Consistency and asymptotic normality for a nonparametric prediction under measurement errors

Mynbaev, Kairat and Martins-Filho, Carlos

Kazakh-British Technical University, University of Colorado - Boulder

2015

Consistency and asymptotic normality for a nonparametric prediction under measurement errors[1]

Kairat Mynbaev

International School of Economics
Kazakh-British Technical University
Tolebi 59
Almaty 050000, Kazakhstan
email: kairat_mynbayev@yahoo.com
Voice: + 7 727 303 7004

and

Carlos Martins-Filho

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

February, 2015

Abstract. Nonparametric prediction of a random variable Y conditional on the value of an explanatory variable X is a classical and important problem in Statistics. The problem is significantly complicated if there are heterogeneously distributed measurement errors on the observed values of X used in estimation and prediction. Carroll et al. (2009) have recently proposed a kernel deconvolution estimator and obtained its consistency. In this paper we use the kernels proposed in Mynbaev and Martins-Filho (2010) to define a class of deconvolution estimators for prediction that contains their estimator as one of its elements. First, we obtain consistency of the estimators under much less restrictive conditions. Specifically, contrary to what is routinely assumed in the extant literature, the Fourier transform of the underlying kernels is not required to have compact support, higher-order restrictions on the kernel can be avoided and fractional smoothness of the involved densities is allowed. Second, we obtain asymptotic normality of the estimators under the assumption that there are two types of measurement errors on the observed values of X. It is apparent from our study that even in this simplified setting there are multiple cases exhibiting different asymptotic behavior. Our proof focuses on the case where measurement errors are super-smooth and we use it to discuss other possibilities. The results of a Monte Carlo simulation are provided to compare the performance of the estimator using traditional kernels and those proposed in Mynbaev and Martins-Filho (2010).

Keywords and phrases. Measurement errors, nonparametric prediction, asymptotic normality, Lipschitz conditions.

AMS-MS Classification. 62F12, 62G07, 62G20.

[1]We thank two referees and an Associate Editor for helpful and stimulating comments. The second author thanks I4-Basis for financial support.

1 Introduction

Nonparametric prediction of a random variable Y conditional on the value of an explanatory variable X is a classical and important problem in Statistics. In a recent paper, Carroll et al. (2009) - further referred to as CDH - consider the problem of predicting the random variable Y via the estimation of $\mu(t) = E(Y|T = t)$. In their setting, T is an observed "future" explanatory variable generated by $T = X + U^F$ where X is the true unobserved explanatory variable and U^F is a measurement error. The prediction problem is complicated by the fact that "past" observations $\{(Y_j, W_j)\}_{j=1}^n$ are such that $W_j = X_j + U_j$ with measurement errors U_j that are different from U^F. Moreover, the U_j themselves may have different distributions. In this context, CDH have suggested a kernel deconvolution estimator for $\mu(t)$ and obtained its consistency.

In this paper, we make a number of contributions to the asymptotic characterization of the CDH estimator. First, using a family of kernels proposed in Mynbaev and Martins-Filho (2010), we define a class of kernel deconvolution estimators that contains the CDH estimator. Viewing the CDH estimator as a member of this class is illuminating as we are able to obtain consistency of its elements under conditions that are much less stringent than those required by Carroll et al. (2009). In particular, the assumption that the Fourier transform (usually, the characteristic function) of the underlying kernel is compactly supported, which is prevalent in the deconvolution estimation literature, is entirely relaxed. We describe the properties and show how to construct these non-compactly supported Fourier transforms of kernels K that satisfy $\int_{\mathbb{R}} |K(x)|dx < \infty$. In addition, we lift two other assumptions used by CDH: i) we do not require that the underlying kernels be of higher-order (r) to obtain a suitable bound for the bias of the estimators; and ii) we make no assumptions on the smoothness of the underlying regression $g(x) \equiv E(Y|X = x)$, nor do we require the density of X to be continuously differentiable of order r. We only require fractional smoothness of the density f_T of T. This is accomplished simply by restricting f_T and μ to belong to suitably defined Besov spaces.

The second contribution of this paper is to establish the asymptotic normality of the elements in our new class of kernel deconvolution estimators. Asymptotic normality of kernel deconvolution estimators for density and regression has been obtained in various contexts. For density estimators see, *inter alia*, Zhang (1990), Fan (1991), Fan (1992), Masry (1993), Fan and Liu (1997), Cator (2001), van Es and Uh (2004) and van Es and Uh (2005). For regression, see Fan et al. (1990), Fan and Masry (1992), Delaigle et al. (2009) and Honda (2010). In addition, nonparametric deconvolution estimation of density and regression under heterogeneous measurement errors has been considered by Delaigle and Meister (2007, 2008) and Meister (2009).

However, there is no asymptotic distributional result for the CDH estimator. This is unfortunate as applied researchers will be interested in not only reporting point estimates for their forecast, but also constructing suitably chosen confidence bands. In our study of asymptotic normality, we focus on the case where the error distributions are super-smooth.[1] In this case, for deconvolution density estimators, asymptotic normality was proved by Zhang (1990) and Fan (1991). However, in these papers the estimator is normalized by a random quantity whose asymptotic behavior is not clear. van Es and Uh (2005) derived an explicit (nonstochastic) asymptotic characterization of the normalizing quantity in terms of the sample size n and bandwidth h used in kernel estimation. Their result was obtained under a specific restriction on the Fourier transform of the kernel, which was shown to be essential in van Es and Uh (2004).

Our method of proof shows that nonparametric deconvolution prediction estimators in our proposed class (including the CDH estimator), despite being much more complex than the deconvolution density estimator considered by van Es and Uh (2005), can be studied using their main ideas. Our errors' structure has the super-smooth property assumed by van Es and Uh (2005) and we adopt the condition they impose on the (seed) kernel and its Fourier transform (see our Assumption 3.3). However, our work is significantly complicated by the need to account for measurement error heterogeneity, a feature they avoid. The Cauchy boundary effect, discovered in van Es and Uh (2004), also manifests itself in our work. Our asymptotic normality study reveals that there are many different cases to be considered for super-smooth errors, and a comprehensive approach demands a much longer article. Here, we develop a framework that allows us to address the peculiarities of the estimators we propose, as well as the CDH estimator, in a subset of such

[1]See Fan and Truong (1993) for a typology of error distributions and examples.

cases and sheds light on how other cases could be handled.

Lastly, we conduct a Monte Carlo study to shed light on the finite sample performance of the CDH estimator relative to other estimators in the class constructed with the kernels proposed in Mynbaev and Martins-Filho (2010). The simulation results seem to indicate improved performance, measured by mean squared error, when the kernels proposed in Mynbaev and Martins-Filho are used.

The remainder of the paper is organized as follows. Section 2 is devoted to establishing the consistency of estimators in the class and section 3 covers asymptotic normality of our estimators. The two convergences require somewhat different approaches and conditions. In particular, while in section 2 we dispense with the assumption that the Fourier transform of the kernel has compact support, in section 3 we have to impose it, for the method of van Es and Uh (2005) to apply. Section 4 describes the Monte Carlo study and discusses the results. Section 5 contains some summary remarks. All proofs are relegated to an appendix.

2 Consistency

Throughout the paper we adopt the following notation. For $p \geq 1$, L_p denotes the space of p-integrable functions g on $\mathbb{R}$ with norm $\|g\|_p = \left(\int_{\mathbb{R}} |g(x)|^p dx\right)^{1/p}$. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, whenever it exists. Thus, if $g \in L_1$, $\mathcal{F}_g(t) = \int_{\mathbb{R}} e^{itx} g(x)dx$ for $t \in \mathbb{R}$. Furthermore, if g is bounded and continuous for some $x \in \mathbb{R}$ and $\mathcal{F}_g(t) \in L_1$, then $\mathcal{F}^{-1}_{\mathcal{F}_g}(x) \equiv g(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\mathcal{F}_g(t)dt$. If g is a density function, $\mathcal{F}_g$ is called a characteristic function and we write $\mathcal{F}_g = \phi_g$. The letter c, with or without subscripts, denotes various inconsequential constants.

2.1 The CDH model and estimator

Let $\{Y_j, W_j\}_{j=1}^n$ be a sample of independent observations from $Y_j = g(X_j) + \varepsilon_j$ and $W_j = X_j + U_j$. W_j represents a version of X_j which is contaminated by the unobserved error U_j. The unobserved X_j have a common density denoted by f_X, the measurement errors U_j are heterogeneously distributed with known densities denoted by f_{U_j} and $\{X_j, U_j, \varepsilon_j\}_{j=1}^n$ are assumed to be mutually independent. $\{\varepsilon_j\}_{j=1}^n$ are unobserved independent disturbances assumed to have zero means. Out-of-sample observations on X are also observed with error and are generated by $T = X + U^F$. X, U^F and ε_j are independent and the density

f_{U^F} of U^F is assumed to be known.

Let f_{TY} denote the joint density of T and Y, f_T denote the marginal density of T and $d_T(x) = \int y f_{TY}(x,y)dy$. Then, nonparametric prediction of Y is attained by estimating

$$\mu(x) \equiv E(Y|T=x) = \frac{d_T(x)}{f_T(x)}.$$

To define the CDH estimator $\tilde{\mu}(x)$, first let

$$\Psi_j(t) = \frac{\phi_{f_{U_j}}(-t)}{\sum_{k=1}^n \left|\phi_{f_{U_k}}(t)\right|^2}. \tag{2.1}$$

We call a kernel any integrable function K on $\mathbb{R}$ such that $\int_{\mathbb{R}} K(t)dt = 1$ and let

$$K_{T,j,h}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\mathcal{F}_K(t)\phi_{f_{U^F}}\left(\frac{t}{h}\right)\Psi_j\left(\frac{t}{h}\right)dt$$

where $h > 0$ is a bandwidth. For this integral and for (2.6) below to exist we assume that K is such that for each $h > 0$

$$\mathcal{F}_K(t)\phi_{U^F}\left(\frac{t}{h}\right)\Psi_j\left(\frac{t}{h}\right) \in L_1 \cap L_2, \text{ for } j = 1, ..., n. \tag{2.2}$$

The estimator of f_T is defined by $\tilde{f}_T(x) = \frac{1}{h}\sum_{j=1}^n K_{T,j,h}\left(\frac{x-W_j}{h}\right)$. For notational simplicity we keep only the variable parameters n and j (T is fixed and it is well understood that the bandwidth h depends on n, i.e., $h = h_n$). Therefore, we write

$$\tilde{f}_T(x) = \sum_{j=1}^n f_{n,j}(x), \quad f_{n,j}(x) \equiv \frac{1}{2\pi h}\int_{\mathbb{R}} \exp\left(it\frac{W_j - x}{h}\right)\mathcal{F}_K(t)\phi_{f_{U^F}}\left(\frac{t}{h}\right)\Psi_j\left(\frac{t}{h}\right)dt. \tag{2.3}$$

The estimator of d_T is defined by $\tilde{d}_T(x) = \frac{1}{h}\sum_{j=1}^n Y_j K_{T,j,h}\left(\frac{x-W_j}{h}\right)$. Using the notation in (2.3) we write

$$\tilde{d}_T(x) = \sum_{j=1}^n d_{n,j}(x), \text{ where } d_{n,j}(x) = Y_j f_{n,j}(x). \tag{2.4}$$

and

$$\tilde{\mu}(x) = \frac{\tilde{d}_T(x)}{\tilde{f}_T(x)} = \frac{\sum_{j=1}^n d_{n,j}(x)}{\sum_{j=1}^n f_{n,j}(x)}. \tag{2.5}$$

For this model CDH have established, under their assumptions (4.1)-(4.4), that for each x such that $f_T(x) > 0$, $\tilde{\mu}(x) = \mu(x) + O_p\left(\left(\frac{v(h)}{n}\right)^{1/2} + h^r\right)$, provided that as $n \to \infty$, $h \to 0$ and $v(h)/n \to 0$, where

$$v(h) = \frac{n}{h}\int \left|\phi_K(t)\phi_{f_{U^F}}\left(\frac{t}{h}\right)\right|^2 / \sum_{k=1}^n \left|\phi_{f_{U_k}}\left(\frac{t}{h}\right)\right|^2 dt. \tag{2.6}$$

Here, r represents the order of the kernel K and the number of bounded derivatives that f_X and g are assumed to possess. Their proof relies on Taylor's Theorem and establishing

$$(a)\, E\tilde{f}_T(x) = \int K(z) f_T(x-hz)dz, \quad (b)\, E\tilde{d}_T(x) = \int K(z) d_T(x-hz)dz, \tag{2.7}$$

and

$$(a)\, V(\tilde{f}_T(x)) = O(v(h)/n), \quad (b)\, V(\tilde{d}_T(x)) = O(v(h)/n). \tag{2.8}$$

Our Lemmas 1 and 2 (see the Appendix) establish (2.7) and (2.8). We provide full proofs to make explicit what assumptions from CDH we are able to omit without loss. In particular, for the order of the variances we extract from their conditions (4.1)-(4.4) only

Assumption 2.1. $\sup_j \max\left\{\left\|f_{W_j}\right\|_\infty, \left\|f_{U_j}\right\|_\infty, EY_j^2\right\} < \infty$.

It follows from (2.8)(b) that if $v(h)/n = O(1)$ and $\int K(z) d_T(x-hz)dz = O(1)$, then $\tilde{d}_T(x) = O_p(1)$.

2.2 A class of estimators for $\mu(x)$

In this section we start by defining a class of estimators for $\mu(x)$ using the family of kernels $\{M_k\}_{k=1,2,\cdots}$ introduced by Mynbaev and Martins-Filho (2010). We need a series of definitions that support the construction of the class and the consistency result we obtain for its elements.

Let $C_{2k}^l = \frac{(2k)!}{(2k-l)!l!}$ for $l = 0, 1, \cdots, 2k$ be the binomial coefficients. For a kernel K and a natural number k, Mynbaev and Martins-Filho (2010) define a class of kernels

$$M_k(x) = -\frac{1}{C_{2k}^k} \sum_{|l|=1}^{k} (-1)^l C_{2k}^{l+k} \frac{1}{|l|} K\left(\frac{x}{l}\right),$$

and call K a seed kernel for M_k. The kernels $\{M_k\}_{k=1,2,\cdots}$ define a class of estimators indexed by k, and in this context we define

$$\hat{\mu}_k(x) = \frac{\hat{d}_{T,k}(x)}{\hat{f}_{T,k}(x)} \text{ for } k = 1, 2, \cdots$$

where

$$\hat{f}_{T,k}(x) = \sum_{j=1}^{n} f_{n,j,k}(x), \quad f_{n,j,k}(x) \equiv \frac{1}{2\pi h} \int_{\mathbb{R}} \exp\left(it\frac{W_j - x}{h}\right) \mathcal{F}_{M_k}(t) \phi_{f_{U^F}}\left(\frac{t}{h}\right) \Psi_j\left(\frac{t}{h}\right) dt \tag{2.9}$$

and

$$\hat{d}_{T,k}(x) = \sum_{j=1}^{n} d_{n,j,k}(x), \quad d_{n,j,k}(x) \equiv Y_j f_{n,j,k}(x). \tag{2.10}$$

We note that

$$\mathcal{F}_{M_k}(t) = \sum_{l=1}^{k} \lambda_{k,l}(\mathcal{F}_K(tl) + \mathcal{F}_K(-tl)) = \sum_{l=1}^{k} 2\lambda_{k,l}\mathcal{F}_K(tl) \tag{2.11}$$

where $\lambda_{k,l} = (-1)^{l+1}(k!)^2/(k+l)!(k-l)!$ and the second equality holds if K is symmetric. Also, if K is symmetric and $k = 1$ we have $\tilde{\mu}(x) = \hat{\mu}_1(x)$. Thus, the CDH estimator is a member of the class of estimators we consider.

2.3 Main results

The properties of nonparametric deconvolution estimators are traditionally obtained by imposing assumptions on the smoothness of the relevant underlying densities and regressions. As pointed out by Mynbaev and Martins-Filho (2010), smoothness can be controlled by finite differences, which can be forward, backward or centered.[2] A centered even-order difference of a function f is defined by

$$\Delta_h^{2k} f(x) = (-1)^k \sum_{l=-k}^{k} (-1)^l C_{2k}^{l+k} f(x+lh), \quad h \in \mathbb{R}.$$

The M_k is designed in such a way as to have the following integral representation for the bias (Mynbaev and Martins-Filho, 2010, equation (22))

$$E\hat{f}_{T,k}(x) - f_T(x) = \frac{(-1)^{k+1}}{C_{2k}^k} \int K(t)\Delta_{-ht}^{2k} f_T(x)dt. \tag{2.12}$$

Next we describe the smoothness characteristic to be used with this representation and the function spaces that contain μ and f_T. A forward even-order difference is defined by

$$\tilde{\Delta}_h^{2k} f(x) = (-1)^k \sum_{l=-k}^{k} (-1)^l C_{2k}^{l+k} f(x+kh+lh).$$

The Sobolev space $W_p^r(\mathbb{R})$, where $1 \le p \le \infty$ and r is a positive integer, is defined as the set of functions on $\mathbb{R}$ with an absolutely continuous derivative $f^{(r-1)}$ and finite norm $\|f\|_{W_p^r} = \left\|f^{(r)}\right\|_p + \|f\|_p$. In this norm, the first part $\left\|f^{(r)}\right\|_p$ characterizes smoothness of f. Now let $1 \le p \le \infty$, $1 \le q \le \infty$, $r > 0$ (r can take

[2] See Besov et al. (1978); Triebel (1983) for comprehensive treatments.

positive fractional values). We need Besov spaces $B^r_{p,q}(\mathbb{R})$, for which the smoothness characteristic is defined by

$$\|f\|_{b^r_{p,q}} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} \left| \tilde{\Delta}^{2k}_h f(x) \right|^p dx \right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}.$$

Here, k is any positive integer satisfying $2k > r$ and in case $p = \infty$ and/or $q = \infty$ the integral(s) is (are) replaced by sup. The norm in $B^r_{p,q}(\mathbb{R})$ is defined by $\|f\|_{B^r_{p,q}} = \|f\|_{b^r_{p,q}} + \|f\|_p$. All these norms with different k are equivalent to one another by Theorem 2.5.13(i) in Triebel (1983). With a translation operator τ_h defined by $(\tau_h f)(x) = f(x+h)$ it is easy to see that $\tilde{\Delta}^{2k}_h f(x) = \Delta^{2k}_h [(\tau_{kh} f)(x)]$. Therefore, for a smoothness characteristic, the centered even-order difference can be used and we write

$$\|f\|_{b^r_{p,q}} = \left\{ \int_{\mathbb{R}} \left[\frac{\left(\int_{\mathbb{R}} \left| \Delta^{2k}_h f(x) \right|^p dx \right)^{1/p}}{|h|^r} \right]^q \frac{dh}{|h|} \right\}^{1/q}.$$

We will also need a Zygmund space $Z^r(\mathbb{R})$ which, by definition, is precisely a Besov space $B^r_{\infty,\infty}(\mathbb{R})$. By Corollary 2.8.2(i) in Triebel (1983) the multiplication by a function $\mu \in Z^\rho(\mathbb{R})$ is bounded in $B^r_{p,q}(\mathbb{R})$ if $\rho > r$. That is,

$$\|\mu f\|_{B^r_{p,q}} \leq c \|\mu\|_{Z^\rho} \|f\|_{B^r_{p,q}}. \tag{2.13}$$

Direct verification that $f \in B^r_{p,q}(\mathbb{R})$ in practice may be difficult. Relationships between different functional spaces may simplify this task. For example, for a natural r one has $W^r_p(\mathbb{R}) \subset B^r_{p,\infty}(\mathbb{R})$, see section 6.2 in Nikolskii (1977).

The following assumption restricts the density f_T and μ to belong to suitably indexed Besov and Zygmund spaces. In addition, a restriction is placed on the seed kernel used to construct the class $\{M_k\}_{k=1,2,\ldots}$. As will be transparent in the remarks following Theorem 1, Assumption 2.2 is much less demanding than those required in Carroll et al. (2009).

Assumption 2.2. *$f_T \in B^r_{\infty,q}$ with some $r > 0$ and $1 \leq q \leq \infty$, $\mu \in Z^\rho(R)$ with some $\rho > r$ and for a seed kernel K*

$$\left(\int |K(t)|^{q'} |t|^{(r+1/q)q'} dt \right)^{q'} < \infty \tag{2.14}$$

where $1/q + 1/q' = 1$.

We now state our main result regarding consistency.

Theorem 1. *If Assumptions 2.1-2.2 hold, for all x such that $f_T(x) > 0$, we have*

1. $\hat{\mu}_k(x) - \mu(x) = O_p\left[h^r + \left(\sum_{l=1}^{k} v(hl)/n\right)^{1/2}\right]$,

2. *if* $h \to 0$ *and* $v(h)/n \to 0$ *as* $n \to \infty$, $\hat{\mu}_k(x) - \mu(x) = o_p(1)$ *for all* $k = 1, 2, \cdots$.

The conclusions of Theorem 1 are basically those of Theorem 4.1 in CDH (take $k = 1$), but there are important differences in our assumptions.

Remarks: 1. Since $EY_j^2 = Eg^2(X_j) + E\varepsilon_j^2$, Assumption 2.1 implies $\sup_j E\varepsilon_j^2 < \infty$ as in condition (4.1) in CDH.

2. The part of their condition (4.2), on the order of the kernel (the seed kernel in our case), is substantially relaxed by our lighter requirement (2.14).

3. One of the requirements in condition (4.3) in CDH is that the regression g and f_X have r bounded derivatives. We impose no smoothness on g. Regarding f_X, we note that given $T = X + U^F$ and independence of X and U^F, the density of T is $f_T(x) = f_X * f_{U^F}(x)$, the convolution of f_X and f_{U^F}. Since $f_{U^F} \in L_1$ and f_X has r bounded derivatives, $f_T(x)$ has bounded derivatives of order r. We, in turn, require only fractional smoothness of f_T, i.e., $f_T \in B^r_{\infty,q}$ for $r > 0$ and μ slightly smoother than f_X, i.e., $\rho > r$.

2.4 Asymptotic behavior of $v(h)/n$

The consistency of $\hat{\mu}_k(x)$ depends on $v(h)/n \to 0$ as $n \to \infty$. The standard approach in the nonparametric deconvolution literature to verify that $v(h)/n \to 0$ is to assume that $\mathcal{F}_K$ has compact support. For example, this is assumed in condition (4.2) in CDH. In what follows, we relax the requirement that $\mathcal{F}_K$ is compactly supported and show that $v(h)$ is finite for positive h or $v(h)/n \to 0$, with a suitably chosen h. To our knowledge, this is the first general study of the asymptotic behavior of $v(h)/n$. The applicability of our result extends beyond the model we consider to various nonparametric deconvolution estimators. We start with the following assumption.

Assumption 2.3. *For any compact set $\mathcal{K} \subset \mathbb{R}$, $\sup_{s\in\mathcal{K}} \Phi_n(s) < \infty$, where $\Phi_n(s) = \frac{n\left|\phi_{f_{U^F}}(s)\right|^2}{\sum_{k=1}^n \left|\phi_{f_{U_k}}(s)\right|^2}$. In addition, $\Phi_n(s)$ has a majorant P in the neighborhood of infinity such that:*

(a) for some positive c_1, c_2 one has $\Phi_n(s) \le c_2 P(s)$ for all $|s| \ge c_1$,

(b) $P(s) = P(-s)$ and for some $c_3 > 0$, $P(s) \le c_3 P'(s)$ holds for all $s \ge c_1$,

(c) $\int_{c_1}^{\infty} \exp(-P(s)) \left(1 + \left|P^{(1)}(s)\right|^2\right) ds < \infty$,

(d) $J(h) \equiv \int_{P(c_1)}^{\infty} \exp\left[-P(hP^{-1}(t))\right] dt < \infty$ for all $0 < h < 1$.

The inverse of P in (d) exists because from (b) it follows that P is strictly increasing on $[c_1, \infty)$, with P^{-1} defined on $[P(c_1), \infty)$. $\Phi_n(s) \le c_2 P(s)$ and $P(s) \le c_3 P'(s)$ can be replaced by their consequence $\Phi_n(s) \le cP^{(1)}(s)$ and still provide enough structure for our applications. For transparency, we prefer to keep the two inequalities. Examples of functions P are $P(s) = \exp(s^\alpha)$ and iterated exponential functions $P_1(s) = e^s$, $P_2(s) = P_1(P_1(s))$, ..., $P_n(s) = P_1(P_{n-1}(s))$. Iterated exponential functions form a scale that covers all imaginable errors. Note that $J(h)$ is monotonic and therefore it is bounded from above when h is bounded away from zero.

Remark: Assumption 2.3 has been developed with growing Φ_n in mind, because the case of a bounded Φ_n is simpler. If $\phi_K \in L_2$ and there is a constant c such that $|\Phi_n(t)| \le c$ for all large n, then it is easy to see that

$$|v(h)| = \left| \int |\phi_K(ht)|^2 \Phi_n(t) dt \right| \le c\frac{1}{h} \int |\phi_K(ht)|^2 d(ht) = c \left\|\phi_K\right\|_{L_2}^2 /h. \tag{2.15}$$

This bound can be made more precise with the help of Theorem 2 in Mynbaev (2012). Namely, if (a) $|\phi_K(t)|^2$ is even and integrable, (b) there exists a bounded continuous ϕ such that $|\Phi_n(t)| \le \phi(t)$ for all large n and the limit $M(\phi) = \lim_{r\to\infty} \frac{1}{2r} \int_{-r}^{r} \phi(t)dt$ exists, then for any $\varepsilon > 0$ there exists h_0 such that for $h \ge h_0$ one has $|v(h)| \le (M(\phi) + \varepsilon) \left\|\phi_K\right\|_{L_2}^2 /h$. This bound is more precise than (2.15) because $M(\phi) \le \sup\phi$. In particular, if ϕ is integrable, then $M(\phi) = 0$.

Theorem 2. *Under Assumption 2.3,*

1. *There exists a function K such that $K \in L_1$ and $\int K(x)dx = 1$, the support of $\mathcal{F}_K$ is not compact and $v(h) < \infty$ for all $0 < h < 1$.*

2. The kernel K from part 1 satisfies $v(h)/n = o(1)$ with suitably chosen $h = h_n$.

Theorem 2 is an existence result. It shows that there exist a rich profusion of well behaved kernels K with Fourier transforms that are not compactly supported for which $v(h)/n = o(1)$. We note that the computation of kernel deconvolution estimators does not require per se a kernel. Rather, what is needed is a Fourier transform which can be associated with a well behaved K. We now give an example of how Theorem 2 can be used when the measurement errors are super-smooth. The example relies on Assumptions 3.1 and 3.2 in section 3 below.

Example. Under Assumptions 3.1 and 3.2 in section 3, we have

$$\begin{aligned}\Phi_n(t) &\sim \frac{A_3^2 |t|^{2a_3} \exp\left(-2\mu_3 |t|^{\lambda_3}\right)}{\delta A_1^2 |t|^{2a_1} \exp\left(-2\mu_1 |t|^{\lambda_1}\right) + (1-\delta) A_2^2 |t|^{2a_2} \exp\left(-2\mu_2 |t|^{\lambda_2}\right)} \\ &\sim \frac{A_3^2}{(1-\delta)A_2^2} |t|^{2(a_3-a_2)} \exp\left(2\mu_2 |t|^{\lambda_2} - 2\mu_3 |t|^{\lambda_3}\right).\end{aligned}$$

For any $\varepsilon > 0$, this is dominated by $P(t) = \exp\left(c\,|t|^{\lambda_2}\right)$, where $c = 2\mu_2 + \varepsilon$ (increasing c_2, we can achieve $c_1 = 1$). Conditions (b) and (c) from Assumption 2.3 are satisfied given Assumption 3.1(2):

$$P^{(1)}(s) = P(s) c\lambda_2 s^{\lambda_2 - 1} \geq c\lambda_2 P(s), \ \ s \geq 1,$$

$$\int_1^\infty \exp\left[-\exp\left(cs^{\lambda_2}\right)\right] \left\{1 + \left[\exp\left(cs^{\lambda_2}\right) c\lambda_2 s^{\lambda_2-1}\right]^2\right\} ds < \infty.$$

As for (d), we note that $P^{-1}(t) = [(\ln t)/c]^{1/\lambda_2}$ and $P(c_1) = e^c$. Thus,

$$J(h) = \int_{e^c}^\infty \exp\left(-\exp\left(c\left|h\left[(\ln t)/c\right]^{1/\lambda_2}\right|^{\lambda_2}\right)\right) dt = \int_{e^c}^\infty \exp\left(-\exp\left(h^{\lambda_2} \ln t\right)\right) dt = \int_{e^c}^\infty \exp\left(-t^{h^{\lambda_2}}\right) dt.$$

Since $\exp\left(t^{h^{\lambda_2}}\right)$ grows faster than any positive power of t, by L'Hôpital's rule the last integral converges for any $0 < h < 1$. By Theorem 2 we can put

$$\mathcal{F}_K(s) = \exp\left(-\exp\left(c\,|s|^{\lambda_2}/2\right)\right), \ |s| \geq 1, \ \mathcal{F}_K(s) = 1 - as^2, \ |s| < 1.$$

a is defined from the sewing condition $1 - a = \mathcal{F}_K(1-0) = \mathcal{F}_K(1+0) = \exp(-\exp(c/2))$, i.e., $a = 1 - \exp(-\exp(c/2))$. It is unnecessary to impose a sewing condition for $\mathcal{F}_K^{(1)}$ at $s = \pm 1$ because all that is needed is the existence of $\mathcal{F}_K^{(1)}$ almost everywhere and its square-integrability.

3 Asymptotic normality

In this section we give sufficient conditions for the estimators $\hat{\mu}_k(x)$ for $k = 1, 2, \cdots$ to converge in distribution. We consider the case of only two different types of measurement errors for the "past" observations on X but allow for the measurement error for the "future" observation on X to differ from both. By imposing conditions on the parameters of the measurement error distributions we are able to obtain asymptotic normality of all $\hat{\mu}_k(x)$ under suitable centering and a non-random normalization that depends only on n and h. The practical relevance of a model with two types of measurement error is discussed in Carroll et al. (2006) and Carroll et al. (2009) where empirical applications are given. The following assumption describes the restrictions we will place on the measurement errors.

Assumption 3.1. *(1) There are only two different measurement errors,*

$$\begin{aligned} \{U_j\}_{j=1}^{m} &\text{ are identically distributed as the random variable } V_1, \\ \{U_j\}_{j=m+1}^{n} &\text{ are identically distributed as the random variable } V_2. \end{aligned} \tag{3.1}$$

The out-of-sample error is denoted $V_3 = U^F$. ϕ_j denotes the characteristic function of V_j and all three errors are assumed super-smooth:

$$\phi_j(t) = (1 + o(1))A_j\, |t|^{a_j} \exp(-\mu_j\, |t|^{\lambda_j}), \quad j = 1, 2, 3; \quad |t| \to \infty, \tag{3.2}$$

where λ_j and μ_j are positive, A_j and a_j are real and $A_j \neq 0$. All error densities are assumed symmetric and therefore their characteristic functions are real and symmetric (although this symmetry condition is not critical).

(2) $\lambda_1 > \lambda_2 > \max\{1, \lambda_3\}$.

Part (2) of Assumption 3.1 excludes a number of cases and we now comment on what is left out. Regarding the past errors, in the super-smooth case we have the following possibilities: the cases $\lambda_1 \lesseqgtr \lambda_2$, $\mu_1 \lesseqgtr \mu_2$, $a_1 \lesseqgtr a_2$, $A_1 \lesseqgtr A_2$ can be combined giving a total of 3^4 combinations. Some of these combinations are trivial. For example, if $\lambda_1 > \lambda_2$, then, because of the dominance of the exponential functions, the relationships between the remaining parameters don't matter. Some combinations can be reduced to others by changing notation. For instance, the case $\lambda_1 = \lambda_2$, $\mu_1 > \mu_2$ is the same as $\lambda_1 = \lambda_2$, $\mu_1 < \mu_2$ up to the notation. Still, by assuming (I) $\lambda_1 > \lambda_2$, we leave out at least the cases

$$\begin{aligned}\text{(II) } \lambda_1 &= \lambda_2, \mu_1 > \mu_2, \quad \text{(III) } \lambda_1 = \lambda_2, \mu_1 = \mu_2, a_1 < a_2, \\ \text{(IV) } \lambda_1 &= \lambda_2, \mu_1 > \mu_2, a_1 = a_2, A_1 \neq A_2.\end{aligned}$$

Case (I) is subdivided further as

$$\text{(Ia) } \lambda_1 > \lambda_2 > \lambda_3, \quad \text{(Ib) } \lambda_1 > \lambda_3 \geq \lambda_2, \quad \text{(Ic) } \lambda_3 \geq \lambda_1 > \lambda_2.$$

We exclude (Ib) and (Ic). Further, the analysis of the estimators $\hat{\mu}_k$, $k = 1, 2, \cdots$ in light of van Es and Uh (2004) shows that under condition (Ia) the cases

$$\text{(i) } \lambda_2 > 1, \quad \text{(ii) } \lambda_2 = 1 \text{ and (iii) } 0 < \lambda_2 < 1$$

are conceptually different. Part (2) of Assumption 3.1 is an intersection of (Ia) and (i).

A full treatment of all possibilities requires a much longer article and the development of alternative methods to handle the asymptotics. However, our results cover a number of interesting situations. For example, in meta-analysis (Walter (1997), Delaigle and Meister (2007)), where samples are obtained by combining data from different studies, it may be unrealistic to assume that V_1 and V_2 come from the same distribution (heterogeneity) or even the same family of distributions ($\lambda_1 \neq \lambda_2$). For linear regression models this possibility is considered by Cheng and Riu (2006) (section 3) where, in our notation, V_1 and V_2 are allowed to have different variances and come from arbitrary families of distributions. Another setting where our results are applicable occurs when the sample used in estimation combines groups of observations subject to different measurement error. See, for example, Altman and Bland (1983) where variables of interest in medicine (blood pressure, cardiac stroke volume) are measured with different instruments or techniques and are subject to non-homogeneous contamination.

In many instances (Kulathinal et al. (2002); Staudenmayer et al. (2008)) it is common to assume that the measurement errors come from the same family of distributions with possibly different parameters. For example, when normality (centered at 0) is assumed, in our typology, we have $\lambda_1 = \lambda_2$ and cases II, III or IV. In this case the change in the asymptotic representation of the product $\phi_3\Psi_j$ (see equation 3.5) would lead to drastic changes in the proof of the asymptotic normality. Finally, note that an informal classification, based on similarity of asymptotic distributions, would be preferable to the formal classification that utilizes relationships between parameters.

Assumption 3.2. *The ratio* $\delta_n = m/n$ *stabilizes at some value* $\delta \in (0,1)$*:* $\lim_{n\to\infty} \delta_n = \delta$.

The situation when one error type prevails in the limit would not be interesting, as then the classical Nadaraya-Watson estimator would likely perform better than the CDH estimator, see (Carroll et al., 2009, Remark 3.1). As a result of Assumption 3.2, the numbers m, $n-m$ and n tend to infinity at the same rate.

To reflect the eternal presence of two past error types, we split (2.3) and (2.4) as

$$\hat{f}_{T,k}(x) = \hat{f}^{\mathrm{I}}_{T,k}(x) + \hat{f}^{\mathrm{II}}_{T,k}(x), \quad \hat{d}_{T,k}(x) = \hat{d}^{\mathrm{I}}_{T,k}(x) + \hat{d}^{\mathrm{II}}_{T,k}(x)$$

where

$$\begin{aligned} \hat{f}^{\mathrm{I}}_{T,k}(x) &= \sum_{j=1}^{m} f_{n,j,k}(x), \quad \hat{f}^{\mathrm{II}}_{T,k}(x) = \sum_{j=m+1}^{n} f_{n,j,k}(x), \\ \hat{d}^{\mathrm{I}}_{T,k}(x) &= \sum_{j=1}^{m} d_{n,j,k}(x), \quad \hat{d}^{\mathrm{II}}_{T,k}(x) = \sum_{j=m+1}^{n} d_{n,j,k}(x). \end{aligned} \tag{3.3}$$

The asymptotic behavior of the product $\phi_3\Psi_j$ at infinity is of crucial importance for asymptotic calculations. By (2.1) and Assumption 3.1 for $j = 1, ..., m$ we have

$$\begin{aligned} (\phi_3\Psi_j)(t) &= \frac{\phi_3(t)\phi_1(-t)}{m\,|\phi_1(t)|^2 + (n-m)\,|\phi_2(t)|^2} \\ &\quad (\lambda_1 > \lambda_2 \text{ implies } \phi_1(t)/\phi_2(t) = o(1), \quad |t| \to \infty) \\ &= (1+o(1))\frac{1}{n-m}\frac{\phi_3(t)\phi_1(-t)}{|\phi_2(t)|^2} \\ &\quad \text{(applying (3.2))} \\ &= \frac{1+o(1)}{n-m}\frac{A_1A_3}{A_2^2}\,|t|^{a_1+a_3-2a_2}\,\frac{\exp(2\mu_2\,|t|^{\lambda_2})}{\exp(\mu_1\,|t|^{\lambda_1} + \mu_3\,|t|^{\lambda_3})}. \end{aligned} \tag{3.4}$$

Similarly, for $j = m+1, ..., n$

$$
\begin{aligned}
(\phi_3 \Psi_j)(t) &= \frac{\phi_3(t)\phi_2(-t)}{m\,|\phi_1(t)|^2 + (n-m)\,|\phi_2(t)|^2} = \frac{1+o(1)}{n-m}\frac{\phi_3(t)}{\phi_2(t)} \\
&= \frac{1+o(1)}{n-m}\frac{A_3}{A_2}\,|t|^{a_3-a_2}\,\frac{\exp(\mu_2\,|t|^{\lambda_2})}{\exp(\mu_3\,|t|^{\lambda_3})}.
\end{aligned} \tag{3.5}
$$

In general, each of the possibilities

$$
|(\phi_3 \Psi_j)(t)| \to \infty, \quad |(\phi_3 \Psi_j)(t)| \to \mathit{const} \neq 0, \quad |(\phi_3 \Psi_j)(t)| \to 0, \quad |t| \to \infty,
$$

for the first set of errors ($j = 1, ..., m$) can be combined with similar possibilities for the second set ($j = m+1, ..., n$). From (3.4), (3.5) and Assumption 3.1 one has

$$
|(\phi_3 \Psi_j)(t)| \to 0, \quad j = 1, ..., m, \quad |(\phi_3 \Psi_j)(t)| \to \infty, \quad j = m+1, ..., n.
$$

Our method applies to all other combinations of 0 and ∞ but we have no results for the case $|(\phi_3 \Psi_j)(t)| \to \mathit{const} \neq 0$. The next assumption is taken from van Es and Uh (2005).

Assumption 3.3. *The kernel K in* (2.3) *is symmetric and its Fourier transform $\mathcal{F}_K$ is supported on* $[-1, 1]$. *With some constants A (real) and α (nonnegative) $\mathcal{F}_K$ satisfies*

$$
\mathcal{F}_K(1-t) = At^{\alpha} + o(t^{\alpha}), \quad t \downarrow 0.
$$

Two of the most used kernels in nonparametric deconvolution estimation satisfy this condition. They are the sinc kernel $K(x) = \frac{\sin x}{\pi x}$ (where $\alpha = 0$, $A = 1$) with $\mathcal{F}_K(t) = I_{[-1,1]}(t)$ (an indicator of the segment $[-1, 1]$) and $K(x) = \frac{48 \cos x}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144 \sin x}{\pi x^5}\left(2 - \frac{5}{x^2}\right)$ (where $\alpha = 3$, $A = 8$) with $\mathcal{F}_K(t) = (1-t^2)^3 I_{[-1,1]}(t)$, which we use in our simulations. Owing to this assumption, the symmetry of $\mathcal{F}_K$, ϕ_3, Ψ_j and from (2.9), (2.11) we have

$$
f_{n,j,k}(x) = \sum_{l=1}^{k} \frac{2\lambda_{k,l}}{\pi h l} \int_0^1 \cos\left(t\frac{W_j - x}{lh}\right) \mathcal{F}_K(t)(\phi_3 \Psi_j)\left(\frac{t}{lh}\right) dt. \tag{3.6}
$$

The following assumption is needed to verify Lyapunov's condition when applying the Central Limit Theorem.

Assumption 3.4. *For some $\kappa > 0$, $\sup_j E\,|Y_j|^{2+\kappa} < \infty$, $\sup_j E\,|W_j|^{(2+\kappa)/\kappa} < \infty$ and for some $c > 0$, $c < E\varepsilon_j^2 < \infty$ for all j.*

We are now ready to state our main result on asymptotic normality of $\hat{\mu}_k$.

Theorem 3. *Suppose Assumptions 3.1-3.4 hold, and fix x such that $f_T(x) \neq 0$. Let $\zeta(h) = \exp(\mu_2 h^{-\lambda_2})$ and $M_{n,h} = \frac{h^{\lambda_2(\alpha+1)+a_2-a_3-1}\zeta(h)}{\sqrt{n-m}}$. Then, if $M_{n,h} \to 0$ and $h \to 0$ as $n \to \infty$ we have*

$$\frac{1}{M_{n,h}}\left(\hat{\mu}_k(x) - \frac{E\hat{d}_{T,k}(x)}{E\hat{f}_{T,k}(x)}\right) \xrightarrow{d} N\left(0, \left(\frac{2kB_0}{\pi(k+1)}\right)^2 \frac{\mu^2(x)\sigma_{11} - 2\mu(x)\sigma_{12} + \sigma_{22}}{f_T^2(x)}\right), \tag{3.7}$$

where $B_0 = \frac{A_3 A\Gamma(\alpha+1)}{A_2(\lambda_2\mu_2)^{\alpha+1}}$, $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du$ is the Γ-function and

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \lim_{n\to\infty} V\left(\begin{pmatrix} Z^0_{n,h}(x) \\ Z^1_{n,h}(x) \end{pmatrix}\right)$$

with

$$Z^p_{n,h}(x) = \frac{1}{\sqrt{n-m}}\sum_{j=m+1}^{n}\left[Y_j^p \cos\left(\frac{W_j - x}{h}\right) - EY_j^p \cos\left(\frac{W_j - x}{h}\right)\right], \text{ for } p = 0, 1. \tag{3.8}$$

Remarks: 1. Theorem 3 reveals one more complicating feature of this class of estimators in comparison with the density deconvolution estimator considered in van Es and Uh (2005). In their deconvolution problem, the asymptotic distribution of the estimator depends only on the asymptotic parameters through a constant similar to our constant B_0. Theorem 3 reveals that when dealing with nonparametric deconvolution for prediction the variance of the asymptotic distribution also depends on the local properties of distributions ($\mu(x)$ and $f_T(x)$).

2. The variance of the asymptotic distribution is proportional to $\left(\frac{k}{1+k}\right)^2$, which is smallest for $k = 1$ and grows to 1 as k increases. Hence, it is minimized at $k = 1$. However, a byproduct of the proof of part 1) in Theorem 1 is that $|E\hat{\mu}_k(x) - \mu(x)| = O(h^r)$. Since $2k > r$, increasing the value of k accommodates larger values of r and faster decay of the bias as $n \to \infty$. The exact relationship of the order of the bias and k is theoretically unknown, but our Monte Carlo simulations suggest that increases in k may produce significant bias reduction that more than compensates the increase in variance suggested by Theorem 3 (see section 4).

3. Under Assumptions 3.1-3.4 instead of upper bounds (2.8) one has exact orders

$$V(\hat{f}_{T,k}(x)) = (1+o(1))\left(\frac{2kB_0}{\pi(k+1)}M_{n,h}\sigma_{11}\right)^2, \quad V(\hat{d}_{T,k}(x)) = (1+o(1))\left(\frac{2kB_0}{\pi(k+1)}M_{n,h}\sigma_{22}\right)^2.$$

4. A critical step in proving Theorem 2 is provided in Lemma 6. It solves a special case of a problem involving convergence of a sequence of random variables to an improper random variable (Mynbaev (2012)).

4 Monte Carlo simulations

In this section we conduct a simulation study to shed some light on the finite sample properties of $\hat{\mu}_k(x)$ for $k > 1$ and its performance relative to $\tilde{\mu}(x) = \hat{\mu}_1(x)$. The main goal is to assess whether or not the desirable experimental properties attained for density estimation by Mynbaev and Martins-Filho (2010) using the class $\{M_k\}$ manifest themselves for a deconvolution prediction in the context of the CDH model. In our simulations, we consider the case where the measurement errors U_j are of only two types and as in (3.1) we denote the first m errors by $V_1 \sim f_{V_1}$ and the remaining $n-m$ errors by $V_2 \sim f_{V_2}$. We consider the following data generating processes (DGP),

1. case: $g_1(x) = 3x + \frac{20}{\sqrt{2\pi}}\exp\left(-100(x-0.5)^2\right)$, $X \sim N(0.5, \sigma_X^2)$, $\sigma_X^2 = 1/3.92^2$ $\epsilon \sim N(0, 0.673)$

2. case: $g_2(x) = \sin(x\pi/2)/(1 + 2x^2(\text{sign}(x)+1))$, $X \sim N(0, \sigma_X^2)$, $\sigma_X^2 = 1$, $\epsilon \sim N(0, 0.09)$

3. case: $g_3(x) = \begin{cases} 0.5 & \text{for } x \leq -1 \\ |x| & \text{for } -1 < x \leq 0 \\ 2 + \log(x) & \text{for } x > 0 \end{cases}$, $X \sim N(0, \sigma_X^2)$, $\sigma_X^2 = 1$, $\epsilon \sim N(0, 0.09)$.

The first two DGPs were considered by CDH. The third involves a g that is not smooth, violating part of CDH's assumption (4.3) but none of our assumptions, as we do not require smoothness of g.

For each DGP we take $U^F \sim f_{V_1}$ and consider three different error structures: (i) f_{V_1} is $N(0, \sigma_1^2)$ and f_{V_2} is $L(0, \sigma_1/\sqrt{2})$ (Laplace) with $\sigma_1^2 = 0.2\sigma_X^2$ (f_{V_1} smoother than f_{V_2}); (ii) f_{V_1} is $N(0, \sigma_1^2)$ and f_{V_2} is $N(0, \sigma_1^2/2)$ with $\sigma_1^2 = 0.2\sigma_X^2$ (f_{V_1} smoother than f_{V_2}); (iii) f_{V_1} is $N(0, \sigma_1^2)$ and f_{V_2} is $N(0, 2\sigma_1^2)$ with $\sigma_1^2 = 0.1\sigma_X^2$ (f_{V_2} smoother than f_{V_1}). We draw 1000 samples from all DGPs and obtain for each sample the root average squared error (RASE) for $\tilde{\mu}(x)$, $\hat{\mu}_2(x)$ and $\hat{\mu}_4(x)$ in equally spaced grids with 41 points in $(0, 1)$ for case 1, and 41 points in $(-2, 2)$ for cases 2 and 3. Each estimator requires the selection of a kernel (a seed kernel in the case of $\hat{\mu}_k(x)$) and a bandwidth. We considered two kernels,

$$K_1(x) = \frac{1}{\sqrt{2\pi}}\exp(-0.5x^2) \text{ and } K_2(x) = \frac{48\cos(x)}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin(x)}{\pi x^5}\left(2 - \frac{5}{x^2}\right).$$

K_2 has a Fourier transform that is compactly supported in $[-1, 1]$ and satisfies condition (4.2) in CDH and the kernel requirements in our Assumptions 2.2 and 3.3. K_1 does not have a compactly supported Fourier transform and therefore does not satisfy (4.2) in CDH and neither does it satisfy our Assumption 3.3.

We follow CDH and select both a bandwidth h and a ridge parameter ρ by minimizing a cross validation criterion. Representing the jackknifed version of each of the estimators considered generically by $\dot{\mu}_J(x)$, we minimize

$$CV(h,\rho) = \sum_{j=1}^{m+1} (Y_j - \dot{\mu}_J(W_j))^2$$

with respect to (h,ρ). The ridge parameter ρ is necessary to avoid division by a number in the vicinity of zero in the definition of the estimators. For each estimator, at grid points x where the denominator was smaller than ρ, it was replaced by ρ. Throughout the simulations the sample sizes are $n = 250, 500$ and we always take $m = n/2$.

We note that in the two cases where f_{V_1} is smoother than f_{V_2}, a simpler version of the CDH estimator (see the estimator defined in their equation (2.8)) is available that requires neither a bandwidth nor a kernel for its calculation. Here, since one of our goals is to contrast the use of K and M_k in estimator performance, even in these cases we always considered the CDH estimator as defined in our equation (2.5) calculated with a kernel K. In the case where f_{V_2} is smoother than f_{V_1}, the simpler version of the CDH estimator is not available, and by necessity the estimator is given by (2.5). Tables 1, 2 and 3 summarize the results of our simulations for regressions $g_1(x)$, $g_2(x)$ and $g_3(x)$, respectively.

As expected from the asymptotic theory, the mean RASE decreases with the sample size n for all estimators considered across all DGPs. Similarly, the median and the lower and upper boundaries for the interquartile range for all estimators decrease with n for all DGPs.

Table 1. Mean (M), Median (D), Interquartile range $[q_{0.25},q_{0.75}]$(I) for Root Average Squared Error with sample size n, m V_1's, kernel K and $g_1(x) = 3x + \frac{20}{\sqrt{2\pi}}\exp\left(-100(x-0.5)^2\right)$

			$f_{V_1} \sim N, f_{V_2} \sim L$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 > \sigma_2^2$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 < \sigma_2^2$, $K = K_1$		
	n	m	M	D	I	M	D	I	M	D	I
$\tilde{\mu}$	250	125	.455	.448	[.370,.531]	.437	.437	[.359,.512]	.493	.469	[.391,.574]
	500	250	.357	.354	[.299,.413]	.350	.345	[.295,.405]	.365	.345	[.303,.451]
$\hat{\mu}_2$	250	125	.412	.402	[.333,.484]	.376	.365	[.301,.441]	.518	.466	[.387,.600]
	500	250	.321	.318	[.270,.368]	.295	.289	[.246,.340]	.366	.345	[.287,.425]
$\hat{\mu}_4$	250	125	.412	.404	[.331,.478]	.366	.353	[.289,.436]	.532	.485	[.399,.609]
	500	250	.306	.300	[.251,.356]	.278	.271	[.228,.324]	.378	.350	[.288,.437]

Table 2. Mean (M), Median (D), Interquartile range $[q_{0.25},q_{0.75}]$(I) for Root Average Squared Error with sample size n, m V_1's, kernel K and $g_2(x) = \sin(x\pi/2)/(1 + 2x^2(\text{sign}(x) + 1))$

			$f_{V_1} \sim N, f_{V_2} \sim L$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 > \sigma_2^2$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 < \sigma_2^2$, $K = K_1$		
	n	m	M	D	I	M	D	I	M	D	I
$\tilde{\mu}$	250	125	.111	.110	[.088,.136]	.116	.117	[.093,.138]	.103	.097	[.077,.124]
	500	250	.103	.104	[.083,.121]	.107	.108	[.090,.123]	.083	.078	[.063,.100]
$\hat{\mu}_2$	250	125	.101	.098	[.075,.125]	.108	.100	[.077,.124]	.106	.099	[.076,.127]
	500	250	.093	.094	[.073,.113]	.095	.094	[.078,.112]	.083	.077	[.060,.101]
$\hat{\mu}_4$	250	125	.095	.091	[.070,.117]	.098	.096	[.073,.120]	.112	.101	[.075,.133]
	500	250	.085	.084	[.065,.105]	.089	.089	[.072,.108]	.083	.078	[.058,.103]

Table 3. Mean (M), Median (D), Interquartile range $[q_{0.25},q_{0.75}]$(I) for Root Average Squared Error with sample size n, m V_1's, kernel K and

$$g_3(x) = \begin{cases} 0.5 & \text{for } x \leq -1 \\ |x| & \text{for } -1 < x \leq 0 \\ 2 + \log(x) & \text{for } x > 0 \end{cases}$$

			$f_{V_1} \sim N, f_{V_2} \sim L$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 > \sigma_2^2$, $K = K_2$			$f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 < \sigma_2^2$, $K = K_1$		
	n	m	M	D	I	M	D	I	M	D	I
$\tilde{\mu}$	250	125	.139	.137	[.111,.162]	.134	.132	[.117,.156]	.172	.162	[.132,.202]
	500	250	.116	.116	[.096,.135]	.116	.115	[.099,.131]	.134	.129	[.109,.156]
$\hat{\mu}_2$	250	125	.112	.108	[.086,.132]	.106	.101	[.084,.125]	.196	.168	[.125,.227]
	500	250	.091	.090	[.074,.107]	.089	.089	[.074,.103]	.133	.123	[.100,.156]
$\hat{\mu}_4$	250	125	.120	.111	[.088,.142]	.110	.100	[.083,.126]	.215	.172	[.131,.240]
	500	250	.093	.090	[.073,.109]	.090	.089	[.074,.103]	.140	.127	[.100,.166]

For all cases in which f_{V_1} is smoother than f_{V_2} the estimators $\hat{\mu}_2$ and $\hat{\mu}_4$ outperform $\tilde{\mu}$ in that mean, median and boundaries for the interquartile range of their RASE are closer to zero than those associated with $\tilde{\mu}$. By the same standards, $\hat{\mu}_4$ tends to perform better than $\hat{\mu}_2$ for the DGPs using g_1 and g_2. In the case where the

DGP uses g_3, the performances of $\hat{\mu}_2$ and $\hat{\mu}_4$ are virtually the same. In all such cases, we have performed estimation using the kernel K_2 for $\tilde{\mu}$ and as a seed for $\hat{\mu}_2$ and $\hat{\mu}_4$. Results are qualitatively similar when using K_1.

As discussed in Remark 2 following Theorem 3 we should expect an increase in variance with k. This is indeed the case in our simulations (results on the variance are available upon request). However, the reduction in bias produced by increasing k more than compensates the increase in variance to produce smaller RASE for all cases in which f_{V_1} is smoother than f_{V_2}.

In the case where f_{V_2} is smoother than f_{V_1} the estimators perform very similarly in terms of mean and median RASE, specially when the sample size is $n = 500$. The most pronounced difference in this case concerns the interquartile ranges which are larger for the estimators $\hat{\mu}_2$ and $\hat{\mu}_4$ when $n = 250$. In this case, all estimations were performed using the kernel K_1 for $\tilde{\mu}$ and as a seed for $\hat{\mu}_2$ and $\hat{\mu}_4$. Results are qualitatively similar when using K_2.

The simulations suggest, confirming what Mynbaev and Martins-Filho (2010) have found in the case of density estimation, that the use of the M_k reduces bias but increases variance. In addition, the impact of the choice of k on root average squared error seems to change with the family of distributions assumed for the measurement errors, complicating further the determination of an optimal choice for k.

To provide a more vivid portrayal of the distribution of RASE across the simulated samples, we have estimated their densities using the gamma kernel density estimator proposed by Chen (2000). Figure 1 provides estimated densities associated with the RASE of each estimator for $n = 500$ and all DGPs under consideration. The left side panels correspond to DGPs that use the regression g_1, the center panels correspond to DGPs that use the regression g_2 and the right side panels correspond to DGPs that use the regression g_3. Top, middle and bottom panels correspond to error structure (i), (ii) and (iii), respectively. It is apparent that the estimated densities for the RASE of estimators $\hat{\mu}_2$ (dashed graph) and $\hat{\mu}_4$ (solid graph) are closer to the vertical axis and exhibit thicker tails to the left and thinner tails to the right if compared to the estimated density for the RASE of $\tilde{\mu}$ (dashed-dotted graph) in the top and middle panels. In the bottom three panels the densities are more similar, but there seems to be evidence of thicker left and right tails for the estimated densities associated with $\hat{\mu}_2$ and $\hat{\mu}_4$.

5 Summary

The literature on nonparametric density and regression estimation in the presence of measurement error has shown that the rates of convergence for kernel based deconvolution estimators are exceedingly slow. Carroll et al. (2009) have shown that when considering nonparametric prediction in the presence of heterogeneous measurement errors it is possible to obtain much faster convergence of the prediction estimator. They propose an estimator that is consistent and appears to have good finite sample properties. In this paper, we show that the consistency result they have obtained can be derived under less restrictive assumptions on the kernel and on the underlying data generating process. In particular, we show how kernels with non-compactly supported Fourier transforms can be constructed and substantially relax requirements on their order and the smoothness of densities of the measurement errors. The gains come from using the class of kernels proposed in Mynbaev and Martins-Filho (2010) and alternative measures of smoothness of the underlying densities to construct a class of nonparametric prediction estimators that includes the estimator proposed by Carroll et al. (2009). We have also obtained the asymptotic normality of estimators in the class for a subset of super-smooth densities. Although our convergence in distribution result does not cover all possible cases, it shows that the insights from van Es and Uh (2005) can be used to study the properties of the estimator. Lastly, our Monte Carlo simulation shows that it might be beneficial in finite samples to use the kernels of Mynbaev and Martins-Filho (2010) to construct deconvolution type estimators. As pointed out in section 3, future research on a comprehensive study of asymptotic normality is needed.

Appendix - Proofs

The following Lemmas 1 and 2 are part of the proof in Carroll et al. (2009). We provide full proofs here just to show that the assumptions from CDH that we omit are not required.

Lemma 1. *Equations* (2.7) *are true.*

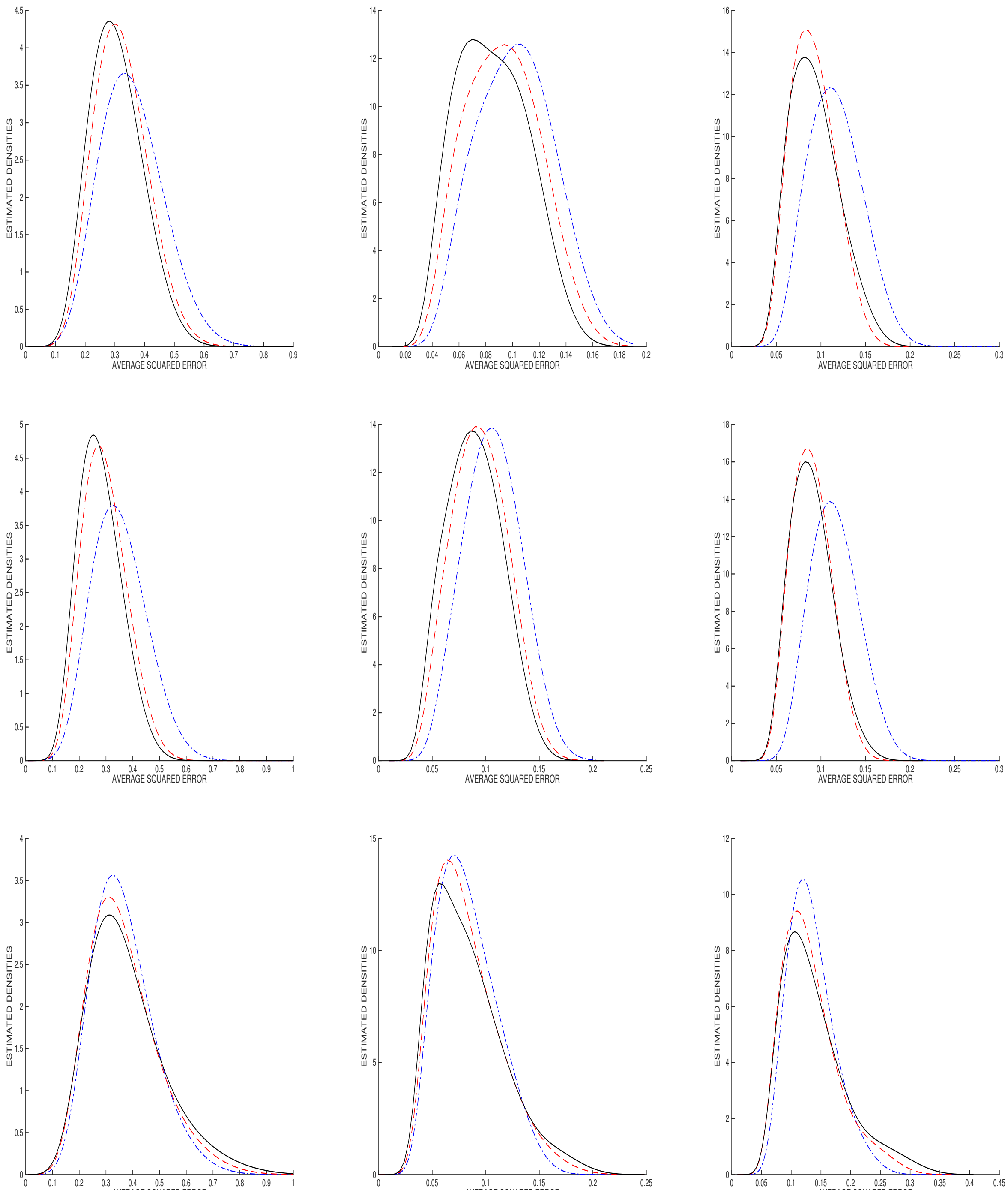

Figure 1: Estimated densities for RASE of estimators $\tilde{\mu}$, $\hat{\mu}_2$ and $\hat{\mu}_4$ using 1000 samples and $n = 500$. $\tilde{\mu}(x)$ $-\cdot-$, $\hat{\mu}_2(x)$ $--$, $\hat{\mu}_4$ $-$. Top panels: $f_{V_1} \sim N, f_{V_2} \sim L$; middle panels: $f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 > \sigma_2^2$; bottom panels: $f_{V_1} \sim N, f_{V_2} \sim N$, $\sigma_1^2 < \sigma_2^2$. Left panels for $g_1(x)$; center panels for $g_2(x)$; right panels for $g_3(x)$.

Proof. (a) By the assumed independence the characteristic functions of W_j satisfy

$$\phi_{f_{W_j}}(t) = \phi_{f_X}(t)\phi_{f_{U_j}}(t),\ \ \phi_{f_T}(t) = \phi_{f_X}(t)\phi_{f_{U^F}}(t). \tag{5.1}$$

For a complex number $a = b + ic$ let $\bar{a} = b - ic$ denote its conjugate. From $a\bar{a} = |a|^2$, $\bar{\phi}_{f_{U_j}}(t) = \phi_{f_{U_j}}(-t)$ and (2.1) one has

$$\sum_{j=1}^{n} \Psi_j(t)\phi_{f_{U_j}}(t) = 1,\ \ \sum_{j=1}^{n} |\Psi_j(t)|^2 = 1/\sum_{j=1}^{n} \left|\phi_{f_{U_j}}(t)\right|^2. \tag{5.2}$$

From (2.3) we have

$$\begin{aligned}
E\tilde{f}_T(x) &= \frac{1}{h}\sum_{j=1}^{n} E\frac{1}{2\pi}\int \exp\left(-it\frac{x - W_j}{h}\right)\mathcal{F}_K(t)\phi_{f_{U^F}}\left(\frac{t}{h}\right)\Psi_j\left(\frac{t}{h}\right)dt \\
&= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\left[\sum_{j=1}^{n} Ee^{itW_j}\Psi_j(t)\right]\phi_{f_{U^F}}(t)\,dt \\
&\text{(using (5.1))} \\
&= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\left[\sum_{j=1}^{n} \phi_{f_X}(t)\,\phi_{f_{U_j}}(t)\,\Psi_j(t)\right]\phi_{f_{U^F}}(t)\,dt \\
&\text{(using (5.2) and (5.1))}
\end{aligned}$$

$$= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\phi_{f_X}(t)\,\phi_{f_{U^F}}(t)\,dt = \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\phi_{f_T}(t)\,dt. \tag{5.3}$$

Note that $\mathcal{F}_{K_h}(t) = \mathcal{F}_K(ht)$ where $K_h(\cdot) = (1/h)\,K(\cdot/h)$. Then,

$$\begin{aligned}
E\tilde{f}_T(x) &= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_{K_h}(t)\phi_{f_T}(t)dt = \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_{K_h * f_T}(t)dt = K_h * f_T(x) \\
&= \int \frac{1}{h}K\left(\frac{x-y}{h}\right)f_T(y)dy = \int K(z)f_T(x - hz)dz
\end{aligned}$$

where $K_h * f_T(x)$ denotes the convolution of K_h and T.

(b) By independence of X_j, ε_j, U^F the variables $Y_j e^{itX_j} = (g(X_j) + \varepsilon_j)e^{itX_j}$ and U^F are independent. Consequently,

$$\begin{aligned}
(EY_j e^{itX_j})(Ee^{itU^F}) &= EY_j e^{it(X_j + U^F)} = EY_j e^{itT} \\
&= E_T[E(Y_j|T = w)e^{itw}] = \int \mu(w)e^{itw}f_T(w)dw = \mathcal{F}_{d_T}(t),
\end{aligned} \tag{5.4}$$

using the law of iterated expectations and the definition of μ. Similarly, for the W_j we have

$$\begin{aligned} EY_j e^{itW_j} &= (EY_j e^{itX_j})(Ee^{itU_j}) = E(g(X_j)+\varepsilon_j)e^{itX_j}\phi_{f_{U_j}}(t) \\ &= Eg(X_j)e^{itX_j}\phi_{f_{U_j}}(t) = EY_j e^{itX_j}\phi_{f_{U_j}}(t). \end{aligned}$$

Because of (5.2), this implies

$$\sum_{j=1}^{n} EY_j e^{itW_j}\Psi_j(t) = Eg(X_j)e^{itX_j} = EY_j e^{itX_j}. \tag{5.5}$$

Now,

$$\begin{aligned} E\tilde{d}_T(x) &= \frac{1}{h}\sum_{j=1}^{n} E\frac{1}{2\pi}\int Y_j e^{-it\frac{x-W_j}{h}}\mathcal{F}_K(t)\phi_{f_{U^F}}\left(\frac{t}{h}\right)\Psi_j\left(\frac{t}{h}\right)dt \\ &= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\left[\sum_{j=1}^{n} EY_j e^{itW_j}\Psi_j(t)\right]\phi_{f_{U^F}}(t)\,dt \\ &= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)EY_j e^{itX_j}\phi_{U^F}(t)\,dt, \text{ by (5.5)} \\ &= \frac{1}{2\pi}\int e^{-itx}\mathcal{F}_K(ht)\mathcal{F}_{d_T}(t)\,dt \text{ by (5.4).} \end{aligned}$$

The last expression is similar to the last expression in (5.3), hence by applying a similar argument, we have $E\tilde{d}_T(x) = \frac{1}{h}\int K\left(\frac{y}{h}\right)d_T(x-y)dy = \int K(z)d_T(x-hz)dz.$ □

Lemma 2. *Under Assumption 2.1* (2.8) *is true. If* $v(h)/n = O(1)$ *and* $\int K(z)d_T(x-hz)dz = O(1)$ *then* $\tilde{d}_T(x) = O_p(1)$.

Proof. $f_{n,j}(x)$ is real-valued, so by (2.3)

$$\begin{aligned} V(f_{n,j}(x)) &\le E\left(f_{n,j}(x)\right)^2 \\ &= \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}Ee^{i(s+t)W_j}\mathcal{F}_K(hs)\phi_{f_{U^F}}(s)\Psi_j(s)\mathcal{F}_K(ht)\phi_{f_{U^F}}(t)\Psi_j(t)dsdt \\ &= \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}Ee^{i(s+t)W_j}\eta_{h,j}(s)\eta_{h,j}(t)dsdt \end{aligned}$$

where we denote $\eta_{h,j}(s) = \mathcal{F}_K(hs)\phi_{f_{U^F}}(s)\Psi_j(s)$. By Assumption 2.1 and boundedness of the Fourier transform in L_2

$$\begin{aligned}
E\left(f_{n,j}(x)\right)^2 &= \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}\int e^{i(s+t)z} f_{W_j}(z)dz\eta_{h,j}(s)\eta_{h,j}(t)dsdt \\
&= \frac{1}{(2\pi)^2}\int\left[\int e^{-i(x-z)s}\eta_{h,j}(s)ds\right]\left[\int e^{-i(x-z)t}\eta_{h,j}(t)dt\right] f_{W_j}(z)dz \\
&= \int\left[\mathcal{F}^{-1}_{\eta_{h,j}}(x-z)\right]^2 f_{W_j}(z)dz \le c_1\left\|\mathcal{F}^{-1}_{\eta_{h,j}}\right\|_2^2 \\
&\le c_2\int\left|\Psi_j(t)\right|^2\left|\mathcal{F}_{K_h}(t)\phi_{f_{U^F}}(t)\right|^2 dt. \qquad (5.6)
\end{aligned}$$

Using this bound and (5.2) we have

$$\begin{aligned}
V(\tilde{f}_T(x)) &= \sum_{j=1}^n V(f_{n,j}(x)) \le c_2\sum_{j=1}^n\int\left|\Psi_j(t)\right|^2\left|\mathcal{F}_{K_h}(t)\phi_{f_{U^F}}(t)\right|^2 dt \\
&= c_2\int\left|\mathcal{F}_K(ht)\phi_{f_{U^F}}(t)\right|^2/\sum_{j=1}^n\left|\phi_{f_{U_j}}(t)\right|^2 dt = \frac{v(h)}{n}c_2.
\end{aligned}$$

which establishes the first equation in (2.8). Denoting $E\varepsilon_j^2 = \sigma_j^2$ and given the fact that the pair (X_j, ε_j) is independent of U_j, we have

$$\begin{aligned}
EY_j^2 e^{i(s+t)W_j} &= E\left(g^2(X_j) + 2g(X_j)\varepsilon_j + \varepsilon_j^2\right)e^{i(s+t)W_j} = \left(Eg^2(X_j)e^{i(s+t)X_j} + \sigma_j^2\phi_{f_{X_j}}(s+t)\right)\phi_{f_{U_j}}(s+t) \\
&= Eg^2(X_j)e^{i(s+t)X_j}\phi_{f_{U_j}}(s+t) + \sigma_j^2\phi_{f_{W_j}}(s+t).
\end{aligned}$$

Then,

$$\begin{aligned}
E\left(d_{n,j}(x)\right)^2 &= E\left(\frac{1}{2\pi}\int e^{-itx}Y_j e^{itW_j}\eta_{h,j}(t)dt\right)^2 \\
&= \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}EY_j^2 e^{i(s+t)W_j}\eta_{h,j}(s)\eta_{h,j}(t)dsdt \\
&= \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}Eg^2(X_j)e^{i(s+t)X_j}\phi_{f_{U_j}}(s+t)\eta_{h,j}(s)\eta_{h,j}(t)dsdt \qquad (5.7) \\
&+ \frac{1}{(2\pi)^2}\int\int e^{-i(s+t)x}\sigma_j^2\phi_{f_{W_j}}(s+t)\eta_{h,j}(s)\eta_{h,j}(t)dsdt. \qquad (5.8)
\end{aligned}$$

Note that because $EY_j^2 = Eg^2(X_j) + \sigma_j^2$, the condition $\sup_j EY_j^2 < \infty$ contained in Assumption 2.1 implies

$\sup_j \sigma_j^2 < \infty$. Therefore, the required bound for (5.8) is obtained as (5.6). For (5.7) we proceed as follows:

$$\begin{aligned}
& \frac{1}{(2\pi)^2} \int \int e^{-i(s+t)x} \left[E g^2(X_j) e^{i(s+t)X_j} \phi_{f_{U_j}}(s+t) \right] \eta_{h,j}(s) \eta_{h,j}(t) ds dt \\
= & \int \left[\int \left(\frac{1}{2\pi} \int e^{-i(x-v-u)s} \eta_{h,j}(s) ds \right)^2 f_{U_j}(u) du \right] g^2(v) f_X(v) dv \\
& \text{(using Assumption 2.1 and boundedness of the Fourier transform)} \\
\leq & \; c_1 \int \left\| \mathcal{F}^{-1} \eta_{h,j} \right\|_2^2 g^2(v) f_X(v) dv \leq c_2 \left\| \eta_{h,j} \right\|_2^2 E g^2(X_j) \\
\leq & \; c_3 \int \left| \Psi_j(t) \right|^2 \left| \mathcal{F}_{K_h}(t) \phi_{f_{U^F}}(t) \right|^2 dt.
\end{aligned}$$

As a result, $V(\tilde{d}_T(x)) = \sum_{j=1}^n V(d_{n,j}(x)) \leq c_4 \int \left| \mathcal{F}_{K_h}(t) \phi_{f_{U^F}}(t) \right|^2 / \sum_{j=1}^n \left| \phi_{f_{U_j}}(t) \right|^2 dt = c_4 \frac{v(h)}{n}$. Note that the bound

$$E \left(\tilde{d}_T(x) \right)^2 = V \left(\tilde{d}_T(x) \right) + \left(E \tilde{d}_T(x) \right)^2 \leq \left(\int K(z) d_T(x - hz) dz \right)^2 + c_4 v(h)/n$$

implies $\tilde{d}_T(x) = O_p(1)$ when the terms on the right are $O(1)$. □

Proof of Theorem 1. By (2.12), Hölder's inequality, changing variables and applying (2.14) we have

$$\begin{aligned}
\left| E \hat{f}_{T,k}(x) - f_T(x) \right| &= c \left| \int K(t) \left| ht \right|^{r+1/q} \frac{\Delta_{-ht}^{2k} f_T(x)}{\left| ht \right|^{r+1/q}} dt \right| \\
&\leq c \left(\int \left| K(t) \right|^{q'} \left| ht \right|^{(r+1/q)q'} dt \right)^{1/q'} \left[\int \left(\frac{\sup_x \left| \Delta_{-ht}^{2k} f_T(x) \right|}{\left| ht \right|^r} \right)^q \frac{dt}{\left| ht \right|} \right]^{1/q} \\
&\leq c h^r \left(\int \left| K(t) \right|^{q'} \left| t \right|^{(r+1/q)q'} dt \right)^{1/q'} \left\| f_T \right\|_{b_{\infty,q}^r} = O(h^r). \qquad (5.9)
\end{aligned}$$

Given that $\mathcal{F}_{M_k}(t) = \sum_{l=1}^k \lambda_{k,m} (\mathcal{F}_K(tl) + \mathcal{F}_K(-tl))$, following the same arguments used to bound $V(\tilde{f}_T)$ in Lemma 2, and the Cauchy-Schwarz inequality we obtain

$$V(\hat{f}_{T,k}(x)) \leq c \left\{ \sum_{l=1}^k \int \left| \mathcal{F}_K(hlt) \right|^2 \left| \phi_{f_{U^F}}(t) \right|^2 / \sum_{j=1}^n \left| \phi_{f_{U_j}}(t) \right|^2 dt \right\} \equiv \frac{c}{n} \sum_{l=1}^k v(hl).$$

Consequently, $\hat{f}_{T,k}(x) - E \hat{f}_{T,k}(x) = O_p \left((\sum_{l=1}^k v(hl)/n)^{1/2} \right)$. Thus,

$$\hat{f}_{T,k}(x) = \left(\hat{f}_{T,k}(x) - E \hat{f}_{T,k}(x) \right) + \left(E \hat{f}_{T,k}(x) - f_T(x) \right) + f_T(x) = f_T(x) + O_p(m_{n,h}).$$

where $m_{n,h} = h^r + (\sum_{l=1}^{k} v(hl)/n)^{1/2}$. By Assumption 2.2 and (2.13) $d_T \in B^s_{\infty,q}$. As above, $\hat{d}_{T,k}(x) = d_T(x) + O_p(m_{n,h})$. The theorem follows from

$$\hat{\mu}_k(x) - \frac{d_T(x)}{f_T(x)} = \frac{[d_T(x) + O_p(m_{n,h})]\, f_T(x) - [f_T(x) + O_p(m_{n,h})]\, d_T(x)}{f_T(x)[f_T(x) + O_p(m_{n,h})]} = O_p(m_{n,h}).$$

□

Proof of Theorem 2. 1. We define $\mathcal{F}_K$ and then take its inverse Fourier transform to obtain K. Let

$$\mathcal{F}_K(s) = \exp(-P(s)/2),\ |s| \geq c_1.$$

For $|s| < c_1$, $\mathcal{F}_K(s)$ can be defined to be any function, as long as the condition $\mathcal{F}_K(0) = 1$ necessary for $\int K(t)dt = 1$ is met. We also need $\mathcal{F}_K$ to be sufficiently smooth, including the sewing conditions at $s = \pm c_1$, for $\mathcal{F}_K$ to belong to the Sobolev space $W_2^1(\mathbb{R})$. Assumption 2.3 (c) provides the required ingredient for the domain $|s| \geq c_1$:

$$\|\mathcal{F}_K\|^2_{W_2^1(|s|\geq c_1)} = \int_{|s|\geq c_1} \left(|\mathcal{F}_K(s)|^2 + \left|\mathcal{F}_K^{(1)}(s)\right|^2 \right) ds = \int_{|s|\geq c_1} \exp(-P(s)) \left(1 + \left|P^{(1)}(s)\right|^2/4\right) ds < \infty.$$

Let $K(x) = \left(\mathcal{F}^{-1}\mathcal{F}_K\right)(x)$. Then, $ixK(x) = \mathcal{F}^{-1}_{\mathcal{F}_K^{(1)}}(x)$. Since the inverse Fourier transform preserves the L_2-norm (up to a constant c), we have

$$\int_{\mathbb{R}} |K(x)|^2 (1 + x^2) dx = c \int_{\mathbb{R}} \left(|\mathcal{F}_K(s)|^2 + \left|\mathcal{F}_K^{(1)}(s)\right|^2 \right) ds < \infty.$$

By Hölder's inequality

$$\int_{\mathbb{R}} |K(x)| dx \leq \left[\int_{\mathbb{R}} |K(x)|^2 (1 + x^2) dx \right]^{1/2} \left(\int_{\mathbb{R}} \frac{dx}{1 + x^2} \right)^{1/2} < \infty.$$

establishing that $K \in L_1$. Since $\mathcal{F}_K \in W_2^1(R)$, $\mathcal{F}_K$ is globally bounded (see Adams and Fournier (2003), Theorem 4.12). By Assumption 2.3(a) the product $|\mathcal{F}_K|^2 \Phi_n$ is locally bounded. Therefore to check that $v(h) < \infty$ for $0 < h < 1$ it suffices to verify that the integral

$$I(h) \equiv \int_{|sh|\geq c_1} |\phi_K(sh)|^2 \Phi_n(s) ds = \int_{|sh|\geq c_1} \exp(-P(sh)) \Phi_n(s) ds$$

is finite for $0 < h < 1$. Using Assumption 2.3(a) we have

$$I(h) \leq c_2 \int_{|sh|\geq c_1} \exp(-P(sh)) P(s) ds \leq c_2 \int_{|s|\geq c_1} \exp(-P(sh)) P(s) ds.$$

Requiring the last integral to be finite for all $0<h<1$ instead of $J(h)<\infty$ would be enough for applications (and the invertibility of P would not be needed). Note that $\exp(-P(sh))$ is used here to suppress the effect of growth of $P(s)$. Using Assumption 2.3(b) and replacing $t=P(s)$ we get by (d) the desired result

$$I(h) \le 2c_2c_3 \int_{s\ge c_1} \exp(-P(sh))P'(s)ds = 2c_2c_3J(h) < \infty.$$

2. If $v(h)\le c$ for all $0<h<1$, then obviously $v(h)/n=o(1)$ with any choice of $h_n\to 0$. Suppose $v(h)$ is unbounded. Take any positive sequence $\varepsilon_n=o(1)$ such that $\varepsilon_n n\to\infty$. Define h_n by $v(h_n)=\varepsilon_n n$ (if there are many such h_n, take the least of them). From the above proof we know that $v(h)$ is bounded from above when h is bounded away from zero. Therefore $\varepsilon_n n\to\infty$ implies $h_n\to 0$. Finally, $v(h_n)/n=\varepsilon_n=o(1)$. □

Representation for $f_{n,j,k}(x)$. For $j=m+1,...,n$ the analysis of expressions defined in (3.6) is based on the representation derived here. Denoting

$$\gamma_n(t) = \frac{1}{n-m}\frac{A_3}{A_2}|t|^{a_3-a_2}\exp(\mu_2|t|^{\lambda_2}-\mu_3|t|^{\lambda_3}), \tag{5.10}$$

write (3.5) as

$$(\phi_3\Psi_j)(t) = (1+u_n(t))\gamma_n(t),\ \ j=m+1,...,n, \tag{5.11}$$

where $u_n(t)\to 0$ as $|t|\to\infty$. Fix $\varepsilon\in(0,1)$ and put $\theta_l=2\lambda_{k,l}/(\pi l)$,

$$s^1_{n,j,k}(x) = \sum_{l=1}^k \frac{\theta_l}{h}\int_0^\varepsilon \cos\left(t\frac{W_j-x}{lh}\right)\mathcal{F}_K(t)(\phi_3\Psi_j)\left(\frac{t}{lh}\right)dt, \tag{5.12}$$

$$s^2_{n,j,k}(x) = \sum_{l=1}^k \frac{\theta_l}{h}\int_\varepsilon^1 \cos\left(t\frac{W_j-x}{lh}\right)\mathcal{F}_K(t)\left[(\phi_3\Psi_j)\left(\frac{t}{lh}\right)-\gamma_n\left(\frac{t}{lh}\right)\right]dt. \tag{5.13}$$

Then from (3.6) we have

$$f_{n,j,k}(x) = s^1_{n,j,k}(x)+s^2_{n,j,k}(x)+\sum_{l=1}^k\frac{\theta_l}{h}\int_\varepsilon^1\cos\left(t\frac{W_j-x}{lh}\right)\mathcal{F}_K(t)\gamma_n\left(\frac{t}{lh}\right)dt. \tag{5.14}$$

In the last integral, all values of $\cos\left(t\frac{W_j-x}{lh}\right)$ for $t\in(\varepsilon,1)$ contribute to its value, which makes it difficult to analyze. One of the main insights of Van Es and Uh was to approximate $\cos\left(t\frac{W_j-x}{lh}\right)$ with $\cos\left(\frac{W_j-x}{lh}\right)$. From $\cos x-\cos y=-2\sin\left(\frac{x+y}{2}\right)\sin\left(\frac{x-y}{2}\right)$ we get

$$\cos\left(t\frac{W_j-x}{lh}\right)-\cos\left(\frac{W_j-x}{lh}\right) = -2\sin\left(\frac{t+1}{2}\frac{W_j-x}{lh}\right)\sin\left(\frac{t-1}{2}\frac{W_j-x}{lh}\right) = R_{j,l}(t)$$

where the remainder $R_{j,l}$ is defined by the right-hand side and, because $|\sin x| \le |x|$, we have

$$|R_{j,l}(t)| \le (|x| + |W_j|)\frac{1-t}{lh}. \tag{5.15}$$

Letting $s^3_{n,j,k}(x) = \sum_{l=1}^{k} \frac{\theta_l}{h} \int_\varepsilon^1 R_{j,l}(t)\mathcal{F}_K(t)\gamma_n\left(\frac{t}{lh}\right) dt$ we rearrange (5.14) and write

$$\begin{aligned} f_{n,j,k}(x) &= s^1_{n,j,k}(x) + s^2_{n,j,k}(x) + s^3_{n,j,k}(x) \\ &\quad + \sum_{l=1}^{k} \frac{\theta_l}{h} \cos\left(\frac{W_j - x}{lh}\right) \int_\varepsilon^1 \mathcal{F}_K(t)\gamma_n\left(\frac{t}{lh}\right) dt. \end{aligned} \tag{5.16}$$

Next, let

$$S^{p,d}_{n,k}(x) = \sum_{j=m+1}^{n} Y_j^p s^d_{n,j,k}(x), \;\; p = 0, 1; \;\; d = 1, 2, 3. \tag{5.17}$$

Now (5.16) and (3.8) yield

$$\begin{aligned} \hat{f}^{\text{II}}_{T,k}(x) - E\hat{f}^{\text{II}}_{T,k}(x) &= \sum_{d=1}^{3} \left[S^{0,d}_{n,k}(x) - ES^{0,d}_{n,k}(x)\right] \\ &\quad + \sum_{j=m+1}^{n} \sum_{l=1}^{k} \frac{\theta_l}{h}(\Lambda_{j,l} - E\Lambda_{j,l}) \int_\varepsilon^1 \mathcal{F}_K(t)\gamma_n\left(\frac{t}{lh}\right) dt, \end{aligned} \tag{5.18}$$

$$\begin{aligned} \hat{d}^{\text{II}}_{T,k}(x) - E\hat{d}^{\text{II}}_{T,k}(x) &= \sum_{d=1}^{3} \left[S^{1,d}_{n,k}(x) - ES^{1,d}_{n,k}(x)\right] \\ &\quad + \sum_{j=m+1}^{n} \sum_{l=1}^{k} \frac{\theta_l}{h}(Y_j\Lambda_{j,l} - EY_j\Lambda_{j,l}) \int_\varepsilon^1 \mathcal{F}_K(t)\gamma_n\left(\frac{t}{lh}\right) dt, \end{aligned} \tag{5.19}$$

where $\Lambda_{j,l} = \cos\left(\frac{W_j - x}{lh}\right) - E\cos\left(\frac{W_j - x}{lh}\right)$. Our goal will be to show that the first three terms in (5.18) and (5.19) are small relative to the last ones, while the last terms are asymptotically normal, up to a factor that depends on n and h. □

Lemma 3. *For $p = 0, 1$ and $\varepsilon \in (0, 1)$ one has, as $n \to \infty$ and $h \to 0$,*

$$S^{p,1}_{n,k}(x) - ES^{p,1}_{n,k}(x) = O_p\left[\frac{1}{\sqrt{n-m}} h^{a_2 - a_3 - 1} \zeta(h)^{\varepsilon^{\lambda_2}(1+o(1))}\right].$$

Proof. From (5.12) and (5.17) by independence of (Y_j, W_j)

$$\begin{aligned} V\left(S^{p,1}_{n,k}(x)\right) &\le \sum_{j=m+1}^{n} E\left[\sum_{l=1}^{k} \frac{\theta_l}{h} Y_j^p \int_0^\varepsilon \cos\left(t\frac{W_j - x}{lh}\right) \mathcal{F}_K(t)(\phi_3\Psi_j)\left(\frac{t}{lh}\right) dt\right]^2 \\ &\quad \text{(bounding cos by 1 and } |\mathcal{F}_K| \text{ by } \|K\|_{L_1}\text{)} \\ &\le \varepsilon^2 \sum_{j=m+1}^{n} E\left(Y_j^p\right)^2 \left[\sum_{l=1}^{k} \frac{|\theta_l|}{h} \|K\|_{L_1} \max_{1\le l\le k} \sup_{|t|\le\varepsilon} \left|(\phi_3\Psi_j)\left(\frac{t}{lh}\right)\right|\right]^2. \end{aligned} \tag{5.20}$$

For a sufficiently small h, the numbers $\varepsilon/(lh)$, $l = 1, ..., k$, belong to the domain where (3.5) holds. By (3.5) and the assumption $\lambda_2 > \lambda_3$, the product $\phi_3\Psi_j$ grows at infinity. Hence, the sup in (5.20) is attained at the boundary of $|t| \leq \varepsilon$ and the max is attained at $l = 1$, so that

$$V\left(S_{n,k}^{p,1}(x)\right) \leq \frac{c(\varepsilon)}{h^2}(n-m)\sup_j E\left(Y_j^p\right)^2 \frac{1}{(n-m)^2}\left(\frac{\varepsilon}{h}\right)^{2(a_3-a_2)} \frac{\exp\left(2\mu_2(\varepsilon/h)^{\lambda_2}\right)}{\exp\left(2\mu_3(\varepsilon/h)^{\lambda_3}\right)}$$

$$\leq \frac{c_2(\varepsilon)}{n-m}h^{2(a_2-a_3-1)}\exp\left\{2\mu_2\left(\frac{\varepsilon}{h}\right)^{\lambda_2}\left[1-\frac{2\mu_3}{2\mu_2}\left(\frac{\varepsilon}{h}\right)^{\lambda_3-\lambda_2}\right]\right\}$$
$$= \frac{c_2(\varepsilon)}{n-m}h^{2(a_2-a_3-1)}\left[\exp\left(\mu_2 h^{-\lambda_2}\right)\right]^{2\varepsilon^{\lambda_2}(1+o(1))}.$$

This bound proves the lemma. □

Lemma 4. *For $p = 0, 1$ and $\varepsilon \in (0, 1)$ one has, as $n \to \infty$ and $h \to 0$,*

$$S_{n,k}^{p,2}(x) - ES_{n,k}^{p,2}(x) = o_p\left[\frac{1}{\sqrt{n-m}}h^{a_2-a_3-1}\zeta(h)\exp\left(-\mu_3 h^{-\lambda_3}\right)\right].$$

Proof. By (5.11) $(\phi_3\Psi_j)(t) - \gamma_n(t) = u_n(t)\gamma_n(t)$. For any $\delta > 0$ there exists $t(\delta) > 0$ such that $|u_n(t)| \leq \delta$ for $|t| \geq t(\delta)$. Let $|t| \geq \varepsilon$ and $h \leq \varepsilon/(kt(\delta))$. Then

$$\left|\frac{t}{lh}\right| \geq \varepsilon\frac{kt(\delta)}{l\varepsilon} \geq t(\delta) \text{ and } \left|u_n\left(\frac{t}{lh}\right)\right| \leq \delta. \tag{5.21}$$

By independence of (Y_j, W_j), (5.13) and (5.17) imply

$$\begin{aligned} V\left(S_{n,k}^{p,2}(x)\right) &\leq \sum_{j=m+1}^{n} E\left[\sum_{l=1}^{k}\frac{\theta_l}{h}Y_j^p\int_\varepsilon^1 \cos\left(t\frac{W_j-x}{lh}\right)\mathcal{F}_K(t)u_n\left(\frac{t}{lh}\right)\gamma_n\left(\frac{t}{lh}\right)dt\right]^2 \\ &\quad \text{(using (5.21))} \\ &\leq (n-m)\sup_j E\left(Y_j^p\right)^2\delta^2\left[\sum_{l=1}^{k}\frac{|\theta_l|}{h}\|K\|_{L_1}\max_{1\leq l\leq k}\sup_{t\in(\varepsilon,1)}\gamma_n\left(\frac{t}{lh}\right)(1-\varepsilon)\right]^2 \\ &= c_1\frac{\delta^2(n-m)}{h^2}\max_{1\leq l\leq k}\sup_{t\in(0,1)}\gamma_n^2\left(\frac{t}{lh}\right). \end{aligned}$$

Since $\lambda_2 > \lambda_3$, the function γ_n is U-shaped, the sup is attained at $|t| = 1$ and the max is attained at $l = 1$. Hence, by (5.10)

$$\begin{aligned} V\left(S_{n,k}^{p,2}(x)\right) &\leq c_2\frac{\delta^2(n-m)}{h^2}\frac{1}{(n-m)^2}h^{2(a_2-a_3)}\exp(2\mu_2 h^{-\lambda_2} - 2\mu_3 h^{-\lambda_3}) \\ &= c_2\frac{\delta^2}{n-m}h^{2(a_2-a_3-1)}\zeta^2(h)\exp(-2\mu_3 h^{-\lambda_3}) \text{ for } h \leq \varepsilon/(kt(\delta)). \end{aligned}$$

Since δ is arbitrarily close to zero, this bound proves the statement. This proof partially explains why ε cannot be set to zero. If t is not bounded away from zero, then (5.21) does not hold. In the proof of Theorem 3 below the positivity of ε is even more important. $\square$

Lemma 5. *Denote* $g(t) = g_\lambda(t) = \left((1-t)^\lambda - 1\right)/t$, *where* $\lambda > 0$ *and* $t \in (0,1)$. *Then,* $g(t) = -\lambda + o(1)$ *as* $t \to 0$, *and*

(a) $-\lambda \le g(t) \le -1$ *in case* $\lambda \ge 1$,

(b) $-1 \le g(t) \le -\lambda$ *in case* $0 < \lambda < 1$.

Proof. The asymptotic behavior as $t \to 0$ follows from $(1-t)^\lambda = 1 - \lambda t + o(t)$.

(a) Let $\lambda \ge 1$. Obviously, $(1-t)^\lambda \le 1 - t$ which implies $g(t) \le -1$. Let $G(t) = 1 - \lambda t$, $H(t) = (1-t)^\lambda$. Then

$$G(0) = H(0) = 1,\ G'(t) = -\lambda \le -\lambda(1-t)^{\lambda-1} = H'(t).$$

Therefore

$$1 - \lambda t = G(t) = G(0) + \int_0^t G'(s)ds \le H(0) + \int_0^t H'(s)ds = H(t) = (1-t)^\lambda,$$

which implies $g(t) \ge -\lambda$.

(b) Let $0 < \lambda < 1$. The inequality $1 - t \le (1-t)^\lambda$ is obvious, so $g(t) \ge -1$. In this case, the inequalities above are reverted and $H'(t) \le G'(t)$ and $g(t) \le -\lambda$. $\square$

The next lemma is similar to (van Es and Uh, 2005, Lemma 5).

Lemma 6. *For* $\varepsilon \in (0,1)$, $\beta \ge 0$ *denote*

$$I_\beta(n,h) = \int_\varepsilon^1 (1-t)^\beta \mathcal{F}_K(t)\gamma_n\left(\frac{t}{h}\right)dt, \quad B_\beta = \frac{A_3 A \Gamma(\alpha+\beta+1)}{A_2(\lambda_2\mu_2)^{\alpha+\beta+1}}.$$

Then

$$I_\beta(n,h) = (1+o(1))\frac{B_\beta}{n-m}h^{\lambda_2(\alpha+\beta+1)+a_2-a_3}\zeta(h).$$

If in the definition of $I_\beta(n,h)$, $\mathcal{F}_K$ *is replaced by* $|\mathcal{F}_K|$, *then the asymptotic behavior remains the same, with* A *replaced by* $|A|$.

Proof. Let $1-t=h^{\lambda_2}v$. Then, $t=1-h^{\lambda_2}v$, $dt=-h^{\lambda_2}dv$ and by definition (5.10)

$$\begin{aligned}
I_\beta(n,h) &= -\frac{h^{\lambda_2}}{n-m}\int_{(1-\varepsilon)h^{-\lambda_2}}^{0} h^{\beta\lambda_2}v^{\beta}\frac{\phi_K(1-h^{\lambda_2}v)}{(h^{\lambda_2}v)^{\alpha}}\left(h^{\lambda_2}v\right)^{\alpha}\frac{A_3}{A_2}\left(\frac{1-h^{\lambda_2}v}{h}\right)^{a_3-a_2} \\
&\quad\times\exp\left[\mu_2\left(\frac{1-h^{\lambda_2}v}{h}\right)^{\lambda_2}-\mu_3\left(\frac{1-h^{\lambda_2}v}{h}\right)^{\lambda_3}\right]dv \\
&= \frac{A_3}{A_2(n-m)}h^{\lambda_2(\alpha+\beta+1)+a_2-a_3}\exp\left(\mu_2h^{-\lambda_2}-\mu_3h^{-\lambda_3}\right)\int_0^{(1-\varepsilon)h^{-\lambda_2}}\Delta_h(v)dv \qquad (5.22)
\end{aligned}$$

where the integrand $\Delta_h(v)$ is defined by

$$\begin{aligned}
\Delta_h(v) &= v^{\alpha+\beta}\frac{\mathcal{F}_K(1-h^{\lambda_2}v)}{(h^{\lambda_2}v)^{\alpha}}\left(1-h^{\lambda_2}v\right)^{a_3-a_2} \\
&\quad\times\exp\left[\mu_2\frac{(1-h^{\lambda_2}v)^{\lambda_2}-1}{h^{\lambda_2}v}v-\mu_3\frac{(1-h^{\lambda_2}v)^{\lambda_3}-1}{h^{\lambda_2}v}h^{\lambda_2-\lambda_3}v\right].
\end{aligned}$$

We need to find the limit of $\Delta_h(v)$, as $h\to 0$. The interval $(0,(1-\varepsilon)h^{-\lambda_2})$ expands to $(0,\infty)$. By Assumption 3.3, $\frac{\mathcal{F}_K(1-h^{\lambda_2}v)}{(h^{\lambda_2}v)^{\alpha}}\to A$ and by Lemma 5 $\frac{(1-h^{\lambda_2}v)^{\lambda_2}-1}{h^{\lambda_2}v}\to-\lambda_2$ and $\frac{(1-h^{\lambda_2}v)^{\lambda_3}-1}{h^{\lambda_2}v}\to-\lambda_3$. Hence,

$$\Delta_h(v) = (1+o(1))Av^{\alpha+\beta}\exp\left(-\mu_2\lambda_2v+\mu_3\lambda_3h^{\lambda_2-\lambda_3}v\right)=(1+o(1))Av^{\alpha+\beta}\exp\left(-\mu_2\lambda_2v\right). \qquad (5.23)$$

Next we need to find an integrable majorant for $\Delta_h(v)$. By Assumption 3.3

$$\sup_{0<t<1}\frac{|\mathcal{F}_K(1-t)|}{t^{\alpha}}<\infty. \qquad (5.24)$$

Further,

$$\left(1-h^{\lambda_2}v\right)^{a_3-a_2}\le\max\left\{1,\varepsilon^{a_3-a_2}\right\}\text{ for }v\in(0,(1-\varepsilon)h^{-\lambda_2}). \qquad (5.25)$$

By Lemma 5, in which we put $t=h^{\lambda_2}v$,

$$\begin{aligned}
&\exp\left[\mu_2\frac{(1-h^{\lambda_2}v)^{\lambda_2}-1}{h^{\lambda_2}v}v-\mu_3\frac{(1-h^{\lambda_2}v)^{\lambda_3}-1}{h^{\lambda_2}v}h^{\lambda_2-\lambda_3}v\right] \\
=&\exp\left[\mu_2g_{\lambda_2}(t)v-\mu_3g_{\lambda_3}(t)h^{\lambda_2-\lambda_3}v\right] \\
\le&\exp\left[\mu_2\max\{-1,-\lambda_2\}v+\mu_3\max\{1,\lambda_3\}h^{\lambda_2-\lambda_3}v\right]. \qquad (5.26)
\end{aligned}$$

We can choose h_0 so that for $h\le h_0$

$$\mu_3\max\{1,\lambda_3\}h^{\lambda_2-\lambda_3}\le-\frac{1}{2}\mu_2\max\{-1,-\lambda_2\}. \qquad (5.27)$$

Then for all $0 < h \le h_0$ (5.24)-(5.27) imply

$$I_{(0,(1-\varepsilon)h^{-\lambda_2})}\Delta_h(v) \le cv^{\alpha+\beta} \exp\left[\frac{1}{2}\mu_2 \max\{-1, -\lambda_2\}v\right]. \tag{5.28}$$

The function on the right is integrable on $(0,\infty)$. By the dominated convergence theorem, from (5.22), (5.23), (5.28) and noting that $\exp\left(\mu_2 h^{-\lambda_2} - \mu_3 h^{-\lambda_3}\right) = \zeta(h)\exp\left(-\mu_3 h^{-\lambda_3}\right)$, we have

$$\begin{aligned}
I_\beta(n,h) &= \frac{(1+o(1))A_3 A}{A_2(n-m)} h^{\lambda_2(\alpha+\beta+1)+a_2-a_3}\zeta(h)\exp\left(-\mu_3 h^{-\lambda_3}\right) \\
&\quad \times \int_0^\infty v^{\alpha+\beta}\exp\left(-\lambda_2\mu_2 v\right)dv \\
&\quad (\text{replacing } \lambda_2\mu_2 v = u) \\
&= \frac{(1+o(1))A_3 A}{A_2(n-m)(\lambda_2\mu_2)^{\alpha+\beta+1}} h^{\lambda_2(\alpha+\beta+1)+a_2-a_3}\zeta(h)\exp\left(-\mu_3 h^{-\lambda_3}\right) \\
&\quad \times \int_0^\infty u^{\alpha+\beta}e^{-u}du,
\end{aligned}$$

which gives the desired result. If $\mathcal{F}_K$ is replaced by $|\mathcal{F}_K|$, then in (5.23) A gets replaced by $|A|$; everything else does not change. □

Lemma 7. *The variable defined in* (5.17) *for* $d = 3$ *satisfies*

$$S_{n,k}^{p,3}(x) - ES_{n,k}^{p,3}(x) = O_p\left[\frac{1}{\sqrt{n-m}}h^{\lambda_2(\alpha+2)+a_2-a_3-2}\zeta(h)\right].$$

Proof. Using (5.15) we estimate one term in (5.17):

$$\begin{aligned}
\left|Y_j^p s_{n,j,k}^3(x)\right| &\le \sum_{l=1}^k \frac{|\theta_l|}{h}\left|Y_j^p\right|\frac{|x|+|W_j|}{lh}\int_\varepsilon^1 (1-t)\left|\mathcal{F}_K(t)\right|\gamma_n\left(\frac{t}{lh}\right)dt \\
&\quad (\text{by Lemma 6 with } \beta = 1) \\
&\le c_1\sum_{l=1}^k \frac{|\theta_l|}{h}\left|Y_j^p\right|\frac{|x|+|W_j|}{lh}\frac{(lh)^{\lambda_2(\alpha+2)+a_2-a_3}}{n-m}\zeta(lh).
\end{aligned}$$

Since $\zeta(lh) = o(\zeta(h))$, we have

$$\left|Y_j^p s_{n,j,k}^3(x)\right| \le c\left|Y_j^p\right|(|x|+|W_j|)\frac{h^{\lambda_2(\alpha+2)+a_2-a_3-2}}{n-m}\zeta(h).$$

By Assumption 3.4 and Hölder's inequality it follows that

$$\begin{aligned} V\left(S_n^{p,3}(x)\right) &\leq \sum_{j=m+1}^{n} E\left[Y_j^p s_{n,j,k}^3(x)\right]^2 \\ &\leq c \sum_{j=m+1}^{n} E\left[\left|Y_j^p\right|(|x|+|W_j|)\right]^2 \left[\frac{h^{\lambda_2(\alpha+2)+a_2-a_3-2}}{n-m}\zeta(h)\right]^2 \\ &\leq c_1 \left[\frac{h^{\lambda_2(\alpha+2)+a_2-a_3-2}}{\sqrt{n-m}}\zeta(h)\right]^2. \end{aligned}$$

This proves the statement. □

Lemma 8. *The variables* $\hat{f}_T^{\mathrm{I}}(x)$ *and* $\hat{d}_T^{\mathrm{I}}$ *defined in* (3.3) *satisfy*

$$\hat{f}_{T,k}^{\mathrm{I}}(x) - E\hat{f}_{T,k}^{\mathrm{I}}(x) = O_p\left(\frac{1}{h\sqrt{n-m}}\right), \quad \hat{d}_{T,k}^{\mathrm{I}}(x) - E\hat{d}_{T,k}^{\mathrm{I}}(x) = O_p\left(\frac{1}{h\sqrt{n-m}}\right).$$

Proof. The product $\phi_3(t)\phi_1(-t)\left|\phi_2(t)\right|^{-2}$ is continuous and vanishes at infinity, as seen from (3.4) and the condition $\lambda_1 > \lambda_2$. Hence, (3.4) implies $|(\phi_3\Psi_j)(t)| \leq \frac{c_1}{n-m}$ and by Assumptions 3.3 and 3.4

$$\begin{aligned} V\left(\hat{d}_{T,k}^{\mathrm{I}}(x)\right) &\leq \sum_{j=1}^{m} E\left[\sum_{l=1}^{k}\frac{|\theta_l|}{h}|Y_j|\,\|K\|_{L_1}\frac{c_1}{n-m}\right]^2 \\ &\leq \frac{c_1 m}{h^2(n-m)^2} \leq \frac{c_2}{h^2(n-m)}. \end{aligned}$$

The bound for $\hat{f}_{T,k}^{\mathrm{I}}(x)$ follows similarly. □

Lemma 9. *The variables from* (3.8) *converge in joint distribution*

$$\begin{pmatrix} Z_{n,h}^0(x) \\ Z_{n,h}^1(x) \end{pmatrix} \xrightarrow{d} N(0,\Sigma), \tag{5.29}$$

where Σ *is from* (3.8) *and* $\sigma_{11} = 1/2$.

Proof. By the Cramér-Wold theorem it suffices, for each $a \in \mathbb{R}^2$, to prove convergence in distribution of $S_n = a_1 Z_{n,h}^0(x) + a_2 Z_{n,h}^1(x)$ to $N(0, a'\Sigma a)$. Denoting

$$\Gamma_j = \cos\left(\frac{W_j - x}{h}\right), \quad X_{nj} = \frac{1}{\sqrt{n-m}}\left[(a_1 + a_2 Y_j)\Gamma_j - E(a_1+a_2Y_j)\Gamma_j\right]$$

we can write $S_n = \sum_{j=m+1}^{n} X_{nj}$. To prove convergence of S_n, we check the conditions of Lyapounov's Theorem (Billingsley, 1995, Theorem 27.3). X_{nj} are obviously independent and satisfy $EX_{nj} = 0$. Further,

$$\begin{aligned} EX_{nj}^2 &= \frac{1}{n-m} V\left((a_1 + a_2 Y_j)\Gamma_j\right) \\ &= \frac{1}{n-m}\left[a_1^2 V(\Gamma_j) + 2a_1 a_2 cov(\Gamma_j, Y_j\Gamma_j) + a_2^2 V(Y_j\Gamma_j)\right]. \end{aligned} \tag{5.30}$$

By independence and the condition $E\varepsilon_j = 0$

$$\begin{aligned} cov(\Gamma_j, Y_j\Gamma_j) &= E\left(\Gamma_j - E\Gamma_j\right)\left[(g(X_j) + \varepsilon_j)\Gamma_j - E(g(X_j) + \varepsilon_j)\Gamma_j\right] \\ &= E\left(\Gamma_j - E\Gamma_j\right)\left[g(X_j)\Gamma_j - Eg(X_j)\Gamma_j\right] \\ &\quad + E\left(\Gamma_j - E\Gamma_j\right)\left[\varepsilon_j\Gamma_j - E\varepsilon_j\Gamma_j\right] = cov(\Gamma_j, g(X_j)\Gamma_j). \end{aligned}$$

Similarly,

$$\begin{aligned} V(Y_j\Gamma_j) &= E(g(X_j) + \varepsilon_j)^2\Gamma_j^2 - \left[E(g(X_j) + \varepsilon_j)\Gamma_j\right]^2 \\ &= E(g^2(X_j) + 2g(X_j)\varepsilon_j + \varepsilon_j)^2\Gamma_j^2 - \left[Eg(X_j)\Gamma_j\right]^2 \\ &= V(g(X_j)\Gamma_j) + E\varepsilon_j^2 E\Gamma_j^2. \end{aligned}$$

From the last three equations we obtain

$$\begin{aligned} EX_{nj}^2 &= \frac{1}{n-m}\left[a_1^2 V(\Gamma_j) + 2a_1 a_2 cov(\Gamma_j, g(X_j)\Gamma_j)\right. \\ &\quad \left. + \, a_2^2 (V(g(X_j)\Gamma_j) + E\varepsilon_j^2 E\Gamma_j^2)\right] \\ &= \frac{1}{n-m}\left[V((a_1 + a_2 g(X_j))\Gamma_j) + a_2^2 E\varepsilon_j^2 E\Gamma_j^2\right]. \end{aligned}$$

By (van Es and Uh, 2005, p.477, line +9)

$$E\Gamma_j^2 \to 1/2, \;\; E\Gamma_j \to 0 \text{ as } n \to \infty. \tag{5.31}$$

From the last two equations we conclude that $\liminf_{n\to\infty}(n-m)EX_{nj}^2 \geq \begin{cases} a_2^2 \liminf_{j\to\infty}\sigma_j^2/2, & a_2 \neq 0; \\ a_1^2/2, & a_2 = 0. \end{cases}$

By independence, it follows that

$$s_n^2 = V(S_n) = \sum_{j=m+1}^{n} EX_{nj}^2 = \frac{1}{n-m}\sum_{j=m+1}^{n}(n-m)EX_{nj}^2 \geq c. \tag{5.32}$$

Now, remembering that $|\Gamma_j| \le 1$, we estimate

$$\begin{aligned}
\left(E|X_{nj}|^{2+\kappa}\right)^{\frac{1}{2+\kappa}} &= \frac{1}{\sqrt{n-m}}\left[E|(a_1+a_2Y_j)\Gamma_j - E(a_1+a_2Y_j)\Gamma_j|^{2+\kappa}\right]^{\frac{1}{2+\kappa}} \\
&\le \frac{2}{\sqrt{n-m}}\left[E|(a_1+a_2Y_j)\Gamma_j|^{2+\kappa}\right]^{\frac{1}{2+\kappa}} \\
&\le \frac{c(a)}{\sqrt{n-m}}\left[1+\left(E|Y_j|^{2+\kappa}\right)^{\frac{1}{2+\kappa}}\right]. \qquad (5.33)
\end{aligned}$$

Use (5.32) and (5.33) to check the Lyapounov condition for S_n:

$$\begin{aligned}
\frac{1}{s_n^{2+\kappa}}\sum_{j=m+1}^{n} E|X_{nj}|^{2+\kappa} &\le \frac{1}{c}(n-m)\left(\frac{c(a)}{\sqrt{n-m}}\right)^{2+\kappa} \\
&\quad \times\left[1+\sup_j\left(E|Y_j|^{2+\kappa}\right)^{\frac{1}{2+\kappa}}\right]^{2+\kappa} \to 0,\ n\to\infty.
\end{aligned}$$

The conclusion is that

$$S_n/s_n \xrightarrow{d} N(0,1). \qquad (5.34)$$

Since in our present notation $Z^p_{n,h}(x) = \frac{1}{\sqrt{n-m}}\sum_{j=m+1}^{n}\left[Y_j^p\Gamma_j - EY_j^p\Gamma_j\right]$, Σ is seen to be equal to

$$\Sigma = \lim_{n\to\infty}\frac{1}{n-m}\sum_{j=m+1}^{n}\begin{pmatrix} V(\Gamma_j) & cov(\Gamma_j, Y_j\Gamma_j) \\ cov(\Gamma_j, Y_j\Gamma_j) & V(Y_j\Gamma_j)\end{pmatrix}.$$

This and (5.30) show that

$$\lim_{n\to\infty} s_n^2 = \lim_{n\to\infty}\sum_{j=m+1}^{n} EX_{nj}^2 = a'\Sigma a.$$

By (5.34) then $S_n \xrightarrow{d} N(0, a'\Sigma a)$ which proves (5.29). □

Lemma 10. *Suppose $n\to\infty$, $h\to 0$ in such a way that $M_{n,h}\to 0$. Then the following asymptotic representations hold:*

$$\hat{f}_{T,k}(x) = E\hat{f}_{T,k}(x) + \frac{2kB_0M_{n,h}}{\pi(k+1)}Z^0_{n,h}(x) + o_p\left(M_{n,h}\right),$$

$$\hat{d}_{T,k}(x) = E\hat{d}_{T,k}(x) + \frac{2kB_0M_{n,h}}{\pi(k+1)}Z^1_{n,h}(x) + o_p\left(M_{n,h}\right).$$

Proof. We consider the proof for $\hat{f}_{T,k}(x)$ as an example, the other case being similar. First we bound a part of the last sum in (5.18). Denoting

$$T_{n,h}(x) = \sum_{j=m+1}^{n}\sum_{l=2}^{k}\frac{\theta_l}{h}(\Lambda_{j,l} - E\Lambda_{j,l})\int_{\varepsilon}^{1} F_K(t)\gamma_n\left(\frac{t}{lh}\right)dt$$

by independence and Lemma 6 we have

$$\begin{aligned}
V(T_{n,h}(x)) &\le \sum_{j=m+1}^{n} E\left(\sum_{l=2}^{k}\frac{\theta_l}{h}(\Lambda_{j,l}-E\Lambda_{j,l})\right)^2 I_0^2(n,lh)\\
&= \sum_{j=m+1}^{n}\sum_{t,s=2}^{k}\theta_s\theta_t E\Lambda_{j,s}\Lambda_{j,t}(1+o(1))\left(\frac{B_0}{n-m}\right)^2\left[(lh)^{\lambda_2(\alpha+1)+a_2-a_3-1}\zeta(lh)\right]^2\\
&\le c\sum_{l=2}^{k}M_{n,lh}^2 = o(1)M_{n,h}^2,\ \ h\to 0,
\end{aligned}$$

where $o(1)$ is uniform in n. This establishes that $T_{n,h}(x)=o_p(M_{n,h})$. This fact, (5.18) and Lemmas 3, 4, 6, 7 give

$$\begin{aligned}
\hat f_{T,k}(x)-E\hat f_{T,k}(x) &= O_p\left[\frac{1}{\sqrt{n-m}}h^{a_2-a_3-1}\zeta(h)^{\varepsilon^{\lambda_2}(1+o(1))}\right]+o_p\left[\frac{1}{\sqrt{n-m}}h^{a_2-a_3-1}\zeta(h)\exp\left(-\mu_3h^{-\lambda_3}\right)\right]\\
&+ O_p\left[\frac{1}{\sqrt{n-m}}h^{\lambda_2(\alpha+2)+a_2-a_3-2}\zeta(h)\right]+o_p(M_{n,h})+(1+o(1))\frac{2kB_0M_{n,h}}{\pi(k+1)}Z_{n,h}^0(x)\\
&= \frac{2kB_0M_{n,h}}{\pi(k+1)}Z_{n,h}^0(x)+o_p(M_{n,h}).
\end{aligned}$$

Here we used the fact that $Z_{n,h}^0(x)$ converges in distribution by Lemma 9 and is therefore $O_p(1)$. Due to the conditions $0<\varepsilon<1$ and $\lambda_2>1$ the expressions in the square brackets are $o_p(M_{n,h})$, as $h\to 0$. □

Proof of Theorem 3. Below we shall use the facts that $E\hat f_{T,k}(x)=f_T(x)+o(1)$ (see (5.9)) and $\hat f_{T,k}(x)=f_T(x)+o_p(1)$ (this follows from (5.9) and Lemma 10) and similar facts for $\hat d_{T,k}(x)$. By Lemma 10

$$\begin{aligned}
\hat\mu_k(x)-\frac{E\hat d_{T,k}(x)}{E\hat f_{T,k}(x)} &= \frac{\hat d_{T,k}(x)}{\hat f_{T,k}(x)}-\frac{E\hat d_{T,k}(x)}{E\hat f_{T,k}(x)}\\
&= \frac{\left[E\hat d_{T,k}(x)+\frac{2kB_0M_{n,h}}{\pi(k+1)}Z_{n,h}^1(x)+o_p\left(M_{n,h}\right)\right]E\hat f_{T,k}(x)}{\hat f_{T,k}(x)E\hat f_{T,k}(x)}\\
&- \frac{\left[E\hat f_{T,k}(x)+\frac{2kB_0M_{n,h}}{\pi(k+1)}Z_{n,h}^0(x)+o_p\left(M_{n,h}\right)\right]E\hat d_{T,k}(x)}{\hat f_{T,k}(x)E\hat f_{T,k}(x)}
\end{aligned}$$

After canceling out some terms and joining the terms of order $o_p(M_{n,h})$ this becomes

$$\begin{aligned}
\hat\mu_k(x)-\frac{E\hat d_{T,k}(x)}{E\hat f_{T,k}(x)} &= \frac{\frac{2kB_0M_{n,h}}{\pi(k+1)}\left[Z_{n,h}^1(x)E\hat f_{T,k}(x)-Z_{n,h}^0(x)E\hat d_{T,k}(x)\right]+o_p\left(M_{n,h}\right)}{\left[f_T(x)+o_p(1)\right]^2}\\
&= \frac{\frac{2kB_0M_{n,h}}{\pi(k+1)}\left[Z_{n,h}^1(x)f_T(x)-Z_{n,h}^0(x)d_T(x)\right]+o_p\left(M_{n,h}\right)}{\left[f_T(x)+o_p(1)\right]^2}.
\end{aligned}$$

The last equation shows that the limit in distribution of the variable on the left side of (3.7) is the same as that of

$$\Omega_{n,h} = \frac{2kB_0}{\pi(k+1)} \left(Z^1_{n,h}(x) - Z^0_{n,h}(x)\mu(x)\right)/f_T(x).$$

By the definition of Σ (3.8)

$$\begin{aligned} V\left(\Omega_{n,h}\right) &= \left(\frac{2kB_0}{\pi(k+1)f_T(x)}\right)^2 \left[V\left(Z^1_{n,h}(x)\right) - 2\mu(x)cov\left(Z^1_{n,h}(x), Z^0_{n,h}(x)\right)\right. \\ &+ \left.\mu^2(x)V\left(Z^0_{n,h}(x)\right)\right] \to \left(\frac{2kB_0}{\pi(k+1)f_T(x)}\right)^2 [\sigma_{22} - 2\mu(x)\sigma_{12} + \mu^2(x)\sigma_{11}]. \end{aligned}$$

□

References

Adams, R. A., Fournier, J., 2003. Sobolev Spaces, 2nd Edition. Vol. 140 of Pure and Applied Mathematics. Elsevier Science B.V., Amsterdam.

Altman, D. G., Bland, J. M., 1983. Measurement in medicine: the analysis of method comparison studies. The Statistician 32, 307–317.

Besov, O., Ilin, V., Nikolski, S., 1978. Integral representations of functions and embedding theorems. Wiley, New York, NY.

Billingsley, P., 1995. Probability and Measure, 3rd Edition. John Wiley and Sons.

Carroll, R. J., Delaigle, A., Hall, P., 2009. Nonparametric prediction in measurement error models. Journal of the American Statistical Association 104, 993–1014.

Carroll, R. J., Ruppert, D., Stefanski, L., Crainiceau, C., 2006. Measurement error in nonlinear models. No. 105 in Monographs in Statistics and Applied Probability. Chapman and Hall, New York, NY.

Cator, E. A., 2001. Deconvolution with arbitrarily smooth kernels. Statistics & Probability Letters 54, 205–214.

Chen, S. X., 2000. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 52, 471–480.

Cheng, C.-L., Riu, J., 2006. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. Technometrics 48, 511–518.

Delaigle, A., Fan, J., Carroll, R. J., 2009. A deisgn-adaptive local polynomial estimator for the errors-in-variables problem. Journal of the American Statistical Association 104, 348–359.

Delaigle, A., Meister, A., 2007. Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. Journal of the American Statistical Association 102, 1416–1426.

Delaigle, A., Meister, A., 2008. Density estimation with heteroscedastic error. Bernoulli 14, 562–579.

Fan, J., 1991. Asymptotic normality for deconvolution kernel density estimators. Sankhya 53, 97–110.

Fan, J., 1992. Deconvolution with supersmooth distributions. Canadian Journal of Statistics 20, 155–169.

Fan, J., Masry, E., 1992. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. Journal of Multivariate Analysis 43, 237–271.

Fan, J., Truong, Y., 1993. Nonparametric regression with errors in variables. Annals of Statistics 21, 1900–1925.

Fan, J., Truong, Y., Wang, Y., 1990. Measurements error regression: a nonparametric approach. Tech. Rep. 2042, University of North Carolina - Institute of Statistics.

Fan, Y., Liu, Y., 1997. A Note on asymptotic normality for deconvolution kernel density estimators. Sankhya 59, 138–141.

Honda, T., 2010. Nonparametric regression for dependent data in the errors-in-variables problem. Journal of Statistical Planning and Inference 140, 3409–3424.

Kulathinal, S. B., Kuulasmaa, K., Gasbarra, D., 2002. Estimation of errors-in-variables regrssion model when the variances of the measurement errors vary between the observations. Statistics in Medicine 21, 1089–1101.

Masry, E., 1993. Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. Journal of Multivariate Analysis 44, 47–68.

Meister, A., 2009. Deconvolution problems in nonparametric statistics. No. 193 in Lecture Notes in Statistics. Springer Verlag.

Mynbaev, K., 2012. Distributions escaping to infinity and the limiting power of the Cliff-Ord test for autocorrelation. ISRN Probability and Statistics (article number 926164), 1–39.

Mynbaev, K., Martins-Filho, C., 2010. Bias reduction in kernel density estimation via Lipschitz condition. Journal of Nonparametric Statistics 22, 219–235.

Nikolskii, S. M., 1977. Priblizhenie funktsii mnogikh peremennykh i teoremy vlozheniya, 2nd Edition. Nauka, Moscow.

Staudenmayer, J., Ruppert, D., Buonnacorsi, J., 2008. Density estimation in the presence of heteroskedastic measurement error. Journal of the American Statistical Association 103, 726–736.

Triebel, H., 1983. Theory of Function Spaces. Birkhauser Verlag, Basel.

van Es, A. J., Uh, H.-W., 2004. Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the Cauchy boundary. Journal of Nonparametric Statistics 16, 261–277.

van Es, B., Uh, H.-W., 2005. Asymptotic normality of kernel-type deconvolution estimators. Scandinavian Journal of Statistics 32, 467–483.

Walter, S. D., 1997. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. Statistics in Medicine 16, 2883–2900.

Zhang, C.-H., 1990. Fourier methods for estimating mixing densities and distributions. Annals of Statistics 18, 806–831.