Ann Inst Stat Math (2014) 66:865–895
DOI 10.1007/s10463-013-0428-7

Root *n* estimates of vectors of integrated density partial derivative functionals

Tiee-Jian Wu · Chih-Yuan Hsu · Huang-Yu Chen · Hui-Chun Yu

Received: 3 February 2012 / Revised: 30 April 2013 / Published online: 5 September 2013

Abstract Based on a random sample of size n from an unknown d-dimensional density f, the nonparametric estimations of a single integrated density partial derivative functional as well as a vector of such functionals are considered. These single and vector functionals are important in a number of contexts. The purpose of this paper is to derive the information bounds for such estimations and propose estimates that are asymptotically optimal. The proposed estimates are constructed in the frequency domain using the sample characteristic function. For every d and sufficiently smooth f, it is shown that the proposed estimates are asymptotically normal, attain the optimal $O_p(n^{-1/2})$ convergence rate and achieve the (conjectured) information bounds. In simulation studies the superior performances of the proposed estimates are clearly demonstrated.

Keywords Bandwidth selection · Characteristic function · Convergence rate · Cross-validation · Multivariate kernel estimate · Nonparametric information bound

1 Introduction

For every $d \geq 1$, let $\mathbf{X}$ be an $n \times d$ data matrix of random vectors $\mathbf{x} = (x_1, \ldots, x_d)$ (we use the row-vector convention throughout the paper) where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent observations drawn from an unknown d-dimensional density $f(\mathbf{x})$. Let us write

$$\psi_{\mathbf{r}} = \int_{\Re^d} f_{\mathbf{r}}(\mathbf{x}) f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathrm{E} f_{\mathbf{r}}(\mathbf{x}_1), \quad |\mathbf{r}| = 2m,\ m = 0, 1, 2, \ldots \tag{1}$$

Research supported by a grant from the National Science Council of Taiwan and by the National Center for Theoretical Sciences (South) of Taiwan.

T.-J. Wu (✉) · C.-Y. Hsu · H.-Y. Chen · H.-C. Yu
Department of Statistics, National Cheng-Kung University, Tainan 70101, Taiwan
e-mail: tjwu@stat.ncku.edu.tw

where $\mathbf{r} = (r_1, \ldots, r_d)$, $r_i \geq 0$ and $f_{\mathbf{r}}(\mathbf{x})$ denotes a partial derivative of f of order $|\mathbf{r}|$ with $|\mathbf{p}| = \sum_{j=1}^{d} p_j$ for any vector $\mathbf{p} = (p_1, \ldots, p_d)$ satisfying $p_j \geq 0$ for all j. If all the components of $\mathbf{r}$ are even, say, $\mathbf{r} = 2\mathbf{m}$ for some $\mathbf{m} = (m_1, \ldots, m_d)$, then using integration by parts, we can express $\psi_{\mathbf{r}} = \psi_{2\mathbf{m}}$ as

$$\psi_{2\mathbf{m}} = (-1)^{|\mathbf{m}|}\theta_{\mathbf{m}} \quad \text{with} \quad \theta_{\mathbf{m}} = \int_{\Re^d} \{f_{\mathbf{m}}(\mathbf{x})\}^2 \, \mathrm{d}\mathbf{x}, \quad |\mathbf{m}| = 0, 1, 2, \ldots. \tag{2}$$

Indeed, (2) holds under Condition ($\mathrm{C}_{2\mathbf{m}}$) where Condition ($\mathrm{C}_{\mathbf{k}}$) $= \{f_{\mathbf{k}} \in L^1(\Re^d)$ and $f_{\mathbf{s}} \in L^1(\Re^d)$ is absolutely continuous for all $\mathbf{s} = (s_1, \ldots, s_d)$ with $0 \leq |\mathbf{s}| \leq |\mathbf{k}| - 1\}$. Let $\otimes$ denote the Kronecker product. For $m = 0, 1, \ldots$, define the vector

$$\boldsymbol{\psi}_{2m} = \int_{\Re^d} \mathrm{D}^{\otimes 2m} f(\mathbf{x}) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \;\; \text{(say)} \; = (\psi_{\mathbf{r}_i})_{1 \leq i \leq d^{2m}}, \; |\mathbf{r}_i| = 2m \text{ for all } i \tag{3}$$

where $\mathbf{r}_i = (r_{i,1}, \ldots, r_{i,d})$ and $\mathrm{D}^{\otimes 2m} f = \mathrm{D} f \otimes \cdots \otimes \mathrm{D} f$ is the $2m$-th Kronecker power of the vector $\mathrm{D} f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_d)$. Thus, $\mathrm{D}^{\otimes 2m} f \in \Re^{d^{2m}}$ is a vector containing all the partial derivatives of order $2m$ and $\boldsymbol{\psi}_{2m} \in \Re^{d^{2m}}$ is one containing all the $\psi_{\mathbf{r}}$'s with $|\mathbf{r}| = 2m$ (including possible multiplicities arising from the fact that for smooth enough f the mixed partial derivatives may be equal).

Nonlinear functionals (1)–(2) and vector functionals (3) are of distinct interest from the viewpoint of actual applications. For example, the functionals $\theta_{\mathbf{m}}$, $1 \leq |\mathbf{m}| \leq 2$, and the vector $\boldsymbol{\psi}_4$ appear in the asymptotically (plug-in) optimal bandwidth for d-dimensional histograms, frequency polygons and density estimates (cf. Scott 1992; Wand and Jones 1994; Duong and Hazelton 2003) and in the rescaling factor for testing of multimodality based on kernel density estimates (cf. Fisher et al. 1994). Also, $\theta_{\mathbf{0}}$ (here and below, $\mathbf{c} = (c, c, \ldots, c, c)$ for any constant c) can be applied to projection pursuit because $\theta_{\mathbf{0}}$ appears in the d-dimensional Friedman–Tukey projection index (cf. Silverman 1986), and $\log\theta_{\mathbf{0}}$ is an upper bound for the d-dimensional negative Shannon entropy (we have $\mathrm{E}\{\log f(\mathbf{x}_1)\} \leq \log\{\mathrm{E} f(\mathbf{x}_1)\} = \log\theta_{\mathbf{0}}$ by Jensen's inequality). Furthermore, let f_j denote the jth marginal density, the functionals $\int_{\Re} f_j^2(x)\,\mathrm{d}x$, $1 \leq j \leq d$ appear in the asymptotic variance of the d-dimensional Wilcoxon-type rank test, in the asymptotic relative efficiency of such rank test relative to the Hotelling's T^2 test (cf. Puri and Sen 1971) and in the weed emergence index for measuring the spread of the probability distribution of the cumulative hydrothermal time at emergence (cf. Cao et al. 2011).

In the past, most work for estimating (2) has focused on the univariate case, while the multivariate case has been largely neglected. This may be due to the fact that it is technically more difficult to calibrate multivariate density partial derivatives (see the beginning of Sect. 2 for further descriptions of the challenge and difficulty). For $d = 1$, some work on estimating (2) includes Hall and Marron (1987, 1991a), Bickel and Ritov (1988), Aldershof (1991), Jones and Sheather (1991), Wu (1995), Cheng (1997), Laurent (1997), Martinez and Olivares (1999), Giné and Mason (2008) and Chacón and Tenreiro (2012), among others. Also, results on estimating (2) with f being supported in [0, 1] are given by Fan (1991), who dealt with a white noise model (see also Donoho and Nussbaum 1990), Goldstein and Messer (1992) and Efromovich and Low (1996a,b). Additional result on estimating (2) with contaminated sample is given by Delaigle and Gijbels (2002).

For $d \geq 2$, Wand (1992) (see also Wand and Jones 1994) proposed an estimate, denoted by $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$ herein, of (2) which is the multivariate version of the estimate of Jones and Sheather (1991). The $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$ utilizes the form (1) and is a kernel-based estimate of (1) with the kernel being a spherically symmetric pdf. Wand and Jones (1994) proved that the MSE (mean squared error) $\mathrm{E}\{\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}} - \theta_{\mathbf{m}}\}^2$ converges to zero quickly as $n \to \infty$ if the same non-adaptive (non-data-driven) bandwidth of the form $An^{-1/(2+d+2|\mathbf{m}|)}$ is used in every coordinate direction where A depends on the kernel and the unknown $\sum_{i=1}^{d} \theta_{\mathbf{m}+\mathbf{e}_i}$ with $\mathbf{e}_i$'s denoting the standard unit vectors in $\Re^d$. Although their result provides very significant insight into the theoretical issue of choosing bandwidth and kernel, their $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$ in general can not achieve the optimal MSE convergence rate $O(n^{-1})$. In addition, the $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$ is not immediately applicable in practice because the asymptotically optimal bandwidth depends on the unknown $\sum_{i=1}^{d} \theta_{\mathbf{m}+\mathbf{e}_i}$. Recently, Duong and Hazelton (2003) and Chacón and Duong (2010) (see also Duong and Hazelton 2005a,b, Chacón and Duong 2011 and Chacón et al. 2011) have shown that it can be beneficial to estimate the functionals (1) satisfying $|\mathbf{r}| = 2m$ for a fixed m all at once, instead of estimating each of them separately; and proposed to estimate the vector functionals (3) by multistage methods which are modifications and generalizations of the aforementioned methods of Wand (1992) and Wand and Jones (1994).

The purpose of this paper is fourfold. First, the nonparametric information bound for estimating a single functional $\psi_{\mathbf{r}}$ with even $|\mathbf{r}|$ (while $\theta_{\mathbf{m}}$ is a special case) is given (see Theorem 3). Secondly, we propose a nonparametric estimate of $\psi_{\mathbf{r}}$ which is asymptotically efficient if the non-adaptive bandwidth in each coordinate direction varies freely in a specific range (see Lemma 1). Thirdly, we propose an adaptive bandwidth selector which falls into the just-mentioned range in probability (see Lemma 2) and thus makes our estimate of $\psi_{\mathbf{r}}$ adaptive and immediately applicable in practice. Moreover, for all d, $\mathbf{r}$ and sufficiently smooth f, our adaptive estimate of $\psi_{\mathbf{r}}$ is shown to be asymptotically normal, attains the optimal $O_p(n^{-1/2})$ error convergence rate and achieves the best possible constant coefficient in this convergence (see Theorem 1). Fourthly, we propose an adaptive estimate of the vector functionals $\boldsymbol{\psi}_{2m}$, as a modification and generalization of our estimate of $\psi_{\mathbf{r}}$, and prove it is asymptotically normal, attains the optimal $O_p(n^{-1/2})$ error convergence rate and achieves the (conjectured) best possible constant covariance matrix in this convergence (see Theorem 2). Sections 2.1 and 2.3 give the details of the proposed estimates. Section 2.2 contains the main theoretical results. In Sect. 3, for $d = 2$ and 3, simulation studies are carried out and the superior performance of the proposed adaptive estimates is clearly demonstrated. Section 4 is devoted to proofs.

2 The proposed method

The proposed estimates are constructed in the Fourier domain and the estimate of $\psi_{\mathbf{r}}$ is mainly an extension to higher d of the univariate estimate of (1.2) of Wu (1995). We have chosen such extension because that univariate estimate achieves the optimal $O_p(n^{-1/2})$ error convergence rate, attains the information bound and performs superiorly in simulation studies. However, the extension is challenging and requires non-trivial work due to complications and difficulties caused by (i) the correlation between the bandwidths in different coordinate directions and (ii) the curse of dimensionality (as termed by Bellman 1961, which describes the rapid growth in the difficulty of problems as d increases, and the cost of an algorithm was observed to grow exponentially in d).

2.1 The proposed estimates

Throughout the paper, $\int$ is shorthand for $\int_{\Re^d}$. For any vectors $\mathbf{t} = (t_1, \ldots, t_d)$ and $\mathbf{p} = (p_1, \ldots, p_d)$, we denote $\mathbf{t}^{\mathbf{p}} = \prod_{j=1}^{d} t_j^{p_j}$ and use $\mathbf{t} \le \mathbf{p}$ to mean $t_j \le p_j$ for all j (similarly, $\mathbf{t} < \mathbf{p}$). Also, for ease of writing, we shall say a vector is large [small, resp.] if its Euclidean norm $\|\cdot\|$ is large [small, resp.]. Furthermore, we use $\phi_g(\mathbf{t}) = \int \exp(i\mathbf{x}\mathbf{t}')g(\mathbf{x})\,\mathrm{d}\mathbf{x}$ to denote the Fourier transform of any $g \in L^1(\Re^d) \cup L^2(\Re^d)$, $R(\mathbf{T})$ the d-dimensional rectangular region $[-T_1, T_1] \times \cdots \times [-T_d, T_d]$, and $R'(\mathbf{T}) = \Re^d \setminus R(\mathbf{T})$ where $\mathbf{T} = (T_1, \ldots, T_d) > \mathbf{0}$. For the rest, we set $|\mathbf{r}| = 2m$ for some fixed $m \ge 0$. For sufficiently smooth f, we can express $\psi_{\mathbf{r}}$ as

$$\psi_{\mathbf{r}} = (2\pi)^{-d}(-1)^m \int \mathbf{t}^{\mathbf{r}} |\phi_f(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t}, \quad |\mathbf{r}| = 2m. \tag{4}$$

Indeed, if f satisfies Condition ($\mathrm{C}_{\mathbf{r}}$) [as defined immediately below (2)], then $\phi_{f_{\mathbf{r}}}(\mathbf{t}) = (-1)^m \mathbf{t}^{\mathbf{r}} \phi_f(\mathbf{t})$ (cf. Hewitt and Stromberg 1969, pages 414–415) and $\phi_{f_{\mathbf{r}} * \tilde{f}}(\mathbf{t}) = \phi_{f_{\mathbf{r}}}(\mathbf{t})\phi_{\tilde{f}}(\mathbf{t}) = (-1)^m \mathbf{t}^{\mathbf{r}} |\phi_f(\mathbf{t})|^2$ where $*$ denotes convolution and $\phi_{\tilde{f}}(\mathbf{t}) = \phi_f(-\mathbf{t})$ with $\tilde{f}$ being defined by $\tilde{f}(\mathbf{x}) = f(-\mathbf{x})$ for all $\mathbf{x}$. Further, if $\mathbf{t}^{\mathbf{r}}|\phi_f(\mathbf{t})|^2 \in L^1(\Re^d)$ (as ensured by requiring $p_0 > m + (d/2)$ in Condition (A) below), then by Fourier inversion formula (see, e.g., Rektorys 1969, page 1136), we get $\psi_{\mathbf{r}} = \int f_{\mathbf{r}}(-\mathbf{x}) f(-\mathbf{x})\,\mathrm{d}\mathbf{x} = (f_{\mathbf{r}} * \tilde{f})(\mathbf{0}) = (2\pi)^{-d} \int \phi_{f_{\mathbf{r}} * \tilde{f}}(\mathbf{t})\,\mathrm{d}\mathbf{t}$, and hence (4). By utilizing the form (4), a kernel-based estimate of $\psi_{\mathbf{r}}$ is obtained if ϕ_f is replaced by $\phi_{\hat{f}}$ in (4) where $\hat{f}$ is the following multivariate product kernel estimate of f:

$$\hat{f}(\mathbf{x}) = (nh_1 \cdots h_d)^{-1} \sum_{i=1}^{n} \left\{ \prod_{j=1}^{d} K((x_j - x_{ij})/h_j) \right\}. \tag{5}$$

Here, the same symmetric univariate kernel K is used in each coordinate direction, but with a different bandwidth h_j for each direction. We remark that any d-dimensional spherically symmetric kernel can be used in (5). The product kernel is used here to avoid the technical difficulties involved in the analysis of unconstrained estimators. For cases where there is much to be gained by selecting a full bandwidth matrix (which contains $d(d+1)/2$ smoothing parameters), the readers are referred to Wand and Jones (1993), Wand and Jones (1994), Duong and Hazelton (2005a,b), Chacón (2009), Chacón and Duong (2010, 2011) and Chacón et al. (2011), among others.

Since the present problem is estimating $\psi_{\mathbf{r}}$, we need have no qualms over using higher-order kernels. We shall choose the "sync kernel" $K_\infty(x) = (\pi x)^{-1} \sin x$,

$-\infty < x < \infty$, a symmetric and infinite-order kernel. For advantages of using K_∞ see Davis (1975, 1977), Ibragimov and Khas'minskii (1982), Devroye (1988, 1992), Hall and Marron (1988) and Wu (1995). Note that $K_\infty \in L^2(\Re) \setminus L^1(\Re)$ and $\phi_{K_\infty}(\cdot) = I_{[-1,1]}(\cdot)$ with $I(\cdot)$ denoting the indicator function. Let

$$\tilde{\phi}(\mathbf{t}) = n^{-1} \sum_{j=1}^{n} \exp(i\mathbf{t}\mathbf{x}_j') \tag{6}$$

denote the sample characteristic function and $\hat{f}_{\mathbf{T}}$ the product sync-kernel estimate resulting from replacing K by K_∞ and h_j by T_j^{-1}, $1 \le j \le d$ in (5). Then $\phi_{\hat{f}_{\mathbf{T}}}(\mathbf{t}) = \tilde{\phi}(\mathbf{t}) \prod_{j=1}^{d} I_{[-T_j, T_j]}(t_j)$. Replacing ϕ_f in (4) by $\phi_{\hat{f}_{\mathbf{T}}}$ results in the family of estimates

$$\tilde{\psi}_{\mathbf{r}}(\mathbf{T}) = (2\pi)^{-d}(-1)^m \int_{R(\mathbf{T})} \mathbf{t}^{\mathbf{r}} |\tilde{\phi}(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t}, \quad |\mathbf{r}| = 2m \tag{7}$$

of $\psi_{\mathbf{r}}$ where the cutoff value $\mathbf{T}$ must satisfy $\min_{1\le j\le d} T_j \to \infty$ and $n^{-1}\mathbf{T}^{\mathbf{m}+\mathbf{1}} \to 0$ as $n \to \infty$ (see Remark 1 below). When $\mathbf{r} = 2\mathbf{m}$, the estimate (7) can be viewed as the sync-kernel, Fourier domain, multivariate version of the "diagonals-in" estimate of (2) proposed by Jones and Sheather (1991) (see Wu 1995 for details). The performance of (7) depends heavily on how well the cutoff value $\mathbf{T}$ can be selected. It was pointed out by Wu and Tsai (2004) (see also Chiu 1991) that $\tilde{\phi}(\mathbf{t})$ at large $\|\mathbf{t}\|$ is dominated by sample variation and does not contain much information about f. Therefore, (7) would have large variation if $\mathbf{T}$ is too large and large negative bias if $\mathbf{T}$ is too small.

For estimating the single functional $\psi_{\mathbf{r}}$ the proposed adaptive cutoff value $\hat{\mathbf{T}}_{\mathbf{r}} = (\hat{T}_{\mathbf{r},1}, \ldots, \hat{T}_{\mathbf{r},d})$ is the minimizer of

$$\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) = 2^{d+1}\mathbf{T}^{\mathbf{r}+\mathbf{1}} \left\{ (n+1) \prod_{i=1}^{d} (r_i + 1) \right\}^{-1} - \int_{R(\mathbf{T})} |\mathbf{t}^{\mathbf{r}}| |\tilde{\phi}(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t}, \quad \mathbf{T} > \mathbf{0} \tag{8}$$

(see Remark 1 below for an explanation) and the proposed adaptive estimate of $\psi_{\mathbf{r}}$ is

$$\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}}) = (2\pi)^{-d}(-1)^m \int_{R(\hat{\mathbf{T}}_{\mathbf{r}})} \mathbf{t}^{\mathbf{r}} |\tilde{\phi}(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t}. \tag{9}$$

Note that $\hat{\mathbf{T}}_{\mathbf{r}}$ and $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ reduce to those proposed in Wu and Tsai (2004) when $\mathbf{r} = \mathbf{0}$, and to those in Wu (1995) when $d = 1$.

Remark 1 By arguments similar to those in deriving (2.3)–(2.5) of Wu (1995) (see Lemma 3 herein), (5)–(8) of Wu (1997) and (10)–(13) of Wu and Tsai (2004), we can quickly see that (8) is an unbiased estimate of a sharp upper bound, up to a constant factor and a constant shift, of the risk $\mathrm{E}[\{\tilde{\psi}_{\mathbf{r}}^{+}(\mathbf{T})\}^{1/2} - \{\psi_{\mathbf{r}}^{+}\}^{1/2}]^2$ where $\psi_{\mathbf{r}}^{+}$ and $\tilde{\psi}_{\mathbf{r}}^{+}(\mathbf{T})$ are quantities resulting from replacing $\mathbf{t}^{\mathbf{r}}$ by $|\mathbf{t}^{\mathbf{r}}|$ in (4) and (7), respectively; and, moreover, when $\mathbf{r} = 2\mathbf{m}$, (8) is the product sync-kernel (recalling the above $\hat{f}_{\mathbf{T}}$), Fourier domain version of the cross-validation (CV) score, up to a constant factor, for estimating $f_{\mathbf{m}}$ with badwidth $\mathbf{h} = (T_1^{-1}, \ldots, T_d^{-1})$ (cf. Chacón and Duong 2013; Wu

1997). Next, suppose $n \geq 2$ is fixed and the data matrix $\mathbf{X}$ is given. Rewrite (8) as $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) = \int_{R(\mathbf{T})} |\mathbf{t}^{\mathbf{r}}|\{2/(n+1) - |\tilde{\phi}(\mathbf{t})|^2\}\, d\mathbf{t}$. By the fact that $\tilde{\phi}(\mathbf{t})$ is continuous and $\tilde{\phi}(\mathbf{0}) = 1$, we get $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) \to 0$ as $\|\mathbf{T}\| \to 0$ and $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) < 0$ for all $\mathbf{T}$ in a small neighborhood about $\mathbf{0}$. Also, after some computations we see that $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) \to \infty$ as $\|\mathbf{T}\| \to \infty$. This, together with the continuity of $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$, implies that a global minimizer of $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ exists over $\mathbf{T} > \mathbf{0}$. See also Stone (1984) for an argument showing the existence of a minimizer of the usual CV score.

Next, we consider the problem of estimating the $\Re^{d^{2m}}$-vector functionals $\boldsymbol{\psi}_{2m}$. In view of (3) and (7)–(9), we propose to estimate $\boldsymbol{\psi}_{2m}$ by

$$\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m}) = (\tilde{\psi}_{\mathbf{r}_i}(\hat{\mathbf{T}}_{2m}))_{1 \leq i \leq d^{2m}} \tag{10}$$

where the cutoff $\hat{\mathbf{T}}_{2m}$ is the minimizer of

$$\mathrm{SCV}_{2m}^{\infty}(\mathbf{T}) = \sum_{\mathbf{r}:|\mathbf{r}|=2m} \mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}), \quad \mathbf{T} > \mathbf{0}. \tag{11}$$

For reasons and advantages of using a common cutoff value for the estimation of all components of $\boldsymbol{\psi}_{2m}$, the readers are referred to Duong and Hazelton (2003) and Chacón and Duong (2010), among others. Note that the arguments in Remark 1 also show, when $n \geq 2$ is fixed and $\mathbf{X}$ is given, a global minimizer of $\mathrm{SCV}_{2m}^{\infty}(\mathbf{T})$ exists over $\mathbf{T} > \mathbf{0}$.

2.2 Main theoretical results

The notion of smoothness of f can be expressed in terms of the decay rate of $|\phi_f(\mathbf{t})|$. Throughout we assume that Condition (A) holds for some $\mathbf{r}$ with $|\mathbf{r}| = 2m$, where

Condition (A) *Relation* (4) *holds and for some finite* $p_0 > m + (d/2)$, *it holds that* $\prod_{i=1}^{d} |t_i|^{p_i} |\phi_f(\mathbf{t})| = O(1)$ *as* $\prod_{i=1}^{d} |t_i|^{p_i} \to \infty$ *for every non-negative vector* $\mathbf{p}$ *satisfying* $|\mathbf{p}| = p_0$.

Remark 2 If Condition (A) holds with $p_0 > |\mathbf{m}| + d$ for some $\mathbf{m}$, then by the Fourier inversion formula, $f_{\mathbf{m}}$ exists and is bounded over $\Re^d$. The following are examples of densities fulfilling or not fulfilling Condition (A): (i) a d-variate normal mixture density satisfies Condition (A) for any $\mathbf{r}$, (ii) a d-variate Pareto or exponential density does not satisfy Condition (A) because it has a jump discontinuity and does not satisfy (4) for any $\mathbf{r}$, and (iii) the Mckay's bivariate gamma with density $f(x_1, x_2) = \{\Gamma(a)\Gamma(b)\}^{-1}\lambda^{a+b}x_1^{a-1}(x_2 - x_1)^{b-1}e^{-\lambda x_2}$, $x_2 > x_1 > 0$, a, b, $\lambda > 0$ and characteristic function $\phi_f(\mathbf{t}) = (1 - it_2/\lambda)^{-b}(1 - i(t_1 + t_2)/\lambda)^{-a}$ (see, e.g., Kotz et al. 2000, page 432) fulfills Condition (A) for every $\mathbf{r}$ if $\min\{a, b\} > |\mathbf{r}| + 1$ and $p_0 = \min\{a, b\}$.

The next lemma shows that $\tilde{\psi}_{\mathbf{r}}(\mathbf{T})$ (see (7)) is asymptotically efficient for estimating $\psi_{\mathbf{r}}$ (see Theorem 3 herein) and $\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T})$ (see (10)) is conjectured to be asymptotically efficient for estimating $\boldsymbol{\psi}_{2m}$ (see Remark 3) if the non-adaptive $\mathbf{T}$ is selected from a specific range.

Lemma 1 *For any d and $\mathbf{r}$ with $|\mathbf{r}| = 2m$ assume Condition* (A) *with $p_0 > 2m + d$. Then as $n \to \infty$,*

$$\mathrm{E}\{n(\tilde{\psi}_{\mathbf{r}}(\mathbf{T}) - \psi_{\mathbf{r}})^2\} \to 4Var\{f_{\mathbf{r}}(\mathbf{x}_1)\} \tag{12}$$

provided that the non-adaptive cutoff value $\mathbf{T}$ satisfies

$$\min_{1\le j\le d} T_j \to \infty, \quad \max_{1\le j\le d} T_j = O(n^{1/(2p_0)}), \tag{13}$$

$$g_{\mathbf{r}}^{+}(\mathbf{T}) = O(n^{-1+\{(2m+d)/(2p_0)\}}) \tag{14}$$

with $g_{\mathbf{r}}^{+}(\mathbf{T}) := \int_{\cup_{j=1}^{d}\{\mathbf{t}:|t_j|>T_j\}} |\mathbf{t}^{\mathbf{r}}||\phi_f(\mathbf{t})|^2\,\mathrm{d}\mathbf{t}\ \left(= \int_{R'(\mathbf{T})} |\mathbf{t}^{\mathbf{r}}||\phi_f(\mathbf{t})|^2\,\mathrm{d}\mathbf{t}\right)$. *Further, for all the $\mathbf{r}_i$'s* (*see* (3)) *assume Condition* (A) *with $p_0 > 2m + d$. Put $\sigma_{ij} = Cov(f_{\mathbf{r}_i}(\mathbf{x}_1), f_{\mathbf{r}_j}(\mathbf{x}_1))$. Then*

$$E\{n(\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T}) - \boldsymbol{\psi}_{2m})'(\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T}) - \boldsymbol{\psi}_{2m})\} \to 4(\sigma_{ij})_{1\le i,j\le d^{2m}}, \quad n \to \infty \tag{15}$$

provided that $\mathbf{T}$ satisfies (13) *and $g_{\mathbf{r}_i}^{+}(\mathbf{T}) = O(n^{-1+\{(2m+d)/(2p_0)\}})$ for all $1 \le i \le d^{2m}$.*

We remark that (15) implies the mean squared Euclidean-norm error $\mathrm{E}\{n\|\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T}) - \boldsymbol{\psi}_{2m}\|^2\}$ converges to $4\sum_{i=1}^{d^{2m}} \sigma_{ii}^2$ as $n \to \infty$. Also, we note that the condition $c_0 n^{1/(2p_0)} \le T_j \le c_1 n^{1/(2p_0)}$ for all $1 \le j \le d$ and all sufficiently large n, where $c_1 \ge c_0 > 0$ are absolute constants, is sufficient for (13)–(14) to hold. The next lemma shows that both $\hat{\mathbf{T}}_{\mathbf{r}}$ and $\hat{\mathbf{T}}_{2m}$ (see (8) and (11)) are in the desired range, as specified in Lemma 1.

Lemma 2 *For any d and $\mathbf{r}$ with $|\mathbf{r}| = 2m$ assume Condition* (A). *Then as $n \to \infty$,*

$$\max_{1\le j\le d} \hat{T}_{\mathbf{r},j} = O_p(n^{1/(2p_0)}), \ g_{\mathbf{r}}^{+}(\hat{\mathbf{T}}_{\mathbf{r}}) = O_p(n^{-1+\{(2m+d)/(2p_0)\}}) \tag{16}$$

(*$g_{\mathbf{r}}^{+}$ is defined in Lemma 1 above*). *Further, for all the $\mathbf{r}_i$'s* (*see* (3)) *assume Condition* (A). *Then*

$$\max_{1\le j\le d} \hat{T}_{2m,j} = O_p(n^{1/(2p_0)}) \tag{17}$$

and

$$g_{\mathbf{r}_i}^{+}(\hat{\mathbf{T}}_{2m}) = O_p(n^{-1+\{(2m+d)/(2p_0)\}}) \quad \text{for all } 1 \le i \le d^{2m}. \tag{18}$$

The main results concerning the asymptotic property of our adaptive estimate $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}})$ and $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m})$ (see (9) and (10)) are contained in the next two theorems, and Lemma 2 is the key to prove them.

Theorem 1 *For any d and $\mathbf{r}$ with $|\mathbf{r}| = 2m$ assume Condition* (A). *Then, the following two assertions hold:*

(i) $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}}) - \psi_{\mathbf{r}} = \begin{cases} O_p(n^{-1+\{(2m+d)/(2p_0)\}}), & \text{if } p_0 < 2m+d, \\ O_p(n^{-1/2}\log^d n), & \text{if } p_0 = 2m+d. \end{cases}$

(ii) *If $p_0 > 2m + d$, then*

$$n^{1/2}\{\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}}) - \psi_{\mathbf{r}}\} \to N\big(0, 4Var\{f_{\mathbf{r}}(\mathbf{x}_1)\}\big) \;\; in\ law.$$

Theorem 2 *For any d and all the $\mathbf{r}_i$'s* (*see* (3)) *assume Condition* (A). *Let $\mathbf{1}_{d^{2m}}$ denote a vector of dimension d^{2m} with all entries equaling 1. Then the following two assertions hold:*

(i) $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m}) - \boldsymbol{\psi}_{2m} = \begin{cases} O_p(n^{-1+\{(2m+d)/(2p_0)\}})\mathbf{1}_{d^{2m}}, & \text{if } p_0 < 2m+d, \\ \\ O_p(n^{-1/2}\log^d n)\mathbf{1}_{d^{2m}}, & \text{if } p_0 = 2m+d. \end{cases}$

(ii) *If $p_0 > 2m + d$, then* (*see* (15) *for the definition of σ_{ij}*)

$$n^{1/2}\{\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m}) - \boldsymbol{\psi}_{2m}\} \to N\big(\mathbf{0}, 4(\sigma_{ij})_{1\le i,j\le d^{2m}}\big) \;\; in\ law.$$

The next theorem gives the information bound, in the sense of Koshevnik and Levit (1976), for any nonparametric estimate $\hat{\psi}_{\mathbf{r}}$ of $\psi_{\mathbf{r}} = \psi_{\mathbf{r}}(f)$. This extends the result of Bickel and Ritov (1988) to the multivariate case.

Theorem 3 *For any d and $\mathbf{r}$ with $|\mathbf{r}| = 2m$, let $\mathcal{F}_d = \{f$: Condition* (A) *holds with $p_0 > 2m + d\}$. Then*

$$\lim_{C\to\infty} \liminf_{n\to\infty} \inf_{\hat{\psi}_{\mathbf{r}}} \sup_{g\in H_n(f,C)} E_g\{n(\hat{\psi}_{\mathbf{r}} - \psi_{\mathbf{r}}(g))^2\} \ge 4Var\{f_{\mathbf{r}}(\mathbf{x}_1)\}$$

where $H_n(f, C) = \{g\colon g \in \mathcal{F}_d,\ \|g^{1/2} - f^{1/2}\|_2 \le Cn^{-1/2}\}$ is a Hellinger ball in the neighborhood of f with $\|\cdot\|_2^2 = \int\{\cdot(\mathbf{x})\}^2\, d\mathbf{x}$.

Remark 3 We note that $f_{\mathbf{r}}$ is bounded if $f \in \mathcal{F}_d$ (see Remark 2). Moreover, Theorems 1–3 indicate that the order p_0, which dominates the decay rate of ϕ_f, is crucial. For $p_0 > 2m + d$, both $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}})$ and $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m})$ are $\sqrt{n}$-consistent and $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}})$ achieves the information bound, i.e., achieves the best possible constant coefficient in this convergence. We conjecture that $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m})$ also does so, i.e., $4(\sigma_{ij})_{1\le,i,j\le d^{2m}}$ is the best possible constant covariance matrix in the convergence in the sense that if $n^{1/2}(\hat{\boldsymbol{\psi}}_{2m} - \boldsymbol{\psi}_{2m}) \to N\big(\mathbf{0}, \boldsymbol{\Sigma}\big)$ in law where $\hat{\boldsymbol{\psi}}_{2m}$ is any vector estimate, then the constant covariance matrix $\boldsymbol{\Sigma}$ exceeds $4(\sigma_{ij})_{1\le,i,j\le d^{2m}}$ by a non-negative definite matrix (a related conjecture is that $4\sum_{i=1}^{d^{2m}} \sigma_{ii}^2$ is the smallest possible limiting mean squared Euclidean-norm error for $\sqrt{n}$-convergent vector estimates. See the remark made immediately below Lemma 1). For smaller p_0, $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}})$ and $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m})$ are still consistent but with slower convergence rate. Borrowing the terminology from Efromovich and Low (1996a,b), we may call $p_0 > 2m + d$ a regular case and $m + (d/2) < p_0 \le 2m + d$ an irregular case. We conjecture that in the irregular case the rates described in Theorem

3(i) and Theorem 2(i) are the best possible rates (in a minimax sense, see Efromovich and Low 1996a,b for details) for adaptive estimates and hence $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}})$ and $\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m})$ are rate-optimal. In the present paper, we do not attempt to prove this conjecture because its proof seems complicated and involved. We will report the complete proof in the future. We believe that Efromovich and Low's methods (who only dealt with the case $d = 1$ and f is compactly supported) can be extended to general d and f, with sufficient additional effort.

Remark 4 If $\mathbf{r} = 2\mathbf{m}$ (then $|\mathbf{m}| = m$), then the estimation of $\psi_{\mathbf{r}}$ reduces to that of $\theta_{\mathbf{m}}$ (see (2)). Now, if Condition ($\mathrm{C}_{\mathbf{m}}$) holds, then $|\phi_{f_{\mathbf{m}}}(\mathbf{t})|^2 = \mathbf{t}^{2\mathbf{m}}|\phi_f(\mathbf{t})|^2$. If, in addition, $f_{\mathbf{m}} \in L^2(\Re^d)$, then $\theta_{\mathbf{m}} = (2\pi)^{-d} \int |\phi_{f_{\mathbf{m}}}(\mathbf{t})|^2\, d\mathbf{t}$ by Parseval's formula. Thus

$$\theta_{\mathbf{m}} = (2\pi)^{-d} \int \mathbf{t}^{2\mathbf{m}}|\phi_f(\mathbf{t})|^2\, d\mathbf{t}. \tag{19}$$

Evidently, the sufficient conditions for (19) are essentially weaker than those for (4). It can be easily seen that if we replace (4) by (19) in Condition (A), then (16) and assertion (i) of Theorem 1 (with $\psi_{\mathbf{r}}$ being replaced by $(-1)^m\theta_{\mathbf{m}}$ there) remain true.

2.3 The modification of the proposed estimates

It is well known that the bandwidth selected by CV has a tendency toward undersmoothing. Thus, $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ (recalling Remark 1) and $\mathrm{SCV}_{2m}^{\infty}(\mathbf{T})$ will occasionally select unduly large cutoff values (i.e., small bandwidths) $\hat{\mathbf{T}}_{\mathbf{r}}$ and $\hat{\mathbf{T}}_{2m}$, respectively. In order to reduce such chance, we propose a modification which extends the ones in Chiu (1992) and Wu (1995, 1997) (from $d = 1$ to general d) and Wu and Tsai (2004) (from $\mathbf{r} = \mathbf{0}$ to general $\mathbf{r}$). Let $\hat{L}_{\mathbf{r}}$ and $\hat{L}_{2m}$ denote the open rays (half-lines) starting from the origin and passing through $\hat{\mathbf{T}}_{\mathbf{r}}$ and $\hat{\mathbf{T}}_{2m}$, respectively. We focus on finding cutoff values smaller than $\hat{\mathbf{T}}_{\mathbf{r}}$ and $\hat{\mathbf{T}}_{2m}$ along the one-dimensional rays $\hat{L}_{\mathbf{r}}$ and $\hat{L}_{2m}$, respectively. The simulation study in Sect. 3 shows that our scheme (see below) along $\hat{L}_{\mathbf{r}}$ and $\hat{L}_{2m}$ is computationally efficient and performs quite well. Therefore, our scheme is sufficient from a practical point of view (here we have also overcome the curse of dimensionality, see the beginning of Sect. 2). In what follows, all derivations and solutions are confined to points on $\hat{L}_{\mathbf{r}}$ or $\hat{L}_{2m}$, as the case may be.

We first consider the modification of $\hat{\mathbf{T}}_{\mathbf{r}}$ along $\hat{L}_{\mathbf{r}}$. The basic idea is to use $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ as the cutoff value unless $\tilde{\phi}(\mathbf{t})$ at larger $\mathbf{t}$ contains significant information about f, while $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}} = \min\{\hat{\mathbf{T}}_{\mathbf{r},\mathrm{loc}}, \hat{\mathbf{T}}_{\mathbf{r},u}\}$. Here $\mathbf{T}_{\mathbf{r},\mathrm{loc}}$ is the smallest local minimizer (which plays a pivotal role, see Hall and Marron 1991b) of $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ and $\hat{\mathbf{T}}_{\mathbf{r},u}$ is a minimizer of $|\hat{B}_{\mathbf{r}}(\mathbf{T}, u)|$, where $\hat{B}_{\mathbf{r}}(\mathbf{T}, u)$ is an estimate of the bias $B_{\mathbf{r}}(\mathbf{T}) = \mathrm{E}\tilde{\psi}_{\mathbf{r}}(\mathbf{T}) - \psi_{\mathbf{r}}$ using a reference density u. Thus $\hat{\mathbf{T}}_{\mathbf{r},u}$ is an estimate of the minimizer of $|B_{\mathbf{r}}(\mathbf{T})|$ (which plays an important role, see e.g., Sheather and Jones 1991). The details are given below.

First, for any fixed n it can be shown that (see Lemma 5 herein)

$$\mathrm{Var}\{\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{t}) - \mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{s})\} \sim 2n^{-2}\prod_{i=1}^{d}\left\{(t_i^{2r_i+1} - s_i^{2r_i+1})/(2r_i + 1)\right\}(4\pi)^d\psi_{\mathbf{0}} \tag{20}$$

as $\mathbf{s} \to \infty$ and $\mathbf{t} - \mathbf{s} \to \infty$. Substituting $\tilde{\psi}_{\mathbf{0}}(\hat{\mathbf{T}}_{\mathbf{0}})$ for $\psi_{\mathbf{0}}$ leads to an estimate $\hat{\mathrm{V}}_{\mathbf{r}}(\mathbf{s}, \mathbf{t})$, say, of (20). Next, (38) herein implies that $B_{\mathbf{r}}(\mathbf{T})$ is not estimable, i.e., there does not exist any unbiased estimate of $B_{\mathbf{r}}(\mathbf{T})$ for each fixed n. Following a multivariate version of the scale-model approach for f in Park and Marron (1990), we use

$$\hat{B}_{\mathbf{r}}(\mathbf{T}, u) = (-1)^m n^{-1} (2\pi)^{-d} \left\{ \prod_{j=1}^{d} \int_{-T_j}^{T_j} t_j^{r_j} \, \mathrm{d}t_j \right\} + \tilde{\psi}_{\mathbf{r}}(\mathbf{T}) - \left\{ \prod_{j=1}^{d} \hat{\sigma}_j^{r_j+1} \right\}^{-1} \psi_{\mathbf{r}}(u) \tag{21}$$

to estimate $B_{\mathbf{r}}(\mathbf{T})$, where u is a reference density with covariance matrix equaling the identity matrix $\mathbf{I}$ and $\hat{\sigma}_j$, the estimate of the s.d. of the marginal pdf f_j, equals min{s.d. of $x_{1j}, \ldots, x_{nj}$, (interquartile range of $x_{1j}, \ldots, x_{nj}$)/1.349} (cf. Silverman 1986, page 47) [as an alternative, the hybrid scale measure by Janssen et al. (1995) may also be used here]. It is easy to see if all r_j's are even, $\hat{B}_{\mathbf{r}}(\mathbf{T}, u)$ is strictly monotone and the equation $\hat{B}_{\mathbf{r}}(\mathbf{T}, u) = 0$ has the unique solution $\hat{\mathbf{T}}_{\mathbf{r},u}$. On the other hand, if some of the r_j's are odd, then on the right side of (21), the first term is zero and, moreover, $\psi_{\mathbf{r}}(u) = 0$ if u is the $N(\mathbf{0}, \mathbf{I})$ density. In this case, $\hat{\mathbf{T}}_{\mathbf{r},u}$ is taken to be the smallest solution to $\hat{B}_{\mathbf{r}}(\mathbf{T}, u) = 0$. We define the following modification of $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ beyond $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ (see Remark 5 below for an explanation):

$$\mathrm{CV}_{\mathbf{r}}^{*}(\mathbf{T}) = \mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T}) + z_{.99} \{\hat{\mathrm{V}}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}, \mathbf{T})\}^{1/2} I[\mathbf{T} > \hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}], \quad \mathbf{T} \in \hat{L}_{\mathbf{r}} \tag{22}$$

where $I[\cdot]$ denotes the indicator function, $z_{.99} = 2.33$ is the .99th quantile of $N(0, 1)$. Other quantiles like $z_{.975}$ may be used (of course, quantiles of t-distribution may be preferred for very small sample size).

The modified cutoff value $\hat{\mathbf{T}}_{\mathbf{r}}^{*}$ is the minimizer of $\mathrm{CV}_{\mathbf{r}}^{*}(\mathbf{T})$. This results in the following modified adaptive estimate (which is recommended in practice) of $\psi_{\mathbf{r}}$, namely,

$$\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}}^{*}) = (2\pi)^{-d} (-1)^m \int_{R(\hat{\mathbf{T}}_{\mathbf{r}}^{*})} \mathbf{t}^{\mathbf{r}} |\tilde{\phi}(\mathbf{t})|^2 \, \mathrm{d}\mathbf{t}. \tag{23}$$

Remark 5 Note that $\mathrm{CV}_{\mathbf{r}}^{*}(\mathbf{T})$ modifies $\mathrm{CV}_{\mathbf{r}}^{\infty}(\mathbf{T})$ beyond $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ by adding the non-negative increasing weight function $z_{.99}\{\hat{\mathrm{V}}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}, \mathbf{T})\}^{1/2}$. Evidently, $\hat{\mathbf{T}}_{\mathbf{r}}^{*} \approx \hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ if the weight function is large in magnitude and increases fast in $\mathbf{T}$, i.e., if $\tilde{\phi}(\mathbf{t})$ is dominated by sample variation. Moreover, we note that both $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{loc}}$ and $\hat{\mathbf{T}}_{\mathbf{r},u}$ will occasionally select unduly large values, and taking $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ as the minimum of these two values can significantly improve the practical performance of the estimate (23). Indeed, our simulation study shows that very frequently $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ equals $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{loc}}$ or $\hat{\mathbf{T}}_{\mathbf{r},u}$ according as the underlying density is far away from normal or not.

Remark 6 Clearly, the inequalities $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}} \leq \hat{\mathbf{T}}_{\mathbf{r}}^{*} \leq \hat{\mathbf{T}}_{\mathbf{r}}$ and $g_{\mathbf{r}}^{+}(\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}) \geq g_{\mathbf{r}}^{+}(\hat{\mathbf{T}}_{\mathbf{r}}^{*}) \geq g_{\mathbf{r}}^{+}(\hat{\mathbf{T}}_{\mathbf{r}})$ hold (see (14)). Now, the first-order relation in (16) holds for both $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$ and $\hat{\mathbf{T}}_{\mathbf{r}}^{*}$. Thus, if $\phi_f(\mathbf{t})$ decays in some regular way (e.g., exponential decay or algebraic decay) so that the second order relation in (16) holds for $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$, then it also hold for $\hat{\mathbf{T}}_{\mathbf{r}}^*$. Consequently, Theorem 1 holds for $\tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{\mathbf{r}}^*)$.

We next consider the modification of $\hat{\mathbf{T}}_{2m}$ along $\hat{L}_{2m}$. The ideas are similar to those in deriving (20)–(23). Put $\hat{\mathbf{T}}_{2m,\mathrm{mod}} = \min\{\hat{\mathbf{T}}_{2m,\mathrm{loc}}, \hat{\mathbf{T}}_{2m,u}\}$. Here $\hat{\mathbf{T}}_{2m,\mathrm{loc}}$ is the smallest local minimizer of $\mathrm{SCV}_{2m}^{\infty}(\mathbf{T})$ and $\hat{\mathbf{T}}_{2m,u}$ is the minimizer of the squared Euclidean distance $\sum_{\mathbf{r}:|\mathbf{r}|=2m} \hat{B}_{\mathbf{r}}^2(\mathbf{T}, u)$, as an estimate of $\sum_{\mathbf{r}:|\mathbf{r}|=2m} B_{\mathbf{r}}^2(\mathbf{T})$ (by arguments similar to those near the end of Remark 1, we can show that a global minimizer exists). For any fixed n it can be shown that as $\mathbf{s} \to \infty$ and $\mathbf{t} - \mathbf{s} \to \infty$ (see Lemma 5 herein),

$$\mathrm{Var}\{\mathrm{SCV}_{2m}^{\infty}(\mathbf{t}) - \mathrm{SCV}_{2m}^{\infty}(\mathbf{s})\} = \sum_{\mathbf{r}_1,\mathbf{r}_2} \sigma_{\mathbf{r}_1,\mathbf{r}_2}^{\infty}$$
$$\sim 2n^{-2}(4\pi)^d \psi_{\mathbf{0}} \sum_{\mathbf{r}_1,\mathbf{r}_2} \left\{ \prod_{i=1}^{2} \prod_{j=1}^{d} \left\{ (t_j^{2r_{i,j}+1} - s_j^{2r_{i,j}+1})/(2r_{i,j}+1) \right\}^{1/2} \right\} \tag{24}$$

where the summation is over $\{(\mathbf{r}_1, \mathbf{r}_2) : |\mathbf{r}_i| = 2m,\ i = 1, 2\}$ and $\sigma_{\mathbf{r}_1,\mathbf{r}_2}^{\infty} = \mathrm{Cov}\big\{\mathrm{CV}_{\mathbf{r}_1}^{\infty}(\mathbf{t}) - \mathrm{CV}_{\mathbf{r}_1}^{\infty}(\mathbf{s}), \mathrm{CV}_{\mathbf{r}_2}^{\infty}(\mathbf{t}) - \mathrm{CV}_{\mathbf{r}_2}^{\infty}(\mathbf{s})\big\}$ (noting that $\sigma_{\mathbf{r},\mathbf{r}}^{\infty}$ reduces to (20)). Substituting $\tilde{\psi}_{\mathbf{0}}(\hat{\mathbf{T}}_{\mathbf{0}})$ for $\psi_{\mathbf{0}}$ leads to an estimate $\hat{\mathrm{V}}_{2m}(\mathbf{s}, \mathbf{t})$, say, of (24). The modified cutoff value $\hat{\mathbf{T}}_{2m}^*$ is the minimizer of

$$\mathrm{SCV}_{2m}^*(\mathbf{T}) = \mathrm{SCV}_{2m}^{\infty}(\mathbf{T}) + z_{.99}\{\hat{\mathrm{V}}_{2m}(\hat{\mathbf{T}}_{2m,\mathrm{mod}}, \mathbf{T})\}^{1/2} I[\mathbf{T} > \hat{\mathbf{T}}_{2m,\mathrm{mod}}], \quad \mathbf{T} \in \hat{L}_{2m} \tag{25}$$

and the modified adaptive estimate (which is recommended in practice) of $\boldsymbol{\psi}_{2m}$ is

$$\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m}^*) = (\tilde{\psi}_{\mathbf{r}_i}(\hat{\mathbf{T}}_{2m}^*))_{1 \le i \le d^{2m}} \tag{26}$$

(recalling (11)). Finally, we mention that for the preceding (24)–(26), results similar to those in Remarks 5–6 also hold with trivial modifications (e.g., replacing $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{mod}}$, $\hat{\mathbf{T}}_{\mathbf{r},\mathrm{loc}}$, $\hat{\mathbf{T}}_{\mathbf{r},u}$ and (23) in Remark 5 by $\hat{\mathbf{T}}_{2m,\mathrm{mod}}$, $\hat{\mathbf{T}}_{2m,\mathrm{loc}}$, $\hat{\mathbf{T}}_{2m,u}$ and (26), respectively; and replacing (16) and Theorem 1 in Remark 6 by (17)–(18) and Theorem 2, respectively).

3 Simulation results

We have carried out simulation studies to compare the performance of (i) our estimate $\tilde{\theta}_{\mathbf{m}}^* = (-1)^{|\mathbf{m}|}\tilde{\psi}_{2\mathbf{m}}(\hat{\mathbf{T}}_{2\mathbf{m}}^*)$ (see (23)) with $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$ (see Sect. 1) in the case of estimating $\theta_{\mathbf{m}} = (-1)^{|\mathbf{m}|}\psi_{2\mathbf{m}}$ (see (2)); and (ii) our estimate $\tilde{\boldsymbol{\psi}}_{2m}^*$ $(=\tilde{\boldsymbol{\psi}}_{2m}(\hat{\mathbf{T}}_{2m}^*)$, see (26)) with $\hat{\boldsymbol{\psi}}_{2m,k}^{\mathrm{DH}}$ (the k-stage estimate by Duong and Hazelton (2003) where a single bandwidth is used in their pilot selector) and $\hat{\boldsymbol{\psi}}_{2m,k}^{\mathrm{CD}}$ (the k-stage estimate by Chacón and Duong (2010) where an unconstrained pilot bandwidth matrix is used) in the case of estimating the vector $\boldsymbol{\psi}_{2m}$, while $N(\mathbf{0}, \mathbf{I})$ is used as both the kernel and reference density (see (21)) for our estimates; and $N(\mathbf{0}, \mathbf{I})$ and $N(\mathbf{0}, \mathbf{S}_n)$ with $\mathbf{S}_n$ denoting the sample covariance

Table 1 Parameters for 11 example normal mixture densities

(a) Bivariate density	$\sum_{j=1}^{k}\omega_j N(\mu_{j1},\mu_{j2},\sigma_{j1}^2,\sigma_{j2}^2,\rho_j)$
#1 Normal I	$N(0,0,1,1,0)$ (i.e., $N(\mathbf{0},\mathbf{I})$)
#2 Normal II	$N\left(0,0,\frac{1}{4},1,0\right)$
#3 Skewed	$\frac{1}{5}N(0,0,1,1,0)+\frac{1}{5}N\left(\frac{1}{2},\frac{1}{2},\left(\frac{2}{3}\right)^2,\left(\frac{2}{3}\right)^2,0\right)$ $+\frac{3}{5}N\left(\frac{13}{12},\frac{13}{12},\left(\frac{5}{9}\right)^2,\left(\frac{5}{9}\right)^2,0\right)$
#4 Kurtotic	$\frac{2}{3}N\left(0,0,1,4,\frac{1}{2}\right)+\frac{1}{3}N\left(0,0,\left(\frac{2}{3}\right)^2,\left(\frac{1}{3}\right)^2,-\frac{1}{2}\right)$
#5 Bimodal I	$\frac{1}{2}N(-2,0,1,1,0)+\frac{1}{2}N(2,0,1,1,0)$
#6 Bimodal II	$\frac{1}{2}N\left(1,-1,\left(\frac{2}{3}\right)^2,\left(\frac{2}{3}\right)^2,\frac{3}{5}\right)+\frac{1}{2}N\left(-1,1,\left(\frac{2}{3}\right)^2,\left(\frac{2}{3}\right)^2,\frac{3}{5}\right)$
#7 Bimodal III	$\frac{1}{2}N\left(1,-1,\left(\frac{2}{3}\right)^2,\left(\frac{2}{3}\right)^2,\frac{7}{10}\right)+\frac{1}{2}N\left(-1,1,\left(\frac{2}{3}\right)^2,\left(\frac{2}{3}\right)^2,0\right)$
#8 Trimodal I	$\frac{9}{20}N\left(-\frac{6}{5},\frac{6}{5},\left(\frac{3}{5}\right)^2,\left(\frac{3}{5}\right)^2,\frac{3}{10}\right)+\frac{9}{20}N\left(\frac{6}{5},-\frac{6}{5},\left(\frac{3}{5}\right)^2,\left(\frac{3}{5}\right)^2,-\frac{3}{5}\right)$ $+\frac{1}{10}N\left(0,0,\left(\frac{1}{4}\right)^2,\left(\frac{1}{4}\right)^2,\frac{1}{5}\right)$
#9 Trimodal II	$\frac{3}{7}N\left(-1,0,\left(\frac{3}{5}\right)^2,\left(\frac{7}{10}\right)^2,\frac{3}{5}\right)+\frac{3}{7}N\left(1,\frac{2\sqrt{3}}{3},\left(\frac{3}{5}\right)^2,\left(\frac{7}{10}\right)^2,0\right)$ $+\frac{1}{7}N\left(1,-\frac{2\sqrt{3}}{3},\left(\frac{3}{5}\right)^2,\left(\frac{7}{10}\right)^2,0\right)$
(b) Trivariate density	$\sum_{j=1}^{k}\omega_j N(\mu_{j1},\mu_{j2},\mu_{j3},\sigma_{j1}^2,\sigma_{j2}^2,\sigma_{j3}^2,\rho_{j12},\rho_{j13},\rho_{j23})$
#10 Normal	$N(0,0,0,1,1,1,0,0,0)$ (i.e., $N(\mathbf{0},\mathbf{I})$)
#11 Skewed	$\frac{1}{5}N(0,0,0,1,1,1,0,0,0)+\frac{1}{5}N\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},(\frac{2}{3})^2,(\frac{2}{3})^2,(\frac{2}{3})^2,0,0,0\right)$ $+\frac{3}{5}N\left(\frac{13}{12},\frac{13}{12},\frac{13}{12},(\frac{5}{9})^2,(\frac{5}{9})^2,(\frac{5}{9})^2,0,0,0\right)$

matrix are used as the kernel and reference density, respectively for the estimates $\hat{\theta}_{\mathbf{m}}^{\mathrm{WJ}}$, $\hat{\boldsymbol{\psi}}_{2m,k}^{\mathrm{DH}}$ and $\hat{\boldsymbol{\psi}}_{2m,k}^{\mathrm{CD}}$. See their papers for details of their reference density approach. We also include in the comparison, as a special case of our $\tilde{\boldsymbol{\psi}}_{2m}^{*}$, the estimate $\tilde{\boldsymbol{\psi}}_{2m,S}^{*}$ which uses the same cutoff value in every coordinate direction and is obtained by minimizing the scores resulting from setting $T_1=\cdots=T_d$ in (11), (21) and (25) and $\hat{T}_{2m,1}^{*}=\cdots=\hat{T}_{2m,d}^{*}$ in (26).

For $d=2$ and $d=3$, we generate 100 replications of data sets of various sizes n ($n=200,\ 500$ for $d=2$ and $n=500,\ 900$ for $d=3$) from each of the normal mixture densities given in Table 1. Densities #1–#4, #10 and #11 are unimodal, #5–#7 bimodal and #8–#9 trimodal. From the (asymptotic) relative efficiency point of view, it is adequate to use the same smoothing parameter in every coordinate direction for all these densities except for densities #2, #4 and #9, while for the latter three densities different smoothing parameters should be used in different coordinate directions (see Wand and Jones 1993 for details. The problem addressed by these two authors, although related to the questions studied in the present paper, is not the same). We remark that densities #1, #5 and #10 are considered by Sain et al. (1994), while densities #2, #3, #4, #6, #7, #8 and #9 are essentially the same as densities (A), (C), (D), (G), (H), (I) and (K), respectively, in Wand and Jones (1993). Figure 1 presents the contour plots of the bivariate densities #1–#9 and Fig. 2 the isosurface plots (see, e.g., Panaretos and Konis 2012) of the trivariate densities #10 and #11. For each sample, we apply the fast Fourier transform (FFT) to evaluate $\tilde{\phi}(\mathbf{t})$. The actual implementation is the same as that in Wu and Tsai (2004), and the details are omitted.

For comparison of the above estimates, say, $\hat{\theta}_n$ or $\hat{\boldsymbol{\psi}}_n$, we choose to compare the sample mean squared relative error MSRE $(\hat{\theta}_n) = k^{-1}\sum_{i=1}^{k}(\hat{\theta}_{ni}/\theta_{\mathbf{m}} - 1)^2$ or the sample mean squared Euclidean-norm relative error MSNRE $(\hat{\boldsymbol{\psi}}_n) = k^{-1}\sum_{i=1}^{k}\|\hat{\boldsymbol{\psi}}_{ni} - \boldsymbol{\psi}_{2m}\|^2/\|\boldsymbol{\psi}_{2m}\|^2$, respectively where $k = 100$ and $\hat{\theta}_{ni}$ and $\hat{\boldsymbol{\psi}}_{ni}$ are the estimates based on the i-th replication.

Figures 3 and 4 plot MSRE $(\hat{\theta}_n)$ for $\hat{\theta}_n = \tilde{\theta}^*_{\mathbf{m}}, \tilde{\theta}^*_{\mathbf{m},S}$ and $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ versus $\mathbf{m}$, $0 \le |\mathbf{m}| \le 3$, for all the above densities. They show that in general the level of difficulty of estimating

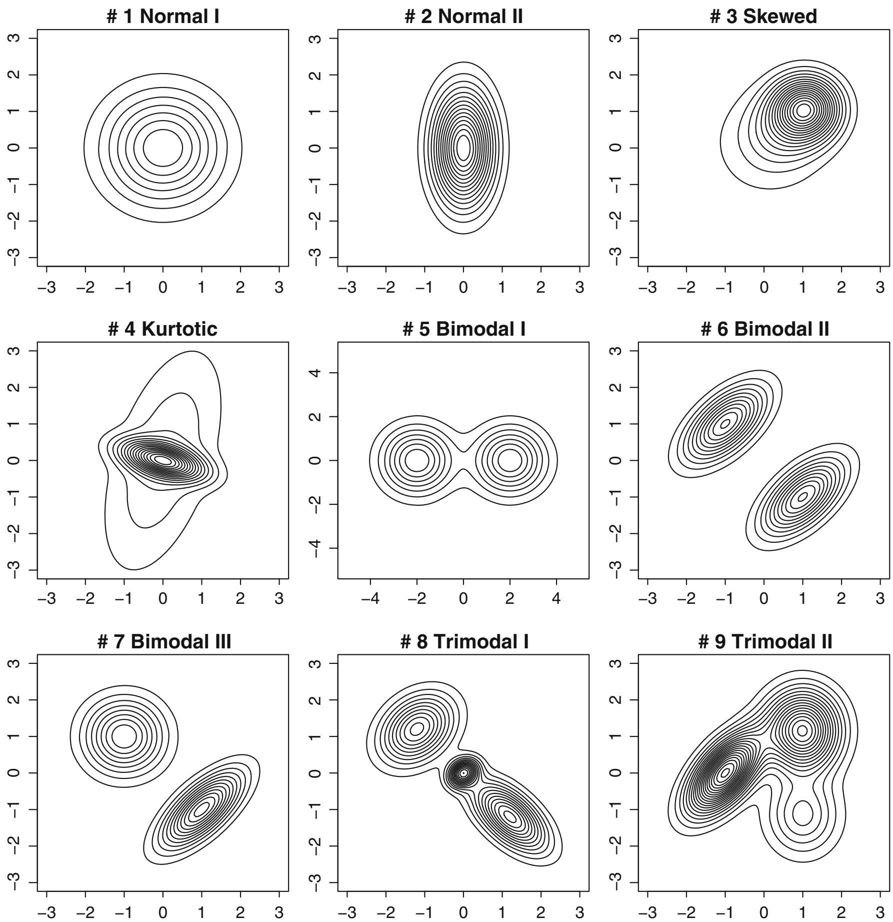

Fig. 1 Contour plots of the bivariate densities #1–#9 defined in Table 1

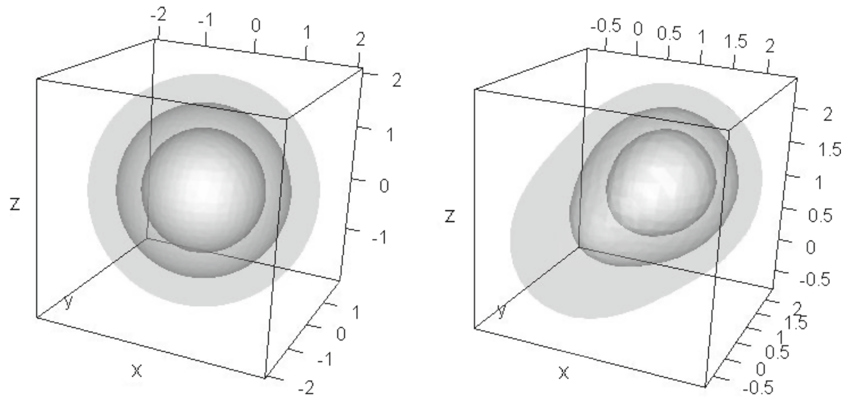

Fig. 2 Isosurface plots of the trivariate densities #10 (*left*) and #11 (*right*) defined in Table 1

$\theta_{\mathbf{m}}$ increases with $|\mathbf{m}|$. Moreover, they clearly show that $\tilde{\theta}^*_{\mathbf{m}}$ is conclusively better than $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ for all the cases except mainly when (i) $n = 200$ and the density is #3, #4 (at $\mathbf{m} = (2, 0), (3, 0)$) or #5 (at $\mathbf{m} = (0, 2), (1, 2), (0, 3)$) and (ii) $n = 500$ and the density is #3 (at $|\mathbf{m}| = 3$) or #10 (at $|\mathbf{m}| = 2$ and $\mathbf{m} = (1, 0, 2), (0, 1, 2)$). For these cases, $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ is (slightly) better than $\tilde{\theta}^*_{\mathbf{m}}$. Thus, overall speaking, $\tilde{\theta}^*_{\mathbf{m}}$ is superior to $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ when the sample size is relatively large or the underlying density is far away from normal (e.g., multimodal), and $\tilde{\theta}^*_{\mathbf{m}}$ is essentially comparable (superior for density #1, but inferior for density #3) to $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ when the sample size is relatively small and the underlying density is not far away from normal (e.g., skewed unimodal). In addition, Figs. 3 and 4 show that overall $\tilde{\theta}^*_{\mathbf{m}}$ is comparable (superior for densities #1, #2, #4, #6, #8, and #9, but inferior for densities #3, #5, #7, #10 and #11) to $\tilde{\theta}^*_{\mathbf{m},S}$. Figures 5 and 6 show the relative biases and s.e.'s (standard errors) of the estimates. They indicate that (i) $\tilde{\theta}^*_{\mathbf{m}}$ and $\tilde{\theta}^*_{\mathbf{m},S}$ have smaller biases, but larger s.e.'s, than $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ for almost all cases except for densities #1–#3, (ii) $\tilde{\theta}^*_{\mathbf{m}}$ has slightly larger s.e.'s than $\tilde{\theta}^*_{\mathbf{m},S}$ in general and (iii) the biases of all estimates tend to increase with $|\mathbf{m}|$ and, in particular, they underestimate $\theta_{\mathbf{m}}$, $|\mathbf{m}| = 3$ quite severely for densities #4 or #8, both are far away from normal.

Tables 2 and 3 show MSNRE($\hat{\boldsymbol{\psi}}_n$) in estimating the vector functionals $\boldsymbol{\psi}_4$ (i.e., $2m = 4$) for $\hat{\boldsymbol{\psi}}_n = \tilde{\boldsymbol{\psi}}^*_4, \tilde{\boldsymbol{\psi}}^*_{4,S}, \hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}, \hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}, \hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$ and $\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,2}$. They show (i) for the sample sizes considered, $\tilde{\boldsymbol{\psi}}^*_{4,S}$ is overall the best among all estimates. In particular, $\tilde{\boldsymbol{\psi}}^*_{4,S}$ is the best for densities #1, #6 and #7, and either the second or third best for the other densities, (ii) $\tilde{\boldsymbol{\psi}}^*_4$ is the best for densities #2 (at $n = 200$), #4, #8 (at $n = 500$) and #9, and the second best for densities #6–#7, (iii) $\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$ performs superiorly for densities #3, #8 and #11 and very well for density #4, (iv) for normal densities #1 and #10, $\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}$ and $\tilde{\boldsymbol{\psi}}^*_{4,S}$ perform the best, followed by $\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}$ and $\tilde{\boldsymbol{\psi}}^*_4$, (v) $\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}$ performs superiorly for density #2 (at $n = 500$), and $\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,2}$ performs very well for densities #3 (at $n = 500$) and #9 and (vi) for density #5, $\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}$ performs the best, followed by $\tilde{\boldsymbol{\psi}}^*_{4,S}$ and $\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$. Tables 2

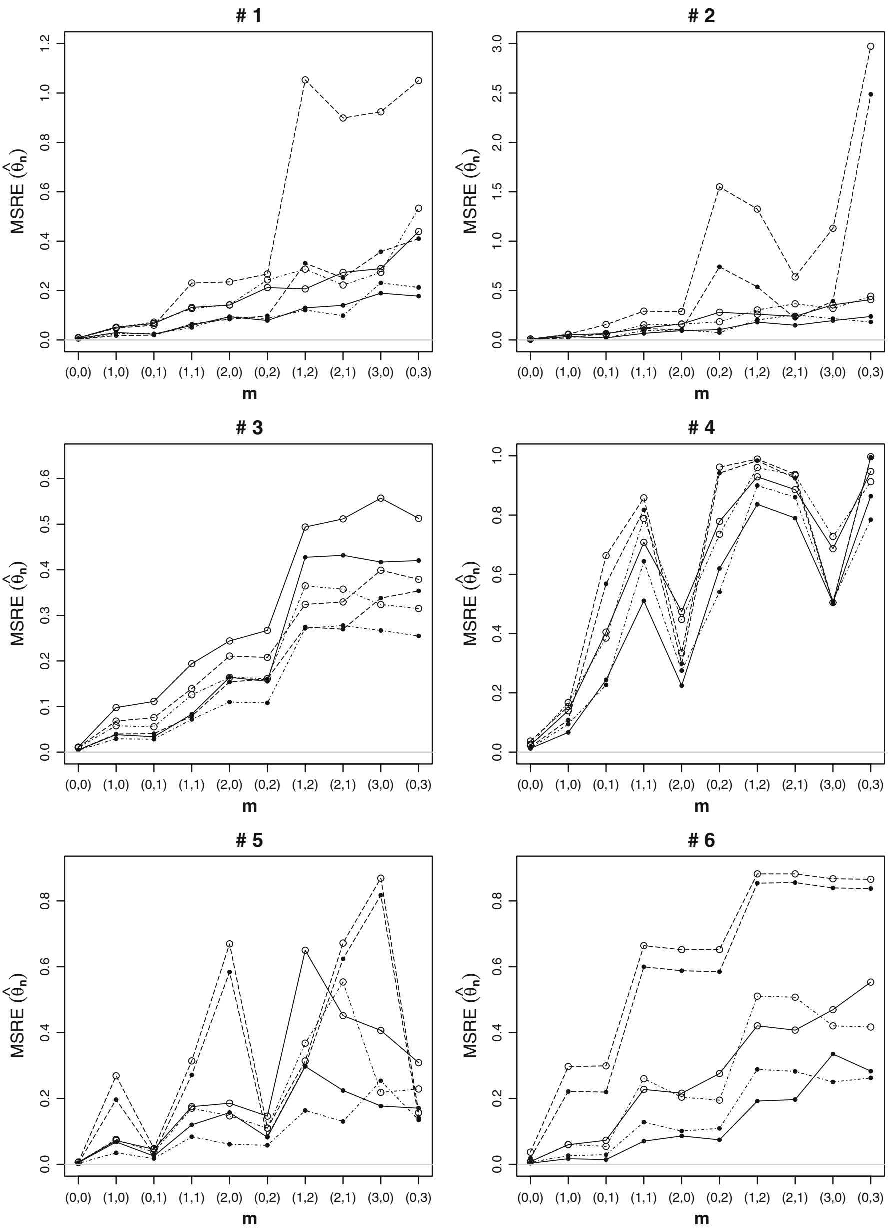

Fig. 3 Plot of the sample mean squared relative error MSRE $(\hat{\theta}_n) = (100)^{-1} \sum_{i=1}^{100} (\hat{\theta}_{ni}/\theta_{\mathbf{m}} - 1)^2$ versus $\mathbf{m}$, $0 \le |\mathbf{m}| \le 3$ for $\hat{\theta}_n = \tilde{\theta}^*_{\mathbf{m}}$ (*solid line*), $\tilde{\theta}^*_{\mathbf{m},S}$ (*dotted-and-dashed line*) and $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ (*dashed line*) based on 100 Monte Carlo replication of samples of size $n = 200$ (*open circle*) and 500 (*soliddot*) from the bivariate densities #1–#9

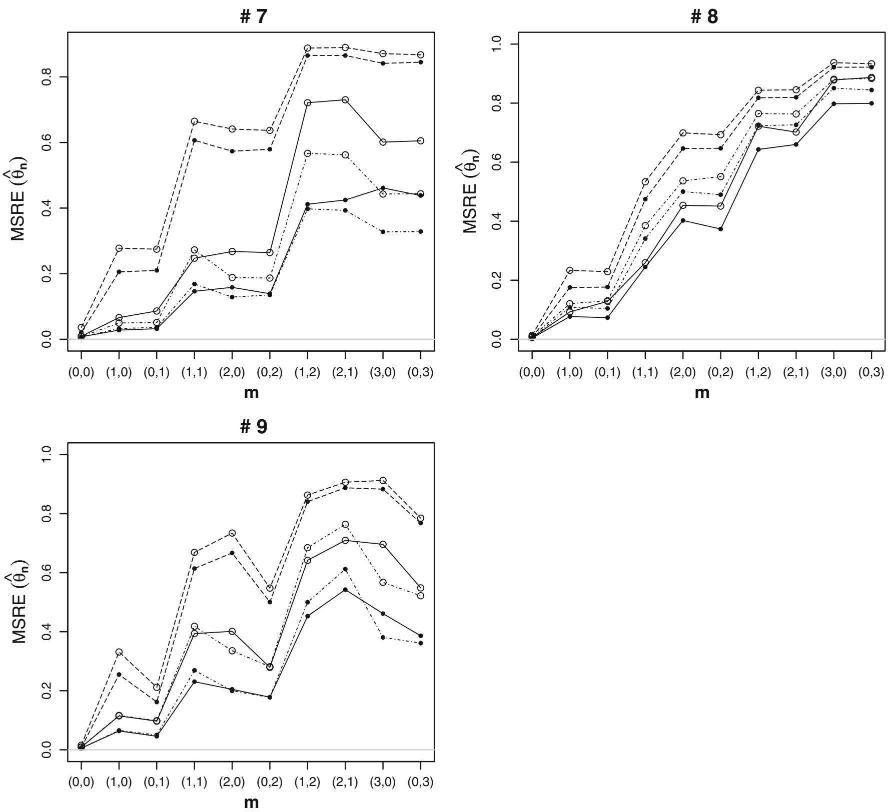

Fig. 3 continued

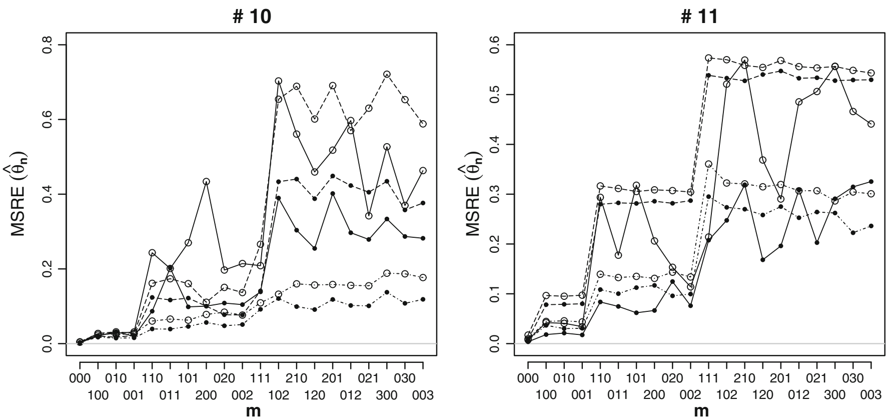

Fig. 4 Plot of the sample mean squared relative error $\text{MSRE}(\hat{\theta}_n) = (100)^{-1} \sum_{i=1}^{100} (\hat{\theta}_{ni}/\theta_{\mathbf{m}} - 1)^2$ versus $\mathbf{m}$, $0 \le |\mathbf{m}| \le 3$ for $\hat{\theta}_n = \tilde{\theta}^*_{\mathbf{m}}$ (*solid line*), $\tilde{\theta}^*_{\mathbf{m},S}$ (*dotted-and-dashed line*) and $\hat{\theta}^{\text{WJ}}_{\mathbf{m}}$ (*dashed line*) based on 100 Monte Carlo replication of samples of size $n = 500$ (*open circle*) and 900 (*solid dot*) from the trivariate densities #10–#11, where $m_1 m_2 m_3$ stands for (m_1, m_2, m_3)

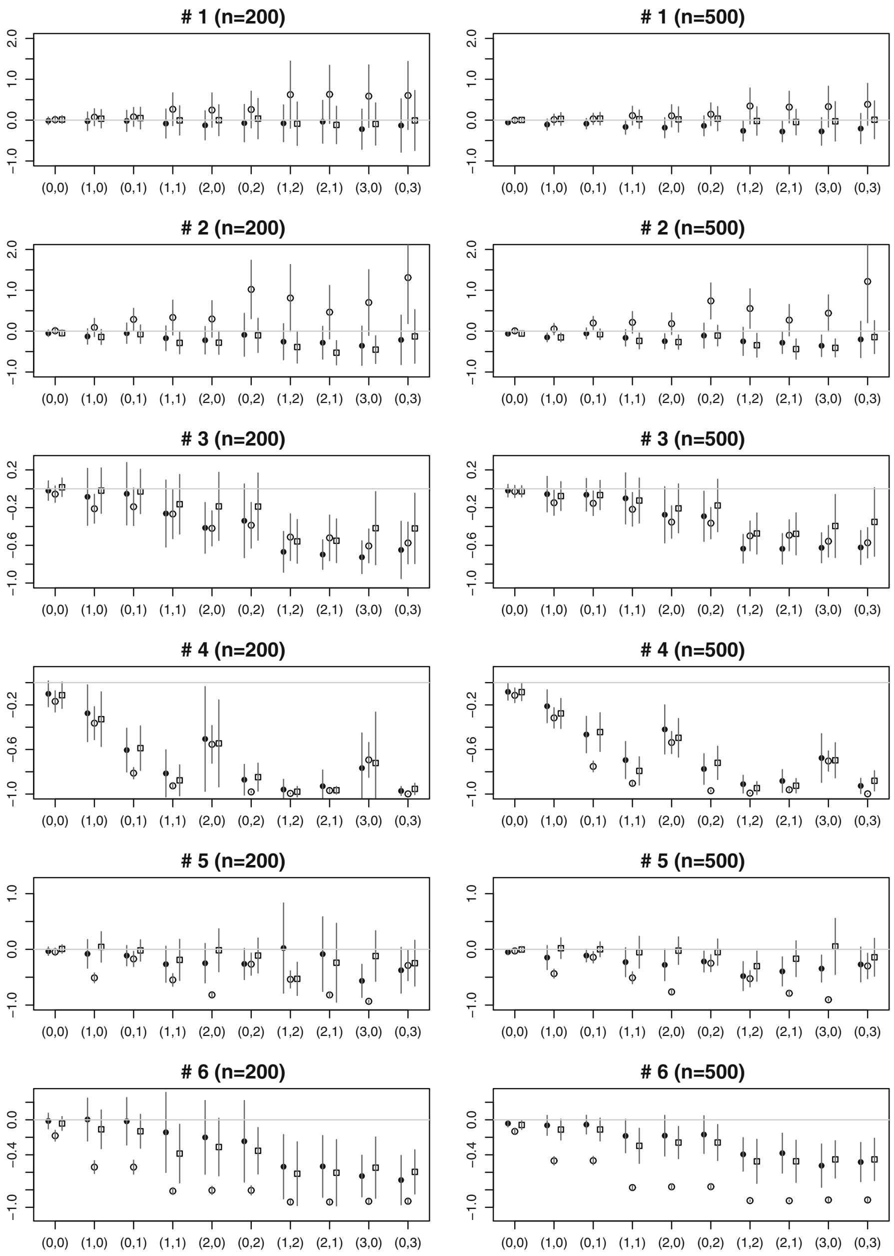

Fig. 5 (The setups are the same as in Fig. 3) Plot of the sample mean relative bias $\bar{B}_{\hat{\theta}_n} = (100)^{-1}\sum_{i=1}^{100}(\hat{\theta}_{ni}/\theta_{\mathbf{m}} - 1) = \bar{\hat{\theta}}_n/\theta_{\mathbf{m}} - 1$ for $\hat{\theta}_n = \tilde{\theta}^*_{\mathbf{m}}$ (*solid circle*), $\tilde{\theta}^*_{\mathbf{m},S}$ (*open square*) and $\hat{\theta}^{\mathrm{WJ}}_{\mathbf{m}}$ (*open circle*) along with $\bar{B}_{\hat{\theta}_n} \pm \hat{\sigma}_n$ (*vertical lines*) where $\hat{\sigma}_n$ is the estimated s.e. of $\hat{\theta}_n/\theta_{\mathbf{m}}$ (which is used, instead of the s.e. of $\bar{\hat{\theta}}_n/\theta_{\mathbf{m}}$, for ease of graphical presentation)

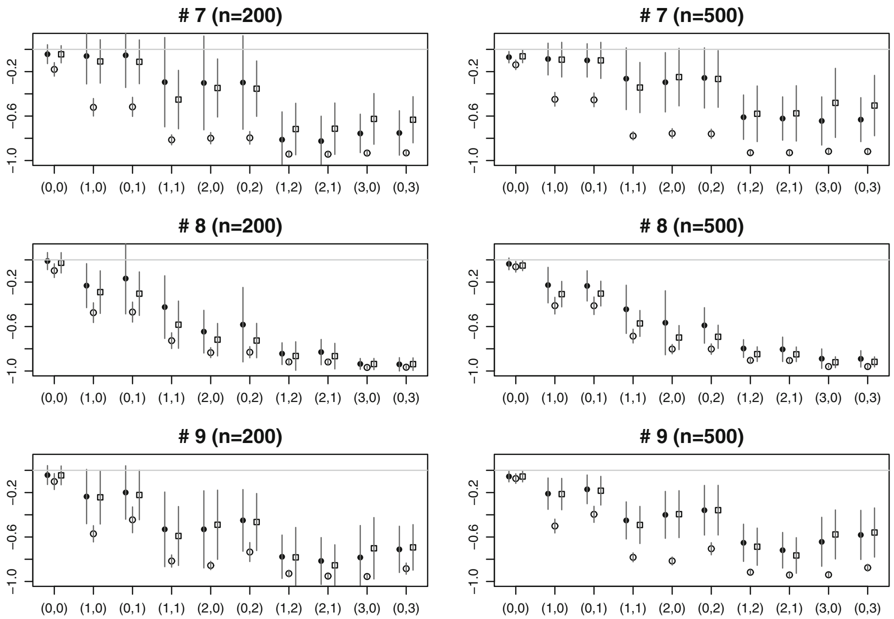

Fig. 5 continued

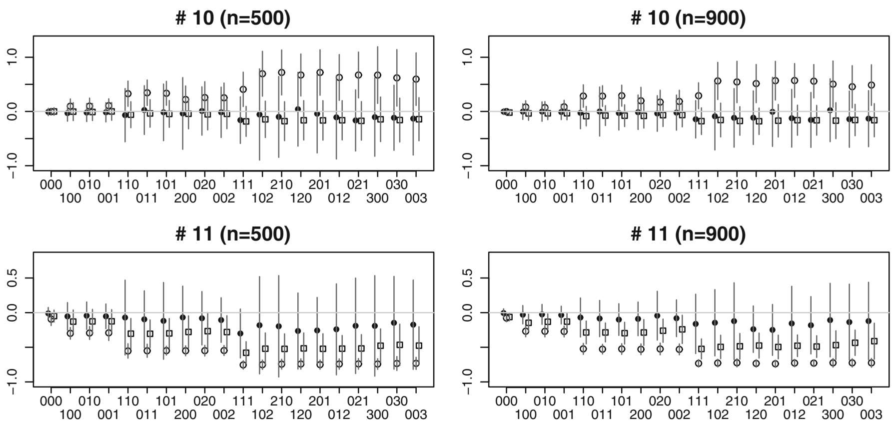

Fig. 6 (The setups are the same as in Fig. 4) For caption see Fig. 4

and 3 also show that for most cases, our estimates have larger s.e.'s and smaller biases than the other estimates. Thus, overall speaking, $\tilde{\boldsymbol{\psi}}_4^*$ and $\tilde{\boldsymbol{\psi}}_{4,S}^*$ perform superiorly or comparably (especially when n is relatively large) according as the underlying density is far away from normal or not.

Finally, we note that Tables 2 and 3 show that in the estimation of $\boldsymbol{\psi}_4$, $\tilde{\boldsymbol{\psi}}_4^*$ performs better than $\tilde{\boldsymbol{\psi}}_{4,S}^*$ for densities #8 (trimodal) and #2, #4 and #9 (while different bandwidths should be used in different coordinate directions, see above). By simulation

Table 2 Simulation results for estimating vector $\boldsymbol{\psi}_4$ with bivariate data

Density	#1		#2		#3	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
$\tilde{\boldsymbol{\psi}}^*_4$	.435 (.059)	.239 (.028)	.185 (.023)	.113 (.012)	.347 (.028)	.169 (.022)
$\tilde{\boldsymbol{\psi}}^*_{4,S}$	.176 (.025)	.084 (.009)	.267 (.014)	.178 (.009)	.163 (.012)	.119 (.012)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}$	.291 (.045)	.121 (.017)	.542 (.056)	.258 (.019)	.166 (.011)	.163 (.008)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}$	.345 (.058)	.135 (.019)	.289 (.054)	.097 (.016)	.194 (.011)	.173 (.008)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$	.749 (.123)	.277 (.047)	1.48 (.207)	.588 (.061)	.161 (.011)	.083 (.007)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,2}$	1.11 (.232)	.392 (.069)	.899 (.181)	.300 (.049)	.243 (.022)	.100 (.007)
Density	#4		#5		#6	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
$\tilde{\boldsymbol{\psi}}^*_4$	.785 (.019)	.595 (.018)	.479 (.079)	.167 (.024)	.315 (.025)	.147 (.011)
$\tilde{\boldsymbol{\psi}}^*_{4,S}$	.882 (.013)	.742 (.011)	.145 (.011)	.070 (.005)	.251 (.014)	.140 (.010)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}$	.899 (.003)	.859 (.003)	.097 (.007)	.064 (.005)	.720 (.005)	.653 (.004)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}$	.986 (.001)	.980 (.001)	.481 (.007)	.404 (.004)	.904 (.001)	.859 (.001)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$	.791 (.008)	.701 (.008)	.255 (.039)	.134 (.024)	.524 (.008)	.426 (.007)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,2}$	.962 (.003)	.928 (.002)	.477 (.031)	.281 (.016)	.668 (.005)	.531 (.005)
Density	#7		#8		#9	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
$\tilde{\boldsymbol{\psi}}^*_4$	.316 (.022)	.196 (.015)	.513 (.018)	.462 (.012)	.419 (.020)	.260 (.016)
$\tilde{\boldsymbol{\psi}}^*_{4,S}$	.288 (.017)	.187 (.012)	.513 (.016)	.474 (.010)	.433 (.018)	.274 (.012)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,1}$	.724 (.006)	.655 (.005)	.646 (.009)	.623 (.006)	.708 (.007)	.649 (.005)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,1}$	.927 (.001)	.895 (.001)	.748 (.006)	.730 (.005)	.663 (.009)	.598 (.006)
$\hat{\boldsymbol{\psi}}^{\mathrm{DH}}_{4,2}$	.556 (.010)	.453 (.010)	.492 (.013)	.470 (.010)	.528 (.012)	.433 (.009)
$\hat{\boldsymbol{\psi}}^{\mathrm{CD}}_{4,2}$	.774 (.005)	.651 (.005)	.679 (.007)	.670 (.005)	.430 (.014)	.329 (.011)

Sample means of the squared Euclidean-norm relative error $D = \|\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_4\|^2/\|\boldsymbol{\psi}_4\|^2$ are given for $n = 200,\ 500$ from each of the bivariate densities #1–#9 (100 replications in each case). The value inside the parentheses is the estimated standard error of $\bar{D}$

studies, we find that this pattern can also be observed in the estimation of the vectors $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_6$ at the sample size n specified in Tables 2 and 3. Moreover, as n becomes large ($n = 900,\ 1{,}600$ in our simulation setup), $\tilde{\boldsymbol{\psi}}^*_{2m}$ will outperform $\tilde{\boldsymbol{\psi}}^*_{2m,S}$, $1 \le m \le 3$, for most of the 11 densities considered. The details of the simulation results are not presented here to save space. The complete results can be found at http://www.stat.ncku.edu.tw/faculty_private/tjwu/.

In summary, Figs. 3 and 4 and Tables 2 and 3 reveal that over a wide range of smooth density shapes and at practical sample sizes, the overall performance of the

Table 3 Simulation results for estimating vector $\boldsymbol{\psi}_4$ with trivariate data

Density	#10		#11	
	$n = 500$	$n = 900$	$n = 500$	$n = 900$
$\tilde{\boldsymbol{\psi}}_4^*$	.345 (.102)	.185 (.036)	.221 (.014)	.156 (.006)
$\tilde{\boldsymbol{\psi}}_{4,S}^*$	.154 (.021)	.113 (.014)	.171 (.008)	.134 (.005)
$\hat{\boldsymbol{\psi}}_{4,1}^{\mathrm{DH}}$	.147 (.014)	.085 (.008)	.211 (.009)	.197 (.006)
$\hat{\boldsymbol{\psi}}_{4,1}^{\mathrm{CD}}$	.169 (.015)	.097 (.008)	.300 (.010)	.246 (.007)
$\hat{\boldsymbol{\psi}}_{4,2}^{\mathrm{DH}}$	.439 (.039)	.241 (.021)	.093 (.007)	.072 (.004)
$\hat{\boldsymbol{\psi}}_{4,2}^{\mathrm{CD}}$	.566 (.048)	.323 (.026)	.279 (.020)	.191 (.011)

The note for Table 2 holds except now $n = 500,\ 900$ and densities are the trivariate #10–#11

proposed estimates $\tilde{\theta}^*_{\mathbf{m}}$, $\tilde{\theta}^*_{\mathbf{m},S}$ $(0 \le |\mathbf{m}| \le 3)$, $\tilde{\boldsymbol{\psi}}^*_{2m}$ and $\tilde{\boldsymbol{\psi}}^*_{2m,S}$ $(2m = 4)$ are quite good. In addition, they reveal that for smooth densities the convergence rates, as $n \to \infty$, of the proposed estimates to their target values are fast. This agrees with the earlier theoretical results. Finally, it should be mentioned that our algorithm is fairly time efficient. For example, using a PC with Pentium D processor running at 2.8 Ghz CPU and 1 GB RAM, it takes only ~7.848 CPU s to finish the computation of all the ten estimates $\tilde{\theta}^*_{\mathbf{m}}$, $0 \le |\mathbf{m}| \le 3$, based on a sample of size $n = 500$ from the bivariate $N(\mathbf{0}, \mathbf{I})$ density.

4 Proofs

For any d-dimensional measurable set $\mathbf{A}$, we set $W_{1,\mathbf{r}}(\mathbf{A}) = \int_{\mathbf{A}} \mathbf{t}^{\mathbf{r}} |\tilde{\phi}_0(\mathbf{t})|^2\, \mathrm{d}\mathbf{t}$, $W_{2,\mathbf{r}}(\mathbf{A}) = \int_{\mathbf{A}} \mathbf{t}^{\mathbf{r}} \phi_f(-\mathbf{t})\tilde{\phi}_0(\mathbf{t})\, \mathrm{d}\mathbf{t}$ and $W_{3,\mathbf{r}}(\mathbf{A}) = \int_{\mathbf{A}} \mathbf{t}^{\mathbf{r}} \{|\tilde{\phi}(\mathbf{t})|^2 - |\phi_f(\mathbf{t})|^2\}\, \mathrm{d}\mathbf{t}$ where $\tilde{\phi}_0 = \tilde{\phi} - \phi_f$. Also, let $W^+_{j,\mathbf{r}}(\mathbf{A})$ be the quantity resulting from replacing $\mathbf{t}^{\mathbf{r}}$ by $|\mathbf{t}^{\mathbf{r}}|$ in $W_{j,\mathbf{r}}(\mathbf{A})$, $j = 1, 2, 3$. Note that $W_{3,\mathbf{r}}(\mathbf{A}) = W_{1,\mathbf{r}}(\mathbf{A}) + 2\mathrm{Re}(W_{2,\mathbf{r}}(\mathbf{A}))$ and $\mathrm{Re}(W_{2,\mathbf{r}}(\mathbf{A})) = W_{2,\mathbf{r}}(\mathbf{A})$ if $\mathbf{A} = -\mathbf{A}$ (similar equations also hold for the $W^+_{j,\mathbf{r}}(\mathbf{A})$'s). The notations $W^0_{j,\mathbf{r}}(\mathbf{A}) = W_{j,\mathbf{r}}(\mathbf{A}) - \mathrm{E}W_{j,\mathbf{r}}(\mathbf{A})$ and $W^{+,0}_{j,\mathbf{r}}(\mathbf{A}) = W^+_{j,\mathbf{r}}(\mathbf{A}) - \mathrm{E}W^+_{j,\mathbf{r}}(\mathbf{A})$ shall be used.

In the sequel, $C > 0$ denotes a finite generic constant that does not depend on n, $(\cdot)_{i,j}$ is shorthand for matrix $(\cdot)_{1\le i,j\le d^{2m}}$, and $\mathbf{1}_{p\times q}$ denotes a $(p \times q)$-matrix with all entries equaling 1. Also, according to our previous notation, $\mathbf{r}^{\mathbf{c}} = \prod_{m=i}^{j} r_m^c$ for any $\mathbf{r} = (r_i, \ldots, r_j)$ and $\mathbf{c} = (c, c, \ldots, c, c)$. Thus, for example, $\mathbf{r}^{\mathbf{1}} = \prod_{m=i}^{j} r_m$, $\mathbf{r}^{-\mathbf{1/2}} = \prod_{m=i}^{j} r_m^{-1/2}$, etc.

Lemma 3 *It holds that $E\tilde{\phi}(\mathbf{t}) = \phi_f(\mathbf{t})$. Also, the following result concerning cumulant holds:*

$$cum(\tilde{\phi}(\mathbf{t}_1), \tilde{\phi}(\mathbf{t}_2)) = \{\phi_f(\mathbf{t}_1 + \mathbf{t}_2) - \phi_f(\mathbf{t}_1)\phi_f(\mathbf{t}_2)\}/n. \tag{27}$$

Hence $Var\{\tilde{\phi}(\mathbf{t})\}$ $(=E|\tilde{\phi}_0(\mathbf{t})|^2) = n^{-1}(1 - |\phi_f(\mathbf{t})|^2)$ and $E|\tilde{\phi}(\mathbf{t})|^2 = (n - 1)n^{-1}|\phi_f(\mathbf{t})|^2 + n^{-1}$.

Lemma 4 *The following hold for any* $\mathbf{t}_1$ *and* $\mathbf{t}_2$*:*

$$\begin{aligned}
&cum(|\tilde{\phi}_0(\mathbf{t}_1)|^2, \tilde{\phi}_0(\mathbf{t}_2)) \\
&\quad = n^{-2}\left\{2|\phi_f(\mathbf{t}_1)|^2\phi_f(\mathbf{t}_2) - \phi_f(-\mathbf{t}_1)\phi_f(\mathbf{t}_1+\mathbf{t}_2) - \phi_f(\mathbf{t}_1)\phi_f(\mathbf{t}_2-\mathbf{t}_1)\right\}; \qquad (28) \\
&cum(\tilde{\phi}_0(\mathbf{t}_1), \tilde{\phi}_0(-\mathbf{t}_1), \tilde{\phi}_0(\mathbf{t}_2), \tilde{\phi}_0(-\mathbf{t}_2)) \\
&\quad = n^{-3}\Big\{2\sum_{\mathbf{a},\mathbf{b}}\phi_f(\mathbf{a})\phi_f(\mathbf{b})\phi_f(-\mathbf{a}-\mathbf{b}) - 6|\phi_f(\mathbf{t}_1)|^2|\phi_f(\mathbf{t}_2)|^2 \\
&\qquad -|\phi_f(\mathbf{t}_1+\mathbf{t}_2)|^2 - |\phi_f(\mathbf{t}_1-\mathbf{t}_2)|^2\Big\} \qquad (29)
\end{aligned}$$

where the summation is over $\{(\mathbf{a},\mathbf{b}) : \mathbf{a} = \mathbf{t}_1 \text{ or } -\mathbf{t}_1, \mathbf{b} = \mathbf{t}_2 \text{ or } -\mathbf{t}_2\}$*. Furthermore, for any* $s \geq 2$*, it holds uniformly in* $\mathbf{t}_1, \ldots, \mathbf{t}_s$ *that*

$$cum(\tilde{\phi}(\mathbf{t}_1), \ldots, \tilde{\phi}(\mathbf{t}_s)) = O(n^{-s+1}), \qquad n \to \infty. \qquad (30)$$

The preceding two lemmas are Lemma 1 and Lemma A.1, respectively, of Wu and Tsai (2004). They are stated here to make this article self-contained.

Lemma 5 *For any* $\mathbf{A} = R(\mathbf{T}),\ R(\mathbf{T}) \setminus R(\mathbf{S}),\ or\ \Re^d$*, where* $\mathbf{S} < \mathbf{T} < \infty$*,*

$$\begin{aligned}
Cov\{W_{3,\mathbf{r}_1}(\mathbf{A}), W_{3,\mathbf{r}_2}(\mathbf{A})\} = {}& 2n^{-2}\{Q_{1,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) - Q_{3,2,\mathbf{r}_1}(\mathbf{A})Q_{3,2,\mathbf{r}_2}(\mathbf{A})\} \\
&+4(n^{-1}-3n^{-2})\{Q_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) - Q_{3,2,\mathbf{r}_1}(\mathbf{A})Q_{3,2,\mathbf{r}_2}(\mathbf{A})\} \\
&+n^{-3}Q_{4,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) \qquad (31)
\end{aligned}$$

and

$$\begin{aligned}
Cov\{W^+_{3,\mathbf{r}_1}(\mathbf{A}), W^+_{3,\mathbf{r}_2}(\mathbf{A})\} = {}& 2n^{-2}\{Q^+_{1,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) - Q^+_{3,2,\mathbf{r}_1}(\mathbf{A})Q^+_{3,2,\mathbf{r}_2}(\mathbf{A})\} \\
&+4(n^{-1}-3n^{-2})\{Q^+_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) \\
&-Q^+_{3,2,\mathbf{r}_1}(\mathbf{A})Q^+_{3,2,\mathbf{r}_2}(\mathbf{A})\} + n^{-3}Q^+_{4,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) \qquad (32)
\end{aligned}$$

where $Q_{j,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A})$ *and* $Q_{3,i,\mathbf{r}}(\mathbf{A})$ *are quantities resulting from replacing* $|\mathbf{t}_1^{\mathbf{r}_1}||\mathbf{t}_2^{\mathbf{r}_2}|$ *and* $|\mathbf{t}^{\mathbf{r}}|$ *by* $\mathbf{t}_1^{\mathbf{r}_1}\mathbf{t}_2^{\mathbf{r}_2}$ *and* $\mathbf{t}^{\mathbf{r}}$ *in* $Q^+_{j,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A})$ *and* $Q^+_{3,i,\mathbf{r}}(\mathbf{A})$*, respectively* ($j = 1, 2$ *and* 4)*, with*

$$\begin{aligned}
Q^+_{1,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) &= \int_{\mathbf{A}}\int_{\mathbf{A}} |\mathbf{t}_1^{\mathbf{r}_1}||\mathbf{t}_2^{\mathbf{r}_2}||\phi_f(\mathbf{t}_1+\mathbf{t}_2)|^2\, d\mathbf{t}_1\, d\mathbf{t}_2, \\
Q^+_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) &= \int_{\mathbf{A}}\int_{\mathbf{A}} |\mathbf{t}_1^{\mathbf{r}_1}||\mathbf{t}_2^{\mathbf{r}_2}|\phi_f(-\mathbf{t}_1)\phi_f(-\mathbf{t}_2)\phi_f(\mathbf{t}_1+\mathbf{t}_2)\, d\mathbf{t}_1\, d\mathbf{t}_2, \\
Q^+_{3,i,\mathbf{r}}(\mathbf{A}) &= \int_{\mathbf{A}} |\mathbf{t}^{\mathbf{r}}||\phi_f(\mathbf{t})|^i\, d\mathbf{t}, \quad i = 1, 2,
\end{aligned}$$

and $Q^+_{4,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A})$ *satisfies* $|Q^+_{4,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A})| \leq 2Q^+_{1,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) + 14Q^+_{3,1,\mathbf{r}_1}(\mathbf{A})Q^+_{3,1,\mathbf{r}_2}(\mathbf{A})$.

Proof of Lemma 5 Let $c(\mathbf{t}_1, \mathbf{t}_2)$ and $d(\mathbf{t}_1, \mathbf{t}_2)$ denote the left-hand sides of (27) and (29), respectively. By Lemmas 3 and 4,

$$\begin{aligned}
\mathrm{Cov}\{W_{1,\mathbf{r}_1}(\mathbf{A}), W_{1,\mathbf{r}_2}(\mathbf{A})\} &= \int_{\mathbf{A}}\int_{\mathbf{A}} \mathbf{t}_1^{\mathbf{r}_1}\mathbf{t}_2^{\mathbf{r}_2}\mathrm{cum}(|\tilde{\phi}_0(\mathbf{t}_1)|^2, |\tilde{\phi}_0(\mathbf{t}_2)|^2)\,\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2 \\
&= \int_{\mathbf{A}}\int_{\mathbf{A}} \mathbf{t}_1^{\mathbf{r}_1}\mathbf{t}_2^{\mathbf{r}_2}\{2c(\mathbf{t}_1, \mathbf{t}_2)c(-\mathbf{t}_1, -\mathbf{t}_2) + d(\mathbf{t}_1, \mathbf{t}_2)\}\,\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2 \\
&= 2n^{-2}\{Q_{1,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) - 2Q_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) \\
&\quad + Q_{3,2,\mathbf{r}_1}(\mathbf{A})Q_{3,2,\mathbf{r}_2}(\mathbf{A})\} + n^{-3}Q_{4,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}), \qquad (33)\\
n\mathrm{Cov}\{W_{2,\mathbf{r}_1}(\mathbf{A}), W_{2,\mathbf{r}_2}(\mathbf{A})\} &= \int_{\mathbf{A}}\int_{\mathbf{A}} \mathbf{t}_1^{\mathbf{r}_1}\mathbf{t}_2^{\mathbf{r}_2}\phi_f(-\mathbf{t}_1)\phi_f(-\mathbf{t}_2)\mathrm{cum}(\tilde{\phi}_0(\mathbf{t}_1), \tilde{\phi}_0(\mathbf{t}_2))\,\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2 \\
&= Q_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A}) - Q_{3,2,\mathbf{r}_1}(\mathbf{A})Q_{3,2,\mathbf{r}_2}(\mathbf{A}) \qquad (34)
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}\{W_{1,\mathbf{r}_1}(\mathbf{A}), W_{2,\mathbf{r}_2}(\mathbf{A})\} &= \int_{\mathbf{A}}\int_{\mathbf{A}} \mathbf{t}_1^{\mathbf{r}_1}\mathbf{t}_2^{\mathbf{r}_2}\phi_f(-\mathbf{t}_2)\mathrm{cum}(|\tilde{\phi}_0(\mathbf{t}_1)|^2, \tilde{\phi}_0(\mathbf{t}_2))\,\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2 \\
&= 2n^{-2}\{Q_{3,2,\mathbf{r}_1}(\mathbf{A})Q_{3,2,\mathbf{r}_2}(\mathbf{A}) - Q_{2,\mathbf{r}_1,\mathbf{r}_2}(\mathbf{A})\}. \qquad (35)
\end{aligned}$$

The proof of Eq. (31) follows by noting that

$$\mathrm{Cov}\{W_{3,\mathbf{r}_1}(\mathbf{A}), W_{3,\mathbf{r}_2}(\mathbf{A})\} = \mathrm{Cov}\{W_{1,\mathbf{r}_1}(\mathbf{A}) + 2W_{2,\mathbf{r}_1}(\mathbf{A}), W_{1,\mathbf{r}_2}(\mathbf{A}) + 2W_{2,\mathbf{r}_2}(\mathbf{A})\}. \qquad (36)$$

Now, (33)–(36) remain true if we replace W by W^+, Q by Q^+ and $\mathbf{t}_k^{\mathbf{r}_k}$ by $|\mathbf{t}_k^{\mathbf{r}_k}|$ $(k = 1, 2)$ throughout. This establishes (32). □

For the rest of the paper, $|\mathbf{r}| = 2m$ and $|\mathbf{r}_i| = 2m$ for all occurrences of $\mathbf{r}$ and $\mathbf{r}_i$, $1 \le i \le d^{2m}$.

Proof of Lemma 1 First, we derive (15). We note that $|Q_{3,2,\mathbf{r}}| \le Q^+_{3,1,\mathbf{r}}$ and $|Q_{2,\mathbf{r}_i,\mathbf{r}_j}| \le Q^+_{3,1,\mathbf{r}_i}Q^+_{3,1,\mathbf{r}_j}$. The condition $p_0 > 2m + d$ entails $Q^+_{3,1,\mathbf{r}}(\Re^d) < \infty$ and $f_{\mathbf{r}}$ is bounded over $\Re^d$ (see Remark 3). Now,

$$(-1)^m(2\pi)^d(\tilde{\psi}_{\mathbf{r}}(\mathbf{T}) - \psi_{\mathbf{r}}) = W_{1,\mathbf{r}}(R(\mathbf{T})) + 2W_{2,\mathbf{r}}(R(\mathbf{T})) - g_{\mathbf{r}}(\mathbf{T}) \qquad (37)$$

where $g_{\mathbf{r}}(\mathbf{T})$ denotes the quantity resulting from replacing $|\mathbf{t}|^{\mathbf{r}}$ by $\mathbf{t}^{\mathbf{r}}$ in $g^+_{\mathbf{r}}(\mathbf{T})$ (noting that $Q_{3,2,\mathbf{r}}(R'(\mathbf{T})) = g_{\mathbf{r}}(\mathbf{T})$). By Lemma 3 and the fact $\mathrm{E}W_{2,\mathbf{r}}(R(\mathbf{T})) = 0$,

$$(-1)^m(2\pi)^d B_{\mathbf{r}}(\mathbf{T}) = n^{-1}\int_{R(\mathbf{T})} \mathbf{t}^{\mathbf{r}}\,\mathrm{d}\mathbf{t} - n^{-1}Q_{3,2,\mathbf{r}}(R(\mathbf{T})) - g_{\mathbf{r}}(\mathbf{T}) \qquad (38)$$

where $B_{\mathbf{r}}(\mathbf{T}) = \mathrm{E}\tilde{\psi}_{\mathbf{r}}(\mathbf{T}) - \psi_{\mathbf{r}}$ denotes the bias. Consequently, $n|B_{\mathbf{r}_i}(\mathbf{T})B_{\mathbf{r}_j}(\mathbf{T})| \to 0$ as $n \to \infty$ under the conditions of the present lemma. Next, by (37),

$(2\pi)^{2d}\mathrm{Var}\{\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T})\} = (\sigma^*_{3,\mathbf{r}_i,\mathbf{r}_j}(R(\mathbf{T}))_{i,j}$ where $\sigma^*_{k,\mathbf{r}_i,\mathbf{r}_j}(\mathbf{A}) = \mathrm{Cov}\{W_{k,\mathbf{r}_i}(\mathbf{A}), W_{k,\mathbf{r}_j}(\mathbf{A})\}$. This, together with (31), (34) and the fact that

$$Q^+_{1,\mathbf{r}_i,\mathbf{r}_j}(R(\mathbf{T})) \le \mathbf{T}^{\mathbf{r}_i}\int_{R(\mathbf{T})}\left|\mathbf{t}_2^{\mathbf{r}_j}\right|\int|\phi_f(\mathbf{t_1}+\mathbf{t_2})|^2\,\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2 \le C\mathbf{T}^{\mathbf{r}_i+\mathbf{r}_j+1} \tag{39}$$

(here and below C does not depend on $\mathbf{T}$ either), yields

$$(2\pi)^{2d}\mathrm{Var}\{\tilde{\boldsymbol{\psi}}_{2m}(\mathbf{T})\} = 4(\sigma^*_{2,\mathbf{r}_i,\mathbf{r}_j}(R(\mathbf{T}))_{i,j} + Cn^{-2}(\mathbf{T}^{\mathbf{r}_i+\mathbf{r}_j+1})_{i,j} + O(n^{-2})\mathbf{1}_{d^{2m}\times d^{2m}}. \tag{40}$$

Now, since $Q^+_{3,1,\mathbf{r}}(\Re^d) < \infty$, we get, by arguments similar to those immediately below (4), $f_\mathbf{r}(\mathbf{x}) = (2\pi)^{-d}\int(-1)^m e^{-i\mathbf{t}\mathbf{x}'}\mathbf{t}^\mathbf{r}\phi_f(\mathbf{t})\,\mathrm{d}\mathbf{t}$ for all $\mathbf{x}$. It follows that

$$\begin{aligned}(2\pi)^{2d}\mathrm{E}\{f_{\mathbf{r}_i}(\mathbf{x}_1)f_{\mathbf{r}_j}(\mathbf{x}_1)\} &= \int\int \mathbf{t}_1^{\mathbf{r}_i}\mathbf{t}_2^{\mathbf{r}_j}\phi_f(\mathbf{t}_1)\phi_f(\mathbf{t}_2)\left\{\int e^{-i(\mathbf{t}_1+\mathbf{t}_2)\mathbf{x}'}f(\mathbf{x})\,\mathrm{d}\mathbf{x}\right\}\mathrm{d}\mathbf{t}_1\,\mathrm{d}\mathbf{t}_2\\ &= Q_{2,\mathbf{r}_i,\mathbf{r}_j}(\Re^d)\end{aligned}$$

and, similarly, $(2\pi)^d\mathrm{E}\{f_\mathbf{r}(\mathbf{x}_1)\} = (-1)^m Q_{3,2,\mathbf{r}}(\Re^d)$. In view of (34),

$$\sigma^*_{2,\mathbf{r}_i,\mathbf{r}_j}(\Re^d) = (2\pi)^{2d}n^{-1}\mathrm{Cov}\{f_{\mathbf{r}_i}(\mathbf{x}_1), f_{\mathbf{r}_j}(\mathbf{x}_1)\}, \tag{41}$$

$$|n\sigma^*_{2,\mathbf{r}_i,\mathbf{r}_j}(\mathbf{A})| \le 2Q^+_{3,1,\mathbf{r}_i}(\mathbf{A})Q^+_{3,1,\mathbf{r}_j}(\mathbf{A}), \quad \mathbf{A} = R'(\mathbf{T}) \text{ or } \Re^d. \tag{42}$$

Equations (41) and (42) and an application of the Cauchy–Schwarz inequality yield

$$\begin{aligned}&|n\sigma^*_{2,\mathbf{r}_i,\mathbf{r}_j}(R(\mathbf{T})) - (2\pi)^{2d}\mathrm{Cov}\{f_{\mathbf{r}_i}(\mathbf{x}_1), f_{\mathbf{r}_j}(\mathbf{x}_1)\}|\\ &\le C\left\{\prod_{k\in\{i,j\}} Q^+_{3,1,\mathbf{r}_k}(R'(\mathbf{T})) + Q^+_{3,1,\mathbf{r}_i}(R'(\mathbf{T}))Q^+_{3,1,\mathbf{r}_j}(\Re^d)\right.\\ &\qquad\left.\times Q^+_{3,1,\mathbf{r}_j}(R'(\mathbf{T}))Q^+_{3,1,\mathbf{r}_i}(\Re^d)\right\}.\end{aligned} \tag{43}$$

Now, the condition that $\min_{1\le j\le d}T_j\to\infty$ as $n\to\infty$ implies $I[\mathbf{t}\in R(\mathbf{T})]|\mathbf{t}^\mathbf{r}||\phi_f(\mathbf{t})|\to|\mathbf{t}^\mathbf{r}||\phi_f(\mathbf{t})|$ at every $\mathbf{t}$. Since $Q^+_{3,1,\mathbf{r}}(\Re^d) < \infty$, we get from the Dominated Convergence Theorem that $Q^+_{3,1,\mathbf{r}}(R(\mathbf{T})) = \int_{\Re^d} I[\mathbf{t}\in R(\mathbf{T})]|\mathbf{t}^\mathbf{r}||\phi_f(\mathbf{t})|d\mathbf{t} \to Q^+_{3,1,\mathbf{r}}(\Re^d)$ and, consequently, $Q^+_{3,1,\mathbf{r}}(R'(\mathbf{T})) = o(1)$. This, together with (38), (40) and (43), leads to (15) immediately. Next, (14) can be derived with trivial modifications (setting $\mathbf{r}_i = \mathbf{r}$ for all $i = 1,\ldots,d^{2m}$) to all the foregoing arguments. This completes the proof. □

Proof of Lemma 2 The proof extends that of Lemma 2 of Wu and Tsai (2004) from $\mathbf{r} = \mathbf{0}$ to general $\mathbf{r}$. First , we prove (17). Throughout, $\sum_\mathbf{r}$ is shorthand for $\sum_{\mathbf{r}:|\mathbf{r}|=2m}$. Define $\tilde{\varphi}^*(\mathbf{T}) = \sum_\mathbf{r}(-1)^m\tilde{\psi}^+_\mathbf{r}(\mathbf{T})$ (see (7) and Remark 1). From (8), (11) and the fact that $\mathrm{SCV}^\infty_{2m}(\hat{\mathbf{T}}_{2m}) \le \mathrm{SCV}^\infty_{2m}(\mathbf{T})$ for all $\mathbf{T} > \mathbf{0}$, we get

$$\sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \hat{\mathbf{T}}_{2m}^{\mathbf{r}+\mathbf{1}} - \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \mathbf{T}^{\mathbf{r}+\mathbf{1}} \leq 2^{-1}(n+1)(\pi)^d (\tilde{\varphi}^*(\hat{\mathbf{T}}_{2m}) - \tilde{\varphi}^*(\mathbf{T})) \tag{44}$$

for all $\mathbf{T} > \mathbf{0}$. Pick any fix any ε from interval $(2^{-1/(4m+2)}, 1)$ and set

$$u_n^{2p_0} = 2n\varepsilon M_0(2m+1)\{(2p_0 - 2m - 1)(2\varepsilon^{2(2m+1)} - 1)\}^{-1} \tag{45}$$

where $\sup_{\mathbf{t}}\{\prod_{i=1}^d |t_i|^{p_i} |\phi(\mathbf{t})|\} \leq M_0^{1/2}$ for all non-negative $\mathbf{p}$ satisfying $|\mathbf{p}| = p_0$. In what follows, we suppress the subscript $2m$ in $\hat{\mathbf{T}}_{2m} = (\hat{T}_{2m,1}, \ldots, \hat{T}_{2m,d})$ and denote $\tilde{\mathbf{r}} = (r_2, \ldots, r_d)$, $\hat{\mathbf{T}}_{-1} = (\hat{T}_2, \hat{T}_3, \ldots, \hat{T}_d)$ (similarly, $\mathbf{T}_{-1}$, etc.), $P_{\hat{\mathbf{t}}_{-1}}[\cdot] = P[\cdot|\hat{\mathbf{T}}_{-1} = \hat{\mathbf{t}}_{-1}]$ and $J_k = R(u_n\varepsilon^{-(k+1)}, \hat{\mathbf{t}}_{-1}) \setminus R(u_n, \hat{\mathbf{t}}_{-1}),\ k = 0, 1, \ldots$. By (44),

$$\begin{aligned}
&P_{\hat{\mathbf{t}}_{-1}}[\hat{T}_1 > (1-\varepsilon)^{-1} u_n] \\
&\quad = P_{\hat{\mathbf{t}}_{-1}}\left[\bigcap_{\mathbf{r}:|\mathbf{r}|=2m} \left\{(\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \hat{T}_1^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} > (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} (1-\varepsilon)^{-1(r_1+1)} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}}\right\}\right] \\
&\quad \leq P_{\hat{\mathbf{t}}_{-1}}\left[\sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \hat{T}_1^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} > (1-\varepsilon)^{-1} \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}}\right] \\
&\quad \leq \sum_{k=0}^{\infty} P_{\hat{\mathbf{t}}_{-1}}\Bigg[2^{-1}(n+1)(\pi)^d (\tilde{\varphi}^*(\hat{\mathbf{T}}) - \tilde{\varphi}^*(u_n, \hat{\mathbf{t}}_{-1})) > \varepsilon \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \hat{T}_1^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}}, \\
&\qquad\qquad u_n\varepsilon^{-k} < \hat{T}_1 \leq u_n \varepsilon^{-(k+1)}\Bigg] \\
&\quad \leq \sum_{k=0}^{\infty} P_{\hat{\mathbf{t}}_{-1}}\Bigg[2^{-(d+1)}(n+1) \sum_{\mathbf{r}} W_{3,\mathbf{r}}^{+,0}(J_k) > \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} \varepsilon^{(r_1+1)(1-k)} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \\
&\qquad\qquad - q_k^*(\hat{\mathbf{t}}_{-1})\Bigg]
\end{aligned} \tag{46}$$

where $q_k^*(\hat{\mathbf{t}}_{-1}) = 2^{-(d+1)}(n+1) \sum_{\mathbf{r}} \int_{J_k} |\mathbf{t}^{\mathbf{r}}| \mathrm{E}|\tilde{\phi}(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t}$. By Lemma 3 and (45),

$$\begin{aligned}
q_k^*(\hat{\mathbf{t}}_{-1}) &= (n+1)(2n)^{-1} \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} (\varepsilon^{-(k+1)(r_1+1)} - 1) u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \\
&\qquad + 2^{-(d+1)} n^{-1}(n^2-1) \sum_{\mathbf{r}} \int_{J_k} |\mathbf{t}^{\mathbf{r}}| |\phi_f(\mathbf{t})|^2 \,\mathrm{d}\mathbf{t} \\
&\leq \frac{n+1}{2n} \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \left\{\varepsilon^{-(k+1)(r_1+1)} + \frac{(n-1)M_0(r_1+1)u_n^{-2p_0}}{2p_0 - r_1 - 1}\right\} \\
&\leq \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \varepsilon^{(1-k)(r_1+1)} \left\{\frac{n+1}{2n}\left\{\varepsilon^{-2(r_1+1)} + \frac{2\varepsilon^{2(r_1+1)} - 1}{2\varepsilon^{2(r_1+1)}}\right\}\right\}
\end{aligned}$$

$$\le \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \varepsilon^{(1-k)(r_1+1)} \{(6\varepsilon^{2(r_1+1)}+1)(8\varepsilon^{2(r_1+1)})^{-1}\}$$
$$\le C_\varepsilon \sum_{\mathbf{r}} (\mathbf{r}+\mathbf{1})^{-\mathbf{1}} u_n^{r_1+1} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \varepsilon^{(1-k)(r_1+1)} \tag{47}$$

for all $n \ge 2 + 4(2\varepsilon^{2(r_1+1)} - 1)^{-1}$, where $C_\varepsilon = \max_{\mathbf{r}} (6\varepsilon^{2(r_1+1)}+1)(8\varepsilon^{2(r_1+1)})^{-1} < 1$; and the second inequality follows from $u_n^{-2p_0} \le (2p_0 - r_1 - 1)(2\varepsilon^{2(2m+1)} - 1)\{2n\varepsilon M_0(r_1+1)\}^{-1}$, $\varepsilon^{2(2m+1)} \le \varepsilon^{2(r_1+1)}$ and $\varepsilon^{(k+1)(r_1+1)} \le \varepsilon$. Now, by $\psi_{\mathbf{0}} < \infty$ and arguments similar to (39),

$$Q^+_{1,\mathbf{r},\mathbf{r}}(J_k) \le C\psi_{\mathbf{0}} u_n^{2r_1+1} \hat{\mathbf{t}}_{-1}^{2\tilde{\mathbf{r}}+\mathbf{1}} \varepsilon^{-(k+1)(2r_1+1)} \tag{48}$$

(here and below, C does not depend on k either). By Cauchy–Schwarz inequality, $Q^+_{3,1,\mathbf{r}}(J_k) \le C\{u_n^{r_1+1}\, \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}} \varepsilon^{-(k+1)(r_1+1)}\}^{1/2}\{Q^+_{3,2,\mathbf{r}}(J_k)\}^{1/2}$ and, consequently,

$$Q^+_{2,\mathbf{r},\mathbf{r}}(J_k) \le \int_{J_k} |\mathbf{t}_2^{\mathbf{r}_2}||\phi_f(-\mathbf{t}_2)|\{Q^+_{3,2,\mathbf{r}}(J_k)\}^{1/2} \left\{\int_{J_k} |\mathbf{t}_1^{\mathbf{r}_1}||\phi_f(\mathbf{t}_1+\mathbf{t}_2)|^2\, d\mathbf{t}_1\right\}^{1/2} d\mathbf{t}_2$$
$$\le \{Q^+_{3,2,\mathbf{r}}(J_k)\}^{1/2} \left\{u_n^{r_1}\varepsilon^{-(k+1)r_1}\hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}}\right\}^{1/2} \psi_{\mathbf{0}}^{1/2} Q^+_{3,1,\mathbf{r}}(J_k)$$
$$\le C\psi_{\mathbf{0}}^{1/2}\{u_n^{2r_1+1}\hat{\mathbf{t}}_{-1}^{2\tilde{\mathbf{r}}+\mathbf{1}}\varepsilon^{-(k+1)(2r_1+1)}\}^{1/2} Q^+_{3,2,\mathbf{r}}(J_k). \tag{49}$$

It follows from the fact $Q^+_{3,2,\mathbf{r}}(J_k) \le M_0 \int_{J_k} |t_1^{r_1-2p_0}||\mathbf{t}_{-1}^{\tilde{\mathbf{r}}}|\, d\mathbf{t} \le C u_n^{-(2p_0-r_1-1)} \hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}}$, (48), (49), Lemma 5 and an application of (generalized) c_r-inequality that

$$\mathrm{Var}\left\{\sum_{\mathbf{r}} W^{+,0}_{3,\mathbf{r}}(J_k)\right\} \le C_{d,m} \sum_{\mathbf{r}} \mathrm{Var}\left\{W^{+,0}_{3,\mathbf{r}}(J_k)\right\}$$
$$\le C_{d,m} \sum_{\mathbf{r}} C\left\{n^{-2}u_n^{2r_1+1}\hat{\mathbf{t}}_{-1}^{2\tilde{\mathbf{r}}+\mathbf{1}}\varepsilon^{-(k+1)(2r_1+1)}\right.$$
$$\left. + n^{-1}u_n^{2r_1+(3/2)-2p_0}\hat{\mathbf{t}}_{-1}^{2\tilde{\mathbf{r}}+(\mathbf{3/2})}\varepsilon^{-(k+1)(2r_1+1)/2}\right\} \tag{50}$$

where $C_{d,m}$ is the cardinality of $\{\mathbf{r} : |\mathbf{r}| = 2m\}$. Applying Chebyshev's inequality ($P[|X| > \sum_{i=1}^s c_i] \le \mathrm{E}X^2/c_j^2$ for any $1 \le j \le s$ if all c_i's are positive) to (46) and using (47) and (50),

$$P_{\hat{\mathbf{t}}_{-1}}[\hat{T}_1 > (1-\epsilon)^{-1}u_n]$$
$$\le \frac{(n+1)^2 C_{d,m}}{2^{2(d+1)}(1-C_\varepsilon)^2} \sum_{k=0}^{\infty}\sum_{\mathbf{r}} \left\{\mathrm{Var}\{W^{+,0}_{3,\mathbf{r}}(J_k)\}/[(\mathbf{r}+\mathbf{1})^{-\mathbf{1}}u_n^{r_1+1}\hat{\mathbf{t}}_{-1}^{\tilde{\mathbf{r}}+\mathbf{1}}\varepsilon^{(1-k)(r_1+1)}]^2\right\}$$
$$\le C(n+1)^2\left\{n^{-2}u_n^{-1}\hat{\mathbf{t}}_{-1}^{-\mathbf{1}}\sum_{\mathbf{r}}(\mathbf{r}+\mathbf{1})^{\mathbf{2}}\sum_{k=0}^{\infty}\varepsilon^{k-4r_1-3}\right.$$

$$+n^{-1}u_n^{-1/2-2p_0}\hat{\mathbf{t}}_{-1}^{-\mathbf{1/2}}\sum_{\mathbf{r}}(\mathbf{r}+\mathbf{1})^{\mathbf{2}}\sum_{k=0}^{\infty}\varepsilon^{[(k-3)(2r_1+3)+4]/2}\Bigg\}$$
$$\le C\{u_n^{-1}\hat{\mathbf{t}}_{-1}^{-\mathbf{1}}+u_n^{-1/2}\hat{\mathbf{t}}_{-1}^{-\mathbf{1/2}}\}. \tag{51}$$

Let $a_n = \mathrm{E}\big\{\big(1 - I[\hat{\mathbf{T}}_{-1}^{\mathbf{1}} > u_n^{-1/2}]\big)P_{\hat{\mathbf{T}}_{-1}}[\hat{T}_1 > (1-\varepsilon)^{-1}u_n]\big\}$. Rewrite (11) as $\mathrm{SCV}_{2m}^{\infty}(\mathbf{T}) = \sum_{\mathbf{r}}\int_{R(\mathbf{T})}|\mathbf{t}^{\mathbf{r}}|\big\{\{2/(n+1) - |\phi_f(\mathbf{t})|^2\} + \{|\phi_f(\mathbf{t})|^2 - |\tilde{\phi}(\mathbf{t})|^2\}\big\}\,d\mathbf{t}$. By the facts $\phi_f(\mathbf{t})$ is continuous, $\phi_f(\mathbf{0}) = 1$ and for any fixed $\mathbf{T} > \mathbf{0}$ it holds that $\sup_{\mathbf{t}\in R(\mathbf{T})}\big||\tilde{\phi}(\mathbf{t})|^2 - |\phi_f(\mathbf{t})|^2\big| \stackrel{a.s.}{\to} 0$ as $n\to\infty$ (see Theorem 2.1 of Csörgő 1981), we get for some $\mathbf{T}_0 > \mathbf{0}$, $\hat{\mathbf{T}} > \mathbf{T}_0$ a.s. as $n\to\infty$. Therefore, $I[\hat{\mathbf{T}}_{-1}^{\mathbf{1}} > u_n^{-1/2}] \stackrel{a.s.}{\to} 1$ as $n\to\infty$ and, consequently, $a_n = o(1)$ by the Bounded Convergence Theorem. It follows from (51) that

$$P[\hat{T}_1 > (1-\varepsilon)^{-1}u_n] = a_n + \mathrm{E}\left\{I\left[\hat{\mathbf{T}}_{-1}^{\mathbf{1}} > u_{n1}^{-1/2}\right]P_{\hat{\mathbf{T}}_{-1}}\left[\hat{T}_1 > (1-\varepsilon)^{-1}u_n\right]\right\}$$
$$= o(1) + O(n^{-1/(8p_0)}) = o(1). \tag{52}$$

Similarly, we can show that $P[\hat{T}_j > (1-\varepsilon)^{-1}u_n] = o(1)$ for all $2 \le j \le d$, and so (17) holds. Next, we prove, (18). Put $g_{2m}^*(\mathbf{T}) = \sum_{\mathbf{r}} g_{\mathbf{r}}^+(\mathbf{T})$ and let $\mathbf{T}_{2m}^0 = (T_{2m,1}^0,\ldots,T_{2m,d}^0)$ denote the minimizer of the score function $\mathrm{S}_{2m}^*(\mathbf{T})$ resulting from replacing $|\tilde{\phi}(\mathbf{t})|^2$ by $|\phi_f(\mathbf{t})|^2$ in (11) (the subscript $2m$ will be suppressed in what follows). Evidently, (18) is proved if we show $g^*(\mathbf{T}) = O(z_n)$ where $z_n = n^{-1+\{(2m+d)/2p_0\}}$. By the equation $\partial\mathrm{S}^*(\mathbf{T}^0)/\partial T_1 = 0$,

$$2\sum_{\mathbf{r}}\{(\mathbf{T}_{-1}^0)^{\tilde{\mathbf{r}}+\mathbf{1}}(n+1)^{-1}(\tilde{\mathbf{r}}+\mathbf{1})^{-\mathbf{1}}\}$$
$$= 2^{-d}\sum_{\mathbf{r}}\int_{R(\mathbf{T}_{-1}^0)}|\mathbf{t}_{-1}^{\tilde{\mathbf{r}}}|\{|\phi_f(T_1^0,\mathbf{t}_{-1})|^2 + |\phi_f(-T_1^0,\mathbf{t}_{-1})|^2\}\,d\mathbf{t}_{-1}$$
$$\le M_0(T_1^0)^{-2p_0}\sum_{\mathbf{r}}\{(\mathbf{T}_{-1}^0)^{\tilde{\mathbf{r}}+\mathbf{1}}(\tilde{\mathbf{r}}+\mathbf{1})^{-\mathbf{1}}\}. \tag{53}$$

It follows that $T_1^0 \le b_n$ where $b_n = \{M_0 n\}^{1/2p_0}$. Similarly, we get $T_j^0 \le b_n$ for all $2 \le j \le d$. This, together with the inequality $2^{d+1}(n+1)^{-1}\sum_{\mathbf{r}}(\mathbf{r}+\mathbf{1})^{-\mathbf{1}}\{\mathbf{b}_n^{\mathbf{r}+\mathbf{1}} - (\mathbf{T}^0)^{\mathbf{r}+\mathbf{1}}\} \ge g^*(\mathbf{T}^0) - g^*(\mathbf{b}_n)$ where $\mathbf{b}_n = (b_n, b_n,\ldots,b_n,b_n)$, implies $g^*(\mathbf{T}^0) = O(z_n)$. Put $\Gamma_n = (\gamma_n,\gamma_n,\ldots,\gamma_n,\gamma_n)$ where $\gamma_n = \gamma n^{1/2p_0}$ and $\gamma \gg$ (much larger than) $\max\{b_n,(1-\varepsilon)^{-1}u_n\}n^{-1/2p_0}$. Then $\mathbf{T}^0 \le \Gamma_n$ and $P[\hat{\mathbf{T}} \le \Gamma_n] \to 1$ by (52). Therefore, eventually in n, $\hat{\mathbf{T}}$ is also the minimizer of $\widetilde{\mathrm{SCV}}_{2m}^{\infty}(\mathbf{T}) = 2^{d+1}(n+1)^{-1}\sum_{\mathbf{r}}(\mathbf{r}+\mathbf{1})^{-\mathbf{1}}\mathbf{T}^{\mathbf{r}+\mathbf{1}} + \tilde{g}^*(\mathbf{T})$ over $\mathbf{0} < \mathbf{T} \le \Gamma_n$, where $\tilde{g}^*(\mathbf{T}) = \sum_{\mathbf{r}}\int_{R(\Gamma_n)\setminus R(\mathbf{T})}|\mathbf{t}^{\mathbf{r}}||\tilde{\phi}(\mathbf{t})|^2\,d\mathbf{t}$. By Lemma 3 and the fact $g^*(\mathbf{T}^0) = O(z_n)$, we get $\tilde{g}^*(\mathbf{T}^0) = O_p(z_n)$ and, consequently,

$$\tilde{g}^*(\hat{\mathbf{T}}) \le 2^{d+1}(n+1)^{-1}\sum_{\mathbf{r}}(\mathbf{r}+\mathbf{1})^{-\mathbf{1}}\big\{(\mathbf{T}^0)^{\mathbf{r}+\mathbf{1}} - \hat{\mathbf{T}}^{\mathbf{r}+\mathbf{1}}\big\} + \tilde{g}^*(\mathbf{T}^0) = O_p(z_n). \tag{54}$$

Define $g_j^*(\mathbf{T}) = \sum_{\mathbf{r}} \int_{\{\mathbf{t}:|t_j|>T_j\}} |\mathbf{t}^{\mathbf{r}}||\phi_f(\mathbf{t})|^2\,d\mathbf{t}$. Then $g_j^*(\mathbf{T}) \le g^*(\mathbf{T}) \le \sum_{j=1}^d g_j^*(\mathbf{T})$ and therefore $g_j^*(\mathbf{T}^0) = O(z_n)$ for all j. Since $g_j^*(\hat{\mathbf{T}})I[\hat{T}_j \ge T_j^0] \le g_j^*(\mathbf{T}^0)$, (18) will be proved if we prove that $L_{1j}L_{2j} = O_p(z_n)$ for all j, where $L_{1j} = I[\hat{T}_j < T_j^0]$ and $L_{2j} = g_j^*(\hat{\mathbf{T}}) - g_j^*(\mathbf{T}^0)$. It can be seen that $L_{1j}L_{2j} = L_{1j}\sum_{\mathbf{r}} Q_{3,2,\mathbf{r}}^+(R(\mathbf{T}^0) \setminus R(\hat{\mathbf{T}}_j^0)) + \hat{\Delta}_j^*$ where $\hat{\mathbf{T}}_j^0$ denotes the vector resulting from replacing the j-th component T_j^0 by $\hat{T}_j$ in $\mathbf{T}^0$ and $0 \le \hat{\Delta}_j^* \le C\sum_{i\ne j} g_i^*(\mathbf{T}^0) = O(z_n)$. By Lemma 3 and Markov inequality, $W_{1,\mathbf{r}}^+(R(\mathbf{T}^0)) = O_p(z_n)$. This, together with (7), (17) and Cauchy–Schwarz inequality, implies

$$
\begin{aligned}
L_{1j}L_{2j} &= L_{1j}\left\{-\sum_{\mathbf{r}} W_{3,\mathbf{r}}^+(R_{\mathbf{T}^0,\hat{\mathbf{T}}_j^0}) + (2\pi)^d(\tilde{\varphi}^*(\mathbf{T}^0) - \tilde{\varphi}^*(\hat{\mathbf{T}}_j^0))\right\} + \hat{\Delta}_j^* \\
&\le L_{1j}\left\{\sum_{\mathbf{r}} W_{1,\mathbf{r}}^+(R(\mathbf{T}^0)) + 2\sum_{\mathbf{r}}\left|W_{2,\mathbf{r}}^+(R_{\mathbf{T}^0,\hat{\mathbf{T}}_j^0})\right| + \tilde{g}^*(\hat{\mathbf{T}})\right\} + \hat{\Delta}_j^* \\
&\le L_{1j}\left\{\sum_{\mathbf{r}} W_{1,\mathbf{r}}^+(R(\mathbf{T}^0)) + \tilde{g}^*(\hat{\mathbf{T}})\right. \\
&\qquad \left. +2\sum_{\mathbf{r}}\left\{Q_{3,2,\mathbf{r}}^+(R_{\mathbf{T}^0,\hat{\mathbf{T}}_j^0})W_{1,\mathbf{r}}^+(R_{\mathbf{T}^0,\hat{\mathbf{T}}_j^0})\right\}^{1/2}\right\} + \hat{\Delta}_j^* \\
&\le O_p(z_n) + O_p(z_n^{1/2})L_{2j}^{1/2}L_{1j} \qquad (55)
\end{aligned}
$$

where $R_{\mathbf{T}^0,\hat{\mathbf{T}}_j^0} = R(\mathbf{T}^0) \setminus R(\hat{\mathbf{T}}_j^0)$. Put $L_{3j} = L_{1j}I[L_{2j} > z_n]$. Multiplying (55) by $L_{2j}^{-1/2}L_{3j}$ yields $L_{2j}^{1/2}L_{3j} \le O_p(z_n^{1/2})$. Hence $L_{2j}L_{3j}^2 = L_{2j}L_{1j}I[L_{2j} > z_n] = O_p(z_n)$. The proof of (18) follows by noting that $L_{2j}L_{1j}I[L_{2j} \le z_n] \le z_n$. Finally, (16) can be derived with trivial modifications (e.g., replacing SCV_{2m}^∞ by $\mathrm{CV}_{\mathbf{r}}^\infty$ and $\hat{\mathbf{T}}_{2m}$ by $\hat{\mathbf{T}}_{\mathbf{r}}$; and dropping $\sum_{\mathbf{r}}$ and $\bigcap_{\mathbf{r}:|\mathbf{r}|=2m}$) to all the foregoing arguments. This completes the proof. □

Proof of Theorem 1 The proof follows from that of Theorem 2 below with trivial modifications (replacing $\hat{\mathbf{T}}_{2m}$ by $\hat{\mathbf{T}}_{\mathbf{r}}$ for a fixed $\mathbf{r}$). □

Proof of Theorem 2 By (54) we get, for all $p_0 > m+d/2$, $\{\tilde{\varphi}^*(\Gamma_n) - \tilde{\varphi}^*(\hat{\mathbf{T}}_{2m})\}I[\hat{\mathbf{T}}_{2m} \le \Gamma_n] = O_p(z_n)$ and, consequently, $\{\tilde{\psi}_{\mathbf{r}}(\Gamma_n) - \tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{2m})\}I[\hat{\mathbf{T}}_{2m} \le \Gamma_n] = O_p(z_n)$ for all $\mathbf{r}$. This, together with the fact $P[\hat{\mathbf{T}}_{2m} \le \Gamma_n] \to 1$ (as established below (53)), leads to $\tilde{\psi}_{\mathbf{r}}(\Gamma_n) - \tilde{\psi}_{\mathbf{r}}(\hat{\mathbf{T}}_{2m}) = O_p(z_n)$ for all $p_0 > m + d/2$. Hence the theorem will be proved if we show that assertions (i) and (ii) remain true if $\hat{\mathbf{T}}_{2m}$ is replaced by Γ_n throughout. Now, by (37),

$$
(-1)^m(2\pi)^d(\tilde{\psi}_{\mathbf{r}}(\Gamma_n) - \psi_{\mathbf{r}}) = W_{1,\mathbf{r}}(R(\Gamma_n)) + 2W_{2,\mathbf{r}}(R(\Gamma_n)) - g_{\mathbf{r}}(\Gamma_n). \qquad (56)
$$

By Lemma 3 and routine computation, $W_{1,\mathbf{r}}(R(\Gamma_n)) = O_p(z_n)$, $g_{\mathbf{r}}(\Gamma_n) = O_p(z_n)$ and $\mathrm{E}|W_{2,\mathbf{r}}(R(\Gamma_n))| \le O(n^{-1/2})Q_{3,1,\mathbf{r}}^+(R(\Gamma_n))$ for all $p_0 > m + d/2$. Further,

$Q^+_{3,1,\mathbf{r}}(R(\Gamma_n)) \le Cn^{(2m+d-p_0)/(2p_0)}$ if $p_0 < 2m+d$ and $Q^+_{3,1,\mathbf{r}}(R(\Gamma_n)) \le C\log^d n$ if $p_0 = 2m+d$. Therefore, assertion (i) remains true with $\hat{\mathbf{T}}_{2m}$ being replaced by Γ_n. For the rest we assume $p_0 > 2m+d$. Now, $|W_{1,\mathbf{r}}(R(\Gamma_n))| = o_p(n^{-1/2})$, $|g_{\mathbf{r}}(\Gamma_n)| = o_p(n^{-1/2})$ and $|W_{2,\mathbf{r}}(R'(\Gamma_n))| = o_p(n^{-1/2})$ (by (42)). Thus, by (56), the vectors $2n^{1/2}(W_{2,\mathbf{r}_1}(\Re^d), \ldots, W_{2,\mathbf{r}_{d^{2m}}}(\Re^d))$ and $n^{1/2}(-1)^m(2\pi)^d(\tilde{\boldsymbol{\psi}}_{2m}(\Gamma_n) - \boldsymbol{\psi}_{2m})$ have the same asymptotic distribution. In view of (41) and by the Cramér–Wold device, it suffices to show the k-th order cumulant of any linear combination $2n^{1/2}\sum_{i=1}^{d^{2m}} a_i W_{2,\mathbf{r}_i}(\Re^d)$, where a_i's are constants, converges to zero for all $k \ge 3$. But this can be seen by noting that such k-th order cumulant is equal to

$$n^{k/2}\sum_{v_1,\ldots,v_{d^{2m}}} C_{k,v_1,\ldots,v_{d^{2m}}}\prod_{i=1}^{d^{2m}} a_i^{v_i}\int_{\Re^d}\cdots\int_{\Re^d}\prod_{i=1}^{d^{2m}}\prod_{j=S_{i-1}+1}^{S_i}\mathbf{t}_j^{\mathbf{r}_i}$$
$$\times\prod_{j=1}^{k}\phi_f(-\mathbf{t}_j)\mathrm{cum}(\tilde{\phi}_0(\mathbf{t}_1), \ldots, \tilde{\phi}_0(\mathbf{t}_k))\,\mathrm{d}\mathbf{t}_1\cdots\mathrm{d}\mathbf{t}_k = O(n^{-(k/2)+1}) \quad (57)$$

where $S_0 = 0$, $S_i = \sum_{j=1}^{i} v_j$, $C_{k,v_1,\ldots,v_{d^{2m}}} > 0$ is a constant not depending on n, the summation is taken over the region $\{(v_1, \ldots, v_{d^{2m}}) : 0 \le v_i \le k,\ i = 1, \ldots, d$ and $\sum_{i=1}^{d^{2m}} v_i = k\}$ and the last equality is from (30). This completes the proof. □

Proof of Theorem 3 The proof straightforwardly extends that of Theorem 2(i) of Bickel and Ritov (1988) from $d = 1$ to general d. Let $\{f^{(v)}, v \ge 1\}$ be a sequence of densities from $\mathcal{F}_d$. We denote $s^{(v)} = (f^{(v)})^{1/2}$ and $s = f^{1/2}$. In order to prove the theorem it suffices to show $\psi_{\mathbf{r}}$ is (Frèchet) pathwise differentiable along paths satisfying

$$\|s^{(v)} - s\|_2 \to 0 \ \text{ and } \ \|(f_{\mathbf{r}}^{(v)} - f_{\mathbf{r}})s\|_2 \to 0, \quad v \to \infty \quad (58)$$

with derivative $4\{f_{\mathbf{r}} - \psi_{\mathbf{r}}\}f^{1/2}$ because then, as at the end of the proof of Theorem 2(i) of Bickel and Ritov (1988), the information bound for $\psi_{\mathbf{r}}$ is $\|2\{f_{\mathbf{r}} - \psi_{\mathbf{r}}\}f^{1/2}\|_2^2 = 4\mathrm{Var}\{f_{\mathbf{r}}(\mathbf{x}_1)\}$, where the last equality can be seen from (1). Write, with some abuse of notation, $\psi_{\mathbf{r}}(s) = \psi_{\mathbf{r}}(f)(=\psi_{\mathbf{r}})$ and $\psi_{\mathbf{r}}(s^{(v)}) = \psi_{\mathbf{r}}(f^{(v)})$. Using integration by parts, we get $\int f_{\mathbf{r}}(\mathbf{x})f^{(v)}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int f(\mathbf{x})f_{\mathbf{r}}^{(v)}(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and, therefore, $\psi_{\mathbf{r}}(s^{(v)}) - \psi_{\mathbf{r}}(s) = 2V_1 + V_2$ with $V_1 = \int f_{\mathbf{r}}(\mathbf{x})\{f^{(v)}(\mathbf{x}) - f(\mathbf{x})\}\,\mathrm{d}\mathbf{x}$ and $V_2 = \int\{f^{(v)}(\mathbf{x}) - f(\mathbf{x})\}\{f_{\mathbf{r}}^{(v)}(\mathbf{x}) - f_{\mathbf{r}}(\mathbf{x})\}\,\mathrm{d}\mathbf{x}$. By (1), (58) and Remark 3,

$$\begin{aligned}
V_2 &= \int\{s^{(v)}(\mathbf{x}) - s(\mathbf{x})\}^2\{f_{\mathbf{r}}^{(v)}(\mathbf{x}) - f_{\mathbf{r}}(\mathbf{x})\}\,\mathrm{d}\mathbf{x} \\
&\quad + 2\int s(\mathbf{x})\{s^{(v)}(\mathbf{x}) - s(\mathbf{x})\}\{f_{\mathbf{r}}^{(v)}(\mathbf{x}) - f_{\mathbf{r}}(\mathbf{x})\}\,\mathrm{d}\mathbf{x} \\
&\le \|s^{(v)} - s\|_2^2 \sup_{\mathbf{x}\in\Re^d}|f_{\mathbf{r}}^{(v)}(\mathbf{x}) + f_{\mathbf{r}}(\mathbf{x})| + 2\|s^{(v)} - s\|_2\,\|(f_{\mathbf{r}}^{(v)} - f_{\mathbf{r}})s\|_2 \\
&= o(\|s^{(v)} - s\|_2) \qquad (59)
\end{aligned}$$

and

$$\begin{aligned} V_1 &= \int f_{\mathbf{r}}(\mathbf{x}) f^{(v)}(\mathbf{x})\, \mathrm{d}\mathbf{x} - \psi_{\mathbf{r}}(s) = \int \{f_{\mathbf{r}}(\mathbf{x}) - \psi_{\mathbf{r}}(s)\}\, f^{(v)}(\mathbf{x})\, \mathrm{d}\mathbf{x} \\ &= \int \{f_{\mathbf{r}}(\mathbf{x}) - \psi_{\mathbf{r}}(s)\}\, \{f^{(v)}(\mathbf{x}) - f(\mathbf{x})\}\, \mathrm{d}\mathbf{x} \\ &= \int \{f_{\mathbf{r}}(\mathbf{x}) - \psi_{\mathbf{r}}(s)\}\, 2s(\mathbf{x})\{s^{(v)}(\mathbf{x}) - s(\mathbf{x})\}\, \mathrm{d}\mathbf{x} \\ &\quad + \int \{f_{\mathbf{r}}(\mathbf{x}) - \psi_{\mathbf{r}}(s)\}\, \{s^{(v)}(\mathbf{x}) - s(\mathbf{x})\}^2\, \mathrm{d}\mathbf{x} \\ &= \langle 2\{f_{\mathbf{r}} - \psi_{\mathbf{r}}\}s,\ s^{(v)} - s\rangle + O(\|s^{(v)} - s\|_2^2) \end{aligned} \tag{60}$$

where $\langle\cdot,\cdot\rangle$ denotes the usual inner product. It follows that as $v \to \infty$,

$$\left|\psi_{\mathbf{r}}(s^{(v)}) - \psi_{\mathbf{r}}(s) - \langle 4\{f_{\mathbf{r}} - \psi_{\mathbf{r}}\}s,\ s^{(v)} - s\rangle\right| = o(\|s^{(v)} - s\|_2).$$

This leads to the desired (Frèchet) derivative of $\psi_{\mathbf{r}}$, and the proof is completed. □

Acknowledgments The authors are grateful to an Associate Editor and two referees for many valuable suggestions which significantly improved the quality and presentation of the paper. In particular, the critical suggestions by the referees of including the study of estimating the vector functionals (3), at both the theoretical- and practical-levels, led the authors to reconstruct the modified procedure of Sect. 2.3 that significantly improved its first version.

References

Aldershof, B. (1991). *Estimation of integrated squared density derivatives*. Mimeo Series, No. 2053, Institute of Statistics, University of North Carolina.

Bellman, R. E. (1961). *Adaptive control processes*. Princeton, NJ: Princeton University Press.

Bickel, P. J., Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Series A*, *50*, 381–393.

Cao, R., Francisco-Fernandez, M., Anand, A., Bastida, F., Gonzalez-Andujar, J. L. (2011). Computing statistical indices for hydrothermal times using weed emergence data. *Journal of Agricultural Science*, *149*, 701–712.

Chacón, J. E. (2009). Data-driven choice of the smoothing parametrization for kernel density estimators. *Canadian Journal of Statistics*, *37*, 249–265.

Chacón, J. E., Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test*, *19*, 375–398.

Chacón, J. E., Duong, T. (2011). Unconstrained pilot selectors for smoothed cross validation. *Australian & New Zealand Journal of Statistics*, *53*, 331–351.

Chacón, J. E., Duong, T. (2013). Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics*, *7*, 499–532.

Chacón, J. E., Tenreiro, C. (2012). Exact and asymptotically optimal bandwidths for kernel estimation of density functionals. *Methodology and Computing in Applied Probability*, *14*, 523–548.

Chacón, J. E., Duong, T., Wand, M. P. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica*, *21*, 807–840.

Cheng, M.-Y. (1997). Boundary aware estimators of integrated density derivative products. *Journal of the Royal Statistical Society: Series B*, *59*, 191–203.

Chiu, S. T. (1991). Bandwidth selection for kernel density estimation. *Annals of Statistics*, *19*, 1883–1905.

Chiu, S. T. (1992). An automatic bandwidth selector for kernel density estimation. *Biometrika*, *79*, 771–782.

Csörgő, S. (1981). Multivariate empirical characteristic functions. *Zeitschrift för Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *55*, 203–229.
Davis, K. B. (1975). Mean square error properties of density estimates. *Annals of Statistics*, *3*, 1025–1030.
Davis, K. B. (1977). Mean integrated square error properties of density estimates. *Annals of Statistics*, *5*, 530–535.
Delaigle, A., Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *Journal of the Royal Statistical Society: Series B*, *64*, 869–886.
Devroye, L. (1988). Asymptotic performance bounds for the kernel estimates. *Annals of Statistics*, *16*, 1162–1179.
Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Annals of Statistics*, *20*, 2037–2056.
Donoho, D. L., Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity*, *6*, 290–323.
Duong, T., Hazelton, M. L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*, *15*, 17–30.
Duong, T., Hazelton, M. L. (2005a). Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimation. *Journal of Multivariate Analysis*, *93*, 417–433.
Duong, T., Hazelton, M. L. (2005b). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, *32*, 485–506.
Efromovich, S., Low, M. (1996a). On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Annals of Statistics*, *24*, 682–686.
Efromovich, S., Low, M. (1996b). On optimal adaptive estimation of a quadratic functional. *Annals of Statistics*, *24*, 1106–1125.
Fan, J. (1991). On the estimation of quadratic functionals. *Annals of Statistics*, *19*, 1273–1294.
Fan, J., Marron, J. S. (1992). Best possible constant for bandwidth selection. *Annals of Statistics*, *20*, 2057–2070.
Fisher, N. I., Mammen, E., Marron, J. S. (1994). Testing for multimodality. *Computational Statistics & Data Analysis*, *18*, 499–512.
Giné, E., Mason, D. M. (2008). Uniform in bandwidth estimation of integral functionals of the density function. *Scandinavian Journal of Statistics*, *35*, 739–761.
Goldstein, L., Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Annals of Statistics*, *20*, 1306–1328.
Hall, P., Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters*, *6*, 109–115.
Hall, P., Marron, J. S. (1988). Choice of kernel order in density estimation. *Annals of Statistics*, *16*, 161–173.
Hall, P., Marron, J. S. (1991a). Lower bounds for bandwidth selection in density estimation. *Probability Theory and Related Fields*, *90*, 149–173.
Hall, P., Marron, J. S. (1991b). Local minima in cross-validation functions. *Journal of the Royal Statistical Society: Series B*, *53*, 245–252.
Hewitt, E., Stromberg, K. (1969). *Real and abstract analysis*. New York: Springer.
Ibragimov, I. A., Khas'minskii, R. Z. (1982). Estimation of distribution density belonging to a class of entire functions. *Theory of Probability and Its Applications*, *27*, 551–562.
Janssen, P., Marron, J. S., Veraverbeke, N., Sarle, W. (1995). Scale measures for bandwidth selection. *Journal of Nonparametric Statistics*, *5*, 359–380.
Jones, M. C., Sheather, S. J. (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics & Probability Letters*, *11*, 511–514.
Koshevnik, Y. A., Levit, B. Y. (1976). On a non-parametric analogue of the information matrix. *Theory of Probability and Its Applications*, *21*, 738–753.
Kotz, S., Balakrishnan, N., Johnson, N. L. (2000). *Continuous multivariate distributions: V. 1: models and applications* (2nd ed.). New York: Wiley.
Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli*, *3*, 181–211.
Martinez, H. V., Olivares, M. M. (1999). Estimation of quadratic functionals of a density. *Statistics & Probability Letters*, *42*, 327–332.
Panaretos, V. M., Konis, K. (2012). Nonparametric construction of multivariate kernels. *Journal of the American Statistical Association*, *107*, 1085–1095.
Park, B. U., Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, *85*, 66–78.

Puri, M. L., Sen, P. K. (1971). *Nonparametric methods in multivariate Analysis*. New York: Wiley.

Rektorys, K. (1969). *Survey of applicable mathematics*. Cambridge, MA: M.I.T. Press.

Sain, S. R., Baggerly, K. A., Scott, D. W. (1994). Cross-Validation of Multivariate Densities. *Journal of the American Statistical Association*, *89*, 807–817.

Scott, D. W. (1992). *Multivariate dendity estimation: theory, practice, and visualization*. New York: John Wiley.

Sheather, S. J., Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B*, *53*, 683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Annals of Statistics*, *12*, 1285–1297.

Wand, M. P. (1992). Error Analysis for general multivariate kernel estimators. *Journal of Nonparametric Statistics*, *2*, 1–15.

Wand, M. P., Jones, M. C. (1993). Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, *88*, 520–528.

Wand, M. P., Jones, M. C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, *9*, 97–166.

Wu, T.-J. (1995). Adaptive root n estimates of integrated squared density derivatives. *Annals of Statistics*, *23*, 1474–1495.

Wu, T.-J. (1997). Root n bandwidth selectors for kernel estimation of density derivatives. *Journal of the American Statistical Association*, *92*, 536–547.

Wu, T.-J., Tsai, M. H. (2004). Root n bandwidth selectors in multivariate kernel density estimation. *Probability Theory and Related Fields*, *129*, 537–558.