



Direct local linear estimation for Sharpe ratio function

Hongmei LIN^{1,2}, Tiejun TONG³, Yuedong WANG⁴ , Wenchao XU^{5*} , and Riquan ZHANG^{2,6}

¹*School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China*

²*Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education, East China Normal University, Shanghai, China*

³*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

⁴*Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, U.S.A.*

⁵*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China*

⁶*School of Statistics, East China Normal University, Shanghai, China*

Key words and phrases: Heteroscedasticity; local likelihood estimation; local linear regression; nonparametric regression; Sharpe ratio function.

MSC 2020: Primary 62G05; secondary 62G08.

Abstract: Nonparametric regression has been widely used to deal with nonlinearity and heteroscedasticity in financial time series. As the ratio of the mean and standard deviation functions, the Sharpe ratio function is one of the most commonly used risk/return measures in financial econometrics. Most existing methods take an indirect procedure, which first estimates the mean and variance functions and then applies these two functions to estimate the Sharpe ratio function. In practice, however, such an indirect procedure can often be less efficient. In this article, we propose a direct method to estimate the Sharpe ratio function by local linear regression. We further establish the asymptotic normality of the proposed estimator, apply Monte Carlo simulations to evaluate its finite sample performance, and compare it with the indirect method. The usefulness of our new method is also illustrated through a real data analysis. *The Canadian Journal of Statistics* 00: 000–000; 2021 © 2021 Statistical Society of Canada

Résumé: La régression non paramétrique est communément utilisée pour traiter la non-linéarité et l'hétéroscédasticité dans les séries chronologiques financières. Par ailleurs, le ratio de Sharpe est l'une des mesures de risque/rendement les plus couramment utilisées en économétrie financière, et ce en raison de son expression comme rapport de fonctions moyenne et écart-type. L'estimation de ce ratio se fait, généralement, en deux temps. Les fonctions moyenne et écart-type sont d'abord évaluées, ensuite, elles sont appliquées pour estimer le ratio de Sharpe. Mais, une telle procédure indirecte peut s'avérer moins efficace en pratique. Comme alternative, les auteurs de cet article font appel à la régression linéaire locale pour une estimation directe du ratio de Sharpe. En plus d'établir la normalité asymptotique de l'estimateur proposé, ils procèdent à des études de simulation Monte Carlo afin d'examiner les performances à distance finie de cette méthode directe, et la comparer à la méthode indirecte. Enfin, ils illustrent l'utilité de la nouvelle méthode par le biais d'une véritable analyse de données. *La revue canadienne de statistique* 00: 000–000; 2021 © 2021 Société statistique du Canada

* Corresponding author: wcx_stat@126.com

1. INTRODUCTION

Consider a heteroscedastic nonparametric regression model

$$y_i = \mu(x_i) + \sigma(x_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where y_i and x_i are the i th observations of a response variable and an explanatory variable, respectively, $\mu(x)$ is the mean function, $\sigma(x)$ is the standard deviation (or volatility) function and bounded away from zero, and ε_i 's are independent and identically distributed (i.i.d.) normal errors with zero mean and unit variance. Estimation of the mean function has been extensively studied for several decades, with the existing methods including, for example, kernel smoothing (Härdle, 1990; Wand & Jones, 1995), smoothing splines (Wahba, 1990; Wang, 2011) and local polynomial regression (Fan & Gijbels, 1996). The variance function is of nearly equal importance in practice and has also been well studied in the literature (Hall & Carroll, 1989; Ruppert et al., 1997; Fan & Yao, 1998; Yu & Jones, 2004; Yuan & Wahba, 2004; Liu, Tong & Wang, 2007; Cai & Wang, 2008; Zhang et al., 2020).

The ratio between the mean and standard deviation functions, $\mu(x)/\sigma(x)$, is referred to as the Sharpe ratio function. In financial econometrics, the Sharpe ratio is one of the most commonly used statistics, expressing the ratio of the excess expected return of an investment to its return volatility. The Sharpe ratio is also used in many other contexts, including performance attribution, test of market efficiency and risk management. The Sharpe ratio is one of the most popular measures of the risk-adjusted return and is often regarded as a gold standard for comparison between different assets or trading strategies. There is a vast literature on the estimation of the Sharpe ratio, e.g., Sharpe (1994), Lo (2002), Tang & Whitelaw (2011) and Leung, Song & Yang (2013). On the other hand, a static Sharpe ratio with a constant standard deviation usually oversimplifies the risk. This motivates us to consider the covariate-dependent Sharpe ratio function, which may provide evidence of the fundamental economics underlying the economy and asset pricing.

The Sharpe ratio function may be modelled as a linear function of time-varying market variables such as credit spreads and yields (Mario & Santa-Clara, 2012). Existing methods for estimating the Sharpe ratio function often follow an indirect procedure (Sharpe, 1994; Lo, 2002), in which the mean and variance functions are estimated separately and then a ratio is applied to estimate the Sharpe ratio function. As a consequence, such an indirect approach can often be less efficient because the smoothing parameters are selected individually for the mean and variance functions. To our knowledge, the only exception was the paper by Kim, Lim & Won (2018) where the nonparametric functions were modelled using cubic splines. Specifically, by solving a sequence of finite-dimensional convex programs with increasing dimensions, Kim, Lim & Won (2018) concentrated only on the computational method but did not study the theoretical properties of their proposed estimators. In this article, we propose a direct method for estimating the Sharpe ratio function based on the local likelihood estimation, and we also develop a novel data-driven method to select the smoothing parameters.

The rest of the article is organized as follows. In Section 2, we propose the new estimation method for the Sharpe ratio function, together with a specific computational procedure. In Section 3, we investigate the statistical properties of the proposed estimators and establish their asymptotic normality. In Section 4, we conduct extensive simulations to evaluate the finite sample performance of our new method and also compare it with the indirect method and the method in Kim, Lim & Won (2018). In Section 5, we apply the proposed method to analyze Treasury bill data, demonstrating that our new method is able to identify the underlying stochastic process in terms of the Sharpe ratio and can also capture the well-known empirical evidence that low-priced assets will always outperform high-priced ones. Finally, we conclude the paper with a future direction in Section 6. We provide the proofs of the theoretical results in the Appendix.

2. LOCAL LINEAR ESTIMATION

2.1. Methodology

Suppose that $\{(x_i, y_i), i = 1, \dots, n\}$ are n independent realizations of (X, Y) that follow the heteroscedastic nonparametric regression model (1). Ignoring the additive constant, the negative log-likelihood function is given by

$$l(\mu, \sigma) = \sum_{i=1}^n \left\{ \frac{1}{2} \left[\frac{y_i - \mu(x_i)}{\sigma(x_i)} \right]^2 + \log \sigma(x_i) \right\}. \quad (2)$$

Let $f(x) = \mu(x)/\sigma(x)$ be the Sharpe ratio function, and denote $g(x) = -\log(\sigma(x))$. Thus, the negative log-likelihood function in (2) can be represented as

$$l(f, g) = \sum_{i=1}^n \left\{ \frac{1}{2} [y_i \exp(g(x_i)) - f(x_i)]^2 - g(x_i) \right\}. \quad (3)$$

Note that by the logarithmic transformation we have removed the positive constraint on the standard deviation function.

With (3), our main objective is then to estimate the Sharpe ratio as the function f that minimizes the negative log-likelihood function $l(f, g)$, where g is treated as a nuisance function. Suppose that f and g have continuous second derivatives. For ease of notation, let $K_h(\cdot) = K(\cdot/h)/h$, where $K(\cdot)$ is a kernel function, with $h > 0$ being a bandwidth. Let $\mu_j = \int u^j K(u) du$ and $\nu_j = \int u^j K^2(u) du$ for $j = 0, 1$ and 2 . In what follows, we propose to estimate f and g nonparametrically by the local linear estimation and apply the backfitting method (Hastie & Tibshirani, 1990; Ansley & Kohn, 1994) for the iterative estimation. With the current estimate $f_{-}(\cdot)$, we minimize the negative local log-likelihood function

$$\sum_{i=1}^n \left\{ \frac{1}{2} [y_i \exp(z_i^T \gamma) - f_{-}(x_i)]^2 - z_i^T \gamma \right\} K_{h_2}(x_i - x) \quad (4)$$

with respect to $\gamma = (\gamma_0, \gamma_1)^T$, where $z_i = (1, (x_i - x)/h_2)^T$. We then assign the minimizer of γ_0 as the updated estimate of $g(x)$. With the updated estimate $g_{-}(\cdot)$, we minimize the following objective function:

$$\sum_{i=1}^n \frac{1}{2} [y_i \exp(g_{-}(x_i)) - w_i^T \beta]^2 K_{h_1}(x_i - x) \quad (5)$$

with respect to $\beta = (\beta_0, \beta_1)^T$, where $w_i = (1, (x_i - x)/h_1)^T$. Note that the term $g(x)$ has been removed from (5) because of its independence of β . We update the estimate of $f(x)$ as the minimizer of β_0 .

Note that the objective function (4) is not quadratic and there is no closed-form solution. We will derive the Hessian matrix of (4) and show that it is positive definite in the neighbourhood of the true value. That is, the objective function (4) is strictly convex in expectation and so a minimizer exists. We propose to compute it by the one-step Newton–Raphson estimator (Cai, Fan & Li, 2000), which is known to be both statistically and computationally efficient when given a good initial estimate.

In addition, we note that the selection of the two bandwidths, h_1 and h_2 , is crucial for the performance of the proposed method. In this article, we apply the leave-one-out cross-validation (LOO-CV) method to select the bandwidths separately at each iteration (Mammen & Park, 2005).

For example, at the estimation step that solves Equation (5) for an estimate of f , the LOO-CV bandwidth is

$$h_1 = \arg \min_{h_1} \sum_{i=1}^n \frac{1}{2} [y_i \exp(g_-(x_i)) - \hat{f}^{(-i)}(x_i)]^2,$$

where $\hat{f}^{(-i)}(x_i)$ is the updated estimate of f based on data with the i th observation excluded.

2.2. Algorithm

To summarize the estimation procedure in Section 2.1, we have a two-step algorithm as follows:

1. *Initialize*: Derive an initial estimate of f .
2. *Cycle*: Alternate between (a) and (b) until convergence.
 - (a) Conditional on the current estimate of f , update g by minimizing the objective function (4), in which the bandwidth h_2 is selected by the LOO-CV method.
 - (b) Conditional on the current estimate of g , update f by minimizing the objective function (5), in which the bandwidth h_1 is selected by the LOO-CV method.

We denote the estimates of f and g at convergence as \hat{f} and \hat{g} , respectively. Moreover, for the initial estimate of f , we first compute the local linear estimate of μ under the homoscedastic model where $\sigma(x) = \sigma$ is a constant function, and then set the initial estimate of f as the ratio of the estimated μ and residual standard deviation. We use the current estimate of g as the initial value in the one-step Newton–Raphson procedure for updating g . This simple approach works very well in our simulations.

3. ASYMPTOTIC PROPERTIES

In this section we study the asymptotic properties of the proposed estimators. Some regularity conditions are needed as follows, where (C1) and (C2) are routinely assumed in the literature of nonparametric regression and (C3) and (C4) are needed when deriving the asymptotic normality for the proposed estimators.

- (C1) The random variable X has a bounded support, and its density function $\varphi(x)$ is Lipschitz-continuous and is bounded away from 0 on the support.
- (C2) The kernel function $K(\cdot)$ is a symmetric density function with bounded support and satisfies the Lipschitz condition.
- (C3) Suppose that $f(\cdot)$ and $g(\cdot)$ have a continuous third derivative.
- (C4) Suppose that $nh_2^6 \rightarrow 0$, $nh_1h_2^4 \rightarrow 0$ and $h_1/h_2 \rightarrow 0$.

Theorem 1. Suppose that conditions (C1)–(C4) hold and $nh_1h_2 \rightarrow \infty$ as $n \rightarrow \infty$. Then for the Sharpe ratio function

$$\sqrt{nh_1} \left(\hat{f}(x) - f(x) - \frac{1}{2} f''(x) \mu_2 h_1^2 \right) \xrightarrow{D} N \left(0, \frac{v_0}{\varphi(x)} \right),$$

and for the nuisance function

$$\sqrt{nh_2} \left(\hat{g}(x) - g(x) - \frac{1}{2} g''(x) \mu_2 h_2^2 \right) \xrightarrow{D} N \left(0, \frac{\varpi(x) v_0}{\{f^2(x) + 1\}^2 \varphi(x)} \right),$$

where $\varpi(x) = E(\varepsilon^4 - 2\varepsilon^2 + 1) + f^2(x)$, and \xrightarrow{D} denotes convergence in distribution.

It immediately follows from Theorem 1 that $\hat{f}(x)$ and $\hat{g}(x)$ are consistent estimators of $f(x)$ and $g(x)$, respectively. We also note that the asymptotic distribution of the estimated Sharpe ratio function, $\hat{f}(x)$, has the same structure as that for the local linear estimator of the mean function. Furthermore, the asymptotic bias and variance of $\hat{f}(x)$ are independent of the nuisance function $g(x)$, which results from the smoothness assumption on the Sharpe ratio function and the approximation error from the local linear regression. For more details, see the proof of Theorem 1 in the Appendix.

4. SIMULATION STUDY

In this section, we investigate the finite-sample performance of our proposed estimator for the Sharpe ratio function by Monte Carlo simulations. We also compare our proposed direct method with the indirect method, which estimates the mean and variance functions separately and then uses the ratio to estimate the Sharpe ratio function, and with the existing direct method in Kim, Lim & Won (2018) where the nonparametric functions were modelled using cubic splines. As a benchmark for comparison, the oracle estimate of the Sharpe ratio function, obtained by assuming the variance function is known, is also included in the simulations.

To assess the estimation accuracy of different methods, we compute the root integrated squared error (RISE):

$$\text{RISE}(\hat{f}(\cdot)) = \left[\int_0^1 (\hat{f}(t) - f(t))^2 dt \right]^{\frac{1}{2}}.$$

Example 1: We generate 500 random samples from the following model as in Wang et al. (2008) and Kim, Lim & Won (2018):

$$y_i = 0.75 \sin(b\pi x_i) + \sigma(x_i)\varepsilon_i, \quad i = 1, \dots, n,$$

where $x_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0,1)$ and $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. We consider a factorial design with five sample sizes: $n = 25, 50, 100, 200$, and 400 , two different frequencies: $b = 1$ and 10 and two different standard deviation functions: $\sigma_1(x) = 0.5$ and $\sigma_2(x) = \sqrt{(x - 1/2)^2 + 1/2}$. With the above settings, we compute the direct estimate of the Sharpe ratio function, $\hat{f}(x)$, by the proposed algorithm in Section 2.2. We also compute the indirect estimates as $\check{f}(x) = \check{\mu}(x)/\sqrt{\check{\sigma}^2(x)}$ and $\check{g}(x) = -\log(\sqrt{\check{\sigma}^2(x)})$, where $\check{\mu}$ and $\check{\sigma}^2$ are estimated separately using the local linear method and the residual-based method (Fan & Gijbels, 1996; Yu & Jones, 2004). To be more specific, we first compute the local linear estimate $\check{\mu}(x)$ based on the original observations; and then with the squared residuals $r_i = (y_i - \check{\mu}(x_i))^2$ for $i = 1, \dots, n$, we estimate $\check{\sigma}^2(x)$ at the intercept $\check{\alpha}_1$, where

$$(\check{\alpha}_1, \check{\alpha}_2) = \arg \min_{\alpha_1, \alpha_2} \sum_{i=1}^n (r_i - \alpha_1 - \alpha_2(x_i - x))^2 K_h(x_i - x),$$

and the bandwidth h is selected using the cross-validation method. For the oracle estimate of the Sharpe ratio function, we first standardize model (1) as $y^* = y/\sigma(x) = f(x) + \varepsilon$, which assumes $\sigma(x)$ is known, and then apply the local linear method to estimate f directly as the Sharpe ratio function, denoted by $\hat{f}^*(x)$.

With 500 simulations, we compute the RISE, and present the boxplots of these RISEs in Figure 1 for $\sigma(x) = \sigma_1(x)$ and in Figure 2 for $\sigma(x) = \sigma_2(x)$. From the boxplots in Figures 1 and 2, it is evident that the direct method has smaller RISE than the indirect method in all settings. The direct estimate performs nearly as well as the oracle estimate with σ known. Specifically,

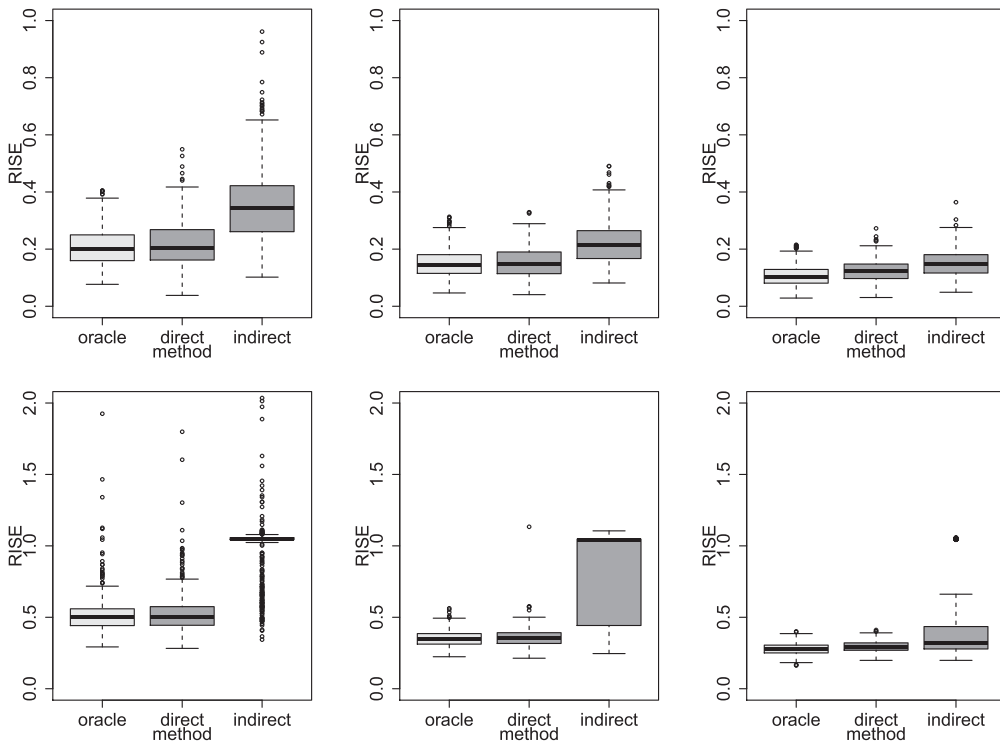


FIGURE 1: Boxplots of the root integrated squared errors (RISEs) from the estimation of the Sharpe ratio function $f(x)$ when $\sigma_1(x) = 0.5$ in Example 1. Top: $b = 1$; bottom: $b = 10$. Left: $n = 100$; middle: $n = 200$; right: $n = 400$.

for small sample sizes, our direct method still performs well compared to the indirect method. However, there are some discarded values (NaNs) in computing RISE for the indirect method when the sample size is smaller than 100. This is because the indirect method produced negative variance estimates, which led to NaNs in computing $\tilde{f}(\cdot)$ for x with negative variance. On average, the indirect method has 9 out of 500 samples with negative variance estimates. It is also noteworthy that the overall performance improves as the sample size increases. Finally, to visualize the curve fitting, we plot in Figure 3 the true curve of the Sharpe ratio function together with the estimated curves corresponding to the minimum, the first quartile, the median, the third quartile and the maximum RISEs for the setting with $b = 10$, $n = 200$ and $\sigma(x) = \sigma_2(x)$. It is evident that our proposed direct method is able to provide a reasonable fit for the Sharpe ratio function. In our algorithm, we set the maximum number of iterations to be 20 and the convergence criterion to be that the difference between the two estimates is less than 0.001. The average number of iterations in this example is 7.4.

Example 2: This simulation example is designed to compare our method with the direct method in Kim, Lim & Won (2018) where the nonparametric functions were modelled using cubic splines. Following their simulation study scenario, we simulate 500 random samples from the following model:

$$y_i = a(x_i + 2 \exp(-16x_i^2)) + (0.4 \exp(-2x_i^2) + 0.2)\varepsilon_i,$$

where $x_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(-2, 2)$ and $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. We consider a factorial design with four sample sizes: $n = 50, 100, 200$, and 400 , and four different amplitudes: $a = 0.5, 1, 2$, and 4 . Thus the

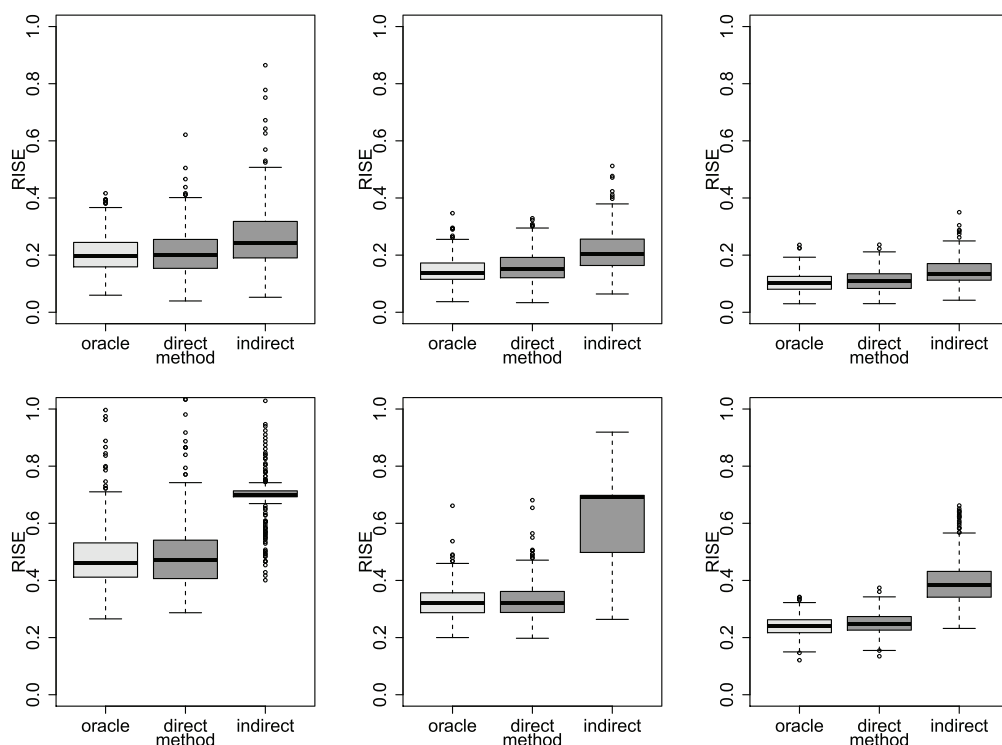


FIGURE 2: Boxplots of the root integrated squared errors (RISEs) from the estimation of the Sharpe ratio function $f(x)$ when $\sigma_2(x) = \sqrt{(x - 1/2)^2 + 1/2}$ in Example 1. Top: $b = 1$; bottom: $b = 10$. Left: $n = 100$; middle: $n = 200$; right: $n = 400$.

Sharpe ratio function is given by

$$f(x) = \frac{a(x_i + 2e^{-16x_i^2})}{0.4 \exp(-2x_i^2) + 0.2}.$$

We compare our estimator $\hat{f}(\cdot)$ with the estimate by Kim, Lim & Won (2018), denoted by \hat{f}_{kim} . The boxplots in Figures 4 and 5 indicate that the proposed method has a smaller RISE than the method in Kim, Lim & Won (2018) when the sample size is larger than 100. Kim, Lim & Won's (2018) method performs better when $n = 50$. It is also noteworthy that the overall performance improves as the sample size increases. Finally, to visualize the curve-fitting, we plot in Figure 6 the true curve of the Sharpe ratio function $f(\cdot)$ and $g(\cdot)$ with the estimated curves for the setting with $a = 0.5$ and $n = 200$.

Overall, our proposed estimation method has a satisfactory performance in many settings. We also note that the average number of non-convergence cases in our algorithm is 5.6 out of 500 simulations. The non-convergence cases are mainly caused by the near-singular Hessian matrix of the objective function in (4) due to data sparsity in certain local regions or the small bandwidth, especially when the sample size is small.

5. APPLICATION TO TREASURY BILL DATA

In this section, we apply our proposed method to analyze the 3-month US Treasury bill (T-bill) from the secondary market rates. The 3-month T-bill rate is the yield received for investing in

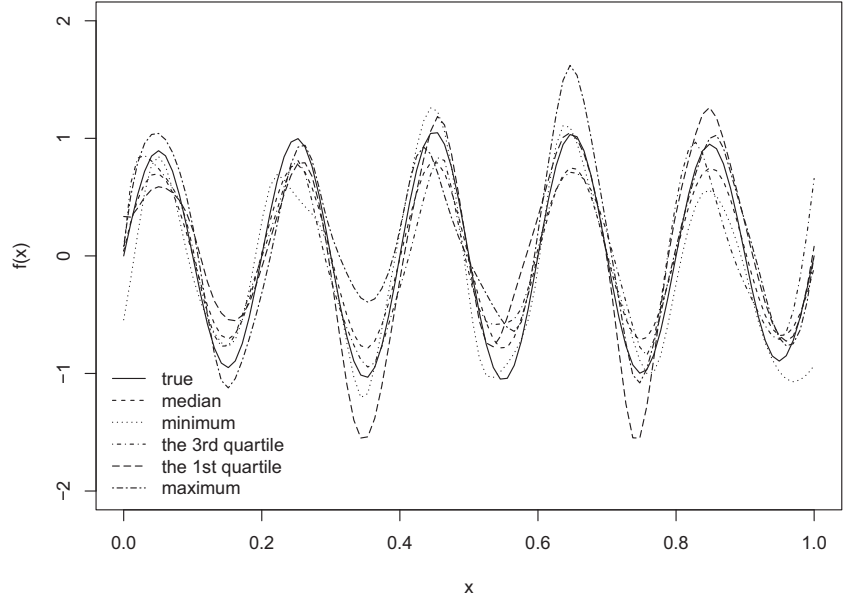


FIGURE 3: Plots of the true Sharpe ratio function (solid line) together with five dotted and dashed lines corresponding to the minimum, the first quartile, the median, the third quartile, and the maximum RISEs under the setting with $b = 10, n = 200$ and $\sigma_2(x) = \sqrt{(x - 1/2)^2 + 1/2}$ in Example 1.

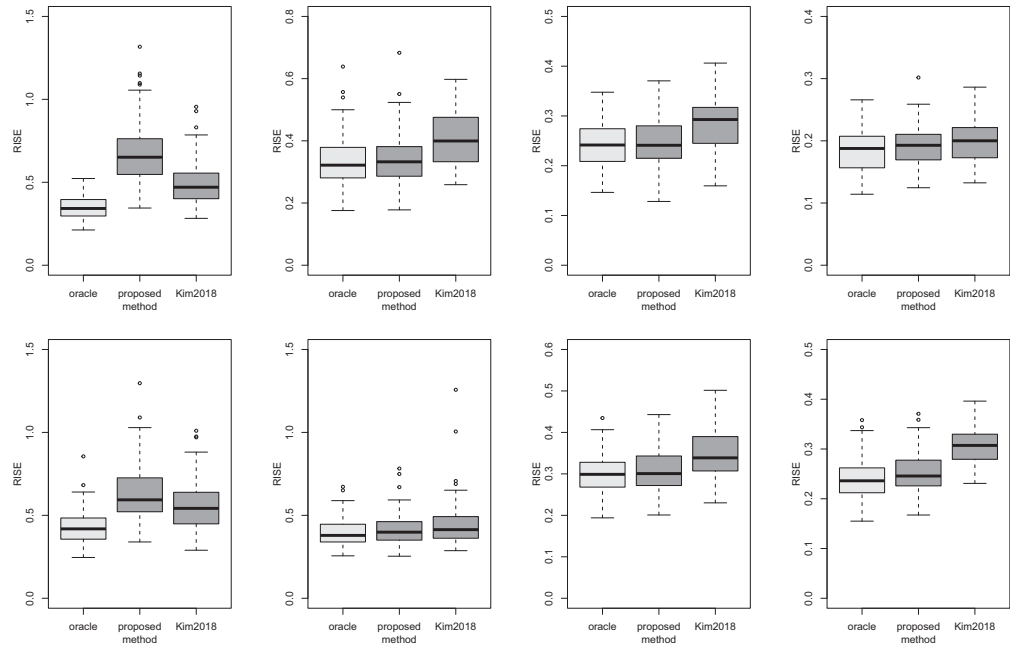


FIGURE 4: Boxplots of the root integrated squared errors (RISEs) from the estimation of the Sharpe ratio function $f(x)$ in Example 2. Top: $a = 0.5$; bottom: $a = 1$. Four panels from left to right correspond to sample size $n = 50, 100, 200$ and 400 , respectively.

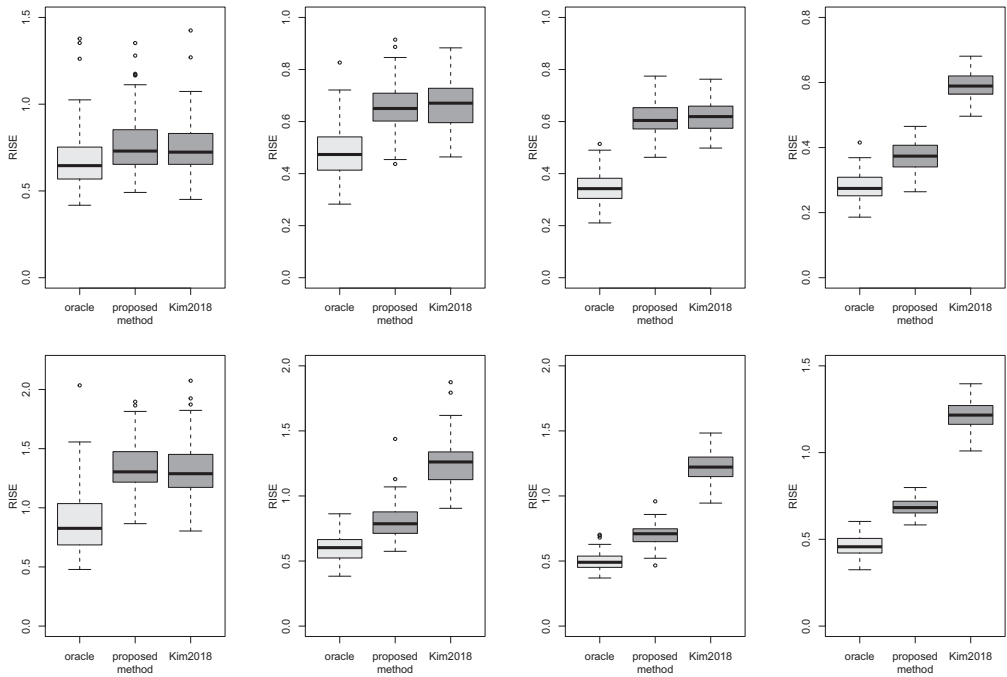


FIGURE 5: Boxplots of the root integrated squared errors (RISEs) from the estimation of the Sharpe ratio function $f(x)$ in Example 2. Top: $a = 2$; bottom: $a = 4$. Four panels from left to right correspond to sample size $n = 50, 100, 200$ and 400 , respectively.

a government-issued treasury security that has a maturity of 3 months. The 3-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. In what follows, we analyze two datasets where μ and σ represent the instantaneous expected rate of return and the volatility, respectively.

The first dataset consists of 398 monthly observations, from January 1962 to February 1995, of the yields of the 3-month US T-bill from the secondary market rates. The data are presented in the left panel of Figure 7, where z_t denote the time series of the yields. As analyzed in Andersen & Lund (1997), Fan & Yao (1998) and Kim, Lim & Won (2018), we first fit an autoregressive (AR) model with the order selected by the Akaike information criterion (AIC). This leads to the AR(3) model as

$$z_t = 1.3869z_{t-1} - 0.6306z_{t-2} + 0.2210z_{t-3} + y_t,$$

where the residuals y_t are plotted against $x_t \equiv z_{t-1}$ in the right panel of Figure 7. We then consider the following heteroscedastic nonparametric regression model, which is a discrete-time approximation to a continuous-time diffusion process model (Kim, Lim & Won, 2018):

$$y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t,$$

where $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. Then, to estimate the Sharpe ratio function $f(x) = \mu(x)/\sigma(x)$, we apply the proposed direct method and plot in Figure 8 the estimated Sharpe ratio function and its 95% point-wise confidence intervals obtained by the bootstrap method. From the estimated curve, the nonlinearity with a slightly increasing trend (up to $x_t = 11$) and then a decreasing trend can be

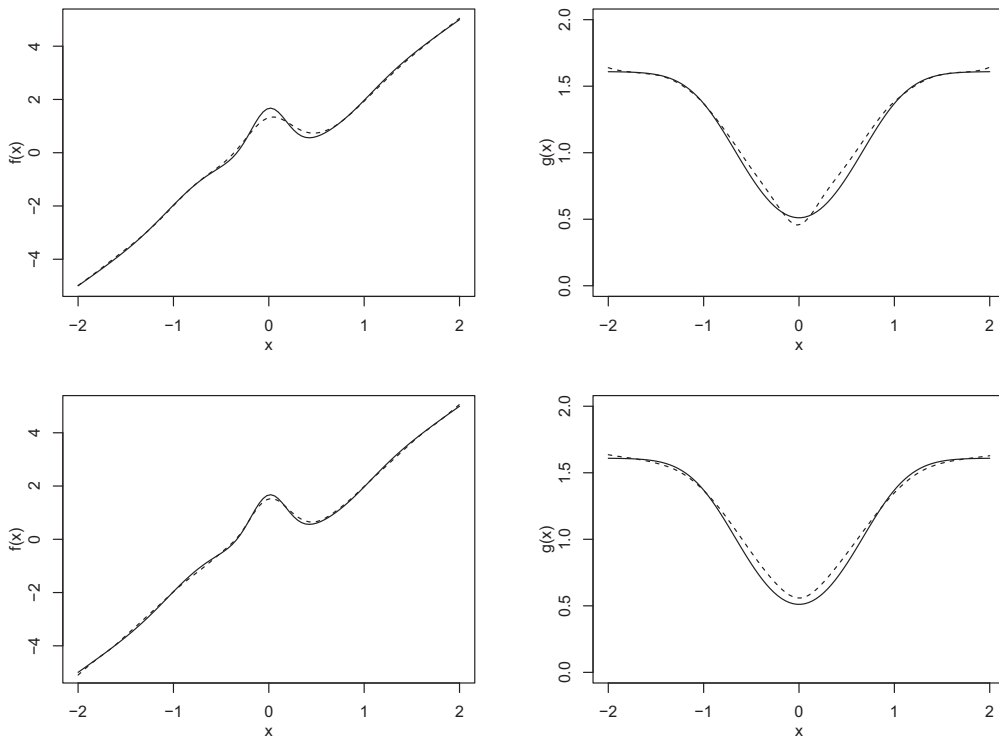


FIGURE 6: Simulation results for the estimation methods in Example 2 ($a = 0.5, n = 200$). Top: Kim, Lim & Won's (2018) method; bottom: our proposed method. Left: the true Sharpe ratio function (solid lines) together with the estimated regression curves (dotted lines); right: the true function $g(x)$ (solid lines) together with the estimated regression curves (dotted lines).

observed. This shows that our proposed method can capture the well-known empirical evidence that low-priced assets always outperform high-priced ones.

The second dataset, presented in the left panel of Figure 9, consists of the yields of the 3-month US T-bill from the secondary market rates between January 2000 and July 2019. The data can be downloaded from the webpage with the link <https://fred.stlouisfed.org/series/TB3MS>, with the length of 235 for this time series $\{z_t\}$. Similarly as in the first study, we also fit an AR model with the order selected by AIC, and regressed the residuals, denoted by y_t , against $x_t \equiv z_{t-1}$. The residuals y_t are then plotted against $x_t \equiv z_{t-1}$ in the right panel of Figure 9. Thus the model that we estimate is

$$z_t - 1.4783z_{t-1} + 0.5846z_{t-2} - 0.3552z_{t-3} + 0.2633z_{t-4} = y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t,$$

where $E(\varepsilon_t|x_t) = 0$ and $\text{Var}(\varepsilon_t|x_t) = 1$. To check the effectiveness of our estimation procedure, we also plot the estimated Sharpe ratio function and its 95% point-wise confidence intervals in Figure 10. From the fitted curve, it is clear that the Sharpe ratio function for the T-bill has a nonlinear trend. Ignoring the damping on the left edge, which is possibly due to the boundary effect, we obtain a similar conclusion that low-priced assets will always outperform high-priced ones.

In order to make a more comprehensive comparison with the estimation method in Kim, Lim & Won (2018) and evaluate the performance of the proposed method for a much shorter time

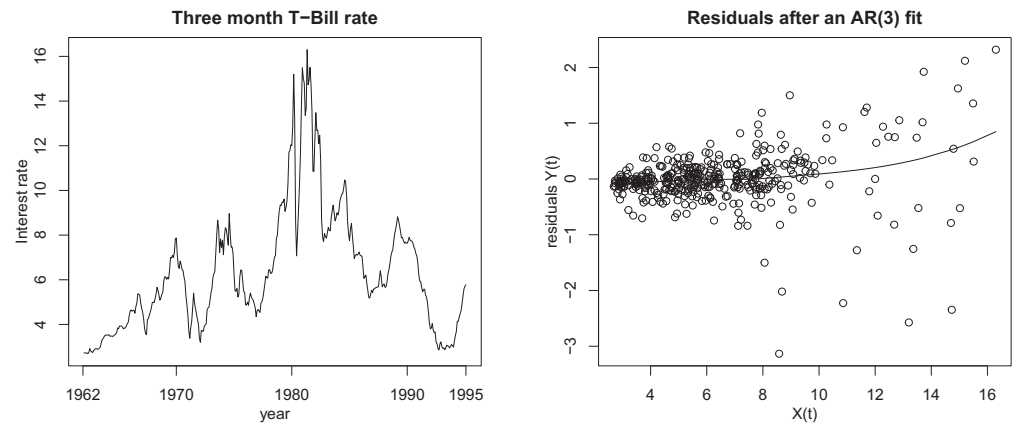


FIGURE 7: The 3-month US Treasury bill data from January 1962 to February 1995. Left: raw data; right: residuals after an AR(3) fit is plotted against $x_t = z_{t-1}$.

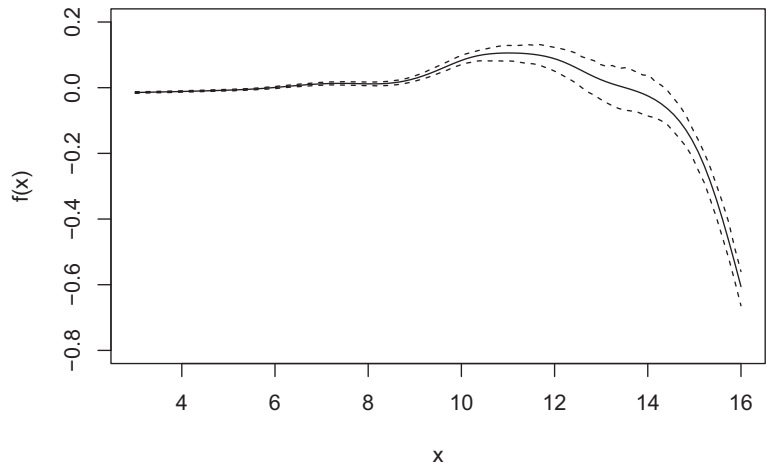


FIGURE 8: The direct estimate of the Sharpe ratio function from the 3-month US Treasury bill data (solid curve) and its 95% confidence intervals (dashed curves).

period, we select the yields of the 3-month US T-bill data from the past 5 years. This dataset consists of 67 monthly observations, from January 2015 to July 2020. The data are presented in the left panel of Figure 11, where z_t denote the time series of the yields. Similarly, after fitting an AR model with the order selected by AIC and regressing the residuals y_t against $x_t \equiv z_{t-1}$, we obtain the model

$$z_t - 1.209z_{t-1} + 0.2371z_{t-2} = y_t = \mu(x_t) + \sigma(x_t)\varepsilon_t,$$

where $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. The residuals y_t are plotted against $x_t \equiv z_{t-1}$ in the right panel of Figure 11. Then we apply the proposed method and Kim, Lim & Won's (2018) method to estimate the Sharpe ratio function, and plot the estimated Sharpe ratio function and its 95% point-wise confidence intervals obtained by the bootstrap method in Figure 12. From the fitted curves, the results of the two methods for analyzing T-bill data show the same trend: that is, low-priced

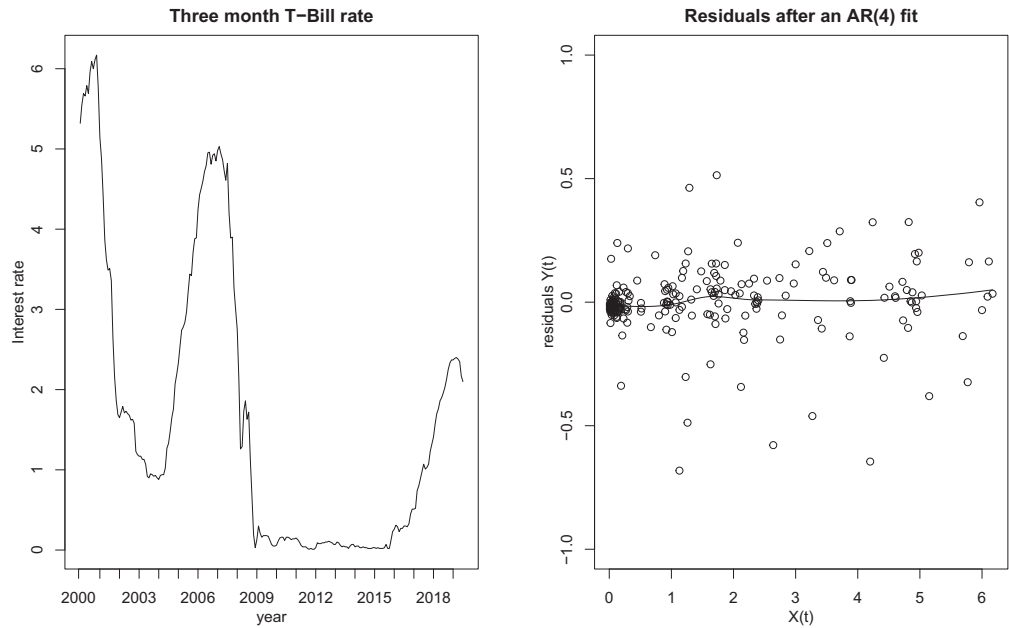


FIGURE 9: The 3-month US Treasury bill data from January 2000 to July 2020. Left: raw data; right: residuals after an AR(2) fit is plotted against $x_t = z_{t-1}$.

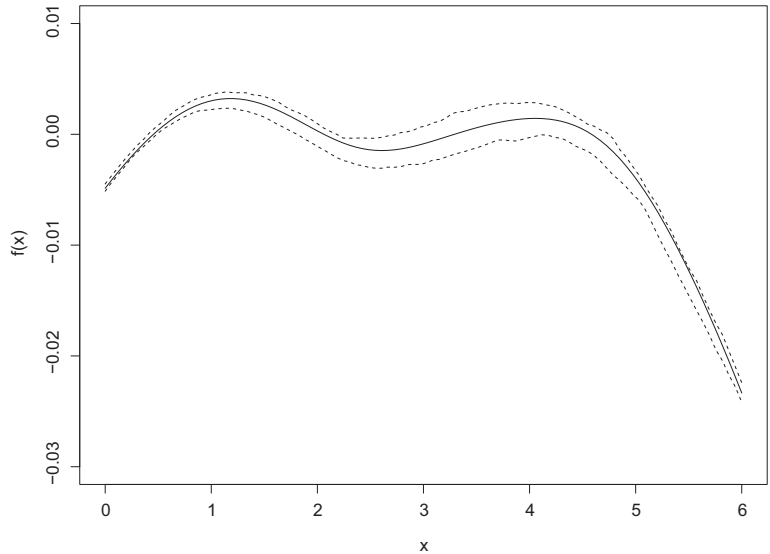


FIGURE 10: The direct estimate of the Sharpe ratio function from the 3-month US Treasury bill data (solid curve) and its 95% confidence intervals (dashed curves).

assets always outperform high-priced ones. Note also that our proposed method has narrower confidence intervals, indicating that our method is more stable. Moreover, our newly proposed method is simple to understand, easy to implement and fast to calculate.

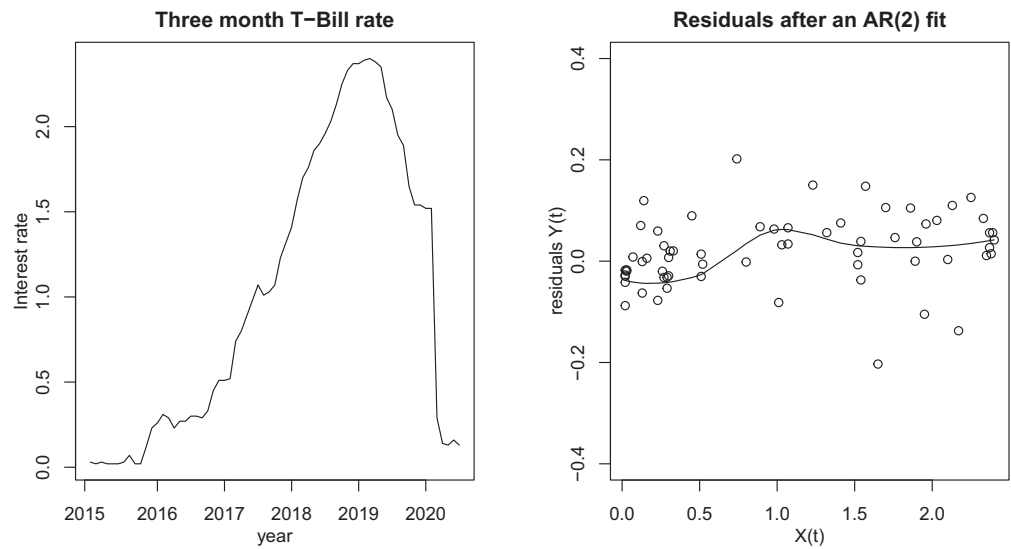


FIGURE 11: The 3-month US Treasury bill data from January 2015 to July 2020. Left: raw data; right: residuals after an AR(2) fit is plotted against $x_t = z_{t-1}$.

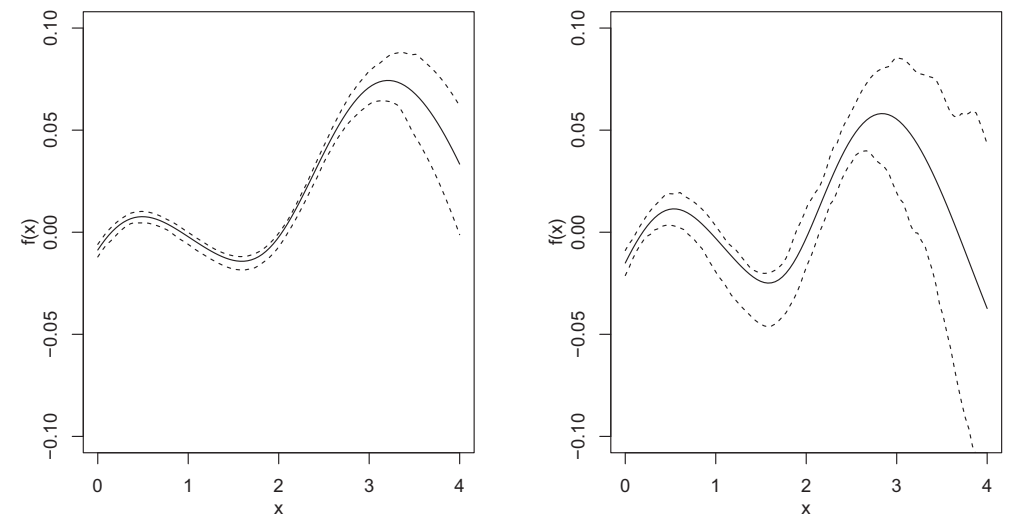


FIGURE 12: The direct estimate of the Sharpe ratio function from the 3-month US Treasury bill data (solid curve) and its 95% confidence intervals (dashed curves). Left: our proposed method; right: Kim, Lim & Won's (2018) method.

6. CONCLUSION

In this article, we have proposed a direct method for estimating the Sharpe ratio function in heteroscedastic models by local linear regression. We developed a backfitting algorithm to compute the estimate. Extensive simulations showed that the algorithm would converge in most cases, but exceptions occur because of the near-singular Hessian matrix of (4) caused by data sparsity in certain local regions or by small bandwidth, especially when the sample size is small.

Under mild conditions, we established the asymptotic normality of the proposed estimators. Simulation results also showed that the direct method is superior to the indirect method, which combines the separate estimates of the mean and variance functions. Moreover, our proposed method performs better than Kim, Lim & Won's (2018) method when the sample size is large, for which we have the theoretical guarantee. The analysis of the 3-month US T-bill data also showed that our proposed method can be potentially applied to identify the underlying stochastic process in terms of the Sharpe ratio and capture the well-known empirical evidence. Finally, we note that, for real-data analysis under the AR model, the residuals are fitted using a heteroscedastic nonparametric regression model to estimate the Sharpe ratio function. In future research, we will also investigate whether a more general model, e.g., a semiparametric model, is needed to estimate the AR coefficient and the Sharpe ratio function simultaneously.

ACKNOWLEDGEMENTS

We thank the editor, the associate editor and the two reviewers for their constructive comments, which led to a significant improvement of this article. Hongmei Lin's research was partially supported by the National Natural Science Foundation of China (11701360, 11971300), the Shanghai Natural Science Foundation (20ZR1421800, 19ZR1420900) and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University). Tiejun Tong's research was partially supported by the National Natural Science Foundation of China (11671338) and the General Research Fund (HKBU12303918). Yuedong Wang's research was partially supported by the National Science Foundation of the United States (DMS-1507620). Wenchao Xu's research was partially supported by China Post-doctoral Science Foundation (2021M693340). Riquan Zhang's research was partially supported by the National Natural Science Foundation of China (11971171).

REFERENCES

- Andersen, T. G. & Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77, 343–377.
- Ansley, C. F. & Kohn, R. (1994). Convergence of the backfitting algorithm for additive models. *Journal of the Australian Mathematical Society*, 57, 316–329.
- Cai, T. T. & Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics*, 36, 2025–2054.
- Cai, Z., Fan, J., & Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888–902.
- Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Fan, J. & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 645–660.
- Hall, P. & Carroll, R. J. (1989). Variance function estimation in regression: The effect of estimating the mean. *Journal of the Royal Statistical Society, Series B*, 51, 3–14.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Boston.
- Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Kim, S., Lim, J., & Won, J. (2018). Nonparametric Sharpe ratio function estimation in heteroscedastic regression models via convex optimization. *21st International Conference on Artificial Intelligence and Statistics*, 1495–1504.
- Leung, T., Song, Q., & Yang, J. (2013). Outperformance portfolio optimization via the equivalence of pure and randomized hypothesis testing. *Finance and Stochastics*, 17, 839–870.
- Liu, A., Tong, T., & Wang, Y. (2007). Smoothing spline estimation of variance functions. *Journal of Computational and Graphical Statistics*, 16, 312–329.
- Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58, 36–52.
- Mammen, E. & Park, B. U. (2005). Bandwidth selection for smooth backfitting in additive models. *The Annals of Statistics*, 33, 1260–1294.

- Mario, P. & Santa-Clara, P. (2012). Multifactor models and their consistency with the ICAPM. *Journal of Financial Economics*, 106, 586–613.
- Ruppert, D., Wand, M. P., Holst, U., & Hossjer, O. (1997). Local polynomial variance-function estimation. *Technometrics*, 39, 262–273.
- Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 21, 49–58.
- Tang, Y. & Whitelaw, R. F. (2011). Time-varying Sharpe ratios and market timing. *The Quarterly Journal of Finance*, 1, 465–493.
- Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- Wang, L., Brown, L. D., Cai, T. T., & Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics*, 36, 646–664.
- Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. Chapman and Hall, London.
- Yu, K. & Jones, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association*, 99, 139–144.
- Yuan, M. & Wahba, G. (2004). Doubly penalized likelihood estimator in heteroscedastic regression. *Statistics and Probability Letters*, 69, 11–20.
- Zhang, J., Wang, H., Zhang, R., & Zhang, J. (2020). Sequential feature screening for generalized linear models with sparse ultra-high dimensional data. *Journal of Systems Science and Complexity*, 33, 510–526.

APPENDIX

This section outlines the main steps for proving Theorem 1. Let $\boldsymbol{\gamma} = (\gamma_0, h_2\gamma_1)^T$, $\tau_n = (nh_2)^{-1/2}$, $\boldsymbol{\xi}(x) = (g(x), h_2g'(x))^T$, and $\boldsymbol{\gamma}^* = \tau_n^{-1} (\gamma_0 - g(x), h_2(\gamma_1 - g'(x)))^T$.

If $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, h_2\hat{\gamma}_1)^T$ minimizes (4), then $\hat{\boldsymbol{\gamma}}^*$ minimizes

$$h_2 \sum_{i=1}^n \left(\frac{1}{2} [y_i \exp(\tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \mathbf{z}_i^T \boldsymbol{\xi}(x)) - \hat{f}(x_i)]^2 - \tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* \right) K_{h_2}(x_i - x) \\ - h_2 \sum_{i=1}^n \mathbf{z}_i^T \boldsymbol{\xi}(x) K_{h_2}(x_i - x)$$

with respect to $\boldsymbol{\gamma}^*$. Since the second term does not depend on $\boldsymbol{\gamma}^*$, it is then equivalent to saying that $\hat{\boldsymbol{\gamma}}^*$ minimizes

$$\tilde{\ell}_n(\boldsymbol{\gamma}^*) = h_2 \sum_{i=1}^n \left(\frac{1}{2} [y_i \exp(\tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \mathbf{z}_i^T \boldsymbol{\xi}(x)) - \hat{f}(x_i)]^2 - \tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* \right) K_{h_2}(x_i - x). \quad (\text{A.1})$$

Note also that

$$y_i \exp(\tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \mathbf{z}_i^T \boldsymbol{\xi}(x)) - \hat{f}(x_i) \\ = y_i \exp(g(x_i)) \left[\tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* - \frac{1}{2} g''(x)(x_i - x)^2 + r_n(x_i, \boldsymbol{\gamma}^*) \right] + f(x_i) - \hat{f}(x_i) + \varepsilon_i, \quad (\text{A.2})$$

where

$$r_n(x_i, \boldsymbol{\gamma}^*) = \exp(\tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \mathbf{z}_i^T \boldsymbol{\xi}(x) - g(x_i)) - 1 - \tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* + \frac{1}{2} g''(x)(x_i - x)^2.$$

Substituting (A.2) into (A.1), $\tilde{\ell}_n(\boldsymbol{\gamma}^*)$ can be decomposed as

$$\tilde{\ell}_n(\boldsymbol{\gamma}^*) = \frac{1}{2} \boldsymbol{\gamma}^{*T} \mathbf{A}_n \boldsymbol{\gamma}^* - \mathbf{W}_n^T \boldsymbol{\gamma}^* + \sum_{j=1}^6 R_{nj}(\boldsymbol{\gamma}^*) + R_{n7},$$

where

$$\begin{aligned}
 A_n &= h_2 \tau_n^2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) K_{h_2}(x_i - x) \mathbf{z}_i \mathbf{z}_i^T, \\
 W_n &= h_2 \tau_n \sum_{i=1}^n \left[1 - y_i \exp(g(x_i)) \varepsilon_i + \frac{1}{2} y_i^2 \exp(2g(x_i)) g''(x)(x_i - x)^2 \right] \mathbf{z}_i K_{h_2}(x_i - x), \\
 R_{n1}(\boldsymbol{\gamma}^*) &= \frac{1}{2} h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) r_n^2(x_i, \boldsymbol{\gamma}^*) K_{h_2}(x_i - x), \\
 R_{n2}(\boldsymbol{\gamma}^*) &= \tau_n h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) r_n(x_i, \boldsymbol{\gamma}^*) \mathbf{z}_i^T \boldsymbol{\gamma}^* K_{h_2}(x_i - x), \\
 R_{n3}(\boldsymbol{\gamma}^*) &= -g''(x) \frac{1}{2} h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) r_n(x_i, \boldsymbol{\gamma}^*) (x_i - x)^2 K_{h_2}(x_i - x), \\
 R_{n4}(\boldsymbol{\gamma}^*) &= h_2 \sum_{i=1}^n y_i \exp(g(x_i)) \varepsilon_i r_n(x_i, \boldsymbol{\gamma}^*) K_{h_2}(x_i - x), \\
 R_{n5}(\boldsymbol{\gamma}^*) &= h_2 \sum_{i=1}^n y_i \exp\{g(x_i)\} \tau_n \mathbf{z}_i^T \boldsymbol{\gamma}^* \{f(x_i) - \hat{f}(x_i)\} K_{h_2}(x_i - x), \\
 R_{n6}(\boldsymbol{\gamma}^*) &= h_2 \sum_{i=1}^n y_i \exp\{g(x_i)\} r_n(x_i, \boldsymbol{\gamma}^*) \{f(x_i) - \hat{f}(x_i)\} K_{h_2}(x_i - x),
 \end{aligned}$$

and R_{n7} is a term that does not depend on $\boldsymbol{\gamma}^*$. This shows that $\hat{\boldsymbol{\gamma}}^*$ minimizes

$$\ell_n(\boldsymbol{\gamma}^*) = \frac{1}{2} \boldsymbol{\gamma}^{*T} A_n \boldsymbol{\gamma}^* - \mathbf{W}_n^T \boldsymbol{\gamma}^* + \sum_{j=1}^6 R_{nj}(\boldsymbol{\gamma}^*).$$

For A_n , noting that

$$\mathbb{E} \left[\tau_n^2 h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) K_{h_2}(x_i - x) \left(\frac{x_i - x}{h_2} \right)^j \right] = [f^2(x) + 1] \varphi(x) \mu_j + o(1),$$

we have

$$A_n = A + o_p(1), \quad \text{where } A = [f^2(x) + 1] \varphi(x) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}.$$

Furthermore, if we can prove that $R_{nj}(\boldsymbol{\gamma}^*) = o_p(1)$ for each $\boldsymbol{\gamma}^*$ and $j = 1, \dots, 6$, then

$$\ell_n(\boldsymbol{\gamma}^*) = \frac{1}{2} \boldsymbol{\gamma}^{*T} A \boldsymbol{\gamma}^* - \mathbf{W}_n^T \boldsymbol{\gamma}^* + o_p(1).$$

By the quadratic approximation theorem, we obtain

$$\hat{\boldsymbol{\gamma}}^* = A^{-1} \mathbf{W}_n + o_p(1).$$

This yields

$$\begin{aligned}\sqrt{nh_2}(\hat{g}_0 - g(x)) &= \frac{\tau_n h_2}{[f^2(x) + 1]\varphi(x)} \sum_{i=1}^n \left[1 - y_i \exp(g(x_i))\varepsilon_i + \frac{1}{2}y_i^2 \exp(2g(x_i))g''(x)(x_i - x)^2 \right] \\ &\quad \times K_{h_2}(x_i - x) + o_p(1).\end{aligned}$$

Note that

$$\begin{aligned}&\frac{1}{2}g''(x)\tau_n h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i))(x_i - x)^2 K_{h_2}(x_i - x) \\ &= \frac{1}{2}g''(x)\tau_n h_2 \sum_{i=1}^n [f^2(x_i) + \varepsilon_i^2](x_i - x)^2 K_{h_2}(x_i - x) \\ &\quad + g''(x)\tau_n h_2 \sum_{i=1}^n f(x_i)\varepsilon_i(x_i - x)^2 K_{h_2}(x_i - x).\end{aligned}$$

By Fan & Gijbels (1996), we have

$$\begin{aligned}\sum_{i=1}^n [f^2(x_i) + \varepsilon_i^2](x_i - x)^2 K_{h_2}(x_i - x) &= nh_2^2 \mu_2 [f^2(x) + 1]\varphi(x) + o_p(nh_2^2), \\ \sum_{i=1}^n f(x_i)\varepsilon_i(x_i - x)^2 K_{h_2}(x_i - x) &= O_p\left(\sqrt{nh_2^3}\right).\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\frac{1}{2}g''(x)\tau_n h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i))(x_i - x)^2 K_{h_2}(x_i - x) \\ &= \tau_n^{-1} \frac{1}{2}g''(x)[f^2(x) + 1]\varphi(x)\mu_2 h_2^2 + o_p(\tau_n^{-1}h_2^2) + O_p(h_2^2) \\ &= \tau_n^{-1} \frac{1}{2}g''(x)[f^2(x) + 1]\varphi(x)\mu_2 h_2^2 + o_p(\tau_n^{-1}h_2^2).\end{aligned}$$

Then

$$\begin{aligned}&\frac{1}{2}g''(x)\tau_n h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i))(x_i - x)^2 K_{h_2}(x_i - x) \\ &= \sqrt{nh_2} \frac{1}{2}g''(x)[f^2(x) + 1]\varphi(x)\mu_2 h_2^2 + o_p\left(\sqrt{nh_2^5}\right)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left\{\tau_n h_2 \sum_{i=1}^n [1 - y_i \exp(g(x_i))\varepsilon_i] K_{h_2}(x_i - x)\right\} &= 0, \\ \text{Var}\left\{\tau_n h_2 \sum_{i=1}^n [1 - y_i \exp(g(x_i))\varepsilon_i] K_{h_2}(x_i - x)\right\} &= \varpi(x)\varphi(x)v_0 + o(1),\end{aligned}$$

where $\varpi(x) = E[(\varepsilon^2 - 1)^2] + f^2(x)$. This establishes the asymptotic distribution of the estimator $\hat{g}(x)$.

Next we shall verify that $R_{nj}(\gamma^*) = o_p(1)$ for each γ^* and $j = 1, \dots, 6$. By the Taylor expansion

$$\begin{aligned} r_n(x_i, \gamma^*) &= z_i^T \xi(x) - g(x_i) + \frac{1}{2} g''(x)(x_i - x)^2 + \frac{1}{2} \exp(\zeta_{i1}) [\tau_n z_i^T \gamma^* + z_i^T \xi(x) - g(x_i)]^2 \\ &= -\frac{1}{6} g'''(\zeta_{i2})(x_i - x)^3 + \frac{1}{2} \exp(\zeta_{i1}) \left[\tau_n z_i^T \gamma^* - \frac{1}{2} g''(\zeta_{i3})(x_i - x)^2 \right]^2, \end{aligned}$$

where ζ_{i1} lies between 0 and $\tau_n z_i^T \gamma^* + z_i^T \xi(x) - g(x_i)$, and ζ_{i2}, ζ_{i3} lie between x and x_i .

Since $K(\cdot)$ has a bounded support, there exists a constant $C_1 > 0$ such that $K(u) = 0$ for any $|u| > C_1$. Let $N_x = \{i : |x_i - x| \leq C_1 h_2\}$. Since $g'''(\cdot)$ is continuous in a neighbourhood of x , when n is sufficiently large, there exists a constant $C_2 > 0$ such that

$$\max_{i \in N_x} |g'''(\zeta_{i2})| \leq C_2, \quad \max_{i \in N_x} |g''(\zeta_{i2})| \leq C_2.$$

We can prove that

$$\max_{i \in N_x} |\zeta_{i1}| \leq \max_{i \in N_x} [\tau_n |z_i^T \gamma^*| + |z_i^T \xi(x) - g(x_i)|] = O_p(\tau_n + h_2^2) = o_p(1),$$

and then $\max_{i \in N_x} \exp(\zeta_{i1}) = O_p(1)$. Therefore, we have $\max_{i \in N_x} |r_n(x_i, \gamma^*)| = O_p(h_2^3 + \tau_n^2)$. Therefore

$$\begin{aligned} R_{n1}(\gamma^*) &= \frac{1}{2} h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) r_n^2(x_i, \gamma^*) K_{h_2}(x_i - x) \\ &\leq \max_{i \in N_x} r_n^2(x_i, \gamma^*) h_2 \sum_{i=1}^n y_i^2 \exp(2g(x_i)) K_{h_2}(x_i - x) \\ &= O_p\left(h_2^6 + \frac{1}{n^2 h_2^2}\right) O_p(nh_2) = O_p\left(nh_2^7 + \frac{1}{nh_2}\right). \end{aligned}$$

Similarly

$$R_{n2}(\gamma^*) = O_p\left(\sqrt{nh_2^7} + \frac{1}{\sqrt{nh_2}}\right), \quad R_{n3}(\gamma^*) = O_p(nh_2^6 + h_2^2).$$

Since

$$E[R_{n4}^2(\gamma^*)|X] = \sum_{i=1}^n y_i^2 \exp(2g(x_i)) r_n^2(x_i, \gamma^*) K_{h_2}^2(x_i - x) = O_p\left(nh_2^7 + \frac{1}{nh_2}\right),$$

we have $R_{n4}(\gamma^*) = O_p\left(\sqrt{nh_2^7} + \frac{1}{\sqrt{nh_2}}\right)$. Based on the conditions that $h_2 \rightarrow 0$, $nh_2 \rightarrow \infty$ and $nh_2^6 \rightarrow 0$, we conclude that $R_{nj}(\gamma^*) = o_p(1)$ for each γ^* and $j = 1, 2, 3, 4$. Next, we observe that

$$\begin{aligned} |R_{n5}(\gamma^*)| &\leq \sup_x |\hat{f}(x) - f(x)| h_2 \tau_n \sum_{i=1}^n |y_i \exp\{g(x_i)\}| K_{h_2}(x_i - x) |z_i^T \gamma^*| \\ &= O_p(\sqrt{nh_2}) \sup_x |\hat{f}(x) - f(x)| \end{aligned}$$

and

$$\begin{aligned} |R_{n6}(\gamma^*)| &\leq \max_{i \in N_x} |r_n(x_i, \gamma^*)| \sup_x |\hat{f}(x) - f(x)| h_2 \sum_{i=1}^n |y_i \exp\{g(x_i)\}| K_{h_2}(x_i - x) \\ &= O_p(h_2^3 + \tau_n^2) O_p(nh_2) \sup_x |\hat{f}(x) - f(x)| = O_p(nh_2^4 + 1) \sup_x |\hat{f}(x) - f(x)|. \end{aligned}$$

When $\sup_x |\hat{f}(x) - f(x)| = o_p\left(\min\left\{\frac{1}{\sqrt{nh_2}}, \frac{1}{nh_2^4}\right\}\right)$, we have $R_{n5}(\gamma^*) = R_{n6}(\gamma^*) = o_p(1)$.

Note that

$$\hat{f}(x) = [1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x \hat{Y},$$

where

$$A_x = \begin{pmatrix} 1 & \frac{x_1 - x}{h_1} \\ \vdots & \vdots \\ 1 & \frac{x_n - x}{h_1} \end{pmatrix}, \quad W_x = \begin{pmatrix} K_{h_1}(x_1 - x) & & \\ & \ddots & \\ & & K_{h_1}(x_n - x) \end{pmatrix}, \quad \hat{Y} = \begin{pmatrix} y_1 \exp(\hat{g}(x_1)) \\ \vdots \\ y_n \exp(\hat{g}(x_n)) \end{pmatrix},$$

and $\beta = (f(x), h_1 f'(x))^T$. Then

$$\begin{aligned} \hat{f}(x) - f(x) &= [1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x (\hat{Y} - A_x \beta) \\ &=: \mathcal{A}_{n1} + \mathcal{A}_{n2} + \mathcal{A}_{n3}, \end{aligned}$$

where

$$\begin{aligned} \mathcal{A}_{n1} &= [1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x [f(x_i) - f(x) - f'(x)(x_i - x)]_{1 \leq i \leq n}, \\ \mathcal{A}_{n2} &= [1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x (\varepsilon_i)_{1 \leq i \leq n}, \\ \mathcal{A}_{n3} &= [1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x \{y_i [\exp(\hat{g}(x_i)) - \exp(g(x_i))]\}_{1 \leq i \leq n}. \end{aligned}$$

In addition, note that

$$[1, 0](A_x^T W_x A_x)^{-1} A_x^T W_x = \frac{1}{n\varphi(x)} [K_{h_1}(x_1 - x), \dots, K_{h_1}(x_n - x)].$$

Hence

$$\begin{aligned} \mathcal{A}_{n1} &= \frac{1 + o_p(1)}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) [f(x_i) - f(x) - f'(x)(x_i - x)] \\ &= \frac{1 + o_p(1)}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \left[\frac{1}{2} f''(x)(x_i - x)^2 + o_p((x_i - x)^2) \right] \\ &= \frac{1}{2} f''(x) \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x)(x_i - x)^2 (1 + o_p(1)) \\ &= \frac{1}{2} f''(x) \mu_2 h_1^2 + o_p(h_1^2), \end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{n2} &= \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i (1 + o_p(1)) \\
&= \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i + o_p(1) \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i \\
&= \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i + o_p(1) \sqrt{E \left[\frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i \right]^2} \\
&= \frac{1}{n\varphi(x)} \sum_{i=1}^n K_h(X_i - x) \varepsilon_i + o_p\left(\frac{1}{\sqrt{nh_1}}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{A}_{n3} &= \frac{1 + o_p(1)}{n\varphi(x)} \sum_{i=1}^n K_{h_1}(x_i - x) y_i [\exp(\hat{g}(x_i)) - \exp(g(x_i))] \\
&= \frac{1 + o_p(1)}{n\varphi(x)} \sum_{i=1}^n K_{h_1}(x_i - x) y_i \exp(g(x_i)) [\exp(\hat{g}(x_i) - g(x_i)) - 1] \\
&= \frac{1}{n\varphi(x)} \sum_{i=1}^n K_{h_1}(x_i - x) y_i \exp(g(x_i)) [\hat{g}(x_i) - g(x_i)] (1 + o_p(1)). \tag{A.3}
\end{aligned}$$

By the proof of the estimator $\hat{g}(x)$, $\hat{g}(x_i) - g(x_i)$ can be approximated by

$$\begin{aligned}
&\frac{1}{n[f^2(x_i) + 1]\varphi(x_i)} \sum_{j=1}^n \left[1 - y_j \exp(g(x_j)) \varepsilon_j + \frac{1}{2} y_j^2 \exp(2g(x_j)) g''(x_i) (x_j - x_i)^2 \right] \\
&\quad \times K_{h_2}(x_j - x_i) + o_p(1). \tag{A.4}
\end{aligned}$$

Substituting (A.4) into (A.3), we have

$$\begin{aligned}
\mathcal{A}_{n3} &\approx \frac{1}{2n^2\varphi(x)} \sum_{i=1}^n \sum_{j: j \neq i}^n \frac{y_i \exp(g(x_i)) g''(x_i)}{[f^2(x_i) + 1]\varphi(x_i)} y_j^2 \exp(2g(x_j)) (x_j - x_i)^2 K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \\
&\quad + \frac{1}{n^2\varphi(x)} \sum_{i=1}^n \sum_{j=1}^n \frac{y_i \exp(g(x_i))}{[f^2(x_i) + 1]\varphi(x_i)} [1 - y_j \exp(g(x_j)) \varepsilon_j] K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \\
&=: D_{j1} + D_{j2}.
\end{aligned}$$

Since x_i and x_j ($i \neq j$) are independent, and $y_i \exp\{g(x_i)\} = f(x_i) + \varepsilon_i$, we have

$$\begin{aligned}
E(D_{j1}) &= \frac{n(n-1)}{2n^2\varphi(x)} E \left\{ \frac{y_i \exp(g(x_i)) g''(x_i)}{[f^2(x_i) + 1]\varphi(x_i)} y_j^2 \exp(2g(x_j)) (x_j - x_i)^2 K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \right\} \\
&= \frac{n-1}{2n\varphi(x)} E \left\{ \frac{f(x_i) g''(x_i)}{[f^2(x_i) + 1]\varphi(x_i)} [f^2(x_j) + \varepsilon_j^2] (x_j - x_i)^2 K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{n-1}{2n\varphi(x)} \iint \frac{f(t_1)g''(t_1)}{[f^2(t_1)+1]\varphi(t_1)} [f^2(t_2)+1](t_2-t_1)^2 K_{h_1}(t_1-x) K_{h_2}(t_2-t_1) \\
&\quad \times \varphi(t_1)\varphi(t_2) dt_1 dt_2 \\
&= \frac{n-1}{2n\varphi(x)} \iint F(t_1)G(t_2)(t_2-t_1)^2 K_{h_1}(t_1-x) K_{h_2}(t_2-t_1) dt_1 dt_2,
\end{aligned}$$

where

$$F(t_1) = \frac{f(t_1)g''(t_1)\varphi(t_1)}{[f^2(t_1)+1]\varphi(t_1)}, \quad G(t_2) = [f^2(t_2)+1]\varphi(t_2).$$

Let $u_1 = (t_1 - x)/h_1$ and $u_2 = (t_2 - t_1)/h_2$; then, $t_1 = x + u_1 h_1$ and $t_2 = x + u_1 h_1 + u_2 h_2$. Therefore

$$\begin{aligned}
&\iint F(t_1)G(t_2)(t_2-t_1)^2 K_{h_1}(t_1-x) K_{h_2}(t_2-t_1) dt_1 dt_2 \\
&= h_2^2 \iint F(x + u_1 h_1) G(x + u_1 h_1 + u_2 h_2) u^2 K(u_1) K(u_2) du_1 du_2 \\
&= h_2^2 F(x) G(x) \mu_2 + o(h_2^2) \\
&= f(x)g''(x)\varphi(x)\mu_2 h_2^2 + o(h_2^2).
\end{aligned}$$

It follows that

$$E(D_{j1}) = \frac{n-1}{2n} f(x)g''(x)\mu_2 h_2^2 + o(h_2^2) = \frac{1}{2} f(x)g''(x)\mu_2 h_2^2 + o(h_2^2).$$

On the other hand

$$\begin{aligned}
D_{j2} &= \frac{K(0)}{n^2 h_2 \varphi(x)} \sum_{i=1}^n \frac{y_i \exp(g(x_i))}{[f^2(x_i)+1]\varphi(x_i)} [1 - y_i \exp(g(x_i))\varepsilon_i] K_{h_1}(x_i - x) \\
&\quad + \frac{1}{n^2 \varphi(x)} \sum_{i=1}^n \sum_{j:j \neq i} \frac{y_i g(x_i)}{[f^2(x_i)+1]\varphi(x_i)} [1 - y_j \exp(g(x_j))\varepsilon_j] K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \\
&=: D_{j21} + D_{j22}.
\end{aligned}$$

Since $y_i \exp(g(x_i)) = f(x_i) + \varepsilon_i$

$$\begin{aligned}
E(D_{j21}) &= \frac{K(0)}{n h_2 \varphi(x)} E \left\{ \frac{y_i \exp(g(x_i))}{[f^2(x_i)+1]\varphi(x_i)} [1 - y_i \exp(g(x_i))\varepsilon_i] K_{h_1}(x_i - x) \right\} \\
&= \frac{-K(0)}{n h_2 \varphi(x)} E \left\{ \frac{E(\varepsilon^3) + f(x_i)}{[f^2(x_i)+1]\varphi(x_i)} K_{h_1}(x_i - x) \right\} \\
&= \frac{-K(0)}{n h_2 \varphi(x)} \int \frac{E(\varepsilon^3) + f(t)}{[f^2(t)+1]\varphi(t)} K_{h_1}(t - x) \varphi(t) dt \\
&= O\left(\frac{1}{n h_2}\right).
\end{aligned}$$

Next we consider D_{j22} . Since $y_j \exp(g(x_j)) = f(x_j) + \varepsilon_j$

$$\begin{aligned} D_{j22} &= \frac{1}{n^2 \varphi(x)} \sum_{i=1}^n \sum_{j:j \neq i} \frac{y_i \exp(g(x_i))}{[f^2(x_i) + 1] \varphi(x_i)} [1 - y_j \exp(g(x_j)) \varepsilon_j] K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \\ &= \frac{1}{n^2 \varphi(x)} \sum_{i=1}^n \sum_{j:j \neq i} \frac{y_i \exp(g(x_i))}{[f^2(x_i) + 1] \varphi(x_i)} [1 - \varepsilon_j^2 - f(x_j) \varepsilon_j] K_{h_1}(x_i - x) K_{h_2}(x_j - x_i) \\ &= \frac{1}{n^2 \varphi(x)} \sum_{i=1}^n \frac{y_i \exp(g(x_i))}{[f^2(x_i) + 1] \varphi(x_i)} K_{h_1}(x_i - x) \sum_{j:j \neq i} [1 - \varepsilon_j^2 - f(x_j) \varepsilon_j] K_{h_2}(x_j - x_i). \end{aligned}$$

Since $j \neq i$, then ε_j is independent of x_i and x_j . By $E(1 - \varepsilon_j^2) = 0$ and $E(\varepsilon_j) = 0$, we have $E(D_{j22}) = 0$ and

$$\begin{aligned} E(D_{j22}^2) &= \frac{1}{n^4 \varphi^2(x)} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j:j \neq i_1, i_2} E \left\{ \frac{y_{i_1} \exp(g(x_{i_1}))}{[f^2(x_{i_1}) + 1] \varphi(x_{i_1})} \frac{y_{i_2} \exp(g(x_{i_2}))}{[f^2(x_{i_2}) + 1] \varphi(x_{i_2})} \right. \\ &\quad \times [1 - y_j \exp(g(x_j)) \varepsilon_j]^2 K_{h_1}(x_{i_1} - x) K_{h_2}(x_j - x_{i_1}) K_{h_1}(x_{i_2} - x) K_{h_2}(x_j - x_{i_2}) \left. \right\} \\ &= O\left(\frac{1}{nh_2}\right). \end{aligned}$$

It follows that $D_{j22} = O_P(1/\sqrt{nh_2})$, and moreover $D_{j2} = O_P(1/\sqrt{nh_2})$. Therefore

$$\hat{f}(x) - f(x) = \frac{1}{2} f''(x) \mu_2 h_1^2 + \frac{1}{n \varphi(x)} \sum_{i=1}^n K_{h_1}(x_i - x) \varepsilon_i + o_p\left(h_1^2 + \frac{1}{\sqrt{nh_1}}\right) + O_p\left(h_2^2 + \frac{1}{\sqrt{nh_2}}\right).$$

Finally, given that $nh_1 h_2^4 \rightarrow 0$ and $h_1/h_2 \rightarrow 0$, the proof of Theorem 1 follows from the central limit theorem.

Proof of the positive definiteness of Hessian matrix. To show that the local minimizer of (4) exists, we derive the Hessian matrix of (4) and prove that it is positive definite in the neighbourhood of the true value. The detailed proof is given as follows.

Let $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^T$, $\mathbf{z}_i = (1, (x_i - x)/h_2)^T$ and $k_i = h_2^{-1} K(x_i - x/h_2)$. The objective function (4) can then be represented as

$$Q(\boldsymbol{\gamma}) = \sum_{i=1}^n \left\{ \frac{1}{2} [y_i \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) - f_-(x_i)]^2 - \mathbf{z}_i^T \boldsymbol{\gamma} \right\} K_{h_2}(x_i - x).$$

The Hessian matrix is

$$H = \frac{\partial^2 Q(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} = \sum_{i=1}^n k_i \{ 2y_i^2 \exp(2\mathbf{z}_i^T \boldsymbol{\gamma}) - y_i \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) f(x_i) \} \mathbf{z}_i \mathbf{z}_i^T.$$

To show that the objective function (4) has a local minimizer, we only need to prove that H is positive definite in expectation. Note that

$$\begin{aligned}
2y_i^2 \exp(2g(x_i)) - y_i \exp(g(x_i))f(x_i) &= \exp(g(x_i))\{2y_i^2 \exp(g(x_i)) - y_i f(x_i)\} \\
&= \frac{\mu(x_i) + \sigma(x_i)\varepsilon_i}{\sigma(x_i)} \times \frac{\mu(x_i) + 2\sigma(x_i)\varepsilon_i}{\sigma(x_i)} \\
&= 2\varepsilon_i^2 + 3\frac{\mu(x_i)}{\sigma(x_i)}\varepsilon_i + \frac{\mu^2(x_i)}{\sigma^2(x_i)}
\end{aligned}$$

with $f(x) = \mu(x)/\sigma(x)$, $g(x) = -\log(\sigma(x))$ and $y = f(x) + \sigma(x)\varepsilon$. We have

$$\begin{aligned}
E(H|x) &= \sum_{i=1}^n k_i z_i z_i^T E \left[2\varepsilon_i^2 + 3\frac{\mu(x_i)}{\sigma(x_i)}\varepsilon_i + \frac{\mu^2(x_i)}{\sigma^2(x_i)} \middle| x_i \right] \\
&= \sum_{i=1}^n k_i z_i z_i^T \left\{ 2E[\varepsilon_i^2|x_i] + 3\frac{\mu(x_i)}{\sigma(x_i)}E[\varepsilon_i|x_i] + \frac{\mu^2(x_i)}{\sigma^2(x_i)} \right\} \\
&= \sum_{i=1}^n k_i z_i z_i^T \left(2 + \frac{\mu^2(x_i)}{\sigma^2(x_i)} \right).
\end{aligned}$$

Hence, it is easy to show that $E(H|x)$ is positive definite. ■

Received 15 February 2020

Accepted 22 May 2021