



MONASH
BUSINESS
SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

**Nonparametric Localized Bandwidth Selection for
Kernel Density Estimation**

Tingting Cheng, Jiti Gao, and Xibin Zhang

February 2016

Working Paper 07/16
(revised 27/14)

Nonparametric Localized Bandwidth Selection for Kernel Density Estimation

TINGTING CHENG

School of Finance, Nankai University

JITI GAO* and XIBIN ZHANG

Department of Econometrics and Business Statistics, Monash University

February 29, 2016

Abstract

As conventional cross-validation bandwidth selection methods do not work properly in the situation where the data are serially dependent time series, alternative bandwidth selection methods are necessary. In recent years, Bayesian based methods for global bandwidth selection have been studied. Our experience shows that a global bandwidth is however less suitable than a localized bandwidth in kernel density estimation based on serially dependent time series data. Nonetheless, a difficult issue is how we can consistently estimate a localized bandwidth. This paper presents a nonparametric localized bandwidth estimator, for which we establish a completely new asymptotic theory. Applications of this new bandwidth estimator to the kernel density estimation of Eurodollar deposit rate and the S&P 500 daily return demonstrate the effectiveness and competitiveness of the proposed localized bandwidth.

KEYWORDS: Density estimation; localized bandwidth; GARCH model.

JEL Classification: C13, C14, C21

**Corresponding author:* Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Telephone: 61 3 99031675. Fax: 61 3 99032007. Email: Jiti.Gao@monash.edu.

1 Introduction

Kernel density estimation is an important tool for exploring the distributional properties of a random variable in an unknown population (Silverman, 1986; Li and Racine, 2007). Such kernel estimation techniques have been widely used in many application studies (see for example, Aït-Sahalia, 1996; Bithell, 1990; Seaman and Powell, 1996; Elgammal, Duraiswami, Harwood, and Davis, 2002). It is known that the performance of a kernel density estimator is mainly determined by its bandwidth. This paper aims to present a nonparametric estimation method for localized bandwidth selection.

There exists a large body of literature on bandwidth selection for kernel density estimation. Jones, Marron, and Sheather (1996) presented surveys on bandwidth selection methods, including the rule-of-thumb, the least-square cross-validation, the biased cross-validation, the plug-in method and a smoothed bootstrapping method. Fan and Yao (2003) discussed bandwidth selection methods for nonparametric density estimation in Section 5.4 of Chapter 5, including the normal reference rule, the cross-validation and the plug-in method. An overview on kernel density estimation and relevant bandwidth selectors can also be found in Sheather (2004). Recently, Heidenreich, Schindler, and Sperlich (2013) compared the existing bandwidth methods on a set of designs and suggested a mixture of the simple plug-in and cross-validation methods. There also exist some investigations on Bayesian sampling approaches to bandwidth estimation (see Zhang, King, and Hyndman, 2006, among others). Besides bandwidth selection for unconditional density estimation, there are also some studies on bandwidth selection for nonparametric conditional mean, conditional density function and conditional distribution function estimation (see for example, Hall, Li, and Racine, 2007; Hall, Racine, and Li, 2004; Li, Lin, and Racine, 2013). However, all these methods mainly aim to choose a global bandwidth, with which a kernel estimator is likely to simultaneously under- and over-smooth the underlying density function in different areas on its support (Sain and Scott, 1996). The recent development on kernel density estimation with localized bandwidth suggests that a small bandwidth should be assigned to the observations in the high-density region and a large bandwidth should be assigned to those in the low-density region. The localized kernel density estimation method attempts to solve this problem by allowing for different bandwidths in different regions on the support of the underlying density (see for example, Brewer, 2000; Gangopadhyay and Cheung, 2002; Kulasekera and Padgett, 2006).

In this paper, we focus on the localized bandwidth estimation for unconditional density esti-

mation. We construct a localized bandwidth through the conditional density of the bandwidth parameter given a data set X_1, X_2, \dots, X_n observed from an unknown density $f(x)$. The resulting bandwidth estimate $h(x)$ is a function of the density point x , which is the value where the density estimator is calculated. The resulting density estimator is given by

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h(x)} K\left(\frac{X_i - x}{h(x)}\right),$$

where $K(\cdot)$ is the kernel function.

[Gangopadhyay and Cheung \(2002\)](#) and [de Lima and Atuncar \(2011\)](#) considered this type of localized bandwidth and derived a closed-form bandwidth estimate, where the prior of bandwidth belongs to a conjugate family of the likelihood.*

However, to the best of our knowledge, there is no asymptotic theory for the above bandwidth estimation method in the current literature. This paper aims to fill this gap and presents an asymptotic theory for the bandwidth estimation method. In particular, we propose a more general nonparametric localized bandwidth estimation method and then establish an asymptotically normal distribution with root- n rate of convergence. [Härdle, Hall, and Marron \(1988\)](#) showed that $n^{3/10}(\hat{h}_{cv} - h_0)$ converges in distribution to a normal random variable for the case where $h_0 \sim n^{-1/5}$ and \hat{h}_{cv} was chosen based on the conventional cross-validation method. In this work, we considerably improve this by showing that the proposed nonparametric localized bandwidth can achieve a much faster rate of convergence at an order of \sqrt{n} . Meanwhile, we will demonstrate that the nonparametric localized bandwidth estimation method has satisfactory finite-sample properties and better out-of-sample performance than those based on its natural competitors when estimating the density of Eurodollar daily deposit rate and the density of S&P 500 daily return.

The main contributions of this paper are summarized as follows.

- (i) We propose a nonparametric localized bandwidth selection method and develop an asymptotic theory for the proposed nonparametric localized bandwidth selection method.
- (ii) We conduct simulation studies to examine the finite-sample performance of our proposed nonparametric localized bandwidth selection, as well as the performance of the resulting kernel density estimators.

*In these papers, as well as in our approach to bandwidth selection, bandwidth is treated as a parameter, whose posterior is employed to derive a posterior estimate of bandwidth. Therefore, it should be stressed here that this approach is different from the Bayesian nonparametric approach, which has been extensively investigated in the literature (see for example, [Lo, 1984](#); [Ghosh and Ramamoorthi, 2003](#); [Hjort, Holmes, Müller, and Walker, 2010](#)).

(iii) We apply the proposed nonparametric localized bandwidth selection method to the kernel density estimation of Eurodollar daily deposit rate and the S&P 500 daily return.

The rest of this paper is organized as follows. Section 2 briefly describes the construction of a nonparametric localized bandwidth selection method. In Section 3, we investigate the asymptotic properties of the proposed nonparametric localized bandwidth selection method. Section 4 presents Monte Carlo simulation studies to evaluate the performance of the nonparametric localized bandwidth selection method. In Sections 5 and 6, two empirical examples are presented to illustrate the application of the proposed nonparametric localized bandwidth selection method. Section 7 concludes the paper. The proofs of the main results are given in Appendix.

2 Nonparametric localized bandwidth

Let X denote an univariate random variable with a density $f(x)$, which is approximated by

$$f(x|h) = f * K_h(x) = \int f(y)K_h(y-x)dy = \mathbb{E}_1 [K_h(X-x)], \quad (1)$$

where $K_h(\cdot) = K(\cdot/h)/h$, and $K(\cdot)$ is the kernel function, and $\mathbb{E}_1[\cdot]$ denotes the expectation under the conditional distribution of X given h . In the rest of this paper, we use $\mathbb{E}_2[\cdot]$ to denote the expectation under the marginal distribution of h and $\mathbb{E}[\cdot]$ denote the full expectation under the joint distribution of (X, h) . In other words, $\mathbb{E}[\cdot] = \mathbb{E}_2(\mathbb{E}_1[\cdot])$.

In fact, $f(x|h)$ can be considered as the density of $X + \xi$, where ξ is a random variable with mean 0 and density $K_h(\cdot)$. When h is small, the difference between $f(x|h)$ and $f(x)$ is practically negligible. Let $\pi(h)$ denote a density of h . We define the conditional density of h given $X = x$ as

$$\pi(h|x) = \frac{f(x|h)\pi(h)}{\int f(x|h)\pi(h)dh}.$$

An estimate of h is the conditional mean and is given by

$$h_0(x) = \int h\pi(h|x)dh = \frac{\int hf(x|h)\pi(h)dh}{\int f(x|h)\pi(h)dh} \equiv \frac{q(x)}{p(x)}, \quad (2)$$

where $q(x) = \int hf(x|h)\pi(h)dh$ and $p(x) = \int f(x|h)\pi(h)dh$.

The resulting kernel density estimate is given by

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_0(x)} K\left(\frac{X_i - x}{h_0(x)}\right).$$

In the rest of this paper, we will focus on the issue of how to estimate $h_0(x)$ nonparametrically. As mentioned in Remark 3 that follows Theorem 2 in Section 3, we also discuss the case where $h_0(x) = a_n b_0(x)$ and the corresponding estimation of $b_0(x)$, where $a_n \rightarrow 0$ as $n \rightarrow \infty$, and $b_0(x)$ is treated as an unknown function of x .

Any inference or computation based on $\pi(h|x)$ cannot be directly conducted because $f(x|h)$ is unknown. When a random sample, denoted as $\{X_i : i = 1, 2, \dots, n\}$, is observed from X , we can approximate $f(x|h)$ by its sample mean:

$$\hat{f}(x|h) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \quad (3)$$

which is known as the kernel estimator of $f(x)$. Therefore, we can estimate $\pi(h|x)$ by

$$\hat{\pi}(h|x) = \frac{\hat{f}(x|h)\pi(h)}{\int \hat{f}(x|h)\pi(h)dh}.$$

Thus, we estimate $h_0(x)$ by

$$h_n(x) = \frac{\int h \hat{f}(x|h)\pi(h)dh}{\int \hat{f}(x|h)\pi(h)dh} \equiv \frac{q_n(x)}{p_n(x)}, \quad (4)$$

which we call the nonparametric localized bandwidth (NLB) estimator of $h_0(x)$, where $q_n(x) = \int h \hat{f}(x|h)\pi(h)dh$ and $p_n(x) = \int \hat{f}(x|h)\pi(h)dh$. We will investigate the asymptotic properties of $h_n(x)$ in the next section.

The resulting kernel density estimator is

$$\hat{f}_n^*(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n(x)} K\left(\frac{X_i - x}{h_n(x)}\right), \quad (5)$$

where $h_n(x)$ is the bandwidth. In the next section, we establish asymptotic properties for the kernel density estimator $\hat{f}_n^*(x)$.

3 Asymptotic theory

The following conditions are required to establish some asymptotic results, although some of them might not be the weakest possible.

Assumption 1. (i) $\{X_i\}$ is a strictly stationary and α -mixing process with mixing coefficient $\alpha(\cdot)$ satisfying $\sum_{j=1}^{\infty} \alpha(j)^{1-2/\delta} < \infty$; and (ii) $\mathbb{E}(|X_i|^\delta) < \infty$ for some constant $\delta > 2$.

Assumption 2. $K(\cdot)$ is a continuous, bounded and symmetric probability density function with $\int u^2 K(u) du < \infty$.

Assumption 3. $\pi(h)$ satisfies

$$\int h^{1-\delta} |hp(x) - q(x)| \Lambda(x, h) \pi(h) dh < \infty \text{ and } \int |hp(x) - q(x)| \pi(h) dh < \infty$$

for the same $\delta > 2$ as in Assumption 1(i), in which $\Lambda(x, h) = \int K^\delta(v) f(x + vh) dv$.

Assumption 4. (i) $f(x)$ is twice differentiable, and the second-order derivative, $f^{(2)}(x)$, is continuous. (ii) There exists a nonnegative function $L(\cdot)$ such that $|K(y) - K(x)| \leq L(x) |y - x|$ for $y \in \mathcal{S}(x, \epsilon) = \{y : |y - x| \leq \epsilon\}$ for some $\epsilon > 0$, in which $L(\cdot)$ satisfies $\int |v|^j L(v) dv < \infty$, for $j = 0, 1$ and 2 .

Assumption 1 is a standard requirement for moments and the mixing coefficient for an α -mixing time series. Assumption 2 is a standard assumption for kernel functions, and the commonly used Gaussian and uniform kernel functions satisfy this assumption. Assumption 3 imposes a requirement on the density $\pi(h)$. Actually as long as $\mathbb{E}[h^{2-\delta}]$ exists, Assumption 3 is satisfied. Assumption 4 (i) imposes smoothness constraints on the unknown density function, which is commonly used in the literature. Assumption 4 (ii) imposes the Lipschitz condition for kernel functions, and this assumption is easily satisfied by many commonly used kernel functions.

Before presenting the main theoretical results, we introduce the following notations:

$$\begin{aligned}
f_u(x) &= \frac{1}{u} \int K\left(\frac{y-x}{u}\right) f(y) dy, \quad f_{uv}(x) = \frac{1}{uv} \int K\left(\frac{y-x}{u}\right) K\left(\frac{y-x}{v}\right) f(y) dy, \\
R_{uv}(x) &= f_{uv}(x) - f_u(x)f_v(x), \quad g_{uv,s}(x) = \frac{1}{uv} \int K\left(\frac{y-x}{u}\right) K\left(\frac{z-x}{v}\right) f_s(y,z) dy dz, \\
G_{uv,s}(x) &= g_{uv,s}(x) - f_u(x)f_v(x), \quad m_{uv,s}(x_a, x_b) = \frac{1}{uv} \int K\left(\frac{y-x_a}{u}\right) K\left(\frac{z-x_b}{v}\right) f_s(y,z) dy dz, \\
S_{uv,s}(x_a, x_b) &= m_{uv,s}(x_a, x_b) - f_u(x_a)f_v(x_b), \quad \int y^2 K(y) dy = \mu_2(K) > 0, \quad R(K) = \int K^2(y) dy,
\end{aligned}$$

where $s = |i - j|$, and $f_{i-j}(y, z)$ denotes the joint density of (X_i, X_j) .

We now present the asymptotic properties of the NLB estimator $h_n(x)$ and its resulting kernel density estimator $\widehat{f}_n^*(x)$. The proofs for Theorems 1–3 are given in Appendix below.

Theorem 1. *Under Assumptions 1–3, we have as $n \rightarrow \infty$,*

$$\sqrt{n}(h_n(x) - h_0(x)) \rightarrow_D \mathcal{N}(0, \Sigma_0(x)), \quad (6)$$

where $\Sigma_0(x) = Q^{-2}(x)\Sigma_L(x)$ and $\Sigma_L(x) = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j)$, in which $Q(x) = p^2(x)$ and

$$\begin{aligned}
\gamma(j) &= \iint [up(x) - q(x)][vp(x) - q(x)] G_{uv,j}(x) \pi(u)\pi(v) dudv, \\
\gamma(0) &= \iint [up(x) - q(x)][vp(x) - q(x)] R_{uv}(x) \pi(u)\pi(v) dudv.
\end{aligned}$$

Remark 1. This theorem shows that $h_n(x)$ is asymptotically normal and is a consistent estimator of $h_0(x)$ with root- n rate of convergence. This property holds for the localized bandwidth at each point x . This fast rate of convergence is not surprising, because [Härdle, Hall, and Marron \(1992\)](#) showed that under certain conditions, the root- n rate of convergence is achievable when a double-smoothing technique is used.

Theorem 2. *Let $N \geq 2$ be a finite integer and let x_1, x_2, \dots, x_N be fixed points. Under Assumptions 1–3, we have as $n \rightarrow \infty$,*

$$(\sqrt{n}(h_n(x_1) - h_0(x_1)), \dots, \sqrt{n}(h_n(x_N) - h_0(x_N))) \rightarrow_D \mathcal{N}(0, \Sigma_N), \quad (7)$$

where $\Sigma_{N,aa} = \Sigma_0(x_a)$, $\Sigma_{N,ab} = Q^{-1}(x_a)\Sigma_v(x_a, x_b)Q^{-1}(x_b)$, $\Sigma_v(x_a, x_b) = \gamma_{ab}(0) + 2\sum_{s=1}^{\infty} \gamma_{ab}(s)$, $\gamma_{ab}(s) = \mathbb{E}[V_i(x_a)V_j(x_b)] = \iint [up(x_a) - q(x_a)][vp(x_b) - q(x_b)] S_{uv,s}(x_a, x_b) \pi(u)\pi(v) dudv$, $\gamma_{ab}(0) = \mathbb{E}[V_i(x_a)V_i(x_b)] = \iint [up(x_a) - q(x_a)][vp(x_b) - q(x_b)] S_{uv,0}(x_a, x_b) \pi(u)\pi(v) dudv$ and

$a, b \in \{1, 2, \dots, N\}$.

Remark 2. This theorem shows that when we consider localized bandwidth around more than one point, the corresponding localized bandwidths are jointly asymptotically normal.

Remark 3. The discussion in Theorems 1 and 2 for the estimators of $h_0(x)$ can be extended to the case where $h = a_n b$, in which $a_n \rightarrow 0$ as $n \rightarrow \infty$, and b is treated as an unknown parameter. In this case, we have

$$h_0(x) = \mathbb{E}_3[h|x] = a_n \mathbb{E}_3[b|x] \equiv a_n b_0(x), \quad (8)$$

where $h_0(x)$ is as defined in (2), and $\mathbb{E}_3[h|x]$ denotes the conditional expectation of h given $X = x$. In this case, we have $b_n(x) = a_n^{-1} h_n(x)$. As a consequence, the convergence rate of $b_n(x) - b_0(x)$ becomes $n^{-\frac{1}{2}} a_n^{-1}$. In the univariate case where $a_n = n^{-\frac{1}{5}}$, the rate of convergence of $b_n(x)$ to $b_0(x)$ reduces to $n^{-\frac{3}{10}}$.

Meanwhile, the rate of $n^{-1/2}$ of $h_n(x) - h_0(x)$ is faster than that of $n^{-3/10}$ of $\hat{h}_{cv} - h_0$ obtained by Härdle et al. (1988) for the case where the bandwidth $h_0 \sim n^{-1/5}$ was treated as an unknown fixed value and selected by the conventional cross-validation method.

Remark 4. For a multivariate kernel density setting, we can still use the proposed NLB method to choose an optimal bandwidth matrix H (see, for example, de Lima and Atuncar (2011)) and then obtain similar asymptotic results to those for the univariate case, although technical details are complicated.

Recall that $\mu_2(K) = \int v^2 K(v) dv$ and $R(K) = \int K^2(v) dv$. We have the following asymptotic normality results.

Theorem 3. *Let Assumptions 1–4 hold. If, in addition, $h_0(x) = a_n b_0(x) \rightarrow 0$ and $nh_0(x) \rightarrow \infty$ as $n \rightarrow \infty$, then for each given x , we have as $n \rightarrow \infty$,*

$$\sqrt{nh_0(x)} \left(\hat{f}_n^*(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \rightarrow_D \mathcal{N}(0, R(K) f(x)).$$

Theorem 3 shows that the conventional normality is achievable for $\hat{f}_n^*(x)$. In Sections 4–6, we evaluate the finite-sample performance and applicability of $h_n(x)$ and $\hat{f}_n^*(x)$.

4 Monte Carlo simulation

In this section, we present several examples to examine the finite-sample performance of NLB. We compare NLB with other bandwidth selection methods, including local cross-validation

proposed by [Hall and Schucany \(1989\)](#), least-square cross-validation discussed by [Rudemo \(1982\)](#) and [Bowman \(1984\)](#) and a Bayesian sampling method proposed by [Zhang, King, and Hyndman \(2006\)](#). For simplicity, we denote local cross-validation as LCV and least-square cross-validation as GCV hereafter.

4.1 Computational aspects

In most situations, a closed form expression of $h_0(x)$ and $h_n(x)$ is not available. We introduce the following Lemmas to approximate $h_0(x)$ and $h_n(x)$. Their proofs are given in the Appendix.

Let

$$\begin{aligned} p_m(x) &= \frac{1}{m} \sum_{i=1}^m f(x + u_i v_i), & q_m(x) &= \frac{1}{m} \sum_{i=1}^m f(x + u_i v_i) v_i, \\ p_{nm}(x) &= \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{v_j} K\left(\frac{X_i - x}{v_j}\right), & q_{nm}(x) &= \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m K\left(\frac{X_i - x}{v_j}\right), \end{aligned} \quad (9)$$

where u_i and v_i , for $i = 1, 2, \dots, m$, are drawn randomly from respectively, $K(u)$ and $\pi(v)$, in which $K(u)$ is the Gaussian kernel and $\pi(v)$ denotes the density of v .

Lemma 1. As $m \rightarrow \infty$, $p_m(x) - p(x) = o_P(1)$ and $q_m(x) - q(x) = o_P(1)$.

Lemma 2. As $m \rightarrow \infty$, $p_{nm}(x) - p_n(x) = o_P(1)$ and $q_{nm}(x) - q_n(x) = o_P(1)$.

Based on Lemmas 1 and 2, we approximate $h_0(x)$ and $h_n(x)$ by $\frac{q_m(x)}{p_m(x)}$ and $\frac{q_{nm}(x)}{p_{nm}(x)}$, respectively.

In this simulation study, we assume that the density of h^2 is the density of an inverse Gamma distribution denoted as $h^2 \sim IG(\alpha, \beta)$ with its probability density function (pdf) expressed as

$$\pi(h^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{h^2}\right)^{\alpha+1} \exp\left\{-\frac{\beta}{h^2}\right\}, \quad (10)$$

where $\theta = (\alpha, \beta)'$ is a vector of two hyperparameters. The density of h is

$$\pi(h) = \frac{2\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{h}\right)^{2\alpha+1} \exp\left\{-\frac{\beta}{h^2}\right\}. \quad (11)$$

4.2 Localized bandwidth

We examine the finite-sample performance of $h_n(x)$ through simulated random samples, where sample sizes are respectively, 250, 750 and 1500, and the number of replications for each sample size is 500. To generate such random samples, we consider the following four data generating processes (DGPs):

DGP1: the standard normal distribution $\mathcal{N}(0, 1)$;

DGP2: a mixture of two normal distributions $0.4\mathcal{N}(-0.6, 0.16) + 0.6\mathcal{N}(0.4, 0.49)$;

DGP3: a Weibull distribution with shape parameter 1.5 and scale parameter 1 denoted as $\mathcal{W}(1.5, 1)$; and

DGP4: a stationary AR(1) process $x_t = 0.2x_{t-1} + u_t$, where $u_t \sim \mathcal{N}(0, 1)$.

For each sample size considered, we calculated $h_0(x)$ and $h_n(x)$ using each of the 500 random samples generated from each distribution, where x takes values on 100 equally spaced grid points on a finite interval. Such intervals are chosen to be $(-3, 3)$ for $\mathcal{N}(0, 1)$, $(-3, 4)$ for the mixture density, $(0, 4.1)$ for $\mathcal{W}(1.5, 1.0)$ and $(-4, 4)$ for the AR(1) process. We then took the average of each bandwidth curve over 500 replications. We plotted these averaged bandwidth curves in Figure 1.

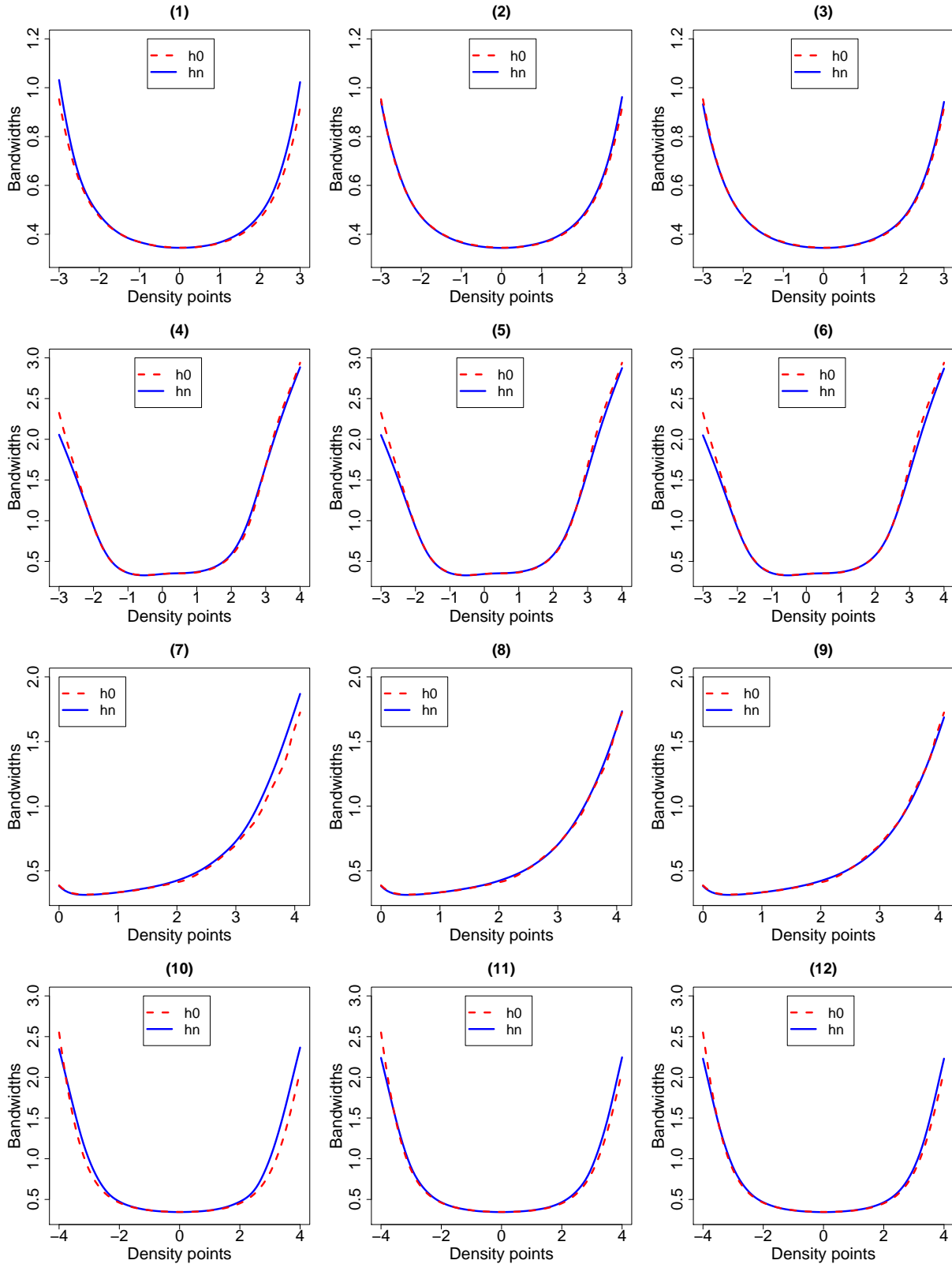
Figure 1(1)–(3) present the bandwidth curves averaged over 500 samples, which were generated from $\mathcal{N}(0, 1)$. Each averaged bandwidth curve indicates that a global bandwidth is inappropriate. However, these bandwidth curves may indicate that bandwidth may be approximately treated as a constant within a certain interval, which may choose for example, $(-2, 2)$. As the true density is the standard Gaussian, this interval is approximately the 95% high density region. Nonetheless, localized bandwidths should be used in both tails of the underlying density.

Figure 1(4)–(6) present the bandwidth curves averaged over 500 samples that were generated from the mixture of two Gaussians. These two averaged bandwidth curves are almost the same regardless different sample sizes. However, the shape of each bandwidth curve indicates that localized bandwidth should be used in the kernel estimator of the underlying mixture density. When x is within a certain interval, which for example, is from -1 to 2 , a high density region, bandwidth can be approximately treated as a constant.

Figure 1(7)–(9) present the bandwidth curves averaged over 500 samples, which were generated from the Weibull distribution. The shape of each bandwidth curve indicates that localized bandwidth should be used when x is larger than a threshold such as 2. Bandwidth can be treated as a constant when x is less than this threshold, which classifies a high density region.

Figure 1(10)–(12) present the bandwidth curves averaged over 500 samples, which were generated from the AR(1) process. The shape of each bandwidth curve indicates that localized bandwidth should be used in the kernel density estimator. When x is within a certain interval, which for example, is from -2 to 2 , a high density region, bandwidth can be approximately treated as a constant.

Figure 1: Averaged $h_0(x)$ and $h_n(x)$ with random samples generated from $\mathcal{N}(0, 1)$ (the first row), $0.4\mathcal{N}(-0.6, 0.16) + 0.6\mathcal{N}(0.4, 0.49)$ (the second row), $\mathcal{W}(1.5, 1)$ (the third row) and AR(1) process (the fourth row). The three columns correspond to sample sizes of 250, 750 and 1500, respectively.



To summarize, the estimated bandwidth curve indicates that localized bandwidth should be used for kernel density estimation, although bandwidth can be approximately treated as a constant in the high density region.

Table 1: Bias, standard deviation and MSE of NLB estimates with random samples generated from four different data generating processes.

	Sample size	NLB			
		Gaussian	Mixture	Weibull	AR(1)
bias	250	0.0149	0.0089	0.0782	0.0613
	750	0.0037	0.0026	0.0394	0.0183
	1500	0.0014	-0.0006	0.0241	0.0137
std	250	0.0962	0.0704	0.2232	0.2229
	750	0.0533	0.0403	0.1834	0.1638
	1500	0.0361	0.0278	0.1541	0.1275
MSE	250	0.0100	0.0053	0.0693	0.0633
	750	0.0029	0.0017	0.0431	0.0314
	1500	0.0014	0.0008	0.0307	0.0205

We now calculate the bias and standard deviation (std) of each bandwidth curve over 500 replications. Let $h_{n,r}(x_i)$ denote $h_n(x_i)$ calculated at x_i using the r th sample, for $r = 1, 2, \dots, 500$, and $i = 1, 2, \dots, 100$. For each sample size, we calculate the bias and standard deviation measures as follows:

$$\text{bias} = \frac{1}{100} \frac{1}{500} \sum_{i=1}^{100} \sum_{r=1}^{500} (h_{n,r}(x_i) - h_0(x_i)), \quad \text{std} = \sqrt{\frac{1}{100} \frac{1}{500} \sum_{i=1}^{100} \sum_{r=1}^{500} (h_{n,r}(x_i) - \bar{h}_n(x_i))^2},$$

where $\bar{h}_n(x_i) = 1/500 \sum_{r=1}^{500} h_{n,r}(x_i)$.

We define the mean squared error (MSE) as the overall average of the squared errors calculated at all grid points and sampling rounds. This measure is expressed as

$$\text{MSE} = \frac{1}{100} \frac{1}{500} \sum_{i=1}^{100} \sum_{r=1}^{500} \text{MSE}_{i,r},$$

where $\text{MSE}_{i,r} = (h_{n,r}(x_i) - h_0(x_i))^2$.

As shown by (9), the computation of $h_0(x)$ and $h_n(x)$ involves simulating m random numbers from each of $K(u)$ and $\pi(v)$, respectively. We choose $m = 5000$ in this study. Table 1 presents the

results of bias, standard deviation and mean squared error obtained based on samples generated from DGP1 to DGP 4.

From Table 1, we can find that for each data generating process, as sample size increases, the bias, standard deviation and mean squared error of $h_n(x)$ decrease. This shows that our proposed NLB estimator $h_n(x)$ has good finite-sample performance.

4.3 Comparison of different bandwidth choices for kernel density estimation

Recall that the local cross-validation is denoted as LCV and the global least-square cross-validation is denoted as GCV. In this section, we compare the performance of NLB with its bandwidth selected by the LCV, GCV and a global Bayesian method through the resulting kernel density estimates.

The LCV method is explained by [Hall and Schucany \(1989\)](#) as follows. Suppose we aim to estimate density $f(x)$ at a point x_0 , which we choose the origin without loss of generality. Construct a small region S_ε centred at $x_0 = 0$ with $\varepsilon > 0$. A localized bandwidth can be obtained by minimizing a local mean integrated squared error defined as

$$CV = \int_{S_\varepsilon} \hat{f}^2 - 2n^{-1} \sum_{j=1}^n I(X_j \in S_\varepsilon) \hat{f}_j(X_j), \quad (12)$$

where \hat{f} is the density estimator whose bandwidth is to be selected, \hat{f}_j is the leave-one-out version of \hat{f} computed with X_j omitted from the sample, and $I(X_j \in S_\varepsilon)$ equals 1 if $X_j \in S_\varepsilon$ and 0 otherwise.

In the univariate case, when the kernel function is $K(z) = 1/2$ for $|z| \leq 1$, and $S_\varepsilon = (-\varepsilon, \varepsilon)$, the LCV given by (12) can be simplified as

$$CV = (2nh)^{-2} \sum_i \sum_j \max\{0, \min(\varepsilon, X_i + h, X_j + h) + \min(\varepsilon, -X_i + h, -X_j + h)\} - \{n(n-1)h\}^{-1} \sum_{i \neq j} \sum I(|X_i - X_j| \leq h, -\varepsilon \leq X_j \leq \varepsilon). \quad (13)$$

The choice of a value of ε is not very important. After transforming the data to make x_0 the origin, we can use (13) to obtain local bandwidth h_{lcV} .

Let h_{bayes} denote the global bandwidth estimated through the Bayesian method proposed by [Zhang et al. \(2006\)](#), and h_{gcv} denote the bandwidth selected by the GCV. We have the following

four kernel density estimates for each data generating process.

$$\begin{aligned}\hat{f}_n^*(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n(x)} K\left(\frac{X_i - x}{h_n(x)}\right), \quad \hat{f}_{\text{lcv}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{\text{lcv}}} K\left(\frac{X_i - x}{h_{\text{lcv}}}\right), \\ \hat{f}_{\text{gcv}}(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{\text{gcv}}} K\left(\frac{X_i - x}{h_{\text{gcv}}}\right), \quad \hat{f}_{\text{bayes}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{\text{bayes}}} K\left(\frac{X_i - x}{h_{\text{bayes}}}\right).\end{aligned}$$

For each bandwidth selection method, the estimation of $f(\cdot)$ is calculated at m equally spaced points x_1, x_2, \dots, x_m . Then we compare the performance of different bandwidth selection methods by computing the corresponding average squared error (ASE):

$$\begin{aligned}\text{ASE}(h_n) &= \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{500} \sum_{r=1}^{500} (\hat{f}_n^*(x_{ri}) - f(x_i)) \right)^2, \quad \text{ASE}(h_{\text{lcv}}) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{500} \sum_{r=1}^{500} (\hat{f}_{\text{lcv}}(x_{ri}) - f(x_i)) \right)^2 \\ \text{ASE}(h_{\text{gcv}}) &= \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{500} \sum_{r=1}^{500} (\hat{f}_{\text{gcv}}(x_{ri}) - f(x_i)) \right)^2, \quad \text{ASE}(h_{\text{bayes}}) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{500} \sum_{r=1}^{500} (\hat{f}_{\text{bayes}}(x_{ri}) - f(x_i)) \right)^2,\end{aligned}$$

where $\hat{f}_n^*(x_{ri})$, $\hat{f}_{\text{lcv}}(x_{ri})$, $\hat{f}_{\text{gcv}}(x_{ri})$ and $\hat{f}_{\text{bayes}}(x_{ri})$ are density estimates calculated at x_i in the r -th replication using NLB, LCV, GCV and global Bayesian methods, respectively.

Table 2: ASE of kernel density estimates using different localized and global bandwidths.

DGP	n	Local		Global	
		$\text{ASE}(h_n)$	$\text{ASE}(h_{\text{lcv}})$	$\text{ASE}(h_{\text{gcv}})$	$\text{ASE}(h_{\text{bayes}})$
Normal	250	0.00062	0.00298	0.00779	0.00516
	750	0.00036	0.00189	0.00584	0.00230
	1500	0.00029	0.00120	0.00410	0.00128
Mixture	250	0.00304	0.00681	0.01059	0.00709
	750	0.00272	0.00352	0.00729	0.00514
	1500	0.00239	0.00279	0.00471	0.00340
Weibull	250	0.00473	0.01629	0.02743	0.02063
	750	0.00436	0.00852	0.01945	0.01162
	1500	0.00424	0.00554	0.01164	0.00808
AR(1)	250	0.00086	0.00214	0.00583	0.00393
	750	0.00058	0.00143	0.00363	0.00190
	1500	0.00049	0.00101	0.00237	0.00104

The results are summarized in Table 2, where we find the following evidence. First, under each DGP and for each bandwidth choice, as the sample size increases, the ASE of a kernel

density estimate decreases. Second, in each case, the NLB method leads to a smallest ASE than any other method. For example, when the DGP is normal and the sample size is 250, the ASE derived through NLB is 0.00062, while the ASE derived through LCV is 0.00298. The ASEs derived GCV and Bayesian are respectively, 0.00779 and 0.00516, which are larger than the former two. This indicates that the NLB method result in a better kernel density estimate in comparison to the other methods. Third, both the NLB and LCV methods lead to smaller ASE values than the GCV and Bayesian methods. This shows that localized bandwidth performs better than a global bandwidth.

5 Estimate and forecast the density of Eurodollar deposit rate

We estimate localized bandwidth for kernel density estimation of Eurodollar deposit rate using our proposed NLB method. The performance of the estimated bandwidth is measured by the resulting kernel density estimator. For comparison purposes, we examine the performance of LCV, GCV and the Bayesian sampling approach of [Zhang, King, and Hyndman \(2006\)](#). The performance of these methods are also examined via the forecasting performance of the resulting kernel density estimates.

The data consist of daily Eurodollar deposit rates in London with deposit maturities of 1, 3 and 6 months, respectively. The data are collected from the website of Federal reserve bank of St. Louis in the U.S. The sample period is from the 4th January 1971 to the 3rd January 1995. Time series plots of these three deposit rates are presented in [Figure 2](#).

5.1 Full sample density estimation

In the kernel density estimator given by (5), we use localized bandwidth in the sense that it depends on the density point. We assume an inverse Gamma prior for the bandwidth parameter given by (10), in which hyperparameters are $\alpha = 1$ and $\beta = 0.05$.

The four kernel density estimates with bandwidth selected by NLB, LCV, GCV and Bayesian methods, as well as the histogram of daily Eurodollar deposit rate of each maturity, are plotted in [Figure 3](#). We find that the estimated density with the four different bandwidths clearly differ from each other, especially in the peak and tail areas. Moreover, we find that the kernel density estimate with bandwidth selected through NLB is much closer to the histogram than other density estimates. This finding indicates that the NLB method may have better performance than the the LCV, GCV and Bayesian methods.

Figure 2: Eurodollar daily deposit rates with maturities of 1, 3 and 6 months: (1) 1-month maturity; (2) 3-month maturity; and (3) 6-month maturity.

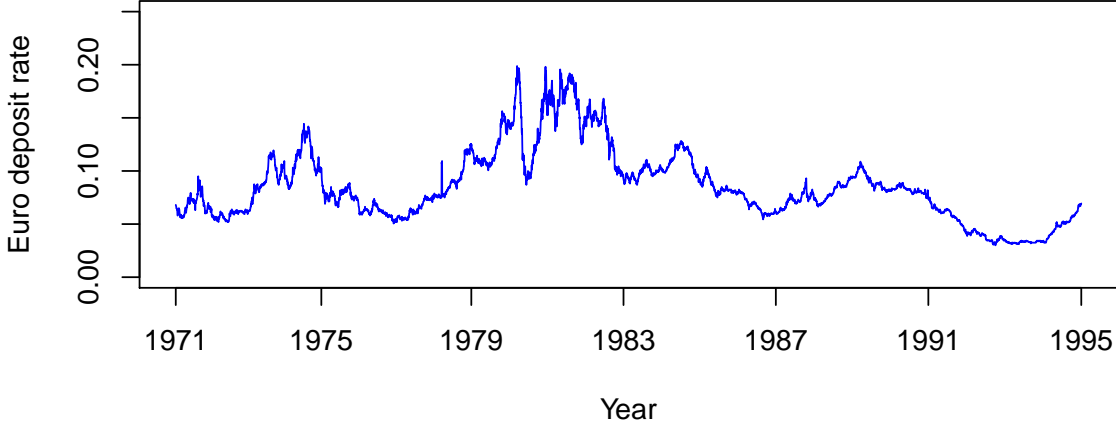
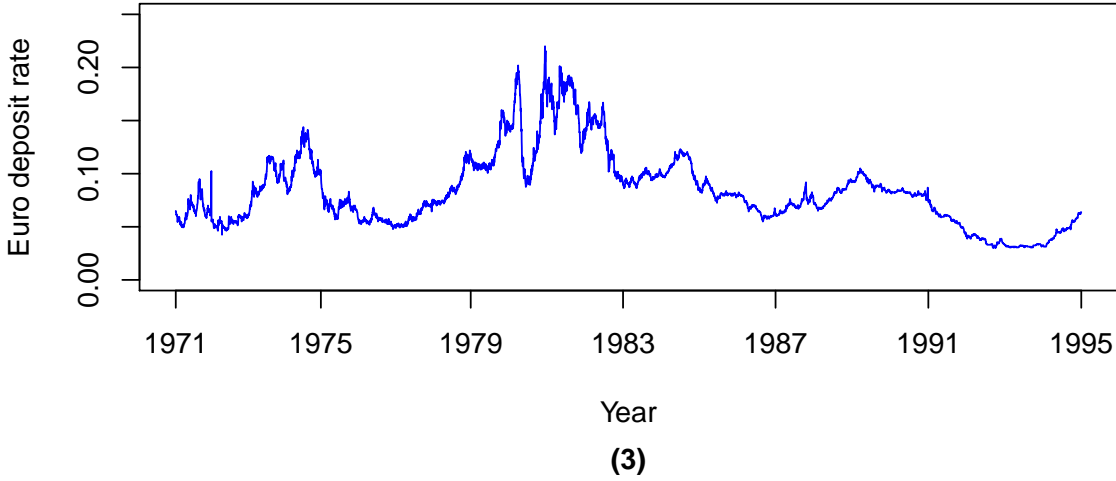
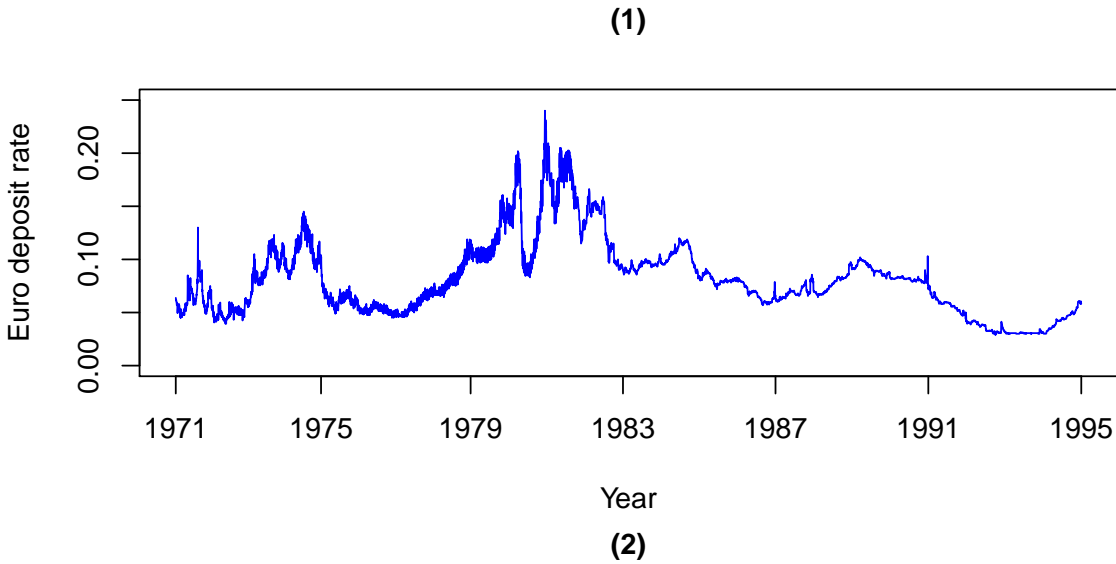
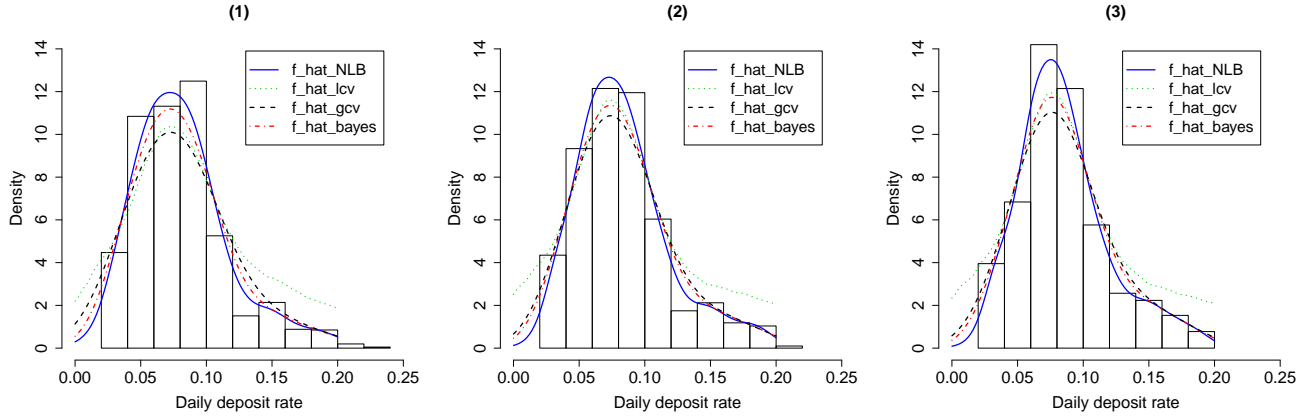


Figure 3: *Density estimates of Eurodollar daily deposit rates with maturities of 1, 3 and 6 months: (1) 1-month maturity; (2) 3-month maturity; and (3) 6-month maturity.*



As the true density of Eurodollar deposit rate is unknown, we employ the scoring rule discussed by [Amisano and Giacomini \(2007\)](#) to examine the out-of-sample performance of each estimated density. In this way, we are able to identify the best performer among a group of competing density estimates.

5.2 Forecasting results

We conducted a rolling-sample procedure to evaluate the performance of density estimate with its bandwidth selected through each of the four methods. For each maturity, let T denote the number of all observations, and let y_t denote the observed deposit rate at day t , for $t = 1, 2, \dots, T$. The first sample contains the first n observations, y_1, y_2, \dots, y_n , and is used to derive the bandwidth through the above-mentioned four methods. The resulting four density estimates are used to forecast the density of y_{n+1} . The second sample contains y_2, y_3, \dots, y_{n+1} , which are obtained by rolling the first sample forward for one step. Using the second sample, we repeat what was done based on the previous sample and forecast the density of y_{n+2} . This rolling procedure continues until the density of y_T is forecast.

We calculated the average logarithmic scores (ALS) over the out-of-sample period for density estimates with the above-mentioned four bandwidth selection methods. Denote $ALS(h_n)$, $ALS(h_{lcv})$, $ALS(h_{gcv})$ and $ALS(h_{bayes})$ as the ALS values obtained through respectively, NLB, LCV,

GCV and global Bayesian methods:

$$\begin{aligned}
 ALS(h_n) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \widehat{f}_n^*(y_{n+r}|h_n(y_{n+r})) = \frac{1}{T-n} \sum_{r=1}^{T-n} \log \left(\frac{1}{n} \sum_{i=r}^{n+r-1} \frac{1}{h_n(y_{n+r})} \phi \left(\frac{y_{n+r} - y_i}{h_n(y_{n+r})} \right) \right), \\
 ALS(h_{lcv}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \widehat{f}_{lcv}(y_{n+r}|h_{lcv}) = \frac{1}{T-n} \sum_{r=1}^{T-n} \log \left(\frac{1}{n} \sum_{i=r}^{n+r-1} \frac{1}{h_{lcv}} \phi \left(\frac{y_{n+r} - y_i}{h_{lcv}} \right) \right), \\
 ALS(h_{gcv}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \widehat{f}_{gcv}(y_{n+r}|h_{gcv}) = \frac{1}{T-n} \sum_{r=1}^{T-n} \log \left(\frac{1}{n} \sum_{i=r}^{n+r-1} \frac{1}{h_{gcv}} \phi \left(\frac{y_{n+r} - y_i}{h_{gcv}} \right) \right), \\
 ALS(h_{baves}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \widehat{f}_{baves}(y_{n+r}|h_{baves}) = \frac{1}{T-n} \sum_{r=1}^{T-n} \log \left(\frac{1}{n} \sum_{i=r}^{n+r-1} \frac{1}{h_{baves}} \phi \left(\frac{y_{n+r} - y_i}{h_{baves}} \right) \right),
 \end{aligned}$$

where $\phi(\cdot)$ is the standard Gaussian density. In terms of the average logarithmic score, the larger it is, the better the corresponding density performs.

Table 3 presents a summary of the total number of observations, size of the rolling sample, and number of rolling samples for each maturity.

Table 3: *A summary of rolling sample facts.*

Maturity	T	Size of rolling sample (n)	Number of rolling samples
1 month	6128	4059	2069
3 months	6129	4060	2069
6 months	6135	4066	2069

Table 4: *Out-of-sample average logarithmic scores of density estimates with bandwidths estimation through four different methods.*

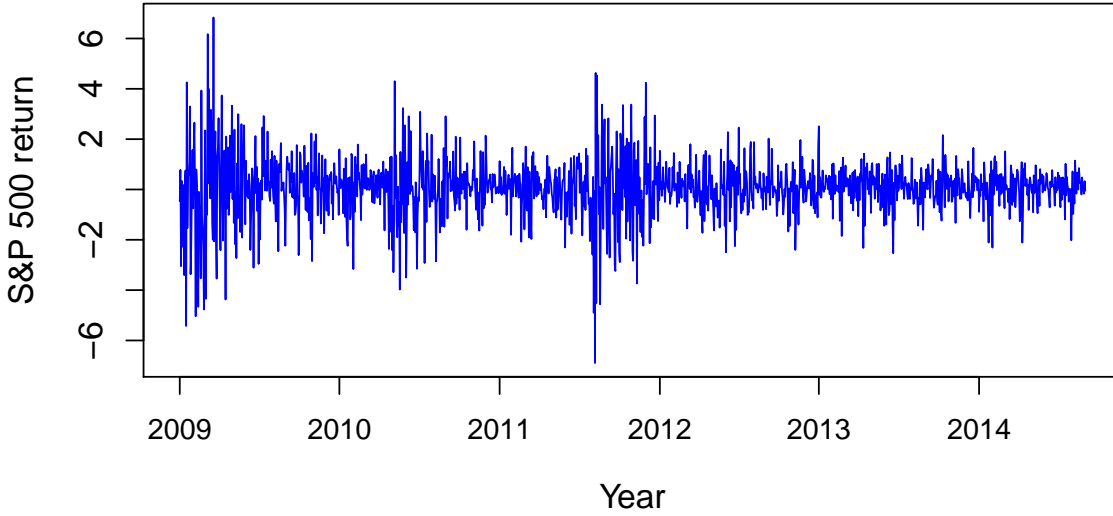
Maturity	Local		Global	
	$ALS(h_n)$	$ALS(h_{lcv})$	$ALS(h_{gcv})$	$ALS(h_{baves})$
1 month	2.4579	2.4323	2.4561	2.3469
3 months	2.4981	2.4949	2.4906	2.4471
6 months	2.6484	2.6008	2.5253	2.4958

Table 4 presents the average logarithmic scores derived through the four density estimates. At the maturities of 1 month and 3 months, the NLB method leads to a slightly better performance of density estimate than its competitors, while the former performs clearly better than the latter ones at the maturity of 6 months.

6 Estimate and forecast the density of S&P 500 daily return

It is important for being able to estimate the density of an asset return in finance. In this section, we estimate the density of the continuously compounded return of the S&P 500 daily closing index. We downloaded the S&P 500 daily closing prices, p_t , during the period from the 2nd January 2009 to the 2nd September 2014 from <http://finance.yahoo.com>. The date t return is calculated as $y_t = \log(p_t/p_{t-1}) \times 100$, and there are $T = 1425$ observations of the return. The time series plot of the return series is presented in Figure 4.

Figure 4: Time series plot of S&P 500 daily returns.



6.1 Density estimation under conditional heteroscedasticity

Let $\mathbf{y} = (y_1, y_2, \dots, y_T)'$ be a vector of T observations of S&P 500 daily returns. We consider a semiparametric GARCH (1,1) model given by

$$y_t = \sigma_t \varepsilon_t, \text{ and } \sigma_t^2 = b_0 + b_1 y_{t-1}^2 + b_2 \sigma_{t-1}^2, \quad (14)$$

where ε_t , for $t = 1, 2, \dots, T$, are independent and follow an unknown distribution with its density denoted as $f(\varepsilon_t)$. Zhang and King (2013) proposed approximating $f(\varepsilon_t)$ by

$$\hat{f}_\varepsilon(\varepsilon_t) = \frac{1}{T} \sum_{i=1}^T \frac{1}{h} \phi\left(\frac{\varepsilon_t - \varepsilon_i}{h}\right), \quad (15)$$

which has the form of kernel density estimator of errors. They suggested using a global bandwidth, as well as a localized version with $h(1 + h_\varepsilon|\varepsilon_t|)$ being assigned to ε_t as its bandwidth, for $t = 1, 2, \dots, T$.

We use our proposed NLB method to estimate localized bandwidth, which is now based on residuals rather than a sample of observations discussed in Section 5. We assume that the prior of h^2 is the inverse Gamma density with its hyperparameter vector, $\theta = (\alpha, \beta)' = (0.8, 0.01)'$. The estimation procedure is as follows.

Step 1: With the localized bandwidth $h_n(\varepsilon_t)$, the kernel-form error density is approximated as

$$\hat{f}_\varepsilon^*(\varepsilon_t) = \frac{1}{T} \sum_{i=1}^T \frac{1}{h_n(\varepsilon_t)} \phi\left(\frac{\varepsilon_t - \varepsilon_i}{h_n(\varepsilon_t)}\right). \quad (16)$$

Step 2: The density of y_t can then be expressed as

$$\hat{f}_Y^*(y_t|\lambda) = \frac{1}{T\sigma_t} \sum_{i=1}^T \frac{1}{h_n(y_t/\sigma_t)} \phi\left(\frac{y_t/\sigma_t - y_i/\sigma_i}{h_n(y_t/\sigma_t)}\right), \quad (17)$$

for $t = 1, 2, \dots, T$, where $\lambda = (b_1, b_2)'$.[†] Therefore, the likelihood of $\mathbf{y} = (y_1, y_2, \dots, y_T)'$ given λ is approximately $\ell(\mathbf{y}|\lambda) = \prod_{t=1}^T \hat{f}_Y^*(y_t|\lambda)$.

Step 3: Derive a semiparametric estimate of $\lambda = (b_1, b_2)'$ by maximizing $\ell(\mathbf{y}|\lambda)$.

After completing these three steps, we derive the localized bandwidth and a semiparametric estimate of λ , which are denoted as $h_n(y_t/\hat{\sigma}_t)$ and $\hat{\lambda} = (\hat{b}_1, \hat{b}_2)'$.

It is of great interest to estimate the one-day-ahead density of the S&P 500 daily return, y_{T+1} . Under the Gaussian kernel GARCH model, the density of y_{T+1} is estimated as

$$\hat{f}_Y^*(y_{T+1}|\hat{\lambda}) = \frac{1}{T\hat{\sigma}_{T+1}} \sum_{i=1}^T \frac{1}{h_n(y_{T+1}/\hat{\sigma}_{T+1})} \phi\left(\frac{y_{T+1}/\hat{\sigma}_{T+1} - y_i/\hat{\sigma}_i}{h_n(y_{T+1}/\hat{\sigma}_{T+1})}\right), \quad (18)$$

which was calculated at 1000 grid points, where $\hat{\sigma}_t^2 = \hat{b}_0 + \hat{b}_1 y_{t-1}^2 + \hat{b}_2 \hat{\sigma}_{t-1}^2$, for $t = 1, 2, \dots, T + 1$.

For comparison purpose, we computed the kernel density estimates of y_{T+1} with the bandwidth selected by the LCV. The estimation procedure is similar to Steps 1–3 described above. The only difference is that $h_n(y_t/\sigma_t)$ is replaced by $h_{\text{lcV}}(y_t/\sigma_t)$ selected through LCV.

Apart from using the localized bandwidth, we considered a global bandwidth for the kernel density estimator of y_{T+1} , where the bandwidth is estimated through Bayesian sampling

[†]When $\{y_t : t = 1, 2, \dots, T\}$ is pre-standardized, [Zhang and King \(2013\)](#) suggested choosing the value of b_0 as $(1 - b_1 - b_2)$ due to identification reasons.

proposed by [Zhang and King \(2013\)](#) and GCV, respectively. In the context of Bayesian sampling, the global bandwidth is treated as a parameter. Therefore, the vector of parameters is $\lambda_{\text{Bayes}} = (h, \sigma_0^2, b_1, b_2)'$. The sampling algorithm of [Zhang and King \(2013\)](#) was carried out to derive the estimate of λ_{Bayes} denoted as $\tilde{\lambda}_{\text{Bayes}} = (\tilde{h}, \tilde{\sigma}_0^2, \tilde{b}_1, \tilde{b}_2)'$. We then calculated the conditional variance $\tilde{\sigma}_t^2$ as

$$\tilde{\sigma}_t^2 = \tilde{b}_0 + \tilde{b}_1 y_{t-1}^2 + \tilde{b}_2 \tilde{\sigma}_{t-1}^2,$$

for $t = 1, 2, \dots, T+1$, where $\tilde{b}_0 = 1 - \tilde{b}_1 - \tilde{b}_2$. The density of y_{T+1} is then calculated according to (18) at 1000 grid points, where h and σ_t are replaced by respectively, \tilde{h} and $\tilde{\sigma}_t$.

We also used GCV method to choose a global bandwidth for the kernel density estimator of y_{T+1} . First, we estimated b_0 , b_1 and b_2 of (14) through the quasi maximum likelihood method under the normality assumption of ε_t , and the resulting estimates are denoted as $\tilde{b}_0^{(\text{gcv})}$, $\tilde{b}_1^{(\text{gcv})}$ and $\tilde{b}_2^{(\text{gcv})}$. We then computed the residuals as $\tilde{\varepsilon}_t = y_t / \tilde{\sigma}_{t,\text{gcv}}$, where

$$\tilde{\sigma}_{t,\text{gcv}}^2 = \tilde{b}_0^{(\text{gcv})} + \tilde{b}_1^{(\text{gcv})} y_{t-1}^2 + \tilde{b}_2^{(\text{gcv})} \tilde{\sigma}_{t-1,\text{gcv}}^2, \quad (19)$$

for $t = 1, 2, \dots, T$.

Second, we chose a global bandwidth denoted as \tilde{h}_{gcv} , for $\{\tilde{\varepsilon}_t : t = 1, 2, \dots, T\}$ using GCV. Third, with the selected bandwidth for the kernel density estimator, we derived an approximate likelihood, which is constructed through (15) and expressed as

$$\tilde{f}_Y(y_t | \lambda) = \frac{1}{T \sigma_t} \sum_{i=1}^T \frac{1}{\tilde{h}_{\text{gcv}}} \phi \left(\frac{y_t / \sigma_t - y_i / \sigma_i}{\tilde{h}_{\text{gcv}}} \right). \quad (20)$$

It was maximized with respect to b_1 and b_2 , where $b_0 = 1 - b_1 - b_2$. Thus, a semiparametric estimate of $(b_1, b_2)'$ was derived and denoted as $(\tilde{b}_1^{(\text{gcv})}, \tilde{b}_2^{(\text{gcv})})'$. This estimation procedure so far is similar to the semiparametric estimation of ARCH models proposed by [Engle and González-Rivera \(1991\)](#).

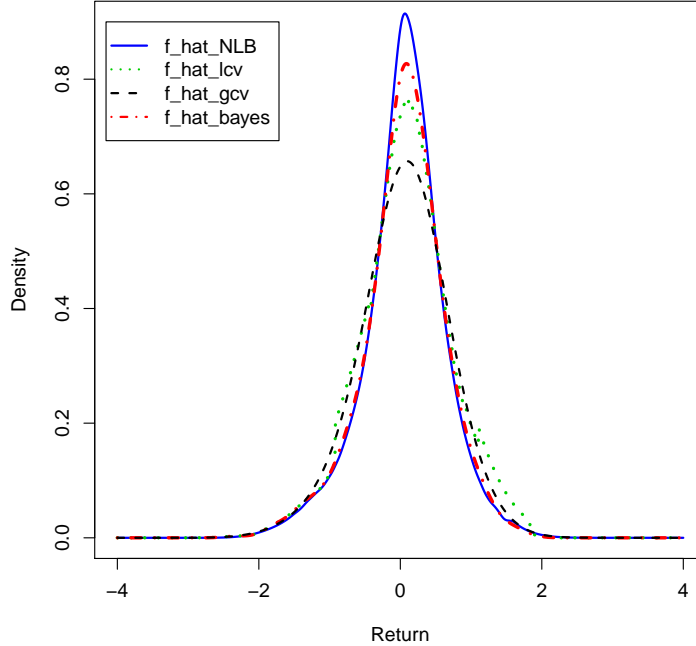
As we are interested in not only the bandwidth of kernel error density, but also the parameter estimates, we used the above-derived parameter estimates to calculate residuals again. Then, GCV method is applied to the updated residuals to derive a bandwidth, which is also denoted as \tilde{h}_{gcv} . With the updated bandwidth being substituted into the likelihood given by (20), we maximized the likelihood and obtained an updated estimate of $(b_1, b_2)'$, denoted as $(\tilde{b}_1^{(\text{gcv})}, \tilde{b}_2^{(\text{gcv})})'$, which is a semiparametric estimate of the parameter vector.

After completing these steps, we calculated the kernel density estimate of y_{T+1} ,

$$\tilde{f}_Y(y_{T+1}|\lambda) = \frac{1}{T\tilde{\sigma}_{T+1,\text{gcv}}} \sum_{t=1}^T \frac{1}{\tilde{h}_{\text{gcv}}} \phi\left(\frac{y_{T+1}/\tilde{\sigma}_{T+1,\text{gcv}} - y_t/\tilde{\sigma}_{t,\text{gcv}}}{\tilde{h}_{\text{gcv}}}\right), \quad (21)$$

at 1000 grid points, where $\tilde{\sigma}_{T+1,\text{gcv}}$ is calculated through (19).

Figure 5: *Density estimates of the one-day-ahead out-of-sample S&P 500 daily return.*



The resulting four kernel estimates of the density of the one-day-ahead S&P 500 daily return are plotted in Figure 5. We find that the kernel density estimate with bandwidth estimated through our NLB differs from the kernel density estimates with bandwidths selected through the LCV, GCV and Bayesian methods, especially in their tail and peak areas.

6.2 Forecasting results

We conducted the same rolling-sample procedure as we did in Section 5.2 to evaluate the out-of-sample performance of each density estimate resulted from each of these four bandwidth estimation methods. The number of all observations is $T = 1425$, and the size of a rolling sample is $n = 1005$, where the first rolling sample is from 2nd January 2013 to 2nd September 2014.

We calculated the average logarithmic scores, which are denoted as $ALS(h_n)$, $ALS(h_{\text{lcv}})$, $ALS(h_{\text{gcv}})$ and $ALS(h_{\text{bayes}})$ for respectively, the NLB, LCV, GCV and Bayesian methods, over the

out-of-sample period:

$$\begin{aligned}
 ALS(h_n) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \hat{f}_n^*(y_{n+r}|h_n(y_{n+r})) = -0.99, \\
 ALS(h_{lcv}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \hat{f}_{lcv}(y_{n+r}|h_{lcv}) = -1.00 \\
 ALS(h_{gcv}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \hat{f}_{gcv}(y_{n+r}|h_{gcv}) = -1.09, \\
 ALS(h_{bayes}) &= \frac{1}{T-n} \sum_{r=1}^{T-n} \log \hat{f}_{bayes}(y_{n+r}|h_{bayes}) = -1.03.
 \end{aligned}$$

By comparing these average scores, we find that the the NLB method leads to the highest score. This indicates that our proposed NLB method outperforms the LCV, GCV and Bayesian sampling methods for bandwidth selection. In addition, both $ALS(h_n)$, $ALS(h_{lcv})$ are larger than $ALS(h_{gcv})$ and $ALS(h_{bayes})$. This shows that the use of localized bandwidths outperforms the use of a global bandwidth.

7 Conclusions

In this paper, we have investigated the asymptotic properties of nonparametric localized bandwidth (NLB) estimator for kernel density estimation for stationary time series. We have proved that the NLB estimator is asymptotically normally distributed with root- n rate of convergence. Monte Carlo simulation studies have shown that the proposed NLB estimator has good finite-sample performance and performs better than the local cross-validation (LCV) and other global bandwidth selection methods.

When estimating the density of Eurodollar deposit through the kernel method, we have found that our proposed NLB method leads to a better performance of the density estimator than the LCV, GCV and Bayesian sampling method in terms of out-of-sample forecasting. In the kernel estimator of the density of S&P 500 daily return under conditional heteroscedasticity, our proposed NLB method outperforms its competitors in terms of out-of-sample forecasting.

In this paper, hyperparameters are chosen subjectively. It is well known that expressing honest prior information can be difficult. [Casella \(2001\)](#) investigated hyperparameter estimation based on the EM algorithm and the Markov chain Monte Carlo simulation. [Atchadé \(2011\)](#) developed an adaptive Monte Carlo strategy for sampling from posterior in empirical Bayes analysis. However, these methods are not applicable to bandwidth estimation discussed in this paper,

mainly because information from the original data may not be sufficient for consistently estimating the hyperparameters. Therefore, a more reliable data-driven approach to hyperparameter estimation will be part of our future research.

Acknowledgements

The authors acknowledge constructive comments from the seminar participants at Monash University, University of Bergen, University of York in England and University of Science and Technology of China in Hefei, in particular to Anastasios Panagiotelis, Dag Tjøstheim, Yaohua Wu, Wenyang Zhang and Lincheng Zhao. This research was supported under the Australian Research Council's *Discovery Projects* Scheme under Grant Numbers: DP1095838 & DP130104229 & DP150101012.

References

- Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. *Review of Financial studies* 9 (2), 385–426.
- Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25 (2), 177–190.
- Atchadé, Y. F., 2011. A computational framework for empirical Bayes inference. *Statistics and Computing* 21 (4), 463–473.
- Bithell, J., 1990. An application of density estimation to geographical epidemiology. *Statistics in Medicine* 9 (6), 691–701.
- Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71 (2), 353–360.
- Brewer, M. J., 2000. A Bayesian model for local smoothing in kernel density estimation. *Statistics and Computing* 10 (4), 299–309.
- Casella, G., 2001. Empirical Bayes Gibbs sampling. *Biostatistics* 2 (4), 485–500.
- de Lima, M. S., Atuncar, G. S., 2011. A Bayesian method to estimate the optimal bandwidth for multivariate kernel estimator. *Journal of Nonparametric Statistics* 23 (1), 137–148.
- Elgammal, A., Duraiswami, R., Harwood, D., Davis, L. S., 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE* 90 (7), 1151–1163.

- Engle, R. F., González-Rivera, G., 1991. Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9 (4), 345–359.
- Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Non- and Parametric Methods*. Springer, New York.
- Gangopadhyay, A., Cheung, K., 2002. Bayesian approach to the choice of smoothing parameter in kernel density estimation. *Journal of Nonparametric Statistics* 14 (6), 655–664.
- Gao, J., 2007. *Nonlinear Time Series: Semi- and Non-Parametric Methods*. Chapman & Hall/CRC, London.
- Ghosh, J. K., Ramamoorthi, R. V., 2003. *Bayesian Nonparametrics*. Springer, New York.
- Hall, P., Li, Q., Racine, J. S., 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. *The Review of Economics and Statistics* 89 (4), 784–789.
- Hall, P., Racine, J., Li, Q., 2004. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99 (468), 1015–1026.
- Hall, P., Schucany, W. R., 1989. A local cross-validation algorithm. *Statistics & Probability Letters* 8 (2), 109–117.
- Härdle, W., Hall, P., Marron, J. S., 1988. How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association* 83 (401), 86–95.
- Härdle, W., Hall, P., Marron, J. S., 1992. Regression smoothing parameters that are not far from their optimum. *Journal of the American Statistical Association* 87 (417), 227–233.
- Heidenreich, N.-B., Schindler, A., Sperlich, S., 2013. Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis* 97 (4), 403–433.
- Hjort, N. L., Holmes, C., Müller, P., Walker, S. G., 2010. *Bayesian Nonparametrics*. Cambridge University Press, Cambridge.
- Jones, M. C., Marron, J. S., Sheather, S. J., 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91 (433), 401–407.
- Kulasekera, K. B., Padgett, W. J., 2006. Bayes bandwidth selection in kernel density estimation with censored data. *Nonparametric Statistics* 18 (2), 129–143.
- Li, Q., Lin, J., Racine, J. S., 2013. Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics* 31 (1), 57–65.
- Li, Q., Racine, J. S., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Lo, A. Y., 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics* 12 (1), 351–357.
- Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics* 9 (2), 65–78.

Sain, S., Scott, D., 1996. On locally adaptive density estimation. *Journal of the American Statistical Association* 91 (436), 1525–1534.

Seaman, D. E., Powell, R. A., 1996. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology* 77 (7), 2075–2085.

Sheather, S. J., 2004. Density estimation. *Statistical Science* 19 (4), 588–597.

Silverman, B. W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Zhang, X., King, M. L., 2013. Gaussian kernel GARCH models. Working paper, Monash University.

URL <https://ideas.repec.org/e/pzh72.html>

Zhang, X., King, M. L., Hyndman, R. J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis* 50 (11), 3009–3031.

Appendix

We now provide the proofs of Lemmas 1–2 and Theorems 1–3.

Proof of Lemma 1.

We have

$$\begin{aligned}
 p(x) &= \int f(x|h)\pi(h)dh = \iint f(y)K_h(y-x)\pi(h)dhd y \\
 &= \iint f(x+uh)K(u)\pi(h)dhd u = \iint f(x+uv)K(u)\pi(v)dvdu = \mathbb{E}[f(x+u_i v_i)], \\
 q(x) &= \int hf(x|h)\pi(h)dh = \iint hf(x+uh)K(u)\pi(h)dhd u \\
 &= \iint v f(x+uv)K(u)\pi(v)dvdu = \mathbb{E}[f(x+u_i v_i)v_i].
 \end{aligned}$$

As $\{u_i, v_i; i = 1, 2, \dots, m\}$ are independent and identically distributed (iid), $f(x+u_i v_i)$ is also iid. Therefore, by the law of large numbers, as $m \rightarrow \infty$, $\frac{1}{m} \sum_{i=1}^m f(x+u_i v_i) - \mathbb{E}[f(x+u_i v_i)] = o_P(1)$. Hence $p_m(x) - p(x) = o_P(1)$. Similarly, $f(x+u_i v_i)v_i$ is also iid. Therefore, by the law of large numbers, as $m \rightarrow \infty$, $\frac{1}{m} \sum_{i=1}^m f(x+u_i v_i)v_i - \mathbb{E}[f(x+u_i v_i)v_i] = o_P(1)$. Hence $q_m(x) - q(x) = o_P(1)$.

Proof of Lemma 2.

Observe that

$$\begin{aligned}
p_n(x) &= \int \widehat{f}(x|h)\pi(h)dh = \int \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)\pi(h)dh \\
&= \frac{1}{n} \sum_{i=1}^n \int \frac{1}{v} K\left(\frac{X_i-x}{v}\right)\pi(v)dv = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_2 \left[\frac{1}{v} K\left(\frac{X_i-x}{v}\right) \right] \\
q_n(x) &= \int h\widehat{f}(x|h)\pi(h)dh = \int h \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)\pi(h)dh \\
&= \frac{1}{n} \sum_{i=1}^n \int K\left(\frac{X_i-x}{v}\right)\pi(v)dv = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_2 \left[K\left(\frac{X_i-x}{v}\right) \right],
\end{aligned}$$

where $\mathbb{E}_2(\cdot)$ denotes the expectation under the marginal distribution of h .

Note that $\{v_j : j = 1, 2, \dots, m\}$ and $\{X_i : i = 1, 2, \dots, n\}$ are mutually independent. By the law of large numbers, we have as $m \rightarrow \infty$ and $n \rightarrow \infty$, $\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{v_j} K\left(\frac{X_i-x}{v_j}\right) - \mathbb{E} \left[\frac{1}{v} K\left(\frac{X_1-x}{v}\right) \right] = o_P(1)$. Similarly, $\frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m K\left(\frac{X_i-x}{v_j}\right) - \mathbb{E} \left[K\left(\frac{X_1-x}{v}\right) \right] = o_P(1)$. Therefore, $q_{nm}(x) - q_n(x) = o_P(1)$.

Proof of Theorem 1.

According to (2) and (4), we have

$$\begin{aligned}
h_n(x) - h_0(x) &= \frac{q_n(x)}{p_n(x)} - \frac{q(x)}{p(x)} = \frac{1}{p_n(x)p(x)} [q_n(x)p(x) - q(x)p_n(x)] \\
&= \frac{1}{p_n(x)p(x)} [q_n(x)p(x) - q(x)p(x) + q(x)p(x) - q(x)p_n(x)] \\
&= \frac{1}{p_n(x)p(x)} [p(x)(q_n(x) - q(x)) - q(x)(p_n(x) - p(x))] = \frac{1}{p_n(x)p(x)} L_n(x),
\end{aligned}$$

where $L_n(x) = p(x)(q_n(x) - q(x)) - q(x)(p_n(x) - p(x))$. Note that $\mathbb{E}_1[\widehat{f}(x|h)] = f(x|h)$ and that

$$\begin{aligned}
\widehat{f}(x|h) - f(x|h) &= \widehat{f}(x|h) - \mathbb{E}_1[\widehat{f}(x|h)] = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) - \frac{1}{nh} \sum_{i=1}^n \mathbb{E}_1 \left[K\left(\frac{X_i-x}{h}\right) \right] \\
&= \frac{1}{nh} \sum_{i=1}^n \left(K\left(\frac{X_i-x}{h}\right) - \mathbb{E}_1 \left[K\left(\frac{X_i-x}{h}\right) \right] \right) = \frac{1}{n} \sum_{i=1}^n U_i(x; h),
\end{aligned}$$

where $U_i(x; h) = \frac{1}{h} \left[K\left(\frac{X_i-x}{h}\right) - \mathbb{E}_1 K\left(\frac{X_i-x}{h}\right) \right]$.

Therefore, we have

$$\begin{aligned}
L_n(x) &= p(x)(q_n(x) - q(x)) - q(x)(p_n(x) - p(x)) \\
&= \int hp(x)\pi(h) [\hat{f}(x|h) - f(x|h)] dh - \int q(x)\pi(h) [\hat{f}(x|h) - f(x|h)] dh \\
&= \int [hp(x) - q(x)] \pi(h) [\hat{f}(x|h) - f(x|h)] dh = \int [hp(x) - q(x)] \pi(h) [\hat{f}(x|h) - \mathbb{E}_1(\hat{f}(x|h))] dh \\
&= \int [hp(x) - q(x)] \pi(h) \left[\frac{1}{n} \sum_{i=1}^n U_i(x; h) \right] dh = \frac{1}{n} \sum_{i=1}^n \int [hp(x) - q(x)] U_i(x; h) \pi(h) dh \\
&= \frac{1}{n} \sum_{i=1}^n V_i(x),
\end{aligned}$$

where $V_i(x) = \int [hp(x) - q(x)] U_i(x; h) \pi(h) dh$.

It is easy to check that

$$\mathbb{E}_1[V_i(x)] = \mathbb{E}_1 \left[\int [hp(x) - q(x)] U_i(x; h) \pi(h) dh \right] = \int [hp(x) - q(x)] \mathbb{E}_1[U_i(x; h)] \pi(h) dh = 0.$$

Note that

$$\begin{aligned}
V_i^2(x) &= \left\{ \int [up(x) - q(x)] U_i(x; u) \pi(u) du \right\} \left\{ \int [vp(x) - q(x)] U_i(x; v) \pi(v) dv \right\} \\
&= \iint [up(x) - q(x)] [vp(x) - q(x)] U_i(x; u) U_i(x; v) \pi(u) \pi(v) dudv.
\end{aligned}$$

Thus, the variance of $V_i(x)$ is given by

$$\text{Var}[V_i(x)] = \mathbb{E}[V_i^2(x)] = \iint [up(x) - q(x)] [vp(x) - q(x)] \mathbb{E}_1[U_i(x; u) U_i(x; v)] \pi(u) \pi(v) dudv,$$

where

$$\begin{aligned}
\mathbb{E}_1[U_i(x; u) U_i(x; v)] &= \frac{1}{uv} \mathbb{E}_1 \left\{ \left[K\left(\frac{X_i - x}{u}\right) - \mathbb{E}_1 K\left(\frac{X_i - x}{u}\right) \right] \left[K\left(\frac{X_i - x}{v}\right) - \mathbb{E}_1 K\left(\frac{X_i - x}{v}\right) \right] \right\} \\
&= \frac{1}{uv} \left(\mathbb{E}_1 \left[K\left(\frac{X_i - x}{u}\right) K\left(\frac{X_i - x}{v}\right) \right] - \mathbb{E}_1 \left[K\left(\frac{X_i - x}{u}\right) \right] \mathbb{E}_1 \left[K\left(\frac{X_i - x}{v}\right) \right] \right) \\
&= \frac{1}{uv} [uv f_{uv}(x) - uv f_u(x) f_v(x)] = f_{uv}(x) - f_u(x) f_v(x) = R_{uv}(x).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}[V_i^2(x)] &= \iint [up(x) - q(x)] [vp(x) - q(x)] \mathbb{E}_1[U_i(x; u) U_i(x; v)] \pi(u) \pi(v) dudv \\
&= \iint [up(x) - q(x)] [vp(x) - q(x)] R_{uv}(x) \pi(u) \pi(v) dudv.
\end{aligned}$$

We denote $\text{Var}[V_i(x)]$ as $\gamma(0) = \iint [up(x) - q(x)] [vp(x) - q(x)] R_{uv}(x) \pi(u) \pi(v) dudv$.

The covariance of $V_i(x)$ and $V_j(x)$ is given by

$$\begin{aligned}
\text{Cov}[V_i(x), V_j(x)] &= \mathbb{E}[V_i(x)V_j(x)] \\
&= \mathbb{E}\left\{\int [up(x) - q(x)] U_i(x; u)\pi(u)du\right\}\left\{\int [vp(x) - q(x)] U_j(x; v)\pi(v)dv\right\} \\
&= \mathbb{E}\left\{\iint [up(x) - q(x)][vp(x) - q(x)] U_i(x; u)U_j(x; v)\pi(u)\pi(v)dudv\right\} \\
&= \iint [up(x) - q(x)][vp(x) - q(x)] \{\mathbb{E}_1[U_i(x; u)U_j(x; v)]\} \pi(u)\pi(v)dudv,
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_1[U_i(x; u)U_j(x; v)] &= \frac{1}{uv} \mathbb{E}_1\left\{\left[K\left(\frac{X_i - x}{u}\right) - \mathbb{E}_1 K\left(\frac{X_i - x}{u}\right)\right]\left[K\left(\frac{X_j - x}{v}\right) - \mathbb{E}_1 K\left(\frac{X_j - x}{v}\right)\right]\right\} \\
&= \frac{1}{uv} \left(\mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)K\left(\frac{X_j - x}{v}\right)\right] - \mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)\right]\mathbb{E}_1\left[K\left(\frac{X_j - x}{v}\right)\right]\right).
\end{aligned}$$

Since $\{X_i\}$ is strictly stationary by Assumption 1, we have $\mathbb{E}_1 K\left(\frac{X_i - x}{u}\right) = \mathbb{E}_1 K\left(\frac{X_j - x}{v}\right) = v f_v(x)$. Therefore $\mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)\right]\mathbb{E}_1\left[K\left(\frac{X_j - x}{v}\right)\right] = uv f_u(x) f_v(x)$, and

$$\mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)K\left(\frac{X_j - x}{v}\right)\right] = \int K\left(\frac{y - x}{u}\right)K\left(\frac{z - x}{v}\right) f_s(y, z) dy dz = uv g_{uv, s}(x),$$

where $s = |i - j|$ and $f_{i-j}(y, z)$ denotes the joint density of (X_i, X_j) .

Similarly, we have

$$\begin{aligned}
\mathbb{E}_1[U_i(x; u)U_j(x; v)] &= \frac{1}{uv} \left(\mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)K\left(\frac{X_j - x}{v}\right)\right] - \mathbb{E}_1\left[K\left(\frac{X_i - x}{u}\right)\right]\mathbb{E}_1\left[K\left(\frac{X_j - x}{v}\right)\right]\right) \\
&= \frac{1}{uv} [uv g_{uv, s}(x) - uv f_u(x) f_v(x)] = g_{uv, s}(x) - f_u(x) f_v(x) = G_{uv, s}(x).
\end{aligned}$$

We therefore have

$$\begin{aligned}
\text{Cov}[V_i(x), V_j(x)] &= \iint [up(x) - q(x)][vp(x) - q(x)] \{\mathbb{E}_1[U_i(x; u)U_j(x; v)]\} \pi(u)\pi(v)dudv \\
&= \iint [up(x) - q(x)][vp(x) - q(x)] G_{uv, s}(x) \pi(u)\pi(v)dudv.
\end{aligned}$$

Denote $\text{Cov}[V_i(x), V_{i+j}(x)]$ as $\gamma(j) = \iint [up(x) - q(x)][vp(x) - q(x)] G_{uv, j}(x) \pi(u)\pi(v)dudv$.

Before we establish the central limit theorem, we need to verify the following condition: $\mathbb{E}|V_i(x)|^\delta < \infty$ for some constant $\delta > 2$ that satisfies $\sum_{j \geq 1} \alpha(j)^{1-2/\delta} < \infty$.

As $V_i(x)$ is a measurable function of X_i , the process $V_i(x)$ possesses the mixing property of $\{X_i\}$. This indicates that $\{V_i\}$ is a sequence with α -mixing coefficient satisfying $\sum_{j \geq 1} \alpha(j)^{1-2/\delta} < \infty$. Denote $g_1(h) = [hp(x) - q(x)] \pi(h)$ and $g_2(h) = U_i(x; h)$. Under Hölder's inequality for $\delta > 1$ and $1/\delta + 1/q = 1$, we

have

$$\begin{aligned} |V_i(x)|^\delta &= \left| \int [hp(x) - q(x)] U_i(x; h) \pi(h) dh \right|^\delta = \left| \int g_1(h) g_2(h) dh \right|^\delta = \left| \int g_1^{1-\frac{1}{\delta}}(h) \cdot g_1^{\frac{1}{\delta}}(h) g_2(h) dh \right|^\delta \\ &\leq \left(\int g_1(h) dh \right)^{\frac{\delta}{q}} \int g_1(h) |g_2(h)|^\delta dh = \left(\int g_1(h) dh \right)^{\delta-1} \int g_1(h) |g_2(h)|^\delta dh. \end{aligned}$$

By Assumption 3, we have for some $0 < C_\delta < \infty$

$$\begin{aligned} \mathbb{E} \left[|V_i(x)|^\delta \right] &= \left(\int g_1(h) dh \right)^{\delta-1} \int g_1(h) \cdot \mathbb{E}_1 \left[|g_2(h)|^\delta \right] dh \\ &\leq C_\delta \left(\int g_1(h) dh \right)^{\delta-1} \cdot \int g_1(h) h^{1-\delta} \left(\int K^\delta(v) f(x+vh) dv \right) dh < \infty. \end{aligned}$$

Lemma A.1 of [Gao \(2007\)](#) implies that

$$|\gamma(j)| = |\text{Cov}(V_t(x), V_s(x))| \leq 10\alpha(j)^{1-2/\delta} \left\{ \mathbb{E}|V_t(x)|^\delta \right\}^{\frac{1}{\delta}} \left\{ \mathbb{E}|V_s(x)|^\delta \right\}^{\frac{1}{\delta}},$$

where $|t-s|=j$. By $\sum_{j \geq 1} \alpha(j)^{1-2/\delta} < \infty$, we have

$$\sum_{j=1}^{\infty} |\gamma(j)| = 2 \sum_{1 \leq t < s \leq n} |\text{Cov}(V_t(x), V_s(x))| \leq \sum_{j=1}^{\infty} 10\alpha(j)^{1-2/\delta} \left\{ \mathbb{E}|V_t(x)|^\delta \right\}^{\frac{1}{\delta}} \left\{ \mathbb{E}|V_s(x)|^\delta \right\}^{\frac{1}{\delta}} < \infty.$$

Assumptions 1–3 ensure that condition (i) is satisfied. Following Theorem 2.20 of [Fan and Yao \(2003\)](#), we have as $n \rightarrow \infty$

$$\begin{aligned} \text{Var} \left[\sqrt{n} L_n(x) \right] &= \text{Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x) \right] = \frac{1}{n} \text{Var} \left[\sum_{i=1}^n V_i(x) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \text{Var} [V_i(x)] + \frac{2}{n} \sum_{1 \leq i < j \leq n} \text{Cov} [V_i(x), V_j(x)] \\ &= \gamma(0) + 2 \sum_{l=1}^{n-1} \left(1 - \frac{l}{n} \right) \gamma(l) \rightarrow \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j) = \Sigma_L(x). \end{aligned}$$

Theorem 2.21 of [Fan and Yao \(2003\)](#) implies that

$$\sqrt{n} L_n(x) \rightarrow_D \mathcal{N}(0, \Sigma_L(x)), \quad (22)$$

where $\Sigma_L(x) = \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j)$. Therefore, we have

$$h_n(x) - h_0(x) = \frac{1}{p_n(x)p(x)} L_n(x).$$

By Assumption 1 and Proposition 2.8 of [Fan and Yao \(2003\)](#), as $n \rightarrow \infty$, we have

$$\begin{aligned} p_n(x) &= \int \hat{f}(x|h)\pi(h)dh = \int \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)\pi(h)dh \\ &= \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{X_i-x}{h}\right)\pi(h)dh \rightarrow_P \int \frac{1}{h} \mathbb{E}_1 \left[K\left(\frac{X_1-x}{h}\right) \right] \pi(h)dh \\ &= \iint \frac{1}{h} K\left(\frac{y-x}{h}\right)\pi(h)f(y)dhd y = p(x). \end{aligned}$$

Denote $p^2(x) = Q(x)$. As $n \rightarrow \infty$, we have $p_n(x)p(x) \rightarrow_P Q(x)$ and $\frac{1}{p_n(x)p(x)} \rightarrow_P Q^{-1}(x)$. Therefore, under Assumptions 1–3, as $n \rightarrow \infty$,

$$\sqrt{n}(h_n(x) - h_0(x)) \rightarrow_D \mathcal{N}(0, \Sigma_0(x)), \quad (23)$$

where $\Sigma_0(x) = Q^{-2}(x)\Sigma_L(x)$. Therefore, we have proved Theorem 1.

Proof of Theorem 2.

Without loss of generality, it suffices to prove the theorem for the case of $N = 2$.

$$\begin{aligned} \sqrt{n}(h_n(x_1) - h_0(x_1)) &= \frac{1}{p_n(x_1)p(x_1)} \sqrt{n}L_n(x_1) = \frac{1}{p_n(x_1)p(x_1)} \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1), \\ \sqrt{n}(h_n(x_2) - h_0(x_2)) &= \frac{1}{p_n(x_2)p(x_2)} \sqrt{n}L_n(x_2) = \frac{1}{p_n(x_2)p(x_2)} \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2). \end{aligned}$$

Note that

$$\begin{aligned} \text{Cov}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1), \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2)\right) &= \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1) \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2)\right) \\ &= \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n V_i(x_1) \sum_{i=1}^n V_i(x_2)\right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(V_i(x_1)V_j(x_2)), \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}[V_i(x_1)V_j(x_2)] &= \mathbb{E}\left\{\int [up(x_1) - q(x_1)]U_i(x_1; u)\pi(u)du\right\}\left\{\int [vp(x_2) - q(x_2)]U_j(x_2; v)\pi(v)dv\right\} \\ &= \mathbb{E}\left\{\iint [up(x_1) - q(x_1)][vp(x_2) - q(x_2)]U_i(x_1; u)U_j(x_2; v)\pi(u)\pi(v)dudv\right\} \\ &= \iint [up(x_1) - q(x_1)][vp(x_2) - q(x_2)]\{\mathbb{E}_1[U_i(x_1; u)U_j(x_2; v)]\}\pi(u)\pi(v)dudv. \\ \mathbb{E}_1[U_i(x_1; u)U_j(x_2; v)] &= \frac{1}{uv} \mathbb{E}_1\left\{\left[K\left(\frac{X_i-x_1}{u}\right) - \mathbb{E}_1 K\left(\frac{X_i-x_1}{u}\right)\right]\left[K\left(\frac{X_j-x_2}{v}\right) - \mathbb{E}_1 K\left(\frac{X_j-x_2}{v}\right)\right]\right\} \\ &= \frac{1}{uv} \left(\mathbb{E}_1\left[K\left(\frac{X_i-x_1}{u}\right)K\left(\frac{X_j-x_2}{v}\right)\right] - \mathbb{E}_1\left[K\left(\frac{X_i-x_1}{u}\right)\right]\mathbb{E}_1\left[K\left(\frac{X_j-x_2}{v}\right)\right]\right). \end{aligned}$$

As $\mathbb{E}_1 \left[K \left(\frac{X_i - x_1}{u} \right) \right] \mathbb{E}_1 \left[K \left(\frac{X_j - x_2}{v} \right) \right] = uv f_u(x_1) f_v(x_2)$, we have

$$\mathbb{E}_1 \left[K \left(\frac{X_i - x_1}{u} \right) K \left(\frac{X_j - x_2}{v} \right) \right] = \int K \left(\frac{y - x_1}{u} \right) K \left(\frac{z - x_2}{v} \right) f_s(y, z) dy dz = uv m_{uv,s}(x_1, x_2),$$

where the notation $m_{uv,s}(x_1, x_2)$ has already been introduced just above Theorem 1, and $s = |i - j|$ and $f_{|i-j|}(y, z)$ denotes the joint density of (X_i, X_j) .

$$\begin{aligned} \mathbb{E}_1 [U_i(x_1; u) U_j(x_2; v)] &= \frac{1}{uv} \left(\mathbb{E}_1 \left[K \left(\frac{X_i - x_1}{u} \right) K \left(\frac{X_j - x_2}{v} \right) \right] - \mathbb{E}_1 \left[K \left(\frac{X_i - x_1}{u} \right) \right] \mathbb{E}_1 \left[K \left(\frac{X_j - x_2}{v} \right) \right] \right) \\ &= \frac{1}{uv} [uv m_{uv,s}(x_1, x_2) - uv f_u(x_1) f_v(x_2)] = m_{uv,s}(x_1, x_2) - f_u(x_1) f_v(x_2) = S_{uv,s}(x_1, x_2), \end{aligned}$$

where $S_{uv,s}(x_1, x_2)$ has also been introduced just above Theorem 1. Consequently, we have

$$\mathbb{E} [V_i(x_1) V_j(x_2)] = \iint [up(x_1) - q(x_1)] [vp(x_2) - q(x_2)] S_{uv,s}(x_1, x_2) \pi(u) \pi(v) dudv.$$

Let $\gamma_2(0) = \mathbb{E} [V_i(x_1) V_i(x_2)] = \iint [up(x_1) - q(x_1)] [vp(x_2) - q(x_2)] S_{uv,0}(x_1, x_2) \pi(u) \pi(v) dudv$ and $\gamma_2(s) = \mathbb{E} [V_i(x_1) V_j(x_2)] = \iint [up(x_1) - q(x_1)] [vp(x_2) - q(x_2)] S_{uv,s}(x_1, x_2) \pi(u) \pi(v) dudv$. Thus, as $n \rightarrow \infty$, we have

$$\begin{aligned} \text{Cov} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1), \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2) \right) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} (V_i(x_1) V_j(x_2)) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \iint [up(x_1) - q(x_1)] [vp(x_2) - q(x_2)] S_{uv,s}(x_1, x_2) \pi(u) \pi(v) dudv \\ &= \gamma_2(0) + \frac{2}{n} \sum_{1 \leq i < j \leq n} \gamma_2(|i - j|) = \gamma_2(0) + 2 \sum_{s=1}^{n-1} \gamma_2(s) \left(1 - \frac{s}{n} \right) \\ &\rightarrow \gamma_2(0) + 2 \sum_{s=1}^{\infty} \gamma_2(s) = \Sigma_v(x_1, x_2). \end{aligned}$$

Define $S = C_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1) + C_2 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2)$, where C_1 and C_2 are constants. We have

$$S = C_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1) + C_2 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{C_1 V_i(x_1) + C_2 V_i(x_2)\} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i.$$

where $Y_i = C_1 V_i(x_1) + C_2 V_i(x_2)$. It is easy to show that $\mathbb{E}(Y_i) = 0$.

$$\begin{aligned} \text{Var}(Y_i) &= \mathbb{E}(Y_i^2) = \mathbb{E} [C_1^2 V_i^2(x_1) + C_2^2 V_i^2(x_2) + 2C_1 C_2 V_i(x_1) V_i(x_2)] \\ &= C_1^2 \mathbb{E} [V_i^2(x_1)] + C_2^2 \mathbb{E} [V_i^2(x_2)] + 2C_1 C_2 \mathbb{E} [V_i(x_1) V_i(x_2)] \\ &= a_1 + a_2 + a_3 = \gamma_y(0), \end{aligned}$$

where

$$\begin{aligned}
a_1 &= C_1^2 \mathbb{E}[V_i^2(x_1)] = C_1^2 \iint [up(x_1) - q(x_1)][vp(x_1) - q(x_1)] R_{uv}(x_1) \pi(u) \pi(v) dudv, \\
a_2 &= C_2^2 \mathbb{E}[V_i^2(x_2)] = C_2^2 \iint [up(x_2) - q(x_2)][vp(x_2) - q(x_2)] R_{uv}(x_2) \pi(u) \pi(v) dudv, \\
a_3 &= 2C_1 C_2 \mathbb{E}[V_i(x_1) V_i(x_2)] = 2C_1 C_2 \iint [up(x_1) - q(x_1)][vp(x_2) - q(x_2)] S_{uv,0}(x_1, x_2) \pi(u) \pi(v) dudv, \\
\text{Cov}(Y_i, Y_j) &= \mathbb{E}\{[C_1 V_i(x_1) + C_2 V_i(x_2)][C_1 V_j(x_1) + C_2 V_j(x_2)]\} \\
&= C_1^2 \mathbb{E}\{V_i(x_1) V_j(x_1)\} + C_1 C_2 \mathbb{E}\{V_i(x_2) V_j(x_1)\} + C_1 C_2 \mathbb{E}\{V_i(x_1) V_j(x_2)\} + C_2^2 \mathbb{E}\{V_i(x_2) V_j(x_2)\} \\
&= b_1 + 2b_2 + b_3 = \gamma_y(s),
\end{aligned}$$

where $s = |i - j|$ and

$$\begin{aligned}
b_1 &= C_1^2 \mathbb{E}\{V_i(x_1) V_j(x_1)\} = C_1^2 \iint [up(x_1) - q(x_1)][vp(x_1) - q(x_1)] G_{uv,s}(x_1) \pi(u) \pi(v) dudv, \\
b_2 &= C_1 C_2 \mathbb{E}\{V_i(x_2) V_j(x_1)\} = C_1 C_2 \iint [up(x_1) - q(x_1)][vp(x_2) - q(x_2)] S_{uv,s}(x_1, x_2) \pi(u) \pi(v) dudv, \\
b_3 &= C_2^2 \mathbb{E}\{V_i(x_2) V_j(x_2)\} = C_2^2 \iint [up(x_2) - q(x_2)][vp(x_2) - q(x_2)] G_{uv,s}(x_2) \pi(u) \pi(v) dudv.
\end{aligned}$$

In the univariate case, we have verified that $\mathbb{E}|V_i(x)|^\delta < \infty$ and $V_i(x)$ is an α -mixing sequence with the mixing coefficient satisfying $\sum_{j \geq 1} \alpha(j)^{1-2/\delta} < \infty$ for some constant $\delta > 2$.

$$\mathbb{E}|Y_i|^\delta = \mathbb{E}|C_1 V_i(x_1) + C_2 V_i(x_2)|^\delta \leq c \left(|C_1|^\delta \mathbb{E}|V_i(x_1)|^\delta + |C_2|^\delta \mathbb{E}|V_i(x_2)|^\delta \right) < \infty,$$

for some constant $c > 0$. In addition, Y_i is a measurable function of $V_i(x_1)$ and $V_i(x_2)$, therefore $\{Y_i\}$ is α -mixing with the mixing coefficient satisfying $\sum_{j \geq 1} \alpha(j)^{1-2/\delta} < \infty$ for some constant $\delta > 2$. By Theorem 2.20 of [Fan and Yao \(2003\)](#), we can obtain that $\text{Var}(S) \rightarrow \gamma_y(0) + 2 \sum_{s=1}^{\infty} \gamma_y(s) = \Sigma_S$.

Therefore, by Theorem 2.21 of [Fan and Yao \(2003\)](#), we obtain that

$$S = C_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1) + C_2 \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2) \rightarrow_D \mathcal{N}(0, \Sigma_S).$$

Thus, as $n \rightarrow \infty$, we have

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1), \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2) \right) \rightarrow_D \mathcal{N}(0, \Sigma(x_1, x_2)), \tag{24}$$

where $\Sigma(x_1, x_2) = \begin{pmatrix} \Sigma_L(x_1) & \Sigma_v(x_1, x_2) \\ \Sigma_v(x_1, x_2) & \Sigma_L(x_2) \end{pmatrix}$.

As $n \rightarrow \infty$, we have $\frac{1}{p_n(x_1)p(x_1)} \rightarrow_P Q^{-1}(x_1)$ and $\frac{1}{p_n(x_2)p(x_2)} \rightarrow_P Q^{-1}(x_2)$.

Let $Q_{n,12}^{-1} = \text{diag}\left(\frac{1}{p_n(x_1)p(x_1)}, \frac{1}{p_n(x_2)p(x_2)}\right)$ and $Q_{12}^{-1} = \text{diag}(Q^{-1}(x_1), Q^{-1}(x_2))$.

We have therefore shown that as $n \rightarrow \infty$,

$$\begin{aligned} (\sqrt{n}(h_n(x_1) - h_0(x_1)), \sqrt{n}(h_n(x_2) - h_0(x_2))) &= Q_{n,12}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_1), \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i(x_2) \right) \\ \rightarrow_D Q_{12}^{-1} \mathcal{N}(0, \Sigma(x_1, x_2)) &= \mathcal{N}(0, \Sigma_{12}), \end{aligned} \quad (25)$$

where $\Sigma_{12} = Q_{12}^{-1} \Sigma(x_1, x_2) Q_{12}^{-1}$.

More generally, we can show that as $n \rightarrow \infty$,

$$(\sqrt{n}(h_n(x_1) - h_0(x_1)), \dots, \sqrt{n}(h_n(x_N) - h_0(x_N))) \rightarrow_D \mathcal{N}(0, \Sigma_N), \quad (26)$$

where $\Sigma_{N,aa} = Q^{-1}(x_a) \Sigma_L(x_a) Q^{-1}(x_a) = \Sigma_0(x_a)$, $\Sigma_{N,ab} = Q^{-1}(x_a) \Sigma_\nu(x_a, x_b) Q^{-1}(x_b)$, $\Sigma_\nu(x_a, x_b) = \gamma_{ab}(0) + 2 \sum_{s=1}^{\infty} \gamma_{ab}(s)$,

$\gamma_{ab}(s) = \mathbb{E}[V_i(x_a) V_j(x_b)] = \iint [up(x_a) - q(x_a)] [vp(x_b) - q(x_b)] S_{uv,s}(x_a, x_b) \pi(u) \pi(v) dudv$ and $\gamma_{ab}(0) = \mathbb{E}[V_i(x_a) V_j(x_b)] = \iint [up(x_a) - q(x_a)] [vp(x_b) - q(x_b)] S_{uv,0}(x_a, x_b) \pi(u) \pi(v) dudv$. Thus, we have proved Theorem 2.

Proof of Theorem 3.

By the definitions of $\hat{f}_n^*(x)$ and $\hat{f}_n(x)$, in view of the function $L(\cdot)$ involved in Assumption 4 and the conditions of Theorem 3, we have

$$\mathbb{E} \left[L \left(\frac{X_i - x}{h_0(x)} \right) \right] = h_0(x) f(x) \int L(u) du + h_0^2(x) f^{(1)}(x) \int u L(u) du + \frac{(1 + o(1)) h_0^3(x) f^{(2)}(x)}{2} \int u^2 L(u) du \quad (27)$$

when $h_0(x) = a_n b_0(x) \rightarrow 0$ as $n \rightarrow \infty$. It follows that

$$\begin{aligned} |\hat{f}_n^*(x) - \hat{f}_n(x)| &= \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n(x)} K \left(\frac{X_i - x}{h_n(x)} \right) - \frac{1}{n} \sum_{i=1}^n \frac{1}{h_0(x)} K \left(\frac{X_i - x}{h_0(x)} \right) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{h_n(x)} K \left(\frac{X_i - x}{h_n(x)} \right) - \frac{1}{h_0(x)} K \left(\frac{X_i - x}{h_0(x)} \right) \right| \\ &= \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{h_n(x)} K \left(\frac{X_i - x}{h_n(x)} \right) - \frac{1}{h_n(x)} K \left(\frac{X_i - x}{h_0(x)} \right) + \frac{1}{h_n(x)} K \left(\frac{X_i - x}{h_0(x)} \right) - \frac{1}{h_0(x)} K \left(\frac{X_i - x}{h_0(x)} \right) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{h_n(x)} \right| \left| K \left(\frac{X_i - x}{h_n(x)} \right) - K \left(\frac{X_i - x}{h_0(x)} \right) \right| + \frac{1}{n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_0(x)} \right) \left| \frac{1}{h_n(x)} - \frac{1}{h_0(x)} \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{h_n(x)} \right| L \left(\frac{X_i - x}{h_0(x)} \right) \left| \frac{X_i - x}{h_n(x)} - \frac{X_i - x}{h_0(x)} \right| + \frac{1}{n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_0(x)} \right) \left| \frac{1}{h_n(x)} - \frac{1}{h_0(x)} \right| \\ &= \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{h_n(x)} \right| L \left(\frac{X_i - x}{h_0(x)} \right) \left| \frac{(X_i - x)(h_n(x) - h_0(x))}{h_0(x) h_n(x)} \right| + \frac{1}{n} \sum_{i=1}^n K \left(\frac{X_i - x}{h_0(x)} \right) \left| \frac{h_n(x) - h_0(x)}{h_0(x) h_n(x)} \right| = O_P(n^{-1/2}), \end{aligned}$$

where the last equality is ensured by (27), Theorem 1, and Assumptions 1 and 2.

It is easy to see that $\widehat{f}_n^*(x) - f(x) = \widehat{f}_n^*(x) - \widehat{f}_n(x) + \widehat{f}_n(x) - f(x)$. Then,

$$\sqrt{nh_0(x)}(\widehat{f}_n^*(x) - f(x)) = \sqrt{nh_0(x)}(\widehat{f}_n^*(x) - \widehat{f}_n(x)) + \sqrt{nh_0(x)}(\widehat{f}_n(x) - f(x)) := I_{1n}(x) + I_{2n}(x),$$

where $I_{1n}(x) = \sqrt{nh_0(x)}(\widehat{f}_n^*(x) - \widehat{f}_n(x)) \leq O_P(n^{-1/2})n^{1/2}\sqrt{h_0(x)} = O_P(1)\sqrt{h_0(x)} = O_P(1)a_n^{1/2}b_0^{1/2}(x) = o_P(1)$.

The bias of $\widehat{f}_n(x)$ is given by

$$\begin{aligned} \text{Bias}\{\widehat{f}_n(x)\} &= \mathbb{E}\left\{\frac{1}{nh_0(x)}\sum_{i=1}^n K\left(\frac{X_i - x}{h_0(x)}\right)\right\} - f(x) = h_0(x)^{-1}\mathbb{E}\left[K\left(\frac{X_1 - x}{h_0(x)}\right)\right] - f(x) \\ &= h_0(x)^{-1}\int K\left(\frac{y-x}{h_0(x)}\right)f(y)dy - f(x) = h_0(x)^{-1}\int f(x+h_0(x)v)K(v)h_0(x)dv - f(x) \\ &= \int\left\{f(x) + f^{(1)}(x)h_0(x)v + \frac{1}{2}f^{(2)}(x)h_0^2(x)v^2 + O(h_0^3(x))\right\}K(v)dv - f(x) \\ &= \left\{f(x) + 0 + \frac{h_0^2(x)}{2}f^{(2)}(x)\int v^2K(v)dv + O(h_0^3(x))\right\} - f(x) = \frac{h_0^2(x)}{2}\mu_2(K)f^{(2)}(x) + o(h_0^2(x)). \end{aligned} \quad (28)$$

The variance of $\widehat{f}_n(x)$ is given by

$$\begin{aligned} \text{Var}\{\widehat{f}_n(x)\} &= \text{Var}\left\{\frac{1}{nh_0(x)}\sum_{i=1}^n K\left(\frac{X_i - x}{h_0(x)}\right)\right\} = \frac{1}{n^2h_0^2(x)}\sum_{i=1}^n \text{Var}\left\{K\left(\frac{X_i - x}{h_0(x)}\right)\right\} \\ &= \frac{1}{nh_0^2(x)}\text{Var}\left\{K\left(\frac{X_1 - x}{h_0(x)}\right)\right\} = \frac{1}{nh_0^2(x)}\left\{\mathbb{E}\left[K^2\left(\frac{X_1 - x}{h_0(x)}\right)\right] - \left[\mathbb{E}K\left(\frac{X_1 - x}{h_0(x)}\right)\right]^2\right\} \\ &= \frac{1}{nh_0^2(x)}\left\{\int K^2\left(\frac{y-x}{h_0(x)}\right)f(y)dy - \left[\int K\left(\frac{y-x}{h_0(x)}\right)f(y)dy\right]^2\right\} \\ &= \frac{1}{nh_0^2(x)}\left\{h_0(x)\int f(x+h_0(x)v)K^2(v)dv - \left[h_0(x)\int f(x+h_0(x)v)K(v)dv\right]^2\right\} \\ &= \frac{1}{nh_0^2(x)}\left\{h_0(x)\int [f(x) + f^{(1)}(\xi)h_0(x)v]K^2(v)dv - O(h_0^2(x))\right\} \\ &= \frac{1}{nh_0(x)}\left\{f(x)\int K^2(v)dv + O(h_0(x)\int |v|K^2(v)dv) - O(h_0(x))\right\} = \frac{1}{nh_0(x)}R(K)f(x) + o\left(\frac{1}{nh_0(x)}\right), \end{aligned} \quad (29)$$

where ξ lies between x and $x + h_0(x)v$.

Based on (28) and (29), we have that

$$\begin{aligned} \mathbb{E}[\widehat{f}_n(x)] - f(x) &= ch_0^2(x) + o(h_0^2(x)). \\ \mathbb{E}[\widehat{f}_n(x) - f(x)]^2 &= \frac{R(K)f(x)}{nh_0(x)} + \frac{1}{4}(\mu_2(K)f^{(2)}(x))^2h_0^4(x) + O(h_0^4(x) + (nh_0(x))^{-1}) \\ &= \frac{R(K)f(x)}{na_nb_0(x)} + \frac{1}{4}(\mu_2(K)f^{(2)}(x))^2a_n^4b_0^4(x) + O(h_0^4(x) + (nh_0(x))^{-1}). \end{aligned}$$

Then, we can further obtain that

$$\begin{aligned}
& \sqrt{nh_0(x)} \left(\widehat{f}_n(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \\
&= \sqrt{nh_0(x)} (\widehat{f}_n(x) - \mathbb{E}[\widehat{f}_n(x)]) + \sqrt{nh_0(x)} \left(\mathbb{E}[\widehat{f}_n(x)] - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \\
&= \sqrt{nh_0(x)} (\widehat{f}_n(x) - \mathbb{E}[\widehat{f}_n(x)]) + o_P \left(h_0^2(x) \sqrt{nh_0(x)} \right) \\
&= \sqrt{nh_0(x)} (\widehat{f}_n(x) - \mathbb{E}[\widehat{f}_n(x)]) + o_P(1) \\
&= \sqrt{nh_0(x)} \frac{1}{nh_0(x)} \sum_{i=1}^n \left(K \left(\frac{X_i - x}{h_0(x)} \right) - \mathbb{E} \left[K \left(\frac{X_i - x}{h_0(x)} \right) \right] \right) + o_P(1) \\
&= \sum_{i=1}^n \frac{1}{\sqrt{nh_0(x)}} \left(K \left(\frac{X_i - x}{h_0(x)} \right) - \mathbb{E} \left[K \left(\frac{X_i - x}{h_0(x)} \right) \right] \right) + o_P(1) = \sum_{i=1}^n Z_{n,i} + o_P(1),
\end{aligned}$$

where $Z_{n,i} = \frac{1}{\sqrt{nh_0(x)}} \left(K \left(\frac{X_i - x}{h_0(x)} \right) - \mathbb{E} \left[K \left(\frac{X_i - x}{h_0(x)} \right) \right] \right)$.

From (29), it is easy to show that $\text{Var} \left(\sum_{i=1}^n Z_{n,i} \right) = R(K) f(x)$. By Liapunov's CLT, we obtain $\sum_{i=1}^n Z_{n,i} \rightarrow_D \mathcal{N}(0, R(K) f(x))$. Therefore, we have $\sqrt{nh_0(x)} \left(\widehat{f}_n(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \rightarrow_D \mathcal{N}(0, R(K) f(x))$, which implies

$$\begin{aligned}
& \sqrt{nh_0(x)} \left(\widehat{f}_n^*(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \\
&= \sqrt{nh_0(x)} (\widehat{f}_n^*(x) - \widehat{f}_n(x)) + \sqrt{nh_0(x)} \left(\widehat{f}_n(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) \\
&= \sqrt{nh_0(x)} \left(\widehat{f}_n(x) - f(x) - \frac{h_0^2(x)}{2} \mu_2(K) f^{(2)}(x) \right) + o_P(1). \\
&\rightarrow_D \mathcal{N}(0, R(K) f(x)),
\end{aligned}$$

We have finally completed the proof of Theorem 3.