Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2015, Article ID 242683, 21 pages
http://dx.doi.org/10.1155/2015/242683

Research Article

Optimal Bandwidth Selection for Kernel Density Functionals Estimation

Su Chen

Department of Mathematical Sciences, The University of Memphis, Memphis, TN 38152, USA

Correspondence should be addressed to Su Chen; schen4@memphis.edu

Received 10 April 2015; Revised 19 June 2015; Accepted 21 June 2015

Academic Editor: Ricardas Zitikis

The choice of bandwidth is crucial to the kernel density estimation (KDE) and kernel based regression. Various bandwidth selection methods for KDE and local least square regression have been developed in the past decade. It has been known that scale and location parameters are proportional to density functionals $\int \gamma(x) f^2(x)dx$ with appropriate choice of $\gamma(x)$ and furthermore equality of scale and location tests can be transformed to comparisons of the density functionals among populations. $\int \gamma(x) f^2(x)dx$ can be estimated nonparametrically via kernel density functionals estimation (KDFE). However, the optimal bandwidth selection for KDFE of $\int \gamma(x) f^2(x)dx$ has not been examined. We propose a method to select the optimal bandwidth for the KDFE. The idea underlying this method is to search for the optimal bandwidth by minimizing the mean square error (MSE) of the KDFE. Two main practical bandwidth selection techniques for the KDFE of $\int \gamma(x) f^2(x)dx$ are provided: Normal scale bandwidth selection (namely, "Rule of Thumb") and direct plug-in bandwidth selection. Simulation studies display that our proposed bandwidth selection methods are superior to existing density estimation bandwidth selection methods in estimating density functionals.

1. Introduction

Suppose that a random variable X with a probability density function (p.d.f.) $f(x)$ belongs to a location-scale family. Let μ and σ be the location and scale parameter of X, respectively. We have $f(x) = (1/\sigma) f_0((x-\mu)/\sigma)$ for some base function $f_0(x)$. If $f(x)$ is a symmetric function, then $f_0(x)$ is usually chosen to be the same class of distribution with mean zero. For instance, if $f(x)$ is the p.d.f. of Normal distribution with mean μ and standard deviation σ, then $f_0(x)$ is usually chosen to be the density of standard Normal distribution. In the nonparametric world, $f(x)$ is not assumed to have any prespecified distributional format. Therefore, μ and σ are unknown and can not be estimated by any distribution based method such as maximum likelihood estimate. Ahmad [1] proposed a nonparametric kernel estimation of location and scale parameters via density functionals estimation with known base functions. The location and scale functions are written in terms of density functionals as follows:

$$\sigma = \frac{\int f_0^2(x)\,dx}{\int f^2(x;\mu,\sigma)\,dx}, \tag{1}$$

$$\mu = \frac{\int x f^2(x;\mu,\sigma)\,dx - \int x f_0^2(x)\,dx}{\int f^2(x;\mu,\sigma)\,dx} \tag{2}$$

$$= \frac{\sigma\left[\int x f^2(x;\mu,\sigma)\,dx - \int x f_0^2(x)\,dx\right]}{\int f_0^2(x;\mu,\sigma)\,dx}. \tag{3}$$

Apparently, the location μ and scale σ only rely on two functionals of unknown density $f(x)$, namely, $\int f^2(x;\mu,\sigma)dx$ and $\int x f^2(x;\mu,\sigma)dx$, if $f_0(x)$ is known. Ahmad [1] showed that the new kernel location and scale estimates had better asymptotic property than MLE. Simulation results in Ahmad and Amezziane [2], a subsequent work of Ahmad [1], indicated that the kernel location and scale estimators have a comparable variability to that of the MLE and smaller than that of Huber's M-estimator. However, it is usually difficult or impossible to know the base density especially in the nonparametric world. Moreover $f(x)$ can be derived in terms of μ and σ if the base density $f_0(x)$ is given. In this case, it becomes a parametric situation and MLE can be considered. From this point of view, Ahmad's scale and location estimates

are not very practical in real world application because the base density function $f_0(x)$ needs to be known first.

Chen [3] proposed kernel-based nonparametric tests of equality of scale and location parameters among K populations based on the kernel scale and location estimators proposed by Ahmad [1]. To test $H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_K$ is equivalent to test $H_0 : \int f_1^2(x)dx = \int f_2^2(x)dx = \cdots = \int f_K^2(x)dx$ according to (1), where σ_i and $f_i(x)$ are the scale and density function of ith population, respectively, and $i = 1, 2, \ldots, K$. Likewise, $H_0 : \mu_1 = \mu_2 = \cdots = \sigma_K$ is equivalent to $H_0 : \int x f_1^2(x)dx = \int x f_2^2(x)dx = \cdots = \int x f_K^2(x)dx$ by (3) if homogeneous scale is assumed. This fact motivates Chen [3] to build test statistics for equality of scale and location on the density functionals estimation of $\int f_i^2(x)dx$ and $\int x f_i^2(x)dx$, respectively. Chen [3] brought a new life to the two kernel density functionals estimations, which were originally introduced to estimate location and scale parameter by Ahmad [1]. When comparing the scale (or location) parameters among K populations, the differences in scale (or location) can be completely determined by $\int f^2(x)dx$ (or $\int x f^2(x)dx$) and $f_0(x)$ becomes irrelevant if we assume K populations are from same distributional family but differed only in locations and/or scales. Thus the assumption of having to know base density $f_0(x)$ as required in kernel scale and location estimation was successfully dropped. To find a good estimate of density functionals $\int f^2(x)dx$ and $\int x f^2(x)dx$ becomes our next concern.

Aubuchon and Hettmansperger [4] proposed a kernel estimation of $\int f^2(x)dx$ by a convolution of kernel density estimation function with the empirical CDF and showed its asymptotic equivalence to Lehmann's estimator (see Lehmann [5] for details) based on the Wilcoxon confidence interval. Ahmad [1] provided two approaches to estimate $\int f^2(x)dx$ and $\int x f^2(x)dx$, one is similar to Aubuchon and Hettmansperger [4] and the other is to approximate the density function $f(x)$ with an orthogonal series expansion, and then estimate the functionals of density. Grübel [6] estimated the density functionals $\int \phi(f(x))dx$ for known $\phi(\cdot)$ under certain conditions through the kernel density estimate of the unknown $f(\cdot)$.

Choice of bandwidth (window width or smoothing parameter) is crucial for every kernel based procedure, such as kernel density estimation and kernel regression. A vast amount of literatures has been devoted in choosing practical optimal bandwidth for techniques built on kernel estimation. Representative surveys of bandwidth selection techniques can be found in Bowman [10], Jones et al. [11], Loader [12], and Wand and Jones [13]. Jones et al. [11] grouped data-based bandwidth selection methods for density estimation into "first generation" method and "second generation" method. The first generation methods, including the least-square cross-validation (LSCV) in Bowman [8] and biased cross-validation in Scott and Terrell [14], suffer from a slow relative rate of convergence to h_{MISE} of order $n^{-1/10}$. Härdle and Marron [15] applied the least squares cross-validation idea to bandwidth selection on Nadaraya-Watson estimator. The second generation methods are mainly based on plug-in techniques. The idea of "plug-in" is to replace f'' with a consistent estimate $\widehat{f}''$ first proposed by Nadaraya [16] and Woodroofe [17]; however, the practical choice of pilot bandwidth was not discussed. Sheather and Jones [9] proposed a refined plug-in method, so-called "solve-the-equation (STE)" plug-in $\widehat{h}_{\text{STE}}$, which has faster rate of convergence of order $n^{-5/14}$ than cross-validation estimators. Smoothed cross-validation (SCV) is also a plug-in type method with pilot bandwidth of format $g = Cn^p h^m$. It was developed by Hall et al. [18]. Müller [19] and Staniswalis [20] employed the idea in the kernel regression. Hall et al. [21] constructed root-n bandwidth selectors and achieved the optimal $n^{-1/2}$ relative rate of convergence by appropriate choice of the parameters in pilot bandwidth $g = Cn^p h^m$. Gasser et al. [22] and Ruppert et al. [23] borrowed the simple direct plug-in idea to local linear regression. Fan and Gijbels [24] applied "cross-validation technique, the Normal-reference method, and the plug-in approach" for the density estimation setting to their corresponding bandwidth selectors for local polynomial regression method.

However, few literature studies the optimal bandwidth for the estimation of $\int f^2(x)dx$ and $\int x f^2(x)dx$, which are two important density functionals for estimating location and scale parameters as discussed in the prior paragraphs. Aubuchon and Hettmansperger [4] chose the bandwidth by removing the $1/n$ bias term. Grübel [6] suggested using the MISE-optimal choice of bandwidth in kernel density estimation. Chen [3] uses the least-square cross-validation bandwidth selection method for density estimation. In this paper, we will derive optimal bandwidth selection of kernel location and scale estimation by minimizing the MSE of the kernel functionals estimation for $\int f^2(x)dx$ and $\int x f^2(x)dx$. This paper will also propose two practical bandwidth selection methods and then compare them with various bandwidth selections for kernel density estimation such as Rule-of-Thumb, direct plug-in (DPI), least square cross-validation, and biased cross-validation (BCV).

For simplicity of illustration, a unified format (i.e., $\int \gamma(x) f^2(x)dx$) of the two density functionals mentioned above will be used throughout the paper. When $\gamma(x) = 1$ and x, it equals $\int f^2(x)dx$ and $\int x f^2(x)dx$, respectively. The paper is organized as follows. The optimal bandwidth for estimation $\int \gamma(x) f^2(x)dx$ in terms of AMSE criterion is derived in Section 2.1. Two practical bandwidth selection methods for $\int \gamma(x) f^2(x)dx$ are provided in Sections 2.2 and 2.3 when $\gamma(x) = 1$ and x. Asymptotic distribution of direct plug-in bandwidth for kernel functionals estimation of $\int \gamma(x) f^2(x)dx$ is given in Section 2.3 as well. Section 3 conducts three simulation studies to explore the properties of proposed bandwidth selection methods and evaluate their performance compared to several classical bandwidth selection methods for kernel density estimation.

2. Main Results

2.1. Optimal Bandwidth Selection. Define $U_1 = \int f^2(x)dx$ and $U_2 = \int xf^2(x)dx$. Let us write U_1 and U_2 in a more general density functionals $U = \int \gamma(x)f^2(x)dx$. Note that U_1 and U_2 are special cases of U, where $\gamma(x)$ is 1 and x, respectively. Suppose $X_1, X_2, \ldots, X_n$ are n independent random variable from a distribution with density function $f(x)$, where $f(x)$ is unknown. Similar to Aubuchon and Hettmansperger [4] and Grübel [6], we obtain the kernel density functionals estimate of U by $\int \gamma(x)\widehat{f}_n(x)dF_n(x)$, where $\widehat{f}_n(x)$ is the kernel density estimate of f and $F_n(x)$ is the empirical CDF. Thus, a kernel density functionals estimate of U is given by

$$\widehat{U} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\gamma(X_i)+\gamma(X_j)}{2}K_h(X_i - X_j), \quad (4)$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$ and $K(\cdot)$ is the kernel function (details can be found in Wand's book). The following theorem provides the mean and variance of $\widehat{U}$ in (4) for fixed h.

Theorem 1. *For $\widehat{U}$ in (4), the expected value and variance of $\widehat{U}$ are given by*

$$E(\widehat{U}) = U + \frac{1}{2}E[\gamma(y)f''(y)]\mu_2(K)h^2 + o(h^2), \quad (5)$$

$$\begin{aligned}\operatorname{Var}(\widehat{U}) &= 4n^{-1}\left\{\int \gamma^2(y)f^3(y)dy - \left(\int \gamma(y)f^2(y)dy\right)^2\right\} \\ &\quad + 2n^{-2}h^{-1}R(K)\int \gamma^2(y)f^2(y)dy \\ &\quad + \left\{o(n^{-1}) + o(n^{-2}h^{-1})\right\},\end{aligned} \quad (6)$$

where $R(g) = \int g^2(y)dy$ and $\mu_2(g) = \int y^2 g(y)dy$.

We prove this in Appendix A. The first term in (6) is nonnegative by Jensen's inequality. Then the MSE of $\widehat{U}$ can be written as follows:

$$\begin{aligned}\operatorname{MSE}(\widehat{U}) &= \operatorname{Var}(\widehat{U}) + \operatorname{Bias}(\widehat{U})^2 \\ &= 4n^{-1}\left\{\int \gamma^2(y)f^3(y)dy - \left(\int \gamma(y)f^2(y)dy\right)^2\right\} \\ &\quad + 2n^{-2}h^{-1}R(K)\int \gamma^2(y)f^2(y)dy \\ &\quad + \frac{1}{4}E^2[\gamma(y)f''(y)]\mu_2^2(K)h^4 \\ &\quad + \left\{o(n^{-2}h^{-1}) + o(h^4)\right\}.\end{aligned} \quad (7)$$

Therefore, the optimal bandwidth selection for density functionals estimation of $U = \int \gamma(x)f^2(x)dx$ is $h_{\mathrm{MSE};U}$, the minimizer of $\operatorname{MSE}(\widehat{U})$. To obtain a closed form of optimal bandwidth for kernel functionals estimation of U, the minimizer of the asymptotic mean square error (AMSE) of $\widehat{U}$ is studied instead. The optimal bandwidth for estimation of U with respect to AMSE criterion is given by

$$h_{\mathrm{AMSE};U} = \left[\frac{2R(K)R(\gamma(x)f(x))}{\left(\int \gamma(x)f''(x)f(x)dx\right)^2\mu_2^2(K)}\right]^{1/5} n^{-2/5}, \quad (8)$$

where

$$\operatorname{MSE}(\widehat{U}) = \operatorname{AMSE}(\widehat{U}) + \left\{o(n^{-1}) + o(h^4)\right\}. \quad (9)$$

However, h_{AMSE} in (8) is not computable since $R(\gamma(x)f(x)) = \int \gamma^2(x)f^2(x)dx$ and $\int \gamma(x)f''(x)f(x)dx$ depend on unknown function $f(x)$. A quick and simple guess of AMSE-optimal bandwidth is "Normal scale" bandwidth. It gives reasonable answers whenever the data are close to Normal. In the next section, Normal scale bandwidth selection will be studied for $\gamma = 1$ and x, respectively.

2.2. Normal Scale Bandwidth Selection. When $\gamma = 1$, $U = \int \gamma(x)f^2(x)dx$ reduces to $U_1 = \int f^2(x)dx$. By (8) in Section 2.1, the bandwidth that minimizes $\operatorname{MSE}(\widehat{U}_1)$ asymptotically is

$$h_{\mathrm{AMSE};U_1} = \left[\frac{2R(K)R(f)}{E^2[f''(y)]\mu_2^2(K)}\right]^{1/5} n^{-2/5}, \quad (10)$$

where $\widehat{U}_1$ is the kernel density functionals estimation of U_1.

Proposition 2. *If f is Normal with mean 0 and variance σ^2 then the Normal scale AMSE-optimal bandwidth selector for $\widehat{U}_1$ is given by*

$$\widehat{h}_{NS;U_1} = \left[\frac{16\sqrt{\pi}R(K)}{\mu_2^2(K)}\right]^{1/5}\widehat{\sigma}n^{-2/5}, \quad (11)$$

where $\widehat{\sigma}$ is some estimate of σ.

The proof of Proposition 2 can be found in Appendix B. If Gaussian kernel is chosen, that is, $K(\cdot)$ is the density of standard Normal distribution, then $R(K) = 1/2\sqrt{\pi}$ and $\mu_2(K) = 1$. Hence (11) is simplified to

$$\widehat{h}_{\mathrm{ROT};U_1} = \sqrt[5]{8}\widehat{\sigma}n^{-2/5} \simeq 1.515717\widehat{\sigma}n^{-2/5}, \quad (12)$$

which can be called "Rule-of-Thumb" (ROT) bandwidth selector for kernel scale estimation.

When $\gamma = x$, U becomes $U_2 = \int xf^2(x)dx$. The bandwidth selector that minimizes $\operatorname{AMSE}(\widehat{U}_2)$ is

$$h_{\mathrm{AMSE};U_2} = \left[\frac{2R(K)R(xf(x))}{E^2[xf''(x)]\mu_2^2(K)}\right]^{1/5} n^{-2/5}, \quad (13)$$

followed by (8) in Section 2.1.

Proposition 3. *If f is Normal with mean μ and variance σ^2 then the Normal scale AMSE-optimal bandwidth selector for $\widehat{U}_2$ is given by*

$$\widehat{h}_{NS;U_2} = \left[\left(2+\frac{\widehat{\sigma}^2}{\widehat{\mu}^2}\right)\frac{8\sqrt{\pi}R(K)}{\mu_2^2(K)}\right]^{1/5}\widehat{\sigma}n^{-2/5}, \tag{14}$$

where $\widehat{\sigma}$ is an estimate of σ and $\widehat{\mu}$ is an estimate of μ. If $\widehat{\sigma} = s$ (sample standard deviation) and $\widehat{\mu} = \overline{x}$ (sample mean) then (14) can be rewritten as

$$\widehat{h}_{NS;U_2} = \left[\left(2+\widehat{c}_v^2\right)\frac{8\sqrt{\pi}R(K)}{\mu_2^2(K)}\right]^{1/5}\widehat{\sigma}n^{-2/5}, \tag{15}$$

where $\widehat{c}_v = s/\overline{x}$ is the coefficient of variation (CV). Particularly when $\mu = 0$ for fixed σ, $\widehat{h}_{NS;U_2}$ goes to infinity.

The proof of Proposition 3 is given in Appendix C. When kernel function $K(\cdot)$ is the density of standard Normal distribution, then the "Rule-of-Thumb" bandwidth selector for kernel location estimation is

$$\widehat{h}_{\mathrm{ROT};U_2} = \sqrt[5]{4\left(2+\widehat{c}_v^2\right)}\widehat{\sigma}n^{-2/5}. \tag{16}$$

Both (14) and (16) infer that the larger the location of $f(x)$ in absolute value is, the smaller the optimal bandwidth is needed. In another word, the optimal-AMSE bandwidth for $N(0,\sigma^2)$ goes to infinity. This fact also not merely applies to Normal with zero mean but also can be extended to any distribution with p.d.f. an even function (symmetric distribution around zero).

Corollary 4. *For any distribution with even density function $f(x)$, that is, $f(-x) = f(x)$, then the optimal-AMSE bandwidth selector is $h_{AMSE;U_2} = \infty$.*

Remarks. (1) The optimal bandwidth for estimation of U_1 is not effected by the location of $f(x)$, but the scale parameter. However, the optimal bandwidth for estimation of U_2 not only depends on scale but also varies along with the location. This fact will be illustrated by the simulation study in Section 3.1. Note that scale parameter is determined by U_1 and location parameter is determined by U_2.

(2) The common choice of $(\widehat{\sigma})$ is sample standard deviation s as in Silverman [7]. However, Wand and Jones [13] recommended the smaller value between s and interquartile range $\widehat{\sigma}_{\mathrm{IQR}}$. Janssen et al. [25] also studied other more sophisticated estimates of σ.

2.3. Direct Plug-In Bandwidth Selection. If the distribution of X_i's, that is, $f(x)$, departs far from Normal distribution, then Normal scale bandwidth selector will be problematic. Note that $R(\gamma(y)f(y))$ and $\int\gamma(y)f''(y)f(y)dy$ in (8) are unknown and need to be estimated to obtain a practical optimal bandwidth selector. A natural estimate of $R(\gamma(y)f(y))$ is

$$\begin{aligned}\widehat{R} &= \int\gamma^2(y)\widehat{f}(y)\,dF_n(y)\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\gamma^2(X_i)+\gamma^2(X_j)}{2}K_g\left(X_i-X_j\right).\end{aligned} \tag{17}$$

Similarly, $\phi = \int\gamma(y)f''(y)f(y)dy$ can be estimated by

$$\begin{aligned}\widehat{\phi} &= \int\gamma(y)\widehat{f}''(y)\,dF_n(y) = \frac{1}{n(n-1)g^3}\\ &\cdot\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\gamma(X_i)+\gamma(X_j)}{2}K''\left\{\frac{\left(X_i-X_j\right)}{g}\right\}.\end{aligned} \tag{18}$$

Replacement of $R(\gamma(y)f(y))$ and $\int\gamma(y)f''(y)f(y)dy$ by $\widehat{R}$ and $\widehat{\phi}$ leads to the *direct plug-in* (DPI) bandwidth selector for $U = \int\gamma(x)f^2(x)dx$:

$$\widehat{h}_{\mathrm{DPI};U} = \left[\frac{2R(K)\widehat{R}}{\mu_2^2(K)\widehat{\phi}^2}\right]^{1/5}n^{-2/5}. \tag{19}$$

Obviously, the kernel density functionals estimates in (17) and (18) rely on the choice of pilot bandwidth g. Simple candidates for pilot bandwidth are to use *Normal scale* bandwidth selector proposed in Section 2.2 for $\int\gamma(x)f^2(x)dx$ or smoothing parameters for traditional density estimate (e.g., ROT, LSCV, BCV, and DPI surveyed in Wand and Jones [13]). The DPI bandwidth selection can be practically computed through the following procedures.

Step 1. Estimate g using the Normal scale bandwidth proposed in Section 2.2 (i.e., $\widehat{h}_{\mathrm{NS};U_1}$ for estimation of U_1 and $\widehat{h}_{\mathrm{NS};U_2}$ for estimation of U_2) or bandwidth selection for density estimation (such as ROT [7], LSCV [8], BVC [14], and DPI [9]).

Step 2. Estimate $R(\gamma(y)f(y))$ and $\int\gamma(y)f''(y)f(y)dy$ using $\widehat{R}(g)$ in (17) and $\widehat{\phi}(g)$ in (18).

Step 3. The DPI bandwidth selection for U is obtained followed by (19).

The performance of these pilot bandwidth selections is compared in terms of MSE of $\widehat{U}$ through Monte Carlo simulation in Section 3.2 (*Simulation Study 2*). Next, we will study the asymptotic distribution of $\widehat{h}_{\mathrm{DPI};U}$. The limiting distribution of practical bandwidth selector is very important in that the rate of convergence is the chief concern.

Proposition 5. *If $\gamma(y)$ and density function $f(y)$ are continuous and satisfy $\int\gamma^2(y)f^2(y)dy < \infty$ and $\int\gamma(y)f''(y)f(y)dy < \infty$, then*

$$n^{1/5}\left(\frac{\widehat{h}_{DPI;U}}{h_{MSE;U}}-1\right)\longrightarrow_D N\left(0,\sigma^2_{DPI;U}\right). \tag{20}$$

The proof of Proposition 5 is provided in Appendix D. Thus the direct plug-in bandwidth selection for functional density estimation $\widehat{h}_{\mathrm{DPI};U}$ has relative convergence rate of order $n^{-1/5}$.

Remarks. Particularly, when $\gamma(y) = 1$, the DPI bandwidth selector for estimation of U_1 is $\widehat{h}_{\mathrm{DPI};U_1} = [2R(K)\widehat{R}_1/\mu_2^2(K)\widehat{\phi}_1^2]^{1/5}n^{-2/5}$, where $\widehat{R}_1 = (1/n(n-1))\sum_{i=1}^n\sum_{j\neq i}^n K_g(X_i - X_j)$ and $\widehat{\phi}_1 = (1/n(n-1)g^3)\sum_{i=1}^n\sum_{j\neq i}^n K''\{(X_i - X_j)/g\}$. Likewise, when $\gamma(y) = y$, the DPI bandwidth selector for estimation of U_2 is $\widehat{h}_{\mathrm{DPI};U_2} = [2R(K)\widehat{R}_2/\mu_2^2(K)\widehat{\phi}_2^2]^{1/5}n^{-2/5}$, where $\widehat{R}_2 = (1/n(n-1))\sum_{i=1}^n\sum_{j\neq i}^n((X_i^2 + X_j^2)/2)K_g(X_i - X_j)$ and $\widehat{\phi}_2 = (1/n(n-1)g^3)\sum_{i=1}^n\sum_{j\neq i}^n((X_i + X_j)/2)K''\{(X_i - X_j)/g\}$.

3. Simulation Study

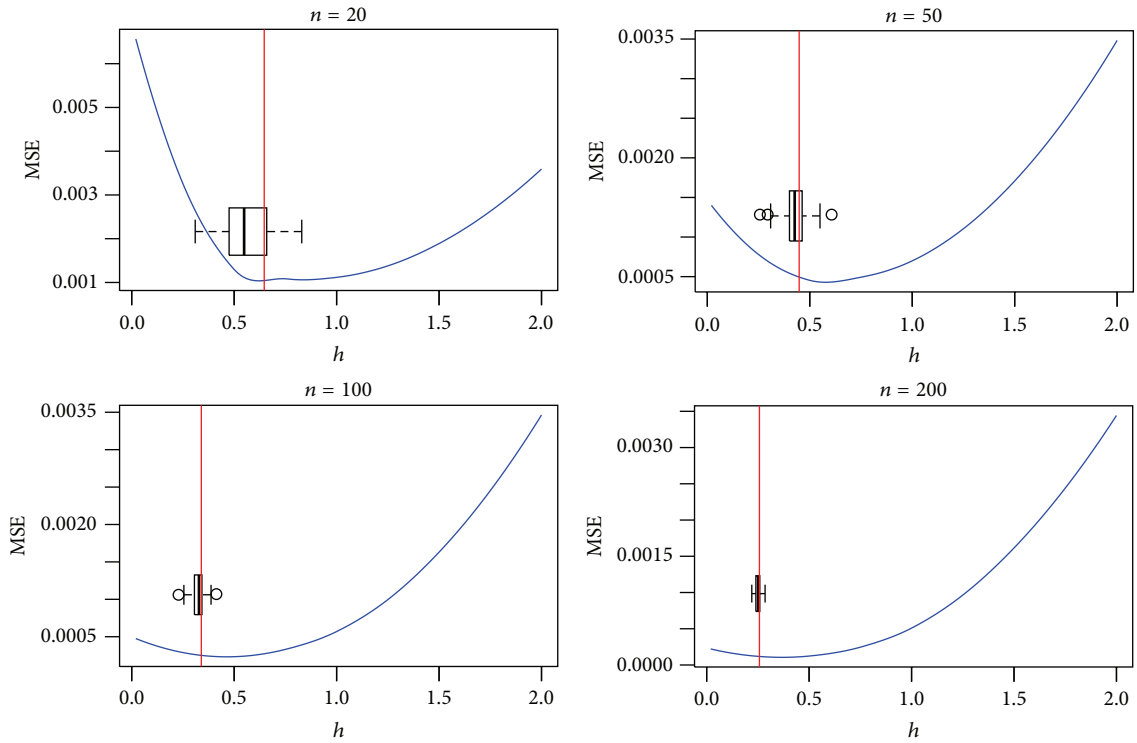
Three simulation studies are carried out to evaluate [*Simulation Study 1*] the accuracy of $\widehat{h}_{\mathrm{NS};U_1}$ and $\widehat{h}_{\mathrm{NS};U_2}$ (Normal scale bandwidth for U_1 and U_2) comparing to $h_{\mathrm{AMSE};U}$ and $h_{\mathrm{MSE};U}$ under normality assumption; [*Simulation Study 2*] the optimal choices of pilot bandwidth for $\widehat{h}_{\mathrm{DPI};U_1}$ and $\widehat{h}_{\mathrm{DPI};U_2}$ in terms of MSE of $\widehat{U}_1$ and $\widehat{U}_2$, respectively; [*Simulation Study 3*] the performance of proposed practical optimal bandwidth selection methods (ROT and DPI proposed in Sections 2.2 and 2.3) versus traditional (classical) bandwidth selection for kernel density estimate in terms of MSE of $\widehat{U}$. As to the choice of kernel function $K(\cdot)$, it has been shown in literatures that the choice of bandwidth overrides the effect of choice of kernel function. So for simplicity, we just use the Gaussian kernel in all the three simulation studies.

3.1. Simulation Study 1. The purpose of this study is threefold (1) to evaluate the performance of $\widehat{h}_{\mathrm{NS};U_1}$ and $\widehat{h}_{\mathrm{NS};U_2}$ when samples are from Normal distribution, (2) to study $h_{\mathrm{MSE};U_1}$ and $h_{\mathrm{MSE};U_2}$ (the optimal bandwidths that minimize the MSE of $\widehat{U}_1$ and $\widehat{U}_2$, resp.) in terms of the location parameter of Normal distribution, and (3) to illustrate numerically that optimal bandwidth that minimizes the MSE of U_2 goes to infinity when location parameter gets closer to zero.

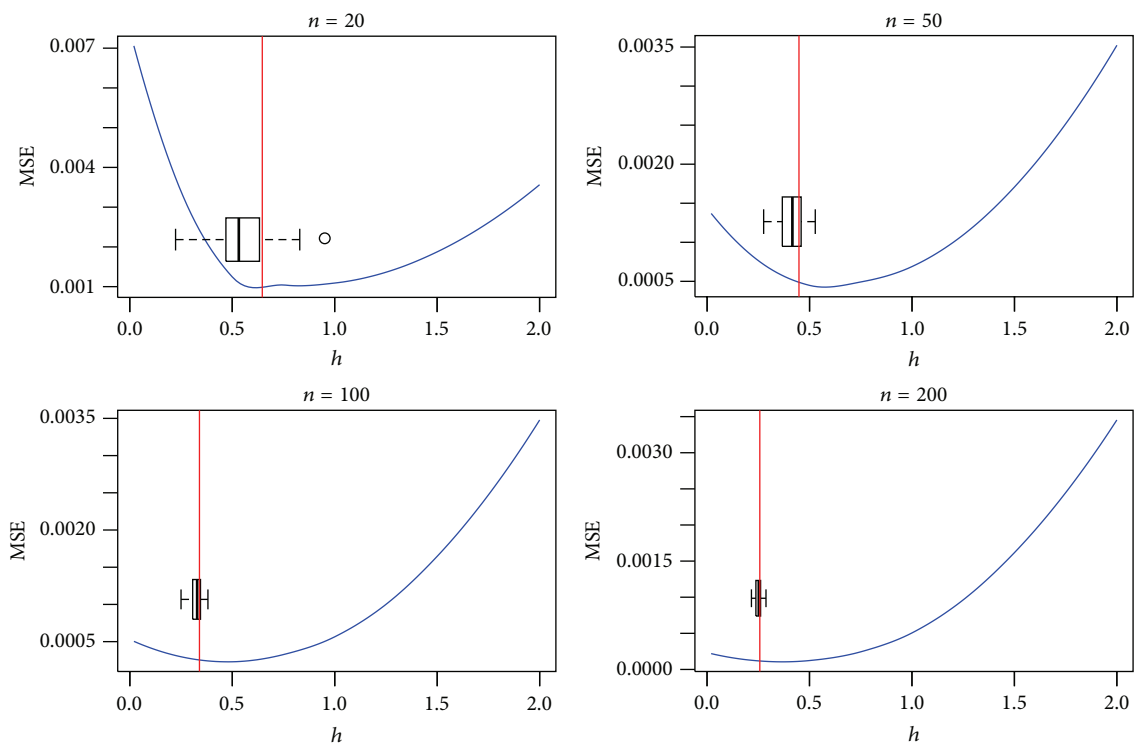
Figure 1 plots the MSE of $\widehat{U}_1$ versus the choice of bandwidth when sample of sizes 20, 50, 100, and 200 is drawn from $N(0, \sqrt{2})$ and $N(1, \sqrt{2})$, respectively (the simulation result is not sensitive to the choice of scale). The blue curve in each subplot represents the MSE($\widehat{U}_1$) as bandwidth ranges from 0 to 2. The minimum point of the blue curve indicates $h_{\mathrm{MSE};U_1}$. The red vertical line in the subplot represents $h_{\mathrm{AMSE};U_1}$ and is computed from (10) by replacing $f(y)$ with the p.d.f. of $N(\mu, \sqrt{2})$, where $\mu = 0$ in Figure 1(a) and $\mu = 1$ in Figure 1(b). $\widehat{h}_{\mathrm{NS};U_1}$ is an estimate of $h_{\mathrm{AMSE};U_1}$ (an asymptotic approximation of $h_{\mathrm{MSE};U_1}$) under normality assumption. Simulation results in Figure 1 show that $\widehat{h}_{\mathrm{NS};U_1}$ tends to have small variance and stabilized around the true $h_{\mathrm{MSE};U_1}$ for normality data. The optimal bandwidth does not change with location parameter μ as shown in Figure 1(a) ($\mu = 0$) and Figure 1(b) ($\mu = 1$) (more simulation results based on location parameters other than 0 and 1 are available upon request.).

Figure 2 plots the MSE of $\widehat{U}_2$ versus the choice of bandwidth when sample of sizes 20, 50, 100, and 200 is drawn from $N(0, \sqrt{2})$, $N(0.5, \sqrt{2})$, $N(1, \sqrt{2})$, and $N(5, \sqrt{2})$, respectively. Similar to Figure 1, the blue curve and red vertical line represent MSE($\widehat{U}_2$) and $h_{\mathrm{AMSE};U_2}$. The boxplot of $\widehat{h}_{\mathrm{NS};U_2}$ is based on the 100 sets of simulated samples of size $n = 10, 50, 100, 200$ from Normal distribution with mean $\mu = 0, 0.5, 1, 5$ and standard deviation $\sqrt{2}$. The red vertical line disappears in Figure 1(a) due to the fact that $h_{\mathrm{AMSE};U_2} \to \infty$ as $\mu \to 0$. Also the MSE of $\widehat{U}_2$ (blue curve) strictly decreases as h rises in Figure 2(a). To conclude from Figure 1(a), the optimal bandwidth for kernel functional estimation of U_2 goes to infinity when mean of underlying distribution is zero, which is consistent to Proposition 3. However, the boxplot in Figure 2(a) infers that the distribution of $\widehat{h}_{\mathrm{NS};U_2}$ is right-skewed with median, 1st quantile, and 3rd quantile around one, which is far departure from the true value $h_{\mathrm{MSE};U_2}$ as well as $h_{\mathrm{AMSE};U_2}$. When μ slightly deviates from zero, just as the case in Figures 2(b) and 2(c), $\widehat{h}_{\mathrm{NS};U_2}$ tends to be less variate (and skewed), overlap with $h_{\mathrm{AMSE};U_2}$ (red vertical line), and get closer to $h_{\mathrm{MSE};U_2}$ (valley of blue curve), especially as the sample size grows. When μ increases up to 1 and above as shown in Figure 2(d), the median of $\widehat{h}_{\mathrm{NS};U_2}$, $h_{\mathrm{AMSE};U_2}$, and $h_{\mathrm{MSE};U_2}$ coincide when sample size is 50+. More simulation results can be found in Supplementary Material available online at http://dx.doi.org/10.1155/2015/242683.

3.2. Simulation Study 2. Several candidate bandwidth selection methods are available to serve as a pilot bandwidth g, such as classical bandwidth selection methods for kernel density estimate (described in Section 2.3 Step 1), the optimal bandwidth for U (U_1 or U_2) on Normal scale references, namely, "$\widehat{h}_{\mathrm{NS};U_1}$" and "$\widehat{h}_{\mathrm{NS};U_2}$" in Section 2.2. This subsection aims to study the pilot bandwidth for $\widehat{R}$ and $\widehat{\phi}$ required to estimate U in (19) (note that U is simplified to U_1 and U_2 when $\gamma(x) = 1$ and x, resp.). Five different choices of pilot bandwidth are studied in this subsection to estimate $\widehat{R}$ and $\widehat{\phi}$ for $\widehat{h}_{\mathrm{DPI};U}$ under five different underlying distributions: (i) Cauchy with location (μ was set to be 1 in simulation study for estimation of U_1 and 5 for estimation of U_2) μ and scale 2/3, (ii) Generalized Pareto with location μ, scale 2/3, and shape 1, (iii) Normal with location μ and scale 2/3, (iv) Mix-Normal I, and (v) Mix-Normal II. Mix-Normal I and Mix-Normal II are weighted mixtures of two Normal distributions: Mix-Normal I is $(1/2)N(-2, 2/3) + (1/2)N(0, 2/3)$ and Mix-Normal I is $(3/4)N(0, 1) + (1/2)N(1.5, 1/3)$. Cauchy distribution is a well-known fat-tailed symmetric distribution. Generalized Pareto with shape 1 is a extremely fat-tailed asymmetric (right-skewed) distribution. Density curves of the two Mix-Normal -distributions are given in Figure 3. It is obvious that both Mix-Normal I and Mix-Normal II are bimodal distributions in contrast to unimodal distributions such as Cauchy, Generalized Pareto, and Normal. Mix-Normal I is symmetric bimodal and Mix-Normal II is asymmetric

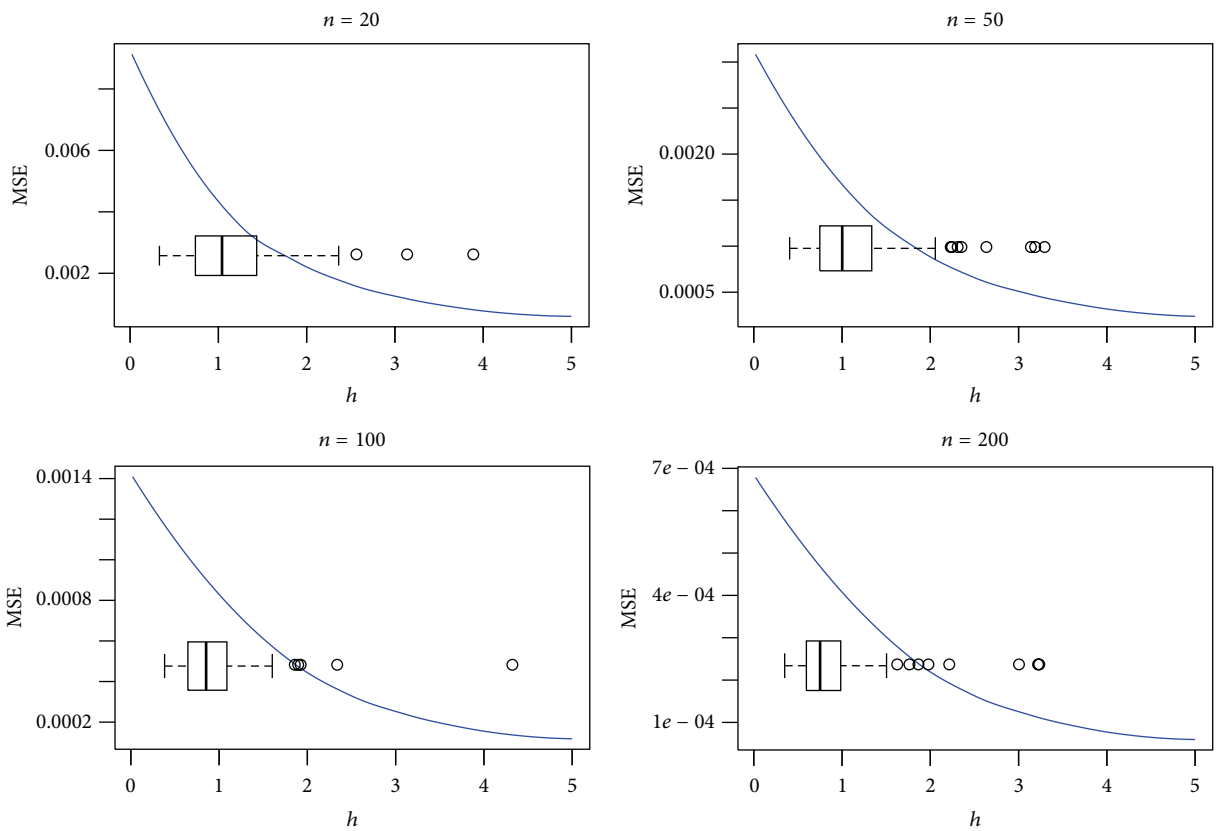


(a) $\mu = 0$

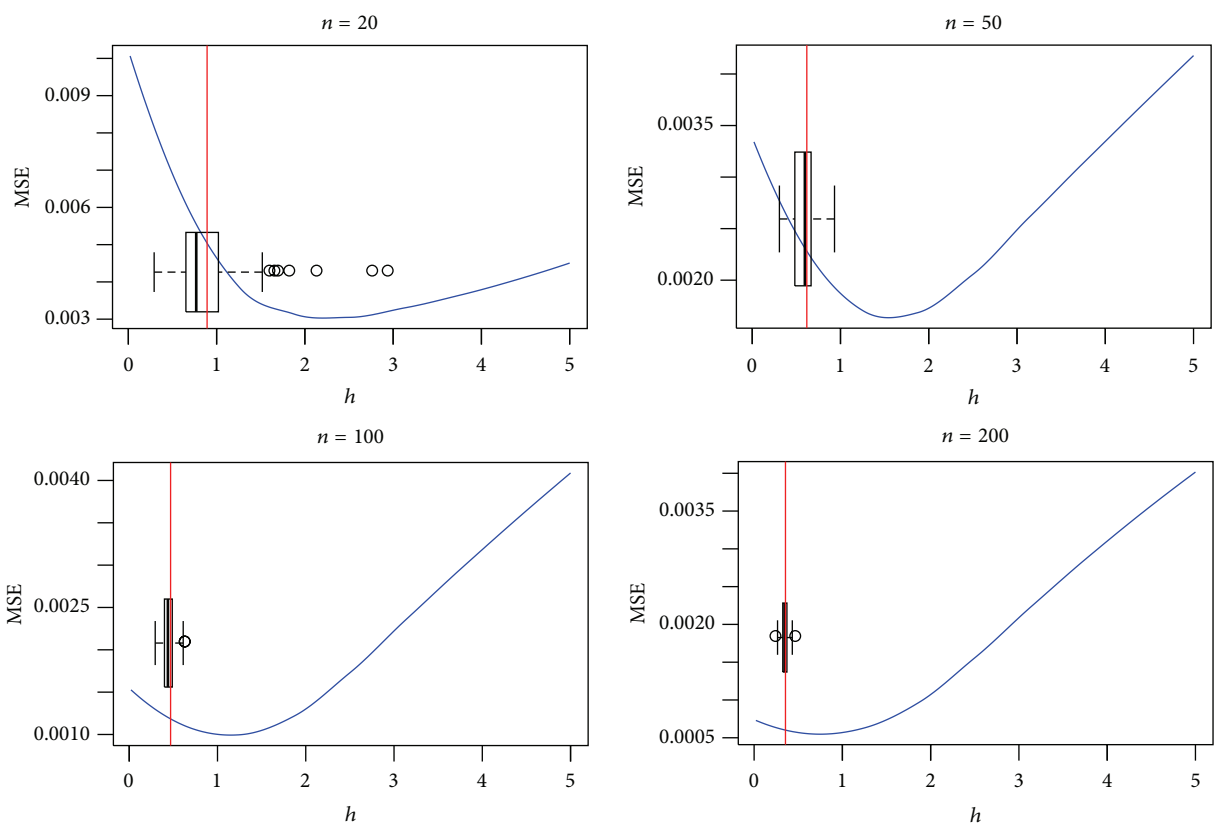


(b) $\mu = 1$

FIGURE 1: The MSE of $\widehat{U}_1$ when underlying distribution is Normal with mean $\mu = 0$ (a), $\mu = 1$ (b) and standard deviation $\sqrt{2}$. The blue curve is the MSE($\widehat{U}_1$) versus the bandwidth h. The boxplot of $\widehat{h}_{\mathrm{ROT};U_1}$ is based on the 100 sets of simulated samples of size $n = 20, 50, 100, 200$ from Normal distribution with mean μ and standard deviation $\sqrt{2}$. The red vertical line represents $h_{\mathrm{AMSE};U_1}$ if f is Normal with known mean 0 and variance σ^2.

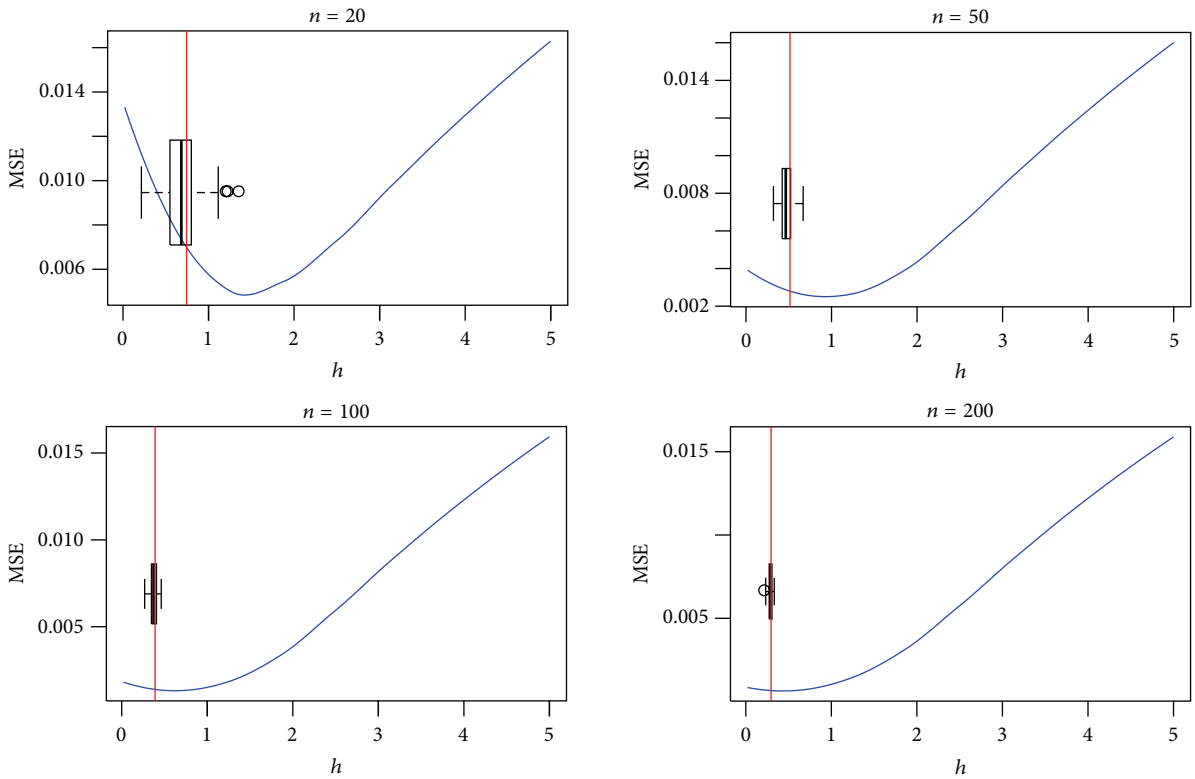


(a) $\mu = 0$

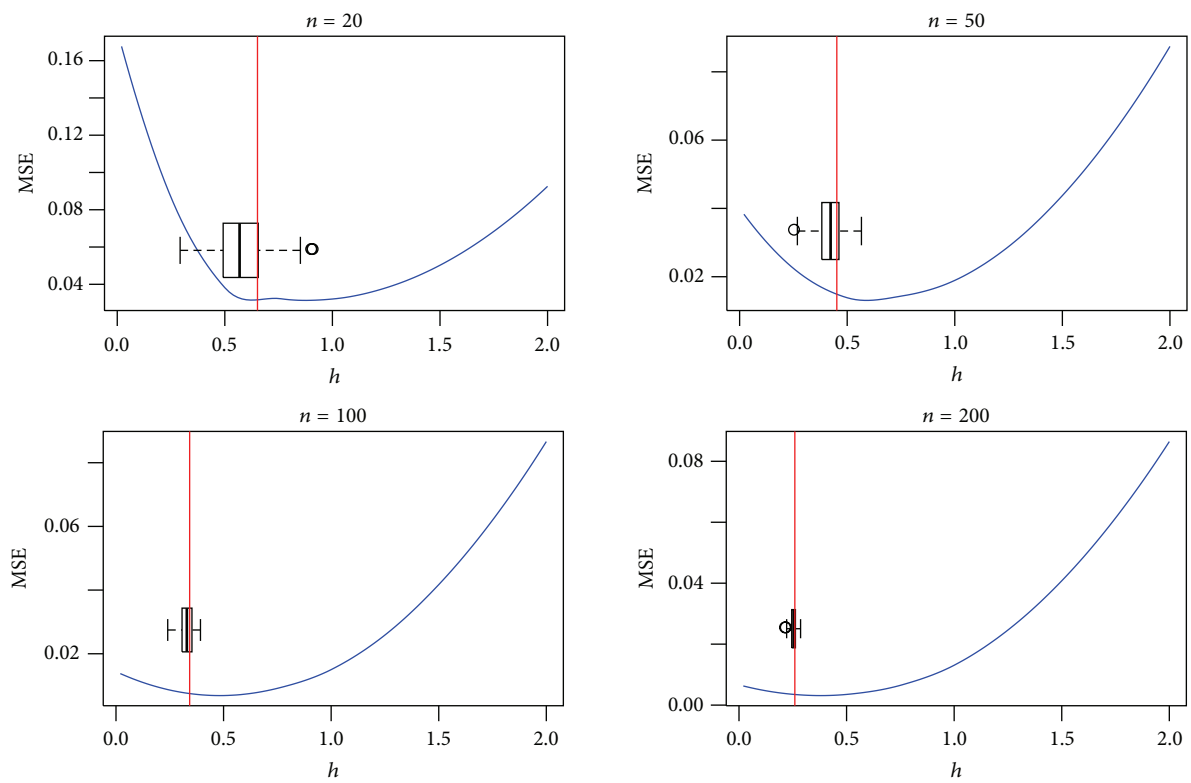


(b) $\mu = 0.5$

Figure 2: Continued.



(c) $\mu = 1$



(d) $\mu = 5$

FIGURE 2: The bandwidth for $\widehat{U}_2$ when underlying distribution is Normal with mean $\mu = 0$ (a), $\mu = 0.5$ (b), $\mu = 1$ (c), and $\mu = 5$ (d) and standard deviation $\sqrt{2}$.

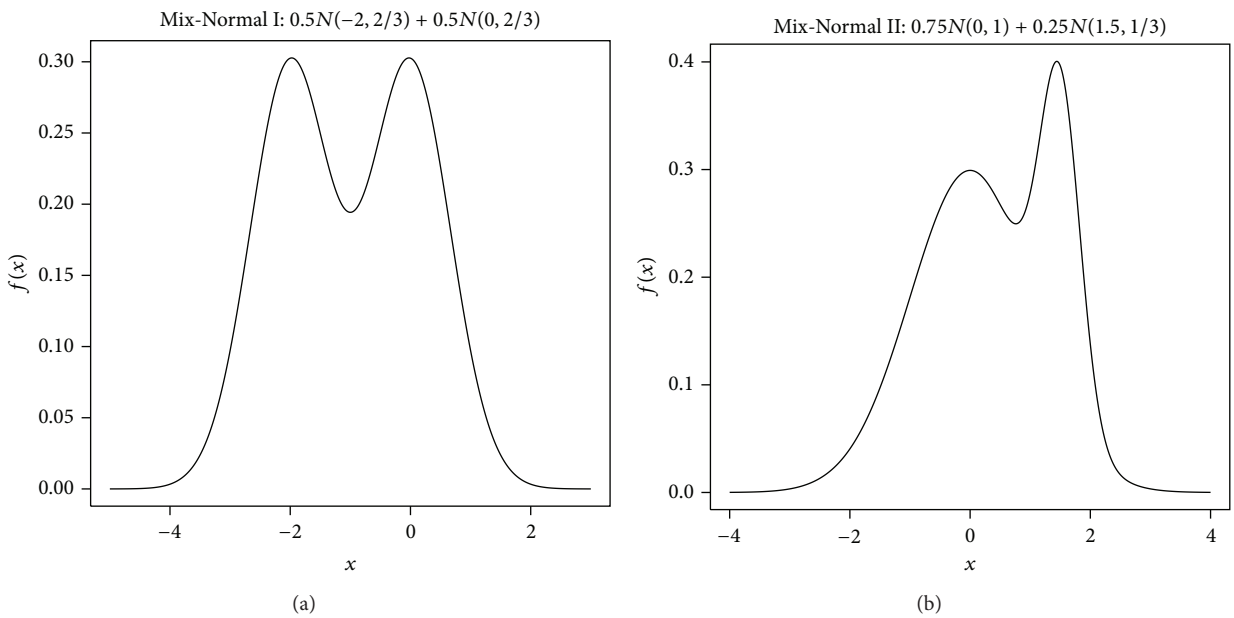


Figure 3: Density curve of (a) Mix-Normal I: $(1/2)N(-2, 2/3) + (1/2)N(0, 2/3)$; (b) Mix-Normal II: $(3/4)N(0, 1) + (1/2)N(1.5, 1/3)$.

(left-skewed) bimodal distribution. The motivation behind the choice of distributions is to see whether the optimal bandwidth is sensitive to the skewness, extreme outliers, and complex shape of the distributions in contrast to Normal.

Figures 4 and 5 compare five candidate pilot bandwidth selection methods in terms of the boxplots of MSE of $\widehat{U}_1$ and $\widehat{U}_2$ when sample size is 100. The five candidates for pilot bandwidth selection considered in this paper are (1) Rule-of-Thumb bandwidth for U_1 and U_2 proposed in Section 2.2 ("1-ROT(s)" means $\widehat{g}_{\mathrm{ROT};U_1}$; "1-ROT(L)" means $\widehat{g}_{\mathrm{ROT};U_2}$); (2) Rule-of-Thumb bandwidth for kernel density estimation (KDE) proposed in Scott [26] ("2-ROT(d)" means $\widehat{g}_{\mathrm{ROT;density}} = 1.06 * \widehat{\sigma} n^{-1/5}$, where $\widehat{\sigma}$ is the minimum of standard deviation and interquantile range); (3) least-square cross-validation bandwidth for density estimation proposed in Bowman [8] ("3-UCV(d)" means $\widehat{g}_{\mathrm{LSCV;density}}$ for density estimation); (4) biased cross-validation for density estimation proposed in Scott and Terrell [14] ("4-BCV(d)" means $\widehat{g}_{\mathrm{BCV;density}}$ for density estimation); (5) direct plug-in bandwidth for density estimation reported by Sheather [27] ("5-DPI(d)" means $\widehat{g}_{\mathrm{DPI;density}}$ for density estimation). As shown in Figures 4 and 5 that $\widehat{g}_{\mathrm{BCV;density}}$ is the *worst* candidate for pilot bandwidth in the density functionals estimation $\int \gamma(x) f(x) dx$. The pilot bandwidth choice $\widehat{g}_{\mathrm{ROT;density}}$ gives the lowest MSE($\widehat{U}_1$) for Cauchy and Normal samples, while $\widehat{g}_{\mathrm{LSCV;density}}$ results in the lowest MSE($\widehat{U}_1$) for Generalized Pareto, Mix-Normal I, and Mix-Normal II samples. Similar conclusions can be found in the pilot bandwidth choice for estimation of U_2; however, $\widehat{g}_{\mathrm{ROT};U_2}$ leads to slightly smaller MSE($\widehat{U}_2$) than $\widehat{g}_{\mathrm{LSCV;density}}$ for Mix-Normal I and Mix-Normal II samples. *Simulation Study 1* illustrates that the Normal reference bandwidth (including $\widehat{h}_{\mathrm{NS};U_2}$ and $\widehat{h}_{\mathrm{ROT};U_2}$) is not a reliable estimate when the location parameter of underlying distribution is close to zero. Therefore, $\widehat{g}_{\mathrm{LSCV;density}}$, rather than $\widehat{h}_{\mathrm{ROT};U_2}$, is recommended to serve as pilot bandwidth in estimation of $\widehat{h}_{\mathrm{DPI};U_2}$.

3.3. Simulation Study 3. This section aims to evaluate the performance of our proposed bandwidth $\widehat{h}_{\mathrm{ROT};U_1}$ (or $\widehat{h}_{\mathrm{ROT};U_2}$) in Section 2.2 and $\widehat{h}_{\mathrm{DPI};U_1}$ (or $\widehat{h}_{\mathrm{DPI};U_2}$) in Section 2.3 and compare with the classical bandwidth selection methods for density estimation, that is, $\widehat{h}_{\mathrm{ROT;density}}$, $\widehat{h}_{\mathrm{LSCV;density}}$, and $\widehat{h}_{\mathrm{DPI;density}}$ in estimation of U_1 (or U_2). *Simulation Study 2* recommends $\widehat{g}_{\mathrm{ROT;density}}$ and $\widehat{g}_{\mathrm{LSCV;density}}$ to be pilot bandwidth to estimate direct plug-in bandwidth for estimation of U_1 and U_2 among the 5 candidate pilot bandwidth methods and 5 different underlying distributions. Therefore, $\widehat{h}_{\mathrm{DPI};U_1}$ (or $\widehat{h}_{\mathrm{DPI};U_2}$) with these two pilot bandwidth selection methods, $\widehat{g}_{\mathrm{ROT;density}}$ and $\widehat{g}_{\mathrm{LSCV;density}}$, is considered separately in this subsection. Direct plug-in bandwidth for estimation of U_1 with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$ is denoted by "$\widehat{h}_{\mathrm{DPI^a(s)}}$" in Table 1 and "2a-DPI(s)" in Figure 1, and with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ is denoted by "$\widehat{h}_{\mathrm{DPI^b(s)}}$" in Table 1 and "2b-DPI(s)" in Figure 1.

The summary statistics (mean, 1st quantile, median, and 3rd quantile) for the three proposed bandwidths and 3 classical bandwidths are provided in Tables 1 and 2 to estimate U_1 and U_2, respectively. Samples of size 50 and 100 from five different underlying distribution (3 unimodal distributions and 2 bimodal distributions) as in *Simulation*

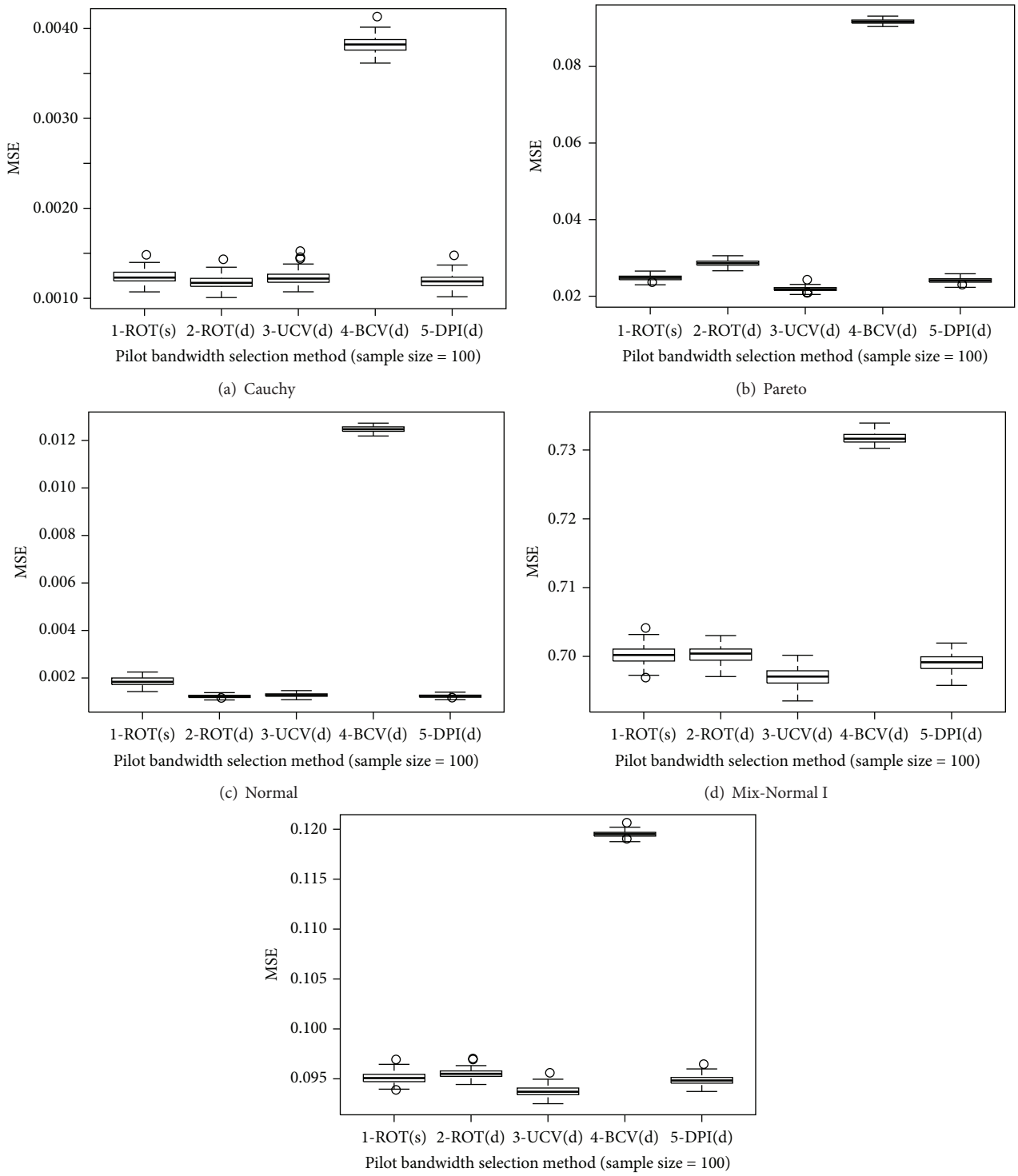


Figure 4: MSE of $\widehat{U}_1$ in terms of the choice of pilot bandwidth g for $\widehat{h}_{\mathrm{DPI};U_1}$; "1-ROT(s)" means $\widehat{g}_{\mathrm{ROT};U_1}$; "2-ROT(d)" means $\widehat{g}_{\mathrm{ROT;density}}$; "3-UCV(d)" means $\widehat{g}_{\mathrm{LSCV;density}}$; "4-BCV(d)" means $\widehat{g}_{\mathrm{BCV;density}}$; "5-DPI(d)" means $\widehat{g}_{\mathrm{DPI;density}}$.

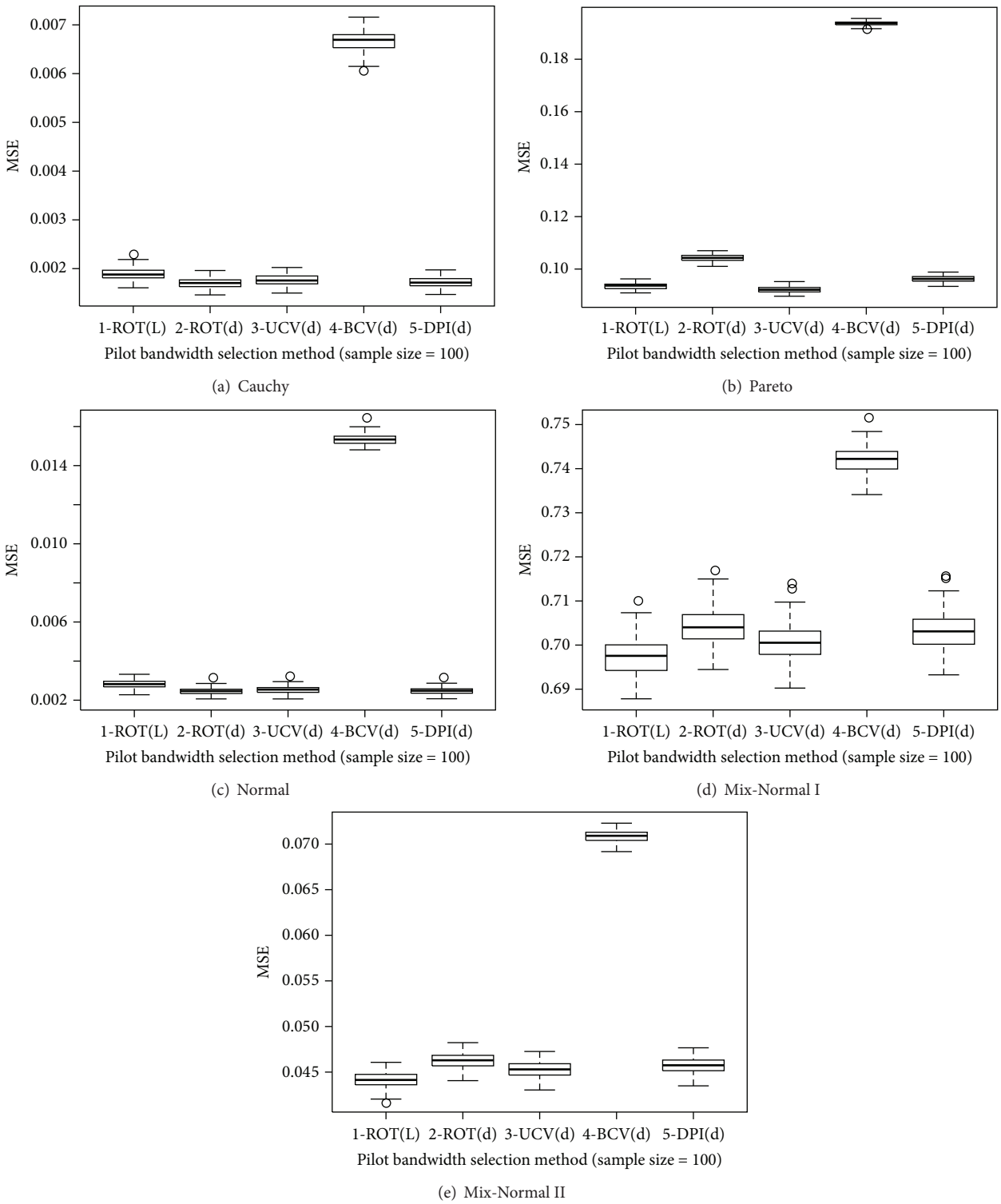


FIGURE 5: MSE of $\widehat{U}_2$ in terms of the choice of pilot bandwidth g for $\widehat{h}_{\mathrm{DPI};U_2}$; "1-ROT(L)" means $\widehat{g}_{\mathrm{ROT};U_2}$; "2-ROT(d)" means $\widehat{g}_{\mathrm{ROT;density}}$; "3-UCV(d)" means $\widehat{g}_{\mathrm{LSCV;density}}$; "4-BCV(d)" means $\widehat{g}_{\mathrm{BCV;density}}$; "5-DPI(d)" means $\widehat{g}_{\mathrm{DPI;density}}$.

Study 2 are considered in both simulation studies of U_1 and U_2 estimation.

In general, the optimal bandwidth for kernel density functionals estimation (estimation of U_1 and U_2 in this paper) is smaller than the one for kernel density estimation under same sample size and underlying distribution as shown in Tables 1 and 2, except for the least square cross-validation bandwidth $\widehat{h}_{\mathrm{LSCV;density}}$ for density estimation on Generalized

Table 1: Optimal bandwidth selection in estimation of U_1 with comparison to classical bandwidth selection methods.

	Bandwidth selection for estimation of U_1											
	Proposed method						Classical bandwidth for KDE					
	$\widehat{h}_{\mathrm{ROT}(s)}$ [1]		$\widehat{h}_{\mathrm{DPI}^a(s)}$ [2]		$\widehat{h}_{\mathrm{DPI}^b(s)}$ [3]		$\widehat{h}_{\mathrm{ROT}(d)}$ [4]		$\widehat{h}_{\mathrm{UCV}(d)}$ [5]		$\widehat{h}_{\mathrm{DPI}(d)}$ [6]	
	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100
Normal												
Mean	.197	.154	.236	.171	.221	.163	.302	.270	.359	.298	.300	.268
1st quantile	.182	.144	.205	.156	.200	.153	.278	.254	.307	.258	.265	.243
Median	.197	.154	.233	.170	.227	.167	.301	.270	.376	.317	.300	.268
3rd quantile	.214	.163	.265	.185	.254	.183	.327	.287	.431	.355	.336	.293
Cauchy												
Mean	.325	.240	.320	.222	.290	.205	.497	.422	.440	.331	.419	.338
1st quantile	.273	.214	.264	.198	.230	.174	.417	.377	.330	.265	.348	.310
Median	.316	.236	.312	.219	.295	.207	.484	.415	.455	.353	.416	.342
3rd quantile	.365	.265	.368	.245	.355	.238	.557	.465	.554	.415	.478	.387
Pareto												
Mean	.442	.332	.263	.178	.145	.098	.676	.583	.207	.143	.374	.284
1st quantile	.329	.267	.210	.151	.103	.078	.502	.469	.140	.106	.298	.241
Median	.422	.320	.251	.174	.137	.091	.646	.563	.200	.131	.359	.278
3rd quantile	.519	.386	.300	.202	.174	.109	.794	.678	.260	.173	.428	.327
Mix-Normal I												
Mean	.378	.289	.389	.279	.349	.253	.578	.507	.522	.413	.495	.417
1st Quantile	.360	.279	.342	.252	.282	.211	.551	.490	.389	.320	.438	.376
Median	.377	.288	.382	.273	.362	.259	.577	.506	.510	.403	.489	.407
3rd Quantile	.397	.298	.431	.309	.428	.304	.607	.524	.658	.510	.544	.460
Mix-Normal II												
Mean	.342	.263	.364	.255	.322	.227	.522	.461	.492	.372	.467	.389
1st Quantile	.321	.251	.318	.229	.254	.188	.490	.441	.339	.279	.413	.347
Median	.342	.262	.361	.254	.345	.231	.523	.461	.509	.355	.464	.387
3rd Quantile	.364	.274	.403	.278	.397	.275	.556	.481	.642	.461	.519	.425

[1]ROT bandwidth for estimation of U_1 given by (12); [2]DPI bandwidth for estimation of U_1 with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$; [3]DPI bandwidth for estimation of U_1 with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$; [4]ROT bandwidth for density estimation proposed in Silverman [7]; [5]LSCV bandwidth for density estimation proposed in Bowman [8]; [6]DPI bandwidth for density estimation proposed in Sheather and Jones [9].

Pareto samples. In another word, kernel density functionals estimation requires less smoothness in the estimation, which exaggerates some characteristics of the sample. For instance, the location and scale estimation will be more sensitive to the outliers than density estimation.

To evaluate the performance of our proposed bandwidth selection methods in contrast to classical bandwidth selection methods for density estimation in estimation of U_1 and U_2, the MSE of $\widehat{U}_1$ (and $\widehat{U}_2$) are computed and compared. Figures 6 and 7 demonstrate the boxplot of MSE of $\widehat{U}_1$ and $\widehat{U}_2$, respectively, in terms of 6 bandwidths shown in Tables 1 and 2 under five different distributions: Normal, Cauchy, Generalized Pareto, Mix-Normal I, and Mix-Normal II. Both figures illustrate that (a) MSE of $\widehat{U}_1$ and $\widehat{U}_2$ decreases as sample size increases from 50 to 100; (b) MSE of $\widehat{U}_1$ and $\widehat{U}_2$ is larger for samples from asymmetric distribution rather than symmetric distribution, from bimodal distribution rather than unimodal distribution. Figure 6 infers that (i) Normal scale bandwidths for both estimation of U_1 ($\widehat{h}_{\mathrm{ROT};U_1}$) and density estimation ($\widehat{h}_{\mathrm{ROT;density}}$) lead to smaller MSE($\widehat{U}_1$) relative to other 4 types of bandwidth selection methods for Normal samples of size 50. When sample size goes up to 100, $\widehat{h}_{\mathrm{ROT};U_1}$ outperforms $\widehat{h}_{\mathrm{ROT;density}}$ with a smaller MSE; (ii) for Cauchy samples with location 1 and scale 2/3, $\widehat{h}_{\mathrm{ROT;density}}$ becomes the worst choice in kernel density functionals estimation of U_1, especially for relative large sample size. Our proposed bandwidth $\widehat{h}_{\mathrm{ROT};U_1}$ results in the smallest MSE in this case; (iii) for Generalized Pareto samples with shape 1, location 1, and scale 2/3, both $\widehat{h}_{\mathrm{ROT};U_1}$ and $\widehat{h}_{\mathrm{ROT;density}}$ perform very poorly. However, $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ gives the smallest MSE for Pareto samples, which can partly be explained by the fact that $\widehat{g}_{\mathrm{LSCV;density}}$ gives the second smallest MSE; (iv) for bimodal distributed samples (including Mix-Normal I and Mix-Normal II), the three proposed bandwidth selection methods for estimation of U_1 completely dominate the three

Table 2: Optimal bandwidth selection in estimation of U_2 with comparison to classical bandwidth selection methods.

	Bandwidth selection for estimation of U_2											
	Proposed method						Classical bandwidth for KDE					
	$\widehat{h}_{\mathrm{ROT}(l)}$ [1]		$\widehat{h}_{\mathrm{DPI}^a(l)}$ [2]		$\widehat{h}_{\mathrm{DPI}^b(l)}$ [3]		$\widehat{h}_{\mathrm{ROT}(d)}$ [4]		$\widehat{h}_{\mathrm{UCV}(d)}$ [5]		$\widehat{h}_{\mathrm{DPI}(d)}$ [6]	
	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100	$n = 50$	100
Normal												
Mean	.086	.067	.236	.173	.221	.163	.303	.271	.358	.298	.300	.270
1st quantile	.079	.063	.203	.156	.198	.150	.278	.256	.296	.256	.263	.244
Median	.086	.067	.236	.173	.229	.170	.304	.272	.380	.320	.302	.272
3rd quantile	.094	.071	.267	.190	.254	.185	.329	.287	.432	.361	.334	.296
Cauchy												
Mean	.137	.104	.312	.221	.283	.200	.488	.422	.428	.320	.410	.337
1st quantile	.114	.092	.258	.194	.223	.164	.406	.373	.317	.238	.342	.298
Median	.136	.103	.304	.217	.285	.201	.483	.416	.431	.335	.405	.339
3rd quantile	.157	.115	.359	.244	.347	.236	.554	.468	.548	.414	.475	.387
Pareto												
Mean	.191	.147	.276	.188	.145	.099	.673	.595	.201	.143	.372	.290
1st quantile	.142	.122	.216	.163	.107	.076	.501	.495	.140	.106	.296	.254
Median	.180	.142	.263	.185	.138	.093	.635	.574	.193	.128	.357	.285
3rd quantile	.219	.166	.318	.208	.173	.114	.775	.673	.245	.174	.430	.322
Mix-Normal I												
Mean	.146	.112	.566	.369	.508	.342	.578	.507	.551	.417	.505	.420
1st quantile	.140	.109	.398	.289	.352	.254	.549	.490	.390	.332	.440	.378
Median	.147	.112	.481	.334	.454	.316	.577	.507	.549	.407	.500	.414
3rd quantile	.154	.116	.634	.391	.593	.382	.608	.523	.719	.500	.563	.458
Mix-Normal II												
Mean	.103	.080	.442	.272	.442	.247	.521	.461	.489	.359	.465	.387
1st quantile	.094	.076	.310	.220	.258	.173	.491	.443	.341	.257	.412	.348
Median	.106	.081	.363	.246	.350	.226	.522	.458	.490	.353	.461	.382
3rd quantile	.114	.087	.468	.282	.470	.278	.555	.481	.637	.456	.514	.419

[1]ROT bandwidth for estimation of U_2 given by (16); [2]DPI bandwidth for estimation of U_2 with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$; [3]DPI bandwidth for estimation of U_2 with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$; [4]ROT bandwidth for density estimation proposed in Silverman [7]; [5]LSCV bandwidth for density estimation proposed in Bowman [8]; [6]DPI bandwidth for density estimation proposed in Sheather and Jones [9].

classical density estimation bandwidth selection methods. Among the three proposed bandwidth, $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ gives minimum MSE($\widehat{U}_1$).

Figure 7 compares the performance of the three proposed practical bandwidth selection methods to three classical density estimation bandwidths in terms of kernel density functionals estimation of U_2. It is shown that $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ outperforms other 5 bandwidth methods in kernel density functionals estimation of U_2 for Normal and Cauchy samples. Similar to estimation of U_1, direct plug-in bandwidth designed for estimation of U_2, that is, $\widehat{h}_{\mathrm{DPI};U_2}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$, beats other candidates for Generalized Pareto samples. The optimal bandwidth selection (with three practical estimates $\widehat{h}_{\mathrm{ROT};U_2}$, $\widehat{h}_{\mathrm{DPI}^a(s)}$, and $\widehat{h}_{\mathrm{DPI}^b(s)}$) for estimation of U_2 proposed in this paper performs significantly better than the bandwidth selection for density estimation in density functionals estimation for Mix-Normal distributions. Among the three proposed bandwidth selection methods for estimation of U_2, $\widehat{h}_{\mathrm{ROT};U_2}$ works better than the other for a mixture of 2 Normal distributions as indicated in Figure 7.

4. Discussion

The optimal bandwidth, along with three practical bandwidth selection methods for kernel density functionals estimation of format $\int \gamma(x) f^2(x)dx$, is discussed in this paper. Necessity and urgency of this study are due to the fact that $\int f^2(x)dx$ and $\int x f^2(x)dx$ are two core portions for scale and location, respectively. Chen [3] shed a light on a novel field of nonparametric analytic method for experimental design relying on kernel density functionals estimation of $\int f^2(x)dx$ and $\int x f^2(x)dx$. Simulation studies in Chen [3] found that kernel-based equality of scale and location tests built on estimation of $\int \gamma(x) f^2(x)dx$ outperform traditional Levene's test of variance and ANOVA test, respectively, in particular

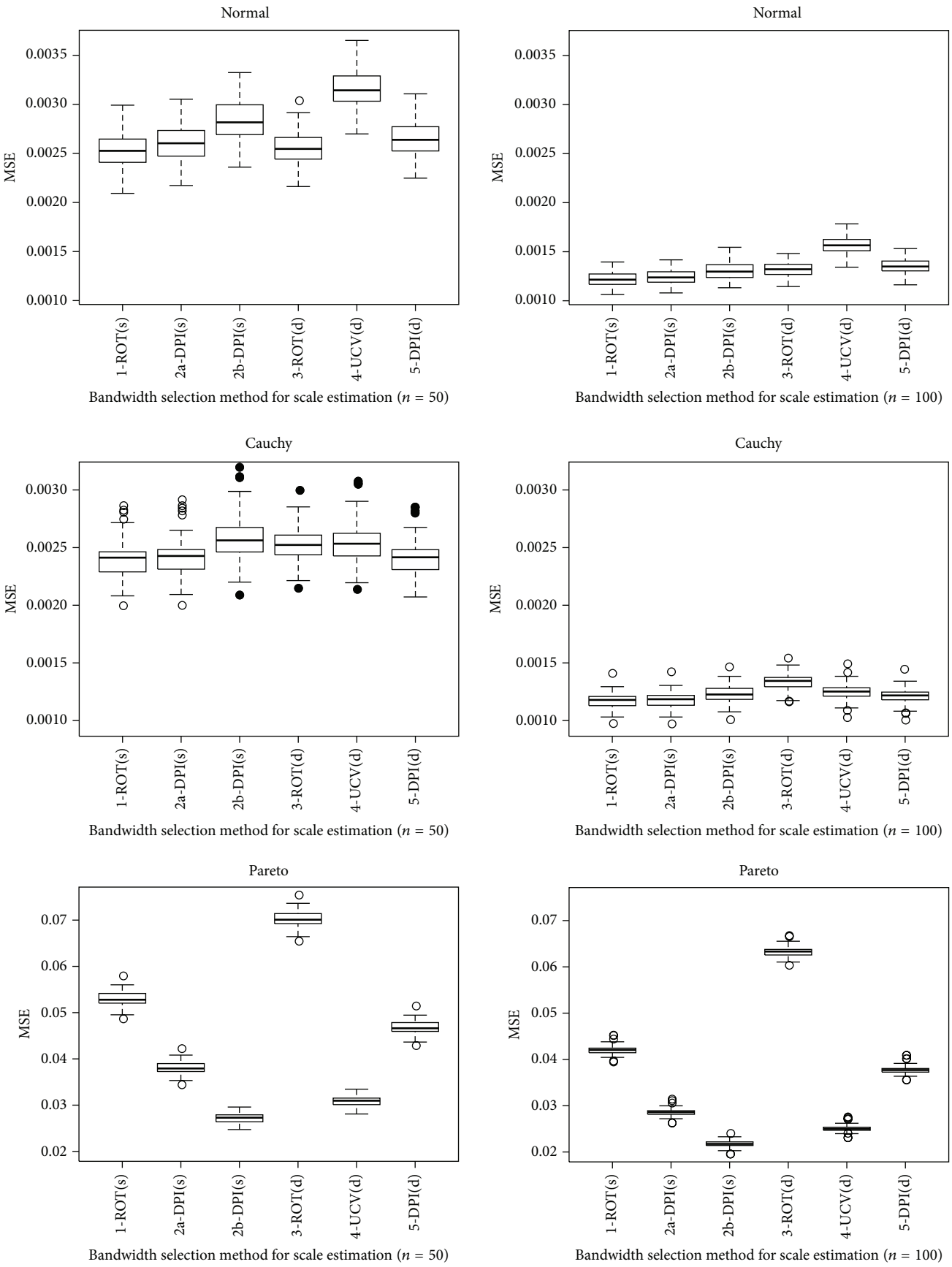


Figure 6: Continued.

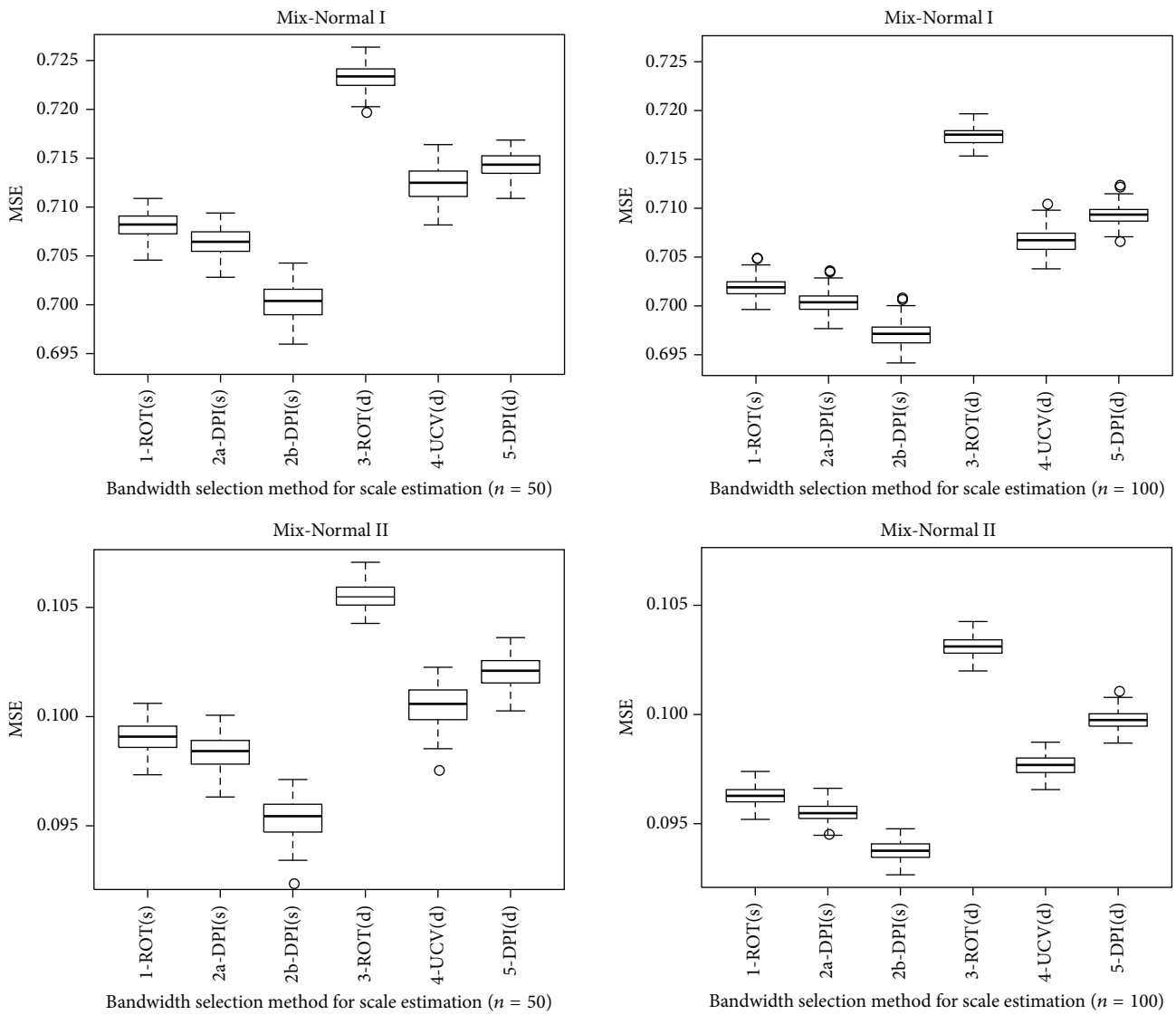


FIGURE 6: The MSE of $\widehat{U}_1$ in terms of bandwidth selection h; "1-ROT(s)" means $\widehat{h}_{\mathrm{ROT};U_1}$; "2a-DPI(s)" means $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$; "2b-DPI(s)" means $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$; "3-ROT(d)" means $\widehat{h}_{\mathrm{ROT;density}}$; "4-UCV(d)" means $\widehat{h}_{\mathrm{LSCV;density}}$; "5-DPI(d)" means $\widehat{h}_{\mathrm{DPI;density}}$.

to fat-tailed distribution, such as Cauchy. Our proposed choice of bandwidth selection methods can be directly applied to Chen's kernel-based equality of location tests, a nonparametric analog of ANOVA test, namely, "kernel-based ANOVA test." As we discussed in Section 1, Chen [3] uses the least-square cross-validation bandwidth which is designed for density estimation. However, the test statistics of the scale and location test are constructed on the kernel functionals estimation of $\int \gamma(x) f^2(x)dx$. Our proposed bandwidth will provide the "kernel-based ANOVA" test a better estimate of kernel functionals estimation of $\int \gamma(x) f^2(x)dx$, which in turn may improve the performance of the "kernel-based ANOVA" test. Kernel-based ANOVA, like other group comparison methods, has a broad application in various fields, such as biomedical sciences, education, and psychology, to compare the differences among 2 or more groups. A real-life example in education by Mimoto and Zitikis [28] is to compare the differences of quantitative abilities between science and nonscience majored students.

To broaden the application of this kernel-based nonparametric experimental design and accelerate the methodology development, optimal bandwidth for kernel density functionals estimation is in need to increase the accuracy of scale or location estimation and its relevant hypothesis testing and confidence intervals. $\int \gamma(x) f^2(x)dx$ is a function of density $f(x)$ and hence the classical bandwidth selection for kernel density estimation, such as ROT, UCV, BCV, and DPI, can serve as a choice of bandwidth h in its kernel density functionals estimation described in (4). However, our simulation study shows that the proposed

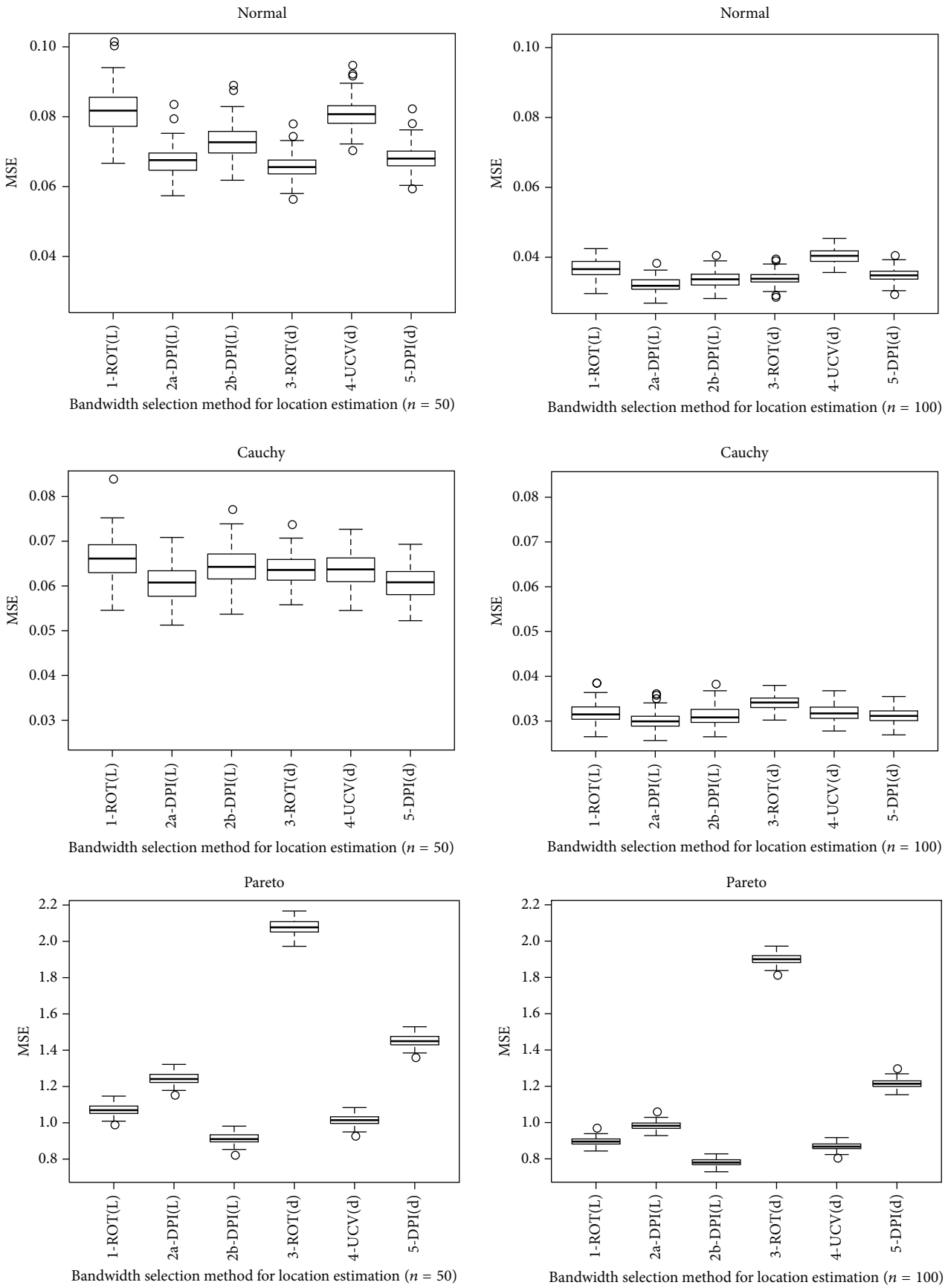


Figure 7: Continued.

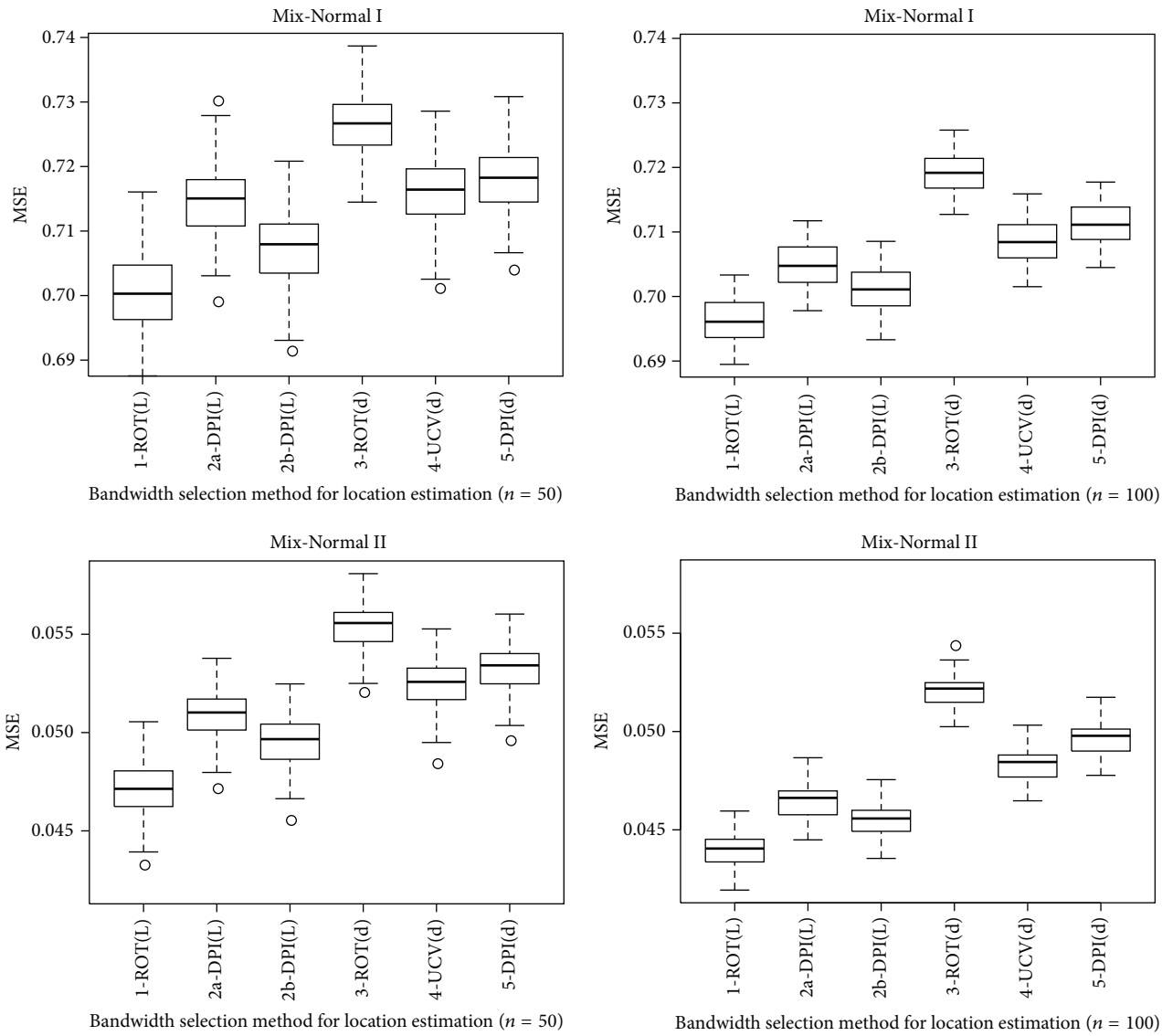


FIGURE 7: The MSE of $\widehat{U}_2$ in terms of bandwidth selection h; "1-ROT(L)" means $\widehat{h}_{\mathrm{ROT};U_2}$; "2a-DPI(L)" means $\widehat{h}_{\mathrm{DPI};U_2}$ with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$; "2b-DPI(L)" means $\widehat{h}_{\mathrm{DPI};U_2}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$; "3-ROT(d)" means $\widehat{h}_{\mathrm{ROT;density}}$; "4-UCV(d)" means $\widehat{h}_{\mathrm{LSCV;density}}$; "5-DPI(d)" means $\widehat{h}_{\mathrm{DPI;density}}$.

optimal bandwidth for $\int \gamma(x) f^2(x)dx$ is superior to the classical bandwidth selection for KDE in estimations U_1 and U_2, two determinant integrations to scale and location parameters. It is also suggested by the simulation study that, in scale estimation, the proposed Rule-of-Thumb bandwidth $\widehat{h}_{\mathrm{ROT};U_1}$ is highly recommended for samples from symmetric unimodal distribution, whereas direct plug-in bandwidth $\widehat{h}_{\mathrm{DPI};U_1}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ is recommended for samples from asymmetric unimodal or bimodal distribution. In location estimation, direct plug-in bandwidth $\widehat{h}_{\mathrm{DPI};U_2}$ with pilot bandwidth $\widehat{g}_{\mathrm{LSCV;density}}$ is still the best choice for asymmetric unimodal distribution, such as Generalized Pareto, whereas direct plug-in bandwidth $\widehat{h}_{\mathrm{DPI};U_2}$ with pilot bandwidth $\widehat{g}_{\mathrm{ROT;density}}$ is recommended for samples from symmetric and unimodal distribution, such as Normal and Cauchy. The proposed Rule-of-Thumb bandwidth $\widehat{h}_{\mathrm{ROT};U_2}$ is recommended for samples from bimodal distribution, such as a mixture of two Normal distributions in location estimation or location related testing.

5. Conclusion

Kernel smoothing method has been actively studied in density estimation and local kernel regression in the past decade. It is a modern data analytic method with its expertise in capturing the local humps and valley in the distributions or regression functions. Few literature applied kernel smoothing techniques in analysis of experimental design,

even the one-way ANOVA model. Ahmad [1] proposed a purely nonparametric method to estimate the location and scale parameter of any unknown distribution using kernel methods. Chen [3] presented a nonparametric nonrank based version of ANOVA using the kernel-based estimate of location (namely, "ANDFE" test) and opened a door to kernel-based nonparametric techniques for experimental design analysis. Chen [3] revealed that smoothing parameter (i.e., bandwidth) substantially affects the size and power of the ANDFE test.

This paper derived an optimal bandwidth for a nonparametric kernel density functionals estimation of location and scale of unknown density $f(x)$ and thus for ANDFE test of equality of location and scale. Thereafter two practical bandwidth selection methods ROT and DPI for location and scale estimation are proposed. Compared to traditional bandwidth selection methods designed for kernel density estimation (e.g., Jones et al. [11]) or kernel regression (e.g., Fan and Gijbels [24]), our proposed bandwidth selection methods demonstrate more accuracy in the estimation of U_1 or U_2. Simulation studies showed that our proposed novel optimal bandwidth method designed for kernel density functionals estimation of scale and location parameters works better than classical bandwidth selection designed for kernel density estimation, in particular bimodal distributions.

Appendices

A. Proof for Theorem 1

$\widehat{U}$ is an U-statistics, and thus by applying the properties of U-statistics in Lee [29], $E(\widehat{U})$ and $\mathrm{Var}(\widehat{U})$ can be computed as follows:

$$\begin{aligned}
E\left(\widehat{U}\right) &= E\left[\frac{\gamma(X_1)+\gamma(X_2)}{2}K_h(X_1-X_2)\right] \\
&= \int \frac{\gamma(y)}{h}K\left(\frac{x-y}{h}\right)f(x)f(y)\,dx\,dy \\
&= \int \frac{\gamma(y)}{h}K(u)f(y+uh)f(y)\,h\,du\,dy \\
&= \int \gamma(y)K(u) \\
&\cdot\left[f(y)+f'(y)uh+\frac{f''(y)}{2}u^2h^2+o\left(h^2\right)\right] \\
&\cdot f(y)\,du\,dy = \int \gamma(y)f^2(y)\,dy \\
&+\frac{h^2}{2}\left[\int \gamma(y)f''(y)f(y)\,dy\right]\left[\int u^2K(u)\,du\right] \\
&+o\left(h^2\right) = U+\frac{1}{2}E\left[\gamma(y)f''(y)\right]\mu_2(K)h^2 \\
&+o\left(h^2\right),
\end{aligned} \tag{A.1}$$

where $\mu_2(K) = \int u^2K(u)du$

$$\begin{aligned}
\mathrm{Var}\left(\widehat{U}\right) &= 2n^{-3}(n-1)\,\mathrm{Var}\left(\frac{\gamma(X_1)+\gamma(X_2)}{2}\right. \\
&\left.\cdot K_h(X_1-X_2)\right)+4n^{-3}(n-1)(n-2) \\
&\cdot \mathrm{cov}\left(\frac{\gamma(X_1)+\gamma(X_2)}{2}\right. \\
&\left.\cdot K_h(X_1-X_2),\frac{\gamma(X_1)+\gamma(X_3)}{2}K_h(X_1-X_3)\right).
\end{aligned} \tag{A.2}$$

To compute the first term of the variance of $\widehat{U}$, $E[(((\gamma(X_1)+\gamma(X_2))/2)K_h(X_1-X_2))^2]$ is calculated as follows:

$$\begin{aligned}
E\left[\left(\frac{\gamma(X_1)+\gamma(X_2)}{2}K_h(X_1-X_2)\right)^2\right] &= \frac{1}{2} \\
&\cdot E\left[\gamma^2(X_2)K_h^2(X_1-X_2)\right]+\frac{1}{2}E\left[\gamma(X_1)\gamma(X_2)\right. \\
&\left.\cdot K_h^2(X_1-X_2)\right] = \frac{1}{2}\int \frac{\gamma^2(y)}{h^2}K^2(u)f(y+uh) \\
&\cdot f(y)\,h\,du\,dy+\frac{1}{2}\int \frac{\gamma(y+uh)\gamma(y)}{h^2}K^2(u) \\
&\cdot f(y+uh)f(y)\,h\,du\,dy = \frac{1}{h}\int \gamma^2(y)f^2(y)\,dy \\
&\cdot \int K^2(u)\,du+o\left(\frac{1}{h}\right) = \frac{R(K)}{h}\int \gamma^2(y) \\
&\cdot f^2(y)\,dy+o\left(\frac{1}{h}\right),
\end{aligned} \tag{A.3}$$

where $R(K) = \int K(u)^2du$. In order to compute the second term of $\mathrm{Var}(\widehat{U})$, that is, the covariance term, $E[(((\gamma(X_1)+\gamma(X_2))/2)K_h(X_1-X_2)(((\gamma(X_1)+\gamma(X_3))/2)K_h(X_1-X_3)]$ needs to be simplified in the following equations:

$$\begin{aligned}
E\left[\frac{\gamma(X_1)+\gamma(X_2)}{2}K_h(X_1-X_2)\frac{\gamma(X_1)+\gamma(X_3)}{2}\right. \\
\left.\cdot K_h(X_1-X_3)\right] = \frac{1}{4h^2}\int\left[\gamma^2(z)+\gamma(z)\gamma(x)\right. \\
\left.+\gamma(z)\gamma(y)+\gamma(x)\gamma(y)\right]K\left(\frac{z-x}{h}\right)K\left(\frac{z-y}{h}\right) \\
\cdot f(x)f(y)f(z)\,dx\,dy\,dz = \frac{1}{4h^2}\int\left[\gamma^2(z)\right.
\end{aligned}$$

$$+\gamma(z)\gamma(z+uh)+\gamma(z)\gamma(z+vh)$$
$$+\gamma(z+uh)\gamma(z+vh)\Big]K(u)K(v)f(z+uh)f(z$$
$$+vh)f(z)h^2du\,dv\,dz=\frac{1}{4}\int\left[4\gamma^2(z)+o(1)\right]$$
$$\cdot K(u)K(v)\left[f(z)+o(1)\right]\left[f(z)+o(1)\right]f(z)\,du\,dv\,dz$$
$$=\int\gamma^2(z)f^3(z)\,dz+o(1). \tag{A.4}$$

Thus the covariance term $\mathrm{cov}(((\gamma(X_1)+\gamma(X_2))/2)K_h(X_1-X_2),((\gamma(X_1)+\gamma(X_3))/2)K_h(X_1-X_3))$ is given by $\int\gamma^2(z)f^3(z)dz-U^2$. Therefore, (6) is proved.

B. Proof for Proposition 2

If f is Normal with mean 0 and variance σ^2, then $R(f)$ and $E[f''(y)]$ in (10) can be calculated in terms of scale parameter σ. Note that

$$\begin{aligned}R(f)&=\int f^2(y)\,dy=\int\frac{1}{2\pi\sigma^2}e^{-y^2/\sigma^2}dy\\&=\int\frac{1}{2\pi\sigma^2}e^{-z^2}\sigma\,dz=\frac{1}{2\sqrt{\pi}\sigma},\\E\left[f''(y)\right]&=\int f''(y)f(y)\,dy\\&=\int\left(\frac{y^2}{\sqrt{2\pi}\sigma^5}e^{-y^2/2\sigma^2}-\frac{1}{\sqrt{2\pi}\sigma^3}e^{-y^2/2\sigma^2}\right)\\&\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-y^2/2\sigma^2}dy=\int\frac{1}{2\pi\sigma^4}\left(\frac{y^2}{\sigma^2}-1\right)e^{-y^2/\sigma^2}dy\\&=\frac{1}{2\pi\sigma^3}\left(\frac{\sqrt{\pi}}{2}-\sqrt{\pi}\right)=-\frac{1}{4\sqrt{\pi}\sigma^3}.\end{aligned} \tag{B.1}$$

Thus

$$\begin{aligned}h_{\mathrm{AMSE};U_1;N(0,\sigma)}&=\left[\frac{2R(K)\left(1/2\sqrt{\pi}\sigma\right)}{\left(1/16\pi\sigma^6\right)\mu_2^2(K)}\right]^{1/5}n^{-2/5}\\&=\left[\frac{16\sqrt{\pi}R(K)}{\mu_2^2(K)}\right]^{1/5}\sigma n^{-2/5}.\end{aligned} \tag{B.2}$$

σ needs to be estimated since it is unknown. Let $\widehat{\sigma}$ be a consistent estimate of σ, then we have Normal scale bandwidth for U_1 followed by (11).

C. Proof for Proposition 3

If f is Normal with mean μ and variance σ^2, then $R(yf(y))$ and $E[yf''(y)]$ in (13) can be calculated in terms of scale parameter σ. Note that

$$\begin{aligned}R(yf(y))&=\int y^2f^2(y)\,dy=\int\frac{y^2}{2\pi\sigma^2}e^{-(y-\mu)^2/\sigma^2}dy\\&=\int\frac{(\mu+z\sigma)^2}{2\pi\sigma^2}e^{-z^2}\sigma\,dz=\frac{1}{2\pi\sigma}\Big(\mu^2\sqrt{\pi}+0+\sigma^2\\&\cdot\frac{\sqrt{\pi}}{2}\Big)=\frac{2\mu^2+\sigma^2}{4\sqrt{\pi}\sigma},\\E\left[yf''(y)\right]&=\int yf''(y)f(y)\,dy\\&=\int y\left(\frac{(y-\mu)^2}{\sqrt{2\pi}\sigma^5}e^{-(y-\mu)^2/2\sigma^2}\right.\\&\left.-\frac{1}{\sqrt{2\pi}\sigma^3}e^{-(y-\mu)^2/2\sigma^2}\right)\frac{1}{\sqrt{2\pi}\sigma}e^{-(y-\mu)^2/2\sigma^2}\\&=\int\frac{\mu+z\sigma}{2\pi\sigma^3}e^{-z^2}\left(z^2-1\right)dz=-\frac{\mu}{4\sqrt{\pi}\sigma^3}.\end{aligned} \tag{C.1}$$

Thus

$$\begin{aligned}&h_{\mathrm{AMSE};U_2;N(\mu,\sigma)}\\&=\left[\frac{2R(K)\left(\left(2\mu^2+\sigma^2\right)/4\sqrt{\pi}\sigma\right)}{\left(\mu^2/16\pi\sigma^6\right)\mu_2^2(K)}\right]^{1/5}n^{-2/5}\\&=\left[\left(2+\frac{\sigma^2}{\mu^2}\right)\frac{8\sqrt{\pi}R(K)}{\mu_2^2(K)}\right]^{1/5}\sigma n^{-2/5}.\end{aligned} \tag{C.2}$$

The Normal scale bandwidth for U_2 is followed by (14) by replacing μ and σ with its corresponding estimate.

D. Proof for Proposition 5

The direct plug-in bandwidth $\widehat{h}_{\mathrm{DPI};U}$ in (19) can be written as

$$\widehat{h}_{\mathrm{DPI};U}=\alpha(K)\left(\frac{\widehat{R}}{\widehat{\phi}^2}\right)^{1/5}n^{-2/5}, \tag{D.1}$$

where $\alpha(K)=[2R(K)/\mu_2^2(K)]^{1/5}$. An analysis of errors involved in the approximation of $h_{\mathrm{MSE};U}$ by $h_{\mathrm{AMSE};U}$ leads to

$$\begin{aligned}h_{\mathrm{MSE};U}&=\alpha(K)\left(\frac{R(\gamma(y)f(y))}{\phi^2}\right)^{1/5}n^{-2/5}\\&+O\left(n^{-4/5}\right).\end{aligned} \tag{D.2}$$

Then the relative error of $\widehat{h}_{\mathrm{DPI};U}$ is

$$\begin{aligned}\frac{\widehat{h}_{\mathrm{DPI};U}}{h_{\mathrm{MSE};U}} - 1 &= \left(\frac{R(\gamma(y) f(y))}{\phi^2}\right)^{-1/5} \\ &\cdot \left[\left(\frac{\widehat{R}}{\widehat{\phi}^2}\right)^{1/5} - \left(\frac{R(\gamma(y) f(y))}{\phi^2}\right)^{1/5}\right] \\ &+ O_P\left(n^{-2/5}\right).\end{aligned} \tag{D.3}$$

By Taylor's theorem,

$$\begin{aligned}\frac{\widehat{h}_{\mathrm{DPI};U}}{h_{\mathrm{MSE};U}} - 1 &= \left(\frac{R}{\phi^2}\right)^{-1/5} \left\{\left(\frac{R}{\phi^2}\right)^{1/5}\right. \\ &\left. + \frac{1}{5}\left(\frac{R}{\phi^2}\right)^{4/5}\left[\frac{\widehat{R}}{\widehat{\phi}^2} - \frac{R}{\phi^2}\right] - \left(\frac{R}{\phi^2}\right)^{1/5}\right\} \\ &+ O_P\left(n^{-2/5}\right) = \frac{1}{5}\left(\frac{R}{\phi^2}\right)^{3/5}\left[\frac{\widehat{R}}{\widehat{\phi}^2} - \frac{R}{\phi^2}\right] \\ &+ O_P\left(n^{-2/5}\right),\end{aligned} \tag{D.4}$$

where $R = R(\gamma(y)f(y))$. Equation (D.4) shows that the convergence rate of $\widehat{h}_{\mathrm{DPI};U}$ depends on the density functionals estimation error of $\widehat{R}/\widehat{\phi}^2$ as well as the approximation error of h_{AMSE} to h_{MSE}. $\widehat{R}$ and $\widehat{\phi}^2$ are U-statistics and thus can be shown easily to follow approximately Normal distribution. The ratio of $\widehat{R}$ and $\widehat{\phi}^2$ is also approximately Normal under certain regular conditions by Hayya et al. [30] with variance given as follows:

$$\begin{aligned}\mathrm{Var}\left(\frac{\widehat{R}}{\widehat{\phi}^2}\right) &= \frac{R^2 \mathrm{Var}\left(\widehat{\phi}^2\right)}{\phi^8} + \frac{\mathrm{Var}\left(\widehat{R}\right)}{\phi^4} \\ &- \frac{2\rho R\sqrt{\mathrm{Var}\left(\widehat{\phi}^2\right)\mathrm{Var}\left(\widehat{R}\right)}}{\phi^6}.\end{aligned} \tag{D.5}$$

If the Normal scale bandwidth selector proposed in Section 2.2 is used as pilot bandwidth ($g = O(n^{-2/5})$) for the estimations $\widehat{R}$ and $\widehat{\phi}^2$, then $\mathrm{Var}(\widehat{R}) = O(n^{-2}g^{-1}) = O(n^{-8/5})$ and $\mathrm{Var}(\widehat{\phi}^2) = O(n^{-4}g^{-10}) = O(1)$. If bandwidth for density estimation is used as pilot bandwidth ($g = O(n^{-1/5})$), then $\mathrm{Var}(\widehat{R}) = O(n^{-2}g^{-1}) = O(n^{-9/5})$ and $\mathrm{Var}(\widehat{\phi}^2) = O(n^{-4}g^{-10}) = O(n^{-2})$. In both cases, $O_P(n^{-2/5})$ term dominates the density functionals estimation error $\widehat{R}/\widehat{\phi}^2 - R/\phi^2$. Thus $n^{1/5}(\widehat{h}_{\mathrm{DPI};U}/h_{\mathrm{MSE};U} - 1) \to_D N(0, \sigma^2_{\mathrm{DPI};U})$.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] I. A. Ahmad, "Nonparametric estimation of the location and scale parameters based on density estimation," *Annals of the Institute of Statistical Mathematics*, vol. 34, no. 1, pp. 39–53, 1982.

[2] I. A. Ahmad and M. Amezziane, "Estimation of location and scale parameters based on kernel functional estimators," in *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P Hettmansperger*, pp. 1–14, World Scientific, 2011.

[3] S. Chen, *Nonparametric ANOVA using Kernel methods [Ph.D. dissertation]*, Department of Statistics, Oklahoma State University, 2013.

[4] J. C. Aubuchon and T. Hettmansperger, "A note on the estimation of the integral of $f^2(x)$," *Journal of Statistical Planning and Inference*, vol. 9, no. 3, pp. 321–331, 1984.

[5] E. L. Lehmann, "Nonparametric confidence intervals for a shift parameter," *The Annals of Mathematical Statistics*, vol. 34, pp. 1507–1512, 1963.

[6] R. Grübel, "Estimation of density functionals," *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 1, pp. 67–75, 1994.

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

[8] A. W. Bowman, "An alternative method of cross-validation for the smoothing of density estimates," *Biometrika*, vol. 71, no. 2, pp. 353–360, 1984.

[9] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society: Series B*, vol. 53, no. 3, pp. 683–690, 1991.

[10] A. W. Bowman, "A comparative study of some kernel-based nonparametric density estimators," *Journal of Statistical Computation and Simulation*, vol. 21, no. 3-4, pp. 313–327, 1985.

[11] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.

[12] C. R. Loader, "Bandwidth selection: classical or plug-in?" *The Annals of Statistics*, vol. 27, no. 2, pp. 415–438, 1999.

[13] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, UK, 1995.

[14] D. W. Scott and G. R. Terrell, "Biased and unbiased cross-validation in density estimation," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1131–1146, 1987.

[15] W. Härdle and J. S. Marron, "Optimal bandwidth selection in nonparametric regression function estimation," *The Annals of Statistics*, vol. 13, no. 4, pp. 1465–1481, 1985.

[16] E. A. Nadaraya, "On the integral mean square error of some nonparametric estimates for the density function," *Theory of Probability and Its Applications*, vol. 19, no. 1, pp. 133–141, 1974.

[17] M. Woodroofe, "On choosing a delta-sequence," *The Annals of Mathematical Statistics*, vol. 41, pp. 1665–1671, 1970.

[18] P. Hall, J. S. Marron, and B. U. Park, "Smoothed cross-validation," *Probability Theory and Related Fields*, vol. 92, no. 1, pp. 1–20, 1992.

[19] H.-G. Müller, "Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators," *Statistics & Decisions*, supplement 2, pp. 193–206, 1985.

[20] J. G. Staniswalis, "Local bandwidth selection for kernel estimates," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 284–288, 1989.

[21] P. Hall, S. J. Sheather, M. C. Jones, and J. S. Marron, "On optimal data-based bandwidth selection in kernel density estimation," *Biometrika*, vol. 78, no. 2, pp. 263–269, 1991.

[22] T. Gasser, A. Kneip, and W. Köhler, "A flexible and fast method for automatic smoothing," *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 643–652, 1991.

[23] D. Ruppert, S. J. Sheather, and M. P. Wand, "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1257–1270, 1995.

[24] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Application*, Chapman & Hall, London, UK, 1996.

[25] P. Janssen, J. S. Marron, N. Veraverbeke, and W. Sarle, "Scale measures for bandwidth selection," *Journal of Nonparametric Statistics*, vol. 5, no. 4, pp. 359–380, 1995.

[26] D. W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, John Wiley & Sons, New York, NY, USA, 1992.

[27] S. J. Sheather, "The performance of six popular bandwidth selection methods on some real datasets," *Computational Statistics*, vol. 7, pp. 225–250, 1992.

[28] N. Mimoto and R. Zitikis, "Czekanowski's index of overlap, its L_p-type extension, and bias reduction," *American Journal of Mathematical and Management Sciences*, vol. 29, no. 1-2, pp. 229–261, 2009.

[29] A. J. Lee, U-*Statistics: Theory and Practice*, vol. 110, Marcel Dekker, New York, NY, USA, 1990.

[30] J. Hayya, D. Armstrong, and N. Gressis, "A note on the ratio of two normally distributed variables," *Management Science*, vol. 21, no. 11, pp. 1338–1341, 1975.