The Korean Economic Review
Volume 41, Number 1, Winter 2025, 141-174.
DOI : 10.22841/kerdoi.2025.41.1.005

Nonparametric Continuous Time Regressions with Functional Coefficients*

Mijung Choi** · Jihyun Kim*** · Nuong Nguyen****

This paper considers a continuous time regression with functional coefficients in conditional mean and variance functions, where the covariate of the regression is assumed to be a general recurrent diffusion. We propose a kernel-based nonparametric estimation for these functional coefficients using discretely sampled data from the underlying continuous time regression. We obtain the limiting behaviors of the proposed estimators through a two-dimensional asymptotic analysis while assuming a shrinking sampling interval and increasing time span and without the stationarity assumption. We demonstrate the feasibility our approach on a short-term interest rate model involving U.S. daily three-month treasury bill rates.

JEL Classification: C14, C22, C58
Keywords: Continuous Time Regression, Recurrent Diffusion, Non/semiparametric Model, Functional Coefficients, Kernel Estimation

I. Introduction

We consider a continuous time regression model with

$$dY_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

where $\mu(x)$ and $\sigma^2(x)$ are the instantaneous conditional mean and variance

Received: Dec. 29, 2023. Revised: July 29, 2024. Accepted: Sept. 25, 2024.
* We thank Nour Meddahi, Stéphane Gregoir, and the two anonymous referees for their many helpful comments.
** First Author, BK21 Post-Doc, Sungkyunkwan University; Associate Research Fellow, Korea Institute for Industrial Economics & Trade, e-mail: mijchoi@kiet.re.kr
*** Corresponding Author, Professor, Department of Economics, Sungkyunkwan University, e-mail: kim.jihyun@skku.edu
**** Co-Author, PhD Student, Department of Economics, University of Kentucky. e-mail: nuong.nguyen@uky.edu

functions, respectively, W is a standard Brownian motion, and the covariate X is a general Harris recurrent diffusion process that allows X to be either stationary or nonstationary. As for the conditional mean and variance functions, we define the functional coefficient specifications as $\mu(x) = \beta(x)x$ and $\sigma^2(x) = \gamma^2(x)x$. We consider nonparametric estimations of the functional coefficients $\beta(x)$ and $\gamma^2(x)$ for the conditional mean and variance functions of instantaneous increments in Y, respectively, based on discretely observed data $(X_{i\Delta}, Y_{i\Delta})_{i=0}^n$ with $T \equiv \Delta n$. Our asymptotic scheme is two dimensional and assumes that the sampling interval Δ decreases, the time span of data T increases, and the bandwidth h may depend on Δ and T.

The continuous time regression (1) have been widely used in the financial econometric literature. The first example is a continuous time predictive regression in finance, where dY often refers to asset returns, with the dividend–price ratio, the earnings–price ratio, or interest rate typically chosen as the predictor. The predictive regression model often considers a linear specification of the conditional mean function as $\mu(x) = \beta_0 + \beta_1 x$ and focuses on the estimation and inference for β_1 under certain problematic characteristics of financial data (see, e.g., Choi, Jacewitz, and Park, 2016; Kim and Park, 2017; Bu, Kim, and Wang, 2023; Ibragimov, Kim, and Skrobotov, 2023, and references therein).

The continuous time regression (1) also includes diffusion models as a special case by letting $Y = X$. Diffusion models have been widely used in the literature to describe the dynamics of underlying economic variables, such as stock prices and bond yields. These studies have used two approaches for modeling the drift $\mu(\cdot)$ and diffusion $\sigma^2(\cdot)$. The first approach is a parametric one, which assumes a specific functional form for the drift and diffusion functions. Most models using this approach exhibit mean-reversion with a linear drift specification as $\mu(x) = \beta_0 + \beta_1 x$ (see, e.g., Vasicek, 1977; Cox, Ingersoll, and Ross, 1985; Chan, Karolyi, Longstaff, and Sanders, 1992; Cai and Hong, 2003). To estimate the parameters in these models, maximum likelihood (see, e.g., Lo, 1988; Pearson and Sun, 1994) or the generalized method of moments (see, e.g., Duffie and Singleton, 1993; Jiang and Knight, 1997; Kim and Meddahi, 2020) can be employed.

However, the specifications of drift and diffusion functions are often chosen for theoretical convenience. Many of the existing diffusion models are not derived from any economic theory; thus, they may not necessarily be consistent with the financial data generation process (Cai and Hong, 2003; Fan, 2005), thus necessitating more flexible modeling techniques to address the issue of misspecification. The second approach, a fully nonparametric one, can serve as an alternative as it avoids explicit functional form specifications.

Some classical references for the fully nonparametric approach include Florens-Zmirou (1993), Ait-Sahalia (1996), Jiang and Knight (1997), and Stanton (1997), where the underlying processes are assumed to have a stationary probability density for the purposes of identification and estimation. However, this assumption might be too restrictive given the presence of nonstationary behaviors in many empirical applications. To address this issue, Bandi and Phillips (2003), Bandi and Phillips (2009), Ait-Sahalia and Park (2016), Kim, Park, and Wang (2021), and Bu, Kim, and Wang (2023) introduce a nonparametric method based on the notion of local time, where recurrence is the only requirement to guarantee the consistency of the drift and diffusion estimators.

In modeling the drift and diffusion functions, functional coefficient models can be used as an alternative in addition to the two aforementioned approaches. Using these models to analyze time series data is not new. The general setting of this framework is introduced in Cai, Fan, and Yao (2000), in which the multivariate regression with functional coefficient is specified as

$$E(Y \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{U} = \boldsymbol{u}) = \sum_{j=1}^{p} a_j(\boldsymbol{u}) x_j ,$$

where $Y \in \mathbb{R}$, $\boldsymbol{X} \in \mathbb{R}^p$, and $\boldsymbol{U} \in \mathbb{R}^k$, where $\boldsymbol{U}$ denotes the smooth variables, $a_j(\cdot)$ denotes the measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$, and $\boldsymbol{x} = (x_1, \ldots, x_p)^T$. In the literature, many time series models can be regarded as particular cases of this general form. For example, Chen and Tsay (1993), Tong (1990), Haggan and Ozaki (1981), Nicholls and Quinn (1982), Chib, Nardari, and Shephard (2002), Zhang and Wu (2012), Han and Lee (2018), and Han and Lee (2020) estimate the functional coefficients using either nonparametric or Bayesian estimation techniques.

We also utilize this modeling technique but in the context of the continuous time regression model. By doing so, our paper contributes an alternative approach to modeling continuous time regressions and diffusion processes that differs from the parametric and fully nonparametric ones used in the literature. This new approach offers two important advantages over the existing fully parametric or nonparametric models. On the one hand, the proposed approach is nearly as flexible as the fully nonparametric approach, thus avoiding the misspecification issue of parametric models. On the other hand, the proposed approach enhances the interpretability of fitted models, thereby offering an advantage over the fully nonparametric approach. In estimating functional coefficients, we employ a nonparametric technique based on locally weighted regression.

Asymptotic theories for the nonparametric estimators of a diffusion model have been established by several authors, including Ait-Sahalia and Park (2016), Bandi and Phillips (2003), Florens-Zmirou (1993), Fan and Zhang (2003), Jiang and Knight (1997), and Ait-Sahalia (1996). In this paper, the approach we used to investigate the asymptotic behaviors of nonparametric estimators for the functional coefficients of the conditional mean and variance functions is closely related to that of Ait-Sahalia and Park (2016) and Bandi and Phillips (2009), where two-dimensional asymptotics are considered. Specifically, this approach allows the time span T to increase and the sampling interval Δ to decrease simultaneously as opposed to more conventional asymptotics that only consider the sample size $n = T/\Delta$. The two-dimensional asymptotics are particularly useful in examining nonstationary diffusion processes as they provide a unified framework that accommodates stationary and nonstationary processes. This analytical framework can also guide the selection of optimal bandwidths for the considered nonparametric estimators (Ait-Sahalia and Park, 2016). We adapt this framework to the functional coefficients of the conditional mean and variance functions and develop asymptotic theories for the functional coefficient estimators.

The rest of this paper is organized as follows. Section 2 introduces the model, its properties, and the underlying assumptions. Section 3 defines the nonparametric estimators for the functional coefficients of the conditional mean and variance functions and develops their asymptotics. Section 4 illustrates the proposed nonparametric methodology through an application to U.S. daily three-month treasury bill data and compares the obtained estimates with those obtained by the CIR model and the fully nonparametric model. Section 5 concludes the paper. The Appendix provides all mathematical proofs.

II. The Model and Preliminaries

We assume that the covariate process X of the continuous time regression (1) is a diffusion process following the stochastic differential equation

$$dX_t = a(X_t)dt + b(X_t)dV_t \tag{2}$$

where a and b^2 represent the drift and diffusion functions, respectively, defined on the domain $D = (\underline{x}, \overline{x})$, and V is a standard Brownian motion with $E[dV_t dW_t] = \rho dt$ for some $\rho \in [-1,1]$. The domain D of X is either $(-\infty,\infty)$ or $(0,\infty)$, which is usually the case for interest rates, the logarithm prices of financial assets, or exchange rates.

Assumption 1. *(a)* $a^2(x) > 0$ *on* D, *(b)* $a(x)$ *and* $b^2(x)$ *are twice continuously differentiable on* D.

Under Assumption 1, a unique weak solution to (2) exists in probability law (see, e.g., Theorem 5.15 in Chapter 5 in Karatzas and Shreve, 1991). Note that the same assumption also appears in Ait-Sahalia and Park (2016) and Bu, Kim, and Wang (2023).

The asymptotics developed in this paper heavily rely on the local time l on $[0,T]$ of the process X at an interior point x of D, which is defined by

$$l(T,x)=\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T 1\{\,|X_s-x|<\varepsilon\}ds$$

Intuitively, $l(T,x)$ is a random quantity that signifies the amount of time spent by X in a neighborhood of x on the time interval $[0,T]$. This quantity may also be interpreted as an occupation density of X at each point $x\in D$, thereby yielding the occupation time formula

$$\int_0^T f(X_t)dt=\int_{-\infty}^{\infty} f(x)l(T,x)dx$$

for any nonnegative measurable function f on $\mathbb{R}$ (see, e.g, Bandi and Phillips, 2003 and Ait-Sahalia and Park, 2016).

Assumption 2. *(a)* $\overline{l}_h(T,x)=\sup_{|u|\leqslant 1} l(T,x+hu)=O_p(l(T,x)^2)$, *and (b)* $l(T,x)=O_p(\mu_T)$ *for some nonrandom sequence* (μ_T).

Assumption 2 (a) regulates the divergence rate of local time in the vicinity of a spatial point. This condition is not essential, and its primary purpose is to simplify our exposition by representing the orders of error terms merely as functions of $l(T,x)$. Assumption 2 (b) is obviously satisfied if $T=\overline{T}$ is fixed. Otherwise, given that $T\to\infty$, the asymptotic behavior of $l(T,x)$ is determined by the recurrence property of the diffusion process X.

The scale density s' of X is defined as

$$s'(x)=\exp\left[-2\int_z^x\frac{a(u)}{b^2(u)}du\right],$$

where the lower limit of the integral can be arbitrarily chosen to be any point $z\in D$.[1] Then, X is a recurrent process if and only if the scale function $s(x)=\int_z^x s'(y)dy$ is unbounded at the boundaries of the domain of X, that is, $s(\underline{x})=-\infty$ and $s(\overline{x})=\infty$. Otherwise, X is said to be transient. For a recurrent process X, we define the speed density as

[1] The scale function s is uniquely identified up to any increasing affine transformation, i.e., if $s(x)$ is a scale function, then so is $as(x)+b$ for any constants $a>0$ and $b\in(-\infty,\infty)$.

$$m(x) = \frac{1}{(b^2 s')(x)},$$

and X becomes positively recurrent if $\int_D m(x)dx < \infty$. Otherwise, X is a null recurrent process (see, e.g., Chapter 5 in Karatzas and Shreve, 1991 for the definitions of $s(\cdot)$ and $m(\cdot)$). When X is a recurrent process, $l(T,x) \to_p \infty$ as $T \to \infty$ at each $x \in D$ given that it visits every point infinitely many times as $T \to \infty$ with a probability of 1 (see, e.g., Chapter VI in Revuz and Yor, 1999 for the properties of Brownian local times). Throughout this paper, we assume that X is a recurrent diffusion that covers stationary and nonstationary diffusions depending upon $\int_D m(x)dx < \infty$ or unbounded.

Assumption 3. *The kernel function* K *satisfies (a)* K *is nonnegative, bounded, twice continuously differentiable, and has support* $[-1,1]$*, and (b)* $\int_{-\infty}^{\infty} K(x)dx = 1$ *and* $\int_{-\infty}^{\infty} xK(x)dx = 0$.

The conditions for the kernel functions in Assumption 3 are standard, except for the boundedness of support, thus allowing us to simplify the proofs of our theorems.

In our asymptotics, we require Δ to be sufficiently small relative to the extremal bounds of various functional transforms of X over time interval $[0,T]$. Similar to Ait-Sahalia and Park (2016), we define

$$T(f) = \max_{0 \leqslant t \leqslant T} |f(X_t)|$$

for a measurable function $f : D \to \mathbb{R}$. Moreover, we define

$$\mathcal{A} = a\mathcal{D} + \frac{1}{2}b^2\mathcal{D}^2, \text{ and } \mathcal{B} = b\mathcal{D},$$

where $\mathcal{D}$ is the differential operator. If we define $f_A = \mathcal{A}f = af' + b^2 f''/2$ and $f_B = \mathcal{B}f = bf'$ for a twice continuously differentiable function f and if f' and f'' denote the first and second derivatives of f, respectively, then we may deduce from Itô's formula that

$$f(X_t) - f(X_s) = \int_s^t f_A(X_u)du + \int_s^t f_B(X_u)dV_u$$

for any $0 \leqslant s \leqslant t$.

III. Nonparametric Estimation of the Functional Coefficients

In this section, we propose the nonparametric estimation of the functional coefficients of the continuous time regression (1) when the covariate process X is a recurrent diffusion that satisfies Assumptions 1 and 2. We assume that the continuous time regression satisfies the following functional coefficient specifications:

Assumption 4. *(a)* $\sigma^2(x) > 0$ *on* D*, (b)* $\mu(x) = \beta(x)x$ *and* $\sigma^2(x) = \gamma^2(x)x$*, and (c)* $\beta(x)$ *and* $\gamma^2(x)$ *are twice continuously differentiable on* D.

Under Assumption 4, one may show that the discrete observation $(Y_{i\Delta}, X_{i\Delta})_{i=0}^{n}$ satisfies the following approximations:

$$Y_{i\Delta} - Y_{(i-1)\Delta} \approx \Delta\beta(X_{(i-1)\Delta})X_{(i-1)\Delta} + \int_{(i-1)\Delta}^{i\Delta} \sigma(X_t)dW_t$$

and

$$(Y_{i\Delta} - Y_{(i-1)\Delta})^2 \approx \Delta\gamma^2(X_{(i-1)\Delta})X_{(i-1)\Delta} + 2\int_{(i-1)\Delta}^{i\Delta} (Y_t - Y_{(i-1)\Delta})\sigma(X_t)dW_t \,.$$

Therefore, the functional coefficients $\beta(x)$ and $\gamma^2(x)$ may be estimated by a local regression with a kernel function

$$\hat{\beta}(x) = \frac{P_T(K,1)}{Q_T(K)} \quad \text{and} \quad \hat{\gamma}^2(x) = \frac{P_T(K,2)}{Q_T(K)}$$

where

$$P_T(K, j) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}(Y_{i\Delta} - Y_{(i-1)\Delta})^j$$

$$Q_T(K) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2$$

In what follows, we develop the asymptotic theory for $\hat{\beta}(x)$ and $\hat{\gamma}^2(x)$.

3.1. The Functional Coefficient of the Conditional Mean Function

To effectively explain our asymptotics, we decompose $\hat{\beta}(x)$ as

$$\hat{\beta}(x) = \hat{\beta}_P(x) + \hat{\beta}_q(x) + \hat{\varepsilon}_\beta(x),$$

where

$$\hat{\beta}_p(x) = \beta(x) + \frac{N_T(K,1)}{Q_T(K)},\ \hat{\beta}_q(x) = \frac{M_T(K,1)}{Q_T(K)},\ \hat{\varepsilon}_\beta(x) = \frac{R_T(K,1)}{Q_T(K)}$$

with

$$N_T(K,1) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 [\beta(X_{(i-1)\Delta}) - \beta(x)]$$
$$M_T(K,1) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta} \sigma(X_t)dW_t$$
$$R_T(K,1) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta} [\mu(X_t) - \mu(X_{(i-1)\Delta})]dt$$

From the above expressions, one can say that $\hat{\beta}_p(x) - \beta(x)$ and $\hat{\beta}_q(x)$ are the bias and variance of the estimator, respectively, while $\hat{\varepsilon}_\beta(x)$ is the approximation error, which is negligible asymptotically as long as Δ is sufficiently small relative to h.

To establish the asymptotic properties of the estimator for $\beta(x)$, we introduce conditions on h and Δ in the following assumption:

Assumption 5. *$h \to_p 0$ and $\Delta \to 0$ such that (a) $h^{-4}\Delta = o_p(1)$, and (b) $\Delta^{1/2}T(\mu_A) = o_p(1)$, $\Delta^{1/2}T(\mu_B) = o_p((hl(T,x))^{1/2})$ uniformly in T as $h \to_p 0$ and $\Delta \to 0$.*

Due to our asymptotic scheme, $\Delta \to 0$ should be sufficiently fast relative to $h \to_p 0$ and $T(\mu_A)$ and $T(\mu_B)$. In particular, our asymptotic results are relevant for the case where Δ is sufficiently small relative to T. For Assumption 5 to hold, $\Delta = o_p(h^4)$ and $\Delta / T = o_p(h)$ if X is stationary and bounded so that $T(\mu_A)$ and $T(\mu_B)$ are constants. In other words, the sampling interval Δ is sufficiently small relative to the bandwidth h and the span of data T. The condition appears to be mild enough to yield asymptotics that are generally relevant for a very wide range of empirical analyses that rely on samples collected from diffusion-type models. For daily observations of over 60 years, such as our empirical analysis in

Section 4, we have $\Delta = 1/252$ and $T^{-1} = 1/60$. Our asymptotics also hold jointly in Δ and T as long as they satisfy Assumption 5 as $\Delta \to 0$ with T being fixed or $T \to \infty$. We do not use sequential asymptotics, which require $\Delta \to 0$ and $T \to \infty$ sequentially.

Lemma 1. *Under Assumptions 1, 2, 3, 4, and 5, we have*

$$\hat{\beta}_p(x) = \beta(x) + \frac{h^2}{2} B_\beta(x) \iota(K_2) + o_p(h^2) + O_p(h^{3/2} l(T,x)^{-1/2}),$$

where $\iota(K_2) = \int x^2 K(x) dx$ *and* $B_\beta(x) = 4x^{-1}\beta'(x) + \beta''(x) + 2\beta'(x) m'(x)/m(x)$, *uniformly in* T *as* $h \to 0$ *and* $\Delta \to 0$, *and*

$$x[hl(T,x)]^{1/2} \hat{\beta}_q(x) \to_d \sigma(x) \iota(K^2)^{1/2} Z$$

as $l(T,x) \to_p \infty$, *where* $\iota(K^2) = \int K^2(x) dx$ *and* Z *is a standard normal random variate independent of* $l(T,x)$. *In addition,*

$$\hat{\varepsilon}_\beta(x) = o_p(h^2).$$

Lemma 1 provides the asymptotic behavior of the bias $\hat{\beta}_p(x) - \beta(x)$, the variance $\hat{\beta}_q(x)$, and the error $\hat{\varepsilon}_\beta(x)$. The bias and error terms are asymptotically negligible as $h \to_p 0$, while the variance term becomes negligible as $hl(T,x) \to_p \infty$.

Proposition 1. *Let Assumptions 1, 2, 3, 4, and 5 hold, and let* $l(T,x) \to_p \infty$ *and* $h = cl(T,x)^r$ *for some constant* $c > 0$. *In this case, we have:*

(i) If $h = cl(T,x)^r$ *with* $r = -1/5$, *then*

$$x[hl(T,x)]^{1/2} \left[\hat{\beta}(x) - \beta(x) - \frac{h^2}{2} \iota(K_2) B_\beta(x) \right] \to_d \sigma(x) \iota(K^2)^{1/2} Z.$$

(ii) If $h = cl(T,x)^r$ *with* $r < -1/5$, *then*

$$x[hl(T,x)]^{1/2} [\hat{\beta}(x) - \beta(x)] \to_d \sigma(x) \iota(K^2)^{1/2} Z.$$

Remark 1. (a) Note that $\hat{\beta}$ is consistent if $l(T,x) \to_p \infty$, which is induced by $T \to \infty$ given that X is recurrent. In particular, $\hat{\beta}$ becomes inconsistent if T is fixed.

(b) The asymptotic bias term B_β appears only when $h = cl(T,x)^r$ with $r = -1/5$ and becomes negligible if $h = cl(T,x)^r$ with $r < -1/5$.

(c) By Proposition 1 (ii), if $h = cl(T,x)^r$ with $r < -1/5$, then

$$\frac{x[hl(T,x)]^{1/2}}{\sigma(x)\iota(K^2)^{1/2}}[\hat{\beta}(x) - \beta(x)] \to_d Z$$

and the limiting normal random variable Z is independent of the local time $l(T,x)$. In other words, $\hat{\beta}(x)$ has a limit of normal mixture. From this result, we can construct confidence intervals for $\hat{\beta}(x)$. We define $\theta(h,T) = (x[hl(T,x)]^{1/2})/(\sigma(x)\iota(K^2)^{1/2})$ and let $z_{1-\alpha}$ denote the $1-\alpha$ quantile of a $N(0,1)$ distribution. We then have

$$\begin{aligned}&\mathbb{P}(\beta(x) \in [\hat{\beta}(x) - \theta(h,T)^{-1} z_{1-\alpha/2}, \hat{\beta}(x) + \theta(h,T)^{-1} z_{1-\alpha/2}])\\ &= \mathbb{P}(\theta(h,T)[\hat{\beta}(x) - \beta(x)] \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]) \to 1-\alpha \quad \text{as} \quad l(T,x) \to \infty\end{aligned}$$

Therefore, the $1-\alpha$ asymptotic confidence interval for $\hat{\beta}(x)$ is given by

$$[\hat{\beta}(x) - \theta(h,T)^{-1} z_{1-\alpha/2}, \hat{\beta}(x) + \theta(h,T)^{-1} z_{1-\alpha/2}]$$

Note that θ involves the two unknown functions $\sigma^2(x)$ and $l(T,x)$, which can be estimated nonparametrically (see, e.g., Ait-Sahalia and Park, 2016; Bu, Kim, and Wang, 2023).

(d) Given the asymptotics of $\hat{\beta}(x)$, we can also establish the asymptotic behavior of the drift function estimator $\hat{\mu}(x) = \hat{\beta}(x)x$ as

$$\frac{[hl(T,x)]^{1/2}}{\sigma(x)\iota(K^2)^{1/2}}[\hat{\mu}(x) - \mu(x)] \to_d Z$$

As in Remark 1 (c), one can also construct the confidence interval for $\hat{\mu}(x)$ as

$$[\hat{\mu}(x) - \lambda(h,T)^{-1} z_{1-\alpha/2}, \hat{\mu}(x) + \lambda(h,T)^{-1} z_{1-\alpha/2}],$$

where $\lambda(h,T) = ([hl(T,x)]^{1/2})/(\sigma(x)\iota(K^2)^{1/2})$.

(e) For a given x, the asymptotic mean squared error (AMSE) of $\hat{\beta}(x)$ can be expressed as the sum of its squared asymptotic bias (ABias) and asymptotic variance (AVar) as

$$\text{AMSE}[\hat{\beta}(x)] = \text{ABias}^2[\hat{\beta}(x)] + \text{AVar}[\hat{\beta}(x)],$$

where $\text{ABias}[\hat{\beta}(x)]$ and $\text{AVar}[\hat{\beta}(x)]$ can be obtained by Lemma 1 as

$$\text{ABias}[\hat{\beta}(x)] = \frac{h^2}{2}\iota(K_2)B_\beta(x) \quad \text{and} \quad \text{AVar}[\hat{\beta}(x)] = \frac{x^{-1}\gamma^2(x)\iota(K^2)}{hl(T,x)}$$

The optimal bandwidth h_β^* that minimizes $\text{AMSE}[\hat{\beta}(x)]$ is derived by taking the first-order condition with respect to h and is given as

$$h_\beta^* = \frac{\iota^{1/5}(K^2)}{\iota^{2/5}(K_2)} x^{-1/5}\gamma^{2/5}(x)B_\beta(x)^{-2/5}l(T,x)^{-1/5}$$

In particular, the optimal bandwidth h_β^* is a function of local time, such as $h_\beta^* = cl(T,x)^{-1/5}$ for some $c > 0$. Note that the divergence rate of the local time is determined by the degree of recurrence. If X is stationary, then $l(T,x)$ diverges at the rate of T, and hence, $h_\beta^* = cT^{-1/5}$ as in the discrete time case. Null recurrent processes are less recurrent than stationary processes. For a Brownian motion, the local time diverges at the rate of $T^{1/2}$, and hence, $h_\beta^* = cT^{-1/10}$.

3.2. The Functional Coefficient of the Conditional Variance Function

Similar to $\hat{\beta}$, we decompose $\hat{\gamma}^2(x)$ as

$$\hat{\gamma}^2(x) = \hat{\gamma}_p^2(x) + \hat{\gamma}_q^2(x) + \hat{\varepsilon}_{\gamma^2}(x)$$

where

$$\hat{\gamma}_p^2(x) = \gamma^2(x) + \frac{N_T(K,2)}{Q_T(K)}, \quad \hat{\gamma}_q^2(x) = \frac{2M_T(K,2)}{Q_T(K)}, \quad \hat{\varepsilon}_{\gamma^2}(x) = \frac{R_T(K,2) + 2S_T(K)}{Q_T(K)}$$

with

$$N_T(K,2) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2[\gamma^2(X_{(i-1)\Delta}) - \gamma^2(x)]$$

$$M_T(K,2) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})\sigma(X_t)dW_t$$

$$R_T(K,2) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})]dt$$

$$S_T(K) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta} \int_{(i-1)\Delta}^{i\Delta} (Y_t - Y_{(i-1)\Delta})\mu(X_t)dt$$

Similar to the conditional mean function estimation, $\hat{\gamma}_p^2(x) - \gamma^2(x)$ and $\hat{\gamma}_q^2(x)$ are the bias and variance of the estimator, respectively, while $\hat{\varepsilon}_{\gamma^2}(x)$ is the approximation error, which is negligible asymptotically as long as Δ is sufficiently small relative to h.

To establish the asymptotic properties of the estimator for $\gamma^2(x)$, we introduce specific conditions on h and Δ in the following assumption:

Assumption 6. *$h \to_p 0$ and $\Delta \to 0$ such that (a) $h^{-4}\Delta = o_p(1)$, and (b) $\Delta^{1/2}T(\sigma_A^2) = o_p(1)$, $\Delta^{1/2}T(\sigma_B^2) = o_p((hl(T,x))^{1/2})$ uniformly in T as $h \to_p 0$ and $\Delta \to 0$, (c) $\Delta T(\mu_A^2)$, $\Delta T(\sigma_A) = o_p(1)$ and $\Delta T(\mu_B^2)$, $\Delta T(\sigma_B^2) = o_p((hl(T,x))^{1/2})$ uniformly in T as $h \to_p 0$ and $\Delta \to 0$, and (d) $\Delta T(a)T(\sigma^{2\prime})$, $\Delta^{1/2}T(b)T(\sigma^{2\prime})$ $= o_p(1)$.*

In this assumption, we only need $\Delta \to 0$ to be sufficiently fast relative to h and the extremal processes of X under various transforms given by the drift and diffusion functions. In particular, the long span assumption $T \to \infty$ is not required.

Lemma 2. *Under Assumptions 1, 2, 3, 4, and 6, we have*

$$\hat{\gamma}_p^2(x) = \gamma^2(x) + \frac{h^2}{2}\iota(K_2)B_{\gamma^2}(x) + o_p(h^2) + O_p(h^{3/2}l(T,x)^{-1/2}),$$

where $B_{\gamma^2}(x) = 4x^{-1}\gamma^{2\prime}(x) + \gamma^{2\prime\prime}(x) + 2\gamma^{2\prime}(x)m'(x)/m(x)$, uniformly in T as $h \to 0$ and $\Delta \to 0$, and

$$x\left[\frac{hl(T,x)}{\Delta}\right]^{1/2} \hat{\gamma}_q^2(x) \to_d \sqrt{2}\sigma^2(x)\iota(K^2)^{1/2} Z$$

where Z is a standard normal random variate independent of $l(T,x)$. In addition,

$$\hat{\varepsilon}_{\gamma^2}(x) = o_p(h^2)$$

In Lemma 2, the asymptotics of the bias $\hat{\gamma}_p^2(x) - \gamma^2(x)$, the asymptotic variance $\hat{\gamma}_q^2(x)$, and the error $\hat{\varepsilon}_{\gamma^2}(x)$ are provided. The bias $\hat{\gamma}_p^2(x) - \gamma^2(x)$ and the error $\hat{\varepsilon}_{\gamma^2}(x)$ are asymptotically negligible as $h \to 0$, while the variance term $\hat{\gamma}_q^2(x)$

converges to zero if $\Delta \to 0$ relative to $h \to_p 0$.

Proposition 2. *Let Assumptions 1, 2, 3, 4, and 6 hold, and let* $h = c[l(T,x)/\Delta]^r$ *for some constant* $c > 0$. *We thus have:*

(i) If $h = c[l(T,x)/\Delta]^{-1/5}$, *then*

$$x\left[\frac{hl(T,x)}{\Delta}\right]^{1/2}\left[\hat{\gamma}^2(x) - \gamma^2(x) - \frac{h^2}{2}\iota(K_2)B_{\gamma^2}(x)\right] \to_d \sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}Z.$$

(ii) If $h = c[l(T,x)/\Delta]^r$ *with* $r < -1/5$, *then*

$$x\left[\frac{hl(T,x)}{\Delta}\right]^{1/2}[\hat{\gamma}^2(x) - \gamma^2(x)] \to_d \sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}Z.$$

Remark 2. (a) The estimator $\hat{\gamma}^2$ is consistent as long as $h \to_p 0$ and $\Delta \to 0$ are sufficiently fast. Unlike $\hat{\beta}$, the consistency of $\hat{\gamma}^2$ does not require $T \to \infty$, and γ^2 can be estimated consistently even when T is fixed.

(b) The asymptotic bias term B_{γ^2} appears only when $h = c[l(T,x)/\Delta]^r$ with $r = -1/5$ and becomes negligible if $h = c[l(T,x)/\Delta]^r$ with $r < -1/5$.

(c) By Proposition 2 (ii), if $h = cl(T,x)^r$ with $r < -1/5$, then

$$\frac{x[hl(T,x)]^{1/2}}{\sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}\Delta^{1/2}}[\hat{\gamma}^2(x) - \gamma^2(x)] \to_d Z$$

and the limiting normal random variable Z is independent of the local time $l(T,x)$. Similar to $\hat{\beta}$, $\hat{\gamma}^2(x)$ has a limit of normal mixture, thereby allowing us to construct confidence intervals for $\hat{\gamma}^2(x)$ based on its asymptotic distribution. We define $\theta(h,T,\Delta) = (x[hl(T,x)]^{1/2})/(\sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}\Delta^{1/2})$ and let $z_{1-\alpha}$ denote the $1-\alpha$ quantile of a $N(0,1)$ distribution. We then have

$$\begin{aligned}&\mathbb{P}(\gamma^2(x) \in [\hat{\gamma}^2(x) - \theta(h,T,\Delta)^{-1}z_{1-\alpha/2}, \hat{\gamma}^2(x) + \theta(h,T,\Delta)^{-1}z_{1-\alpha/2}])\\ &= \mathbb{P}(\theta(h,T,\Delta)[\hat{\gamma}^2(x) - \gamma^2(x)] \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]) \to 1-\alpha \quad \text{as} \quad l(T,x)/\Delta \to \infty,\end{aligned}$$

and hence, the $1-\alpha$ asymptotic confidence interval for $\hat{\gamma}^2(x)$ is given by

$$[\hat{\gamma}^2(x) - \theta(h,T,\Delta)^{-1}z_{1-\alpha/2}, \hat{\gamma}^2(x) + \theta(h,T,\Delta)^{-1}z_{1-\alpha/2}].$$

(d) Given the asymptotic distribution of $\hat{\gamma}^2(x)$, we can also establish asymptotics for $\hat{\sigma}^2(x) = \hat{\gamma}^2(x)x$ as

$$\frac{[hl(T,x)]^{1/2}}{\sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}\Delta^{1/2}}[\hat{\sigma}^2(x)-\sigma^2(x)]\to_d Z$$

If we define $\lambda(h,T,\Delta)=([hl(T,x)]^{1/2})/(\sqrt{2}\sigma^2(x)\iota(K^2)^{1/2}\Delta^{1/2})$, then the confidence interval for $\hat{\sigma}^2(x)$ is given by

$$[\hat{\sigma}^2(x)-\lambda(h,T,\Delta)^{-1}z_{1-\alpha/2},\hat{\sigma}^2(x)+\lambda(h,T,\Delta)^{-1}z_{1-\alpha/2}].$$

(e) For a given x, the AMSE of $\hat{\gamma}^2(x)$ can be expressed as the sum of its squared ABias and AVar as

$$\text{AMSE}[\hat{\gamma}^2(x)]=\text{ABias}^2[\hat{\gamma}^2(x)]+\text{AVar}[\hat{\gamma}^2(x)],$$

where $\text{ABias}[\hat{\gamma}^2(x)]$ and $\text{AVar}[\hat{\gamma}^2(x)]$ can be obtained by Lemma 2 as

$$\text{ABias}[\hat{\gamma}^2(x)]=\frac{h^2}{2}\iota(K_2)B_{\gamma^2}(x) \quad \text{and} \quad \text{AVar}[\hat{\gamma}^2(x)]=\frac{2\gamma^4(x)\iota(K^2)\Delta}{hl(T,x)}$$

The optimal bandwidth $h^*_{\gamma^2}$ that minimizes $\text{AMSE}[\hat{\gamma}^2(x)]$ is given by

$$h^*_{\gamma^2}=\sqrt[5]{2}\frac{\iota(K^2)^{1/5}}{\iota(K_2)^{2/5}}\left(\frac{\sigma^2(x)}{x}\right)^{2/5}B_{\gamma^2}(x)^{-2/5}\Delta^{1/5}l(T,x)^{-1/5}$$

Unlike h^*_β, the optimal bandwidth $h^*_{\gamma^2}$ depends not only local time but also on Δ, such as $h^*_\beta=c(\Delta/l(T,x))^{1/5}$ for some $c>0$. Recall that $T=n\Delta$ and the sample size n can diverge as long as $\Delta\to 0$ regardless of T being fixed or not. For a stationary X, the optimal bandwidth becomes a function of sample size n as $h^*_{\gamma^2}=cn^{-1/5}$. In general, the optimal bandwidth is not given by the sample size alone.

IV. Empirical Analysis

4.1. Data

As an empirical application, we use our approach to model interest rate dynamics under a diffusion setting with $Y=X$. Specifically, we consider the U.S. three-month treasury bill yield data collected from the Federal Reserve Bank of St. Louis. These data include daily secondary market rates spanning from January 1954 to

April 2018, amounting to a total of 16,052 observations. No specific adjustments have been made for weekends or holidays.

[Figure 1] U.S. daily 3-month treasury bill rate from January 1954 to April 2018

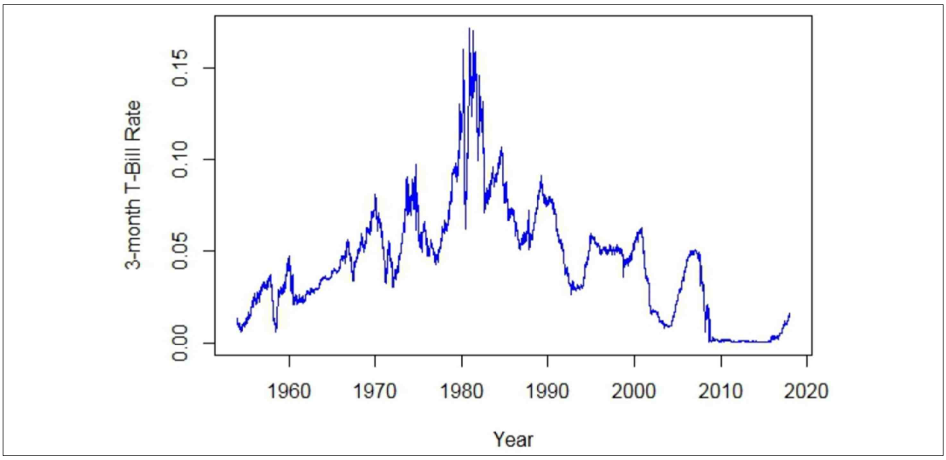
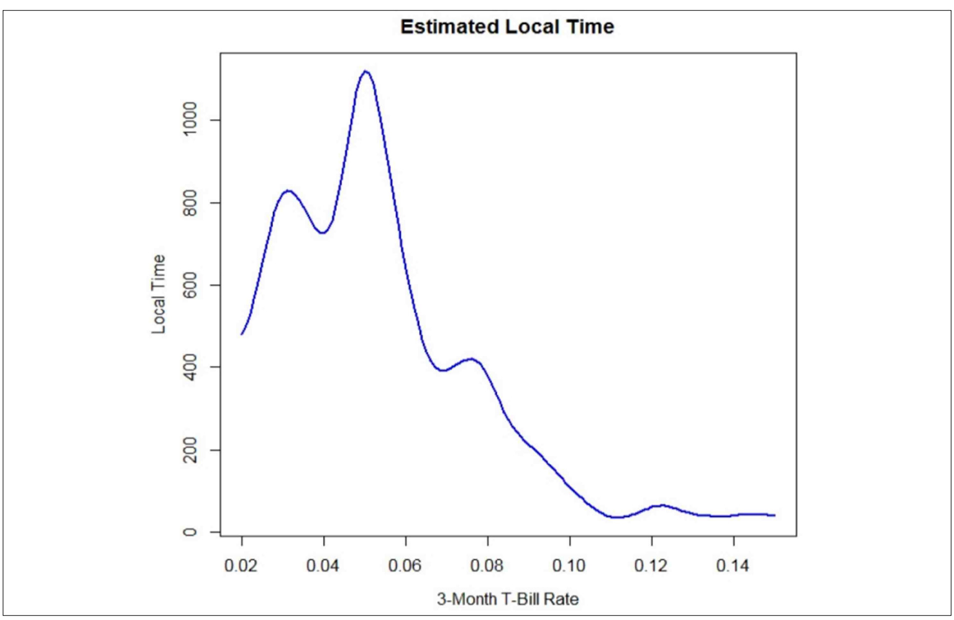


Figure 1 presents a time series plot of the data, while Figure 2 shows the estimated local time of the treasury bill process to offer an insight into its distribution. The estimated local time peaks around 0.05, and the treasury bill process makes most of its visit at levels between approximately 0.02 and 0.08 during the sample period.

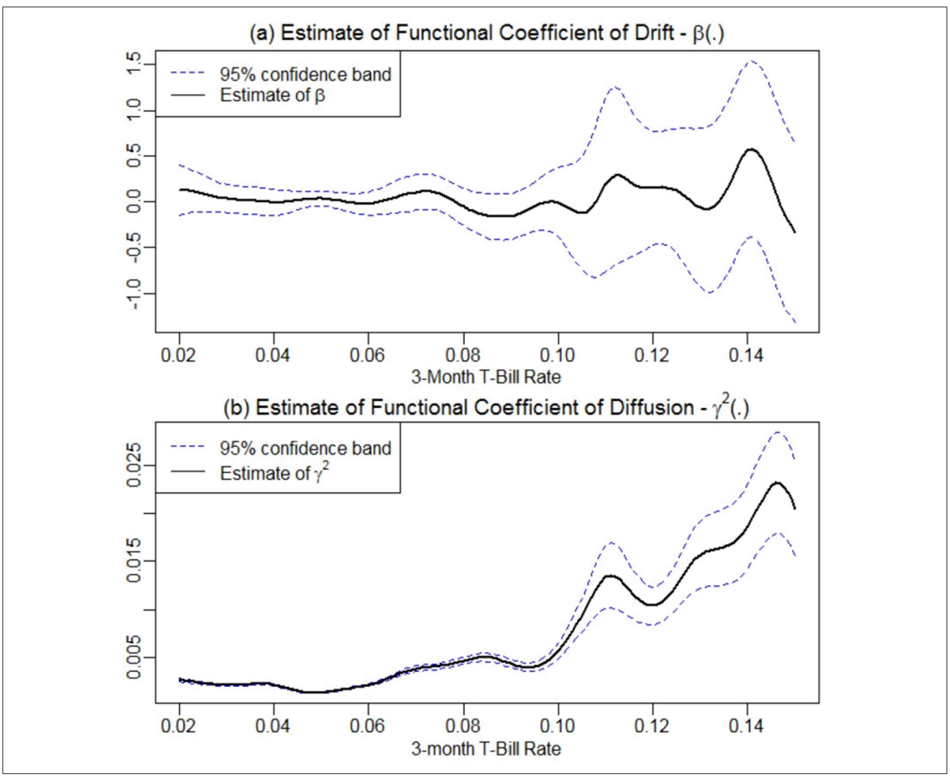
[Figure 2] The estimates of local time



4.2. Estimation Results

Figure 3 illustrates the estimated functional coefficients for the drift and diffusion terms, which are computed using the estimators introduced in Sections 3.1 and 3.2. This figure also includes the 95% pointwise confidence intervals for these estimates derived from the asymptotics. The coefficients for the drift and diffusion functions are visibly nonconstant. Specifically, we observe an increase in the functional coefficient of the diffusion term corresponding to higher rate levels. Meanwhile, the confidence intervals, which serve as indicators of estimation quality, are narrower at relatively lower interest rates and widen at higher interest rates due to limited observations as expected from Figure 2.

[Figure 3] Estimated functional coefficients of drift and diffusion terms with 95% confidence band



We then compare our model with the conventional models proposed in the literature. Among many existing models, we consider the fully nonparametric CIR model proposed by Cox, Ingersoll, and Ross (1985), which has a linear drift and allows complete flexibility in the functional forms of the drift and diffusion functions, denoted as Model 2. We denote the functional coefficient model as

Model 1 given that its flexibility lies between these two approaches. Each of these models provides a distinct framework for capturing the evolution of a stochastic process over time. Analyzing their properties, estimating their parameters, and comparing their abilities to fit empirical data can offer some insights into the underlying dynamics of the process being modeled.

Model CIR: $dX_t = \kappa(\mu - X_t)dt + \sigma\sqrt{X_t}dW_t$
Model 1 (Functional-coefficient): $dX_t = \beta(X_t)X_t dt + \gamma(X_t)\sqrt{X_t}dW_t$
Model 2 (Fully-nonparametric): $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$

Model CIR is estimated using the maximum likelihood method and is given by

$$dX_t = 0.0544(0.0452 - X_t)dt + 0.0674\sqrt{X_t}dW_t$$

For Model 1, using the estimated functional coefficients $\hat{\beta}(x)$ and $\hat{\gamma}^2(x)$, we derive the estimated drift and diffusion functions as $\hat{\beta}(x)x$ and $\hat{\gamma}^2(x)x$, respectively. We then compute the 95% pointwise confidence bands for these estimated functions using the formula obtained from our asymptotics.

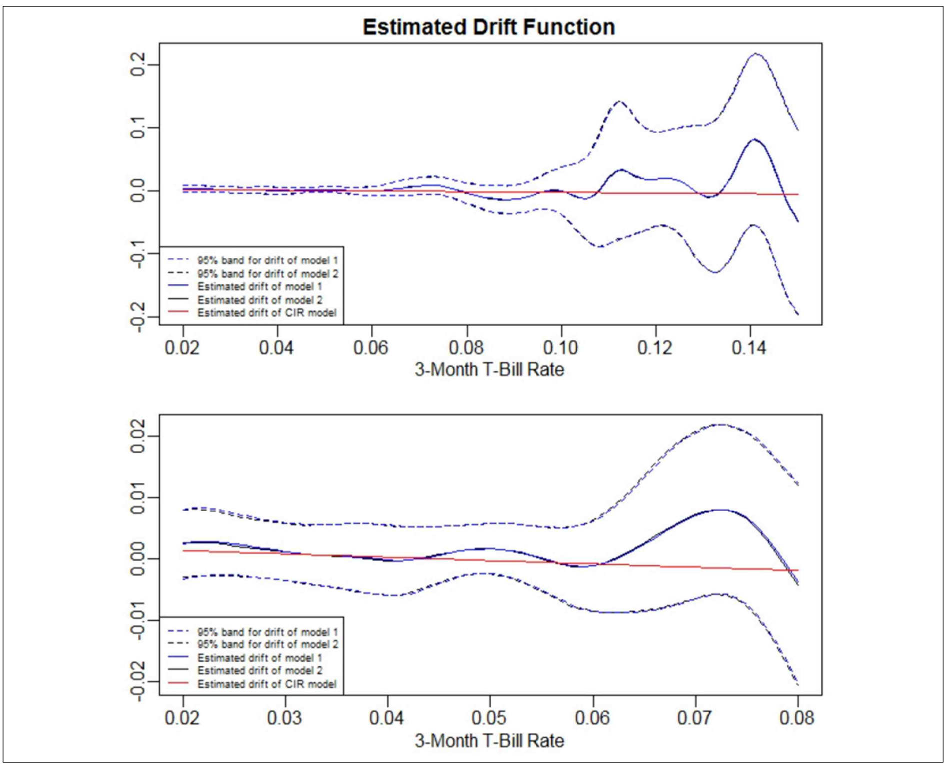
For Model 2, we employ local constant estimation and use the asymptotics developed in Ait-Sahalia and Park (2016) to estimate $\mu(x)$ and $\sigma^2(x)$ and to compute the 95% pointwise confidence intervals for these estimators.

Our theory is developed not only for a diffusion process but also for a general continuous time regression with functional coefficients. Therefore, the factors in the functional coefficients $\beta(\cdot)$ and $\gamma(\cdot)$ of Model 1 need not be limited to X_t; they can be any observable diffusion process. However, in our empirical exercise, the functional coefficient model (Model 1) is set to be a diffusion process so that we can compare three different specifications, namely, Models CIR, 1, and 2.

We employ the Gaussian kernel function in the nonparametric estimation of Models 1 and 2. Previous studies widely acknowledge that any reasonable kernel function yields almost optimal results (Stanton, 1997). However, for a given kernel function, the selection of an optimal bandwidth plays an important role in the performance of a nonparametric estimator (Fan and Zhang, 2003). In the previous section, we construct the optimal bandwidth for each of the proposed estimators, but the computation procedure involves an intensive estimation of not only functional coefficients but also their first- and second-order derivatives. Here, we utilize a simple rule of thumb for bandwidth selection proposed by Silverman (1986) as $h^* \approx 1.06\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of the sample, and n is the number of observations. This bandwidth minimizes the mean integrated squared error of the estimated density function.

Figures 4 and 5 report the estimated drift and diffusion functions from three different models. Upon observing these figures, we find that the nonparametric estimates of Models 1 and 2 closely resemble each other for the drift and diffusion functions. The drift term in the CIR model is specified as a linear mean-reverting function $\kappa(\mu - x)$ with $\kappa > 0$, indicating that the process X_t tends toward μ linearly. In Figure 4, the estimated CIR drift is represented by the red straight line with a negative slope. However, for Models 1 and 2, the nonparametric estimates of the drift do not exhibit linearity. Nevertheless, the estimated CIR drift falls within the 95% pointwise confidence bands for nonparametric estimators, indicating that we cannot definitively reject the linear specification of the CIR drift.

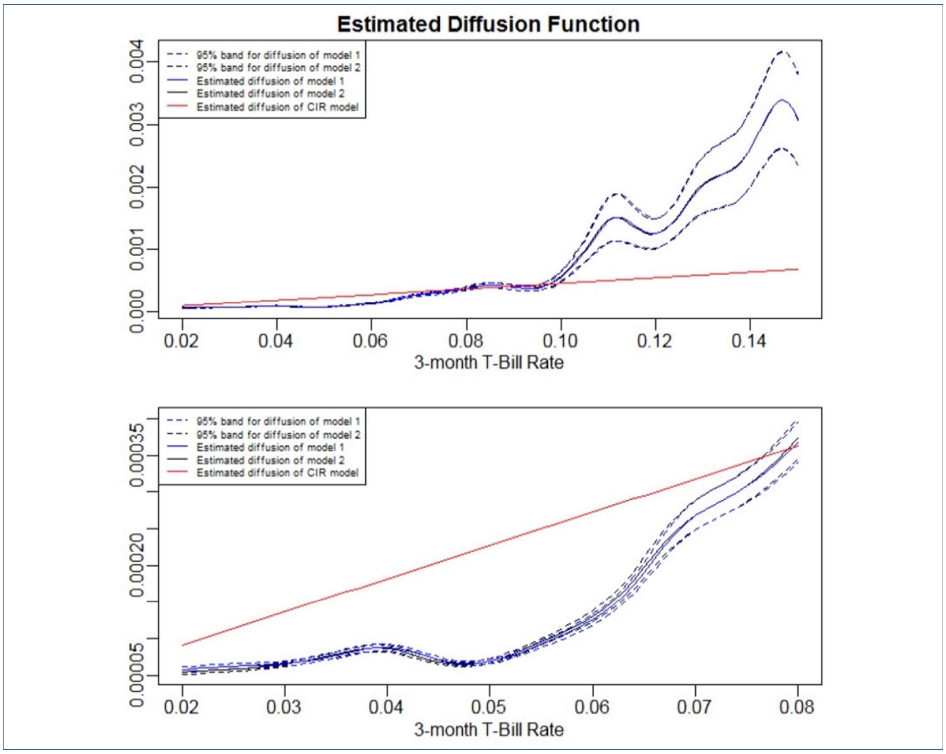
[Figure 4] Comparison of estimated drift functions from three models



This finding is not unexpected given that many diffusion models in the existing literature still assume linearity in the drift function, and previous empirical studies on the functional forms of drift have not decisively ascertained deviations from linearity as shown in Cai and Hong (2003). However, the reliability of the estimated drifts from Models 1 and 2 might be questioned due to the choice of optimal bandwidth. Here, we use a rule-of-thumb constant bandwidth instead of the formula for an optimal local bandwidth for each estimator derived in Section 3.

Furthermore, ensuring the consistency of the drift estimator at a particular spatial level x necessitates the process to visit that level an infinite number of times in the limit. The confidence bands that are computed using the asymptotic distribution may also yield suboptimal results. An alternative solution for constructing confidence bands is to utilize the bootstrap method, which is outside scope of our paper.

[Figure 5] Comparison of estimated diffusion functions from three models



Regarding the nonparametric estimates of the diffusion function, Figure 5 illustrates that they do not follow a linear form akin to the estimated CIR diffusion. Furthermore, the 95% confidence bands for the diffusion estimates do not encompass the estimated diffusion of the CIR model, which may offer grounds for rejecting the linear specification of the CIR diffusion function. However, this result might be unreliable due to the previously discussed limitations.

V. Conclusion

In this paper, we consider a continuous time regression model in which the conditional mean and variance functions involve functional coefficients. We propose the nonparametric estimators for these functional coefficients and provide their asymptotic properties by allowing $h \to_p 0$, $\Delta \to 0$, and $T \to \infty$ to change simultaneously under specific conditions. We find that the limiting distributions of these estimators are Gaussian under some regularity conditions. To ensure the consistency of the estimated functional coefficient $\beta(\cdot)$ of the conditional mean function, we need to satisfy $h \to_p 0$, $\Delta \to 0$, and $T \to \infty$. However, for the functional coefficient $\gamma^2(\cdot)$ of the conditional variance function being consistently estimated, we only need $h \to_p 0$ and $\Delta \to 0$, while T can be fixed. Additionally, the derived asymptotic properties facilitate the determination of optimal bandwidths for the estimators of the functional coefficients. The optimal bandwidth for each estimator can be computed utilizing an estimate for local time along with the estimates for the respective functional coefficients and their first- and second-order derivatives. Further details on the computation of optimal bandwidths may be addressed in subsequent studies.

We conduct an empirical analysis by employing the proposed nonparametric estimators to estimate the functional coefficients of the drift and diffusion terms using U.S. daily 3-month treasury bill data. When comparing the estimation results obtained from the CIR model and the fully nonparametric model, we observe strong similarities between our model and the fully nonparametric model. Additionally, based on the estimated drift and diffusion functions from the three models, we find no conclusive evidence to reject the linear specification of the CIR drift. However, we do have indications to reject the functional form of the CIR diffusion. These results have several limitations, such as the use of a rule of thumb for bandwidth, limited observations, and the construction of confidence bands using the asymptotics.

Appendix

Proof of Lemma 1. By successively applying Lemmas 12 and 6 in Ait-Sahalia and Park (2016) with $f = K$ and $g = x^2$, we have

$$\begin{aligned} Q_T(K) &= \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2 \\ &= \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right)X_t^2\,dt + O_p(h^{-2}\Delta l(T,x)) \\ &= x^2 l(T,x) + o_p(l(T,x)) + O_p(h^{-2}\Delta l(T,x)) \end{aligned}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. Moreover, by Assumption 5, we obtain

$$Q_T(K) = x^2 l(T,x)\,[1 + o_p(1)] \tag{A.1}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. We also have

$$\begin{aligned} N_T(K,1) = \frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2\beta(X_{(i-1)\Delta}) - x^2\beta(x)] \\ -\beta(x)\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2 - x^2] \end{aligned}$$

By Lemmas 12 and 9 in Ait-Sahalia and Park (2016), we have

$$\begin{aligned} &\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2\beta(X_{(i-1)\Delta}) - x^2\beta(x)] \\ &= \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right)[X_t^2\beta(X_t) - x^2\beta(x)]dt + O_p(h^{-2}\Delta l(T,x)) \\ &= \frac{h^2}{2}\iota(K_2)\left[(2\beta(x) + 4x\beta'(x) + \beta''(x)x^2) + (4x\beta(x) + 2\beta'(x)x^2)\frac{m'(x)}{m(x)}\right]l(T,x) \\ &+ o_p(h^2 l(T,x)) + O_p(h^{3/2} l(T,x)^{1/2}) + O_p(h^{-2}\Delta l(T,x)) \end{aligned}$$

and

$$\frac{\Delta}{h}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2 - x^2]$$

$$
\begin{aligned}
&= \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right)(X_t^2 - x^2)dt + O_p(h^{-2}\Delta l(T,x)) \\
&= \frac{h^2}{2}\iota(K_2)\left(2 + 4x\frac{m'(x)}{m(x)}\right)l(T,x) + o_p(h^2 l(T,x)) + O_p(h^{3/2} l(T,x)^{1/2}) \\
&\quad + O_p(h^{-2}\Delta l(T,x))
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
N_T(K,1) = \frac{h^2}{2}\iota(K_2)&\left[(4x\beta'(x) + \beta''(x)x^2) + 2\beta'(x)x^2\frac{m'(x)}{m(x)}\right]l(T,x) \\
&+ o_p(h^2 l(T,x)) + O_p(h^{3/2} l(T,x)^{1/2}) + O_p(h^{-2}\Delta l(T,x))
\end{aligned} \tag{A.2}
$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. Then, it follows from (A.1) and (A.2) that

$$
\hat{\beta}_p(x) - \beta(x) = \frac{h^2}{2}\iota(K_2)B_\beta(x) + o_p(h^2) + O_p(h^{3/2} l(T,x)^{-1/2})
$$

uniformly in T as $h \to 0$ and $\Delta \to 0$ under Assumption 5 (a).

We now consider $M_T(K,1)$. Let M_T be a continuous martingale defined as $M_T = \sqrt{h} M_T(K,1)$ so that

$$
M_T = \frac{1}{\sqrt{h}}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}\sigma(X_t)dW_t
$$

The quadratic variation $[M]$ of M at time T can be written as

$$
\begin{aligned}
[M]_T &= \frac{1}{h}\sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}\sigma^2(X_t)dt \\
&= \frac{\Delta}{h}\sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2\sigma^2(X_{(i-1)\Delta}) \\
&\quad + \frac{1}{h}\sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})]dt
\end{aligned} \tag{A.3}
$$

The first term in (A.3) satisfies

$$
\frac{\Delta}{h}\sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2\sigma^2(X_{(i-1)\Delta})
$$

$$
\begin{aligned}
&= \frac{1}{h}\int_0^T K^2\left(\frac{X_t - x}{h}\right) X_t^2 \sigma^2(X_t)dt + O_p(h^{-2}\Delta l(T,x)) \\
&= x^2\sigma^2(x)\iota(K^2)l(T,x)) + o_p(l(T,x)) + O_p(h^{-2}\Delta l(T,x))
\end{aligned}
$$

by applications of Lemmas 12 and 6 in Ait-Sahalia and Park (2016) with $f = K^2$ and $g = x^2\sigma^2(x)$. Due to Assumption 5, this term becomes $x^2\sigma^2(x)\iota(K^2)l(T,x) + o_p(l(T,x)) + o_p(h^2 l(T,x))$.

In addition, the second term of (A.3)

$$
\frac{1}{h}\sum_{i=1}^n K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2 \int_{(i-1)\Delta}^{i\Delta}[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})]dt = o_p(h^2 l(T,x))
$$

due to Lemma 11 in Ait-Sahalia and Park (2016). Therefore,

$$
[M]_T = x^2\sigma^2(x)\iota(K^2)l(T,x)[1 + o_p(1)] \tag{A.4}
$$

uniformly in T as $h \to 0$ and $\Delta \to 0$.

Meanwhile, we have

$$
\begin{aligned}
[W,M]_T &= \frac{1}{\sqrt{h}}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}\sigma(X_t)dt \\
&= \frac{\Delta}{\sqrt{h}}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\sigma(X_{(i-1)\Delta}) \\
&\quad + \frac{1}{\sqrt{h}}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}[\sigma(X_t) - \sigma(X_{(i-1)\Delta})]dt \\
&= \sqrt{h}\left[\frac{\Delta}{h}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\sigma(X_{(i-1)\Delta})\right. \\
&\quad \left. + \frac{1}{h}\sum_{i=1}^n K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}[\sigma(X_t) - \sigma(X_{(i-1)\Delta})]dt\right] \\
&= \sqrt{h}x\sigma(x)l(T,x)[1 + o_p(1)]
\end{aligned}
\tag{A.5}
$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. Therefore, it follows from (A.4) and (A.5) that

$$
\frac{[W,M]_T}{[M]_T} = \frac{\sqrt{h}x\sigma(x)l(T,x)[1 + o_p(1)]}{x^2\sigma^2(x)\iota(K^2)l(T,x)[1 + o_p(1)]} = o_p(h^{1/2}) \tag{A.6}
$$

and

$$\frac{[W,M]_T}{[W]_T} = \frac{\sqrt{h}x\sigma(x)l(T,x)[1+o_p(1)]}{T} = o_p(h^{1/2}T^{-1}) \tag{A.7}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. From (A.4), (A.6), and (A.7), we can deduce that

$$l(T,x)^{-1/2} M_T \to_d x\sigma(x)\iota(K^2)^{1/2} Z, \tag{A.8}$$

where Z is a standard normal random variate independent of $l(T,x)$. Furthermore, given that $M_T = \sqrt{h}M_T(K,1)$, we have

$$\begin{aligned}
[hl(T,x)]^{1/2}\hat{\beta}_q(x) &= \frac{l(T,x)^{1/2}}{Q_T(K)}\sqrt{h}M_T(K,1) \\
&= l(T,x)^{1/2}\frac{M_T}{Q_T(K)} \\
&= \frac{l(T,x)^{1/2}M_T}{x^2 l(T,x)[1+o_p(1)]} \\
&= x^{-2}l(T,x)^{-1/2}M_T[1+o_p(1)] \to_d x^{-1}\sigma(x)\iota(K^2)^{1/2}Z
\end{aligned}$$

or

$$[hl(T,x)]^{1/2}\hat{\beta}_q(x) \to_d x^{-1/2}\gamma(x)\iota(K^2)^{1/2}Z \tag{A.9}$$

For $\hat{\varepsilon}_\beta(x)$, by applying Lemma 11 in Ait-Sahalia and Park (2016) under Assumptions 5 (a) and (b), we have

$$R_T(K,1) = o_p(h^2 l(T,x))$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. Consequently,

$$\hat{\varepsilon}_\beta(x) = \frac{o_p(h^2 l(T,x))}{x^2 l(T,x)[1+o_p(1)]} = o_p(h^2) \tag{A.10}$$

The proof of Lemma 1 is thus complete. □

Proof of Proposition 1. Proposition 1 can be easily deduced from Lemma 1, so the details are omitted. □

Proof of Lemma 2. We have

$$N_T(K,2)=\frac{\Delta}{h}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2\gamma^2(X_{(i-1)\Delta})-x^2\gamma^2(x)]$$
$$-\gamma^2(x)\frac{\Delta}{h}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2-x^2]$$

By Lemmas 12 and 9 in Ait-Sahalia and Park (2016), we have

$$\frac{\Delta}{h}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2\gamma^2(X_{(i-1)\Delta})-x^2\gamma^2(x)]$$
$$=\frac{1}{h}\int_0^T K\left(\frac{X_t-x}{h}\right)[X_t^2\gamma^2(X_t)-x^2\gamma^2(x)]dt+O_p(h^{-2}\Delta l(T,x))$$
$$=\frac{h^2}{2}\iota(K_2)\left[(2\gamma^2(x)+4x\gamma^{2\prime}(x)+\gamma^{2\prime\prime}(x)x^2)+(4x\gamma^2(x)+2\gamma^{2\prime}(x)x^2)\frac{m'(x)}{m(x)}\right]l(T,x)$$
$$+o_p(h^2l(T,x))+O_p(h^{3/2}l(T,x)^{1/2})+O_p(h^{-2}\Delta l(T,x))$$

and

$$\frac{\Delta}{h}\sum_{i=1}^{n}K\left(\frac{X_{(i-1)\Delta}-x}{h}\right)[X_{(i-1)\Delta}^2-x^2]$$
$$=\frac{1}{h}\int_0^T K\left(\frac{X_t-x}{h}\right)(X_t^2-x^2)dt+O_p(h^{-2}\Delta l(T,x))$$
$$=\frac{h^2}{2}\iota(K_2)\left(2+4x\frac{m'(x)}{m(x)}\right)l(T,x)+o_p(h^2l(T,x))+O_p(h^{3/2}l(T,x)^{1/2})$$
$$+O_p(h^{-2}\Delta l(T,x))$$

Therefore,

$$N_T(K,2)=\frac{h^2}{2}\iota(K_2)\left[(4x\gamma^{2\prime}(x)+\gamma^{2\prime\prime}(x)x^2)+2\gamma^{2\prime}(x)x^2\frac{m'(x)}{m(x)}\right]l(T,x)$$
$$+o_p(h^2l(T,x))+O_p(h^{3/2}l(T,x)^{1/2})+O_p(h^{-2}\Delta l(T,x)) \qquad \text{(B.1)}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$. Then, it follows from (B.1) and (A.1) that

$$\hat{\gamma}_p^2(x) - \gamma^2(x) = \frac{h^2}{2}\left(4x^{-1}\gamma^{2\prime}(x) + \gamma^{2\prime\prime}(x)x^2) + 2\gamma^{2\prime}(x)\frac{m'(x)}{m(x)}\right)\iota_2(K)$$
$$+o_p(h^2) + O_p(h^{3/2}l(T,x)^{-1/2})$$

We now move to the second part $\hat{\gamma}_q^2(x)$. We define M as a continuous martingale such that

$$M_T = \sqrt{\frac{2}{h\Delta}}\sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})\sigma(X_t)dW_t$$

Then quadratic variation $[M]$ at T is given by

$$[M]_T = \frac{2}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta} - x}{h}\right) X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2\sigma^2(X_t)dt \tag{B.2}$$

We decompose

$$\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2\sigma^2(X_t)dt = \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2 dt$$
$$+\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})]dt \tag{B.3}$$

and use Itô's formula to deduce that

$$\int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2 dt = \frac{\Delta^2}{2}\sigma^2(X_{(i-1)\Delta})$$
$$+2\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)(Y_t - Y_{(i-1)\Delta})\mu(X_t)dt$$
$$+\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)[\sigma^2(X_t) - \sigma^2(X_{(i-1)\Delta})]dt$$
$$+2\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)(Y_t - Y_{(i-1)\Delta})\sigma(X_t)dW_t \tag{B.4}$$

We then obtain from (B.2), (B.3), and (B.4) that

$$[M]_T = A_T + B_T + C_T + D_T + E_T \tag{B.5}$$

where

$$A_T = \frac{\Delta}{h}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^4(X_{(i-1)\Delta})$$

$$B_T = \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)(Y_t - Y_{(i-1)\Delta})\mu(X_t)dt$$

$$C_T = \frac{2}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)[\sigma^2(X_t)-\sigma^2(X_{(i-1)\Delta})]dt$$

$$D_T = \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)(Y_t - Y_{(i-1)\Delta})\sigma(X_t)dW_t$$

$$E_T = \frac{2}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \int_{(i-1)\Delta}^{i\Delta}(Y_t - Y_{(i-1)\Delta})^2[\sigma^2(X_t)-\sigma^2(X_{(i-1)\Delta})]dt$$

For the first term A_T, we successively apply Lemmas 12 and 6 in Ait-Sahalia and Park (2016) and obtain

$$\begin{aligned} A_T &= \frac{1}{h}\int_0^T K^2\left(\frac{X_t - x}{h}\right) X_t^2\sigma^4(X_t)dt + O_p(h^{-2}\Delta l(T,x)) \\ &= x^2\sigma^4(x)\iota(K^2)l(T,x) + o_p(l(T,x)) + O_p(h^{-2}\Delta l(T,x)) \\ &= x^2\sigma^4(x)\iota(K^2)l(T,x) + o_p(l(T,x)) + o_p(h^2 l(T,x)) \\ &= x^2\sigma^4(x)\iota(K^2)l(T,x)[1+o_p(1)] \end{aligned} \tag{B.6}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$ under Assumption 6. For B_T and D_T, we apply Lemma 14

$$\begin{aligned} B_T &= \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)(Y_t - Y_{(i-1)\Delta})\mu(X_t)dt \\ &= \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)^2\mu^2(X_t)dt \\ &\quad + \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta - t)^2\mu(X_t)\sigma(X_t)dW_t \\ &= o_p(l(T,x)) \end{aligned}$$

under Assumptions 6 (a) and (c).

Similarly, we have

$$\begin{aligned} D_T &= \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)(Y_t-Y_{(i-1)\Delta})\sigma(X_t)dW_t \\ &= \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)^2\mu(X_t)\sigma(X_t)dW_t \\ &\quad + \frac{4}{h\Delta}\sum_{i=1}^{n} K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta}^2 \sigma^2(X_{(i-1)\Delta})\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)^2\sigma^2(X_t)dW_t \\ &= o_p(l(T,x)) \end{aligned}$$

In sum,

$$B_T, D_T = o_p(l(T,x)) \tag{B.7}$$

It is also straightforward to obtain

$$C_T = o_p(l(T,x)) \tag{B.8}$$

from Lemma 14 in Ait-Sahalia and Park (2016) under Assumptions 6 (a) and (b).

For the last term E_T, we note that

$$|\sigma^2(X_t)-\sigma^2(X_{(i-1)\Delta})| \leqslant T(\sigma^{2\prime})\,|X_t - X_{(i-1)\Delta}| \tag{B.9}$$

and that

$$\begin{aligned} |X_t - X_{(i-1)\Delta}| &= \left|\int_{(i-1)\Delta}^{t} a(X_s)ds + \int_{(i-1)\Delta}^{t} b(X_s)dV_s\right| \\ &\leqslant \left|\int_{(i-1)\Delta}^{t} a(X_s)ds\right| + \left|\int_{(i-1)\Delta}^{t} b(X_s)dV_s\right| \end{aligned} \tag{B.10}$$

for all $1 \leqslant i \leqslant n$ and $t \in [(i-1)\Delta, i\Delta]$. We also have

$$\left|\int_{(i-1)\Delta}^{t} a(X_s)ds\right| \leqslant \int_{(i-1)\Delta}^{t} |a(X_s)|\,ds \leqslant \int_{(i-1)\Delta}^{i\Delta} T(a)bs = \Delta T(a) \tag{B.11}$$

and

$$\int_{(i-1)\Delta}^{t} b(X_s)dV_s = O_p\left(\Delta^{1/2}T(b)\sqrt{\log(T/\Delta)}\right) \tag{B.12}$$

uniformly for all $1 \leqslant i \leqslant n$ and $t \in [(i-1)\Delta, i\Delta]$ by the modulus of continuity for a continuous martingale (see Lemma B.2 of Kim and Park, 2017).

Combining (B.9), (B.10), (B.11), and (B.12), we have

$$|\sigma^2(X_t)-\sigma^2(X_{(i-1)\Delta})|=o_p(1) \tag{B.13}$$

uniformly for all $1\leqslant i\leqslant n$ and $t\in[(i-1)\Delta,i\Delta]$ under the condition of Assumption 6 (d).

By (B.13), we can deduce that

$$E_T=o_p\left(\frac{1}{h\Delta}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(Y_t-Y_{(i-1)\Delta})^2\,dt\right)$$

We then use (B.4) to write

$$\begin{aligned}G&=\frac{1}{h\Delta}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(Y_t-Y_{(i-1)\Delta})^2\,dt\\&=\frac{\Delta}{2h}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\sigma^2(X_{(i-1)\Delta})\\&\quad+\frac{2}{h\Delta}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)(Y_t-Y_{(i-1)\Delta})\mu(X_t)dt\\&\quad+\frac{1}{h\Delta}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)[\sigma^2(X_t)-\sigma^2(X_{(i-1)\Delta})]dt\\&\quad+\frac{2}{h\Delta}\sum_{i=1}^{n}K^2\left(\frac{X_{(i-1)\Delta}-x}{h}\right)X_{(i-1)\Delta}^2\int_{(i-1)\Delta}^{i\Delta}(i\Delta-t)(Y_t-Y_{(i-1)\Delta})\sigma(X_t)dW_t\end{aligned}$$

By successively applying Lemmas 12, 6, 15, and 14 in Ait-Sahalia and Park (2016) as in (B.6), (B.7), and (B.8), we may obtain

$$E_T=o_p(l(T,x)) \tag{B.14}$$

uniformly in T as $h\to 0$ and $\Delta\to 0$. Thus far, we already have all components of $[M]_T$ from (B.6), (B.7), (B.8), and (B.14). Therefore,

$$[M]_T=x^2\sigma^4(x)\iota(K^2)l(T,x)[1+o_p(1)] \tag{B.15}$$

uniformly in T as $h\to 0$ and $\Delta\to 0$.

We now consider $[W,M]_T$. We have

$$
\begin{aligned}
[W,M]_T &= \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta} \int_{(i-1)\Delta}^{i\Delta} (Y_t - Y_{(i-1)\Delta})\sigma(X_t)dt \\
&= \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta} \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)\mu(X_t)\sigma(X_t)dt \\
&\quad + \sqrt{\frac{2}{h\Delta}} \sum_{i=1}^{n} K\left(\frac{X_{(i-1)\Delta}-x}{h}\right) X_{(i-1)\Delta} \int_{(i-1)\Delta}^{i\Delta} (i\Delta - t)\sigma^2(X_t)dW_t
\end{aligned}
$$

Due to Lemma 13 in Ait-Sahalia and Park (2016), under Assumption 6 (c), it follows that

$$
[W,M]_T = o_p(h^{1/2} l(T,x)) \tag{B.16}
$$

We then deduce from (B.15) and (B.16) that

$$
\frac{[W,M]_T}{[M]_T} = \frac{o_p(h^{1/2} l(T,x))}{x^2\sigma^2(x)\iota(K^2) l(T,x)[1+o_p(1)]} = o_p(h^{1/2})
$$

and

$$
\frac{[W,M]_T}{[W]_T} = \frac{o_p(h^{1/2} l(T,x))}{T} = o_p(h^{1/2}T^{-1})
$$

Together with (B.15), we can deduce that

$$
l(T,x)^{-1/2} M_T \to_d x\sigma^2(x)\iota(K^2)^{1/2} Z \tag{B.17}
$$

where Z is a standard normal random variate independent of $l(T,x)$. Therefore,

$$
\begin{aligned}
\left[\frac{hl(T,x)}{\Delta}\right]^{1/2} \hat{\gamma}_q^2(x) &= \left[\frac{hl(T,x)}{\Delta}\right]^{1/2} \frac{2M_T(K,2)}{Q_T(K)} = \frac{\sqrt{2}l(T,x)^{1/2}}{Q_T(K)} \sqrt{\frac{2h}{\Delta}} M_T(K,2) \\
&= \frac{\sqrt{2}l(T,x)^{1/2} M_T}{x^2 l(T,x)[1+o_p(1)]} = \sqrt{2}x^{-2} l(T,x)^{-1/2} M_T [1+o_p(1)] \\
&\to_p \sqrt{2}x^{-1}\sigma^2(x)\iota(K^2)^{1/2} Z
\end{aligned} \tag{B.18}
$$

For the last part $\hat{\varepsilon}_{\gamma^2}(x)$, we consider $R_T(K,2)$ and $S_T(K)$ in sequel. We have

$$R_T(K,2) = o_p(h^2 l(T,x)) \tag{B.19}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$ due to Lemma 11 in Ait-Sahalia and Park (2016) under Assumptions 6 (a) and (b), and we obtain

$$S_T(K) = o_p(h^2 l(T,x)) \tag{B.20}$$

uniformly in T as $h \to 0$ and $\Delta \to 0$ due to Lemma 13 in Ait-Sahalia and Park (2016) under Assumptions 6 (a) and (c). By combining (A.1), (B.19), and (B.20), we obtain

$$\hat{\varepsilon}_{\gamma^2}(x) = \frac{R_T(K,2) + 2S_T(K)}{Q_T(K)} = \frac{o_p(h^2 l(T,x))}{x^2 l(T,x)[1+o_p(1)]} = o_p(h^2) \tag{B.21}$$

The proof is therefore complete.

□

Proof of Proposition 2. It is straightforward to deduce Proposition 2 from Lemma 2, so the details are omitted. □

References

Ait-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivatives," *Econometrica*, 64, 527–560.

Ait-Sahalia, Y., and J. Y. Park (2016), "Bandwidth Selection and Asymptotic Properties of Local Nonparametric Estimators in Possibly Nonstationary Continuous-time Models," *Journal of Econometrics*, 192(1), 119–138.

Bandi, F., and P. C. Phillips (2009), "Nonstationary Continuous-time Processes," *Handbook of Financial Econometrics*, 1, 139–202.

____________________ (2003), "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71(1), 241–283.

Bu, R., J. Kim, and B. Wang (2023), "Uniform and Lp Convergences for Nonparametric Continuous Time Regressions with Ssemiparametric Application," *Journal of Econometrics*, 235(2), 1934–1954.

Cai, Z., J. Fan, and Q. Yao (2000), "Functional-coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95(451), 941–956.

Cai, Z., and Y. Hong (2003), *Nonparametric Methods in Continuous-time Finance: A Selective Review*, Elsevier.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992), "An Empirical Comparison of Alternative Models of the Short-term Interest Rate," *The Journal of Finance*, 47(3), 1209–1227.

Chen, R., and R. S. Tsay (1993), "Functional-coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88(421), 298–308.

Chib, S., F. Nardari, and N. Shephard (2002), "Markov Chain Monte Carlo Methods for Stochastic Volatility Models," *Journal of Econometrics*, 108(2), 281–316.

Choi, Y., S. Jacewitz, and J. Y. Park (2016), "A Reexamination of Stock Return Predictability," *Journal of Econometrics*, 192(1), 168–189.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53(2), 385–407.

Duffie, D., and K. J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61(4), 929–952.

Fan, J. (2005), "A Selective Overview of Nonparametric Methods in Financial Econometrics," *Statistical Science*, 20(4), 317–337.

Fan, J., and C. Zhang (2003), "A Reexamination of Diffusion Estimators with Applications to Financial Model Validation," *Journal of the American Statistical Association*, 98(461), 118–134.

Florens-Zmirou, D. (1993), "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30(4), 790–804.

Haggan, V., and T. Ozaki (1981), "Modelling Nonlinear Random Vibrations using an Amplitude-dependent Autoregressive Time Series Model," *Biometrika*, 68(1), 189–196.

Han, H., and E. Lee (2020), "Triple Regime Stochastic Volatility Model with Threshold and Leverage Effect," *The Korean Economic Review*, 36(2), 481–509.

Han, H., and N. K. Lee (2018), "Modeling the Dynamics between Stock Price and Dividend: An Endogenous Regime Switching Approach," *The Korean Economic Review*, 34(2), 213–235.

Ibragimov, R., J. Kim, and A. Skrobotov (2023), "New Robust Inference for Predictive Regressions," *Econometric Theory*, forthcoming.

Jiang, G. J., and J. L. Knight (1997), "A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-term Interest Rate Model," *Econometric Theory*, 13(5), 615–645.

Karatzas, I., and E. Shreve (1991), *Brownian Motion and Stochastic Calculus*, Springer New York, NY.

Kim, J., and N. Meddahi (2020), "Volatility Regressions with Fat Tails," *Journal of Econometrics*, 218(2), 690–713.

Kim, J., and J. Y. Park (2017), "Asymptotics for Recurrent Diffusions with Application to High Frequency Regression," *Journal of Econometrics*, 196(1), 37–54.

Kim, J., J. Y. Park, and B. Wang (2021), "Estimation of Volatility Functions in Jump Diffusions using Truncated Bipower Increments," *Econometric Theory*, 37(5), 926–958.

Lo, A. W. (1988), "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4(2), 231–247.

Nicholls, D. F., and B. G. Quinn (1982), *Random Coefficient Autoregressive Models an Introduction*, Springer New York, NY.

Pearson, N. D., and T.-S. Sun (1994), "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross model," *The Journal of Finance*, 49(4), 1279–1304.

Phillips, P. C. (2015), "Pitfalls and Possibilities in Predictive Regression," *Cowles Foundation Discussion Papers*, 2438.

Revuz, D., and M. Yor (1999), *Continuous Martingales and Brownian Motion*, Springer Berlin, Heidelberg.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

Stambaugh, R. F. (1999), "Predictive Regressions," *Journal of Financial Economics*, 54(3), 375–421.

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *The Journal of Finance*, 52(5), 1973–2002.

Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5(2), 177–188.

Zhang, T., and W. B. Wu (2012), "Inference of Time-varying Regression Models," *The Annals of Statistics*, 40(3).

함수적 계수를 가진 연속시간 회귀모형의 비모수적 추정법*

최 미 정** · 김 지 현*** · Nuong Nguyen****

초 록 본 연구에서는 연속시간 회귀모형의 조건부 기대값과 조건부 분산이 함수적 계수를 가지는 모형을 고려하고 있다. 이때 함수적 계수와 설명변수는 정상성 혹은 비정상성을 가질 수 있는 일반적인 확산과정(diffusion process)을 따른다. 본 연구에서 고려되고 있는 연속시간 회귀모형의 추정을 위해 이산적으로 수집된 자료를 사용하여 커널 방법에 기반한 비모수적 방법을 제시하고 있다. 일반적인 상황 하에서 제안된 추정량의 일치성과 극한 분포를 도출하였고 해당 회귀모형과 추정방법을 활용하여 미국의 단기 금리 모형의 추정에 활용하였다.

핵심 주제어: 연속시간 회귀모형, 확산과정, 비정상 시계열, 함수적 계수, 비모수적 추정
경제학문헌목록 주제분류: C14, C22, C58

투고 일자: 2023. 12. 29. 심사 및 수정 일자: 2024. 7. 29. 게재 확정 일자: 2024. 9. 25.
* 논문에 대한 유익한 조언을 해준 Nour Meddahi 교수와 Stéphane Gregoir 교수, 그리고 익명의 두 심사자께 깊은 감사를 표한다.
** 제1저자, 성균관대학교 BK21 박사후연구원; 산업연구원 부연구위원, e-mail: mijchoi@kiet.re.kr
*** 교신저자, 성균관대학교 경제학부 교수, e-mail: kim.jihyun@skku.edu
**** 공동저자, University of Kentucky 경제학과 박사과정, e-mail: nuong.nguyen@uky.edu