Average derivative estimation under measurement error*

Hao Dong†
Department of Economics
Southern Methodist University

Taisuke Otsu‡
Department of Economics
London School of Economics

Luke Taylor§
Department of Economics and Business Economics
Aarhus University

September 2, 2020

Abstract

In this paper, we derive the asymptotic properties of the density-weighted average derivative estimator when a regressor is contaminated with classical measurement error and the density of this error must be estimated. Average derivatives of conditional mean functions are used extensively in economics and statistics, most notably in semiparametric index models. As well as ordinary smooth measurement error, we provide results for supersmooth error distributions. This is a particularly important class of error distribution as it includes the Gaussian density. We show that under either type of measurement error, despite using nonparametric deconvolution techniques and an estimated error characteristic function, we are able to achieve a $\sqrt{n}$-rate of convergence for the average derivative estimator. Interestingly, if the measurement error density is symmetric, the asymptotic variance of the average derivative estimator is the same irrespective of whether the error density is estimated or not. The promising finite sample performance of the estimator is shown through a Monte Carlo simulation.

*Financial support from the ERC Consolidator Grant (SNP 615882) (Otsu) and the AUFF Starting Grant (26852) (Taylor) is gratefully acknowledged .

†E-mail: haod@smu.edu. Address: Department of Economics, Southern Methodist University, 3300 Dyer Street, Dallas, TX 75275, US.

‡E-mail: t.otsu@lse.ac.uk. Address: Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

§E-mail: lntaylor@econ.au.dk. Address: Department of Economics and Business Economics, Fuglesangs Allé 4 building 2631, 12 8210 Aarhus V, Denmark.

1 Introduction

Average derivatives of conditional mean functions are used extensively in economics and statistics, most notably in semiparametric index models. This paper studies asymptotic properties of the density-weighted average derivative estimator when a regressor is contaminated with classical measurement error and the density of this error must be estimated. As well as ordinary smooth measurement error, we provide results for supersmooth error distributions, which cover the popular Gaussian density. We show that under this ill-posed inverse problem, despite using nonparametric deconvolution techniques and an estimated error characteristic function, we are able to achieve a $\sqrt{n}$-rate of convergence for the average derivative estimator. Interestingly, if the measurement error density is symmetric, the asymptotic variance of the average derivative estimator is the same irrespective of whether the error density is estimated or not.

Since the seminal work of Powell, Stock and Stoker (1989), average derivatives have enjoyed much popularity. They have found primary use in estimating coefficients in single-index models, where Powell, Stock and Stoker (1989) showed that average derivatives identify the parameters of interest up-to-scale. A key benefit of average derivative estimators is their ability to achieve a $\sqrt{n}$-rate of convergence despite being constructed using nonparametric techniques. They have also been employed to great effect in the estimation of consumer demand functions (see, for example, Blundell, Duncan and Pendakur, 1998, and Yatchew, 2003) and sample selection models (for example, Das, Newey and Vella, 2003). Additionally, several testing procedures have made use of these estimators (see, for example, Härdle, Hildenbrand and Jerison, 1991, and Racine, 1997). See also Li and Racine (2007, Sec. 8.3) and Horowitz (2009, Sec. 2.6) for a review, and Cattaneo, Crump and Jansson (2010, 2014) for recent developments on small bandwidth asymptotics in this setting.

Although the literature on average derivative estimation is rich in econometrics and statistics, we emphasize that the majority of papers focus on the case where the regressors are correctly measured. A notable exception is Fan (1995), which extends the $\sqrt{n}$-consistency result to allow for regressors contaminated with classical measurement error from the class of ordinary smooth distributions, for example, gamma or Laplace. In Fan (1995), it was shown that average derivative estimators, constructed using deconvolution techniques, were able to retain the $\sqrt{n}$-rate of convergence enjoyed by their correctly measured counterparts. However, this result relied on knowledge of the true error distribution and did not cover the case of supersmooth error densities, which includes Gaussian error. Our major contributions are to extend Fan's (1995) result to the case where the measurement error density is known to be supersmooth and to the case where the measurement error is unknown and of either smoothness type.

Extending these results to supersmooth measurement error is not a trivial extension, and it is not clear a priori whether a $\sqrt{n}$-rate can be achieved in this case. Indeed, in many estimation and testing problems, convergence rates and asymptotic distributions are fundamentally different between ordinary smooth and supersmooth error densities (see, for example, Fan, 1991, van Es and Uh, 2005, Dong and Otsu, 2018, and Otsu and Taylor, 2020). Furthermore, no result has been provided regarding the asymptotic properties of average derivative estimators in the more realistic situation where the measurement error density is unknown. Much recent work in the errors-in-variables literature has been aimed at relaxing the assumption of a known measurement error distribution, and deriving the asymptotic properties of estimators and test statistics in this setting (see, for example, Delaigle, Hall and Meister, 2008, Dattner, Reiß and Trabs, 2016, and Kato and Sasaki, 2018).

This paper also contributes to the literature on measurement error problems. In contrast to measurement error analysis in parametric or nonparametric models (see, Carroll *et al.*, 2006, and Meister, 2009, for a survey), the literature on measurement error analysis for semiparametric models is relatively thin. You and Chen (2006) and Liang *et al.* (2008) studied semiparametric varying coefficients and partially linear models with mismeasured covariates, respectively. See also a survey by Chen, Hong and Nekipelov (2011) for a survey on measurement error problems in nonlinear models.

It should be emphasized that although this paper deals with a theoretical problem, we believe the results presented are useful in many areas of applied work. Binary choice models are ubiquitous in economic research and settings where a regressor of interest is measured with (potentially) classical measurement error are not uncommon. For example, Dong and Lewbel (2015) estimate a semiparametric binary choice model for an individual's migration decision where income is the variable of interest - which is generally considered to suffer from measurement error. In a similar manner to Schennach (2004), using a measure of income from a recent year as a repeated noisy measurement would allow the use of the average derivative estimator of this paper.

In semiparametric models more generally, measurement error is rife in empirical work. For example, Schennach and Hu (2013) applied measurement error techniques in a semiparametric framework to study the relationship between investment behavior and market value. Indeed, Erickson and Whited (2000) showed that the reason Tobin's Q theory of investment (Brainard and Tobin, 1968) could not be empirically substantiated was due to classical measurement error in the market value variable, and, once this was accounted for, the theoretical predictions matched the empirical findings. In another interesting application of semiparametric methods, Mamuneas, Savvides and Stengos (2006) analyze the effect of labor market share on economic growth. The labor market share of national income is widely believed to suffer from measurement error; this is seen by the Bureau of Labor Statistics and Bureau of Economic Analysis each publishing different measures of the wage share in the US. Conveniently, this provides researchers with two repeated measurements of labor market share which can be used to correct the measurement error issue, as in the estimator of this paper.

In response to the slow convergence rates achieved by nonparametric deconvolution techniques (Fan, 1991, and Fan and Truong, 1993), practitioners may be tempted to shy away from the use of these estimators in the face of classical measurement error. However, by showing that a parametric rate of convergence can still be obtained even in the worst-case scenario of supersmooth error and an estimated error characteristic function, we hope to encourage greater use of nonparametric estimation in applied work when covariates are contaminated. Moreover, since the curse of dimensionality (which plagues all nonparametric estimators) is exacerbated in the presence of measurement error, the potential gain from using average derivatives is increased when regressors are mismeasured. In particular, in the case of ordinary smooth error densities, the convergence rate of deconvolution estimators, although slower than standard nonparametric estimators, remains polynomial. However, for supersmooth densities, this convergence typically deteriorates to a $\log(n)$ rate.

In the next section, we describe the setup of our model, discuss the assumptions imposed, and provide our main result. Section 3 provides details of a Monte Carlo simulation and Section 4 concludes. All mathematical proofs are relegated to the Appendix.

2 Main result

2.1 Setup and estimator

Consider the nonparametric errors-in-variables model

$$\begin{aligned} Y &= g(X^*) + u, \qquad E[u|X^*] = 0, \\ X &= X^* + \epsilon, \end{aligned} \tag{1}$$

where Y is a scalar dependent variable, X^* is an unobservable error-free scalar covariate, X is an observable covariate, u is a regression error term, and ϵ is a measurement error on the covariate. Suppose the density function f of X^* and the regression function g are continuously differentiable. We are interested in estimating the density-weighted average derivative

$$\theta = E[g'(X^*)f(X^*)] = -2E[Yf'(X^*)], \tag{2}$$

where g' and f' are the first-order derivatives of g and f, respectively. The second equality follows from integration by parts (see Lemma 2.1 of Powell, Stock and Stoker, 1989).

The key use of such density-weighted average derivatives is in single-index models and partially linear single-index models. Taking $g(X) = g(X_1'\beta, X_2)$ for some unknown link function g with $X = (X_1, X_2)$, we obtain the partially linear case; when X_2 is removed, this becomes the single-index model. Such specifications are very general and cover a wide variety of regression models, e.g., binary choice models, truncated and censored dependent variable models, and duration models (see Ichimura, 1993, for a more detailed discussion). They can also be used as a simple dimension reduction solution to the curse of dimensionality.

For identification purposes, it is necessary to make some normalization restriction on β, since any scaling factor can be subsumed into g. Hence, the parameter β is only identified up to scale. Due to the linear index structure, the density-weighted average derivative identifies this scaled β.

If we directly observe X^*, θ can be estimated by the sample analog $-\frac{2}{n}\sum_{j=1}^n Y_j \tilde{f}'(X_j^*)$, where $\tilde{f}'$ is a nonparametric estimator of the derivative f'. However, if X^* is unobservable, this estimator is infeasible. Whereas, when the density function f_ϵ of the measurement error ϵ is *known* (and ordinary smooth), Fan (1995) suggested estimating θ by evaluating the joint density $h(x, y)$ of (X^*, Y) and the derivative $f'(x)$ in the expression

$$\theta = -2 \iint y f'(x) h(x, y) dx dy, \tag{3}$$

by applying the deconvolution method. Let $\mathrm{i} = \sqrt{-1}$ and f^{ft} be the Fourier transform of a function f. If f_ϵ is known, based on the i.i.d. sample $\{Y_j, X_j\}_{j=1}^n$ of (Y, X), the densities f and h can be estimated by

$$\tilde{f}(x) = \frac{1}{nb_n}\sum_{j=1}^n \mathbb{K}\left(\frac{x - X_j}{b_n}\right), \qquad \tilde{h}(x, y) = \frac{1}{nb_n^2}\sum_{j=1}^n \mathbb{K}\left(\frac{x - X_j}{b_n}\right) K_y\left(\frac{y - Y_j}{b_n}\right),$$

respectively, where b_n is a bandwidth, K_y is an (ordinary) kernel function and $\mathbb{K}$ is a deconvolution kernel function defined as

$$\mathbb{K}(x) = \frac{1}{2\pi}\int e^{-\mathrm{i}tx}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b_n)}dt.$$

By plugging these estimators into (3), Fan (1995) proposed an estimator of θ and studied its asymptotic properties (again, when f_ϵ^{ft} is known and ordinary smooth).

In this paper, we extend Fan's (1995) result to the cases where (i) f_ϵ is unknown and symmetric around zero but repeated measurements on X^* are available, and (ii) f_ϵ is known and supersmooth. Since the second result is obtained as a by-product of the first one, we hereafter focus on the first case.

Suppose we have two independent noisy measurements of the error-free variable X^*, i.e.,

$$X_j = X_j^* + \epsilon_j \quad \text{and} \quad X_j^r = X_j^* + \epsilon_j^r,$$

for $j = 1, \ldots, n$. Under the assumption that f_ϵ is symmetric, its Fourier transform f_ϵ^{ft} can be estimated by

$$\hat{f}_\epsilon^{\text{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^{n} \cos\{t(X_j - X_j^r)\} \right|^{1/2}, \tag{4}$$

(Delaigle, Hall and Meister, 2008). By plugging in this estimator, the densities f and h can be estimated by

$$\hat{f}(x) = \frac{1}{nb_n} \sum_{j=1}^{n} \hat{\mathbb{K}} \left(\frac{x - X_j}{b_n} \right), \qquad \hat{h}(x,y) = \frac{1}{nb_n^2} \sum_{j=1}^{n} \hat{\mathbb{K}} \left(\frac{x - X_j}{b_n} \right) K_y \left(\frac{y - Y_j}{b_n} \right),$$

where

$$\hat{\mathbb{K}}(x) = \frac{1}{2\pi} \int e^{-\mathrm{i}tx} \frac{K^{\text{ft}}(t)}{\hat{f}_\epsilon^{\text{ft}}(t/b_n)} dt.$$

The parameter θ can then be estimated by

$$\begin{aligned} \hat{\theta} &= -2 \int y \hat{f}'(x) \hat{h}(x,y) dx dy \\ &= -\frac{2}{n^2 b_n^3} \sum_{j=1}^{n} \sum_{k=1}^{n} Y_k \int \hat{\mathbb{K}}' \left(\frac{x - X_j}{b_n} \right) \hat{\mathbb{K}} \left(\frac{x - X_k}{b_n} \right) dx, \end{aligned} \tag{5}$$

where $\hat{f}'$ and $\hat{\mathbb{K}}'$ are the first-order derivatives of $\hat{f}$ and $\hat{\mathbb{K}}'$, respectively, and the second equality follows from $\int y K_y((y - Y_k)/b_n) dy = b_n Y_k$. Here we have derived the estimator for the case of a continuous Y. However, our estimator $\hat{\theta}$ in (5) can be applied to the case of a discrete Y as well.[1,2]

Throughout this paper, we focus on the case of a single covariate to keep the notation simple. The proposed method, however, can easily adapt to the multivariate case. In particular, when there are multiple covariates and D_x of them are mismeasured, i.e.,

$$Y = g(X^*, Z) + u,$$

where $X^* = (X_1^*, \ldots, X_{D_x}^*)$ is a vector of D_x unobserved covariates and $Z = (Z_1, \ldots, Z_{D_z})$ is a vector of D_z observed covariates, the parameters of interest are

$$\begin{aligned} \theta_{x,d_1} &= E\left[\left.\frac{\partial g(x,z)}{\partial x_{d_1}}\right|_{(X^*,Z)} f_{X^*,Z}(X^*,Z) \right] = -2E\left[Y \left.\frac{\partial f_{X^*,Z}(x,z)}{\partial x_{d_1}}\right|_{(X^*,Z)} \right], \\ \theta_{z,d_2} &= E\left[\left.\frac{\partial g(x,z)}{\partial z_{d_2}}\right|_{(X^*,Z)} f_{X^*,Z}(X^*,Z) \right] = -2E\left[Y \left.\frac{\partial f_{X^*,Z}(x,z)}{\partial z_{d_2}}\right|_{(X^*,Z)} \right], \end{aligned}$$

for $d_1 = 1, \ldots, D_x$ and $d_2 = 1, \ldots, D_z$, and can be written as

$$\begin{aligned} \theta_{x,d_1} &= -2 \iint y \frac{\partial f_{X^*,Z}(x,z)}{\partial x_{d_1}} h(x,y,z) dx dy dz, \\ \theta_{z,d_2} &= -2 \iint y \frac{\partial f_{X^*,Z}(x,z)}{\partial z_{d_2}} h(x,y,z) dx dy dz, \end{aligned}$$

for the joint densities $f_{X^*,Z}$ and h of (X^*, Z) and (X^*, Y, Z), respectively. Instead of $X_{d_1}^*$, we observe two noisy measurements

$$X_{d_1} = X_{d_1}^* + \epsilon_{d_1} \quad \text{and} \quad X_{d_1}^r = X_{d_1}^* + \epsilon_{d_1}^r,$$

where ϵ_{d_1} and $\epsilon_{d_1}^r$ are two independent measurement errors. If ϵ_{d_1} and $\epsilon_{d_1}^r$ are identically distributed and $f_{\epsilon_{d_1}}$ is symmetric, $f_{\epsilon_{d_1}}^{\text{ft}}$ can be estimated by

$$\hat{f}_{\epsilon_{d_1}}^{\text{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^{n} \cos\{t(X_{d_1,j} - X_{d_1,j}^r)\} \right|^{1/2}.$$

[1]Our deconvolution approach can be extended to more general weighted averages, say $\theta_w = E[g'(X^*)w(X^*)]$, where $w(\cdot)$ is a continuously differentiable known weight function (with $w(\cdot)/f(\cdot)$ bounded). In this case, integration by parts yields

$$\theta_w = -\int y\{w'(x) + w(x)f'(x)/f(x)\}h(x,y)dxdy,$$

and θ_w can be estimated by the sample analog $\hat{\theta}_w = -\int y\{w'(x) + w(x)\hat{f}'(x)/\hat{f}(x)\}\hat{h}(x,y)dxdy$. Although a detailed analysis will be more cumbersome, analogous results to the density-weighted case can be derived.

[2]Although it is beyond the scope of this paper, it would be interesting to extend our approach to the case where f_ϵ is possibly asymmetric. In this case, we could construct an estimator for θ by estimating f_ϵ^{ft} using the methods in Li and Vuong (1998) or Comte and Kappus (2015). These estimators take more complicated forms than (4), and technical arguments will be substantially different from ours.

Let $K_z : \mathbb{R}^{D_z} \to \mathbb{R}$ be an ordinary kernel function for correctly measured covariates Z, and $\hat{\mathbb{K}}_x : \mathbb{R}^{D_x} \to \mathbb{R}$ be a deconvolution kernel function for mismeasured covariates X^*. To simplify our analysis, we use a product deconvolution kernel as in Fan and Masry (1992). In particular, we assume that ϵ is mutually independent so that $f_\epsilon^{\rm ft}(t) = \prod_{d_1=1}^{D_x} f_{\epsilon_{d_1}}^{\rm ft}(t_{d_1})$. Let $K : \mathbb{R} \to \mathbb{R}$ be an ordinary univariate kernel. Then the product deconvolution kernel is defined as $\hat{\mathbb{K}}_x(x) = \prod_{d_1=1}^{D_x} \hat{\mathbb{K}}_{d_1}(x_{d_1})$, where

$$\hat{\mathbb{K}}_{d_1}(x_{d_1}) = \frac{1}{2\pi} \int e^{-\mathrm{i}tx_{d_1}} \frac{K^{\rm ft}(t)}{f_{\epsilon_{d_1}}^{\rm ft}(t/b_n)} dt.$$

Then, $f_{X^*,Z}$ and h can be estimated by

$$\begin{aligned} \tilde{f}_{X^*,Z}(x,z) &= \frac{1}{nb_n^{D_x+D_z}} \sum_{j=1}^n \hat{\mathbb{K}}_x\left(\frac{x-X_j}{b_n}\right) K_z\left(\frac{z-Z_j}{b_n}\right), \\ \tilde{h}(x,y,z) &= \frac{1}{nb_n^{D_x+D_z+1}} \sum_{j=1}^n \hat{\mathbb{K}}_x\left(\frac{x-X_j}{b_n}\right) K_z\left(\frac{z-Z_j}{b_n}\right) K_y\left(\frac{y-Y_j}{b_n}\right), \end{aligned}$$

and θ_{x,d_1} and θ_{z,d_2} can be estimated by

$$\begin{aligned} \hat{\theta}_{x,d_1} &= -\frac{2}{n^2 b_n^{2D_x+2D_z+1}} \sum_{j=1}^n \sum_{k=1}^n Y_k \iint \frac{\partial \hat{\mathbb{K}}_x\left(\frac{x-X_j}{b_n}\right)}{\partial x_{d_1}} \hat{\mathbb{K}}_x\left(\frac{x-X_k}{b_n}\right) K_z\left(\frac{z-Z_j}{b_n}\right) K_z\left(\frac{z-Z_k}{b_n}\right) dxdz, \\ \hat{\theta}_{z,d_2} &= -\frac{2}{n^2 b_n^{2D_x+2D_z+1}} \sum_{j=1}^n \sum_{k=1}^n Y_k \iint \hat{\mathbb{K}}_x\left(\frac{x-X_j}{b_n}\right) \hat{\mathbb{K}}_x\left(\frac{x-X_k}{b_n}\right) \frac{\partial K_z\left(\frac{z-Z_j}{b_n}\right)}{\partial z_{d_2}} K_z\left(\frac{z-Z_k}{b_n}\right) dxdz. \end{aligned}$$

We expect that analogous results to our main theorem can be established for this estimator as well. See Footnote 7 in Appendix A for more details.

2.2 Asymptotic properties

We now investigate the asymptotic properties of the average derivative estimator $\hat{\theta}$ in (5). Let $G = gf$. For ordinary smooth measurement error densities, we impose the following assumptions.

Assumption OS.

(1) *$\{Y_j, X_j, X_j^r\}_{j=1}^n$ is an i.i.d. sample of (Y, X, X^r) satisfying (1). $g(\cdot) = E[Y|X^* = \cdot]$ has p continuous, bounded, and integrable derivatives. The density function $f(\cdot)$ of X^* has $(p+1)$ continuous, bounded, and integrable derivatives, where p is a positive integer satisfying $p > \alpha + 1$.*

(2) (ϵ, ϵ^r) *are mutually independent and independent of* (Y, X^*)*, the distributions of* ϵ *and* ϵ^r *are identical, absolutely continuous with respect to the Lebesgue measure, and the characteristic function* f_ϵ^{ft} *is of the form*

$$f_\epsilon^{\text{ft}}(t) = \frac{1}{\sum_{v=0}^{\alpha} C_v |t|^v} \quad \textit{for all } t \in \mathbb{R},$$

for some finite constants $C_0, \ldots, C_\alpha$ *with* $C_0 \neq 0$ *and a positive integer* α*.*

(3) K *is differentiable to order* $(\alpha + 1)$ *and satisfies*

$$\int K(x)dx = 1, \qquad \int x^p K(x)dx \neq 0, \qquad \int x^l K(x)dx = 0, \ \textit{for all } l = 1, \ldots, p-1.$$

Also K^{ft} *is compactly supported on* $[-1, 1]$*, symmetric around zero, and bounded.*

(4) $n^{-1/2} b_n^{-2(1+3\alpha)} \log(b_n^{-1})^{-1/2} \to 0$*, and* $n^{1/2} b_n^p \to 0$ *as* $n \to \infty$*.*

(5) $Var(r(X, Y)) < \infty$*, where*

$$r(x, y) = \sum_{v=0}^{\alpha} (-\mathrm{i})^v C_v \{ y f^{(v+1)}(x) - G^{(v+1)}(x) \},$$

for almost every (x, y)*.*

The i.i.d. restriction on the data from Assumption (1) is standard in the literature and is imposed merely for ease of derivation rather than necessity. The second part of this assumption requires sufficient smoothness from the regression function and density function of X relative to the smoothness of the measurement error. Assumption (2) is the conventional ordinary smooth assumption for the measurement error.[3] Assumption (3) requires a kernel function of order p to remove the bias term from the nonparametric estimator. The first part of Assumption (4) requires that the bandwidth does not decay to zero too quickly as $n \to \infty$. This is necessary to ensure the asymptotic linear representation using the Hájek projection; the particular rate depends on the parameters of the measurement error characteristic function. The second part of Assumption (4) ensures the bandwidth approaches zero sufficiently fast to remove the asymptotic bias from the nonparametric estimator. Finally, Assumption (5) is a high-level assumption on the boundedness of the asymptotic variance of the average derivative estimator.

For the supersmooth case, we impose the following assumptions.

[3] We note that the convergence rates of $\hat{f}$ and $\hat{h}$ are polynomial orders under Assumption OS (2) and additional regularity conditions (Delaigle, Hall and Meister, 2008), and these rates may be faster than $n^{-1/4}$ under certain cases with properly chosen bandwidths. However, under Assumption SS (2) below for the supersmooth case, $\hat{f}$ and $\hat{h}$ typically converge at logarithmic rates, which are always slower than $n^{-1/4}$.

Assumption SS.

(1) $\{Y_j, X_j, X_j^r\}_{j=1}^n$ *is an i.i.d. sample of* (Y, X, X^r) *satisfying (1).* $g(\cdot) = E[Y|X^* = \cdot]$ *and the Lebesgue density* $f(\cdot)$ *of* X^* *are infinitely differentiable, and all derivatives of* g *and* f *are bounded.*

(2) (ϵ, ϵ^r) *are mutually independent and independent of* (Y, X^*)*, the distributions of* ϵ *and* ϵ^r *are identical, absolutely continuous with respect to the Lebesgue measure, and the characteristic function* f_ϵ^{ft} *is of the form*

$$f_\epsilon^{\text{ft}}(t) = Ce^{-\mu|t|^\gamma} \quad \text{for all } t \in \mathbb{R},$$

for some positive constants C *and* μ*, and positive even integer* γ*.*

(3) K *is infinitely differentiable and satisfies*

$$\int K(x)dx = 1, \qquad \int x^l K(x)dx = 0, \ \text{for all } l \in \mathbb{N}.$$

Also K^{ft} *is compactly supported on* $[-1, 1]$*, symmetric around zero, and bounded.*

(4) $b_n \to 0$ *and* $n^{-1/2} b_n^{-2} e^{6\mu b_n^{-\gamma}} \log(b_n^{-1})^{-1/2} \to 0$ *as* $n \to \infty$*.*

(5) $Var(r(X, Y)) < \infty$*, where*

$$r(x, y) = \sum_{h=0}^{\infty} \frac{\mu^h}{\mathrm{i}^{h\gamma} C h!} \{y f^{(h\gamma+1)}(x) - G^{(h\gamma+1)}(x)\},$$

for almost every (x, y)*.*

Many of the same comments as for the ordinary smooth case apply to this setting. However, the second part of Assumption (1) is more restrictive and appears to be necessary. As discussed in Meister (2009), one can show that 'the class of infinitely differentiable functions still contains a comprehensive nonparametric class of densities' (p. 44), including, of course, Gaussian and mixtures of Gaussians. For the regression function, all polynomials satisfy this restriction, as well as circular functions, exponentials, and products or sums of such smooth functions. Nevertheless, we admit that infinite order differentiability on g and f is a major limitation; in this sense, our results illustrate how theoretically and empirically challenging it is to achieve $\sqrt{n}$-consistency in the supersmooth case. Assumption (2) is the conventional supersmooth assumption for the measurement error, with the non-standard additional constraint on γ being even. Although this rules out the Cauchy distribution (where $\gamma = 1$), importantly, this still contains the canonical

Gaussian distribution as well as Gaussian mixtures. van Es and Gugushvili (2008) imposed the same constraint. Assumption (3) requires an infinite-order kernel function; these are often required in supersmooth deconvolution problems.[4] Meister (2009) discusses their construction and notes that the commonly used sinc kernel, $K(x) = \frac{\sin(x)}{\pi x}$, satisfies the requirements. Assumption (4) requires the bandwidth to decay to zero at a logarithmic rate. In particular, because an infinite-order kernel is used, we can ignore concerns regarding the bias from the nonparametric estimator and choose a bandwidth of at least $b_n = O((\log n)^{-1/\gamma})$ to satisfy this assumption.

Based on these assumptions, our main result is stated as follows.

Theorem. *Suppose Assumption OS or SS holds true. Then*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 4Var(r(X, Y))).$$

The most important aspect of this result is the $\sqrt{n}$-convergence of the estimator. Before this result, Powell, Stock and Stoker (1989) showed the same rate of convergence in the case of correctly measured regressors, and Fan (1995) confirmed this result for ordinary smooth error in the regressors when the error distribution is known. The above theorem shows that the convergence rate of these average derivative estimators does not change when measurement error is introduced. In particular, it does not change in the severely ill-posed case of supersmooth error, nor does it change when the measurement error distribution is estimated. On the other hand, the asymptotic variance takes different forms for the ordinary and supersmooth cases.[5]

Interestingly, as outlined in the Appendix, the asymptotic variance depends on the symmetry of the measurement error density. When the measurement error is symmetric around zero, remainder terms associated with the estimation error of the measurement error characteristic function vanish, and the asymptotic variance is the same as if the measurement error distribution is known.[6]

As a by-product of the proof, we also establish the asymptotic distribution of Fan's (1995) estimator for θ when the distribution of ϵ is known and supersmooth.

[4] Although infinite-order kernels are less common outside supersmooth deconvolution problems, papers such as Devroye (1992) and Politis and Romano (1995) advocate the use of infinite-order kernels for nonparametric density estimation (without measurement error) and spectral density estimation, respectively.

[5] Also, for $\sqrt{n}$-consistency, the conditions on the bandwidth (i.e., Assumptions OS (4) and SS (4)) are not adaptive with respect to the unknown measurement error distribution. Although it is beyond the scope of this paper, it would be interesting to see whether adaptive estimation is possible in this setting (see, Butucea and Comte, 2009, for linear functional estimation when the measurement error density is known).

[6] More precisely, symmetry of the measurement error density is crucial to guarantee the last equality in (12) to show that the second term in (11) vanishes. However, it is an open question whether symmetry of the measurement error density is necessary for such a phenomenon regarding the asymptotic variance of $\hat{\theta}$. To answer this question would require extending our asymptotic analysis for average derivative estimators to the case where f_ϵ^{ft} is estimated by methods such as those in Li and Vuong (1998) or Comte and Kappus (2015) to allow for a possibly asymmetric f_ϵ.

Corollary. *Suppose Assumption SS holds true without the repeated measurement X^r. Then the estimator $\tilde{\theta}$, defined by replacing $\hat{\mathbb{K}}$ in (5) with $\mathbb{K}$, satisfies*

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N(0, 4Var(r(X, Y))).$$

2.3 Bandwidth choice and implementation

To construct our estimator, it is necessary to choose a bandwidth. To this end, we employ the approach of Bissantz *et al.* (2007). In that paper, they show that the dependence between the bandwidth and the L^∞ distance between a deconvolution kernel density estimate and the true density is quite different when the bandwidth is smaller than the optimal choice in comparison to when it is larger than the optimal choice. It appears that this phenomenon is also found in our setting. Moreover, other deconvolution estimation problems also appear to exhibit this pattern; for example, Kato and Sasaki (2018) find the same result in a regression context.

The bandwidth selection procedure involves two steps. First, select a pilot bandwidth, b_n^0, that is oversmoothing. The exact value is not critical providing that it is larger than the optimal choice; we use the plug-in bandwidth of Delaigle and Gijbels (2004) and, as suggested in Bissantz *et al.* (2007), we multiply this by two to ensure it is large enough. With this pilot bandwidth, create a grid of potential bandwidths $b_{n,j} = b_n^0(j/J)$ for $j = 1, \ldots, J$ and denote $\hat{\theta}(b_{n,j})$ as the estimator using the j-th bandwidth in this grid. In the second step, choose the largest bandwidth $b_{n,j}$ such that $d(b_{n,j}, b_{n,j-1}) = |\hat{\theta}(b_{n,j}) - \hat{\theta}(b_{n,j-1})|$ is larger than $\rho d_{J-1,J}$ for some $\rho > 1$. Following Kato and Sasaki (2018) we choose $J = 4 \log n$ and $\rho = 0.4 \log n$. Evidence of the suitability of this bandwidth selection procedure is provided in Section 3.

In summary, we suggest the following steps to construct our average derivative estimator.

1. Construct an estimator for the characteristic function of the measurement error, $f_\epsilon^{\text{ft}}(t)$, using
$$\hat{f}_\epsilon^{\text{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^{n} \cos\{t(X_j - X_j^r)\} \right|^{1/2}.$$

2. Choose an initial bandwidth parameter that is oversmoothing, for example, two times the plug-in bandwidth of Delaigle and Gijbels (2004). Create a grid of candidate bandwidths $b_{n,j} = b_n^0\{j/J\}$ for $j = 1, \ldots, J$ with $J = 4 \log n$.

3. For each $j = 1, \ldots, 4 \log n$, compute the average derivative estimator
$$\hat{\theta}(b_{n,j}) = -\frac{2}{n^2 b_{n,j}^3} \sum_{j=1}^{n} \sum_{k=1}^{n} Y_k \int \hat{\mathbb{K}}' \left(\frac{x - X_j}{b_n} \right) \hat{\mathbb{K}} \left(\frac{x - X_k}{b_n} \right) dx,$$

and choose as the final estimator $\hat{\theta}(b_{n,j})$ for which $b_{n,j}$ is the largest bandwidth such that $|\hat{\theta}(b_{n,j}) - \hat{\theta}(b_{n,j-1})| > 0.4 \log n$.

3 Simulation

In this section, we analyze the small-sample properties of our average derivative estimator across three different models. In each model, the dependent variable, Y, is generated as follows

$$
\begin{aligned}
Y &= \tau(\tilde{Y}), \\
\tilde{Y} &= \beta_1 X^* + \beta_2 X^{*2} + U,
\end{aligned}
$$

where U is drawn from $N(0,1)$ independently of X^* and $(\beta_1, \beta_2) = (1,1)$. In the 'linear model' we have $\tau(\tilde{Y}) = \tilde{Y}$. The 'Probit model' sets $\tau(\tilde{Y}) = \mathcal{I}(\tilde{Y} \geq 0)$, where $\mathcal{I}(A)$ is the indicator function for the event A. Finally, the 'Tobit model' has $\tau(\tilde{Y}) = \tilde{Y}\mathcal{I}(\tilde{Y} \geq 0)$. We draw X^* from $N(0,1)$ and assume it is unobservable. However, we observe two noisy measurements $X = X^* + \epsilon_1$ and $X^r = X^* + \epsilon_2$, where (ϵ_1, ϵ_2) are mutually independent and independent of (X^*, U). For the densities of (ϵ_1, ϵ_2), we consider two settings. In the ordinary smooth case, (ϵ_1, ϵ_2) have a zero mean Laplace distribution with variance of $1/3$. For the supersmooth case, (ϵ_1, ϵ_2) have a normal distribution with zero mean and variance of $1/3$.

We report results for two sample sizes, $n = \{250, 500\}$, and compare our results to the weighted average derivative estimator of Powell, Stock and Stoker (1989), PSS henceforth, and to the infeasible version of our estimator where the error characteristic function is known. Note that the PSS estimator is not designed to deal with the problem of measurement error.

The bandwidth for the PSS estimator is chosen using the method of Powell and Stoker (1996), and the infeasible estimator uses the same bandwidth selection procedure as the estimator with an unknown error characteristic function. Throughout this study, we employ the infinite-order flat-top kernel of McMurry and Politis (2004). This kernel is defined by its Fourier transform

$$
K^{\mathrm{ft}}(t) = \begin{cases} 1 & \text{if } |t| \leq 0.05, \\ \exp\left\{ \frac{-\exp(-1/(|t|-0.05)^2)}{(|t|-1)^2} \right\} & \text{if } 0.05 < |t| < 1, \\ 0 & \text{if } |t| \geq 1, \end{cases}
$$

and satisfies Assumptions OS (3) and SS (3). All results are based on 1000 Monte Carlo replications.

Figure 3.1 gives a graphical representation of the bandwidth selection procedure discussed in

Section 2.3 for the ordinary smooth case; analogous results for the supersmooth case are given in Figure 3.2. Panel (a) of Figure 3.1 plots the absolute error of the average derivative estimator as a function of the bandwidth, $b_{n,j}$. The dashed line marks the optimal bandwidth that minimizes the absolute error; for this sample, the optimal bandwidth is 0.16. Panel (b) of Figure 3.1 plots $d(b_{n,j}, b_{n,j-1})$ as a function of $b_{n,j}$. The dashed line marks the selected bandwidth (equal to 0.19). Taken together, these plots suggest that the approach of Bissantz *et al.* (2007) is well-suited to choosing the optimal bandwidth for the average derivative estimator.

In Tables 1-3, we report the absolute bias (Bias), the standard deviation over Monte Carlo replications (SD), and the mean squared error (MSE) for each of the three estimators of $\theta = E[g'(X^*)f(X^*)]$. Note that each measure of performance is multiplied by 100 to ease comparison. In the linear model $\theta = 0.282$, in the probit model $\theta = 0.057$, and in the Tobit model $\theta = 0.266$. The estimator from this paper is denoted AD, the estimator of this paper with a known error distribution is denoted ADK, and the estimator of Powell, Stock and Stoker (1989) is denoted PSS.

There are several features of the results worthy of discussion. In all settings, the MSE results under ordinary smooth contamination are better than for supersmooth measurement error. This reflects the more difficult problem of deconvolution with supersmooth error that has been found throughout the literature on measurement error. However, it is interesting to note that the difference in MSE between these two cases is driven by the larger bias in the supersmooth setting. This is likely caused by the larger bandwidth that is typically chosen when the measurement error has a supersmooth distribution, as can be seen for representative samples in Figures 3.1 and 3.2.

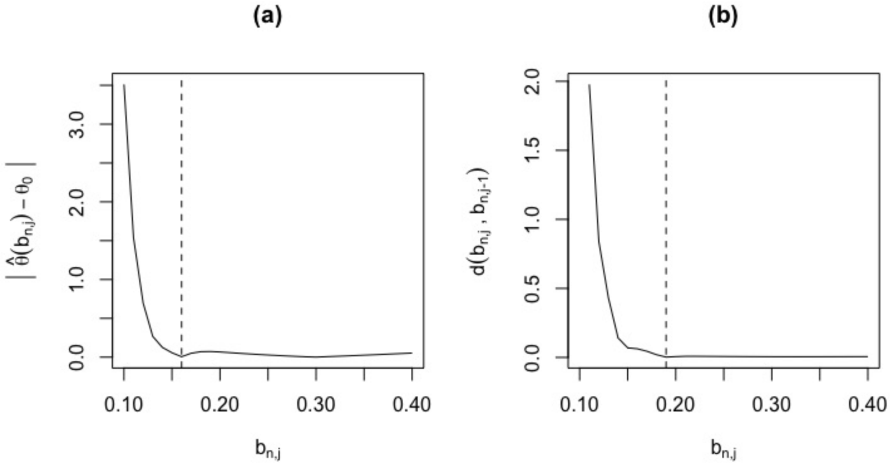

Figure 3.1: (a) Relationship between the bandwidth and the absolute error of the average derivative estimator for a representative sample from the linear model with Laplace error and sample size of 250. The dashed line indicates the optimal bandwidth for this sample, equal to 0.16. (b) Relationship between the bandwidth and the absolute distance between two estimators using consecutive bandwidths, i.e. $d(b_{n,j}, b_{n,j-1})$, for the linear model with Laplace error and sample size of 250. The dashed line indicates the selected bandwidth for this sample, equal to 0.19.

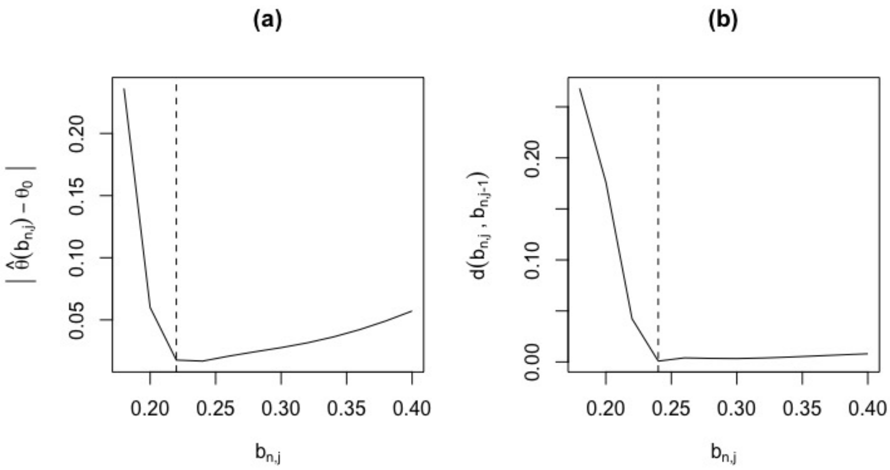

Figure 3.2: (a) Relationship between the bandwidth and the absolute error of the average derivative estimator for a representative sample from the linear model with Gaussian error and sample size of 250. The dashed line indicates the optimal bandwidth for this sample, equal to 0.22. (b) Relationship between the bandwidth and the absolute distance between two estimators using consecutive bandwidths, i.e. $d(b_{n,j}, b_{n,j-1})$, for the linear model with Gaussian error and sample size of 250. The dashed line indicates the selected bandwidth for this sample, equal to 0.24.

Table 1: Linear Model

$n = 250$	**Ordinary Smooth**			**Supersmooth**		
	Bias	**SD**	**MSE**	**Bias**	**SD**	**MSE**
AD	2.87	7.94	0.71	4.10	7.74	0.77
ADK	2.76	7.90	0.70	4.04	7.84	0.78
PSS	9.16	4.70	1.06	10.1	4.42	0.84
$n = 500$						
AD	0.38	5.79	0.34	0.82	6.36	0.41
ADK	0.31	5.86	0.34	0.87	6.23	0.40
PSS	9.25	3.43	0.97	10.32	3.19	1.17

Table 2: Probit model

$n = 250$	**Ordinary Smooth**			**Supersmooth**		
	Bias	**SD**	**MSE**	**Bias**	**SD**	**MSE**
AD	1.92	1.92	0.08	2.40	1.68	0.09
ADK	1.94	1.94	0.08	2.40	1.66	0.08
PSS	2.31	1.38	0.07	2.67	1.29	0.09
$n = 500$						
AD	0.99	1.69	0.04	1.38	1.61	0.04
ADK	0.98	1.70	0.04	1.39	1.57	0.04
PSS	2.40	1.02	0.07	2.64	1.01	0.08

Table 3: Tobit Model

$n = 250$	**Ordinary Smooth**			**Supersmooth**		
	Bias	**SD**	**MSE**	**Bias**	**SD**	**MSE**
AD	1.68	6.21	0.41	2.41	6.31	0.46
ADK	1.74	6.10	0.40	2.39	6.42	0.47
PSS	7.06	3.74	0.64	7.78	3.66	0.74
$n = 500$						
AD	0.20	4.76	0.23	0.42	5.24	0.28
ADK	0.22	4.68	0.22	0.44	5.29	0.28
PSS	7.12	2.74	0.58	7.90	2.60	0.69

Unsurprisingly, the bias and the standard deviation (and, consequently, the MSE) for our estimator and the same estimator with a known error distribution improve with the sample size. However, the bias for the estimator of PSS does not shrink with the sample size. Moreover, for all settings, this bias is substantially larger than that of the two estimators which account for measurement error. In the case of the linear model with ordinary smooth error and a sample size of 500, the bias from the PSS estimator is almost 25 times that of the estimator in this paper with an estimated error characteristic function. This highlights the importance of accounting for measurement error.

When comparing the standard deviations for each of these estimators, it is not so surprising that the PSS estimator has less variation; this reflects the differences between standard kernel estimation and deconvolution kernel estimation. However, across each of the specifications considered, this difference in standard error is not enough to allow the PSS estimator to dominate in terms of MSE.

As shown in our theoretical results in Section 2, the estimator using a known error characteristic function has the same asymptotic distribution as the estimator using an estimated error characteristic function. Interestingly, in a finite sample, there appears to be almost no difference in MSE between the two estimators even at relatively small sample sizes. This is encouraging for empirical applications which typically require estimation of the error characteristic function.

4 Conclusion

We derive the asymptotic properties of the density-weighted average derivative when a regressor is contaminated with classical measurement error and the density of this error must be estimated. Average derivatives of conditional mean functions are an important statistic in economics and have, consequently, received considerable attention in the previous literature. They are used most notably in semiparametric index models, for example, limited dependent variable models and duration models.

We characterize the asymptotic distribution of our average derivative estimator for both ordinary smooth and supersmooth measurement error. Moreover, we show that under either type of error, despite using nonparametric deconvolution techniques - which have notoriously slow convergence rates - and an estimated error characteristic function, we are able to achieve a $\sqrt{n}$-rate of convergence. Interestingly, if the measurement error density is symmetric, the asymptotic variance of the average derivative estimator is the same irrespective of whether the error density is estimated or not. Finally, we show that the finite sample performance of our estimator is encouraging. In particular, in comparison to the estimator of Powell, Stock and Stoker (1989) which estimates the average derivative when the regressors are perfectly measured, the deconvolution estimator dominates in terms of bias and MSE. Furthermore, we show that there is very little loss in performance of our estimator when the error characteristic function is estimated, mirroring the asymptotic results obtained.

A Proof of theorem (supersmooth case)

Since the arguments are similar, we first present a proof for the supersmooth case. Some lemmas for this proof are presented in Appendix B. In Appendix C, we provide a proof for the ordinary smooth case by explaining in detail the parts of the proof that differ to the supersmooth setting.

Let $\hat{\xi}(t) = \frac{1}{n}\sum_{l=1}^{n} \xi_l(t)$ for $\xi_l(t) = \cos(t(X_l - X_l^r))$, and $\xi(t) = |f_\epsilon^{\mathrm{ft}}(t)|^2$. Note that $\hat{f}_\epsilon^{\mathrm{ft}}(t) = |\hat{\xi}(t)|^{1/2}$ and $f_\epsilon(t) = |\xi(t)|^{1/2}$. By expansions around $\hat{\xi}(t/b_n) = \xi(t/b_n)$, we obtain

$$\begin{aligned}
\hat{\mathbb{K}}(x) &= \mathbb{K}(x) + A_1(x) + R_1(x), \\
\hat{\mathbb{K}}'(x) &= \mathbb{K}'(x) + A_2(x) + R_2(x),
\end{aligned}$$

where

$$\begin{aligned}
A_1(x) &= -\frac{1}{4\pi}\int e^{-\mathrm{i}tx} K^{\mathrm{ft}}(t) \left\{ \frac{\hat{\xi}(t/b_n) - \xi(t/b_n)}{|\xi(t/b_n)|^{3/2}} \right\} dt, \\
A_2(x) &= \frac{\mathrm{i}}{4\pi}\int e^{-\mathrm{i}tx} t K^{\mathrm{ft}}(t) \left\{ \frac{\hat{\xi}(t/b_n) - \xi(t/b_n)}{|\xi(t/b_n)|^{3/2}} \right\} dt, \\
R_1(x) &= -\frac{1}{4\pi}\int e^{-\mathrm{i}tx} K^{\mathrm{ft}}(t) \left\{ \frac{1}{|\tilde{\xi}(t/b_n)|^{1/2}} - \frac{1}{|\xi(t/b_n)|^{1/2}} \right\} \left\{ \frac{\hat{\xi}(t/b_n) - \xi(t/b_n)}{|\xi(t/b_n)|} \right\} dt \\
&\quad -\frac{1}{2\pi}\int e^{-\mathrm{i}tx} K^{\mathrm{ft}}(t) \left\{ \frac{1}{|\hat{\xi}(t/b_n)|^{1/2}} - \frac{1}{|\xi(t/b_n)|^{1/2}} \right\} \left\{ \frac{|\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2}}{|\xi(t/b_n)|^{1/2}} \right\} dt, \\
R_2(x) &= \frac{\mathrm{i}}{4\pi}\int e^{-\mathrm{i}tx} t K^{\mathrm{ft}}(t) \left\{ \frac{1}{|\tilde{\xi}(t/b_n)|^{1/2}} - \frac{1}{|\xi(t/b_n)|^{1/2}} \right\} \left\{ \frac{\hat{\xi}(t/b_n) - \xi(t/b_n)}{|\xi(t/b_n)|} \right\} dt \\
&\quad +\frac{\mathrm{i}}{2\pi}\int e^{-\mathrm{i}tx} t K^{\mathrm{ft}}(t) \left\{ \frac{1}{|\hat{\xi}(t/b_n)|^{1/2}} - \frac{1}{|\xi(t/b_n)|^{1/2}} \right\} \left\{ \frac{|\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2}}{|\xi(t/b_n)|^{1/2}} \right\} dt,
\end{aligned}$$

for some $\tilde{\xi}(t/b_n) \in (\hat{\xi}(t/b_n), \xi(t/b_n))$.

Here, $A_1(x)$ and $A_2(x)$ are the Fréchet derivatives of $\hat{\mathbb{K}}(x)$ and $\hat{\mathbb{K}}'(x)$ as functionals of $\hat{\xi}(t/b_n)$ at $\xi(t/b_n)$, which characterize the dominant components of the approximation errors of $\hat{\mathbb{K}}(x)$ and $\hat{\mathbb{K}}'(x)$ to $\mathbb{K}(x)$ and $\mathbb{K}'(x)$, respectively. Also $R_1(x)$ and $R_2(x)$ are the remainder terms, which are of higher order than $A_1(x)$ and $A_2(x)$.

Then, we can decompose [7]

[7] When there are multiple covariates and D_x of them are mismeasured, let $\hat{\xi}_{d_1}(t) = \frac{1}{n}\sum_{l=1}^{n} \xi_{d_1,l}(t)$ for $\xi_{d_1,l}(t) = \cos(t(X_{d_1,l} - X_{d_1,l}^r))$, and $\xi_{d_1}(t) = |f_{\epsilon_{d_1}}^{\mathrm{ft}}(t)|^2$ for $d_1 = 1, \ldots, D_x$. By similar arguments as in the univariate case, for $d_1 = 1, \ldots, D_x$, we obtain

$$\hat{\mathbb{K}}_{d_1}(x_{d_1}) = \mathbb{K}_{d_1}(x_{d_1}) + A_{d_1,1}(x_{d_1}) + R_{d_1,1}(x_{d_1}) \quad \text{and} \quad \hat{\mathbb{K}}'_{d_1}(x_{d_1}) = \mathbb{K}'_{d_1}(x_{d_1}) + A_{d_1,2}(x_{d_1}) + R_{d_1,2}(x_{d_1}),$$

where $A_{d_1,k}$ and $R_{d_1,k}$ are separately defined by replacing $\hat{\xi}$ and ξ by $\tilde{\xi}_{d_1}$ and ξ_{d_1} respectively, for some $\tilde{\xi}_{d_1} \in$

$$\hat{\theta} = -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int \hat{\mathbb{K}}'\left(\frac{x-X_j}{b_n}\right)\hat{\mathbb{K}}\left(\frac{x-X_k}{b_n}\right)dx = S + T_1 + \cdots + T_6, \quad (6)$$

where

$$\begin{aligned}
S &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx \\
&\quad -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)A_1\left(\frac{x-X_k}{b_n}\right)dx \\
&\quad -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int A_2\left(\frac{x-X_j}{b_n}\right)\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx, \\
T_1 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)R_1\left(\frac{x-X_k}{b_n}\right)dx, \\
T_2 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int R_2\left(\frac{x-X_j}{b_n}\right)\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx, \\
T_3 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int A_2\left(\frac{x-X_j}{b_n}\right)A_1\left(\frac{x-X_k}{b_n}\right)dx, \\
T_4 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int R_2\left(\frac{x-X_j}{b_n}\right)A_1\left(\frac{x-X_k}{b_n}\right)dx, \\
T_5 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int A_2\left(\frac{x-X_j}{b_n}\right)R_1\left(\frac{x-X_k}{b_n}\right)dx, \\
T_6 &= -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int R_2\left(\frac{x-X_j}{b_n}\right)R_1\left(\frac{x-X_k}{b_n}\right)dx.
\end{aligned}$$

First, we show that $T_1, \ldots, T_6$ are asymptotically negligible, i.e.,

$$T_1, \ldots, T_6 = o_p(n^{-1/2}). \quad (7)$$

$(\hat{\xi}_{d_1}, \xi_{d_1})$. Then, by letting

$$\begin{aligned}
A_{x,1}(x) &= \sum_{\delta\in\Delta}\prod_{d=1}^{D_x}\mathbb{K}_d^{\delta_d}(x_d)A_{d,1}^{1-\delta_d}(x_d), \quad A_{x,2}^{d_1}(x) = \sum_{\delta\in\Delta}\left\{\prod_{d\neq d_1}\mathbb{K}_d^{\delta_d}(x_d)A_{d,1}^{1-\delta_d}(x_d)\right\}(\mathbb{K}'_{d_1})^{\delta_{d_1}}(x_{d_1})A_{d_1,2}^{1-\delta_{d_1}}(x_{d_1}), \\
R_{x,1}(x) &= \hat{\mathbb{K}}_x(x) - \mathbb{K}_x(x) - A_{x,1}(x), \quad R_{x,2}^{d_1}(x) = \frac{\partial\hat{\mathbb{K}}_x(x)}{\partial x_{d_1}} - \frac{\partial\mathbb{K}_x(x)}{\partial x_{d_1}} - A_{x,2}^{d_1}(x),
\end{aligned}$$

with $\Delta = \{(\delta_1, \ldots, \delta_{D_x}) : \delta_{d_1} = 0, 1 \text{ for } d_1 = 1, \ldots, D_x\} \setminus \{(1, \ldots, 1)\}$, we obtain

$$\hat{\mathbb{K}}_x(x) = \mathbb{K}_x(x) + A_{x,1}(x) + R_{x,1}(x) \quad \text{and} \quad \frac{\partial\hat{\mathbb{K}}_x(x)}{\partial x_{d_1}} = \frac{\partial\mathbb{K}_x(x)}{\partial x_{d_1}} + A_{x,2}^{d_1}(x) + R_{x,2}^{d_1}(x).$$

Thus, we can obtain a similar decomposition of $\hat{\theta}_{x,d_1}$ and $\hat{\theta}_{z,d_2}$ as (6), by which analogous results to our main theorem can be established following a similar route to the univariate case.

For T_2, we decompose $T_2 = T_{2,1} + T_{2,2}$, where

$$T_{2,1} = -\frac{\mathrm{i}}{2\pi n^2 b_n^3}\sum_{j=1}^n\sum_{k=1}^n Y_k \iint \left[\begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} tK^{\mathrm{ft}}(t)\{|\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2}\} \\ \times|\xi(t/b_n)|^{-1}\{\hat{\xi}(t/b_n) - \xi(t/b_n)\}\end{array}\right] dt\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx,$$

$$T_{2,2} = -\frac{\mathrm{i}}{\pi n^2 b_n^3}\sum_{j=1}^n\sum_{k=1}^n Y_k \iint \left[\begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} tK^{\mathrm{ft}}(t)\{|\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2}\} \\ \times|\xi(t/b_n)|^{-1/2}\{|\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2}\}\end{array}\right] dt\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx.$$

For $T_{2,1}$, we have

$$\begin{aligned}
|n^{1/2}T_{2,1}| &= \left|\frac{1}{2\pi n^{3/2} b_n^2}\sum_{j=1}^n\sum_{k=1}^n Y_k \int \left\{\begin{array}{c} e^{\mathrm{i}t\left(\frac{X_j-X_k}{b_n}\right)} tK^{\mathrm{ft}}(t)K^{\mathrm{ft}}(-t)\{|\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2}\} \\ \times|\xi(t/b_n)|^{-3/2}\{\hat{\xi}(t/b_n) - \xi(t/b_n)\}\end{array}\right\} dt\right| \\
&= O_p\left(n^{1/2}b_n^{-2}\sup_{|t|\le b_n^{-1}}\left|\{|\tilde{\xi}(t)|^{-1/2} - |\xi(t)|^{-1/2}\}|\xi(t)|^{-3/2}\{\hat{\xi}(t) - \xi(t)\}\right|\right) \\
&= O_p\left(n^{1/2}b_n^{-2}e^{4\mu b_n^{-\gamma}}\varrho_n^2\right) = o_p(1),
\end{aligned}$$

where the first equality follows from a change of variables, the second equality follows from $\left|e^{\mathrm{i}t\left(\frac{X_j-X_k}{b_n}\right)}\right| = 1$, $\frac{1}{n}\sum_{k=1}^n |Y_k| = O_p(1)$, and $\int |tK^{\mathrm{ft}}(t)K^{\mathrm{ft}}(-t)| < \infty$ (by Assumption SS (3)), the third equality follows from the definition of $\tilde{\xi}(t)$, Assumption SS (2), and Lemma 1 (in Appendix B below), and the last equality follows from Assumption SS (4). A similar argument yields $T_{2,2} = o_p(n^{-1/2})$, and thus $T_2 = o_p(n^{-1/2})$. Also, using similar arguments as for T_2, we obtain $T_1 = o_p(n^{-1/2})$.

For T_3, note that

$$\begin{aligned}
|n^{1/2}T_3| &= \left|\frac{1}{(4\pi)^2 n^{3/2} b_n^2}\sum_{j=1}^n\sum_{k=1}^n Y_k \int \left\{\begin{array}{c} e^{\mathrm{i}t\left(\frac{X_k-X_j}{b_n}\right)} tK^{\mathrm{ft}}(t)K^{\mathrm{ft}}(-t)|\xi(t/b_n)|^{-3/2}|\xi(-t/b_n)|^{-3/2} \\ \times\{\hat{\xi}(t/b_n) - \xi(t/b_n)\}\{\hat{\xi}(-t/b_n) - \xi(-t/b_n)\}\end{array}\right\} dt\right| \\
&= O_p\left(n^{1/2}b_n^{-2}\left(\sup_{|t|\le b_n^{-1}}\left|\frac{|\hat{\xi}(t)|^{1/2} - |\xi(t)|^{1/2}}{|\xi(t/b_n)|^{1/2}}\right|\right)^2\right) \\
&= O_p\left(n^{1/2}b_n^{-2}e^{2\mu b_n^{-\gamma}}\varrho_n^2\right) = o_p(1),
\end{aligned}$$

where the first equality follows from a change of variables, the second equality follows from $\left|e^{\mathrm{i}t\left(\frac{X_k-X_j}{b_n}\right)}\right| = 1$, $\frac{1}{n}\sum_{k=1}^n |Y_k| = O_p(1)$, and $\int |tK^{\mathrm{ft}}(t)K^{\mathrm{ft}}(-t)| < \infty$, the third equality follows from Lemma 1 and Assumption SS (2), and the last equality follows from Assumption SS (4).

For T_4, we decompose $T_4 = T_{4,1} + T_{4,2}$, where

$$
\begin{aligned}
T_{4,1} &= \frac{\mathrm{i}}{8\pi^2 n^2 b_n^3} \sum_{j=1}^n \sum_{k=1}^n Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} t K^{\mathrm{ft}}(t) \{ |\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1} \left\{ \hat{\xi}(t/b_n) - \xi(t/b_n) \right\} \end{array} \right\} dt \\ \times \int e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t) |\xi(t/b_n)|^{-3/2} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} dt \end{array} \right] dx, \\
T_{4,2} &= \frac{\mathrm{i}}{4\pi^2 n^2 b_n^3} \sum_{j=1}^n \sum_{k=1}^n Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} t K^{\mathrm{ft}}(t) \{ |\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1/2} \{ |\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2} \} \end{array} \right\} dt \\ \times \int e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t) |\xi(t/b_n)|^{-3/2} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} dt \end{array} \right] dx.
\end{aligned}
$$

For $T_{4,1}$, we have

$$
\begin{aligned}
|n^{1/2} T_{4,1}| &= O_p\left(n^{1/2} b_n^{-2} \sup_{|t| \le b_n^{-1}} \left| \{ |\tilde{\xi}(t)|^{-1/2} - |\xi(t)|^{-1/2} \} |\xi(t)|^{-5/2} \{ \hat{\xi}(t) - \xi(t) \}^2 \right| \right) \\
&= O_p\left(n^{1/2} b_n^{-2} e^{5\mu b_n^{-\gamma}} \varrho_n^3 \right) = o_p(1),
\end{aligned}
$$

where the first equality follows from a change of variables, $\left| e^{\mathrm{i}t\left(\frac{X_j - X_k}{b_n}\right)} \right| = 1$, $\frac{1}{n}\sum_{k=1}^n |Y_k| = O_p(1)$, and $\int |t K^{\mathrm{ft}}(t) K^{\mathrm{ft}}(-t)| dt < \infty$ (by Assumption SS (3)), the second equality follows from the definition of $\tilde{\xi}(t)$, Assumption SS (2), and Lemma 1, and the last equality follows from Assumption SS (4). A similar argument yields $T_{4,2} = o_p(n^{-1/2})$, and thus $T_4 = o_p(n^{-1/2})$. Also, similar arguments as used for T_4 imply $T_5 = o_p(n^{-1/2})$.

For T_6, we decompose $T_6 = T_{6,1} + T_{6,2} + T_{6,3} + T_{6,4}$, where

$$
\begin{aligned}
T_{6,1} &= \frac{\mathrm{i}}{8\pi^2 n^2 b_n^3} \sum_{j=1}^n \sum_{k=1}^n Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} t K^{\mathrm{ft}}(t) \{ |\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} \end{array} \right\} dt \\ \times \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t) \{ |\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} \end{array} \right\} dt \end{array} \right] dx \\
T_{6,2} &= \frac{\mathrm{i}}{4\pi^2 n^2 b_n^3} \sum_{j=1}^n \sum_{k=1}^n Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} t K^{\mathrm{ft}}(t) \{ |\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1/2} \{ |\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2} \} \end{array} \right\} dt \\ \times \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t) \{ |\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ |\xi(t/b_n)|^{-1} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} \end{array} \right\} dt \end{array} \right] dx \\
T_{6,3} &= \frac{\mathrm{i}}{4\pi^2 n^2 b_n^3} \sum_{j=1}^n \sum_{k=1}^n Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} t K^{\mathrm{ft}}(t) \{ |\tilde{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1} \{ \hat{\xi}(t/b_n) - \xi(t/b_n) \} \end{array} \right\} dt \\ \times \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t) \{ |\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2} \} \\ \times |\xi(t/b_n)|^{-1/2} \{ |\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2} \} \end{array} \right\} dt \end{array} \right] dx
\end{aligned}
$$

$$T_{6,4} = \frac{\mathrm{i}}{2\pi^2 n^2 b_n^3} \sum_{j=1}^{n} \sum_{k=1}^{n} Y_k \int \left[\begin{array}{c} \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} tK^{\mathrm{ft}}(t)\{|\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2}\} \\ \times|\xi(t/b_n)|^{-1/2}\{|\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2}\} \end{array} \right\} dt \\ \times \int \left\{ \begin{array}{c} e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} K^{\mathrm{ft}}(t)\{|\hat{\xi}(t/b_n)|^{-1/2} - |\xi(t/b_n)|^{-1/2}\} \\ \times|\xi(t/b_n)|^{-1/2}\{|\hat{\xi}(t/b_n)|^{1/2} - |\xi(t/b_n)|^{1/2}\} \end{array} \right\} dt \end{array} \right] dx.$$

Since $T_{6,2}$ and $T_{6,3}$ are cross-product terms, it is enough to focus on $T_{6,1}$ and $T_{6,4}$. For $T_{6,1}$, we have

$$\begin{aligned} |n^{1/2} T_{6,1}| &= O_p\left(n^{1/2} b_n^{-2} \sup_{|t| \le b_n^{-1}} \left| \{|\tilde{\xi}(t)|^{-1/2} - |\xi(t)|^{-1/2}\}^2 |\xi(t)|^{-2} \{\hat{\xi}(t) - \xi(t)\}^2 \right| \right) \\ &= O_p\left(n^{1/2} b_n^{-2} e^{6\mu b_n^{-\gamma}} \varrho_n^4 \right) = o_p(1), \end{aligned}$$

where the first equality follows from a change of variables, $\left| e^{\mathrm{i}t\left(\frac{X_j - X_k}{b_n}\right)} \right| = 1$, $\frac{1}{n}\sum_{k=1}^n |Y_k| = O_p(1)$, and $\int |tK^{\mathrm{ft}}(t)K^{\mathrm{ft}}(-t)| dt < \infty$ (by Assumption SS (3)), the second equality follows from the definition of $\tilde{\xi}(t)$, Assumption SS (2), and Lemma 1, and the last equality follows from Assumption SS (4). A similar argument yields $T_{6,4} = o_p(n^{-1/2})$, and thus $T_6 = o_p(n^{-1/2})$.

Combining these results, we obtain (7).

We now consider the term S in (6). Let $d_j = (Y_j, X_j, \xi_j)$ and

$$p_n(d_j, d_k, d_l) = q_n(d_j, d_k, d_l) + q_n(d_j, d_l, d_k) + q_n(d_k, d_j, d_l) + q_n(d_k, d_l, d_j) + q_n(d_l, d_j, d_k) + q_n(d_l, d_k, d_j),$$

where

$$q_n(d_j, d_k, d_l) = -\frac{1}{3b_n^3} \left(\begin{array}{c} \int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right) Y_k \mathbb{K}\left(\frac{x-X_k}{b_n}\right) dx \\ +\frac{\mathrm{i}}{4\pi} \int \left\{ \int e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)} Y_k \mathbb{K}\left(\frac{x-X_k}{b_n}\right) dx \right\} \left\{ \frac{\xi_l(t/b_n) - E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}} \right\} tK^{\mathrm{ft}}(t) dt \\ -\frac{1}{4\pi} \int \left\{ \int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right) Y_k e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)} dx \right\} \left\{ \frac{\xi_l(t/b_n) - E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}} \right\} K^{\mathrm{ft}}(t) dt \end{array} \right).$$

We then decompose

$$S = n^{-2}(n-1)(n-2)U + S_1 + S_2 + S_3 + S_4,$$

where

$$
\begin{aligned}
U &= \binom{n}{3}^{-1}\sum_{j=1}^{n}\sum_{k=j+1}^{n}\sum_{l=k+1}^{n} p_n(d_j,d_k,d_l),\\
S_1 &= \frac{6}{n^3}\sum_{j=1}^{n}\sum_{k=j+1}^{n}[q_n(d_j,d_j,d_k)+q_n(d_k,d_k,d_j)], \qquad S_2=\frac{6}{n^3}\sum_{j=1}^{n}\sum_{k=j+1}^{n}[q_n(d_j,d_k,d_j)+q_n(d_k,d_j,d_k)],\\
S_3 &= \frac{6}{n^3}\sum_{j=1}^{n}\sum_{k=j+1}^{n}[q_n(d_j,d_k,d_k)+q_n(d_k,d_j,d_j)], \qquad S_4=\frac{6}{n^3}\sum_{j=1}^{n} q_n(d_j,d_j,d_j).
\end{aligned}
$$

We show that

$$
S_1,\dots,S_4 = o_p(n^{-1/2}), \tag{8}
$$

in the following way.

For S_1, decompose

$$
\begin{aligned}
&|n^{1/2}S_1|\\
=\;& O(n^{-5/2}b_n^{-3})\left[\begin{array}{r}
\left|\sum_{j=1}^{n}\sum_{k=j+1}^{n}\int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)Y_k\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx\right|\\
+\left|\sum_{j=1}^{n}\sum_{k=j+1}^{n}\frac{\mathrm{i}}{4\pi}\int\left\{\int e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)}Y_k\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx\right\}\left\{\frac{\xi_l(t/b_n)-E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}tK^{\mathrm{ft}}(t)dt\right|\\
+\left|\sum_{j=1}^{n}\sum_{k=j+1}^{n}\frac{1}{4\pi}\int\left\{\int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)Y_k e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)}dx\right\}\left\{\frac{\xi_l(t/b_n)-E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}K^{\mathrm{ft}}(t)dt\right|
\end{array}\right]\\
\equiv\;& S_{1,1}+S_{1,2}+S_{1,3}.
\end{aligned}
$$

To bound $S_{1,1}$, we write

$$
\begin{aligned}
S_{1,1} &= O(n^{-5/2}b_n^{-2})\left|\sum_{j=1}^{n}\sum_{k=j+1}^{n}Y_k\left\{\int\frac{1}{b_n}e^{-\mathrm{i}(s+t)x/b_n}dx\right\}\iint \mathrm{i}se^{\mathrm{i}\left(\frac{tX_k+sX_j}{b_n}\right)}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/b_n)}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b_n)}dsdt\right|\\
&= O_p\left(n^{-1/2}b_n^{-2}e^{2\mu b_n^{-\gamma}}\right)=o_p(1),
\end{aligned}
$$

where the second equality follows from a change of variables, $\left|e^{\mathrm{i}\left(\frac{tX_k+sX_j}{b_n}\right)}\right|=1$, $\frac{1}{n}\sum_{k=1}^{n}|Y_k|=O_p(1)$, and Assumption SS (3), and the last equality follows from Assumption SS (4). A similar argument as used for T_3 can be used to show $S_{1,2}=O_p\left(n^{-1/2}b_n^{-2}e^{4\mu b_n^{-\gamma}}\varrho_n\right)=o_p(1)$. Furthermore, the same arguments can be used to show $S_2,S_3,S_4=o_p(n^{-1/2})$.

We now analyze the main term U. Let $r_n(d_j)=E[p_n(d_j,d_k,d_l)|d_j]$ and $\hat{U}=\theta+\frac{3}{n}\sum_{j=1}^{n}\{r_n(d_j)-E[r_n(d_j)]\}$. By Ahn and Powell (1993, Lemma A.3), if

$$
E[p_n(d_j,d_k,d_l)^2]=o(n), \tag{9}
$$

then it holds

$$\begin{aligned} U &= E[r_n(d_j)] + \frac{3}{n}\sum_{j=1}^{n}\{r_n(d_j) - E[r_n(d_j)]\} + o_p(n^{-1/2}) \\ &= \theta + \frac{3}{n}\sum_{j=1}^{n}\{r_n(d_j) - E[r_n(d_j)]\} + o_p(n^{-1/2}), \end{aligned} \tag{10}$$

where the second equality follows from $E[r_n(d_j)] = \theta$ (i.e., the bias term is exactly zero due to the infinite differentiability in Assumption SS (1) and infinite-order kernel in Assumption SS (3)).

For (9), note that

$$\begin{aligned} &E[p_n(d_j,d_k,d_l)^2] \leq \frac{1}{3b_n^6}E\left[\left\{\int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)Y_k\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx\right\}^2\right] \\ &+\frac{1}{3b_n^6}E\left[\left\{\frac{\mathrm{i}}{4\pi}\int\left\{\int e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)}Y_k\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx\right\}\left\{\frac{\xi_l(t/b_n)-E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}tK^{\mathrm{ft}}(t)dt\right\}^2\right] \\ &+\frac{1}{3b_n^6}E\left[\left\{\frac{1}{4\pi}\int\left\{\int \mathbb{K}'\left(\frac{x-X_j}{b_n}\right)Y_ke^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)}dx\right\}\left\{\frac{\xi_l(t/b_n)-E[\xi_l(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}K^{\mathrm{ft}}(t)dt\right\}^2\right] \\ &\equiv P_1+P_2+P_3. \end{aligned}$$

For P_1,

$$\begin{aligned} P_1 &= \frac{1}{3b_n^4}\iiiint\left\{\int \mathbb{K}'(z)\mathbb{K}\left(z+\frac{s_j+t_j-s_k-t_k}{b_n}\right)dz\right\}^2 E[Y^2|X^*=s_k] \\ &\quad \times f(s_k)f(s_j)f_\epsilon(t_k)f_\epsilon(t_j)ds_kds_jdt_kdt_j \\ &= \frac{1}{12\pi^2b_n^4}\iint\left\{\iint e^{-\mathrm{i}(w_1+w_2)\left(\frac{s_j-s_k}{b_n}\right)}E[Y^2|X^*=s_k]f(s_k)f(s_j)ds_kds_j\right\} \\ &\quad \times\frac{w_1w_2|K^{\mathrm{ft}}(w_1)|^2|K^{\mathrm{ft}}(w_2)|^2}{|f_\epsilon^{\mathrm{ft}}(w_1/b_n)|^2|f_\epsilon^{\mathrm{ft}}(w_2/b_n)|^2}dw_1dw_2 \\ &= O\left(b_n^{-4}e^{4\mu b_n^{-\gamma}}\right), \end{aligned}$$

where the first equality follows by the change of variables $z=\frac{x-s_j-t_j}{b_n}$, the second equality follows by Lemma 2, and the penultimate equality follows from Assumption SS (2). Thus, Assumption SS (4) guarantees $P_1 = o(n)$.

For P_2, note that Lemma 2 implies $\int\left\{\frac{\mathrm{i}}{4\pi}\int te^{-\mathrm{i}tz}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b_n)^3}\{\xi_l(t/b_n)-E[\xi_l(t/b_n)]\}dt\right\}\mathbb{K}(z-$

$c)dz = \frac{\mathrm{i}}{4\pi}\int \frac{we^{-\mathrm{i}wc}|K^{\mathrm{ft}}(w)|^2}{|f_\epsilon^{\mathrm{ft}}(w/b_n)|^4}\{\xi_l(w/b_n) - E[\xi_l(w/b_n)]\}dw$. Then we can write

$$\begin{aligned}
P_2 &= \frac{1}{3b_n^6}E\left[Y_k^2\left\{\int\left\{\frac{\mathrm{i}}{4\pi}\int te^{-\mathrm{i}tz}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b_n)^3}\{\xi_l(t/b_n) - E[\xi_l(t/b_n)]\}dt\right\}\mathbb{K}\left(\frac{x-X_k}{b_n}\right)dx\right\}^2\right] \\
&= \frac{1}{3b_n^6}\int\cdots\int E\left[\left\{\frac{\mathrm{i}}{4\pi}\int\int\left\{\begin{array}{c} te^{-\mathrm{i}t\left(\frac{x-s_j-u_j}{b_n}\right)}K^{\mathrm{ft}}(t) \\ \times f_\epsilon^{\mathrm{ft}}(t/b_n)^{-3}\{\xi_l(t/b_n) - E[\xi_l(t/b_n)]\}\end{array}\right\}dt\mathbb{K}\left(\frac{x-s_k-u_k}{b_n}\right)dx\right\}^2\right] \\
&\quad \times E[Y^2|X^* = s_k]f(s_k)f(s_j)f_v(u_k)f_v(u_j)ds_kds_jdu_kdu_j \\
&= \frac{1}{12\pi^2 b_n^4}\int\int\left\{\int\int e^{-\mathrm{i}(w_1+w_2)\left(\frac{s_j-s_k}{b_n}\right)}E[Y^2|X^*=s_k]f(s_k)f(s_j)ds_kds_j\right\} \\
&\quad \times\frac{w_1w_2|K^{\mathrm{ft}}(w_1)|^2|K^{\mathrm{ft}}(w_2)|^2}{|f_\epsilon^{\mathrm{ft}}(w_1/b_n)|^6|f_\epsilon^{\mathrm{ft}}(w_2/b_n)|^6}E[\{\xi_l(w_1/b_n) - E[\xi_l(w_1/b_n)]\}\{\xi_l(w_2/b_n) - E[\xi_l(w_2/b_n)]\}]dw_1dw_2 \\
&= O\left(b_n^{-4}e^{12\mu b_n^{-\gamma}}\log(b_n^{-1})^{-1}\right) = o(n),
\end{aligned}$$

where the third equality follows from a similar argument as for P_1 combined with Kato and Sasaki (2018, Lemma 4) to bound $\{\xi_l(w_1/b_n) - E[\xi_l(w_1/b_n)\}$, and the last equality follows from Assumption SS (4). The order of P_3 can be shown in an almost identical manner, and we obtain (9).

Combining (6), (7), (8), (10), and a direct calculation to characterize $r_n(d_j) = E[p_n(d_j, d_k, d_l)|d_j]$, it follows that

$$\begin{aligned}
&\sqrt{n}(\hat{\theta} - \theta) = \frac{3}{\sqrt{n}}\sum_{j=1}^n\{r_n(d_j) - E[r_n(d_j)]\} + o_p(1), \\
&= \frac{2}{\sqrt{n}b_n^3}\sum_{j=1}^n\{\eta_j - E[\eta_j]\} - \frac{1}{2\pi\sqrt{n}b_n^3}\sum_{j=1}^n\int\Delta(t)\left\{\frac{\xi_j(t/b_n) - E[\xi_j(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}K^{\mathrm{ft}}(t)dt + o_p(1), \qquad (11)
\end{aligned}$$

where

$$\begin{aligned}
\eta_j &= \int\mathbb{K}\left(\frac{x-X_j}{b_n}\right)\left\{E\left[YK'\left(\frac{x-X^*}{b_n}\right)\right] - Y_jE\left[K'\left(\frac{x-X^*}{b_n}\right)\right]\right\}dx, \\
\Delta(t) &= \int\left\{\mathrm{i}tE\left[e^{-\mathrm{i}t\left(\frac{x-X}{b_n}\right)}\right]E\left[YK\left(\frac{x-X^*}{b_n}\right)\right] - E\left[K'\left(\frac{x-X^*}{b_n}\right)\right]E\left[Ye^{-\mathrm{i}t\left(\frac{x-X}{b_n}\right)}\right]\right\}dx.
\end{aligned}$$

For the first term in (11), note that

$$
\begin{aligned}
\frac{\eta_j}{b_n^3} &= \frac{1}{b_n^2}\int \mathbb{K}(z)\{q_1(X_j+b_nz)-Y_jq_2(X_j+b_nz)\}dz \\
&= \frac{1}{b_n}\int \mathbb{K}(z)\int\{G^{(1)}-Y_jf^{(1)}\}(x^*)K\left(\frac{X_j+b_nz-x^*}{b_n}\right)dx^*dz \\
&= \int \mathbb{K}(z)\int\{Y_jf^{(1)}-G^{(1)}\}\left(X_j+b_n(z-\tilde{x}^*)\right)K(\tilde{x}^*)d\tilde{x}^*dz \\
&= \int \mathbb{K}(z)\left\{\sum_{l=0}^{+\infty}\frac{b_n^l}{l!}\{Y_jf^{(l+1)}-G^{(l+1)}\}(X_j)\int(z-\tilde{x}^*)^lK(\tilde{x}^*)d\tilde{x}^*\right\}dz \\
&= \sum_{l=0}^{+\infty}\frac{b_n^l}{l!}\int \mathbb{K}(z)z^ldz\{Y_jf^{(l+1)}-G^{(l+1)}\}(X_j) \\
&= \sum_{h=0}^{\infty}\frac{\mu^h}{\mathrm{i}^{h\gamma}Ch!}\{Y_jf^{(h\gamma+1)}-G^{(h\gamma+1)}\}(X_j),
\end{aligned}
$$

where the first equality follows by $q_1(x)=E\left[YK'\left(\frac{x-X^*}{b_n}\right)\right]$ and $q_2(x)=E\left[K'\left(\frac{x-X^*}{b_n}\right)\right]$ and the change of variable $z=\frac{x-X_j}{b_n}$, the second equality follows by integration-by-parts, the third equality follows by the change of variable $\tilde{x}^*=\frac{X_j+b_nz-x^*}{b_n}$, the fourth equality follows by the infinite differentiability of G and f, the fifth equality follows by using the binomial theorem and the infinite-order property of the kernel function K, and the last equality follows by Lemma 4.

Let $\Xi_j(t)=\frac{\xi_j(t/b_n)-E[\xi_j(t/b_n)]}{|\xi(t/b_n)|}$. For the second term in (11), we have

$$
\begin{aligned}
&\int \Delta(t)\left\{\frac{\xi_j(t/b_n)-E[\xi_j(t/b_n)]}{|\xi(t/b_n)|^{3/2}}\right\}K^{\mathrm{ft}}(t)dt \\
=\;& \mathrm{i}\int tf^{\mathrm{ft}}(t/b_n)\left\{\iint e^{-\mathrm{i}tx/b_n}K\left(\frac{x-x^*}{b_n}\right)G(x^*)dxdx^*\right\}\Xi_j(t)K^{\mathrm{ft}}(t)dt \\
&-\int G^{\mathrm{ft}}(t/b_n)\left\{\iint e^{-\mathrm{i}tx/b_n}K'\left(\frac{x-x^*}{b_n}\right)f(x^*)dxdx^*\right\}\Xi_j(t)K^{\mathrm{ft}}(t)dt \\
=\;& \mathrm{i}\int tf^{\mathrm{ft}}(t/b_n)\left\{\iint e^{-\mathrm{i}tx/b_n}K\left(\frac{x-x^*}{b_n}\right)G(x^*)dxdx^*\right\}\Xi_j(t)K^{\mathrm{ft}}(t)dt \\
&-\mathrm{i}\int tG^{\mathrm{ft}}(t/b_n)\left\{\iint e^{-\mathrm{i}tx/b_n}K\left(\frac{x-x^*}{b_n}\right)f(x^*)dxdx^*\right\}\Xi_j(t)K^{\mathrm{ft}}(t)dt \\
=\;& \mathrm{i}b_n\int tK^{\mathrm{ft}}(t)f^{\mathrm{ft}}(t/b_n)K^{\mathrm{ft}}(-t)G^{\mathrm{ft}}(-t/b_n)\Xi_j(t)dt \\
&-\mathrm{i}b_n\int tK^{\mathrm{ft}}(-t)f^{\mathrm{ft}}(-t/b_n)K^{\mathrm{ft}}(t)G^{\mathrm{ft}}(t/b_n)\Xi_j(t)dt \\
=\;& 0, \qquad (12)
\end{aligned}
$$

where the first equality follows from the definition of $\Delta(t)$, the second equality follows by integration-by-parts, that is $\int e^{-\mathrm{i}tx/b_n}K'\left(\frac{x-x^*}{b_n}\right)dx=\mathrm{i}t\int e^{-\mathrm{i}tx/b_n}K\left(\frac{x-x^*}{b_n}\right)dx$, the third equality follows from a change of variables, and the last equality follows from symmetry of $\xi_j(t)$ and

$\xi(t)$ (which implies symmetry of $\Xi_j(t)$).

Therefore, the conclusion follows by the central limit theorem.

B Lemmas

Lemma 1. *[Kato and Sasaki, 2018, Lemma 4] Under Assumption SS,*

$$\sup_{|t|\le b_n^{-1}} |\hat{f}_\epsilon^{\mathrm{ft}}(t) - f_\epsilon^{\mathrm{ft}}(t)| = O_p(\varrho_n),$$

where $\varrho_n = n^{-1/2}\log(b_n^{-1})^{1/2}$.

Lemma 2. *Under Assumption SS (1) and (3), it holds*

$$\int \mathbb{K}'(z)\mathbb{K}(z+c)dz = \frac{\mathrm{i}}{2\pi}\int \frac{we^{-\mathrm{i}wc}|K^{\mathrm{ft}}(w)|^2}{|f_\epsilon^{\mathrm{ft}}(w/b_n)|^2}dw,$$

for any constant c.

Proof. Observe that

$$\begin{aligned}
\int \mathbb{K}'(z)\mathbb{K}(z-c)dz &= \int \left(\frac{-\mathrm{i}}{2\pi}\int w_1 e^{-\mathrm{i}w_1 z}\frac{K^{\mathrm{ft}}(w_1)}{f_\epsilon^{\mathrm{ft}}(w_1/b_n)}dw_1\right)\left(\frac{1}{2\pi}\int e^{-\mathrm{i}w_2 z}\frac{e^{-\mathrm{i}w_2 c}K^{\mathrm{ft}}(w_2)}{f_\epsilon^{\mathrm{ft}}(w_2/b_n)}dw_2\right)dz \\
&= \frac{-\mathrm{i}}{2\pi}\iint \left(\frac{1}{2\pi}\int e^{-\mathrm{i}(w_1+w_2)z}dz\right)\frac{w_1 e^{-\mathrm{i}w_2 c}K^{\mathrm{ft}}(w_1)K^{\mathrm{ft}}(w_2)}{f_\epsilon^{\mathrm{ft}}(w_1/b_n)f_\epsilon^{\mathrm{ft}}(w_2/b_n)}dw_1 dw_2 \\
&= \frac{\mathrm{i}}{2\pi}\int \frac{we^{-\mathrm{i}wc}|K^{\mathrm{ft}}(w)|^2}{|f_\epsilon^{\mathrm{ft}}(w/b_n)|^2}dw,
\end{aligned}$$

where the last equality follows by $\int \delta(w-b)f(w)dw = f(b)$ with Dirac delta function $\delta(w) = \frac{1}{2\pi}\int e^{-\mathrm{i}wx}dx$. □

Lemma 3. *Under Assumption SS (1)-(3), it holds*

$$\left|\int \mathbb{K}'(z)\mathbb{K}(z)dz\right| = O(e^{2\mu b_n^{-\gamma}}).$$

Proof. By Lemma 2, we have

$$\left|\int \mathbb{K}'(z)\mathbb{K}(z)dz\right| = \frac{1}{2\pi}\left|\int \frac{w|K^{\mathrm{ft}}(w)|^2}{|f_\epsilon^{\mathrm{ft}}(w/b_n)|^2}dw\right| = O\left(\left(\inf_{|w|\le b_n^{-1}}|f_\epsilon^{\mathrm{ft}}(w)|\right)^{-2}\right),$$

where the second equality follows from the compactness of the support of K^{ft} (Assumption SS (3)). The conclusion follows by Assumption SS (2). □

Lemma 4. *Under Assumptions SS (1)-(3), it holds*

$$\int \mathbb{K}(z)z^p dz = \begin{cases} \frac{\mu^{p/\gamma}p!}{b_n^p \mathrm{i}^p C(p/\gamma)!} & \textit{for } p = h\gamma \textit{ with } h = 0, 1, \dots, \\ 0 & \textit{for other positive integers.} \end{cases}$$

Proof. First, note that

$$\begin{aligned}
\mathbb{K}(z) &= \frac{1}{2\pi C}\int e^{-\mathrm{i}tz} e^{\mu|t/b_n|^\gamma} K^{\mathrm{ft}}(|t|)dt = \sum_{h=0}^{+\infty} \frac{\mu^h}{Ch! b_n^{h\gamma}} \left\{ \frac{1}{2\pi} \int e^{-\mathrm{i}tz} |t|^{h\gamma} K^{\mathrm{ft}}(|t|)dt \right\} \\
&= \sum_{h=0}^{+\infty} \frac{\mu^h}{Ch!(-\mathrm{i}b_n)^{h\gamma}} \left\{ \frac{1}{2\pi} \int e^{-\mathrm{i}tz} (K^{(h\gamma)})^{\mathrm{ft}}(|t|)dt \right\} \\
&= \sum_{h=0}^{+\infty} \frac{\mu^h}{Ch!(-\mathrm{i}b_n)^{h\gamma}} \left\{ \frac{1}{2\pi} \int e^{-\mathrm{i}tz} (K^{(h\gamma)})^{\mathrm{ft}}(t)dt \right\} = \sum_{h=0}^{+\infty} \frac{\mu^h}{Ch!(-\mathrm{i}b_n)^{h\gamma}} K^{(h\gamma)}(z),
\end{aligned}$$

where the first equality follows by Assumption SS (2) and $K^{\mathrm{ft}}(t) = K^{\mathrm{ft}}(-t)$, the second equality follows by $e^u = \sum_{h=0}^{+\infty} \frac{u^h}{h!}$, the third equality follows by $(K^{(l)})^{\mathrm{ft}}(t) = (-\mathrm{i}t)^l K^{\mathrm{ft}}(t)$ (see, e.g., Lemma A.6 of Meister, 2009), the fourth equality follows by $(K^{(h\gamma)})^{\mathrm{ft}}(-t) = (K^{(h\gamma)})^{\mathrm{ft}}(t)$, which uses $K^{\mathrm{ft}}(t) = K^{\mathrm{ft}}(-t)$, $(K^{(l)})^{\mathrm{ft}}(t) = (-\mathrm{i}t)^l K^{\mathrm{ft}}(t)$, and the assumption that γ is even. Thus, we have

$$\int \mathbb{K}(z)z^p dz = \sum_{h=0}^{+\infty} \frac{\mu^h}{Ch!(-\mathrm{i}b_n)^{h\gamma}} \int z^p K^{(h\gamma)}(z)dz,$$

and the conclusion follows by Assumption SS (3) and integration-by-parts. □

C Proof of theorem (ordinary smooth case)

The steps in this proof are the same as that for the supersmooth case, as such, we only explain parts of the proof that differ. Furthermore, in the proof of the supersmooth case we endeavor to obtain expressions in terms of f_ϵ^{ft} wherever possible. This allows us to skip to the final step in each asymptotic argument, and requires input only of the relevant form for f_ϵ^{ft}. This proof also leverages much of the work from Fan (1995) but extends this by allowing for an estimated measurement error density.

As in the proof of the supersmooth case, we have

$$\hat{\theta} = -\frac{2}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n} Y_k \int \hat{\mathbb{K}}'\left(\frac{x-X_j}{b_n}\right)\hat{\mathbb{K}}\left(\frac{x-X_k}{b_n}\right) dx = S + T_1 + \cdots + T_6,$$

where $S, T_1, \ldots, T_6$ are defined in Section A. We were able to show that

$$|n^{1/2}T_2| = O\left(n^{1/2}b_n^{-2}\left(\sup_{|t|\le b_n^{-1}} |\hat{f}_\epsilon^{\mathrm{ft}}(t) - f_\epsilon^{\mathrm{ft}}(t)|\right)^2 \left(\inf_{|t|\le b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|^4\right)^{-1}\right).$$

Then we have $T_2 = o_p(n^{-1/2})$ by Lemma 1 and Assumption OS (2). The rest of $T_1, T_3, \ldots T_6$ are shown to be of order $o_p(n^{-1/2})$ in a similar way.

Again, decompose $S = n^{-2}(n-1)(n-2)U + S_1 + \cdots + S_4$, where all objects are defined in the proof of the supersmooth case. We can show the asymptotic negligibility of $S_1, \ldots, S_4$ as follows. We again decompose $|n^{1/2}S_1| = S_{1,1} + S_{1,2} + S_{1,3}$. To bound $S_{1,1}$, we write

$$S_{1,1} = O_p\left(n^{-1/2}b_n^{-2}\left(\inf_{|t|\le b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right) = o_p(1).$$

where the second equality follows from Assumption OS (2) and (4). Recall from the proof of the supersmooth case

$$S_{1,2} = O_p\left(n^{-1/2}b_n^{-2}\left(\sup_{|t|\le b_n^{-1}} |\hat{f}_\epsilon^{\mathrm{ft}}(t) - f_\epsilon^{\mathrm{ft}}(t)|\right)\left(\inf_{|t|\le b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|^4\right)^{-1}\right) = o_p(1).$$

The asymptotic negligibility of $S_{1,3}$ can be shown in an almost identical way. The same arguments can also be used to show $S_2, S_3, S_4 = o_p(n^{-1/2})$.

As in the supersmooth case, we also need to show $E[p_n(d_j, d_k, d_l)^2] = o(n)$ in order to write $U = E[r_n(d_j)] + \frac{3}{n}\sum_{j=1}^{n}\{r_n(d_j) - E[r_n(d_j)]\} + o_p(n^{-1/2})$. We begin by decomposing $E[p_n(d_j, d_k, d_l)^2] = P_1 + P_2 + P_3$, where these objects are defined in the supersmooth proof. For

P_1,

$$P_1 = O\left(b_n^{-4}\left(\inf_{|w|\leq b_n^{-1}}|f_\epsilon^{\mathrm{ft}}(w)|^2\right)^{-2}\right) = o(n),$$

by Assumption OS (2) and (4). For P_2, we can write

$$P_2 = O\left(b_n^{-4}\left(\inf_{|w|\leq b_n^{-1}}|f_\epsilon^{\mathrm{ft}}(w)|^6\right)^{-2}\log(b_n)^{-2}\right) = o(n),$$

by Assumption OS (2) and (4). The order of P_3 can be shown in an almost identical manner.

Fan (1995, Corollary 3.4) implies $E[r_n(d_j)] = \theta + o(n^{-1/2})$. Then, it follows

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{3}{\sqrt{n}}\sum_{j=1}^{n}\{r_n(d_j) - E[r_n(d_j)]\} + o_p(1),$$

and the remainder of the proof for the supersmooth case can be applied except that Lemma 3.1 of Fan (1995) is used instead of our Lemma 4 to characterize $\int \mathbb{K}(z)z^l dz$.

References

[1] Ahn, H. and J. L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics*, 58, 3-29.

[2] Bissantz, N., Dümbgen, L., Holzmann, H. and A. Munk (2007) Non-parametric confidence bands in deconvolution density estimation, *Journal of the Royal Statistical Society: Series B*, 69, 483-506.

[3] Blundell, R., Duncan, A. and K. Pendakur (1998) Semiparametric estimation and consumer demand, *Journal of Applied Econometrics*, 13, 435-461.

[4] Brainard, W. C. and J. Tobin (1968) Econometric models: their problems and usefulness. Pitfalls in financial model building, *American Economic Review*, 58, 99-122.

[5] Butucea, C. and F. Comte (2009) Adaptive estimation of linear functionals in the convolution model and applications, *Bernoulli*, 15, 69-98.

[6] Carroll, R. J., Ruppert, D., Stefanski, L. A. and C. M. Crainiceanu (2006) *Measurement Error in Nonlinear Models*, 2nd edition, Chapman & Hall/CRC.

[7] Cattaneo, M. D., Crump, R. K. and M. Jansson (2010) Robust data-driven inference for density-weighted average derivatives, *Journal of the American Statistical Association*, 105, 1070-1083.

[8] Cattaneo, M. D., Crump, R. K. and M. Jansson (2014) Small bandwidth asymptotics for density-weighted average derivatives, *Econometric Theory*, 30, 176-200.

[9] Chen, X., Hong, H. and D. Nekipelov (2011) Nonlinear models of measurement errors, *Journal of Economic Literature*, 49, 901-937.

[10] Comte, F. and J. Kappus (2015) Density deconvolution from repeated measurements without symmetry assumption on the errors, *Journal of Multivariate Analysis*, 140, 31-46.

[11] Das, M., Newey, W. K. and F. Vella (2003) Nonparametric estimation of sample selection models, *Review of Economic Studies*, 70, 33-58.

[12] Dattner, I., Reiß, M. and M. Trabs (2016) Adaptive quantile estimation in deconvolution with unknown error distribution, *Bernoulli*, 22, 143-192.

[13] Delaigle, A. and I. Gijbels (2004) Practical bandwidth selection in deconvolution kernel density estimation, *Computational Statistics & Data Analysis*, 45, 249-267.

[14] Delaigle, A., Hall, P. and A. Meister (2008) On deconvolution with repeated measurements, *Annals of Statistics*, 36, 665-685.

[15] Devroye, L. (1992) A note on the usefulness of superkernels in density estimates, *Annals of Statistics*, 20, 2037-2056.

[16] Dong, Y. and A. Lewbel (2015) A simple estimator for binary choice models with endogenous regressors, *Econometric Reviews*, 34, 82-105.

[17] Dong, H. and T. Otsu (2018) Nonparametric estimation of additive model with errors-in-variables, Working paper.

[18] Erickson, T. and T. M. Whited (2000) Measurement error and the relationship between investment and q, *Journal of Political Economy*, 108, 1027-1057.

[19] Fan, J. (1991) On the optimal rates of convergence for nonparametric deconvolution problems, *Annals of Statistics*, 19, 1257-1272.

[20] Fan, J. and E. Masry (1992) Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes, *Journal of Multivariate Analysis*, 43, 237-271.

[21] Fan, J. and Y. K. Truong (1993) Nonparametric regression with errors in variables, *Annals of Statistics*, 21, 1900-1925.

[22] Fan, Y. (1995) Average derivative estimation with errors-in-variables, *Journal of Nonparametric Statistics*, 4, 395-407.

[23] Härdle, W., Hildenbrand, W. and M. Jerison (1991) Empirical evidence on the law of demand, *Econometrica*, 1525-1549.

[24] Horowitz, J. L. (2009) *Semiparametric and Nonparametric Methods in Econometrics*, Springer.

[25] Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58, 71-120.

[26] Kato, K. and Y. Sasaki (2018) Uniform confidence bands in deconvolution with unknown error distribution, *Journal of Econometrics*, 207, 129-161.

[27] Li, Q. and J. S. Racine (2007) *Nonparametric Econometrics*, Princeton University Press.

[28] Li, T. and Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators, *Journal of Multivariate Analysis*, 65, 139-165.

[29] Liang, H., Thurston, S. W., Ruppert, D., Apanasovich, T. and R. Hauser (2008) Additive partial linear models with measurement errors, *Biometrika*, 95, 667-678.

[30] Mamuneas, T. P., Savvides, A. and T. Stengos (2006) Economic development and the return to human capital: a smooth coefficient semiparametric approach, *Journal of Applied Econometrics*, 21, 111-132.

[31] McMurry, T. L. and Politis, D. N. (2004) Nonparametric regression with infinite order flat-top kernels, *Journal of Nonparametric Statistics*, 16, 549-562.

[32] Meister, A (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.

[33] Otsu, T. and L. Taylor (2020) Specification testing for errors-in-variables models, *Econometric Theory,* first published online 19 June 2020. DOI: https://doi.org/10.1017/S0266466620000262.

[34] Politis, D. N. and J. P. Romano (1995) Bias-corrected nonparametric spectral estimation, *Journal of Time Series Analysis*, 16, 67-103.

[35] Powell, J. L., Stock, J. H. and T. M. Stoker (1989) Semiparametric estimation of index coefficients, *Econometrica*, 57, 1403-1430.

[36] Powell, J.L. and T. M. Stoker (1996) Optimal bandwidth choice for density-weighted averages, *Journal of Econometrics*, 75, 291-316.

[37] Racine, J. (1997) Consistent significance testing for nonparametric regression, *Journal of Business & Economic Statistics*, 15, 369-378.

[38] Schennach, S. M. (2004) Estimation of nonlinear models with measurement error, *Econometrica*, 72, 33-75.

[39] Schennach, S. M. and Hu, Y. (2013) Nonparametric identification and semiparametric estimation of classical measurement error models without side information, *Journal of the American Statistical Association*, 108, 177-186.

[40] van Es, B. and S. Gugushvili (2008) Weak convergence of the supremum distance for supersmooth kernel deconvolution, *Statistics & Probability Letters*, 78, 2932-2938.

[41] van Es, B. and H.-W. Uh (2005) Asymptotic normality for kernel type deconvolution estimators, *Scandinavian Journal of Statistics*, 32, 467-483.

[42] Yatchew, A. (2003) *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press.

[43] You, J. and G. Chen (2006) Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model, *Journal of Multivariate Analysis*, 97, 324-341.