



# A new method for estimating Sharpe ratio function via local maximum likelihood

Wenchao Xu<sup>a</sup>, Hongmei Lin<sup>b</sup>, Tiejun Tong<sup>c</sup> and Riquan Zhang<sup>d\*</sup>

<sup>a</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, People's Republic of China; <sup>b</sup>School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, People's Republic of China; <sup>c</sup>Department of Mathematics, Hong Kong Baptist University, Hong Kong, People's Republic of China; <sup>d</sup>Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, and School of Statistics, East China Normal University, Shanghai, People's Republic of China

## ABSTRACT

The Sharpe ratio function is a commonly used risk/return measure in financial econometrics. To estimate this function, most existing methods take a two-step procedure that first estimates the mean and volatility functions separately and then applies the plug-in method. In this paper, we propose a direct method via local maximum likelihood to simultaneously estimate the Sharpe ratio function and the negative log-volatility function as well as their derivatives. We establish the joint limiting distribution of the proposed estimators, and moreover extend the proposed method to estimate the multivariate Sharpe ratio function. We also evaluate the numerical performance of the proposed estimators through simulation studies, and compare them with existing methods. Finally, we apply the proposed method to the three-month US Treasury bill data and that captures a well-known covariate-dependent effect on the Sharpe ratio.

## ARTICLE HISTORY

Received 8 November 2021  
Accepted 13 August 2022

## KEYWORDS

Direct method;  
heteroscedastic  
non-parametric regression;  
joint limiting distribution;  
local polynomial smoothing;  
Sharpe ratio function

## 1. Introduction

In financial analysis, the Sharpe ratio is one of the most popular measures about the risk-adjusted return, which is defined as the difference between the return of an investment and the risk-free return, divided by the volatility (or standard deviation) of the investment. The Sharpe ratio is now commonly used as a gold standard to compare different assets or trading strategies, where the one with a higher Sharpe ratio provides a better return for the same risk. Moreover, it has also been extended to many other contexts, including performance attribution, tests of market efficiency, and risk management [18,21]. On the other hand, however, a static Sharpe ratio with a constant standard deviation may oversimplify the risk due to the serial correlation or the phases of business cycle [18]. This thus motivates to consider the covariate-dependent Sharpe ratio function, also referred to as the Sharpe

**CONTACT** Hongmei Lin  [xuemeijiyi@163.com](mailto:xuemeijiyi@163.com)

\*Present address: School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, People's Republic of China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/02664763.2022.2114431>.

ratio function, which can provide extra evidence to the fundamental economics underlying the economy and asset pricing; see [16,17,19,25].

Let  $(Y, X)$  be a two-dimensional random vector. The Sharpe ratio function is defined as  $f(x) = \mu(x)/\sigma(x)$ , where  $\mu(x) = \mathbb{E}(Y | X = x)$  and  $\sigma^2(x) = \text{Var}(Y | X = x) > 0$  are the conditional mean and variance functions, respectively. More specifically, we consider the heteroscedastic non-parametric model

$$Y = \mu(X) + \sigma(X)\varepsilon, \quad (1)$$

where  $\varepsilon$  is the random error satisfying  $\mathbb{E}(\varepsilon | X) = 0$  and  $\text{Var}(\varepsilon | X) = 1$ . We further assume that  $X$  and  $\varepsilon$  are independent of each other. Model (1) has been extensively studied in the literature; see, for example, [10,26,29,31] and the references therein. The previous literature mainly focuses on the estimation of  $\mu(x)$  or  $\sigma^2(x)$  separately. Common methods for estimating  $\mu(x)$  include, for example, local polynomial regression [10], polynomial spline regression [7,15], and smoothing splines [26,29]. While for the estimation of  $\sigma^2(x)$ , there are three main approaches: the direct method [14], the residual-based methods [8,11,30,32], and the difference-based methods [3,9,22,28].

Model (1) is closely related to the following continuous-time diffusion process model:

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t, \quad (2)$$

where  $\mu(\cdot)$  and  $\sigma(\cdot)$  are the drift and diffusion functions of the process  $\{X_t\}$  respectively, and  $W_t$  is the standard Brownian motion that is used to model the stochastic behavior of economic variables, including, for example, interest rates, exchange rates, and stock prices; see [1,2,23]. By employing a discrete-time approximation to the continuous-time process, model (2) will reduce to model (1). In Section 7, we analyze the three-month US Treasury bill data from the secondary market. The three-month US Treasury bill rate is the yield received for investing in a government-issued treasury security that has a maturity of three months. The three-month Treasury yield is included on the shorter end of the yield curve and is important when looking at the overall US economy. The secondary market rates are annualized using a 360-day year of bank interest and quoted on a discount basis. Specifically, the rates are calculated as unweighted averages of closing bid rates quoted by at least five dealers in the secondary market. The rates are posted on a bank discount basis, but are converted into continuously compounded yields prior to analysis. Andersen and Lund [2] applied model (2) to analyze this dataset from 5 January 1962 to 31 March 1995, where  $\mu$  and  $\sigma$  are, respectively, the instantaneous expected rate of return and the volatility. We focus on estimating the covariate-dependent Sharpe ratio  $f = \mu/\sigma$  in Section 7.

Suppose that  $\{(Y_i, X_i) : i = 1, \dots, n\}$  are independent and identically distributed (i.i.d.) data from model (1). To estimate the Sharpe ratio function  $f(x)$ , most existing methods take a two-step procedure that first estimates the mean function  $\mu(x)$  and volatility function  $\sigma^2(x)$  separately and then applies the plug-in method. Nevertheless, such a two-step procedure is often less efficient since two smoothing parameters have to be involved for estimating  $\mu(x)$  and  $\sigma^2(x)$  individually. In addition, when one is also interested to estimate the first- or higher-order derivative of  $f(x)$ , the indirect methods will often be difficult to implement. Recently, [16] proposed a direct maximum likelihood estimation with a roughness penalty for the Sharpe ratio function based on a parameterization of the likelihood in terms of  $f(x) = \mu(x)/\sigma(x)$  and the inverse volatility function  $1/\sigma(x)$  when the random

error  $\varepsilon$  is normally distributed. Moreover, by reparameterizing the volatility function as  $g(x) = -\log\{\sigma(x)\}$ , [17] proposed to estimate  $f(x)$  and  $g(x)$  iteratively based on the local linear method.

In this paper, we propose a general framework using the local maximum likelihood to jointly estimate  $f(x)$  and  $g(x)$  as well as their derivatives. Compared to [17], our new framework has several advantages. First, our joint estimation only needs one bandwidth for simultaneously estimating  $f$  and  $g$ . Second, our proposed method can also be used to estimate the derivatives of the Sharpe ratio function and apply to the non-Gaussian distribution of  $\varepsilon$  as well. Third, by establishing the joint limiting distribution for the proposed estimators, our new framework can also optimally estimate the conditional mean function  $\mu(x)$  and the conditional variance function  $\sigma^2(x)$ . In particular, the leading terms on the asymptotic bias and variance of our new estimator for  $\sigma^2(x)$  will be the same as those in [30].

The rest of the paper is organized as follows. Section 2 describes the estimation procedure of  $f(x)$  and  $g(x)$  or their derivatives by combining the maximum likelihood estimation and the local polynomial smoothing. Section 3 establishes the joint limiting distribution for the local maximum likelihood estimators induced by the normally distributed random errors. Section 4 proposes new estimators for  $\mu(x)$  and  $\sigma^2(x)$  and derives their joint limiting distribution. Section 5 extends the proposed method to estimate the multivariate Sharpe ratio function. Section 6 conducts extensive simulations to assess the finite sample performance of the proposed estimators. We will also mention the rule of thumb and the leave-one-out cross-validation for the bandwidth selection. Section 7 analyzes a three-month US Treasury bill data using our new method. Section 8 concludes the paper, and the technical results, including the proofs of theorems, are given in the online supplemental material.

## 2. Estimation method

In this section, we propose to estimate the Sharpe ratio function  $f(x)$  and the negative log-volatility function  $g(x)$  simultaneously, where

$$f(x) = \mu(x)/\sigma(x) \quad \text{and} \quad g(x) = -\log\{\sigma(x)\}. \quad (3)$$

Note that the logarithm transformation in  $g(x)$  removes the positivity constraint on the volatility function, as also adopted by [17,30,32]. Moreover, unlike most existing literature, we do not impose the normality assumption on the random errors so as to be more realistic.

For simplicity, let the density function of  $\varepsilon$  be  $\psi(\cdot)$ . Then accordingly, the density function of  $\mu_\varepsilon + \sigma_\varepsilon\varepsilon$  is  $\sigma_\varepsilon^{-1}\psi(\sigma_\varepsilon^{-1}(\cdot - \mu_\varepsilon))$  for any  $\mu_\varepsilon \in \mathbb{R}$  and  $\sigma_\varepsilon > 0$ , and by which the conditional density function of  $Y$  given  $X = x$  is

$$\frac{1}{\sigma(x)}\psi\left\{\frac{Y - \mu(x)}{\sigma(x)}\right\} = \exp\{g(x)\}\psi(Y\exp\{g(x)\} - f(x)),$$

where the right-hand side is reparameterized with  $f$  and  $g$  defined in (3). Moreover, the log-likelihood function of the sample  $\{(Y_i, X_i) : i = 1, \dots, n\}$  is given by

$$\ell_n^*(f, g) = \sum_{i=1}^n [\log \psi(Y_i \exp\{g(X_i)\} - f(X_i)) + g(X_i)]. \quad (4)$$

For each given point  $x$ , we use the local polynomial smoothing [10] to estimate the  $\nu$ th derivatives  $f^{(\nu)}(x)$  and  $g^{(\nu)}(x)$ . Suppose that the  $(p + 1)$ th derivatives of  $f$  and  $g$  exist at point  $x$ , where  $p \geq \nu$ . Then for any  $X_i$  in a neighborhood of  $x$ , we can approximate  $f(X_i)$  and  $g(X_i)$  locally by

$$f(X_i) \approx \sum_{l=0}^p \frac{f^{(l)}(x)}{l!} (X_i - x)^l \quad \text{and} \quad g(X_i) \approx \sum_{l=0}^p \frac{g^{(l)}(x)}{l!} (X_i - x)^l.$$

By the above approximations, together with Equation (4), leads to a local maximum likelihood problem that minimizes

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n [-\log \psi \{Y_i \exp(\mathbf{Z}_i^\top \boldsymbol{\gamma}) - \mathbf{Z}_i^\top \boldsymbol{\beta}\} - \mathbf{Z}_i^\top \boldsymbol{\gamma}] K_h(X_i - x) \quad (5)$$

with respect to  $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^{p+1}$ , where  $\boldsymbol{\beta} = (f(x), hf'(x), \dots, h^p f^{(p)}(x)/p!)^\top$ ,  $\boldsymbol{\gamma} = (g(x), hg'(x), \dots, h^p g^{(p)}(x)/p!)^\top$ ,  $K(\cdot)$  is a kernel function,  $K_h(\cdot) = K(\cdot/h)/h$  with  $h > 0$  as the bandwidth, and

$$\mathbf{Z}_i = \left\{ 1, \frac{X_i - x}{h}, \dots, \left( \frac{X_i - x}{h} \right)^p \right\}^\top.$$

Let also  $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \arg \min_{\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^{p+1}} \ell_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ . It then yields the local polynomial estimators of  $f^{(\nu)}(x)$  and  $g^{(\nu)}(x)$  as

$$\hat{f}^{(\nu)}(x) = h^{-\nu} \nu! \mathbf{e}_{\nu+1}^\top \hat{\boldsymbol{\beta}} \quad \text{and} \quad \hat{g}^{(\nu)}(x) = h^{-\nu} \nu! \mathbf{e}_{\nu+1}^\top \hat{\boldsymbol{\gamma}},$$

where  $\mathbf{e}_{\nu+1}$  is the standard unit vector with 1 in the  $(\nu + 1)$ th component and 0 elsewhere.

For illustration, we also provide two examples with the error distribution given, as well as derive their negative local log-likelihood functions.

**Example 2.1:** When  $\varepsilon$  follows the standard normal distribution, by ignoring additive constants, we have

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \left[ \frac{1}{2} \{Y_i \exp(\mathbf{Z}_i^\top \boldsymbol{\gamma}) - \mathbf{Z}_i^\top \boldsymbol{\beta}\}^2 - \mathbf{Z}_i^\top \boldsymbol{\gamma} \right] K_h(X_i - x). \quad (6)$$

**Example 2.2:** When  $\varepsilon$  follows the Laplace distribution, i.e.  $\psi(\varepsilon) = 2^{-1/2} \exp(-\sqrt{2}|\varepsilon|)$ , by ignoring additive constants we have

$$\ell_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^n \{ \sqrt{2} |Y_i \exp(\mathbf{Z}_i^\top \boldsymbol{\gamma}) - \mathbf{Z}_i^\top \boldsymbol{\beta}| - \mathbf{Z}_i^\top \boldsymbol{\gamma} \} K_h(X_i - x). \quad (7)$$

In the remainder of the paper, for simplicity, we will only consider the estimates of  $f^{(\nu)}(x)$  and  $g^{(\nu)}(x)$  obtained by minimizing the negative local log-likelihood (6), i.e. where  $\varepsilon$  is normally distributed. Future research is needed for the joint estimation that minimizes (5) for a general error distribution or (7) for the Laplace error distribution.

For the special case of  $\nu = p = 0$ , one can apply the local constant smoothing to estimate the Sharpe ratio function and the resulting estimator is  $\hat{f}^{\text{NW}}(x) = \hat{\beta}_0$ , where

$$(\hat{\beta}_0, \hat{\gamma}_0) = \arg \min_{\beta_0, \gamma_0} \sum_{i=1}^n \left[ \frac{1}{2} \{Y_i \exp(\gamma_0) - \beta_0\}^2 - \gamma_0 \right] K_h(X_i - x).$$

By simple calculations, the above optimization problem will yield an explicit solution as

$$\hat{f}^{\text{NW}}(x) = \frac{\hat{\mu}_{\text{NW}}(x)}{\sqrt{\hat{\sigma}_{\text{NW}}^2(x)}},$$

where

$$\hat{\mu}_{\text{NW}}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)} \quad \text{and} \quad \hat{\sigma}_{\text{NW}}^2(x) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i^2}{\sum_{i=1}^n K_h(X_i - x)} - \hat{\mu}_{\text{NW}}^2(x).$$

Note that  $\hat{\mu}_{\text{NW}}(x)$  is the Nadaraya–Watson estimator of  $\mu(x)$ , and  $\hat{\sigma}_{\text{NW}}^2(x)$  is an improved version of the direct estimator of  $\sigma^2(x) = \mathbb{E}(Y^2 | X = x) - \mu^2(x)$  proposed by [14]. This shows that the direct estimator  $\hat{f}^{\text{NW}}(x)$  is a special two-step estimator in which the estimates of  $\mu(x)$  and  $\sigma^2(x)$  used a common bandwidth and a common kernel function.

### 3. Asymptotic properties

In this section, we derive the joint limiting distribution of  $\hat{f}^{(v)}(x)$  and  $\hat{g}^{(v)}(x)$  obtained by minimizing the negative local log-likelihood (6). For each presentation, we first introduce some notations. Let  $\varphi(\cdot)$  be the marginal density function of the covariate  $X$ . Let

$$\mu_j = \int u^j K(u) du \quad \text{and} \quad \nu_j = \int u^j K^2(u) du, \quad j = 0, 1, \dots$$

and  $\mathbf{c}_p = (\mu_{p+1}, \dots, \mu_{2p+1})^\top$ . Let also  $\mathbf{S}$  be a  $(p+1) \times (p+1)$  matrix with  $(j, k)$ -element  $\mu_{j+k-2}$ ,  $\mathbf{S}^*$  be a  $(p+1) \times (p+1)$  matrix with  $(j, k)$ -element  $\nu_{j+k-2}$ , and  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution. In addition, to derive the joint limiting distribution, some regularity conditions are also needed as follows.

- (C1) Both  $f$  and  $g$  have a continuous  $(p+1)$ th derivative in a neighborhood of  $x$ .
- (C2) The kernel function  $K(\cdot)$  is a symmetric density function with a compact support.
- (C3) The density function  $\varphi(x)$  is differentiable and positive in a neighborhood of  $x$ .
- (C4) The error  $\varepsilon$  satisfies  $\mathbb{E}(\varepsilon) = 0$ ,  $\text{Var}(\varepsilon) = 1$ , and  $\mathbb{E}(|\varepsilon|^{4+\delta}) < \infty$  for some  $\delta > 0$ .

**Theorem 3.1:** *Assume that  $p - v$  is an odd number. Under Conditions (C1)–(C4), if  $nh^{3p+4} \rightarrow 0$  and  $nh \rightarrow \infty$ , then for any interior point  $x$  from the support of  $\varphi(\cdot)$ , we have*

$$\sqrt{nh^{2v+1}} \left\{ \begin{pmatrix} \hat{f}^{(v)}(x) \\ \hat{g}^{(v)}(x) \end{pmatrix} - \begin{pmatrix} f^{(v)}(x) \\ g^{(v)}(x) \end{pmatrix} - \mathbf{e}_{v+1}^\top \mathbf{S}^{-1} \mathbf{c}_p \frac{v!}{(p+1)!} \begin{pmatrix} f^{(p+1)}(x) \\ g^{(p+1)}(x) \end{pmatrix} h^{p+1-v} \right\} \\ \xrightarrow{\mathcal{D}} N \left( 0, \mathbf{e}_{v+1}^\top \mathbf{S}^{-1} \mathbf{S}^* \mathbf{S}^{-1} \mathbf{e}_{v+1} \frac{v!^2}{\varphi(x)} \boldsymbol{\Sigma}(x) \right),$$

where

$$\boldsymbol{\Sigma}(x) = \frac{1}{4} \begin{pmatrix} f^2(x) \text{Var}(\varepsilon^2) - 4f(x) \mathbb{E}(\varepsilon^3) + 4 & f(x) \text{Var}(\varepsilon^2) - 2\mathbb{E}(\varepsilon^3) \\ f(x) \text{Var}(\varepsilon^2) - 2\mathbb{E}(\varepsilon^3) & \text{Var}(\varepsilon^2) \end{pmatrix}. \quad (8)$$

The proof of Theorem 3.1 is given in the Supplemental Material. When  $p - \nu$  is an even number, Theorem 3.1 still holds. But in this case, we have  $\mathbf{e}_{\nu+1}^\top \mathbf{S}^{-1} \mathbf{c}_p = 0$  owing to the symmetry of the kernel  $K$  so that a term of order  $O(h^{p+2-\nu})$  will appear in the asymptotic bias. For details, see Theorem 3.1 in [10]. While for choosing the order  $p$  of the polynomial, [10] further demonstrated that the polynomial matching an odd number of  $p - \nu$  is better than the polynomial matching an even number of  $p - \nu$ . In view of this, we will focus only on an odd  $p - \nu$  in the remainder of the paper, in particular for  $p = \nu + 1$ . Moreover, Condition (C4) guarantees that the proposed estimation and Theorem 3.1 hold no matter whether or not the random error is normally distributed. In this special case when  $\varepsilon \sim N(0, 1)$ ,  $\boldsymbol{\Sigma}(x)$  can be simplified as

$$\boldsymbol{\Sigma}(x) = \frac{1}{2} \begin{pmatrix} f^2(x) + 2 f(x) & \\ f(x) & 1 \end{pmatrix}.$$

In practice, it is often of particular interest to estimate the Sharpe ratio function  $f(x)$ . Specifically, for  $\nu = 0$ , we can apply the local linear fitting with  $p = 1$ . Let the local linear estimators of  $f(x)$  and  $g(x)$  be  $\hat{f}(x)$  and  $\hat{g}(x)$ , respectively. Then by Theorem 3.1, we have the following corollary.

**Corollary 3.2:** *Under Conditions (C1)–(C4) with  $p = 1$ , if  $nh^7 \rightarrow 0$  and  $nh \rightarrow \infty$ , then for any interior point  $x$  from the support of  $\varphi(\cdot)$ , we have*

$$\sqrt{nh} \begin{pmatrix} \hat{f}(x) - f(x) - 0.5f''(x)\mu_2h^2 \\ \hat{g}(x) - g(x) - 0.5g''(x)\mu_2h^2 \end{pmatrix} \xrightarrow{D} N \left( 0, \frac{\nu_0}{\varphi(x)} \boldsymbol{\Sigma}(x) \right).$$

Corollary 3.2 provides the joint limiting distribution of  $\{\hat{f}(x), \hat{g}(x)\}$ , which exhibits the correlation between the estimated Sharpe ratio and volatility functions. In contrast, Theorem 1 in [17] only provides the marginal limiting distributions for the two estimators. Moreover, the asymptotic variances of  $\hat{f}(x)$  and  $\hat{g}(x)$  are also different from those in [17], mainly because  $\hat{f}(x)$  and  $\hat{g}(x)$  share the same bandwidth rather than two different bandwidths in [17].

Note also that the asymptotic bias and variance of  $\hat{f}^{(\nu)}(x)$  do not depend on  $g(x)$  and its derivatives. This implies that  $g(x)$  can be regarded as a nuisance function when estimating  $f(x)$  and its derivatives. Define the equivalent kernel by

$$K_v^*(u) = \mathbf{e}_{\nu+1}^\top \mathbf{S}^{-1} (1, u, \dots, u^\nu)^\top K(u).$$

Note that  $(-1)^\nu K_v^*$  is a kernel of orders  $(\nu, p + 1)$  as defined by [12]. It can be readily shown that  $\mathbf{e}_{\nu+1}^\top \mathbf{S}^{-1} \mathbf{c}_p = \int u^{p+1} K_v^*(u) du$  and  $\mathbf{e}_{\nu+1}^\top \mathbf{S}^{-1} \mathbf{S}^* \mathbf{S}^{-1} \mathbf{e}_{\nu+1} = \int K_v^{*2}(u) du$ .

Also, by Theorem 3.1, the asymptotic mean squared error (AMSE) of  $\hat{f}^{(\nu)}(x)$  is

$$\begin{aligned} \text{AMSE}\{\hat{f}^{(\nu)}(x)\} &= \nu!^2 \left\{ \int u^{p+1} K_v^*(u) du \frac{f^{(p+1)}(x)}{(p+1)!} \right\}^2 h^{2p+2-2\nu} \\ &\quad + \frac{\nu!^2}{nh^{2\nu+1}} \int K_v^{*2}(u) du \frac{\boldsymbol{\Sigma}_{11}(x)}{\varphi(x)}, \end{aligned}$$

where  $\Sigma_{11}(x) = 0.25f^2(x) \text{Var}(\varepsilon^2) - f(x)\mathbb{E}(\varepsilon^3) + 1$ . The optimal variable bandwidth minimizing the AMSE of  $\hat{f}^{(\nu)}(x)$  is

$$h_{\text{opt}}(x) = C_{\nu,p}(K) \left[ \frac{\Sigma_{11}(x)}{\varphi(x)\{f^{(p+1)}(x)\}^2} \right]^{1/(2p+3)} n^{-1/(2p+3)},$$

where

$$C_{\nu,p}(K) = \left[ \frac{(p+1)!^2(2\nu+1) \int K_{\nu}^{*2}(u) du}{2(p+1-\nu)\{ \int u^{p+1} K_{\nu}^*(u) du \}^2} \right]^{1/(2p+3)}.$$

Moreover, one may also consider a constant bandwidth by minimizing the asymptotic mean integrated squared error  $\text{AMISE}(\hat{f}^{(\nu)}) = \int \text{AMSE}\{\hat{f}^{(\nu)}(x)\}w(x) dx$  with a weight function  $w \geq 0$ . Then according, it yields the optimal constant bandwidth as

$$h_{\text{opt}} = C_{\nu,p}(K) \left[ \frac{\int \Sigma_{11}(x)w(x)/\varphi(x) dx}{\int \{f^{(p+1)}(x)\}^2 w(x) dx} \right]^{1/(2p+3)} n^{-1/(2p+3)}. \quad (9)$$

Finally, as claimed in [24], the proposed local polynomial estimator  $\hat{f}^{(\nu)}(x)$  with bandwidth  $h = h_{\text{opt}}(x)$  or  $h_{\text{opt}}$  enjoys the optimal rate of convergence  $n^{-(p+1-\nu)/(2p+3)}$ .

**Remark 3.1 (Point-wise Confidence Band):** To construct a point-wise confidence band for  $f(x)$ , we assume  $\varepsilon \sim N(0, 1)$ . Then by Corollary 3.2, an approximate  $(1 - \alpha)100\%$  confidence interval for  $f(x)$  is

$$\hat{f}(x) - \frac{1}{2}\hat{f}''(x)v_2h^2 \pm z_{1-\alpha/2} \sqrt{\frac{v_0}{nh\hat{\varphi}(x)} \left\{ \frac{1}{2}\hat{f}^2(x) + 1 \right\}}, \quad (10)$$

where  $z_{1-\alpha/2}$  denotes the  $(1 - \alpha)$ th quantile of the standard normal distribution,  $\hat{f}''(x)$  is a local linear estimator of  $f''(x)$ , and  $\hat{\varphi}(x)$  is a kernel density estimator of  $\varphi(x)$ . Note that the biased estimator in (10) involves an estimate of  $f''(x)$ . By (9), however, the optimal rate of convergence of  $\hat{f}^{(\nu)}(x)$  is  $n^{-(p+1-\nu)/(2p+3)}$ , which becomes slower as  $\nu$  gets large. In other words, our proposed estimator for the higher-order derivative may be unreliable or unstable due to the slow convergence rate, especially when  $n$  is small. In view of this, we will apply the regression bootstrap [13] to construct the point-wise confidence band for  $f(x)$  in Section 7. Also as mentioned in Section 8, an interesting future direction is to construct a simultaneous confidence band for  $f(x)$ .

#### 4. Simultaneous estimation of $\mu(\cdot)$ and $\sigma^2(\cdot)$

As a by-product, the proposed estimation in Section 2 can also be used to estimate the conditional mean function  $\mu(x)$  and the conditional variance function  $\sigma^2(x)$ , although they are not the main focus of this paper. Specifically, noting that  $\mu(x) = f(x) \exp\{-g(x)\}$  and  $\sigma(x) = \exp\{-g(x)\}$ , we can estimate them by

$$\hat{\mu}(x) = \hat{f}(x) \exp\{-\hat{g}(x)\} \quad \text{and} \quad \hat{\sigma}^2(x) = \exp\{-2\hat{g}(x)\}.$$

The following theorem establishes the joint limiting distribution of  $\hat{\mu}(x)$  and  $\hat{\sigma}^2(x)$ .

**Theorem 4.1:** Assume that the second derivatives of  $\mu(x)$  and  $\sigma^2(x)$  exist and are continuous in a neighborhood of an interior point  $x$  from the support of  $\varphi(\cdot)$ . Then under Conditions (C2)–(C4), if  $nh^7 \rightarrow 0$  and  $nh \rightarrow \infty$ , we have

$$\sqrt{nh} \begin{pmatrix} \hat{\mu}(x) - \mu(x) - 0.5\{f''(x)\sigma(x) - \mu(x)g''(x)\}\mu_2h^2 \\ \hat{\sigma}^2(x) - \sigma^2(x) + \sigma^2(x)g''(x)\mu_2h^2 \end{pmatrix} \xrightarrow{\mathcal{D}} N\left(0, \frac{\nu_0}{\varphi(x)}\Psi(x)\right),$$

where

$$\Psi(x) = \begin{pmatrix} \sigma^2(x) & \sigma^3(x)\mathbb{E}(\varepsilon^3) \\ \sigma^3(x)\mathbb{E}(\varepsilon^3) & \sigma^4(x)\text{Var}(\varepsilon^2) \end{pmatrix}.$$

The proof of Theorem 4.1 is given in the Supplemental Material. In the special case when  $\mathbb{E}(\varepsilon^3) = 0$ , e.g. when the error density is symmetric about zero, Theorem 4.1 shows that  $\hat{\mu}(x)$  and  $\hat{\sigma}^2(x)$  are asymptotically independent. Theorem 4.1 also provides the asymptotic biases and variances of  $\hat{\mu}(x)$  and  $\hat{\sigma}^2(x)$ . Specifically, for  $\hat{\mu}(x)$ , the leading term in the asymptotic variance is

$$\frac{1}{nh} \frac{\sigma^2(x)\nu_0}{\varphi(x)},$$

which is the same as that of the local linear estimator for  $\mu(x)$  in [10].

While for  $\hat{\sigma}^2(x)$ , we compare it with the residual-based estimator of  $\sigma^2(x)$  proposed by [11]. From Theorem 4.1, the leading terms in the asymptotic bias and variance are, respectively,

$$\text{bias}\{\hat{\sigma}^2(x)\} : -\sigma^2(x)g''(x)\mu_2h^2 \quad \text{and} \quad \text{Var}\{\hat{\sigma}^2(x)\} : \frac{1}{nh} \frac{\sigma^4(x)\text{Var}(\varepsilon^2)\nu_0}{\varphi(x)},$$

which are the same as those in [30]. Let also  $\hat{\mu}_{\text{LL}}(x)$  be the local linear estimator of  $\mu(x)$  with a bandwidth  $h_1 > 0$  and  $\hat{r}_i = \{Y_i - \hat{\mu}_{\text{LL}}(X_i)\}^2$ . Then by [11], the estimator of  $\sigma^2(x)$  is given as  $\hat{\sigma}_{\text{FY}}^2(x) = \hat{\alpha}_0$ , where

$$(\hat{\alpha}_0, \hat{\alpha}_1) = \arg \min_{\alpha_0, \alpha_1} \sum_{i=1}^n \{\hat{r}_i - \alpha_0 - \alpha_1(X_i - x)\}^2 K_h(X_i - x). \tag{11}$$

Moreover, it also follows from [11] that

$$\sqrt{nh} \left\{ \hat{\sigma}_{\text{FY}}^2(x) - \sigma^2(x) - \frac{h^2}{2} \mu_2 \frac{d^2}{dx^2} \sigma^2(x) + o(h_1^2 + h^2) \right\} \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^4(x)\text{Var}(\varepsilon^2)\nu_0}{\varphi(x)}\right),$$

which shows that the two estimators  $\hat{\sigma}^2(x)$  and  $\hat{\sigma}_{\text{FY}}^2(x)$  have exactly the same asymptotic variance.

### 5. Multiple covariate case

This section extends the proposed method in Section 2 to the case of multiple covariates. Let  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^\top$  be the  $d$ -dimensional covariate vector with the marginal density

function  $\varphi$ . We consider the multivariate non-parametric regression model

$$Y_i = \mu(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (12)$$

where  $\varepsilon_i$  are i.i.d. random variables with zero mean and unit variance and are independent of  $\mathbf{X}_i$ . For model (12), the existing literature has mainly focused on the estimation of  $\mu(\cdot)$  and/or  $\sigma^2(\cdot)$ . To name a few, [7] proposed a tensor-product polynomial spline estimator for  $\mu(\cdot)$  and showed that it achieves the optimal rate of convergence as defined in [24]; [20] proposed a local linear kernel-weighted least squares estimator for  $\mu(\cdot)$ ; and [6] established a minimax rate of convergence for estimating  $\sigma^2(\cdot)$  and showed that it can be achieved by a difference-based estimator.

In what follows, we estimate the multivariate Sharpe ratio function  $f(\mathbf{x}) = \mu(\mathbf{x})/\sigma(\mathbf{x})$  based on the sample data  $\{(Y_i, \mathbf{X}_i) : i = 1, \dots, n\}$ . Let  $K$  be a  $d$ -variate non-negative kernel function, and define

$$K_{\mathbf{B}}(\mathbf{u}) = \frac{1}{|\mathbf{B}|}K(\mathbf{B}^{-1}\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d,$$

where  $\mathbf{B}$  is a non-singular  $d \times d$  bandwidth matrix and  $|\mathbf{B}|$  represents its determinant. For simplicity, one can take, for example, a diagonal bandwidth matrix  $\mathbf{B} = \text{diag}\{h_1, \dots, h_d\}$  with  $h_i > 0$ ,  $i = 1, \dots, d$ . We further use the local linear smoothing to estimate  $f(\mathbf{x})$  and the nuisance function  $g(\mathbf{x}) = \log\{1/\sigma(\mathbf{x})\}$  simultaneously. When  $\varepsilon_i$  follows the standard normal distribution, by (6) we can define the local linear estimators of  $f(\mathbf{x})$  and  $g(\mathbf{x})$  as  $\hat{f}(\mathbf{x}) = \hat{a}_0$  and  $\hat{g}(\mathbf{x}) = \hat{b}_0$ , respectively, where

$$\begin{aligned} & (\hat{a}_0, \hat{\mathbf{a}}_1, \hat{b}_0, \hat{\mathbf{b}}_1) \\ &= \arg \min_{a_0, \mathbf{a}_1, b_0, \mathbf{b}_1} \sum_{i=1}^n \left( \frac{1}{2} [Y_i \exp\{b_0 + \mathbf{b}_1^\top (\mathbf{X}_i - \mathbf{x})\} - \{a_0 + \mathbf{a}_1^\top (\mathbf{X}_i - \mathbf{x})\}]^2 \right. \\ & \quad \left. - \{b_0 + \mathbf{b}_1^\top (\mathbf{X}_i - \mathbf{x})\} K_{\mathbf{B}}(\mathbf{X}_i - \mathbf{x}) \right) \end{aligned}$$

As a potential application, [25] considered the estimation of the conditional Sharpe ratio for the market returns. In their analysis, both the mean and volatility functions of stock market returns were modeled as functions of four predetermined financial variables, including the Baa-Aaa spread, the commercial paper-Treasury spread, the one-year Treasury yield, and the dividend yield. The collected data were monthly and covered the period from April 1953 to November 2010. For more details, see also Section 3.1 of [25]. Lastly, to establish the joint limiting distribution of the estimators  $\hat{f}(\mathbf{x})$  and  $\hat{g}(\mathbf{x})$ , we need the following regularity conditions.

- (B1) All second-order partial derivatives of  $f$  and  $g$  are continuous in the neighborhood of  $\mathbf{x}$ , and the density function  $\varphi$  is differentiable and positive in the neighborhood of  $\mathbf{x}$ .
- (B2) The kernel function  $K$  is a compactly supported multivariate density function such that  $\int K(\mathbf{u}) \, d\mathbf{u} = 1$ ,  $\int \mathbf{u}K(\mathbf{u}) \, d\mathbf{u} = 0$ , and  $\int \mathbf{u}\mathbf{u}^\top K(\mathbf{u}) \, d\mathbf{u} = \mu_2(K)I_d$ , where  $\mu_2(K) > 0$  and  $I_d$  is a  $d \times d$  identity matrix.

(B3) The sequence of bandwidth matrix  $\mathbf{B}$  is such that each entry of  $\mathbf{B}\mathbf{B}^\top$  tends to zero,  $n|\mathbf{B}| \rightarrow \infty$ , and  $n|\mathbf{B}|\{\text{tr}(\mathbf{B}\mathbf{B}^\top)\}^3 \rightarrow 0$  as  $n \rightarrow \infty$  with  $\mathbf{B}$  remaining non-singular, where  $\text{tr}(\mathbf{B})$  stands for the trace of the matrix  $\mathbf{B}$ .

**Theorem 5.1:** *Let  $\mathbf{x}$  be an interior point from the support of  $\varphi(\cdot)$ . Under Conditions (B1)–(B3) and (C4), we have*

$$\sqrt{n|\mathbf{B}|} \begin{pmatrix} \hat{f}(\mathbf{x}) - f(\mathbf{x}) - 0.5\mu_2(K)\text{tr}\{\mathbf{H}_f(\mathbf{x})\mathbf{B}\mathbf{B}^\top\} \\ \hat{g}(\mathbf{x}) - g(\mathbf{x}) - 0.5\mu_2(K)\text{tr}\{\mathbf{H}_g(\mathbf{x})\mathbf{B}\mathbf{B}^\top\} \end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \frac{\nu_0(K)}{\varphi(\mathbf{x})} \boldsymbol{\Sigma}(\mathbf{x})\right),$$

where  $\mathbf{H}_f(\mathbf{x})$  and  $\mathbf{H}_g(\mathbf{x})$  are the Hessian matrices of  $f$  and  $g$  evaluated at  $\mathbf{x}$  respectively,  $\nu_0(K) = \int K^2(\mathbf{u}) \, d\mathbf{u}$ , and

$$\boldsymbol{\Sigma}(\mathbf{x}) = \frac{1}{4} \begin{pmatrix} f^2(\mathbf{x}) \text{Var}(\varepsilon_1^2) - 4f(\mathbf{x})\mathbb{E}(\varepsilon_1^3) + 4 & f(\mathbf{x}) \text{Var}(\varepsilon_1^2) - 2\mathbb{E}(\varepsilon_1^3) \\ f(\mathbf{x}) \text{Var}(\varepsilon_1^2) - 2\mathbb{E}(\varepsilon_1^3) & \text{Var}(\varepsilon_1^2) \end{pmatrix}.$$

The proof of Theorem 5.1 is given in the Supplemental Material. When the covariate is univariate with  $d = 1$ , Theorem 5.1 reduces to Corollary 3.2. We note, however, that due to the curse of dimensionality, estimating the multivariate Sharpe ratio function may require a large sample size, especially when  $d$  is very large.

## 6. Simulation studies

### 6.1. Bandwidth selection

Bandwidth selection is an important issue in local polynomial modeling to balance the trade-off between the estimation bias and variance. For the bandwidth  $h$  in  $\hat{f}^{(v)}(x)$ , we follow the same rule of thumb (ROT) as recommended by [10,30]. The idea is to substitute the unknown quantities in the expression of the asymptotically optimal constant bandwidth (9). Specifically, we first fit two polynomials of order  $p + 3$  globally for  $f(x)$  and  $g(x)$  using the log-likelihood function (4) and they yield the fitted curves as

$$\check{f}(x) = \check{\beta}_0 + \check{\beta}_1x + \dots + \check{\beta}_{p+3}x^{p+3} \quad \text{and} \quad \check{g}(x) = \check{\gamma}_0 + \check{\gamma}_1x + \dots + \check{\gamma}_{p+3}x^{p+3}.$$

Taking  $\varepsilon \sim N(0, 1)$  and  $w(x) = \varphi(x)w_0(x)$  for some specific function  $w_0$ , the two terms  $\int \boldsymbol{\Sigma}_{11}(x)w(x)/\varphi(x) \, dx$  and  $\int \{f^{(p+1)}(x)\}^2w(x) \, dx$  can be estimated by  $\int \{0.5\check{f}^2(x) + 1\}w_0(x) \, dx$  and  $n^{-1} \sum_{i=1}^n \{\check{f}^{(p+1)}(X_i)\}^2w_0(X_i)$ , respectively. Finally, by (9) we have the ROT bandwidth as

$$\check{h}_{\text{ROT}} = C_{v,p}(K) \left[ \frac{\int \{0.5\check{f}^2(x) + 1\}w_0(x) \, dx}{\sum_{i=1}^n \{\check{f}^{(p+1)}(X_i)\}^2w_0(X_i)} \right]^{1/(2p+3)}.$$

When estimating the Sharpe ratio function  $f(x)$ , an alternative method for selecting the bandwidth is to apply the leave-one-out cross-validation (LOOCV) as in [17]. Specifically, the LOOCV bandwidth is given as

$$\check{h}_{\text{CV}} = \arg \min_h \sum_{i=1}^n \left( \frac{1}{2} [Y_i \exp\{\hat{g}_h^{(-i)}(X_i)\} - \hat{f}_h^{(-i)}(X_i)]^2 - \hat{g}_h^{(-i)}(X_i) \right),$$

where  $\hat{f}_h^{(-i)}(X_i)$  and  $\hat{g}_h^{(-i)}(X_i)$  are the estimators of  $f$  and  $g$  based on the bandwidth  $h$  with the  $i$ th observation excluded. However, the LOOCV cannot be used to select the bandwidth

for the derivatives of  $f(x)$ . In what follows, we compare the impact of the two bandwidths on the estimation of the Sharpe ratio function.

## 6.2. Simulations for the Sharpe ratio function

In this section, we conduct simulation studies to investigate the finite sample performance of the proposed estimator for the Sharpe ratio function and compare it with existing methods. For a fair comparison, we consider the same two models as studied by [11,16,28]. To evaluate the estimation accuracy of different estimators, we also define the root integrated squared error (RISE) of  $\hat{f}(x)$  as

$$\text{RISE}(\hat{f}) = \left[ \int \{\hat{f}(x) - f(x)\}^2 dx \right]^{1/2}.$$

**Example 6.1:** Following [11,16], we simulate 100 random samples of size  $n$  from the model

$$Y_i = a(X_i + 2 \exp(-16X_i^2)) + (0.4 \exp(-2X_i^2) + 0.2)\varepsilon_i, \quad i = 1, \dots, n,$$

where  $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[-2, 2]$  and  $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ . Accordingly, the Sharpe ratio function is

$$f(x) = a(x + 2 \exp(-16x^2))/(0.4 \exp(-2x^2) + 0.2).$$

We also consider four different sample sizes,  $n = 100, 200, 350,$  or  $500$ , and four different values of the coefficient,  $a = 0.5, 1, 2,$  or  $4$ . For each setting, we estimate  $f(x)$  by the local linear fitting as described in Section 2 with the Epanechnikov kernel and the ROT or LOOCV bandwidth proposed in Section 6.1. We then compare the new estimator with the residual-based estimator [11], the difference-based estimator [3,  $r = 1$ ], and the joint estimator [16]. For the latter three, we use the R and MATLAB codes available at [https://github.com/won-j/joint\\_estim](https://github.com/won-j/joint_estim) provided by [16]. The mean and standard deviation of RISEs for the five estimators are summarized in Table 1. It is worth noting that there are some discarded values (NaNs) in computing the RISEs for the residual- and difference-based methods, since the non-negativity constraint of the estimated variance functions is not imposed by these two methods. From Table 1, for each combination of  $a$  and an estimator of  $f(x)$ , the mean and standard deviation of RISE decrease as  $n$  increases. For each combination of  $n$  and an estimator of  $f(x)$ , the mean and standard deviation of RISE increase as  $a$  increases. When  $a = 0.5$  or  $a = 1$ , the two new estimators provide a similar performance with no matter which bandwidth is used. When  $a = 2$  or  $a = 4$ , the estimator with the ROT bandwidth performs better than that with the LOOCV bandwidth. It is also evident that the new and joint estimators perform better than the residual- and difference-based estimators in all settings. In addition, our new estimator will significantly outperform the joint estimator when  $a$  is large and  $n$  is small; whereas, for other settings, the two estimators are comparable. To conclude, our proposed estimator is competitive compared to the other three estimators and thus can be recommended for practical use.

**Example 6.2:** Following [16,28], we simulate 100 random samples of size  $n$  from the model

$$Y_i = 0.75 \sin(b\pi X_i) + ((X_i - 0.5)^2 + 0.5)^{1/2} \varepsilon_i, \quad i = 1, \dots, n,$$

where  $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$  and  $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ . Thus, the Sharpe ratio function is given by

$$f(x) = 0.75 \sin(b\pi x) / \sqrt{(x - 0.5)^2 + 0.5}.$$

**Table 1.** The mean and standard deviation (in parentheses) of RISEs for the five estimators of  $f(x)$  in Example 6.1.

$a$	Method	$n = 100$	$n = 200$	$n = 350$	$n = 500$
0.5	new(ROT)	1.590(0.964)	0.985(0.378)	0.770(0.249)	0.605(0.151)
	new(LOOCV)	1.555(0.850)	1.051(0.361)	0.817(0.236)	0.649(0.194)
	residual	2.135(1.311)	1.321(0.646)	1.024(0.579)	0.774(0.247)
	difference	2.006(1.045)	1.525(1.215)	1.072(0.479)	0.931(0.542)
	joint	1.577(1.978)	0.970(0.574)	0.733(0.255)	0.630(0.224)
1	new(ROT)	2.766(1.675)	1.966(0.868)	1.390(0.379)	1.216(0.332)
	new(LOOCV)	2.990(1.598)	2.067(0.647)	1.612(0.467)	1.381(0.481)
	residual	4.443(3.380)	2.519(1.428)	1.843(0.913)	1.535(0.690)
	difference	3.883(3.557)	2.590(1.361)	1.935(0.709)	1.776(1.175)
	joint	3.495(1.728)	2.105(1.302)	1.501(0.736)	1.185(0.646)
2	new(ROT)	2.387(0.340)	2.236(0.302)	2.077(0.318)	1.985(0.297)
	new(LOOCV)	4.584(1.017)	3.684(0.906)	3.105(0.556)	2.815(0.398)
	residual	9.547(6.884)	5.659(2.839)	3.773(1.885)	3.059(1.276)
	difference	6.894(4.485)	5.793(4.328)	3.965(1.989)	3.185(1.164)
	joint	5.425(2.641)	4.377(7.997)	2.639(1.073)	2.035(0.761)
4	new(ROT)	5.532(0.641)	5.126(0.733)	4.681(0.756)	4.430(0.643)
	new(LOOCV)	11.508(5.327)	8.647(4.749)	7.156(2.250)	6.694(1.828)
	residual	15.285(7.657)	10.855(5.689)	8.390(3.431)	6.724(2.440)
	difference	12.845(2.934)	9.672(5.635)	7.831(4.569)	6.941(3.653)
	joint	9.153(4.238)	6.803(2.771)	4.816(1.603)	4.303(1.640)

We also consider four sample sizes  $n = 50, 100, 200, 500$ , and four different values of  $b$  from 0, 4, 10 to 20. As  $b$  increases, the Sharpe ratio function  $f(x)$  gets rougher and thus estimating it becomes more and more difficult. For each setting, we then compare the proposed estimator with the three existing ones as in Example 6.1. Table 2 presents the mean and standard deviation of RISEs for the five estimators. For each combination of  $b$  and an estimator of  $f(x)$ , the mean and standard deviation of RISE decrease as  $n$  increases. For each combination of  $n$  and an estimator of  $f(x)$ , the mean and standard deviation of RISE increase as  $b$  increases, which verifies that estimating  $f(x)$  gets worse as  $b$  increases. The two new estimators provide a similar performance with no matter which bandwidth is used. Note also that the new and joint estimators perform better than the residual- and difference-based estimators in all settings except for the combinations of  $n = 50$  and  $a = 4, 10$ , or 40. Moreover, our new estimator is always the best when  $b = 0$ .

### 6.3. Simulations for the first-order derivative

In this section, we investigate the finite sample performance of the proposed estimator for the higher-order derivative of  $f(x)$ . We only focus on the estimation of the first order derivative  $f'(x)$  since estimating the second or higher-order derivative is more complex. Specifically, we estimate  $f'(x)$  by minimizing (6) with  $p = 2$ , the Epanechnikov kernel, and the ROT bandwidth. To the best of our knowledge, there is no literature on estimating the first derivative of the Sharpe ratio function. Note that  $f'(x) = \{\mu'(x)\sigma(x) - \mu(x)\sigma'(x)\}/\sigma^2(x)$ . To compare with our new estimator  $\hat{f}^{(1)}(x)$ , we consider the indirect estimator

$$\hat{f}'_{ID}(x) = \frac{\hat{\mu}'_{LL}(x)\hat{\sigma}_{FY}(x) - \hat{\mu}_{LL}(x)\hat{\sigma}'_{FY}(x)}{\hat{\sigma}_{FY}^2(x)},$$

**Table 2.** The mean and standard deviation (in parentheses) of RISEs for the five estimators of  $f(x)$  in Example 6.2.

$b$	Method	$n = 50$	$n = 100$	$n = 200$	$n = 500$
0	new(ROT)	0.267(0.151)	0.177(0.089)	0.108(0.044)	0.071(0.029)
	new(LOOCV)	0.245(0.141)	0.151(0.074)	0.116(0.055)	0.068(0.028)
	residual	0.390(0.194)	0.241(0.120)	0.163(0.055)	0.093(0.033)
	difference	0.366(0.182)	0.240(0.114)	0.160(0.053)	0.093(0.033)
	joint	0.350(0.356)	0.320(1.077)	0.127(0.044)	0.091(0.031)
4	new(ROT)	0.602(0.348)	0.333(0.069)	0.250(0.052)	0.188(0.038)
	new(LOOCV)	0.713(0.287)	0.349(0.104)	0.301(0.047)	0.169(0.037)
	residual	0.581(0.368)	0.384(0.119)	0.316(0.072)	0.230(0.046)
	difference	0.490(0.162)	0.380(0.110)	0.312(0.074)	0.225(0.046)
	joint	0.603(0.352)	0.354(0.123)	0.307(0.519)	0.184(0.238)
10	new(ROT)	0.741(0.053)	0.611(0.141)	0.373(0.060)	0.229(0.033)
	new(LOOCV)	0.748(0.101)	0.698(0.033)	0.416(0.055)	0.228(0.028)
	residual	0.846(0.155)	0.786(0.308)	0.724(0.027)	0.697(0.026)
	difference	0.844(0.136)	0.779(0.097)	0.736(0.035)	0.707(0.040)
	joint	0.921(0.442)	0.514(0.291)	0.323(0.078)	0.212(0.033)
20	new(ROT)	0.842(0.150)	0.729(0.043)	0.589(0.060)	0.335(0.051)
	new(LOOCV)	0.770(0.071)	0.727(0.030)	0.705(0.011)	0.314(0.036)
	residual	0.824(0.206)	0.740(0.042)	0.712(0.015)	0.700(0.008)
	difference	0.826(0.150)	0.745(0.041)	0.721(0.033)	0.703(0.010)
	joint	1.050(0.331)	0.700(0.073)	0.556(0.132)	0.360(0.059)

**Table 3.** The mean and standard deviation (in parentheses) of RISEs for the two estimators of  $f'(x)$  in Example 6.1.

$a$	Method	$n = 100$	$n = 200$	$n = 350$	$n = 500$
0.5	new	5.243 (1.302)	4.518 (0.710)	4.084 (0.476)	3.892 (0.314)
	indirect	18.111 (28.931)	17.157 (27.470)	12.016 (17.128)	6.630 (6.216)
1	new	10.765 (1.517)	8.835 (1.216)	8.576 (1.100)	8.250 (0.633)
	indirect	39.494 (47.531)	24.996 (42.199)	18.341 (24.271)	12.528 (9.568)

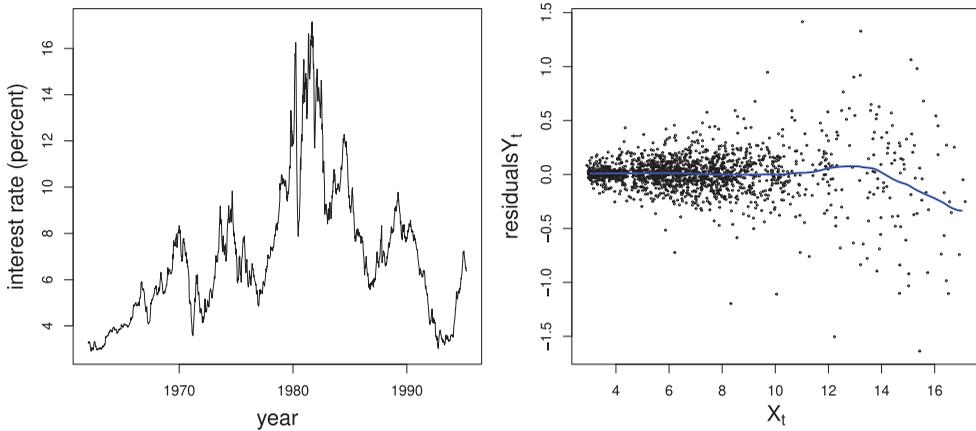
where  $\hat{\sigma}_{FY}(x) = \sqrt{\hat{\sigma}_{FY}^2(x)}$ ,  $\hat{\sigma}'_{FY}(x) = \sqrt{\hat{\sigma}_{FY}^{\prime 2}(x)}$ ,  $\hat{\mu}'_{LL}(x)$  is the local linear estimator of  $\mu'(x)$ , and  $\hat{\sigma}_{FY}^{\prime 2}(x) = \hat{\alpha}_1$  is the local linear estimator of  $d\sigma^2(x)/dx$  defined in (11). To evaluate the estimation accuracy of the estimators, we also consider the root integrated squared error (RISE) of  $\hat{f}^{(1)}(x)$  as

$$RISE(\hat{f}^{(1)}) = \left[ \int \{\hat{f}^{(1)}(x) - f'(x)\}^2 dx \right]^{1/2}.$$

We consider to estimate  $f'(x)$  in Example 6.1 with  $a = 0.5$  or  $1$  and in Example 6.2 with  $b = 0$  or  $4$  in Section 6.2, respectively. It is worth noting that there are some discarded values (NaNs) in computing the RISEs for the indirect method. Tables 3 and 4 present the mean and standard deviation of RISEs for the two estimators in Examples 6.1 and 6.2. From Table 3, for each combination of  $a$  and an estimator of  $f'(x)$ , the mean and standard deviation of RISE decrease as  $n$  increases. The same phenomena are also observed in Table 4. Moreover, our new estimator performs better than the indirect estimator in all settings in Tables 3 and 4 except for  $(b, n) = (4, 500)$ .

**Table 4.** The mean and standard deviation (in parentheses) of RISEs for the two estimators of  $f'(x)$  in Example 6.2.

$b$	Method	$n = 50$	$n = 100$	$n = 200$	$n = 500$
0	new	1.518 (0.799)	1.084 (0.517)	0.815 (0.350)	0.514 (0.242)
	indirect	7.123 (14.752)	3.847 (6.751)	1.589 (0.914)	0.832 (0.475)
4	new	9.976 (2.721)	5.900 (1.607)	4.517 (1.831)	3.043 (0.854)
	indirect	10.419 (19.296)	7.665 (13.888)	4.636 (6.041)	3.035 (0.763)



**Figure 1.** The three-month US Treasury Bill data from 5 January 1962 to 31 March 1995. Left: the raw data. Right: the residuals after an AR(5) fit is plotted against  $X_t$ .

### 7. Application to treasury bill data

We now apply the proposed method to analyze three datasets from the three-month US Treasury bill (T-bill) in the secondary market. The first dataset consists of 1735 weekly observations from 5 January 1962 to 31 March 1995, which has been previously analyzed by [2,11,16] and among others. The time series data, denoted by  $\{z_t\}$ , are presented in the left panel of Figure 1.

Following [11,16], we first fit a fifth-order autoregressive model to  $\{z_t\}$ , which yields the AR(5) model as

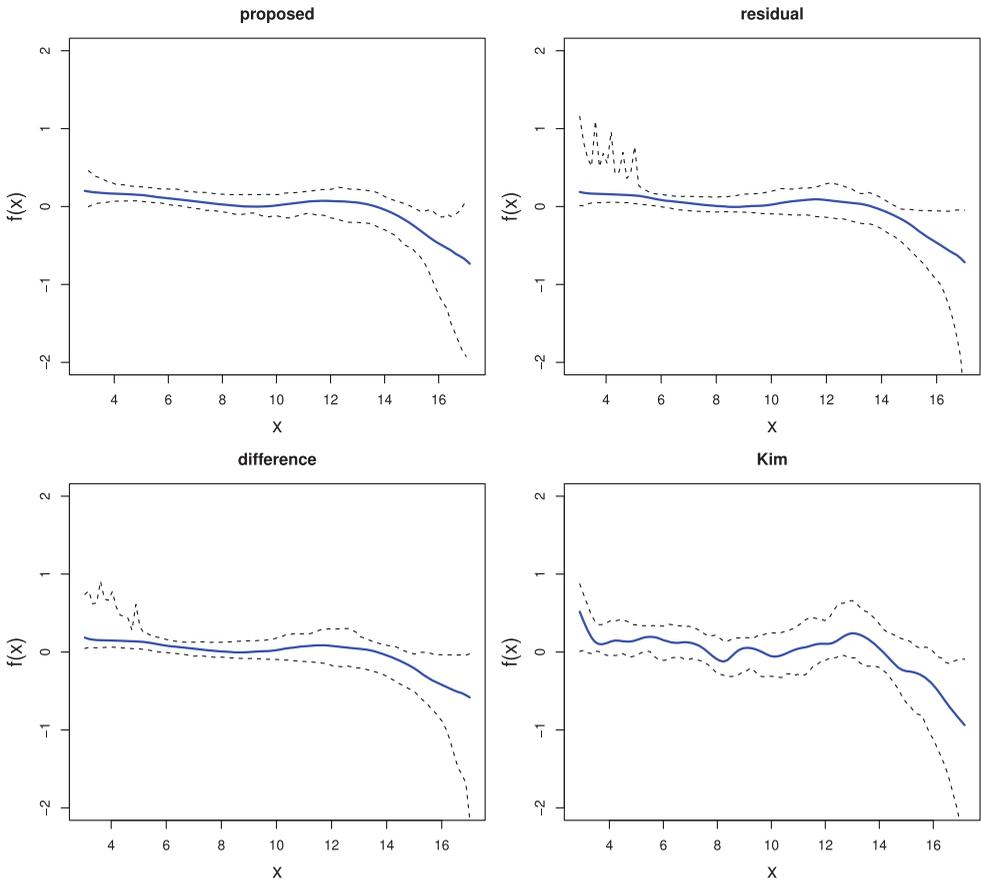
$$z_t = 1.3252z_{t-1} - 0.2800z_{t-2} - 0.0263z_{t-3} + 0.0276z_{t-4} - 0.0472z_{t-5} + Y_t.$$

The residuals  $Y_t$  are then plotted against  $X_t = z_{t-1}$  in the right panel of Figure 1. We further consider a model that is a discrete-time approximation to the continuous-time diffusion process model (2), e.g. [2,16]:

$$Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t, \tag{13}$$

where  $\mathbb{E}(\varepsilon_t | X_t) = 0$  and  $\text{Var}(\varepsilon_t | X_t) = 1$ . Finally, we apply the proposed method with the ROT bandwidth to model (13) to estimate the Sharpe ratio function  $f(x) = \mu(x)/\sigma(x)$ .

While for comparison, we also estimate the Sharpe ratio function by three existing methods, and plot the estimated functions in Figure 2 with their associated 95% point-wise confidence bands using the regression bootstrap [13]. It is evident that our new estimator



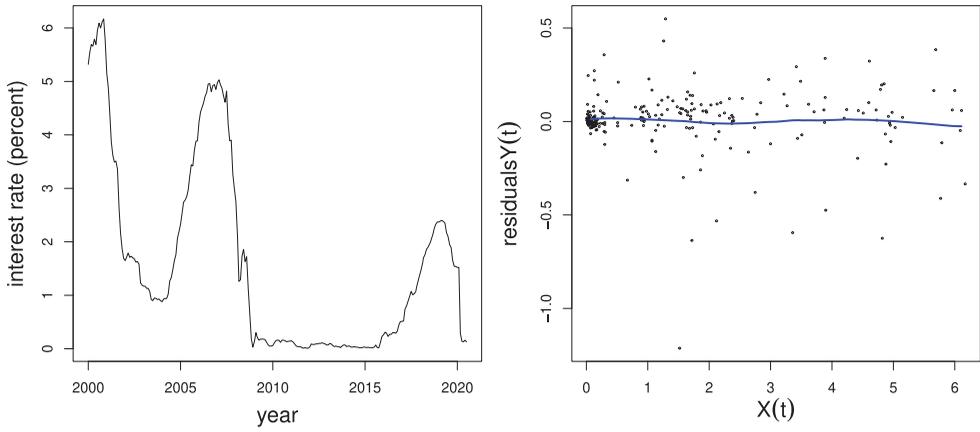
**Figure 2.** The estimated Sharpe ratio functions (solid lines) from the three-month US Treasury Bill data by the four methods: the new, residual-based, difference-based, and joint estimators. The dashed lines represent the associated 95% point-wise bootstrap confidence bands.

and its bootstrap confidence band are more stable than the others. In particular, our new estimator can capture the well-known empirical evidence that the low-priced assets always outperform the high-priced ones since the former has a larger Sharpe ratio than the latter from Figure 2.

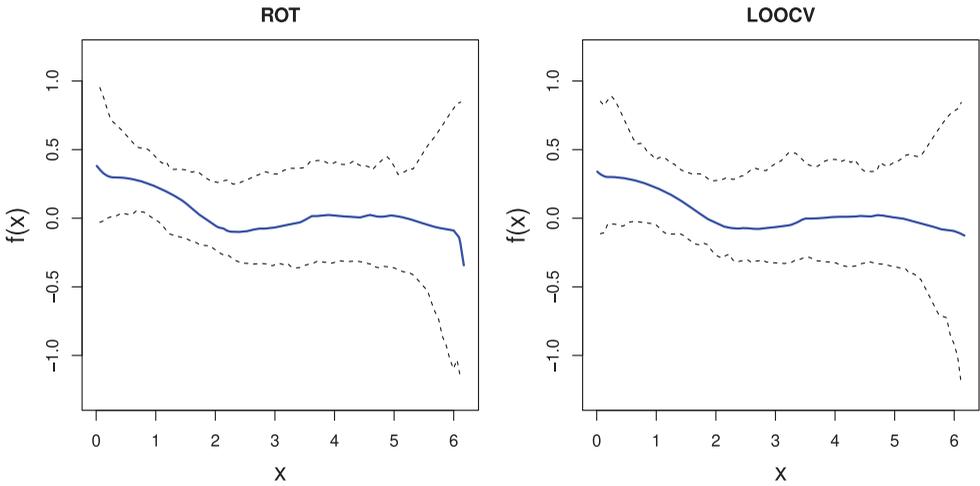
The second dataset, presented in the left panel of Figure 3 and denoted by  $\{z_t\}$ , consists of 247 monthly observations from 1 January 2000 to 1 July 2020. Following [17], after fitting an AR(4) model and regressing the residuals  $Y_t$  against  $X_t = z_{t-1}$ , we obtain the model

$$z_t - 1.3948z_{t-1} + 0.4248z_{t-2} - 0.2125z_{t-3} + 0.1940z_{t-4} = Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t,$$

where  $\mathbb{E}(\varepsilon_t | X_t) = 0$  and  $\text{Var}(\varepsilon_t | X_t) = 1$ . The residuals  $Y_t$  are plotted against  $X_t$  in the right panel of Figure 3. To compare the two bandwidths in Section 6.1, we apply the proposed method with the ROT and LOOCV bandwidths to estimate the Sharpe ratio function, which yields  $\hat{h}_{\text{ROT}} = 1.218$  and  $\hat{h}_{\text{CV}} = 1.36$ . Figure 4 displays the two estimated Sharpe ratio functions and the associated 95% point-wise confidence bands. From the fitted curves, the two bandwidths perform similarly. We can observe that the Sharpe ratio



**Figure 3.** The three-month US Treasury Bill data from 1 January 2000 and 1 July 2020. Left: the raw data. Right: the residuals after an AR(4) fit is plotted against  $X_t$ .



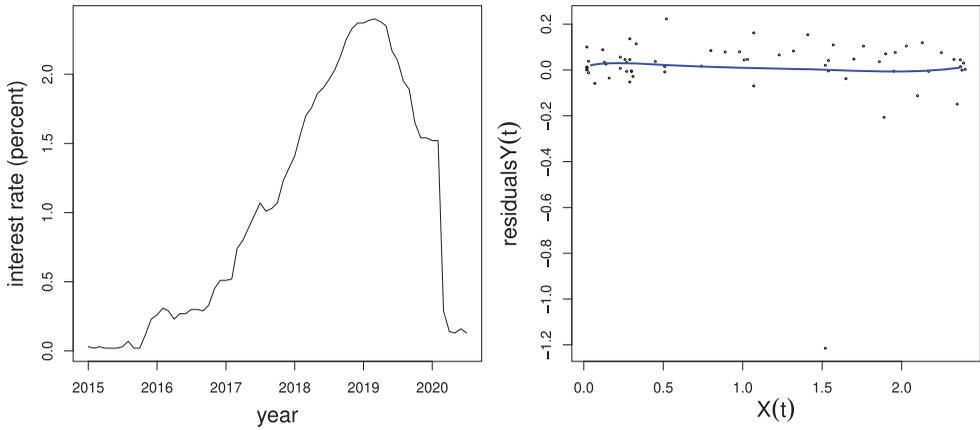
**Figure 4.** The estimated Sharpe ratio functions (solid lines) from the three-month US Treasury Bill data by the new estimator with the ROT (left) or LOOCV (right) bandwidth. The dashed lines represent the associated 95% point-wise bootstrap confidence bands.

function for the T-bill has a nonlinear trend, which also shows that the low-priced assets perform better than the high-priced ones.

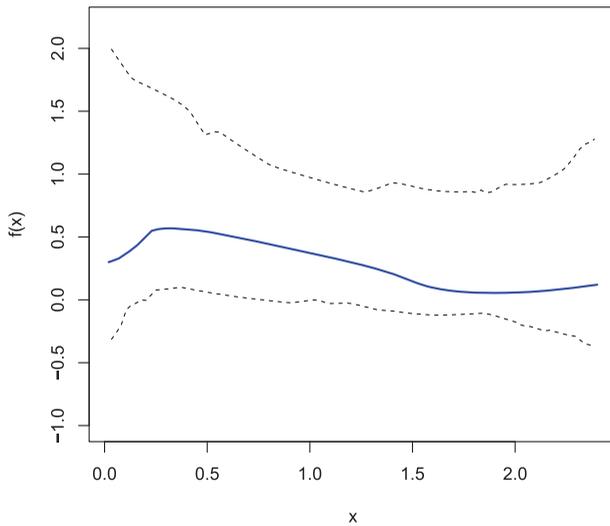
In order to evaluate the performance of the proposed method for a much shorter time period, we select the yields of the three-month US T-bill data from 1 January 2015 to 1 July 2020. This dataset consists of 67 monthly observations and is presented in the left panel of Figure 5, where  $z_t$  denotes the time series of the yields. Following [17], after fitting an AR(2) model and regressing the residuals  $Y_t$  against  $X_t = z_{t-1}$ , we obtain the model

$$z_t - 1.2198z_{t-1} + 0.2301z_{t-2} = Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t,$$

where  $\mathbb{E}(\varepsilon_t | X_t) = 0$  and  $\text{Var}(\varepsilon_t | X_t) = 1$ . The residuals  $Y_t$  are plotted against  $X_t$  in the right panel of Figure 5. Figure 6 plots the estimated Sharpe ratio function with the LOOCV



**Figure 5.** The three-month US Treasury Bill data from 1 January 2015 to 1 July 2020. Left: the raw data. Right: the residuals after an AR(2) fit is plotted against  $X_t$ .



**Figure 6.** The estimated Sharpe ratio function (solid line) from the three-month US Treasury Bill data by the new estimator. The dashed lines represent the associated 95% point-wise bootstrap confidence band.

bandwidth and its 95% point-wise confidence band. The bandwidth selected by LOOCV is 1.33. Figure 6 also indicates that the low-priced assets perform better than the high-priced ones.

### 8. Discussion

In this paper, we propose a direct method via local maximum likelihood for estimating the Sharpe ratio function or its derivatives in the heteroscedastic non-parametric model. We further establish the asymptotic normal distribution for the proposed estimator under

some regularity conditions, and show that it can be competitive compared to existing methods using simulated data and real market data.

Along with the paper, there exist a few interesting extensions. Firstly, it is of interest to investigate the asymptotic properties and numerical performance for the estimators obtained by minimizing the general negative log-likelihood (5) or (7). Secondly, note that there have been extensive studies on the construction of the simultaneous confidence band for  $\mu(x)$  and/or  $\sigma^2(x)$ ; see, for example, [4,5,27] and among others. Inspired by this, another interesting but challenging problem can be to construct a simultaneous confidence band for the Sharpe ratio function. Lastly, our new method for estimating the multivariate Sharpe ratio function in Section 5 may suffer from the curse of dimensionality. To overcome this problem, it might be necessary to impose structural assumptions on the multivariate Sharpe ratio function, e.g. single index modeling. A deeper and detailed investigation of these issues warrants further studies.

## Acknowledgments

The authors thank the editor, associate editor, and two referees for their constructive suggestions and comments that have substantially improved an earlier version of this paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

Wenchao Xu's research was partially supported by the China Postdoctoral Science Foundation (2021M693340) and the National Natural Science Foundation of China (12101591). Hongmei Lin's research was partially supported by the National Natural Science Foundation of China (12171310) and the Shanghai Natural Science Foundation (20ZR1421800). Tiejun Tong's research was partially supported by the General Research Fund (HKBU12303421, HKBU12303918), the Initiation Grant for Faculty Niche Research Areas of Hong Kong Baptist University (RC-FNRA-IG/20-21/SCI/03), and the National Natural Science Foundation of China (1207010822). Riquan Zhang's research was partially supported by the National Science Foundation of China (11971171,11831008), and the Basic Research Project of Shanghai Science and Technology Commission (22JC1400800).

## References

- [1] Y. Ait-Sahalia, *Testing continuous-time models of the spot interest rate*, Rev. Financ. Stud. 9 (1996), pp. 385–426.
- [2] T.G. Andersen and J. Lund, *Estimating continuous-time stochastic volatility models of the short-term interest rate*, J. Econom. 77 (1997), pp. 343–377.
- [3] L.D. Brown and M. Levine, *Variance estimation in nonparametric regression via the difference sequence method*, Ann. Stat. 35 (2007), pp. 2219–2232.
- [4] L. Cai, R. Liu, S. Wang, and L. Yang, *Simultaneous confidence bands for mean and variance functions based on deterministic design*, Stat. Sin. 29 (2019), pp. 505–525.
- [5] L. Cai and L. Yang, *A smooth simultaneous confidence band for conditional variance function*, Test 24 (2015), pp. 632–655.
- [6] T.T. Cai, M. Levine, and L. Wang, *Variance function estimation in multivariate nonparametric regression with fixed design*, J. Multivar. Anal. 100 (2009), pp. 126–136.
- [7] H. Chen, *Polynomial splines and nonparametric regression*, J. Nonparametr. Stat. 1 (1991), pp. 143–156.

- [8] L.H. Chen, M.Y. Cheng, and L. Peng, *Conditional variance estimation in heteroscedastic regression models*, J. Stat. Plan. Inference 139 (2009), pp. 236–245.
- [9] W. Dai, T. Tong, and L. Zhu, *On the choice of difference sequence in a unified framework for variance estimation in nonparametric regression*, Stat. Sci. 32 (2017), pp. 455–468.
- [10] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London, 1996.
- [11] J. Fan and Q. Yao, *Efficient estimation of conditional variance functions in stochastic regression*, Biometrika 85 (1998), pp. 645–660.
- [12] T. Gasser, H.G. Müller, and V. Mammitzsch, *Kernels for nonparametric curve estimation*, J. R. Stat. Soc.: Ser. B 47 (1985), pp. 238–252.
- [13] W. Härdle and A.W. Bowman, *Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands*, J. Am. Stat. Assoc. 83 (1988), pp. 102–110.
- [14] W. Härdle and A. Tsybakov, *Local polynomial estimators of the volatility function in nonparametric autoregression*, J. Econom. 81 (1997), pp. 223–242.
- [15] J. Huang, *Local asymptotics for polynomial spline regression*, Ann. Stat. 31 (2003), pp. 1600–1635.
- [16] S.J. Kim, J. Lim, and J.H. Won, *Nonparametric Sharpe Ratio Function Estimation in Heteroscedastic Regression Models via Convex Optimization*, in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, JMLR, Cambridge, MA, 2018, pp. 1495–1504.
- [17] H. Lin, T. Tong, Y. Wang, W. Xu, and R. Zhang, *Direct local linear estimation for Sharpe ratio function*, Can. J. Stat. 50 (2022), pp. 36–58.
- [18] A.W. Lo, *The statistics of Sharpe ratios*, Financ. Anal. J. 58 (2002), pp. 36–52.
- [19] P. Maio and P. Santa-Clara, *Multifactor models and their consistency with the ICAPM*, J. Financ. Econ. 106 (2012), pp. 586–613.
- [20] D. Ruppert and M.P. Wand, *Multivariate locally weighted least squares regression*, Ann. Stat. 22 (1994), pp. 1346–1370.
- [21] W.F. Sharpe, *The Sharpe ratio*, J. Portf. Manag. 21 (1994), pp. 49–58.
- [22] Y. Shen, C. Gao, D. Witten, and F. Han, *Optimal estimation of variance in nonparametric regression with random design*, Ann. Stat. 48 (2020), pp. 3589–3618.
- [23] R. Stanton, *A nonparametric model of term structure dynamics and the market price of interest rate risk*, J. Finance 52 (1997), pp. 1973–2002.
- [24] C.J. Stone, *Optimal global rates of convergence for nonparametric regression*, Ann. Stat. 10 (1982), pp. 1040–1053.
- [25] Y. Tang and R.F. Whitelaw, *Time-varying Sharpe ratios and market timing*, Q. J. Finance 1 (2011), pp. 465–493.
- [26] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.
- [27] J. Wang and L. Yang, *Polynomial spline confidence bands for regression curves*, Stat. Sin. 19 (2009), pp. 325–342.
- [28] L. Wang, L.D. Brown, T.T. Cai, and M. Levine, *Effect of mean on variance function estimation in nonparametric regression*, Ann. Stat. 36 (2008), pp. 646–664.
- [29] Y. Wang, *Smoothing Splines: Methods and Applications*, Chapman & Hall, London, 2011.
- [30] K. Yu and M.C. Jones, *Likelihood-based local linear estimation of the conditional variance function*, J. Am. Stat. Assoc. 99 (2004), pp. 139–144.
- [31] M. Yuan and G. Wahba, *Doubly penalized likelihood estimator in heteroscedastic regression*, Stat. Probab. Lett. 69 (2004), pp. 11–20.
- [32] F.A. Ziegelmann, *Nonparametric estimation of volatility functions: The local exponential estimator*, Econ. Theory 18 (2002), pp. 985–991.