offsets the truncation bias.[2]

The intuition for the difficulty of estimating the drift function relative to the diffusion function is long established in the literature. While the diffusion function estimator in equation (4) requires the sampling interval $\delta \rightarrow 0$ for convergence, the drift function estimator in equation (3) also requires the sampling period $T \rightarrow \infty$ for convergence. This is consistent with the insight of Merton (1980), who points out that when the sampling interval is small, even though the diffusion term can be estimated very precisely, the estimate of the drift coefficient tends to have low precision. Formally, we note that for the diffusion process in equation (1), the drift term is of order dt and the diffusion term is of order $\sqrt{dt}$, as $(dw_t)^2 = dt + O((dt)^2)$ (i.e., the diffusion term has lower order than the drift term for infinitesimal changes in time). Therefore, the local-time dynamics of the sampling path reflect more of the properties of the diffusion term than those of the drift term. Not surprisingly, approximations of the drift function from high frequency data can be very nonrobust, and the estimates can be very sensitive to the sampling path.

In addition, a well-known property of the short rate is its high persistence over time.[3] Consequently, interest rates tend to stay around certain levels for an extended time period. Such a property leads to an undesirable feature that interest rate observations, over even a reasonably long time period, can only offer us restricted or truncated information of the short rate distribution. As a result, interest rate observations over a limited time period appear to be a "truncated" sample, as phrased in Abhyankar and Basu (2001) and Li et al. (2004). Evidence in Pritsker (1998), Chapman and Pearson (2000), and Jones (2003) suggests that as a result of the strong persistence of interest rates, identifying the drift function requires a large number of observations with a given sampling frequency or equivalently a long sample period. The hope thus rests in extending the sample period instead of sampling frequency in order to provide a reliable estimate of the drift function. Unfortunately, the time period of historical observations of interest rate is inevitably limited.

In this paper, we propose a new nonparametric estimator of the short rate diffusion that can greatly reduce the bias of the estimator proposed in Stanton (1997).

[2]Estimators based on higher order approximations are also constructed in Stanton (1997) with a potential of reducing the discretization bias. Fan and Zhang (2003) derive explicit expressions of the asymptotic behavior of both higher order drift and diffusion estimators. They show that while the high order estimators can reduce approximation errors in asymptotic biases, their asymptotic variances escalate nearly exponentially with the order of approximation.

[3]The statistical issue involved in the nonparametric estimation of the drift function from highly persistent data is the optimal choice of bandwidth, which can be substantially different from that under the i.i.d. condition. Simulations in Chapman and Pearson (2000) and the present study both confirm that the optimal choice of bandwidth helps to reduce the spurious bias. However, the improvement is limited.

Specifically, it uses information from a panel of yields with different maturities instead of a single time series of short rate observations. As illustrated in our simulations, the pooled data offer important incremental information for the drift function identification and estimation.[4] The proposed method is similar in spirit to the Hjort and Glad (1995) approach, which uses a parametric pilot estimator and a nonparametric "correction factor" for improved functional estimation. Briefly, suppose one intends to estimate the conditional mean $\mathrm{E}(Y|X=x)=\mu(x)$ where $Y_i{=}\mu(X_i){+}\epsilon_i$, with ϵ_i being the disturbance. Hjort and Glad (1995) propose a semiparametric estimator that combines a parametrically estimated pilot with a nonparametrically estimated correction factor, using the Nadaraya-Watson method. The parametric pilot can be thought of as a prior for the shape of $\mu(x)$, whereas the correction factor adjusts the pilot if it does not satisfactorily capture the shape of $\mu(x)$. Consequently, the estimator behaves like the parametric pilot if the parametric functional form is sufficiently close to the true conditional mean $\mu(x)$, or it resembles the Nadaraya-Watson estimator otherwise. The advantage of the two-step procedure is that when the parametric pilot is a good approximation of the conditional mean, then the "correction factor" will be a nice smooth or even flat function, and thus easier to estimate nonparametrically than the conditional mean $\mu(x)$ itself. Hjort and Glad (1995) show that when the correction factor is less rough than the conditional mean, the two-step procedure results in an improved nonparametric estimator with reduced bias, a finding supported by our simulations results. The bias reduction is mainly due to the fact that the nonparametric estimation in the second step benefits from a robust choice of the bandwidth. In general, when the function to be estimated is highly nonlinear, bandwidth choice poses a serious challenge and is critical for the accuracy of nonparametric estimation. On the other hand, when the function to be estimated is relatively flat or constant, choosing the optimal bandwidth is much easier, leading to both high efficiency and small bias in nonparametric estimation.

While our estimator applies to both the drift and diffusion functions, as illustrated in the empirical application, we focus on the drift function in the theoretical development and the simulations. In the first step, the drift function is estimated using the pooled data from J different interest rate diffusion processes, including the short rate diffusion process. The bias correction is then performed in the second step. Formally, we define $\mu_p(r)=\sum_{j=1}^{J}\mu_j(r)\omega_j(r)$, where $\omega_j(r)=(p_j(r))/(\sum_{j=1}^{J}p_j(r))$ and $p_j(r)$ is the density function of the jth interest rate process, with $j=1,\ldots,J$. That is, $\mu_p(r)$ is the function being estimated by pooling

[4]Various studies in the finance literature have used extraneous data to improve the efficiency of term structure estimation. For example, Abaffy, Bertocchi, Dupacova, Giacometti, Huskova, and Moriggia (2003) evaluate the position of 11 European Union members in the Euro bond market by assuming that their underlying yield curves can be nonparametrically modeled as a sum of an individual factor and a common factor. The common factor captures cross-country similarities, which result from the elimination of exchange rate risk due to the launching of the Euro. In a parametric framework, Ang and Bekaert (2002) and Lee and Li (2005) supplement U.S. short rate observations with interest rate data from other western countries (U.K. and Germany in Ang and Bekaert (2002) and France, U.K., Italy, Germany, and Japan in Lee and Li (2005)) when estimating the U.S. term structure of interest rates.

the data. Note that $\mu_p(r)$ is identical to the short rate drift $\mu_1(r)$ if the individual drift functions are the same.

Our proposed estimator of the drift function is based on the identity

$$\mu_1(r) \quad = \quad \mu_p(r)\frac{\mu_1(r)}{\mu_p(r)} \quad = \quad \mu_p(r)c(r), \tag{7}$$

where $c(r)$ is referred to as the "correction factor" and is defined on a set where $\mu_p(r) \neq 0$ (see Hjort and Glad (1995), Glad (1998)). If $\mu_p(r)$ is known, then the correction factor $c(r)$ can be estimated nonparametrically by $\hat{c}(r)$:

$$\hat{c}(r) \quad = \quad \frac{1}{\delta}\frac{\sum_{t=0}^{n-1}\left(\frac{\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)}{\mu_p\left(r_{t\delta}^{(1)}\right)}\right)K_h\left(r_{t\delta}^{(1)} - r\right)}{\sum_{t=0}^{n-1}K_h\left(r_{t\delta}^{(1)} - r\right)}, \tag{8}$$

since

$$\mathrm{E}\left[\frac{\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)}{\delta\mu_p\left(r_{t\delta}^{(1)}\right)}\middle| r_{t\delta}^{(1)} = r\right] \quad = \quad c(r).$$

However, in empirical applications, the functional form of $\mu_p(r)$ is unknown. Hence a feasible estimator of the drift function is obtained by replacing $\mu_p(r)$ with its consistent estimator. Under the framework described above, $\mu_p(r)$ can be estimated nonparametrically by pooling all the observations of the J bond yields; for example,

$$\hat{\mu_p}(r) \quad = \quad \frac{1}{\delta}\frac{\sum_{j=1}^{J}\sum_{t=0}^{n-1}\left(r_{(t+1)\delta}^{(j)} - r_{t\delta}^{(j)}\right)K_{h_p}\left(r_{t\delta}^{(j)} - r\right)}{\sum_{j=1}^{J}\sum_{t=0}^{n-1}K_{h_p}\left(r_{t\delta}^{(j)} - r\right)}. \tag{9}$$

Consistency of the pooled estimator $\hat{\mu_p}(r)$ is established in Appendix A under regularity conditions. With a consistent estimator of $\mu_p(r)$ in equation (9), our proposed estimator of the drift function of the short rate $\mu_1(r)$ is given by $\hat{\mu_1}(r) = \hat{\mu_p}(r)\hat{c}(r)$ or

$$\hat{\mu_1}(r) \quad = \quad \hat{\mu_p}(r)\left[\frac{1}{\delta}\frac{\sum_{t=0}^{n-1}\left(\frac{\left(r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}\right)}{\hat{\mu_p}\left(r_{t\delta}^{(1)}\right)}\right)K_h\left(r_{t\delta}^{(1)} - r\right)}{\sum_{t=0}^{n-1}K_h\left(r_{t\delta}^{(1)} - r\right)}\right]. \tag{10}$$

The proposed estimator is designed to reduce the bias of the estimator proposed by Stanton (1997). Intuitively, if the drift functions are identical (i.e., $\mu_1(\cdot) = \mu_2(\cdot) = \cdots = \mu_J(\cdot) = \mu_p(\cdot)$), then $\hat{c}(r)$ is an estimate of unity and the proposed estimator is essentially the pooled pilot $\hat{\mu}_p(r)$. In this case, the proposed estimator would be "optimal" and substantially more efficient than the conventional estimators that use a single time series of short rate observations. In general, when

$\mu_p(r)$ is not too distant from $\mu_1(r)$, the correction factor $c(r)$ will be less rough than $\mu_1(r)$, hence much easier to estimate nonparametrically. The formal results for the bias and variance of the proposed drift function estimator are given in the following proposition.

Proposition 2. Suppose that $\delta = \delta(T,n) \to 0$, $h = h(T,n) \to 0$, and $h_p = h_p(T,n) \to 0$ as $T,n \to \infty$; $Th \to \infty$, $Th_p \to \infty$, and the "auxiliary processes" in equation (2) also satisfy the regularity conditions listed in Appendix A, then the basic statistical properties of $\hat{\mu_1}(r)$ are:

$$\text{(11)} \qquad \mathrm{E}\left[\hat{\mu_1}(r) - \mu_1(r)\right] \simeq \frac{h^2}{2}\left(c''(r) + 2c'(r)\frac{p_1'(r)}{p_1(r)}\right)\mu_p(r)\int z^2K(z)dz + \frac{\delta}{2}\left(c'(r)c(r) + \frac{1}{2}\sigma_1^2(r)c''(r)\right)\mu_p(r) + o(h^2+\delta) \quad \text{and}$$

$$\text{(12)} \qquad \mathrm{var}\left[\hat{\mu_1}(r)\right] \simeq \frac{\sigma_1^2(r)\int K^2(z)dz}{(Th)p_1(r)} + O((JTh_p)^{-1} + h^2h_p^2 + h_p^2\delta) + o((Th)^{-1}).$$

Compared to the results in Proposition 1, it is clear that the asymptotic biases of the two estimators can substantially differ. The bias of the proposed estimator depends on the slope and curvature of the correction factor $c(r)$ instead of the slope and curvature of the drift function $\mu_1(r)$. Consequently, when the pooled pilot is identical to the drift of the short rate, that is, when $\mu_p(r) = \mu_1(r)$, the correction factor $c(r) = (\mu_1(r))/(\mu_p(r))$ is a straight line, hence $c'(r) = c''(r) = 0$. This also reduces the effect of the marginal density function on the bias of the estimator. In addition, by pooling the data, our estimator uses a pooled density estimator (in the numerator of the pooled drift function estimator) instead of the conventional Nadaraya-Watson estimator. Similarly, if $\mu_p(r)$ is not too different from $\mu_1(r)$ in functional shape, the correction factor will oscillate around a constant and, as a result, the proposed estimator will be less biased. In comparison to Proposition 1, the variance of the proposed estimator tends to be higher because of the additional estimation step. Note that the added terms in the variance expression are of the order $O((JTh_p)^{-1})$ or lower. Thus the variance increase relative to the Stanton (1997) estimator is essentially negligible with a reasonably large J or a slowly converging smoothing parameter h_p.

Proposition 3. In addition to the assumptions of Proposition 2, further suppose that $nh^5 \to 0$, $\sqrt{nh}\delta \to 0$, $Jnh_p^5 \to 0$, and $\sqrt{Jnh_p}\delta \to 0$; then the proposed drift function estimator is asymptotically normally distributed:

$$\text{(13)} \qquad \sqrt{Th}\left(\hat{\mu_1}(r) - \mu_1(r)\right) \to N\left(0, \frac{\sigma_1^2(r)\int K^2(z)dz}{p_1(r)}\right).$$

Proposition 3 shows that the estimators $\hat{\mu}_1(r)$ and $\tilde{\mu}_1(r)$ have the same asymptotic distribution. Intuitively, there exists a long enough sampling period T beyond which any marginal information is inconsequential in further reducing the bias of the estimator proposed by Stanton (1997). Thus any bias correction by the proposed estimator must be in finite samples, as discussed above.

The proposed approach is particularly relevant for the estimation of the short rate diffusion, as yields with different maturities are available along the yield curve. One key question is whether the model specifications in equations (1) and (2) are consistent under the term structure framework. In Appendix B, we show that in the presence of measurement error or idiosyncratic shocks to bond prices of a specific maturity, yields of different maturities can follow diffusions with imperfectly correlated Brownian motions. In addition, we use affine and nonlinear models to illustrate that drift functions of yields of different maturities tend to be similar in functional shape. This is consistent with the empirical literature that documents evidence of systematic comovements among yields with different maturities. For example, the principal component analysis in Litterman and Scheinkman (1991) shows that the short rate is a major factor driving the dynamics of yield curve. Moreover, since the bias correction of the proposed estimator is most effective in the area where the drift function is itself highly nonlinear, it can improve the drift function estimation toward the boundaries where the spurious biases are most pronounced.

III. Simulations

In this section, we perform Monte Carlo simulations to evaluate the finite sample properties of the proposed estimator using the Vasicek (1977) and CIR (1985) models. The Vasicek model for the short rate and the auxiliary model are specified as follows:

$$\begin{aligned} (14) \qquad dr_t^{(1)} &= 0.261(0.0717 - r_t^{(1)})dt + 0.02237 dw_t^{(1)}, \\ dr_t^{(2)} &= \kappa * 0.261(0.0717 - r_t^{(2)})dt + \sqrt{\gamma} * 0.02237 dw_t^{(2)}, \quad \text{and} \\ dw_t^{(1)} dw_t^{(2)} &= \rho dt, \end{aligned}$$

where the parameter values are set equal to those in Aït-Sahalia (1999), which are estimated using monthly observations of the Federal funds rate from January 1963 through December 1998. The CIR (1985) model for the short rate and the auxiliary model are specified as follows:

$$\begin{aligned} (15) \qquad dr_t^{(1)} &= 0.8537(0.08571 - r_t^{(1)})dt + 0.1566\sqrt{r^{(1)}} dw_t^{(1)}, \\ dr_t^{(2)} &= \kappa * 0.8537(0.08571 - r_t^{(2)})dt + \sqrt{\gamma} * 0.1566\sqrt{r^{(2)}} dw_t^{(2)}, \quad \text{and} \\ dw_t^{(1)} dw_t^{(2)} &= \rho dt, \end{aligned}$$

where the parameter values of the short rate model are set equal to those in Chapman and Pearson (2000). Without loss of generality, we only consider the case of $J = 2$ (i.e., there is only one additional auxiliary diffusion process). The parameters κ, γ, and ρ in the auxiliary model are set to different values in our simulations in order to investigate the effects of different factors. In total, the following five cases are considered in our simulations:

Case I (benchmark). $\kappa = \gamma = 1$ and $\rho = 0$. This is the ideal situation for the use of the proposed method, since both the drift and diffusion functions of the auxiliary model are identical to those of the short rate model. The additional data from the auxiliary model can be simply viewed as observations over an extended time period.

Case II (mean reversion effect). $\gamma=1$, $\rho=0$, $\kappa=0.75$, 1.25, and 1.5. In this case, the short rate model and the auxiliary model have the same long-run mean, but different speeds of mean reversions. This case studies the effect of the mean-reversion level of the auxiliary process, or the dissimilarity of the functional shapes of $\mu_1(r)$ and $\mu_2(r)$, on the performance of the proposed estimator.

Case III (diffusion effect). $\kappa = 1$, $\rho = 0$, $\gamma = 0.75$, 1.25, and 1.5. In this case, the short rate process and the auxiliary model have identical drift functions, but different diffusion functions. The auxiliary process is more (less) volatile when $\gamma > (<)1$.

Case IV (correlation effect). $\kappa=\gamma=1$ and $\rho=0.2$, 0.4, and 0.6. This case allows us to examine the impact of the level of correlation between the short rate model and the auxiliary model on the performance of the proposed drift function estimator.

Case V (mixed models). Instead of restricting both the short rate model and the auxiliary model to have the same specification as in Cases I through IV, in this simulation we assume that the short rate $r_t^{(1)}$ follows a Vasicek process, while the auxiliary interest rate $r_t^{(2)}$ follows a CIR (1985) process. This case is particularly interesting because in empirical applications the short rate model and the auxiliary model may have different drift and diffusion functions.

In each case, we first simulate 31 years of daily data for both the short rate model and the auxiliary model using the Milshtein discretization scheme. Observations in the first year are discarded to eliminate the start-up effects, which results in a total of 7,500 daily observations over 30 years. With the simulated sampling paths for the short rate model and the auxiliary model, nonparametric drift function estimators are then implemented over the support of $r = [0, 20\%]$. Note that the estimator in Stanton (1997) only uses observations of the short rate diffusion process, while the proposed estimator uses observations from both the short rate and the auxiliary diffusion processes. Throughout the simulations, a Gaussian kernel is used with the bandwidth that minimizes the integrated squared error (ISE).[5]

[5]Note that implementation of the proposed estimator requires the selection of two smoothing parameters, namely, h_p in the first step and h in the second step. The smoothing parameters can be selected using the blockwise cross-validation method, either simultaneously or sequentially. The simultaneous procedure involves choosing h_p and h to minimize the usual "leave block-b out" cross-validation function:

$$\sum_{t=0}^{n-1}\left(\hat{\mu_1}^{(b)}(r_{t\delta}^{(1)}) - \left[\frac{r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}}{\delta}\right]\right)^2,$$

where $\hat{\mu_1}^{(b)}(r_{t\delta}^{(1)})$ is the proposed estimator of the drift function leaving out a block of b observations. Note that this block should also be left out of the cross-validation function in the estimation of the pilot function. On the other hand, the sequential procedure involves first choosing h_p to minimize the cross-validation function of $\hat{\mu_p}(r)$ and then choosing h to minimize the cross-validation function of $\hat{\mu_1}(r)$.

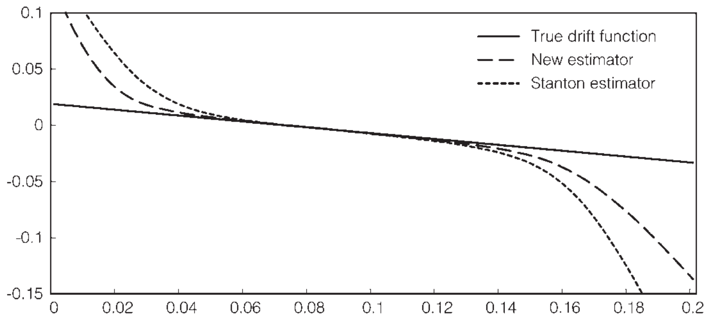
The drift function estimates of the Vasicek (1977) and CIR (1985) short rate processes for the benchmark case, averaged across 5,000 replications, are plotted in Figures 1 and 2, together with the true drift function. For the purpose of comparing our results with those in Chapman and Pearson (2000) and illustrating the effect of the bandwidth choice, the naive bandwidth is used for both the Stanton (1997) estimator and the proposed estimator in Figure 1. The results for the Stanton (1997) estimator are consistent with the findings in Chapman and Pearson (2000). Namely, the drift displays significant nonlinearity toward the boundaries of the support. Chapman and Pearson (2000) argue that such nonlinearity in the estimated drift function results mainly from two sources: the truncation of the observed interest rates within the interval $[r_{\text{min}}^{(1)}, r_{\text{max}}^{(1)}]$ and the limited sampling period. While the proposed estimator has less nonlinearity toward the boundaries, it remains a poor estimate of the linear drift function with the use of the naive bandwidth.

FIGURE 1

Drift Function Estimates of the Vasicek and CIR Models
(with naive bandwidth h_{iid})

Figure 1 plots the short rate drift function estimates based on the Stanton (1997) estimator and the proposed estimator for the Vasicek (1977) and CIR (1985) models. The estimates are based on 7,500 daily observations with 5,000 replications. The naive bandwidth h_{iid} is used for both estimators.

Graph A. Drift Function Estimates of the Vasicek Model



Graph B. Drift Function Estimates of the CIR Model

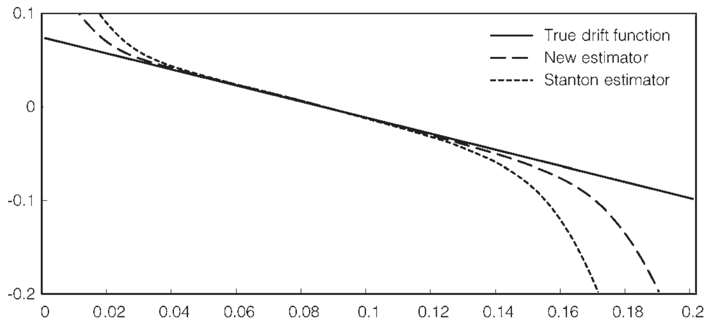


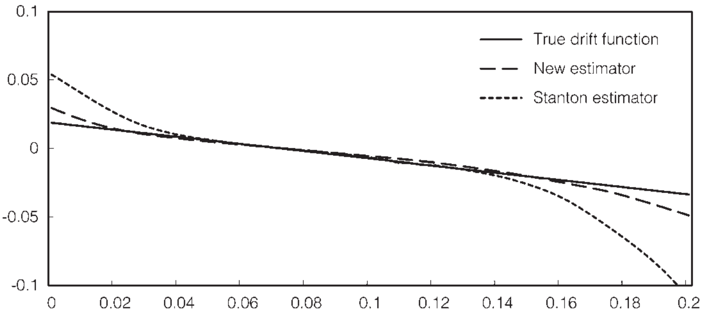
Figure 2 plots the average drift function estimates of the Stanton (1997) estimator and the proposed estimator for the Vasicek (1977) and CIR (1985) models

FIGURE 2

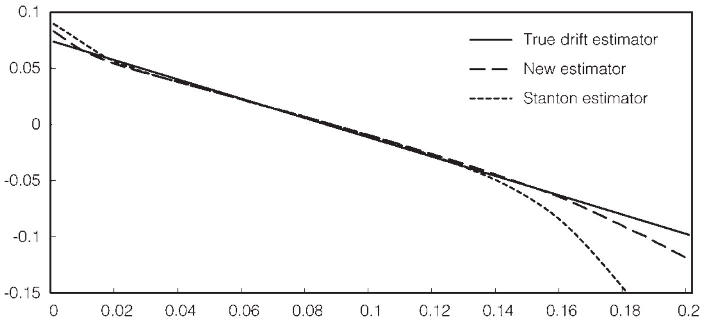
Drift Function Estimates of the Vasicek and CIR Models
(optimal bandwidth h_{opt})

Figure 2 plots the short rate drift function estimates based on the Stanton (1997) estimator and the proposed estimator for the Vasicek (1977) and CIR (1985) models. The estimates are based on 7,500 daily observations with 5,000 replications. The optimal bandwidth h_{opt} that minimizes the RMISE is used for both estimators.

Graph A. Drift Function Estimates of the Vasicek Model



Graph B. Drift Function Estimates of the CIR Model



using the ISE-minimizing bandwidth.[6] Two observations are noted. First, there is a substantial improvement for both the Stanton (1997) estimator and the proposed estimator. The improvement highlights the importance of the choice of the smoothing parameter. While the naive bandwidth used in Figure 1 assumes the sample data are i.i.d., the optimal bandwidth takes into account the serial dependence of the data. Second, compared to the estimator proposed by Stanton (1997), the proposed drift function estimator exhibits a much less "spurious" nonlinearity with smaller biases toward the boundaries of the support. That is, in addition to the improvement due to the use of optimal bandwidth, the proposed estimator can further reduce the bias due to the use of extraneous data. Since in this simulation the auxiliary model is identical to the short rate model, the effect of using the auxiliary model is equivalent to doubling the sample size or the sampling period. The performance of the proposed estimator is similar to that of the Stanton (1997)

[6]To limit computational burden, the smoothing parameters for the pooled and the proposed estimator (h_p and h) are chosen by sequentially minimizing the ISEs of $\hat{\mu_p}(r)$ and $\hat{\mu_1}(r)$. Note that an iterated procedure to simultaneously determine the optimal choice of (h_p and h) would perform at least as well as the suboptimal sequential procedure.

estimator in Chapman and Pearson (2000), where 15,000 or 60 years of daily observations are used. Chapman and Pearson (2000) document that the bias of the nonparametric drift function estimate reduces as the sampling period extends.

Table 1 reports the mean integrated absolute bias (MIAB) and the root mean integrated squared error (RMISE) of alternative drift function estimates for the Vasicek (1977) and CIR (1985) processes. Panel A reports the MIAB and RMISE of the Stanton (1997) estimator, and Panel B reports those of the proposed estimator for the benchmark case (Case I). Again, for both estimators, it is clear that the use of optimal bandwidth is important. When the optimal bandwidth is used, the bias of the proposed estimator is only slightly over 20% of the bias of the Stanton estimator for both the CIR (1985) and Vasicek (1977) models, according to the MIAB. The RMISE, as a measure of overall efficiency, is also substantially reduced as a result of bias reduction.

TABLE 1

The Simulation Results of Vasicek and CIR Models

Table 1 reports the simulation results of the proposed estimator, in comparison with the Stanton method, for the drift function of the Vasicek (1977) and CIR (1985) models. MIAB and RMISE denote mean integrated absolute bias and root mean integrated squared error, respectively.

Model	MIAB	RMISE	MIAB	RMISE	MIAB	RMISE
Panel A. Benchmark Stanton (1997) Estimator						
	h_{iid}		h_{opt}			
Vasicek	1.7877	7.8529	0.2436	2.3305		
CIR	1.1571	9.7516	0.3575	3.8268		
Panel B. Benchmark New Pooled Estimator (Case I: $\kappa = \gamma = 1$, $\rho = 0$)						
	h_{iid}		h_{opt}			
Vasicek	0.6695	5.1900	0.0584	0.8485		
CIR	0.4336	4.3481	0.0848	1.7137		
Panel C. Effect of Mean Reversion (Case II: $\gamma = 1$, $\rho = 0$)						
	$\kappa = 0.75$		$\kappa = 1.25$		$\kappa = 1.5$	
Vasicek	0.0559	0.8407	0.0562	1.0313	0.0630	1.0604
CIR	0.1047	1.7176	0.1179	1.9078	0.1625	2.1532
Panel D. Effect of Diffusion (Case III: $\kappa = 1$, $\rho = 0$)						
	$\gamma = 0.75$		$\gamma = 1.25$		$\gamma = 1.5$	
Vasicek	0.0892	1.2927	0.0501	0.8989	0.0473	0.8014
CIR	0.1468	2.1645	0.0516	1.6330	0.0396	1.5379
Panel E. Effect of Correlation (Case IV: $\kappa = \gamma = 1$)						
	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$	
Vasicek	0.0605	1.1160	0.0626	1.3305	0.0738	1.5040
CIR	0.0891	1.9327	0.0946	2.1656	0.1204	2.4133
Panel F. Effect of Mixed Models (Case V: $\kappa = \gamma = 1$, $\rho = 0$)						
	Auxiliary Model: CIR					
Vasicek	0.1079	1.1454				

The MIAB and RMISE of the proposed estimator for the remaining cases are also reported in Table 1. As in the benchmark case, each simulation involves 5,000 replications. Panel C of Table 1 reports the performance of the proposed estimator when the short rate and auxiliary models have the same type of process but a different level of mean reversion (Case II). The results suggest that the proposed estimator continues to perform well even when the shape of the auxiliary drift $\mu_2(r)$ is different from that of the short rate drift $\mu_1(r)$. Panel D of Table 1 reports the performance of the proposed estimator when the diffusion coefficients $\sigma_1^2(r)$ and $\sigma_2^2(r)$ are dissimilar (Case III). The results suggest that in general the performance of the proposed estimator improves as the volatility of the auxiliary model increases. This result is expected because higher volatility implies less persistence of the auxiliary process, thus observations from the auxiliary model offer more incremental information for the estimation of the pooled drift function. Panel E of Table 1 reports the performance of the proposed estimator for different levels of correlation between the Brownian motions $\omega_t^{(1)}$ and $\omega_t^{(2)}$ (Case IV). Clearly, as ρ increases there is a deterioration in the performance of the proposed estimator. Since the two processes have identical drift and diffusion functions, the two interest rate paths move more closely with each other as ρ increases. We also set $\rho = 0.90, 0.95$, and 0.99 in our simulations; the results confirm that the gain of using extraneous information diminishes as ρ approaches 1.

In all the above simulations, we have restricted the short rate model and the auxiliary model to have the same functional form for both the drift and diffusion functions. To examine how the proposed estimator performs when the short rate model and auxiliary model have different specifications, we consider the case where the auxiliary model follows the CIR (1985) process while the model to be estimated follows the Vasicek (1977) process. The Vasicek model is simulated using the parameter values in equation (14), and the auxiliary CIR process is simulated using the parameters in equation (15) with $\kappa = \gamma = 1$ and $\rho = 0$. In this case, the short rate model and the auxiliary model not only have different functional forms for the diffusion, but also different parameter values for the drift. The MIAB and RMISE of the proposed estimator are reported in Panel F of Table 1. Both MIAB and RMISE are higher than the benchmark case. Still, although the MIAB is about twice that of the benchmark case, it is less than half of the bias of the estimator proposed in Stanton (1997). The RMISE of the proposed estimator is also considerably lower than that of the Stanton (1997) estimator.

The performance of our proposed estimator in these finite sample experiments is quite satisfactory. The incorporation of additional observations from auxiliary interest rate processes has the effect of expanding the local information at the boundaries and thus leads to bias reduction. The simulation results also show that the proposed estimator performs well relative to the Stanton (1997) estimator even when the drift and diffusion functions of the short rate and auxiliary models are different. This is important because in empirical applications, it is likely that the short rate and the auxiliary models have different drift and/or diffusion functions.

IV. Empirical Application

A. Data

The data in our empirical analysis consist of 12,704 daily observations of the U.S. 3-month T-bill rates from January 1954 to November 2004. This data set, to our knowledge, has the longest sampling period compared to those in existing studies. The 3-month T-bill yields are used as a proxy of the short rate. In addition, we use yields of the 6-month T-bill, the 1-year T-bill, and the 3-, 5-, and 10-year T-notes. There are 10,676 daily observations for each of the additional series from February 1962 to November 2004. In other words, we have 12,704 observations on the short rate process or "the model" we intend to estimate, and 10,676 observations on each of the five additional diffusion processes or "the auxiliary models" that we use to improve the estimation of the short rate process.

Descriptive statistics of the data are reported in Table 2. The average yields of different maturities suggest that the yield curve is overall upward sloping, and the standard deviations of the daily yield changes suggest that the yield curve is more volatile over the short end than the long end. The minimum and maximum observations reflect the wide range of interest rates over our sampling period. For instance, the 3-month T-bill yields have a minimum value of 0.55% and a maximum value of 17.14%. Yields with maturities longer than 3 months tend to have higher minimum values than the yields of 3-month T-bills, while the maximum values are comparable across maturities. As in Aït-Sahalia (1996a), we report the autocorrelations of the monthly interest rate series in Table 2. Although the autocorrelation coefficients of the interest rate are very high, they are all significantly different from 1. Moreover, those of the day-to-day changes are generally small and not consistently positive or negative. To test whether the short rate follows a unit root process, we perform the augmented Dickey-Fuller nonstationarity test. The test statistic for the 3-month T-bill rates has a value of −2.86, which is lower than the 10% critical value of −2.57. That is, the null hypothesis of nonstationarity is rejected at the 10% significance level. Similar results are obtained for the time series of yields with longer maturities.

TABLE 2

Summary Statistics of Interest Rates

Panel A of Table 2 reports the summary statistics of the daily interest rates with maturities of 3 months, 6 months, 1 year, 3 years, 5 years, and 10 years. Panel B reports the summary statistics of the daily interest rate changes. The summary statistics of the 3-month T-bill yields are based on the daily observations from January 1954 to November 2004, while those of other interest rates are based on the daily observations from February 1962 to November 2004. The mean for the daily change of interest rate has a magnitude of 10^{-4}. The autocorrelations are calculated from the monthly interest rate data.

							Autocorrelations of Monthly Series				
τ	Mean	StdDev	Skew	Kurt	Min	Max	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
Panel A. Summary Statistics of Daily Interest Rates											
r3M	5.226	2.848	1.060	1.684	0.55	17.14	0.967	0.914	0.869	0.832	0.799
r6M	5.910	2.730	0.927	1.415	0.80	15.93	0.969	0.918	0.874	0.840	0.809
r1Y	6.365	2.932	0.945	1.301	0.88	17.31	0.970	0.921	0.879	0.846	0.819
r3Y	6.795	2.741	0.907	0.989	1.34	16.59	0.976	0.938	0.906	0.881	0.857
r5Y	7.009	2.630	0.959	0.878	2.08	16.27	0.980	0.947	0.920	0.897	0.875
r10Y	7.229	2.515	0.969	0.656	3.13	15.84	0.983	0.958	0.937	0.917	0.897
Panel B. Summary Statistics of Daily Interest Rate Changes											
Δr3M	–0.609	0.102	0.202	26.11	–1.27	1.34	0.440	0.149	0.113	0.101	0.131
Δr6M	–0.618	0.090	0.296	23.74	–1.10	1.17	0.424	0.133	0.103	0.081	0.098
Δr1Y	–0.758	0.092	–0.194	21.39	–1.08	1.10	0.434	0.086	0.082	0.042	0.037
Δr3Y	–0.665	0.081	–0.180	13.26	–0.79	0.92	0.400	0.002	0.015	0.001	–0.059
Δr5Y	–0.440	0.076	–0.305	11.42	–0.77	0.72	0.392	–0.013	0.009	–0.002	–0.073
Δr10Y	0.103	0.068	–0.273	10.08	–0.75	0.65	0.349	–0.034	0.020	0.053	–0.055

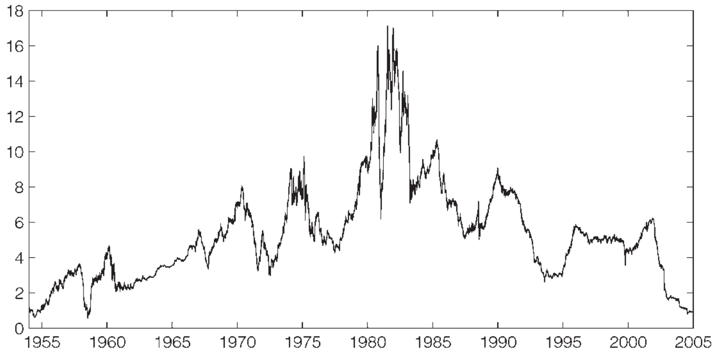
Figure 3 plots the time series of the daily 3-month Treasury yields and the day-to-day changes in Graphs A and B, respectively. The time-series plot confirms the wide range of the 3-month T-bill yields over the sampling period, and the first difference reflects some large changes of the 3-month T-bill yields from day to day. The visibly large daily changes of 3-month T-bill yields are associated with the high levels of yields during the late 1970s and early 1980s, indicating different volatility at different interest rate levels.

FIGURE 3

Daily Yields and Daily Yield Changes of U.S. Three-Month T-Bills

Graphs A and B of Figure 3 plot the time series of the daily 3-month Treasury bill yields and their daily changes, respectively. The sample period is from January 4, 1954 to November 12, 2004.

Graph A. Daily U.S. 3-Month T-Bill Rates (Jan. 1954–Nov. 2004)



Graph B. Daily Change of U.S. 3-Month T-Bill Rates (Jan. 1954–Nov. 2004)

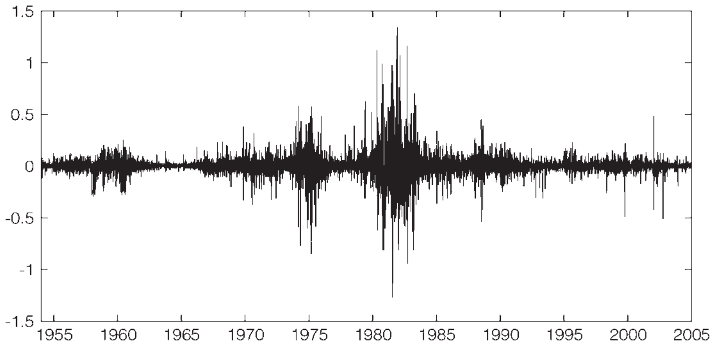


Table 3 also reports the correlations and the principal components of daily interest rate changes. The correlation matrix suggests that the changes of interest rates along the yield curve are highly correlated. For instance, the daily changes of 3-month and 6-month T-bill yields have a correlation of nearly 85%. The daily changes of 3-month T-bill and 10-year T-note yields, however, have a correlation of about 50%. The principal component analysis reported in Table 3 is similar to those in existing studies. Namely, three factors mainly drive the dynamics of yield curve: the level, the slope, and the curvature. In particular, the short rate explains more than 80% of the total variation of the yield curve dynamics, suggesting the importance of modeling the short rate process.

TABLE 3

Correlations and Principal Components of Daily Interest Rate Changes

Panel A of Table 3 reports the correlation matrix of the daily changes of interest rates with different maturities. Panel B reports the principal components of the daily changes of interest rates, as well as the loadings of principal components on the changes in yields. The last column represents the percentage of total variation of the yield curve explained by each of the individual principal components.

Panel A. Correlations of Daily Interest Rate Changes

	Δr3M	Δr6M	Δr1Y	Δr3Y	Δr5Y	Δr10Y
Δr3M	1.0000					
Δr6M	0.8564	1.0000				
Δr1Y	0.7462	0.8720	1.0000			
Δr3Y	0.6076	0.7451	0.8546	1.0000		
Δr5Y	0.5684	0.7063	0.8121	0.9401	1.0000	
Δr10Y	0.5039	0.6343	0.7434	0.8754	0.9251	1.0000

Panel B. Principal Components of Daily Interest Rate Changes

Factor	Factor Loadings						Variation (%)
1 (level)	8.8759	8.6152	9.0454	7.6667	7.0018	5.8645	0.8085
2 (slope)	–5.1705	–2.3409	0.2146	2.8958	3.2404	3.2790	0.1211
3 (curvature)	2.1203	–1.4016	–2.5534	0.2010	0.9380	1.4054	0.0336
4	0.7978	–2.2113	1.4981	0.4569	–0.2175	–0.6072	0.0178
5	–0.0270	0.2172	–0.7848	1.7024	0.2252	–1.5623	0.0128
6	–0.0048	0.0429	–0.0851	0.7668	–1.3537	0.6892	0.0062

B. Nonparametric Estimation Results of the Short Rate Process

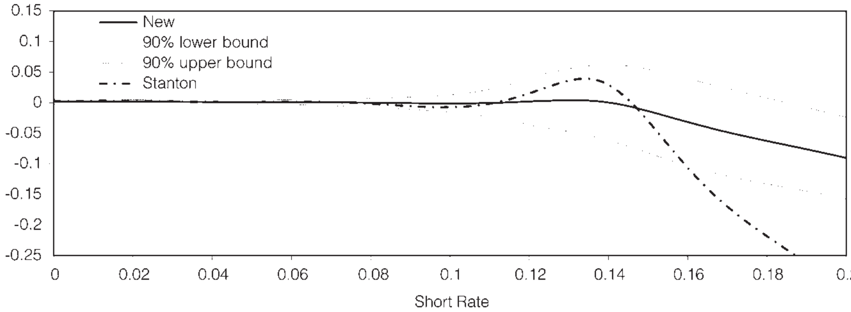
With the panel of yields, we implement the proposed estimator of the short rate drift function. Graph A of Figure 4 plots the nonparametric drift function estimate based on the proposed estimator using the yields of all maturities, together with the Stanton (1997) nonparametric estimate, which uses only the 3-month T-bill yields. The 90% point-wise confidence bands of the proposed estimator are also plotted. The plots show that the two drift function estimates are visibly different for interest rates beyond 12%. Noticeably, the Stanton (1997) drift function estimate is highly nonlinear, with strong mean reversion at high levels of interest rate, consistent with empirical results in the existing literature. In comparison, the new drift function estimate is more flat and has a substantially lower mean reversion at high levels of interest rate. The difference is statistically significant, as the Stanton (1997) estimate lies outside the 90% confidence band of the new estimate for interest rates above 15.75%. Considering the fact that the short rate proxy in our sample ranges from 0.55% to 17.14%, the statistical difference at high levels of the short rate is of practical relevance. Moreover, the results indicate that information from the yields of additional maturities has a significant effect on the estimation of the drift function, especially at the boundaries of the support. Note that as seen from the summary statistics in Table 2, the additional yields with maturities longer than 3 months tend to have higher minimum values than the 3-month T-bill yields. These yields thus do not provide much additional information for the drift function estimation at the lower boundary of the support. On the other hand, the additional yields have maximum values similar to the maximum value of the 3-month T-bill yields, thus providing incremental information for the drift function estimation at the upper boundary of the support.

FIGURE 4

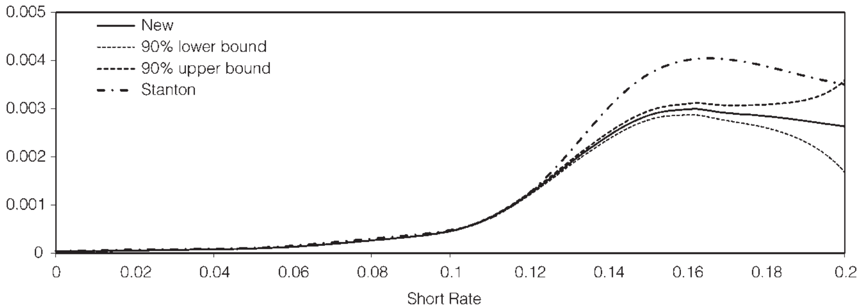
Nonparametric Drift and Diffusion Function Estimates

Graphs A and B of Figure 4 plot the short rate drift and diffusion function estimates based on the proposed estimator using the yields of all maturities, together with the Stanton (1997) nonparametric estimate of the drift function using only the 3-month T-bill yields. The 90% point-wise confidence bands of the new estimator are also plotted.

Graph A. Drift Function Estimates



Graph B. Diffusion Function Estimates



Since the proposed estimator uses additional yields, an interesting question is how much information is actually extracted from these yields to improve the estimation of the short rate drift function. In other words, how many years of short rate observations would be needed for the Stanton (1997) estimator to have the same efficiency as the proposed estimator? To answer this question, we use the MISE as the yardstick of estimation efficiency. First, we estimate the nonparametric diffusion process for each of the six time series of yields using the Stanton (1997) estimator. Second, with the estimated diffusion processes and the correlation matrix of the yields as reported in Table 3, we simulate 50 years of 3-month T-bill yields and 42 years of yields for the additional five maturities. With the simulated panel of yields, we implement the proposed estimator for the short rate drift function. The MISE is calculated based on 5,000 replications. Finally, the Stanton (1997) drift function estimator is implemented using the simulated 3-month T-bill yields over an extended sampling period longer than 50 years. The MISE for the Stanton (1997) drift function estimator, decreasing steadily as the sampling period extends, is also computed based on 5,000 replications. The simulations indicate that with 145 years of short rate observations, the Stanton (1997) drift function estimator matches approximately, in terms of MISE, the efficiency of the proposed estimator. In other words, the proposed estimator of the drift function achieves the same efficiency as the Stanton (1997) estimator implemented with approximately 145 years of daily short rate observations.

We also implement the proposed estimator for the diffusion function using the yields from all maturities. The new diffusion function estimate is plotted in Graph B of Figure 4, together with the Stanton (1997) estimate using only the 3-month T-bill yields. Note that the diffusion function estimates have a much narrower 90% confidence band relative to the drift function estimates, confirming higher efficiency of the diffusion function estimation. Both the Stanton (1997) and the new diffusion function estimates are highly nonlinear, increasing sharply first as the short rate increases and then dropping off slightly toward the upper boundary of the support. Compared to the Stanton (1997) estimate, however, the new diffusion function estimate exhibits a less dramatic increase as the short rate increases. The main difference is that at the high level of the short rate, the Stanton (1997) estimate is clearly outside the 90% confidence band of the new diffusion function estimate. When the short rate is above 12%, the new diffusion function estimate is well below the Stanton (1997) diffusion function estimate. This is consistent with the weaker mean reversion of the new drift function estimate (see Graph A of Figure 4).

C. Impact of the Short Rate Process on the Valuation of Derivatives

In this section, we examine the economic implications of the short rate process estimated using different approaches. We focus on the valuation of both zero-coupon bonds and interest rate derivatives. The risk-neutral process corresponding to the short rate diffusion defined in equation (1) is given by

$$(16) \qquad dr_t^{(1)} = \left(\mu\left(r_t^{(1)}\right) - \lambda\left(r_t^{(1)}\right)\right)dt + \sigma\left(r_t^{(1)}\right)d\tilde{w}_t^{(1)},$$

where $\lambda(r_t^{(1)})=\lambda_0(r_t^{(1)})\sigma(r_t^{(1)})$ is the market price of interest rate risk and $\tilde{w}_t^{(1)}$ is a standard Brownian motion under the equivalent martingale measure Q. It is clear from equation (16) that the drift function of the interest rate enters the risk-neutral process directly. We estimate the nonparametric market price of interest rate risk following the procedure in Jiang (1998). Since the market price of interest rate risk is fully determined by the short rate, it can be estimated from any two non-dividend-paying assets. Let $y(r_t^{(1)}, \tau_i)$ $(i = 1, 2)$ denote the yields of zero-coupon bonds at t with maturity $\tau_i = T_i - t$, and

$$(17) \qquad dy\left(r_t^{(1)}, \tau_i\right) = \alpha(r_t^{(1)}, \tau_i)dt + \kappa(r_t^{(1)}, \tau_i)d\tilde{w}_t^{(1)}.$$

An estimator of $\lambda_0(r_t^{(1)})$ is derived in Jiang (1998) as

$$(18) \quad \hat{\lambda}_0\left(r_t^{(1)}\right) = \left[y_d\left(r_t^{(1)}, \tau_1, \tau_2\right) + \frac{1}{2}\left(\tau_1^2\kappa^2\left(r_t^{(1)}, \tau_1\right) - \tau_2^2\kappa^2\left(r_t^{(1)}, \tau_2\right)\right) + \tau_2\alpha\left(r_t^{(1)}, \tau_2\right) - \tau_1\alpha\left(r_t^{(1)}, \tau_1\right)\right] \Big/ \left[\tau_2\kappa\left(r_t^{(1)}, \tau_2\right) - \tau_1\kappa\left(r_t^{(1)}, \tau_1\right)\right],$$

where $\tau_i = T_i - t$ $(i = 1, 2)$, $y_d(r_t^{(1)}, \tau_1, \tau_2) = y(r_t^{(1)}, \tau_1) - y(r_t^{(1)}, \tau_2)$ is the yield spread between maturities τ_1 and τ_2. In our analysis, we use the 3-month and 10-year yields to estimate the market price of interest rate risk. For the purpose of comparison, we need the estimate of $\lambda_0(r_t^{(1)})$ for both the Stanton (1997) estimator and the proposed new estimator. Consistent with the Stanton (1997) method, the 3-month yield process is estimated using only 3-month T-bill yields, and the 10-year yield process is estimated using only 10-year T-bond yields. Also consistent with our proposed method, yields of all maturities are used for the estimation of both 3-month and 10-year yield processes. Given the estimated 3-month and 10-year yield processes, the estimate of $\lambda_0(r_t^{(1)})$ is obtained from equation (18).

The prices of interest rate derivative securities can be computed based on the simulation of the risk-neutral process in equation (16). Again, simulations of the sample path are based on the Milshtein scheme. In financial applications of the Monte Carlo simulation methods, a number of variance reduction methods have been proposed (see discussions in, e.g., Boyle, Broadie, and Glasserman (1997)). In our simulation, we employ the antithetic variate method to reduce the sample variance. We focus on the impact of the short rate process on the prices of bonds and interest rate caps. Since the functional forms of the drift are mainly debated at the relatively high levels of short rate, the boundary behavior has potentially the largest impact on the valuation of caps.

1. Valuation of Bond Prices

To investigate the impact of the short rate process on bond prices, we simulate the zero-coupon bond prices using the Stanton (1997) estimate and the new estimate of the drift and diffusion functions. The price of a zero-coupon bond with face value \$1 is given by

$$P\left(r_t^{(1)}, t, T\right) = \mathrm{E}_t^Q\left[e^{\left\{-\int_t^T r_u^{(1)} du\right\}}\right],$$

where the interest rate paths are simulated based on the risk-neutral process, with 5,000 replications. Converting the prices to yields, the average yield curves with different values of spot rate are plotted in Figure 5. The 90% confidence bands of the yield curves based on the proposed estimator of the short rate process are also calculated and plotted. With the spot rate at the 10% level, there is no clear difference between the yield curve based on the Stanton (1997) estimator and that based on the new estimator. With the spot rate at either very high or very low levels (15% and 5%), however, the yield curves based on the Stanton (1997) estimator are outside the 90% confidence bands. As expected, due to the stronger mean reversion of the Stanton (1997) drift function estimate, the yield curves based on the Stanton (1997) estimator are significantly lower than those based on the new estimator when the spot rate is high at 15%, but higher than those based on the new estimator when the spot rate is low at 5%. The differences in yield curves are of economic significance, since they also translate into significant differences in bond prices.

2. Valuation of Interest Rate Caps

The value of an interest rate cap is determined by its cash flow over its contract period. The cash flows of an interest rate cap with a notional principal equal to \$100 at time t are

$$100 \times \max[(y(t - \Delta t\,; t) - y_S)\Delta t\,; 0], \tag{19}$$

where $t = t_1, t_2, \ldots, t_n$ are the payment dates (reset of Δt occurs in advance before the payment date), t_n is the last payment date and is often referred to as the cap tenor, n is the number of payments, $y(t - \Delta t\,;\, t)$ is the annualized cap interest rate over the period $(t - \Delta t\,; t]$, and y_S is the annualized cap strike rate. Similarly, the cap prices are calculated from the simulated risk-neutral process with 5,000 replications.

The cap prices are reported in Table 4 for different strike prices (in basis points (bps)), tenors (in years), and annualized spot rates. As shown in Table 4, the cap prices based on the proposed estimator are in general lower than those based on the Stanton (1997) estimator, except when the interest rate is very high ($r_0 = 15\%$). As expected, both the drift and diffusion functions affect the valuation of interest rate caps. While a higher volatility tends to increase the value of caps, a stronger mean reversion has a negative effect on the value when the spot rate is high but a positive effect when the spot rate is low. The differences in cap prices can be explained by the fact that the proposed estimator has an overall lower estimate of the diffusion function compared to the Stanton (1997) estimator. In the meantime, at high interest rate levels, the proposed estimator of the drift function exhibits a much weaker mean reversion than the Stanton (1997) estimator of the drift function. It is noted that for the out-of-the money contracts with long maturity (e.g., the 5-year contract with 50 bps strike price) with spot rate of 5%, the differences between the prices based on the proposed estimates and those based on the Stanton (1997) estimates are more than two standard deviations. While the differences for other contracts are in general within two standard deviations, they are high in numerical values. For instance, for the 5-year at-the-money contracts, the difference is in the range of \$0.38 to \$1.04. For the 5-year contracts with strike price 25 bps below the current spot rate (the in-the-money contracts), the difference is in the range of \$0.37 to \$0.99.

V. Conclusion

In this paper, we propose a new nonparametric estimation method for the short rate diffusion process using information from a panel of yields with different maturities. Our proposed estimator can reduce the bias of the nonparametric

FIGURE 5

Simulated Yield Curves under Alternative Short Rate Processes

Figure 5 plots the simulated yield curves, with different starting interest rate levels, based on the short rate process estimated by the proposed method and the Stanton (1997) method. The 90% point-wise confidence bands of the yield curves based on the new estimator are also plotted.

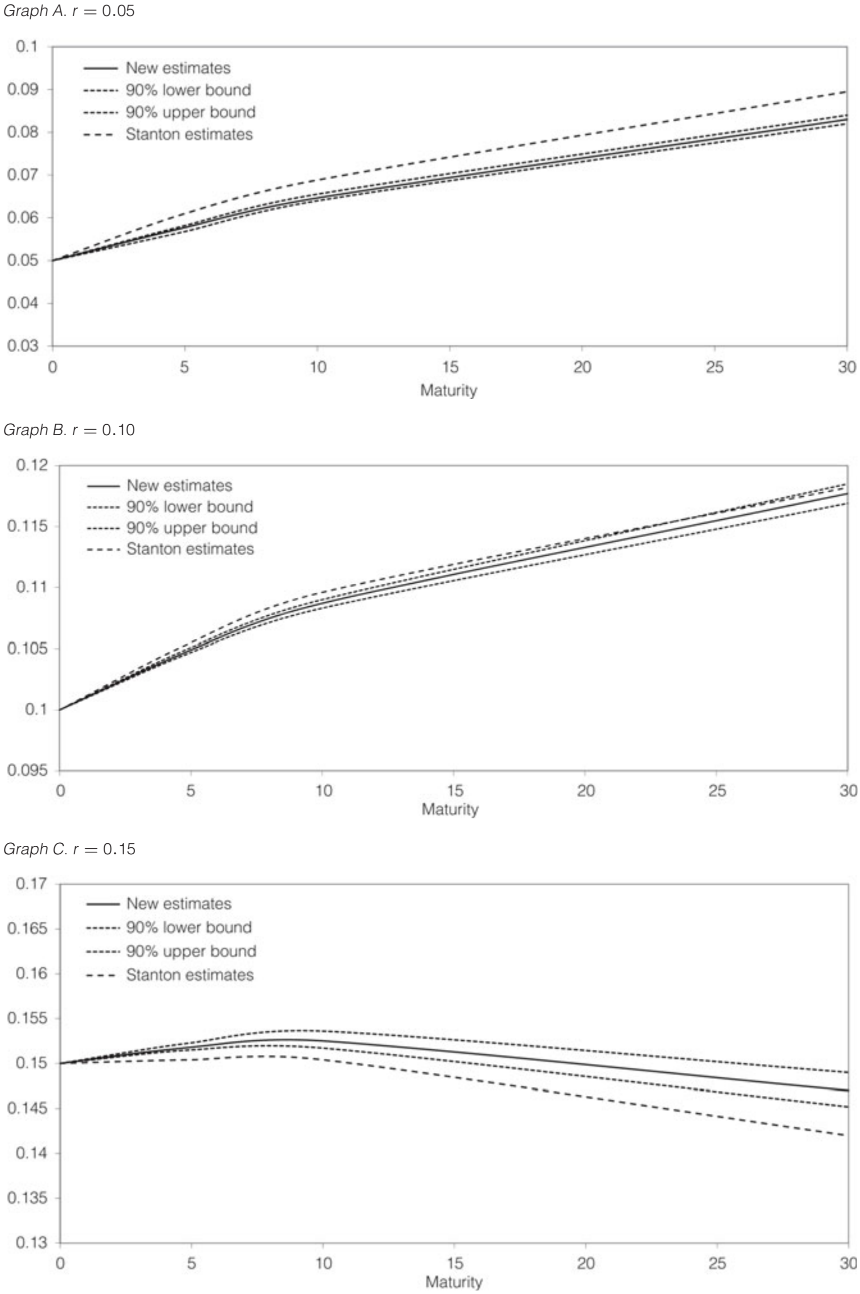


TABLE 4

Valuation of Interest Rate Caps under Alternative Short Rate Processes

Table 4 reports the interest rate cap prices under alternative short rate processes. The numbers in parentheses are standard errors of the nonparametric prices based on the proposed estimates.

			Strike Price (basis points)				
Spot Rate	Cap Tenor	Model	–50	–25	0	25	50
0.05	3	New estimates	2.9242	2.3173	1.7822	1.3309	0.9898
			(0.2075)	(0.3011)	(0.4099)	(0.4780)	(0.4952)
		Stanton	3.3710	2.7599	2.2088	1.7350	1.3448
	5	New estimates	6.2958	5.2953	4.3878	3.5972	2.9309
			(0.3467)	(0.4768)	(0.7144)	(0.7705)	(0.8879)
		Stanton	7.3831	6.3666	5.4275	4.5845	3.8469
0.10	3	New estimates	2.5818	2.0476	1.5873	1.2059	0.9048
			(0.3859)	(0.4812)	(0.5583)	(0.5904)	(0.5727)
		Stanton	2.7330	2.2075	1.7508	1.3663	1.0553
	5	New estimates	5.2007	4.3367	3.5727	2.9188	2.3715
			(0.8318)	(0.9760)	(1.1049)	(1.1808)	(1.2031)
		Stanton	5.5648	4.7126	3.9536	3.2956	2.7354
0.15	3	New estimates	2.6133	2.2121	1.8607	1.5583	1.3039
			(0.9889)	(1.0255)	(1.0295)	(0.9992)	(1.0147)
		Stanton	2.5246	2.1293	1.7818	1.4819	1.2282
	5	New estimates	5.4816	4.8662	4.3269	3.8337	3.4116
			(2.2268)	(2.2623)	(2.2556)	(2.2110)	(2.1381)
		Stanton	5.0021	4.3968	3.8566	3.3813	2.9667

estimator proposed by Stanton (1997). As shown in Chapman and Pearson (2000), the biases of the Stanton (1997) estimator can produce spurious nonlinearities in the drift function. Simulation results show that the proposed method significantly attenuates the spurious nonlinearities of the drift function and improves estimation efficiency. We apply the proposed method to the U.S. short rate diffusion process using a panel of six U.S. Treasury yields with maturities ranging from 3 months to 10 years. Our empirical results corroborate previous findings in the literature about the nonlinearity of the drift function. That is, the short rate process exhibits a strong mean reverting property at high levels. However, the level of mean reversion is substantially weaker than previously documented in the literature. We further show that the differences in the short rate drift and diffusion function estimates have significant economic implications on the pricing of bonds and interest rate derivative securities.

Appendix A. Proofs of Propositions

For simplicity of notation, the superscripts indicating the short rate model and the auxiliary model are omitted in the proofs wherever there is no confusion. Also for notational convenience, $r_{(t+1)}$ and r_t are used to denote $r_{(t+1)\delta}$ and $r_{t\delta}$, respectively. In proving Propositions 1, 2, and 3, we require the following regularity conditions for the diffusion processes and the kernel function:

1. Each of the J interest rate diffusion processes satisfies: i) assumption A1 in Nicolau (2003) or conditions 1 and 2 in Bandi and Phillips (2003) (i.e., each process has a unique strong solution); ii) assumption A2 in Nicolau (2003) (i.e., each process has an invariant density $p_j(r), j = 1, \ldots, J$); and iii) assumption A4 in Nicolau (2003) (i.e., each process is ρ-mixing).

2. The kernel function $K(z)$ is bounded and real-valued, with the following characteristics: i) $\int K(z)\,dz = 1$, ii) $K(z)$ is symmetric about 0, iii) $\int z^2 K(z)dz < \infty$, iv) $|z|K(|z|) \to 0$ as $|z| \to \infty$, and v) $\int K^2(z)dz \leq \infty$.

Proof of Proposition 1. Based on the drift function estimator of Stanton (1997) in equation (3), we have

$$\text{(A-1)}\quad \tilde{\mu}(r) - \mu(r) = \frac{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{r_{(t+1)} - r_t}{\delta}\right) - \tilde{p}(r)\mu(r)}{\tilde{p}(r)}$$

$$= \frac{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right)}{p(r)}\left(1 - \frac{\tilde{p}(r) - p(r)}{\tilde{p}(r)}\right)$$

$$= \frac{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right)}{p(r)} + o_p(1),$$

where $K_h(u) = (1/h)K(u/h)$. The term $((\tilde{p}(r) - p(r))/(\tilde{p}(r)))$ is of order $o_p(1)$ (see, e.g., Pagan and Ullah ((1999), p. 101), Racine and Li (2004), and Sam and Ker (2006)). Hence, ignoring the term $o_p(1)$, we obtain

$$\mathrm{E}\left[\tilde{\mu}(r) - \mu(r)\right] \simeq \frac{1}{np(r)}\mathrm{E}\left[\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right)\right]$$

$$= \frac{1}{np(r)}\mathrm{E}\left[\sum_{t=0}^{n-1} K_h(r_t - r)\,\mathrm{E}_t\left[\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right]\right].$$

It follows from lemma 7 in Nicolau (2003) (see also Hansen and Scheinkman (1995)) that

$$\mathrm{E}_t\left[\frac{r_{(t+1)} - r_t}{\delta}\right] = \mu(r_t) + \frac{1}{2}\left(\mu'(r_t)\mu(r_t) + \frac{\sigma^2(r_t)}{2}\mu''(r_t)\right)\delta + o(\delta).$$

Let $q(r_t) = ½(\mu'(r_t)\mu(r_t) + ½\sigma^2(r_t)\mu''(r_t))$ and with the change of variable $z = (r_t - r)/h$, we have

$$\mathrm{E}\left[\tilde{\mu}(r) - \mu(r)\right] \simeq \frac{1}{np(r)}\mathrm{E}\left[\sum_{t=0}^{n-1} K_h(r_t - r)\left(\mu(r_t) + q(r_t)\delta - \mu(r)\right)\right]$$

$$= \frac{1}{p(r)}\int K(z)(\mu(r + hz) + q(r + hz)\delta - \mu(r))p(r + hz)dz.$$

Thus,

$$\text{(A-2)}\quad \mathrm{E}\left[\tilde{\mu}(r) - \mu(r)\right] \simeq \frac{h^2}{2}\left(\mu''(r) + 2\mu'(r)\frac{p'(r)}{p(r)}\right)\int z^2 K(z)dz$$

$$+ \frac{\delta}{2}\left(\mu'(r)\mu(r) + \frac{\sigma^2(r)}{2}\mu''(r)\right).$$

Now, turning to the derivation of variance, from equation (A-1), and ignoring the term of order $o_p(1)$, we have

$$\begin{aligned}\operatorname{var}[\tilde{\mu}(r)] &\simeq \operatorname{var}\left[\frac{\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right)}{p(r)}\right]\\ &= \frac{1}{p^2(r)}\frac{1}{Th}\operatorname{var}\left[\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} v(r_t, r_{(t+1)})\right],\end{aligned}$$

where

$$v(r_t, r_{(t+1)}) = \sqrt{\frac{\delta}{h}} K\left(\frac{r_t - r}{h}\right)\left(\frac{\left(r_{(t+1)} - r_t\right)}{\delta} - \mu(r)\right).$$

Let $\dot{v}(r_t, r_{(t+1)}) = v(r_t, r_{(t+1)}) - \mathrm{E}[v(r_t, r_{(t+1)})]$, and we have $\mathrm{E}[\dot{v}(r_t, r_{(t+1)})] = 0$ and $\mathrm{E}[\dot{v}^2(r_t, r_{(t+1)})] = \mathrm{E}[v^2(r_t, r_{(t+1)})] - (\mathrm{E}[v(r_t, r_{(t+1)})])^2$, with

$$\begin{aligned}\mathrm{E}[v(r_t, r_{(t+1)})] &= \sqrt{\frac{\delta}{h}}\,\mathrm{E}\left[\frac{K(r_t - r)}{h}\left(\frac{\left(r_{(t+1)} - r_t\right)}{\delta} - \mu(r)\right)\right]\\ &= \sqrt{h\delta}\,O(h^2 + \delta) \quad \to \quad 0,\\ \mathrm{E}\left[v^2(r_t, r_{(t+1)})\right] &= \frac{\delta}{h}\,\mathrm{E}\left[K^2\left(\frac{r_t - r}{h}\right)\mathrm{E}_t\left[\left(\frac{r_{(t+1)} - r_t}{\delta} - \mu(r)\right)^2\right]\right].\end{aligned}$$

It follows from lemma 7 in Nicolau (2003) that

$$\begin{aligned}\mathrm{E}_t\left[\left(\frac{r_{(t+1)} - r_t}{\delta}\right)^2\right] &= \sigma^2(r) + \mu^2(r)\delta + O(\delta), \qquad \text{and}\\ \mathrm{E}\left[v^2(r_t, r_{(t+1)})\right] &= \int K^2(z)dz\left(\sigma^2(r + hz) + O(h^2 + \delta)\right)p(r)\\ &= \sigma^2(r)p(r)\int K^2(z)dz + O(h^2 + \delta) \quad < \quad \infty.\end{aligned}$$

Next, we derive the autocovariance of $v(r_t, r_{(t+1)})$. For $t = 0$,

$$\begin{aligned}&\mathrm{E}[v(r_0\,; r_1)v(r_1\,; r_2)] =\\ &\quad \frac{\delta}{h}\,\mathrm{E}\left[K\left(\frac{r_0 - r}{h}\right)K\left(\frac{r_1 - r}{h}\right)\left(\frac{r_1 - r_0}{\delta} - \mu(r)\right)\left(\frac{r_2 - r_1}{\delta} - \mu(r)\right)\right].\end{aligned}$$

By the law of iterated expectations,

$$\begin{aligned}&\mathrm{E}\left[v(r_0\,; r_1)v(r_1\,; r_2)\right]\\ &= \frac{\delta}{h}\,\mathrm{E}\left[K^2\left(\frac{r_0 - r}{h}\right)((\mu(r_0 - \mu(r))^2 + O(\delta)\right]\\ &= \delta\int K^2(z)dz\left((\mu(r) + O(h) - \mu(r))^2 + O(\delta)\right)(p(r) + O(h))\\ &= O(h^2 + \delta)\delta \quad \to \quad 0.\end{aligned}$$

Under regularity condition that the short rate process is ρ-mixing, for $t \geq 1$, $|\,\mathrm{E}[v(r_0\,;r_1)v(r_t\,;r_{(t+1)})]| \leq |\,\mathrm{E}[v(r_0\,;r_1)v(r_1\,;r_2)]| \to 0$. Finally, it follows from lemma 9 in Nicolau (2003) that

$$\text{(A-3)} \qquad \operatorname{var}[\tilde{\mu}(r)] \simeq \frac{1}{(Th)p(r)^2}\left(\mathrm{E}\left[\dot{v}^2\left(r_t, r_{(t+1)}\right)\right] + o(1)\right) = \frac{\sigma^2(r)\int K^2(z)dz + o(1)}{(Th)p(r)}.$$

This concludes the proof of Proposition 1. □

Before proving Propositions 2 and 3, we first prove the lemma below. The proof of Proposition 2 makes use of the properties of $\hat{\mu}_p(r)$, while the proof of Proposition 3 depends on the weak consistency of $\hat{\mu}_p(r)$. Under regularity conditions on model (1) and auxiliary model (2), the bias and variance of the pooled estimator of the drift function are essentially the same as the estimator in Stanton (1997). We thus only prove the weak consistency of the pooled drift function estimator $\hat{\mu}_p(r)$ in the following lemma:

Lemma. Under the assumptions of Proposition 1, the pooled estimator $\hat{\mu}_p(r)$ defined in equation (9) is weakly consistent; that is, $\hat{\mu}_p(r) - \mu_p(r) = \epsilon_N \to 0$ in probability as $N \to \infty$, where $N = JT$.

Proof of Lemma. From equation (9), we have

$$\hat{\mu}_p(r) = \frac{\frac{1}{nJ}\sum_{j=1}^{J}\sum_{t=0}^{n-1}\left(\frac{r_{(t+1)}^{(j)} - r_t^{(j)}}{\delta}\right)K_{h_p}\left(r_t^{(j)} - r\right)}{\frac{1}{nJ}\sum_{j=1}^{J}\sum_{t=0}^{n-1}K_{h_p}\left(r_t^{(j)} - r\right)} = \frac{\frac{1}{nJ}\sum_{j=1}^{J}\sum_{t=0}^{n-1}\left(\frac{r_{(t+1)}^{(j)} - r_t^{(j)}}{\delta}\right)K_{h_p}\left(r_t^{(j)} - r\right)}{\frac{1}{J}\sum_{j=1}^{J}\hat{p}_j(r)}.$$

Using the fact that

$$\frac{1}{n}\sum_{t=0}^{n-1}\left(\frac{r_{(t+1)}^{(j)} - r_t^{(j)}}{\delta}\right)K_{h_p}\left(r_t^{(j)} - r\right) = \hat{\mu}_j(r)\hat{p}_j(r),$$

we can rewrite $\hat{\mu}_p(r)$ as $\hat{\mu}_p(r) = (\sum_{j=1}^{J}\hat{\mu}_j(r)\hat{p}_j(r))/(\sum_{j=1}^{J}\hat{p}_j(r))$. Applying Slutsky's theorem, we have

$$\operatorname{plim}\hat{\mu}_p(r) = \frac{\sum_{j=1}^{J}\operatorname{plim}\hat{\mu}_j(r)\operatorname{plim}\hat{p}_j(r)}{\sum_{j=1}^{J}\operatorname{plim}\hat{p}_j(r)} = \frac{\sum_{j=1}^{J}\mu_j(r)p_j(r)}{\sum_{j=1}^{J}p_j(r)} = \mu_p(r),$$

which proves weak consistency of the pooled estimator of the drift function.

Proof of Proposition 2. Applying Taylor series expansion to $(\hat{\mu_p}(r))/(\hat{\mu_p}(r_t))$ around $(\mu_p(r))/(\mu_p(r_t))$, we have the following result for the drift function estimator in equation (10):

$$\hat{\mu}(r) \simeq \frac{1}{\hat{p}(r)}(A_n + B_n),$$

where

$$A_n = \frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\frac{(r_{(t+1)} - r_t)}{\delta}\left(\frac{\mu_p(r)}{\mu_p(r_t)}\right) \quad \text{and}$$

$$B_n = \frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\frac{(r_{(t+1)} - r_t)}{\delta}\left(\frac{\hat{\mu_p}(r) - \mu_p(r)}{\mu_p(r_t)} - \frac{\hat{\mu_p}(r_t) - \mu_p(r_t)}{\mu_p(r_t)}\frac{\mu_p(r)}{\mu_p(r_t)}\right).$$

Using the fact that $\mu(r) = \mu_p(r)c(r)$, and ignoring the term of order $o_p(1)$, we have

$$\text{(A-4)} \qquad \mathrm{E}\left[\hat{\mu}(r) - \mu(r)\right] \simeq \frac{\mathrm{E}\left[A_n - \mu_p(r)c(r)\hat{p}(r)\right]}{p(r)} + \frac{\mathrm{E}\left[B_n\right]}{p(r)}.$$

For the numerator of the first term on the right-hand side of equation (A-4), we have,

$$\begin{aligned}
&\mathrm{E}\left[A_n - \mu_p(r)c(r)\hat{p}(r)\right] \\
&= \frac{\mu_p(r)}{n}\mathrm{E}\left[\sum_{t=0}^{n-1} K_h(r_t - r)\,\mathrm{E}_t\left[\frac{(r_{(t+1)} - r_t)}{\delta\mu_p(r_t)} - c(r)\right]\right] \\
&= \frac{h^2}{2}\left(c''(r)p(r) + 2c'(r)p'(r)\right)\mu_p(r)\int z^2K(z)dz + l(r)p(r)\delta + o(h^2 + \delta),
\end{aligned}$$

where $l(r_t) = ½(c'(r_t)c(r_t) + ½\sigma^2(r_t)c''(r_t))$. From the properties of the pooled drift function estimator, we have that $\mathrm{E}[B_n]$ is of order $O(h_p^4 + h_p^2\delta)$. Hence,

$$\text{(A-5)} \qquad \mathrm{E}\left[\hat{\mu}(r) - \mu(r)\right] \simeq \frac{h^2}{2}\left(c''(r) + 2c'(r)\frac{p_1'(r)}{p_1(r)}\right)\mu_p(r)\int z^2K(z)dz + \frac{\delta}{2}l(r) + o(h^2 + \delta).$$

Now, turning to the derivation of variance, for example,

$$\text{(A-6)} \qquad \mathrm{var}\left[\hat{\mu}(r)\right] = \mathrm{var}\left[\frac{A_n}{\hat{p}(r)}\right] + 2\,\mathrm{cov}\left[\frac{A_n}{\hat{p}(r)}, \frac{B_n}{\hat{p}(r)}\right] + \mathrm{var}\left[\frac{B_n}{\hat{p}(r)}\right].$$

Note that

$$\begin{aligned}
\mathrm{var}\left[\frac{A_n}{\hat{p}(r)}\right] &\simeq \mathrm{var}\left[\frac{A_n - \hat{p}(r)\mu(r)}{p(r)}\right] \\
&= \frac{\mu_p^2(r)}{p^2(r)}\mathrm{var}\left[\frac{1}{n}\sum_{t=0}^{n-1} K_h(r_t - r)\left(\frac{(r_{(t+1)} - r_t)}{\delta\mu_p(r_t)} - c(r)\right)\right] \\
&= \frac{1}{p^2(r)}\frac{1}{Th}\mathrm{var}\left[\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_t, r_{(t+1)})\right],
\end{aligned}$$

where

$$g(r_t, r_{(t+1)}) = \mu_p(r)\sqrt{\frac{\delta}{h}}K\left(\frac{r_t - r}{h}\right)\left(\frac{(r_{(t+1)} - r_t)}{\delta\mu_p(r_t)} - c(r)\right).$$

Let $\dot{g}(r_t, r_{(t+1)}) = g(r_t, r_{(t+1)}) - \mathrm{E}[g(r_t, r_{(t+1)})]$, and we have $\mathrm{E}[\dot{g}(r_t, r_{(t+1)})] = 0$ and $\mathrm{E}[\dot{g}^2(r_t, r_{(t+1)})] = \mathrm{E}[g^2(r_t, r_{(t+1)})] - (\mathrm{E}[g(r_t, r_{(t+1)})])^2$. Similar to the proof of Proposition 1, we have $\mathrm{E}[g(r_t, r_{(t+1)})] = \mu_p(r)\sqrt{h\delta}O(h^2+\delta) \to 0$, $\mathrm{E}[g^2(r_t, r_{(t+1)})] = \left(\sigma^2(r)\int K^2(z)dz\right) p(r) + O(h^2+\delta) < \infty$.

Next, we derive the autocovariance function of $g(r_t, r_{(t+1)})$. For $t = 0$,

$$\mathrm{E}[g(r_0\,;r_1)g(r_1\,;r_2)] = \mu_p^2(r)\frac{\delta}{h}\,\mathrm{E}\left[K\left(\frac{r_0-r}{h}\right)K\left(\frac{r_1-r}{h}\right)\left(\frac{r_1-r_0}{\delta\mu_p(r_0)}-c(r)\right)\left(\frac{r_2-r_1}{\delta\mu_p(r_1)}-c(r)\right)\right].$$

Again similar to the proof of Proposition 1, we have $\mathrm{E}[g(r_0\,;r_1)g(r_1\,;r_2)] = \mu_p^2(r)O(h^2+\delta)$ $\delta \to 0$, and $|\,\mathrm{E}[g(r_0\,;r_1)\,g(r_t\,;r_{(t+1)})]| \le |\,\mathrm{E}[g(r_0\,;r_1)\,g(r_1\,;r_2)]| \to 0$, for $t \ge 1$. It follows from lemma 9 in Nicolau (2003) that

$$\text{(A-7)}\quad \mathrm{var}\left[\frac{A_n}{\hat{p}(r)}\right] \simeq \frac{1}{Th}\frac{1}{p(r)^2}\,\mathrm{E}\left[\dot{g}^2(r_t, r_{(t+1)})\right] + o(1) = \frac{\sigma^2(r)R(K)+o(1)}{(Th)p(r)}.$$

To derive the covariance between $(A_n)/(p(r))$ and $(B_n)/(p(r))$, we have

$$\text{(A-8)}\quad \mathrm{cov}[A_n, B_n] = \frac{1}{n^2}\sum_{t=0}^{n-1}\mathrm{E}[g_1(r_t)g_2(r_t)] + \frac{2}{n^2}\sum_{t<t'}\mathrm{E}[g_1(r_t)g_2(r_{t'\delta})] - \mathrm{E}[A_n]\,\mathrm{E}[B_n],$$

where

$$g_1(r_t) = K_h(r_t - r)\left(\frac{\left(r_{(t+1)}-r_t\right)}{\delta}\,\frac{\mu_p(r)}{\mu_p(r_t)}\right) \qquad \text{and}$$

$$g_2(r_t) = K_h(r_t - r)\left(\frac{\left(r_{(t+1)}-r_t\right)}{\delta}\left(\frac{\hat{\mu}_p(r)-\mu_p(r)}{\mu_p(r_t)} - \frac{\hat{\mu}_p(r_t)-\mu_p(r_t)}{\mu_p(r_t)}\,\frac{\mu_p(r)}{\mu_p(r_t)}\right)\right).$$

It can be shown that the second term is negligible, whereas the first term, $(1/n^2)\,\mathrm{E}[g_1(r_t)$ $g_2(r_t)]$, is of the order $o((Th)^{-1})$, and the last term $\mathrm{E}[A_n]\,\mathrm{E}[B_n] = O(1)O(h_p^4 + h_p^2\delta)$.

Finally, using the statistical properties of the pooled estimator of the drift function, for example, $\mathrm{E}[\hat{\mu}_p(r) - \mu_p(r)] = O(h_p^2 + \delta)$ and $h_p^4 \propto (Nh_p)^{-1}$, where $N = JT$, it leads to the following result:

$$\text{(A-9)}\quad \mathrm{var}[B_n] = O\left((Nh_p)^{-1}\right).$$

This concludes the proof of Proposition 2. □

Proof of Proposition 3. Denote $\hat{B}_p(r) = \hat{\mu}_p(r) - \mu_p(r)$. By weak consistency of the pooled estimator of the drift function from the previous lemma, $\hat{B}_p(r) = \epsilon_N \to 0$ in the limit. Thus, we have

$$\text{(A-10)}\quad \hat{\mu}(r) - \mu(r) = \frac{(\hat{\mu}(r)-\mu(r))\,\hat{p}(r)}{\hat{p}(r)} = \frac{\left(\hat{\mu}(r)-\mu(r)-c(r)\hat{B}_p(r)+c(r)\hat{B}_p(r)\right)\hat{p}(r)}{\hat{p}(r)}.$$

Hence,

$$\sqrt{\frac{Th}{R(K)}\hat{p}(r)}\left(\frac{\hat{\mu}(r)-\mu(r)}{\sigma(r)}\right) = \frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}g(r_t\,;r_{(t+1)})}{\sigma(r)\sqrt{R(K)\hat{p}(r)}} + \sqrt{Th}(C_n) + \sqrt{Th}(D_n),$$

where $R(K) = \int K^2(z)dz,\ g(r_t\,;r_{(t+1)})$ is as defined in the proof of Proposition 2,

$$C_n = \frac{1}{\sigma(r)\sqrt{R(K)\hat{p}(r)}}\frac{1}{nh}\sum_{t=0}^{n-1} K\left(\frac{r_t - r}{h}\right)\left(\frac{r_{(t+1)} - r_t}{\delta\mu_p(r_t)} - c(r)\right)(\epsilon_N), \qquad \text{and}$$

$$D_n = \frac{1}{\sigma(r)\sqrt{R(K)\hat{p}(r)}}\frac{1}{nh}\sum_{t=0}^{n-1} K\left(\frac{r_t - r}{h}\right)\left(\frac{r_{(t+1)} - r_t}{\delta\mu_p(r_t)}\frac{\mu_p(r)}{\mu_p(r_t)} - c(r)\right)(\epsilon_N).$$

Applying Slutsky's theorem, we have that $\sqrt{Th}(C_n) = \sqrt{Th}(D_n) = o_p(1)$, and thus these terms are asymptotically negligible. Since

$$\frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g\left(r_t\,;r_{(t+1)}\right)}{\sigma(r)\sqrt{R(K)\hat{p}(r)}} \quad \text{and} \quad \frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g\left(r_t\,;r_{(t+1)}\right)}{\sigma(r)\sqrt{R(K)p(r)}}$$

have the same asymptotic distribution, it is sufficient to show that

$$\frac{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g\left(r_t\,;r_{(t+1)}\right)}{\sigma(r)\sqrt{R(K)p(r)}} \quad \to \quad N(0,1).$$

Note that from the proof of Proposition 2, we have $\mathrm{E}[g(r_t\,;r_{(t+1)})] = \mu_p(r)\sqrt{h\delta}\,\{O(h^2+\delta)\}$. Hence,

$$\text{(A-11)} \qquad \mathrm{E}\left[\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g\left(r_t\,;r_{(t+1)}\right)\right] = \mu_p(r)(\sqrt{nh^5}+\sqrt{nh}\delta)O(\sqrt{\delta}) \quad \to \quad 0.$$

Under the assumptions of Proposition 3 and from the proof of Proposition 2, we have

$$\text{(A-12)} \qquad \mathrm{E}[g^2(r_t\,;r_{(t+1)})] \quad \to \quad \sigma^2(r)p(r)\int K^2(z)dz \qquad \text{and}$$

$$\text{(A-13)} \qquad \mathrm{E}[g(r_0\,;r_1)g(r_t\,;r_{(t+1)})] \quad \to \quad 0 \quad \forall t \geq 1.$$

It follows from equations (A-11), (A-12), (A-13), and lemma 9 of Nicolau (2003) that

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} g(r_t\,;r_{(t+1)}) \quad \to \quad N\left(0, \sigma^2(r)p(r)\int K^2(z)dz\right).$$

This concludes the proof of Proposition 3. □

Appendix B. Drift Functions of Yields with Different Maturities

In this appendix, we show that model specifications in equations (1) and (2) are consistent under the term structure model framework and illustrate that the drift functions of yields with different maturities are similar in shape. Let r_t be the instantaneous interest rate with the following risk-neutral process:

$$\text{(B-1)} \qquad dr_t = (\mu(r_t) - \lambda(r_t))dt + \sigma(r_t)d\tilde{w}_t,$$

where $\tilde{w}_t$ is a standard Brownian motion under the equivalent martingale measure Q, and $\lambda(r_t)$ is the market price of interest rate risk. The bond price $P(r_t, \tau_i) = \mathrm{E}_t^Q[\exp^{-\int_t^{t+\tau_i} r_u du}]$ is a function of the short rate r_t and maturity $\tau_i = T_i - t, i = 1, 2, \ldots$. Similarly, the yield with maturity τ_i is a function of the short rate r_t and τ_i, i.e., $r_t^{(i)} = y_i(r_t) = -(1/\tau_i)\ln P(r_t, \tau_i)$. Following Ingersoll ((1987), p. 394), the price of a zero-coupon bond with maturity τ_i evolves according to $(dP(r_t, \tau_i))/(P(r_t, \tau_i)) = \xi(r_t, \tau_i)dt + \nu(r_t, \tau_i)d\tilde{w}_t$ where, by Itô's lemma, $\xi(r_t, \tau_i)P = ½\sigma^2(r_t)P_{rr} + \mu(r_t)P_r + P_t$, and $\nu(r_t, \tau_i)P = \sigma(r_t)P_r$. Further following Ingersoll (1987), the risk premium is proportional to the standard deviation (i.e., $\xi(r_t, \tau_i) = r_t + \lambda(r_t)P_r/P$). This leads to the following partial differential equation (PDE) for bond price:

$$\text{(B-2)} \qquad \frac{1}{2}\sigma^2(r_t)P_{rr} + (\mu(r_t) - \lambda(r_t))P_r + P_t - rP = 0,$$

with boundary condition $P(r_t, 0) = 1$. Since $r_t^{(i)} = -(1/\tau_i)\ln P(r_t, \tau_i)$, by Itô's lemma, we have

$$dr_t^{(i)} = \left(\left(\frac{1}{2}\nu^2(r_t, \tau_i) - \xi(r_t, \tau_i)\right)\Big/\tau_i - \frac{1}{\tau_i^2}\ln P\right)dt + (\nu(r_t, \tau_i)/\tau_i)d\tilde{w}_t.$$

Note that $-(1/\tau_i)\ln P = r_t^{(i)}$ and $r_t = y_i^{-1}(r_t^{(i)})$, where $y_i^{-1}(\cdot)$ denotes the inverse function of $y_i(\cdot)$, thus both the drift and diffusion in the above process are functions of $r_t^{(i)}$. That is,

$$\text{(B-3)} \qquad dr_t^{(i)} = \mu_i\left(r_t^{(i)}\right)dt + \sigma_i\left(r_t^{(i)}\right)d\tilde{w}_t.$$

Moreover, we assume that the bond yields are observed during the observation interval with error, such as measurement error, market microstructure noise, and idiosyncratic shock to a specific maturity. That is, the observed yields are given by $\hat{r}_t^{(i)} = r_t^{(i)} + \int_{t-1}^{t}\nu_{i,u}dw_{i,u}^M$, where the error terms are driven by Brownian motions $w_{i,t}^M$ that are uncorrelated across maturities. Thus, we have

$$\text{(B-4)} \qquad d\hat{r}_t^{(i)} = \mu_i(r_t^{(i)})dt + \hat{\sigma}_i(r_t^{(i)})d\tilde{w}_t^{(i)},$$

where $\hat{\sigma}_i(r_t^{(i)})d\tilde{w}_t^{(i)} = \sigma_i(r_t^{(i)})d\tilde{w}_t + \nu_{i,t}dw_{i,t}^M$. That is, the processes of yields with different maturities are no longer perfectly correlated.

In the following, we illustrate graphically the similarities of the drift functions of yields with different maturities using two models: i) the well-known affine model considered by Duffie and Kan (1996) and Dai and Singleton (2000), and ii) the nonlinear model of Aït-Sahalia (1996b). The affine model is restrictive but has the advantage that it has closed-form solutions for bond prices and yields.

The general affine model is specified as

$$\text{(B-5)} \qquad dr_t = (\alpha - \beta r_t)dt + \sqrt{\sigma_0 + \sigma_1 r_t}dw_t.$$

The model reduces to the Vasicek (1977) model when $\sigma_1 = 0$ and the CIR (1985) model when $\sigma_0 = 0$. The bond yield with maturity τ_i has the structure $r_t^{(i)} = a(\tau_i) + b(\tau_i)r_t$. The coefficients $a(\tau_i)$ and $b(\tau_i)$ are solved from the PDE in equation (B-2) as

$$\begin{aligned} a(\tau_i) &= \frac{2\alpha}{\sigma_1\tau_i}\ln\left(\frac{g(\tau_i)}{2\gamma e^{(\beta+\lambda-\gamma)\tau_i/2}}\right) \\ &\quad + \frac{\sigma_0}{2\sigma_1\tau_i}\left[\frac{(\beta+\lambda-\gamma)^2(g(\tau_i)-2\gamma)}{2\gamma g(\tau_i)} + 4(\beta+\lambda)\ln\frac{g(\tau_i)}{2\gamma} + 4\gamma(\beta+\lambda+\gamma)\tau_i\right], \\ b(\tau_i) &= \frac{2(1-e^{-\gamma\tau_i})}{\tau_i g(\tau_i)}, \end{aligned}$$

where $g(\tau_i) = 2\gamma + (\beta + \lambda - \gamma)(1 - e^{-\gamma\tau_i})$, $\gamma = \sqrt{(\beta + \lambda)^2 + 2\sigma_1}$, and λ denotes the market price of interest rate risk. This implies that

$$dr_t^{(i)} = \left(-a'(\tau_i)dt - b'(\tau_i)r_t + b(\tau_i)(\alpha - \beta r_t)\right) dt + b(\tau_i)\sqrt{\sigma_0 + \sigma_1 r_t}dw_t.$$

Substituting $r_t = (r_t^{(i)} - a(\tau_i))/b(\tau_i)$, we have an affine diffusion process for $r_t^{(i)}$.

The nonlinear model used in our illustration is the one specified in Aït-Sahalia (1996b):

$$\text{(B-6)} \qquad dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1})dt + \sigma r_t^\gamma dw_t.$$

The bond price $P(r_t, \tau_i)$ no longer has a closed-form solution; we obtain the solution by solving the PDE in equation (B-2) numerically. The subroutine pdepe of Matlab is used to obtain both the bond price P and its partial derivative with respect to the short rate P_r.

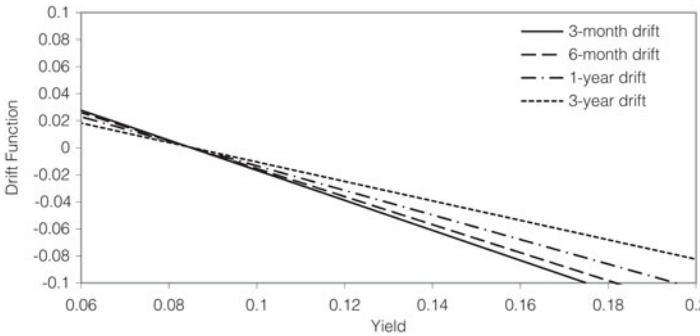
Graphs A and B of Figure B1 plot the drift functions of yields for the affine and nonlinear models with four different maturities (i.e., τ_i = 3 months, 6 months, 1 year, and 3 years). The parameter values of the affine model are obtained from Chapman and Pearson (2000): α=0.0731, β=0.853, $\sqrt{\sigma_1}$=0.1566, and λ is set to 0 for simplicity. The parameter values of the nonlinear model are obtained from Aït-Sahalia (1996b): $\alpha_0 = -0.00464$; $\alpha_1 = 0.0433$; $\alpha_2 = -0.1143$; $\alpha_3 = 0.00013$; $\sigma = 0.00968$; $\gamma = 2.073$, except that we set higher values for α_2 and α_3 to have more nonlinearity; and λ is set to 0 for simplicity. We also consider alternative values for the parameters, and the drift functions remain similar.

FIGURE B1

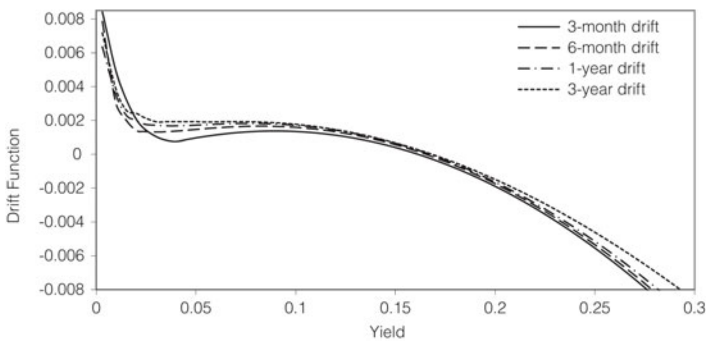
Drift Functions of Bond Yields

Graphs A and B of Figure B1 plot drift functions of yields with maturities equal to 3 months, 6 months, 1 year, and 3 years under, respectively, the CIR (1985) model and the Aït-Sahalia (1996b) model. Parameter values of the models are given in Appendix B.

Graph A. Drift Functions of Bond Yields under the CIR Model



Graph B. Drift Functions of Bond Yields under the Nonlinear Model



References

Abaffy, J.; M. Bertocchi; J. Dupacova; R. Giacometti; M. Huskova; and V. Moriggia. "A Nonparametric Model for Analysis of the Euro Bond Market." *Journal of Economic Dynamics and Control*, 27 (2003), 1113–1131.

Abhyankar, A., and D. Basu. "Does Conditioning Information Matter in Estimating Continuous Time Interest Rate Diffusions?" *Journal of Financial and Quantitative Analysis*, 36 (2001), 335–344.

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

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

Aït-Sahalia, Y. "Transition Densities for Interest Rate and Other Nonlinear Diffusions." *Journal of Finance*, 54 (1999), 1361–1395.

Andersen, T. G., and J. Lund. "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate." *Journal of Econometrics*, 77 (1997), 343–377.

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

Backus, D.; S. Foresi; and S. Zin. "Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing." *Journal of Business and Economic Statistics*, 16 (1998), 13–26.

Bandi, F. M., and P. C. B. Phillips. "Fully Nonparametric Estimation of Scalar Diffusion Models." *Econometrica*, 71 (2003), 241–283.

Banon, G. "Nonparametric Identification for Diffusion Processes." *S.I.A.M. Journal of Control and Optimization,* 16 (1978), 380–395.

Boyle, P.; M. Broadie; and P. Glasserman. "Monte Carlo Methods for Security Pricing." *Journal of Economic Dynamics and Control*, 21 (1997), 1267–1321.

Brennan, M. J., and E. S. Schwartz. "A Continuous Time Approach to the Pricing of Bonds." *Journal of Banking and Finance*, 3 (1979), 133–155.

Canabarro, E. "Where Do One-Factor Interest Rate Models Fail?" *Journal of Fixed Income*, 5 (1995), 31–52.

Chan, K. C.; A. G. Karolyi; F. A. Longstaff; and A. B. Sanders. "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate." *Journal of Finance*, 47 (1992), 1209–1227.

Chapman, D., and N. Pearson. "Is the Short Rate Drift Actually Nonlinear?" *Journal of Finance*, 55 (2000), 355–388.

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

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

Cox, J. C. "Notes on Option Pricing I: Constant Elasticity of Variance Diffusion." Working Paper, Stanford University (1975).

Cox, J. C.; J. E. Ingersoll; and S. A. Ross. "An Analysis of Variable Rate Loan Contracts." *Journal of Finance*, 35 (1980) 389–403.

Cox, J. C.; J. E. Ingersoll; and S. A. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53 (1985), 385–407.

Dai, Q., and K. J. Singleton. "Specification Analysis of Term Structure Models." *Journal of Finance*, 55 (2000), 1943–1978.

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

Fan, J., and C. Zhang. "A Reexamination of Diffusion Estimators with Applications to Financial Model Validation." *Journal of the American Statistical Association*, 98 (2003), 118–134.

Florens-Zmirou, D. "On Estimating the Diffusion Coefficient from Discrete Observations." *Journal of Applied Probability*, 30 (1993), 790–804.

Glad, I. K. "Parametrically Guided Non-Parametric Regression." *Scandinavian Journal of Statistics*, 25 (1998), 649–668.

Hansen, L., and J. Scheinkman. "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes." *Econometrica*, 63 (1995), 767–804.

Hjort, N. L., and I. K. Glad. "Nonparametric Density Estimation with a Parametric Start." *The Annals of Statistics*, 23 (1995), 882–904.

Hong, Y., and H. Li. "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates." *Review of Financial Studies*, 18 (2005), 37–84.

Ingersoll, J. E., Jr. *Theory of Financial Decision Making*. Lanham, MD: Rowman and Littlefield (1987).

Jiang, G. J. "Nonparametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Price of Derivative Securities." *Journal of Financial and Quantitative Analysis*, 33 (1998), 465–497.

Jiang, G. J., and J. L. Knight. "A Nonparametric Approach to the Estimation of Diffusion Processes with an Application to a Short-Term Interest Rate Model." *Econometric Theory*, 13 (1997), 615–645.

Jones, C. S. "Nonlinear Mean Reversion in the Short-Term Interest Rate." *Review of Financial Studies*, 16 (2003), 793–843.

Lee, M., and W. Li. "Drift and Diffusion Function Specification for Short-Term Interest Rates." *Economics Letters*, 86 (2005), 339–346.

Li, M.; N. D. Pearson; and A. M. Poteshman. "Conditional Estimation of Diffusion Processes." *Journal of Financial Economics*, 74 (2004), 31–66.

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

Merton, R. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics*, 8 (1980), 323–361.

Nicolau, J. "Bias Reduction in Nonparametric Diffusion Coefficient Estimation." *Econometric Theory*, 19 (2003), 754–777.

Pagan, A., and A. Ullah. *Nonparametric Econometrics*. New York, NY: Cambridge University Press (1999).

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

Pritsker, M. "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models." *Review of Financial Studies*, 11 (1998), 449–487.

Racine, J., and Q. Li. "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data." *Journal of Econometrics*, 119 (2004), 99–130.

Rice, J. A. "Boundary Modification for Kernel Regression." *Communications in Statistics, Theory and Methods*, 13 (1984), 893–900.

Sam, A. G., and A. P. Ker. "Nonparametric Regression under Alternative Data Environments." *Statistics and Probability Letters*, 76 (2006), 1037–1046.

Stanton, R. "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk." *Journal of Finance*, 52 (1997), 1973–2002.

Vasicek, O. A. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics*, 5 (1977), 177–188.