Nonparametric Estimation of the Short Rate Diffusion Process from a Panel of Yields¹

Abdoul G. Sam and George J. Jiang

April 2007

¹Sam is from the Department of AED Economics, The Ohio State University, Columbus OH, 43210. E-mail: sam.7@osu.edu. Jiang is from the Department of Finance, Eller College of Management, University of Arizona, Tucson AZ, 85721. E-mail: gjiang@email.arizona.edu. We would like to thank Matt Pritsker, Alfonso Flores-Lagunes, Kei Hirano, Alan Ker and seminar participants at the University of Arizona, the World Congress of Econometrics Society in London, U.K. for helpful comments and suggestions. The usual disclaimer applies.
Nonparametric Estimation of the Short Rate Diffusion Process from a Panel of Yields

ABSTRACT

In this paper, we propose a nonparametric estimator of the short rate diffusion process using observations from a panel of yields. The proposed estimator can greatly reduce the finite sample bias of the nonparametric estimator proposed in Stanton (1997) based on 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 the US short rate process using a panel of six Treasury yields. With 42 years’ 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.
I. Introduction

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

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

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):

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

where both the drift function and diffusion function are specified to capture potential nonlinearities. Restrictions on the parameters 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 model in Pearson and Sun (1994), the constant elasticity of volatility (CEV) model of Cox (1975) and Cox, Ingersoll and Ross (1980) ($\alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = 0, \gamma = 3/2$), the CEV models of Chan, Karolyi, Longstaff and Sanders (1992) ($\alpha_2 = \alpha_3 = 0$), etc.

While the main appeal of parametric models is the tractability and a 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, Karolyi, Longstaff and Sanders (1992), Aït-Sahalia (1996b), Andersen and Lund (1997), and Hong and Li (2005) among others. 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 Zhin (1998) and Canabarro (1995).
For the above reasons, nonparametric modeling of the short rate dynamics has received considerable attention in recent years. Various nonparametric estimators of the drift and diffusion functions have been proposed in the finance literature, building on theoretical developments in the statistics literature. Briefly, in the statistical literature, Basawa and Prakasa-Rao (1980) and Florens-Zmirou (1993) propose nonparametric estimators of the diffusion function based on a continuous-time record of observations, while Banon (1978) examines the estimation of the drift function when the diffusion function is either a known function or a constant. 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. Nicolau (2003) refines the diffusion estimator in Florens-Zmirou (1993) and Jiang and Knight (1997) to reduce the finite sample bias. 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 the 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) suggests 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). The Monte Carlo simulations in Chapman and Pearson (2000) show that the nonparametric drift function estimator proposed by Stanton (1997) can produce spurious nonlinearities.
even when the underlying drift function is truly linear. What is more troublesome is the fact that the spurious nonlinearity has a pattern similar to that of the empirically estimated 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 patterns of 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.

It is known in the literature that while increasing the sampling frequency is helpful for the identification and estimation of the diffusion function, it is the increase of 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 of interest rates, identifying the drift function requires a large number of sampling observations for a given sampling frequency or equivalently a long sampling 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 of the short rate diffusion process. The key difference between the proposed estimator and existing parametric and nonparametric estimators of the short rate diffusion is that the new method uses information from a panel of yields with different maturities instead of a single time series of short rate observations. 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 interest rates with different maturities are identical, then pooling the data of all the yields will result in the optimal estimator of the short rate drift function. In practice, while the drift functions of interest rates with different maturities may be similar in functional shape, they are unlikely 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 estimator of the drift function is similar to the drift function of the
short rate in functional shape, the correction factor is a smooth function and 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 by Stanton (1997). The simulations confirm that using a panel of yields with different maturities, the proposed estimator can lead to significant efficiency gains relative to the conventional estimators that simply rely on a single time series of short rate observations. In particular, spurious nonlinearities or biases toward the boundaries are substantially reduced.

While simulations in the existing literature cast doubts on the strong nonlinearity of the drift function at high levels, there is no conclusive evidence of linearity for the drift function. Considering the behavior of the short rate drift at low and medium levels, a linear drift function would suggest a random walk process for the short-term interest rate. As pointed out by Chapman and Pearson (2000), “there are strong theoretical reasons to believe that short rates cannot exhibit the asymptotically explosive behavior implied by a random walk model”. The empirical results in Chapman and Pearson (2000) based on weighted least squares (WLS) estimation also provide quite different point estimates of the drift function, but the results are inconclusive about the linearity or nonlinearity of the drift. Similarly, Li, Pearson and Poteshman (2004) note that the evidence of failing to reject linearity of the drift function should by no means be interpreted as the evidence of accepting linear drift.

To provide further evidence for the dynamics of the short rate process, we apply the proposed method to US data. 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 typical 20 to 30 years of data used in most existing studies. In addition, we use 5 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. We show through simulations that using the panel of yields, including the short rate observations, our proposed estimator of the drift function results in substantial efficiency gain. Using the mean integrated squared-error (MISE) as a yardstick of estimation efficiency, we find that the efficiency of the new drift function estimator is equivalent to that of the nonparametric estimator proposed by 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 levels of interest rate. However, the mean reversion is significantly weaker than that documented by Stanton (1997) and Jiang (1998). More importantly, we further show that the difference in the drift and diffusion function estimates has significant economic effects on the pricing of interest rate derivatives.

The remainder of the paper is structured as follows. In the next section, we propose a new nonparametric estimator of the drift function of the short rate diffusion and derive its properties. Monte Carlo simulations are performed in section III to assess the finite sample properties of the proposed estimator. In section IV, an empirical application of the proposed estimator is undertaken using US interest rate data. Economic implications are also examined using simulated prices of bonds and interest rate derivatives. Section V concludes. Proofs of all propositions are collected in the Appendix.

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

We consider the following diffusion process for the short rate \( r_t^{(1)} \):

\[
dr_t^{(1)} = \mu_1(r_t^{(1)})dt + \sigma_1(r_t^{(1)})d w_t^{(1)}, \quad \text{(The model)}
\]

where \( w_t^{(1)} \) is a standard Brownian motion, \( \mu_1(.) \) and \( \sigma^2_1(.) \) 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:

\[
dr_t^{(j)} = \mu_j(r_t^{(j)})dt + \sigma_j(r_t^{(j)})dw_t^{(j)}, \quad \text{(The auxiliary models)}
\]

where \( \mu_j(.) \) and \( \sigma^2_j(.) \) 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 (1) as “The model” since it is the model of interest to us and the one we intend to estimate, and the models in (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 three-month Treasury bill, then the auxiliary rates could be the yields with maturities longer than three 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 follow 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.

Nonparametric modeling of the short rate diffusion has generated a great deal of interest in recent years because it imposes no restrictions on the functional forms of the drift and diffusion functions. Various nonparametric estimators of the drift and diffusion functions have been proposed in the finance literature, see, e.g., Aït-Sahalia (1996a), Jiang and Knight (1997), Stanton (1997), and Bandi and Phillips (2003). 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 (1) based on the first-order approximation of the discretized process:

\[
\tilde{\mu}_1(r) = \frac{1}{\delta} \sum_{t=0}^{n-1} \left( r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)} \right) K_h \left( r_{t\delta}^{(1)} - r \right) \frac{1}{\sum_{t=0}^{n-1} K_h \left( r_{t\delta}^{(1)} - r \right)},
\]

(3)

\[
\tilde{\sigma}_1^2(r) = \frac{1}{\delta} \sum_{t=0}^{n-1} \left( r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)} \right)^2 K_h \left( r_{t\delta}^{(1)} - r \right) \frac{1}{\sum_{t=0}^{n-1} K_h \left( r_{t\delta}^{(1)} - r \right)}
\]

(4)

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

**Proposition 1:** Suppose that \( h \to 0, Th \to \infty, \) and \( \delta \to 0, \) given the regularity conditions for the short rate process (1) and the kernel function \( K(u) \) (see Appendix), we have

\[
E[\tilde{\mu}_1(r) - \mu_1(r)] = \frac{h^2}{2} m(K)[\mu''_1(r)] + 2\mu'_1(r) p'_1(r) + \frac{q(r)}{p_1(r)} \delta + o(h^2)
\]

(5)

\[
Var[\tilde{\mu}_1(r)] = \frac{\sigma_1^2(r) R(K)}{Thp_1(r)} + o((Th)^{-1})
\]

(6)
where \( m(K) = \int z^2 K(z) dz \), \( R(K) = \int K^2(z) dz \), \( q(r) = \frac{1}{2} \{ \mu'_1(r) \mu_1(r) + \frac{1}{2} \sigma_1^2(r) \mu'_1(r) \} \), and \( p_1(r) \) is the marginal density of the short rate.

Equation (5) shows the finite sample bias of the nonparametric drift function estimator proposed by Stanton (1997). The first term of the bias of the drift function estimator suggests that even when the discretization interval \( \delta \) is sufficiently small, the bias of the nonparametric estimator proposed in Stanton (1997) can be large in regions where the slope and/or curvature of the underlying drift function are noticeable. The slope of the marginal density function of short rate also affects the bias. Thus, the nonparametric estimator of the drift function proposed in Stanton (1997) has the potential of generating spurious nonlinearities as illustrated by the simulation results in Chapman and Pearson (2000). Their simulation is based on the CIR model which by specification has a linear drift function. Chapman and Pearson (2000) point out that a combination of the “truncation” of the observed short rates and a small sample creates artificial patterns of nonlinearity near the boundaries of the support. Chapman and Pearson (2000) also account for the “boundary effect” using the jackknife kernel proposed in Rice (1984) in their simulation 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. 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. 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.

What is more troublesome is the fact that the spurious nonlinearity has a pattern similar to that of the nonparametric estimate of the drift function of the US short rate diffusion process in Stanton (1997) and Jiang (1998). According to these studies, the drift function exhibits a very high level of
mean reversion when the short rate is above the 14% level. Conley, Hansen, Luttmer, and Scheinkman (1997) report similar results with the estimated drift function being nonzero only for rates below 3% or above 11%.

The intuition for the difficulty of estimating the drift function relative to the diffusion function has long been established in the literature. While the diffusion function estimator in (4) requires the sampling interval \( \delta \to 0 \) for convergence, the drift function estimator in (3) also requires the sampling period \( T \to \infty \) 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. The intuition is also confirmed in Aït-Sahalia (1996a) using the Geometric Brownian motion process as an example. Note that for the diffusion process in (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 reflects more of the properties of the diffusion term than those of the drift term, which suggests the possibility of identifying the diffusion term from high-frequency observations. For the same reason, the drift term cannot be estimated precisely based on the local-time dynamics of such sampling paths without further constraints. Not surprisingly, approximations of the drift function from high frequency data can be very non-robust and the estimates can be very sensitive to the sampling path.

A well-known property of the short rate is its high persistence over time.\(^1\) 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, Pearson and Poteshman (2004). This is evident in the plot of the daily 3-month T-bill yields in

\(^1\)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.
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 sampling observations with a given sampling frequency or equivalently a long sampling period. The hope thus rests in extending the sampling 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 by Stanton (1997). 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 offers important incremental information for the drift function identification and estimation. 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 $E(Y|X = x) = \mu(x)$ where $Y_i = \mu(X_i) + \epsilon_i$ with $\epsilon_i$ being the disturbance. Hjort and Glad (1995) propose a semiparametric estimator which 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 resembles the Nadaraya-Watson estimator otherwise. The advantage of the two-step procedure is that

---

2For instance, the yields are never below 2% for the entire period of 1960s to 1990s. The high interest rate observations only occur during the late 1970s and early 1980s with relatively fewer observations of large daily interest rate changes. Extending the sampling period, however, beyond this period back to the 1950s or to early 2000s offers a significant number of interest rates observations below 2% with the lowest observation at 0.55%.

3Various studies in the finance literature have used extraneous data to improve the efficiency of term structure estimation. For example, Abaffy et al (2003) evaluate the position of eleven 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 is resulted 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 US short rate observations with interest rate data from other western countries (UK and Germany in Ang and Bekaert (2002), and France, UK, Italy, Germany, and Japan in Lee and Li (2005)) when estimating the US term structure of interest rates.
when the parametric pilot is a good approximation of the conditional mean, then the “correction factor” will be a nice smooth or even a 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. This is because the nonparametric estimation in the second step is likely to benefit 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. The choice of bandwidth becomes even more critical when the data exhibits high dependence or when there are not enough observations over a certain range of the support. 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.

In this paper, instead of using a parametric pilot we use a nonparametric pilot in the first step. This nonparametric function is estimated using the pooled data from $J$ different interest rate diffusion processes, including the short rate diffusion process. A clear advantage using the nonparametric pilot is that it eliminates the impact of specific parametric functions. Moreover, while the yields of different maturities may be determined by different economic factors, they tend to be correlated with systematic co-movements. Therefore, the pooled estimator of the drift function is likely to be similar in functional shape to the drift function of the short rate process.

While our estimator applies to both the drift and diffusion functions as illustrated in the empirical application, we focus on the drift function in both the theoretical development and the simulations. This is motivated by the difficulty of identifying and estimating the nonparametric drift function as pointed out earlier. Formally, we define $\mu_p(r) = \sum_{j=1}^{J} \mu_j(r) \omega_j(r)$ where $\omega_j(r) = \frac{p_j(r)}{\sum_{j=1}^{J} p_j(r)}$ and $p_j(r)$ is the density function of the $j^{th}$ interest rate process, with $j = 1, \ldots, J$. That is, $\mu_p(r)$ is the function being estimated by pooling the data. Note that $\mu_p(r)$ is identical to the short rate drift $\mu_1(r)$ if the individual density functions are the same.
Our proposed estimator of the drift function is based on the following identity

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

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

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

since \(E \left[ \frac{(r_{(t+1)\delta} - r_{t\delta})}{\delta \mu_p(r_{t\delta})} | r_{t\delta} \right] = c(r)\). 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, i.e.

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

Consistency of the pooled estimator \(\hat{\mu}_p(r)\) is established in the Appendix under regularity conditions.

With a consistent estimator of \(\mu_p(r)\) in (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

$$
\hat{\mu}_1(r) = \hat{\mu}_p(r) \left[ \frac{1}{\delta} \sum_{t=0}^{n-1} \left[ \frac{(r_{(t+1)\delta} - r_{t\delta})}{\mu_p(r_{t\delta})} \right] K_h(r_{t\delta} - r) \right]
$$  \(10\)

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 \(\hat{\mu}_p(r)\) which, in this scenario, would be the “optimal” estimator. In this case, the proposed estimator is substantially more efficient than the conventional estimators of the drift function using 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. As a result, the proposed estimator will have a smaller bias relative to the estimator proposed by Stanton (1997). The formal results for the bias and variance of the proposed drift function estimator are given in the following proposition.

**Proposition 2**: Suppose that $h_p \to 0$, $Th_p \to \infty$, and that the “auxiliary processes” in (2) also satisfy the regularity conditions listed in the appendix. Then given the assumptions of Proposition 1, we have:

\[
E[\hat{\mu}_1(r) - \mu_1(r)] = \frac{h^2}{2} m(K)[c''(r) + 2c'(r)p_1'(r)]\mu_p(r) + \frac{l(r)}{p_1(r)}\mu_p(r) \delta + o(h^2) \tag{11}
\]

\[
Var[\hat{\mu}_1(r)] = \frac{\sigma^2_1(r)R(K)}{(Th)p_1(r)} + O((JTh_p)^{-1} + h^2h_p^2 + h^2p\delta) + o((Th)^{-1}) \tag{12}
\]

where $l(r) = \frac{1}{2}\{c'(r)c(r) + \frac{1}{2}\sigma^2_1(r)c''(r)\}$.

Compared to the results in Proposition 1, it is clear that the biases of the two estimators can substantially differ. The bias of the proposed estimator depends on the slope and curvature of the correction factor $c(r)$ while the bias of the Stanton estimator is a function of the slope and curvature of the true 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 $c(r) = \frac{\mu_1(r)}{\mu_p(r)}$ is a straight line, and hence $c'(r) = c''(r) = 0$. As a result, the finite sample bias of the proposed estimator is reduced to a negligible order. Similarly, if $\mu_p(r)$ is not too different from $\mu_1(r)$ in functional shape, the correction factor will oscillate around a constant, thus having less curvature than $\mu_1(r)$. As a result, the proposed estimator will be less biased than the estimator proposed by Stanton (1997). The variance of the proposed estimator tends to be higher than that of the estimator proposed by Stanton (1997) (see Proposition 1) 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 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 $nh^{5} \to 0$, $\sqrt{nh}\delta \to 0$, $Jnh^{5} \to 0$, and $\sqrt{Jnh}\delta \to 0$, then the proposed drift function estimator is asymptotically normally distributed:

\[
\sqrt{Th}(\hat{\mu}_1(r) - \mu_1(r)) + o_p(1) \to N(0, \Theta(r)) \quad \text{where} \quad \Theta(r) = \frac{\sigma^2_1(r)R(K)}{\tilde{p}_1(r)} \tag{13}
\]
Proposition 3 shows the asymptotic equivalency of $\hat{\mu}_1(r)$ and $\tilde{\mu}_1(r)$ under slightly stronger conditions than those in Propositions 1 and 2. Both estimators 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 approach proposed in this paper is particularly relevant for the estimation of the short rate diffusion as yields with different maturities are available along the yield curve. Furthermore, while the yields of different maturities may be determined by different economic factors or market forces, empirical evidence suggests that the yield curve tends to behave systematically. In particular, there are systematic co-movements among yields with different maturities.\textsuperscript{4} This is further supported by empirical evidence of the principal component analysis in Litterman and Scheinkman (1991). Litterman and Scheinkman (1991) report that there are three major factors driving the dynamics of the US yield curve, namely the level of short rate, the slope of the yield curve, and the curvature of the yield curve, a result for which we also find support. In our sample, the short rate accounts for nearly 80% variation of the whole 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.

\textsuperscript{4}For instance, when the short rate is specified as in (16) under the risk-neutral measure, the bond price of a zero coupon bond with maturity $\tau$ will evolve according to

$$
\frac{dP(r_t, \tau)}{P(r_t, \tau)} = \xi(r_t, \tau)dt + \nu(r_t, \tau)d\tilde{w}_t
$$

where, by Itô's lemma, we have

$$
\xi(r_t, \tau)P = \frac{1}{2} \sigma^2(r_t) P_{rr} + (\mu_3(r_t) - \lambda(r_t)) P_r + P_t
$$

$$
\nu(r_t, \tau) = \sigma(r_t) P_r
$$

That is, the drift and diffusion functions of bond yields with different maturities share similar structure. In addition, when observed bond yields contain market microstructure noises with measurement errors, the processes of these yields are no longer perfectly correlated and combining yields of different maturities offers additional information for the estimation of the short rate process.
III. Simulations

In this section, we perform simulations to assess the finite sample performance of the proposed drift function estimator. The simulations are designed with two goals. The first goal is to revisit the findings of Chapman and Pearson (2000). To evaluate the finite sample properties of the nonparametric estimator proposed by Stanton (1997), Chapman and Pearson simulate interest rate data from a CIR process and find that the drift function estimate is linear only in the center of the support, deviating severely from the true (linear) drift on both boundaries, in particular for interest rates above 14%. The second goal of the simulation is to investigate whether the use of additional interest rate data from similar diffusion processes can attenuate such biases.

We focus on both the Vasicek and CIR models in our simulation. Without loss of generality, we only consider the case of $J = 2$, i.e., there is only one additional auxiliary diffusion process. The Vasicek process for the short rate process as well as the auxiliary model are specified as follows:

\[
\begin{align*}
\frac{dr_t^{(1)}}{dt} &= 0.261(0.0717 - r_t^{(1)})dt + 0.02237dw_t^{(1)} \\
\frac{dr_t^{(2)}}{dt} &= \kappa \cdot 0.261(0.0717 - r_t^{(2)})dt + \sqrt{\gamma} \cdot 0.02237dw_t^{(2)} \\
dw_t^{(1)}dw_t^{(2)} &= \rho dt,
\end{align*}
\]

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 process for the short rate as well as the auxiliary model are specified as follows:

\[
\begin{align*}
\frac{dr_t^{(1)}}{dt} &= 0.8537(0.08571 - r_t^{(1)})dt + 0.1566\sqrt{r_t^{(1)}}dw_t^{(1)} \\
\frac{dr_t^{(2)}}{dt} &= \kappa \cdot 0.8537(0.08571 - r_t^{(2)})dt + \sqrt{\gamma} \cdot 0.1566\sqrt{r_t^{(2)}}dw_t^{(2)} \\
dw_t^{(1)}dw_t^{(2)} &= \rho dt,
\end{align*}
\]

where the parameter values of the short rate model are set equal to those in Chapman and Pearson (2000).

\footnote{Chapman and Pearson (2000) only used the CIR square-root diffusion process in their simulations.}
The parameters $\kappa$, $\gamma$, $\rho$ in the auxiliary model are set at different values in our simulations in order to investigate the effect 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): $\gamma = 1$, $\rho = 0$, $\kappa = 0.75, 1.25, \text{ and } 1.5$. In this case, the short rate model and the auxiliary model have the same long-run mean, but different speed of mean reversion. 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 of the short rate drift function.

- Case III (diffusion effect): $\kappa = 1$, $\rho = 0$, $\gamma = 0.75, 1.25, \text{ and } 1.5$. In this case, the short rate process and the auxiliary 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 $\rho = 0.2, 0.4, \text{ and } 0.6$. This case allows us to examine the impact of the level of the correlation between the short rate model and the auxiliary model on the performance of the proposed drift function estimator. As mentioned earlier, the yields of different maturities along the yield curve tend to be highly correlated with systematic co-movements.

- 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 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. Observations in the first year are discarded to eliminate the start-up effects, which results in a total number of 7,500 daily observations over 30 years. For the Vasicek processes, starting values are drawn from the marginal normal distributions and subsequent values are obtained from the transitional bivariate normal density. For the CIR processes, the starting values are drawn from the marginal Gamma distributions and the subsequent values are simulated using the Milshtein discretization scheme. The Milshtein discretization scheme has faster convergence rate than the Euler discretization scheme with almost sure convergence to the continuous sampling path (see Talay, 1996). When \( \rho \neq 0 \), the short rate model and the auxiliary model are simulated jointly. With the simulated sampling paths for the short rate model and the auxiliary model, the Stanton estimator and the proposed estimator for the drift function of the short rate process are then implemented over the support of \( r = [0, 20\%] \).

Note that the Stanton estimator 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).

The drift function estimates of the Vasicek and CIR 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 and illustrating the effect of the bandwidth choice, the naive bandwidth is used for both the Stanton estimator and the proposed estimator in Figure 1. The results for the Stanton estimator are consistent with the findings in Chapman and

\[ \sum_{t=0}^{n-1} \left( \hat{\mu}_1^{(-b)}(r_{1\delta}^{(1)}) - \left[ \frac{r_{(t+1)\delta}^{(1)} - r_{t\delta}^{(1)}}{\delta} \right] \right)^2 \]

where \( \hat{\mu}_1^{(-b)}(r_{1\delta}^{(1)}) \) 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 \( \hat{\mu}_p(x) \) and then choosing \( h \) to minimize the cross-validation function of \( \hat{\mu}_1(x) \).
Pearson (2000). Namely, the estimate displays significant nonlinearity at 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 \( [r_{min}^{(1)}, r_{max}^{(1)}] \), and the limited sampling period. While the proposed estimator has less nonlinearity towards the boundaries, it remains a poor estimate of the linear drift function with the use of the naive bandwidth.

Figure 2 plots the average drift function estimates of the Stanton estimator and the proposed estimator for the Vasicek and CIR models using the ISE-minimizing bandwidth. Two observations are noted. First, there is a substantial improvement for both the Stanton estimator and the proposed estimator compared to those in Figure 1 where the naive bandwidth is used. The difference 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), our proposed drift function estimator exhibits much less “spurious” nonlinearities 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 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.

Tables 1 and 2 report the mean integrated absolute bias (MIAB) and the root mean integrated squared error (RMISE) of alternative drift function estimates for the Vasicek and CIR processes, respectively. Panel A reports the MIAB and RMISE of the Stanton estimator, and Panel B reports those of the proposed estimator for the benchmark case or case I. Again, for both estimators, it is clear that the use of optimal bandwidth is important. There is a significant reduction in bias and a gain in effi-

---

7To limit computational burden, the smoothing parameters for the pooled and the proposed estimator \((h_p, h)\) are chosen by sequentially minimizing the integrated squared errors of \(\hat{\mu}_p(r)\) and \(\hat{\mu}_1(r)\). Note that an iterated procedure to simultaneously determine the optimal choice of \((h_p, h)\) would perform at least as well as the suboptimal sequential procedure.
ciency when the optimal bandwidth is used. The results also confirm that the proposed estimator has substantially lower bias relative to the Stanton estimator for both the Vasicek and CIR models. When the optimal bandwidth is used for both estimators, the bias of the proposed estimator is only slightly over 20% of the bias of the Stanton estimator for both the CIR and the Vasicek models according to the MIAB. The RMISE, as a measure of overall efficiency, is also substantially reduced as a result of bias correction.

The MIAB and RMISE of the proposed estimator for the remaining cases are also reported in Tables 1 and 2. As in the benchmark case, each simulation involves 5,000 replications. Panel C of Tables 1 and 2 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 $\mu_1(r)$. For the CIR model with $k = 1.5$, although the bias of the proposed estimator has more than doubled relative to the benchmark case, it remains considerably smaller than that of the Stanton estimator. Note that in this case the correction factor oscillates around a constant, thus having less curvature than the underlying true drift of the short rate.

The use of extraneous information leads to bias reduction in the drift function estimation. Panel D of Tables 1 and 2 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 the performance of the proposed estimator in general improves as the volatility of the auxiliary model increases. This result is expected because an increased 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 Tables 1 and 2 reports the performance of the proposed estimator for different levels of correlation between the Brownian motions $\omega_l^{(1)}$ and $\omega_l^{(2)}$ (case IV). Clearly, as $\rho$ 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 $\rho$ increases. In general, as $\rho$ increases the auxiliary process offers less extraneous information for the estimation of the drift function.
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 a CIR process while the model to be estimated follows a the Vasicek process. The Vasicek model is simulated using the parameter values in (14) and the auxiliary CIR process is simulated using the parameters in (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 when the auxiliary model is identical to the short rate model. Still, although the MIAB is about twice that of the benchmark case, it is less than half of the bias of the estimator proposed by Stanton (1997). The root-mean integrated squared error (RMISE) of the proposed estimator is also considerably lower than that of the Stanton 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 toward the boundaries and thus leads to bias reduction. The simulation results also show that the proposed estimator performs well relative to the Stanton estimator even when the drift and diffusion functions of the short rate and auxiliary models are not identical. 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 consists of 12,704 daily observations of the US 3-month T-bill rates from January 1954 to November 2004. This dataset, 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 also 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 3. 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. Both skewness and kurtosis indicate that interest rates are non-normally distributed. 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 3. Although the autocorrelation coefficients of the interest rate are very high, they are all significantly different from one. Moreover, those of the day-to-day changes are generally small and are 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 augmented Dickey-Fuller stationarity test statistic of the 3-month T-bill rates has a value of -2.86 which is compared to the 10% critical value of -2.57. That is, the null hypothesis of nonstationarity is rejected at the 10% significance level for the 3-month T-bill time series. Since the test is known to have low power, even a slight rejection means that stationarity of the series is very likely. Similar results are obtained for the time series of yields with longer maturities.

Figure 3 plots the time series of the daily 3-month Treasury yields and the day-to-day changes in panels 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 70’s and early 80’s, indicating different behavior of interest rates at different interest rate levels.
Table 4 also reports the correlation matrix 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 4 is similar to those in existing studies. Namely, there are mainly three factors driving 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.

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. Panel 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 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 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 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 practically relevant. 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 where there is typically less information. Note that as seen from the summary statistics in Table 3, the additional yields with maturities longer than 3 months, which we use to implement the proposed estimator, 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.

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 estimator to have the same efficiency as the proposed estimator? To answer this question, we use the mean integrated squared error (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 estimator. Second, with the estimated diffusion processes and the correlation matrix of the yields as reported in Table 4, 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 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 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 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, by incorporating additional information from a panel of yields, has the same efficiency as the Stanton estimator implemented with approximately 145 years of daily short rate observations.

We also implement the proposed estimator for the diffusion function using yields of all maturities. The new diffusion function estimate is plotted in Panel B of Figure 4, together with the Stanton 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 the higher efficiency of the diffusion function estimation. Both the Stanton 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 estimate based on the Stanton (1997) estimator,
however, the new diffusion function estimate exhibits a less dramatic increase in volatility as the short rate increases. The main difference is that at high level of the short rate, the Stanton estimate is clearly outside of the 90% confidence band of the new diffusion function estimate. When the short rate is above 12%, the diffusion estimate based on the proposed method is well below the Stanton diffusion function estimate. This is consistent with the weaker mean reversion of the new drift function estimate (see Panel 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 (1) is given by

\[
dr_t = (\mu(r_t) - \lambda(r_t))dt + \sigma(r_t)d\tilde{w}_t
\]  

(16)

where \( \lambda(r_t) = \lambda_0(r_t)\sigma(r_t) \) is the market price of interest rate risk, and \( \tilde{w}_t \) is a standard Brownian motion under the equivalent martingale measure \( Q \). It is clear from (16) that the drift function of the interest rate enters directly into its risk-neutral counterpart.

The market price of interest rate risk can be nonparametrically estimated 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. Suppose \( Y(r_t, \tau_i) (i = 1, 2) \) represents the yields of zero-coupon bonds at \( t \) with maturity \( \tau_i = T_i - t \), and

\[
dY(r_t, \tau_i) = \alpha(r_t, \tau_i)dt + \kappa(r_t, \tau_i)d\tilde{w}_t
\]  

(17)
Following Itô’s lemma and using (16), an estimator of time-stationary $\lambda_0(r_t)$ can be derived as

$$\hat{\lambda}_0(r_t) = \frac{Y_d(r_t, \tau_1, \tau_2) + \frac{1}{2}(\tau_1^2 \kappa^2(r_t, \tau_1) - \tau_2^2 \kappa^2(r_t, \tau_2)) + \tau_2 \alpha(r_t, \tau_2) - \tau_1 \alpha(r_t, \tau_1)}{\tau_2 \kappa(r_t, \tau_2) - \tau_1 \kappa(r_t, \tau_1)}$$

(18)

where $\tau_i = T_i - t$ ($i = 1, 2$), $Y_d(r_t, \tau_1, \tau_2) = Y(r_t, \tau_1) - Y(r_t, \tau_2)$ is the yield spread between maturities $\tau_1$ and $\tau_2$. 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 $\lambda_0(r_t)$ for both the Stanton 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 processes. Given the estimated 3-month and 10-year yield processes, the estimate of $\lambda_0(r_t)$ is obtained from (18).

The prices of interest rate derivative securities can be computed based on the simulation of the risk-neutral process in (16). Again, simulations of the sample path are based on the Milstein scheme. In financial applications of the Monte Carlo simulation methods, a number of variance reduction methods have been proposed, e.g. the control variate approach, the antithetic variate method, the moment matching method, the importance sampling method, the conditional Monte Carlo methods, and quasi-random Monte Carlo methods (see, 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 bond prices as well as the prices of 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.

Given the bond yield process in (17), the bond price process is

$$\frac{dP(r_t, \tau_i)}{P(r_t, \tau_i)} = \xi(r_t, \tau_i)dt + \nu(r_t, \tau_i)d\tilde{w}_t$$

with $\xi(r_t, \tau_i) = -\tau_i \alpha(r_t, \tau_i) + \frac{1}{2}\tau_i^2 \kappa(r_t, \tau_i)$ and $\nu(r_t, \tau_i) = -\tau_i \kappa(r_t, \tau_i)$. Under standard assumptions in the literature (see e.g. Ingersoll (1987)), absence of arbitrage in the economy implies that

$$\xi(r_t, \tau_i) = r_t + \lambda_0(r_t) \nu(r_t, \tau_i)$$

for a non-dividend paying asset. With the process for yield yields with maturities $\tau_1$ and $\tau_2$, the market price of interest rate risk $\lambda_0(r_t)$ can be estimated using (18).
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 estimate and the new estimate of the drift and diffusion functions. The price of a zero-coupon bond with face value \( P(r_t, t, T) = 1 \) is given by

\[
P(r_t, t, T) = E_t^Q \{ \exp \{ - \int_t^T r_u du \} \}
\]

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 starting values of interest 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. At the 10% interest rate level, there is no clear difference between the yield curve based on the Stanton (1997) estimator and that based on the new estimator. At both the high and low levels of interest rates (15% and 5%), however, the yield curves based on the Stanton (1997) estimator are outside of the 90% confidence bands. As expected, due to the stronger mean reversion of the Stanton drift function estimate, the yield curves based on the Stanton estimator are significantly lower than those based on the new estimator when the initial interest rate is high at 15% but higher than those based on the new estimator when the initial interest rate is low at 5%.

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:

\[
100 \times \max[(Y(t - \Delta t; t) - Y_S)\Delta t; 0]
\]

where \( t = t_1, t_2, \ldots, t_n \) are the payment dates (reset of \( \Delta t \) 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 5 for different strike prices (in basis points), tenors (in years), and annualized spot rates. As shown in Table 5, the cap prices based on the proposed estimator are in general lower than those based on the Stanton estimator, except when the interest rate is very high ($r_0 = 15\%$). The above observations can be explained by the fact that the proposed estimator has an overall lower estimate of the diffusion function compared to the Stanton estimator. On the other hand, at high interest rate levels, the proposed estimator of the drift function exhibits a much weaker mean reversion than the Stanton estimator of the drift function. 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 current interest rate is high but a positive effect when the current interest rate is low. 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) and a current interest rate level of 5\%, the differences between the prices based on the proposed estimates and those based on the Stanton 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 contracts with strike price equal to the current spot rate (the at-the-money contracts), the difference is in the range of $0.38\text{ to }1.04$. 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 $0.37\text{ to }1.07$.

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 estimators are designed to reduce the finite sample bias of the nonparametric estimator proposed by Stanton (1997). The biases of the Stanton estimator are due to the persistent dependence of interest rates and a limited sampling period, and can produce spurious nonlinearities in the drift function as shown in Chapman and Pearson (2000). The 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 estimate the US short rate diffusion process using a panel of six US Treasury yields with
maturities ranging from 3 months to 10 years. We find that using observations of the panel of six Treasury yields over the past 42 years, the proposed drift function estimator achieves the same efficiency as the estimator proposed by Stanton (1997) would using 145 years of daily short rate observations. Our empirical results on the estimates of the short rate process 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 speed of the 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 when pricing bonds and other interest rate derivative securities.
References


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


Appendix: Proofs of Propositions

For simplicity of notations, 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 assumption A1 in Nicolau (2003) or conditions 1 and 2 in Bandi and Phillips (2003), i.e., each process has a unique strong solution.

(2) Each of the \( J \) interest processes satisfies assumptions A1 and A2 in Nicolau (2003), i.e. each process has an invariant density \( p_j(r) \), \( j = 1, \ldots, J \).

(3) Under assumption A4 in Nicolau (2003), the short rate diffusion process is \( \rho \)-mixing.

(4) The Kernel function \( K(z) \) is bounded, 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) \( |z| K(|z|) \rightarrow 0 \) as \( |z| \rightarrow \infty \), and (v) \( \int K^2(z) \, dz \leq \infty \).

**Proof of Proposition 1**: Based on the drift function estimator proposed by Stanton (1997) in (3), we have

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

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

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

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

Hence,

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

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

\[
= \frac{1}{np(r)} E \left( \sum_{t=0}^{n-1} K_h(r_t - r) \left[ \mu(r_t) + \frac{1}{2} \left( \mu'(r_t) \mu(r_t) + \sigma^2(r_t) \mu''(r_t) \right) \right] \right)
\]

\[
= \frac{1}{np(r)} E \left( \sum_{t=0}^{n-1} K_h(r_t - r) \left[ \mu(r_t) + q(r_t)\delta - \mu(r) \right] \right)
\]
Under the assumption that the short rate process is
have

\[
E\left(K(z)[\mu(r + hz) + q(x + hz)\delta - \mu(r)]p(r + hz)dz\right)
= \frac{1}{p(r)} \int K(z)[\mu(r + hz) + q(x + hz)\delta - \mu(r)]p(r + hz)dz
= \frac{h^2}{2} \left[ \mu''(r) + 2\mu'(r) \frac{p'(r)}{p(r)} \right] m(K) + \frac{q(r)}{p(r)} \delta + o(h^2 + \delta).
\]

where \(m(K)\) is as defined in proposition 1.

Now turning to the variance of the drift function estimator in (3), we have:

\[
Var(\hat{\mu}(r)) = \frac{1}{p^2(r)} Var\{\frac{1}{n} \sum_{t=0}^{n-1} K_h(r_t - r) \left[ \frac{(r_{t+1} - r_t)}{h} - \mu(r) \right]\}
= \frac{1}{p^2(r)} \frac{1}{Th} Var\{\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} v(r_t, r_{t+1})\}
\]

where \(v(r_t, r_{t+1}) = \sqrt{\frac{\delta}{h}} K(r_t - r) [\frac{(r_{t+1} - r_t)}{\delta} - \mu(r)]\). Introducing the notation \(\hat{\nu}(r_t, r_{t+1}) = v(r_t, r_{t+1}) - E(v(r_t, r_{t+1}))\), we have the mean and variance given by \(E(\hat{\nu}(r_t, r_{t+1})) = 0\) and \(E(\hat{\nu}^2(r_t, r_{t+1})) = E(v^2(r_t, r_{t+1})) - [E(v(r_{t+1}, r_{t+1}))]^2\), where

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

\[
E(v^2(r_t, r_{t+1})) = \frac{\delta}{h} E\left(\frac{K^2(r_t - r)}{h} E\left(\left[ \frac{(r_{t+1} - r_t)}{\delta} - \mu(r) \right]^2\right)_{r_t}\right)
= \int K^2(z) \left[ \sigma^2(r + hz) + O(h^2 + \delta) \right] dz
= \sigma^2(r) R(K)p(r) + O(h^2 + \delta) < \infty.
\]

We further derive the autocovariance of \(v(r_t, r_{t+1})\). For \(t = 0\),

\[
E(v(r_0; r_1)v(r_1; r_2)) = \frac{\delta}{h} E\left(K\left(\frac{r_0 - r}{h}\right)K\left(\frac{r_1 - r}{h}\right)\left\{ \frac{r_1 - r_0}{\delta} - \mu(r) \right\}\left\{ \frac{r_2 - r_1}{\delta} - \mu(r) \right\}\right)
\]

By the law of iterated expectation,

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

Under the assumption that the short rate process is \(\rho\)-mixing (see assumption 4 in Nicolau (2003), we have \(t \geq 1\)

\[
|E(v(r_0; r_1)v(r_1; r_{t+1}))| \leq |E(v(r_0; r_1)v(r_1; r_2))| \to 0. \quad (20)
\]

32
Finally using the results in lemma 9 of Nicolau (2003), it follows that

$$Var(\hat{\mu}(r)) \simeq \frac{1}{(Thp(r)^2)} (E[\hat{\nu}^2(r_t, r_{t+1})] + o(1)) = \frac{\sigma^2(r)R(K) + o(1)}{(Th)p(r)}.$$ 

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 finite sample properties of the pooled estimator of the drift function $\hat{\mu}_p(r)$ in (9), while the proof of Proposition 3 depends on the weak consistency of the pooled estimator of the drift function. Under the regularity conditions on the model in (1) and the auxiliary models in (2), the finite sample properties (i.e., bias and variance) of the pooled estimator of the drift function are essentially the same as the Stanton estimator. The following lemma thus only proves the weak consistency of the pooled estimator of the drift function $\hat{\mu}_p(r)$.

**Lemma:** The pooled estimator $\hat{\mu}_p(r)$ defined in (9) is weakly consistent, that is $\hat{\mu}_p(r) - \mu_p(r) = \epsilon_N \rightarrow 0$ in probability as $N \rightarrow \infty$.

**Proof of Lemma:** Based on the pooled estimator of the drift function $\hat{\mu}_p(r)$ in (9), we have

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

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

Using the fact that $\sum_{t=0}^{n-1} \left( \frac{r_{j(t+1)}^{(j)} - r_t^{(j)}}{\delta} \right) K_{hp}\left( r_t^{(j)} - r \right) = \hat{\mu}_j(r)\hat{p}_j(r)$, we can rewrite $\hat{\mu}_p(r)$ as

$$\hat{\mu}_p(r) = \frac{\sum_{j=1}^{J} \hat{\mu}_j(r)\hat{p}_j(r)}{\sum_{j=1}^{J} \hat{p}_j(r)}$$

Applying the Slutsky’s theorem, we have

$$\text{plim} \hat{\mu}_p(r) = \frac{\sum_{j=1}^{J} \text{plim} \hat{\mu}_j(r) \text{plim} \hat{p}_j(r)}{\sum_{j=1}^{J} \text{plim} \hat{p}_j(r)}$$

$$= \frac{\sum_{j=1}^{J} \mu_j(r)p_j(r)}{\sum_{j=0}^{J} p_j(r)} = \mu_p(r)$$

which proves the weak consistency of the pooled estimator of the drift function.
Proof of Proposition 2: Applying Taylor series expansion to $\frac{\hat{\mu}(r)}{\mu_p(r)}$ around $\mu_p(r_1)$, we have the following expansion of the drift function estimator in (10)

$$
\hat{\mu}(r) \approx \frac{1}{n\hat{p}(r)} \sum_{t=0}^{n-1} K_h(r_t - r) \left( r_{t+1} - r_t \right) \left[ \frac{\mu_p(r)}{\mu_p(r_t)} \right] 
+ \frac{1}{n\hat{p}(r)} \sum_{t=0}^{n-1} K_h(r_t - r) \left( r_{t+1} - r_t \right) \left[ \frac{\hat{\mu}_p(r) - \mu_p(r)}{\mu_p(r_t)} - \frac{\hat{\mu}_p(r_t) - \mu_p(r)}{\mu_p(r_t)} \right] 
= \frac{1}{\hat{p}(r)} \left[ A_n + B_n \right]
$$

where $A_n$ and $B_n$ correspond to, respectively, the first and second terms of the previous equation. Thus, we can write

$$
E(\hat{\mu}(r) - \mu(r)) = \frac{1}{p(r)} E(A_n - \mu_p(r)c(r)\hat{p}(r)) + \frac{1}{p(r)} E(B_n) + o_p(1)
$$

$$
E(A_n - \mu_p(r)c(r)\hat{p}(r)) = \frac{\mu_p(r)}{n} E \left( \sum_{t=0}^{n-1} K_h(r_t - r) \left( r_{t+1} - r_t \right) \left[ \frac{\mu_p(r)}{\mu_p(r_t)} - c(r) \right] \right)
= \frac{\mu_p(r)}{n} E \left( \sum_{t=0}^{n-1} K_h(r_t - r) \left[ c(r_t) + l(r_t)\delta - c(r) \right] \right)
= \mu_p(r) \int K(z) \left[ c(r + hz) + l(x + hz)\delta - c(r) \right] p(r + hz) dz
= \mu_p(r) \left( l(r)\delta + \frac{h^2}{2} m(k) \left[ c''(r) p(r) + 2c'(r) p'(r) \right] \right) + o(h^2 + \delta)
$$

It can be seen that $E(B_n) = 0$ after ignoring terms of the orders $h_p^4$, $h_p^2\delta$ and smaller. Hence

$$
E(\hat{\mu}(r) - \mu(r)) = \mu_p(r) \left[ l(r)\delta + \frac{h^2}{2} m(k) \left[ c''(r) + 2c'(r) \frac{p'_1(r)}{p_1(r)} \right] \right] + o(h^2 + \delta)
$$

(27)

where $l(r)$ is as defined in proposition 2.

Now turning to the variance, we have:

$$
Var(\hat{\mu}(r)) = Var(\frac{A_n}{\hat{p}(r)}) + Var(\frac{B_n}{\hat{p}(r)}) + 2Q
$$

where $Q = Cov(\frac{A_n}{\hat{p}(r)}, \frac{B_n}{\hat{p}(r)})$ and

$$
Var \left( \frac{A_n}{\hat{p}(r)} \right) \approx Var \left( \frac{A_n - \hat{p}(r)\mu(r)}{p(r)} \right) 
= \frac{\mu_p^2(r)}{p^2(r)} \left[ \sum_{t=0}^{n-1} K_h(r_t - r) \left( r_{t+1} - r_t \right) \left[ \frac{\mu_p(r)}{\mu_p(r_t)} - c(r) \right] \right]
= \frac{1}{p^2(r)} \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} g(r_t, r_{t+1})
$$
where \( g(r_t, r_{(t+1)}) = \mu_p(r) \sqrt{\frac{\delta}{h}} K(\frac{r_t - r}{h}) \left( \frac{(r_{(t+1)} - r_t)}{\delta \mu_p(r_t)} - c(r) \right) \). Introducing the notation \( \dot{g}(r_t, r_{(t+1)}) = g(r_t, r_{(t+1)}) - E(g(r_t, r_{(t+1)})) \), we have the mean and variance given by \( E(\dot{g}(r_t, r_{(t+1)})) = 0 \) and \( E(\dot{g}^2(r_t, r_{(t+1)})) = E(g^2(r_t, r_{(t+1)})) - [E(g(r_t, r_{(t+1)}))]^2 \), where

\[
E(g(r_t, r_{(t+1)})) = \mu_p(r) \sqrt{\frac{\delta}{h}} E \left\{ \frac{K(r_t - r)}{h} \left( \frac{(r_{(t+1)} - r_t)}{\delta \mu_p(r_t)} - c(r) \right) \right\} 
= \mu_p(r) \sqrt{h \delta} O(h^2 + \delta) \to 0.
\]

\[
E(g^2(r_t, r_{(t+1)})) = \mu_p^2(r) \frac{\delta}{h} E \left[ K^2(\frac{r_t - r}{h}) E \left\{ \frac{(r_{(t+1)} - r_t)}{\delta \mu_p(r_t)} - c(r) \right\}^2 | r_t \right]
= \mu_p^2(r) \int K^2(z) \left[ \sigma^2(r + hz) + O(h^2 + \delta) \right] \frac{G(r + hz)}{\mu^2_p(r + hz)} dz
= \sigma^2(r) R(K) p(r) + O(h^2 + \delta) < \infty.
\]

We further derive the autocovariance function of \( g(r_t \delta, r_{(t+1) \delta}) \). For \( t = 0 \),

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

By the law of iterated expectation, we have

\[
E(g(r_0; r_1)g(r_1; r_2)) = \mu_p^2(r) \frac{\delta}{h} E \left[ K^2(\frac{r_0 - r}{h}) \right] \left\{ (c(r_0) - c(r))^2 + O(\delta) \right\}
= \mu_p^2(r) \delta \int K^2(z) \left[ (c(r) + O(h) - c(r))^2 + O(\delta) \right] [p(r) + O(h)] dz
= \mu_p^2(r) O(h^2 + \delta) \delta \to 0.
\]

Under assumption 4 in Nicolau (2003), the process is \( \rho \)-mixing, hence for \( t \geq 1 \)

\[
|E(g(r_0; r_1)g(r_t; r_{(t+1)}))| \leq |E(g(r_0; r_1)g(r_1; r_2))| \to 0. \tag{28}
\]

Further using the results in lemma 9 of Nicolau (2003), it follows that

\[
Var \left[ \frac{A_n}{p(r)} \right] \simeq \frac{1}{(Th)} p(r)^2 \left( E[g^2(r_t, r_{(t+1)})] + o(1) \right) = \frac{\sigma^2(r) R(K) + o(1)}{(Th)p(r)}.
\]

Next turning to the covariance term \( Q \simeq Cov \left( \frac{A_n}{p(r)}, \frac{B_n}{p(r)} \right) \), we have

\[
Cov(A_n, B_n) = \frac{1}{n^2} \sum_{t=0}^{n-1} E(g_1(r_t)g_2(r_t)) + \frac{2}{n^2} \sum_{t < t'} E(g_1(r_t)g_2(r_{t'})) - E[A_n]E[B_n]
\]

where \( g_1(r_t) = K_h(r_t - r) \frac{(r_{(t+1)} - r_t)}{\delta} \frac{\mu_p(r)}{\mu_p(r_t)} \), and
\[
g_2(r_t) = K_h(r_t - r) \frac{(r_{(t+1)} - r_t)}{\delta} \left[ \frac{\bar{\mu}(r) - \mu_p(r)}{\mu_p(r_t)} - \frac{\bar{\mu}(r)}{\mu_p(r_t)} \right] \frac{\mu_p(r)}{\mu_p(r_t)}.
\]

We focus first on the second term of the covariance expression. By the law of iterated expectation, it follows that for \( t = 0 \) and \( t' = 1 \), \( E(g_1(r_0)g_2(r_\delta)) = O(h^2 h_p^2 + h^2 \delta) \), which is directly from the fact
that $E(\hat{\mu}_p(r) - \mu_p(r)) = O(h_p^2 + \delta)$. Again by assumption 4 in Nicolau (2003), the process is $\rho$-mixing, thus, for $t' > 1$

$$|E(g_1(r_0)g_2(r_{t'}\delta))| \leq |E(g_1(r_0)g_2(r_1))| = O(h^2 h_p^2 + h^2 \delta) \to 0. \quad (29)$$

It follows that $\frac{2}{nh} \sum_{t<t'} E(g_1(r_t)g_2(r_{t}\delta))$ is negligible. Furthermore, the first term in the covariance expression, $\frac{1}{nh} E(g_1(r_t)g_2(r_t))$, is of the order $o((Th)^{-1})$ while the last term $E(A_n)E(B_n) = O(1)O(h^4_p + h^2 \delta).

Finally, using the finite sample properties of the pooled estimator of the drift function, i.e., $E(\hat{\mu}_p(r) - \mu_p(r)) = O(h_p^2 + \delta)$ and $h_p^4 \propto (Nh_p)^{-1}$, similar calculations lead to the following result:

$$\text{Var}(B_n) = O((Nh_p)^{-1}) \quad (30)$$

where $N = JT$. This concludes the proof of Proposition 2.

**Proof of proposition 3:** Let $\hat{B}_p(r) = \hat{\mu}_p(r) - \mu_p(r)$. Note that by weak consistency of the pooled estimator of the drift function from the Lemma, $\hat{B}_p(r) = \epsilon_N \to 0$ in the limit. Thus we have

$$\hat{\mu}(r) - \mu(r) = \frac{(\hat{\mu}(r) - \mu(r))}{\hat{p}(r)} \hat{p}(r)$$

$$= \frac{\mu_p(r)}{nh\hat{p}(r)} \sum_{t=0}^{n-1} K(r_{t+1}\delta - r_{t}\delta) - c(r)\}$$

$$+ \frac{1}{nh\hat{p}(r)} \sum_{t=0}^{n-1} K(r_t - r_{t+1}\delta - c(r))\}$$

$$- \frac{1}{nh\hat{p}(r)} \sum_{t=0}^{n-1} K(r_t - r_{t+1}\delta - c(r))\}$$

Hence

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

where $g(r_t; r_{t+1})$ is defined as in the proof of Proposition 2, and

$$C_n = \frac{1}{\sigma(r) \sqrt{R(K)\hat{p}(r)}} \frac{1}{nh} \sum_{t=0}^{n-1} K(r_t - r_{t+1}\delta - c(r))\}$$

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

36
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 
\[
\frac{1}{\sigma(r)} \sum_{t=0}^{n-1} \frac{g(r_t; r_{t+1})}{\sqrt{R(K)p(r)}}
\]
and 
\[
\frac{1}{\sigma(r)} \sum_{t=0}^{n-1} \frac{g(r_t; r_{t+1})}{\sqrt{R(K)p(r)}}
\]
have the same asymptotic distribution, it is sufficient to show that 
\[
\frac{1}{\sigma(r)} \sum_{t=0}^{n-1} \frac{g(r_t; r_{t+1})}{\sqrt{R(K)p(r)}} \rightarrow N(0, 1)
\]
Note that from the proof of Proposition 2, we have 
\[
E(g(r_t; r_{t+1})) = \mu_p(r) \sqrt{h} \delta \{O(h^2 + \delta)\}
\]

Hence
\[
E\left( \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} g(r_t; r_{t+1}) \right) = \mu_p(r)(\sqrt{nh^5} + \sqrt{nh}\delta)O(\sqrt{\delta}) \rightarrow 0 \quad (31)
\]

Under the assumptions of Proposition 3 and from the proof of Proposition 2, we have
\[
E(g^2(r_t; r_{t+1})) \rightarrow \sigma^2(r)R(K)p(r) \quad (32)
\]
and
\[
E(g(r_0; r_1)g(r_t; r_{t+1})) \rightarrow 0 \quad \forall t \geq 1 \quad (33)
\]

It follows from equations (31),(32),(33), and Lemma 9 of Nicolau (2003) that:
\[
\frac{1}{\sqrt{n}} \sum_{t=0}^{n} g(r_t; r_{t+1}) \rightarrow N(0, \sigma^2(r)R(K)p(r)) \quad (34)
\]

This concludes the proof of Proposition 3.
Table 1. The Simulation Results of Vasicek Model

This table reports the simulation results of the proposed estimator, in comparison with the Stanton method, for the drift function of the Vasicek model. MIAB and RMISE denote mean integrated absolute bias and root mean integrated squared error, respectively.

<table>
<thead>
<tr>
<th>MIAB</th>
<th>RMISE</th>
<th>MIAB</th>
<th>RMISE</th>
<th>MIAB</th>
<th>RMISE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: Benchmark Stanton (1997) estimator</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{iid}$</td>
<td>$h_{opt}$</td>
<td>1.7877</td>
<td>7.8529</td>
<td>0.2436</td>
<td>2.3305</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Benchmark new pooled estimator (Case I: $\kappa = \gamma = 1, \rho = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{iid}$</td>
<td>$h_{opt}$</td>
<td>0.6695</td>
<td>5.1900</td>
<td>0.0584</td>
<td>0.8485</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Effect of mean reversion (Case II: $\gamma = 1, \rho = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa = 0.75$</td>
<td>$\kappa = 1.25$</td>
<td>$\kappa = 1.5$</td>
<td>0.0559</td>
<td>0.8407</td>
<td>0.0562</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel D: Effect of diffusion (Case III: $\kappa = 1, \rho = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma = 0.75$</td>
<td>$\gamma = 1.25$</td>
<td>$\gamma = 1.5$</td>
<td>0.0892</td>
<td>1.2927</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel E: Effect of correlation (Case IV: $\kappa = \gamma = 1$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = 0.2$</td>
<td>$\rho = 0.4$</td>
<td>$\rho = 0.6$</td>
<td>0.0605</td>
<td>1.1160</td>
<td>0.0626</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel F: Effect of mixed models (Case V: $\kappa = \gamma = 1, \rho = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auxiliary model: CIR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1079</td>
<td>1.1454</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The Simulation Results of CIR Model

This table reports the simulation results of the proposed estimator, in comparison with the Stanton method, for the drift function of the CIR model. MIAB and RMISE denote mean integrated absolute bias and root mean integrated squared error, respectively.

<table>
<thead>
<tr>
<th>MIAB</th>
<th>RMISE</th>
<th>MIAB</th>
<th>RMISE</th>
<th>MIAB</th>
<th>RMISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Benchmark Stanton (1997) estimator</td>
<td>1.1571</td>
<td>9.7516</td>
<td>0.3575</td>
<td>3.8268</td>
<td></td>
</tr>
<tr>
<td>Panel B: Benchmark new pooled estimator (Case I: $\kappa = \gamma = 1$, $\rho = 0$)</td>
<td>0.4336</td>
<td>4.3481</td>
<td>0.0848</td>
<td>1.7137</td>
<td></td>
</tr>
<tr>
<td>Panel C: Effect of mean reversion (Case II: $\gamma = 1$, $\rho = 0$)</td>
<td>$\kappa = 0.75$</td>
<td>0.1047</td>
<td>1.7176</td>
<td>$\kappa = 1.25$</td>
<td>0.1179</td>
</tr>
<tr>
<td>Panel D: Effect of diffusion (Case III: $\kappa = 1$, $\rho = 0$)</td>
<td>$\gamma = .75$</td>
<td>0.1468</td>
<td>2.1645</td>
<td>$\gamma = 1.25$</td>
<td>0.0516</td>
</tr>
<tr>
<td>Panel E: Effect of correlation (Case IV: $\kappa = \gamma = 1$)</td>
<td>$\rho = 0.2$</td>
<td>0.0891</td>
<td>1.9327</td>
<td>$\rho = 0.4$</td>
<td>0.0946</td>
</tr>
</tbody>
</table>
Table 3. Summary Statistics of Interest Rates

Panels A reports the summary statistics of the daily interest rates with maturities of 3-month, 6-month, 1-year, 3-year, 5-year and 10-year. 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.

<table>
<thead>
<tr>
<th>τ</th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>Min</th>
<th>Max</th>
<th>ρ1</th>
<th>ρ2</th>
<th>ρ3</th>
<th>ρ4</th>
<th>ρ5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Summary statistics of daily interest rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r3M</td>
<td>5.226</td>
<td>2.848</td>
<td>1.060</td>
<td>1.684</td>
<td>0.55</td>
<td>17.14</td>
<td>0.967</td>
<td>0.914</td>
<td>0.869</td>
<td>0.832</td>
<td>0.799</td>
</tr>
<tr>
<td>r6M</td>
<td>5.910</td>
<td>2.730</td>
<td>0.927</td>
<td>1.415</td>
<td>0.80</td>
<td>15.93</td>
<td>0.969</td>
<td>0.918</td>
<td>0.874</td>
<td>0.840</td>
<td>0.809</td>
</tr>
<tr>
<td>r1Y</td>
<td>6.365</td>
<td>2.932</td>
<td>0.945</td>
<td>1.301</td>
<td>0.88</td>
<td>17.31</td>
<td>0.970</td>
<td>0.921</td>
<td>0.879</td>
<td>0.846</td>
<td>0.819</td>
</tr>
<tr>
<td>r3Y</td>
<td>6.795</td>
<td>2.741</td>
<td>0.907</td>
<td>0.989</td>
<td>1.34</td>
<td>16.59</td>
<td>0.976</td>
<td>0.938</td>
<td>0.906</td>
<td>0.881</td>
<td>0.857</td>
</tr>
<tr>
<td>r5Y</td>
<td>7.009</td>
<td>2.630</td>
<td>0.959</td>
<td>0.878</td>
<td>2.08</td>
<td>16.27</td>
<td>0.980</td>
<td>0.947</td>
<td>0.920</td>
<td>0.897</td>
<td>0.875</td>
</tr>
<tr>
<td>r10Y</td>
<td>7.229</td>
<td>2.515</td>
<td>0.969</td>
<td>0.656</td>
<td>3.13</td>
<td>15.84</td>
<td>0.983</td>
<td>0.958</td>
<td>0.937</td>
<td>0.917</td>
<td>0.897</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: Summary statistics of daily interest rate changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δr3M</td>
<td>-0.609</td>
<td>0.102</td>
<td>0.202</td>
<td>26.11</td>
<td>-1.27</td>
<td>1.34</td>
<td>0.440</td>
<td>0.149</td>
<td>0.113</td>
<td>0.101</td>
<td>0.131</td>
</tr>
<tr>
<td>Δr6M</td>
<td>-0.618</td>
<td>0.090</td>
<td>0.296</td>
<td>23.74</td>
<td>-1.10</td>
<td>1.17</td>
<td>0.424</td>
<td>0.133</td>
<td>0.103</td>
<td>0.081</td>
<td>0.098</td>
</tr>
<tr>
<td>Δr1Y</td>
<td>-0.758</td>
<td>0.092</td>
<td>-0.194</td>
<td>21.39</td>
<td>-1.08</td>
<td>1.10</td>
<td>0.434</td>
<td>0.086</td>
<td>0.082</td>
<td>0.042</td>
<td>0.037</td>
</tr>
<tr>
<td>Δr3Y</td>
<td>-0.665</td>
<td>0.081</td>
<td>-0.180</td>
<td>13.26</td>
<td>-0.79</td>
<td>0.92</td>
<td>0.400</td>
<td>0.002</td>
<td>0.015</td>
<td>0.001</td>
<td>-0.059</td>
</tr>
<tr>
<td>Δr5Y</td>
<td>-0.440</td>
<td>0.076</td>
<td>-0.305</td>
<td>11.42</td>
<td>-0.77</td>
<td>0.72</td>
<td>0.392</td>
<td>-0.013</td>
<td>0.009</td>
<td>-0.002</td>
<td>-0.073</td>
</tr>
<tr>
<td>Δr10Y</td>
<td>0.103</td>
<td>0.068</td>
<td>-0.273</td>
<td>10.08</td>
<td>-0.75</td>
<td>0.65</td>
<td>0.349</td>
<td>-0.034</td>
<td>0.020</td>
<td>0.053</td>
<td>-0.055</td>
</tr>
</tbody>
</table>
Table 4. Correlation Matrix and Principal Components of Daily Interest Rate Changes

Panels A reports the correlation matrix of the daily changes of interest rates with different maturities. Panel B reports the principal components 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.

<table>
<thead>
<tr>
<th>Panel A: Correlation matrix of daily interest rate changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta r_{3M} )</td>
</tr>
<tr>
<td>( \Delta r_{3M} )</td>
</tr>
<tr>
<td>( \Delta r_{6M} )</td>
</tr>
<tr>
<td>( \Delta r_{1Y} )</td>
</tr>
<tr>
<td>( \Delta r_{3Y} )</td>
</tr>
<tr>
<td>( \Delta r_{5Y} )</td>
</tr>
<tr>
<td>( \Delta r_{10Y} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Principal Components of Daily Interest Rate Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>1 (level)</td>
</tr>
<tr>
<td>2 (slope)</td>
</tr>
<tr>
<td>3 (curvature)</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>
Table 5. Valuation of Interest Rate Caps under Alternative Short Rate Processes

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

<table>
<thead>
<tr>
<th>Spot Rate</th>
<th>Cap Tenor</th>
<th>Model</th>
<th>Strike Price (Basis Points)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>-50</td>
</tr>
<tr>
<td>0.05</td>
<td>3</td>
<td>New estimates</td>
<td>2.9242</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanton</td>
<td>3.3710</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.2075)</td>
</tr>
<tr>
<td>0.10</td>
<td>3</td>
<td>New estimates</td>
<td>6.2958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanton</td>
<td>7.3831</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.3467)</td>
</tr>
<tr>
<td>0.15</td>
<td>3</td>
<td>New estimates</td>
<td>5.2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanton</td>
<td>5.5648</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.8318)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.9889)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.5246</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanton</td>
<td>2.6133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.2268)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.4816</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stanton</td>
<td>5.0021</td>
</tr>
</tbody>
</table>
Figure 1. Drift function estimates of the Vasicek and CIR models
(with naive bandwidth $h_{iid}$)

This figure plots the short rate drift function estimates based on the Stanton (1997) estimator and the proposed estimator for the Vasicek and CIR models. The estimates are based on 7,500 daily observations with 5,000 replications. The naive bandwidth $h_{iid}$ is used for both estimators.
Figure 2. Drift function estimates of the Vasicek and CIR models
(with optimal bandwidth $h_{opt}$)

This figure plots the short rate drift function estimates based on the Stanton (1997) estimator and the proposed estimator for the Vasicek and CIR 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.
Panels A and B 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.
Figure 4. Nonparametric drift and diffusion function estimates

Panels A and B plot the short rate drift and diffusion function estimates based on the proposed estimator using the yields of all maturities, together with the Stanton 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.
Figure 5. Simulated yield curves under alternative short rate processes

This figure 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 method. The 90% point-wise confidence bands of the yield curves based on the new estimator are also plotted.