Kernel estimation when density does not exist

Victoria Zinde-Walsh*

McGill University and CIREQ
Victoria.zinde-walsh@mcgill.ca
(514) 398 4834

April 19, 2005

Abstract

Nonparametric kernel estimation of density is widely used. However, many of the pointwise and global asymptotic results for the estimator are not available unless the density is continuous and appropriately smooth; in kernel estimation for discrete-continuous cases smoothness is required for the continuous variables. Some situations of interest may not satisfy the smoothness assumptions. In this paper the asymptotic process for the kernel estimator is examined by means of the generalized functions and generalized random processes approach according to which density and its derivatives can be defined as generalized functions. The limit process for the kernel estimator of density (whether density exists or not) is characterized in terms of a generalized Gaussian process. Conditional mean and its derivatives can be expressed as values of functionals involving generalized density; this approach makes it possible to extend asymptotic results, in particular those for asymptotic bias, to models with non-smooth density.

*The support of the Social Sciences and Humanities Research Council of Canada (SSHRC), the *Fonds québecois de la recherche sr la société et la culture* (FRQSC) is gratefully acknowledged. I thank participants of the COMPSTAT 2004 and RSS 2004 meetings, participants of the seminar at Vanderbilt University and Dalhousie University and Yanqin Fan for very useful comments and suggstions.

1 Introduction

Many non- or semi-parametric estimators utilize nonparametric kernel estimators of density. For the properties of these estimators and in particular for asymptotic results, the existence of density and some smoothness properties of the density are routinely assumed (see, e.g Pagan and Ullah, 1999 for a review). In the literature that develops asymptotic results for estimators of conditional mean, its derivative and average derivatives, assumptions on smoothness of the density function are made as well as those about smoothness of conditional mean (see Pagan, Ullah for review). However, while the conditional mean is often smooth (even linear) and satisfies various types of conditions on derivatives that follow from some theoretical model, there is no theoretical basis for assuming density smoothness. Indeed, there are examples where it is natural to assume that the density of some variables is discontinuous.

Suppose that an income distribution is smooth with density $f(x)$. Define by x_a the after-tax income; if the income becomes taxable at point a it will be related to x as

$$x_a = \begin{cases} x & \text{for } x < a; \\ (1-r)x & \text{for } x \geq a. \end{cases}$$

Then the density function f_a for x_a is

$$f_a(x_a) = \begin{cases} f(x_a) & \text{for } x_a < a; \\ (1-r)^{-1} f(\frac{x_a}{1-r}) & \text{for } x_a \geq a \end{cases}$$

and as long as $(1-r)f(x_a) \neq f(\frac{x_a}{1-r})$ the density f_a is discontinuous. Similarly discontinuity of density will arise if x_a includes some lump-sum tax or transfer. In a model of household consumption or work decision by a spouse it is reasonable to assume that the dependence of conditional expectation on x_a is continuous and possibly smooth. Other examples of smooth model dependence on variables whose distribution is not smooth also arise naturally. Thus relaxing smoothness assumptions on densities may be desirable even at the price of imposing stronger smoothness assumptions on conditional expectation.

This paper starts by addressing the question: What does a kernel estimator estimate, pointwise and globally, if density is not continuous or does not exist?

The case where some of the variables are discrete has been examined in the literature. Results on joint estimation in combined discrete-continuous cases (see, e.g. Ahmad and Cerrito, 1994) extend the area of coverage of kernel density estimation to cases when the set of variables can be partitioned into two subsets: one of variables in which the distribution is absolutely continuous and the density is at least twice continuously differentiable, and the other of discrete variables. This leaves out cases where the density exists, but is not continuous for non-discrete variables, or even if continuous is not differentiable in some of the continuous variables. Therefore the same questions as above apply to the non-discrete variables in the estimation of the joint discrete-continuous density.

Here in order to extend the concept of density where it may not exist as an ordinary (locally summable) function, we use the apparatus of generalized functions, or distributions as Laurent Schwartz (1950) called them; generalized functions are useful in cases of non-differentiability since they allow characterization of generalized derivatives when derivatives do not exist as ordinary functions (e.g. for a non-differentiable distribution function). Some useful references are Halperin (1952) who provided an English introductory version of L.Schwartz's lectures and Gel'fand and Shilov (v. 1,2, 1964) where the main introductory results on generalized functions are collected. Gel'fand and Vilenkin (v.4, 1964) is the main reference for generalized random processes. Reference is made also to Sobolev's (1992) monograph which makes use of generalized functions in approximation of multivariate integrals; in the introductory chapter the book provides results on kernel averaging and also useful diagrams of embedding mappings of spaces of generalized functions. The results used in the present paper are collected in Appendix B.

Generalized functions are widely used in mathematics and physics for solving differential equations. P.C.B.Phillips drew attention to the usefulness of the generalized functions approach for cases of non-differentiability in econometrics; in 1991 he proposed it in application to the asymptotics of the LAD estimator and then in 1995 successfully applied it to derive the limit process for nonstationary LAD regression. In Zinde-Walsh (2002) the limit process for the least median of squares estimator is described in terms of a generalized Gaussian process; Zinde-Walsh and Phillips (2003) derived the generalized Gaussian random process that represents the derivative of the fractional Wiener process; Zinde-Walsh and Phillips (2004) provide a description of limit processes for various extremum estimators with non-differentiable criterion functions as generalized Gaussian random proceses.

Here we call "density" the generalized derivative, f, which may or may not be an ordinary function, of the distribution function, F; it represents a linear continuous functional on a space D of special "test functions" so that for any $\psi \in D$ the value of the functional (f, ψ) is well defined[1]. The kernel density estimator as the sample size goes to ∞ and the bandwidth parameter goes to zero converges in probability to the generalized derivative of the distribution function which may or may not exist as an ordinary function; if "density" exists as an ordinary function and is continuous at point x the estimator converges to the value of the density function, $f(x)$.

The kernel estimator has a limit process that can be described as a generalized random process; under the usual assumptions it can be represented as a generalized Gaussian process. A full characterization of this generalized Gaussian process is provided here. The kernel estimator of the generalized density function has the following interpretations: (1) the integral of the estimator consistently estimates the ordinary distribution function; (2) if the estimator is smoothed it consistently estimates the value of the density functional applied to the smoothing function; (3) the kernel estimator is useful for estimation of values of functionals involving density as long as they can be properly defined. We examine conditional expectation, its derivatives, average derivatives; as long as the conditional expectation function is sufficiently smooth we can utilize the generalized functions approach to derive expansions for the asymptotic bias functional.

The paper is organized as follows. Section 2 provides interpretation of the distribution function as a generalized function and of the density as its generalized derivative; local properties are defined and the kernel estimator is interpreted as an estimator of the generalized density. In Section 3 the limit process for the kernel estimator is derived and shown to be a generalized Gaussian process. Section 4 discusses relaxing smoothness assumptions for some estimators used in non and semi-parametric estimation. Appendix A provides proofs of the results of the paper. Appendix B gives a collection of definitions and results about generalized functions and generalized random processes that are used here.

[1]Unless F is absolutely continuous the generalized derivative may not coincide with the pointwise derivative even when it is an ordinary function.

2 Distribution function and density as generalized functions and the kernel estimator

2.1 Distribution function, density and its derivatives as generalized functions

The definitions and results about generalized functions are collected in Appendix B; this subsection specializes these results to distribution functions.

Define for a random vector $x \in R^k$ its distribution function $F(x)$; it can be defined as a monotonic and cadlag bounded function, thus it is an ordinary (locally summable) function on R^k as discussed here in Appendix B (BI.2.a,b).

We start with the univariate case $k = 1$. As a generalized function $F(x)$ can be represented by a functional on the generic function space D. That space could be the space K of infinitely differentiable functions with finite support, or the space S of infinitely differentiable functions that go to zero at infinity faster than any power, or any of the spaces D_m of m times continuously differentiable functions (with finite support) defined in B I.2. For any $\psi \in D$ define the value of the functional as in BI.2.a:

$$(F, \psi) = \int F(x)\psi(x)dx. \tag{1}$$

Then as long as D contains continuously differentiable functions (e.g. coincides with K, S, or any D_m, $m \geqslant 1$) we can define density as a generalized derivative of F :

$$(f, \psi) = (F', \psi) = -(F, \psi') = -\int F(x)\psi'(x)dx. \tag{2}$$

A similar relation holds for multivariate densities. If the density exists as an ordinary function for a distribution function $F(x_1, ..., x_k)$ then the density can be defined as an ordinary function

$$f(x) = f(x_1, ..., x_k) = \frac{\partial^k F(x_1, ..., x_k)}{\partial x_1 ... \partial x_k}; \tag{3}$$

$f(x)$ integrates to $F(x)$ if it is absolutely continuous. Whether (3) exists as an ordinary function or not, it can be defined as a generalized function: for any $\psi \in D(R^k)$, where functions in D are suitably differentiable ($D \subseteq D_k$, see BI.2.c),

$$(f, \psi) = (-1)^k (F, \frac{\partial^k \psi(x_1, ..., x_k)}{\partial x_1 ... \partial x_k}). \tag{4}$$

Generalized derivatives of the density function are defined by formulas similar to (4), e.g. in the univariate case the generalized derivative of the density function, f', is given for any $\psi \in D$ (as long as $D \subseteq D_2$) by

$$(f', \psi) = -(f, \psi') = (F, \psi'').$$

2.2 Distribution function and density function locally at a point

For local properties of the distribution and density functions (generalized functions) we can consider the distribution around the point of interest x for $\tilde{x}$ in some small $h-$neighbourhood of x. We introduce a kernel function, K.

ASSUMPTION A.

a. $K(w)$ is an ordinary bounded function on R^k; $\int K(w)dw = 1$;

b. Support of K belongs to $[-1, 1]^k$;

c. $K(w)$ is an $l-$th order kernel: for $w = (w_1, ...w_k)$ the integral

$$\int w_1^{j_1} ... w_k^{j_k} K(w) dw_1 ... dw_k \begin{cases} = 0 & \text{if } \Sigma j_i < l; \\ \neq 0 & \text{for some } (j_1, ..., j_k) \text{ with } \Sigma j_i = l. \end{cases}$$

and is finite.

If $l = 1$ Assumption 1 reduces to a. and b. The finite support assumption can be relaxed and is introduced to simplify assumptions and derivations; K is not restricted to be symmetric or non-negative.

For $k = 1$ define $K_h(\tilde{x}, x) = \frac{1}{h} K(\frac{\tilde{x}-x}{h})$. Note that $\int K_h(\tilde{x}, x) d\tilde{x} = \int K(w)dw$. Then for any h and any fixed x define

$$F_{hK}(x) \equiv \int F(\tilde{x}) K_h(\tilde{x}, x) d\tilde{x} = \frac{1}{h} \int F(\tilde{x}) K(\frac{\tilde{x} - x}{h}) d\tilde{x}. \tag{5}$$

This provides the value of F_{hK} at x as a weighted average of values of the function $F(\tilde{x})$ in the h neighbourhood[2]. Once x is allowed to vary $F_{hK}(x)$

[2] Averaging of a generalized function by a kernel function was considered by e.g. Sobolev (1992) who provided the proof of a statement similar to our Theorem 1.

can be viewed as a functional (generalized function) defined for functions $\psi(x) \in D$ (but it is also an ordinary function of x).

An analogous construction applies to the multivariate case. Where there is no ambiguity we shall write, for the multivariate case and for $h = (h_1, ...h_k)$

$$\begin{aligned} x &= (x_1, ...x_k); w = (w_1, ..., w_k); dw = dw_1...dw_k; \\ \frac{\tilde{x} - x}{h} &= \left(\frac{\tilde{x}_1 - x_1}{h_1}, ...\frac{\tilde{x}_k - x_k}{h_k}\right). \end{aligned} \tag{6}$$

Define for kernel function K

$$K_h(\tilde{x}, x) = \frac{1}{\Pi h_i} K(\frac{\tilde{x}_1 - x_1}{h_1}, ..., \frac{\tilde{x}_k - x_k}{h_k}) = \frac{1}{\Pi h_i} K(\frac{\tilde{x} - x}{h});$$

then $\int K_h(\tilde{x}, x) d\tilde{x} = \int K(w) dw$. The functional $F_{hK}(x)$ is defined similarly to the univariate case:

$$F_{hK}(x) = \int F(\tilde{x}) K_h(\tilde{x}, x) d\tilde{x}.$$

The following theorem establishes convergence of $F_{hK}(x)$ to $F(x)$ as $h \to 0$ as generalized functions. To distinguish convergence of generalized functions (weak convergence of linear continuous functionals on the space D) from ordinary pointwise convergence we denote it by $\Rightarrow$ as opposed to $\to$. To distinguish between different spaces on which the functionals are defined we could subscript $\Rightarrow$ by the corresponding space, e.g. $\Rightarrow_{D_n}$. Usually it is clear for which spaces the the convergence holds and the subscript is omitted. For the multivariate case we employ the notation $h = \max\{h_i\}$.

Theorem 1 *As $h \to 0$ for K that satisfies Assumption A*

$$F_{hK}(x) \Rightarrow F(x),$$

in other words, for any $\psi(x) \in D$ with $D \subset D_0$ we have

$$(F_{hK}(x), \psi(x)) \to (F(x), \psi(x)),$$

if F is continuous at x then $F_{hK}(x) \to F(x)$.

Proof. See Appendix A.

Next, consider the generalized derivative $f_{hK}(x)$ of $F_{hK}(x)$ that corresponds to the generalized density function locally to point x. Write for the univariate case for a given x

$$f_{hK}(x) = F'_{hK}(x). \tag{7}$$

Similarly

$$f_{hK}(x) = \frac{\partial^k F_{hK}(x)}{\partial x_1 ... \partial x_k}$$

in the multivariate case. Of course, if F were absolutely continuous in the neighbourhood of x with ordinary density function $f(x)$ this would be

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{1}{h}\int F(\tilde{x})K(\frac{\tilde{x}-x}{h})d\tilde{x}\right) &= \frac{\partial}{\partial x}\left(\int F(x+hw)K(w)dw\right) \\ &= \int f(x+hw)K(w)dw \end{aligned}$$

and similarly in the mulivariate case.

Assuming that K is a continuously differentiable function (7) is

$$\frac{\partial}{\partial x}\left(\frac{1}{h}\int F(\tilde{x})K(\frac{\tilde{x}-x}{h})d\tilde{x}\right) = -\frac{1}{h^2}\int F(\tilde{x})K'(\frac{\tilde{x}-x}{h})d\tilde{x},$$

thus

$$f_{hK}(x) = -\frac{1}{h}F_{hK'}(x) \tag{8}$$

is an ordinary function. Similarly

$$f_{hK}(x) = (-1)^k \frac{1}{(\Pi h_i)^2}\int F(\tilde{x})\frac{\partial^k K(\frac{\tilde{x}-x}{h})}{\partial x_1 ... \partial x_k}d\tilde{x} = (-1)^k \frac{1}{\Pi h_i}F_{hK'}(x) \tag{9}$$

in the multivariate case.

If K is not assumed to be differentiable (e.g. is a rectangular kernel) the definition (8) still holds but can be understood only as equality of generalized functions implying for $\psi \in D$:

$$(f_{hK}(x), \psi(x)) = (-\frac{1}{h}F_{hK'}(x), \psi(x)) = \frac{1}{h}(F_{hK}(x), \psi'(x))$$

for all $\psi \in D$, $D \subseteq D_1$; a similar relation can be written for the multivariate case. Thus note that as long as either F or K is continuously differentiable $f_{hK}(x)$ is an ordinary function; in any case it is a generalized function.

We say that a sequence of generalized functions $g_n = O(r(n))$ for $n \to \infty$ and some number sequence $r(n)$ if for any $\psi \in D$ we get $(g_n, \psi) = O(r(n))$.

Theorem 2 *As $h \to 0$ and assuming that K satisfies Assumption A*

(a) convergence of generalized functions on R^k,

$$f_{hK}(x) \Rightarrow f(x),$$

holds for $D \subset D_k$; if $F(x)$ is absolutely continuous in Ω and $f(x)$ is continuous at x this coincides with ordinary convergence: $f_{hK}(x) \to f(x)$ in Ω;

(b) if $\psi \in D_{l+k}$ and K is a kernel of order l

$$f_{hK}(x) - f(x) = O(h^l).$$

Proof. See Appendix A.

Thus with the help of the function K (kernel function) and for $h \to 0$ we have constructed sequences of generalized functions (which may be ordinary functions e.g. if K is differentiable) that have support in a neighbourhood of x and as generalized functions of x converge to the generalized derivative of the distribution function. The convergence rate can be controlled by appropriate selection of K and space D (as follows from (b) of Theorem 2).

The following example illustrates the case where ordinary convergence of f_{hK} does not hold.

Example 1 *Suppose that in some region Ω the distribution function can be defined as $F(x) = \alpha I(x - y)$ for some fixed y; K is differentiable. Then it is easy to compute that as $h \to 0$ convergence of ordinary functions*

$$hf_{hK}(x) \to \begin{cases} 0, & \text{if } x \neq y; \\ \alpha K(0), & \text{if } x = y \end{cases}$$

holds. Thus in this case ordinary convergence for f_{hK} does not obtain. It is easy to verify that as generalized functions $f_{hK}(x) \Rightarrow \alpha\delta(x - y)$, where the generalized function δ is Dirac's δ-function:

$$(\delta(x - y), \psi(x)) = \psi(y).$$

2.3 Kernel estimator and its relation to generalized density

Consider a multivariate density; recall the notation in (6). From (9) integrating by parts we have

$$\begin{aligned} f_{hK}(x) &= (-1)^k \frac{1}{(\Pi h_i)^2} \int F(\tilde{x}) \frac{\partial^k K(\frac{\tilde{x}-x}{h})}{\partial x_1 ... \partial x_k} d\tilde{x} \\ &= \frac{1}{\Pi h_i} \int K(\frac{\tilde{x}-x}{h}) dF(\tilde{x}) = E_{\tilde{x}} K_h(\tilde{x}, x). \end{aligned}$$

If f_{hK} is not an ordinary function the equality may hold as equality of generalized functions only. A natural estimator for $f_{hK}(x)$ follows from the fact that it is an expectation; a sample average is used in estimation. The estimator based on a random sample of n observations $\{x_i\}$ from the distribution of x is

$$\widehat{f_{hK}(x)} = \frac{1}{n\Pi h_i} \sum_{i=1}^{n} K(\frac{x_i - x}{h}) \tag{10}$$

and $E\widehat{f_{hK}(x)} = f_{hK}(x)$.

We can thus interpret the kernel density estimator as an estimator of the value of the local generalized density functional whether density exists as an ordinary function or not.

For a sequence of generalized random functions $\{g_n\}$ we write $g_n \approx o_p(\alpha(n))$ (or $O_p(\alpha(n))$) if for any $\psi \in D$ the sequence of ordinary random variables $(g_n, \psi)/\alpha(n)$ converges in probability to zero (or is uniformly bounded in probability for any bounded neighbourhood of zero in D).

Theorem 3 *As $n \to \infty$ for any fixed h and K that satisfies Assumption A*

$$\widehat{f_{hK}(x)} - f_{hK}(x) \approx O_p(n^{-\frac{1}{2}}). \tag{11}$$

If all the bandwidths $h_i \to 0$ as $n \to \infty$ the conditions of Theorem 2 are satisfied for $\Omega = R^k$ and $n\Pi_{i\in c} h_i \to \infty$

$$\widehat{f_h(x)} - f(x) \approx o_p(1) \tag{12}$$

as a generalized random process. If f is continuous at x then

$$\widehat{f_{hK}(x)} - f(x) = o_p(1).$$

Proof. See Appendix A.

We see that the kernel estimator of density for fixed h converges as a generalized random process with the standard rate to $f_{hK}(x) = EK_{hx}$. As $h \to 0$ convergence in probability of the generalized random variables to $f(x)$ (the generalized density function) obtains.

3 Limit process for kernel estimator of generalized density

We now describe the limit process for the kernel estimator as a generalized random process whether the density exists or not. Note that since $E\widehat{f_{hK}(x)} = f_{hK}(x)$ part (b) of Theorem 2 provides the convegence rate for the generalized bias function of the kernel estimator.

The following theorem provides results for the limit process of the kernel density estimator.

Theorem 4 *For a kernel function K satisfying Assumption A, if the bandwidth $h \to 0$ as $n \to \infty$ with $n\Pi_{i\in c}h_i \to \infty$and $h^{2l+k}n \to 0$ the sequence of generalized random processes $(n\Pi_{i\in c}h_i)^{\frac{1}{2}}\left(\widehat{f_{hK}(x)} - f(x)\right)$ converges to a generalized Gaussian process with mean functional zero and covariance functional C which for any (linearly independent) $\psi_1, \psi_2 \in D(\Omega)$ provides*

$$\begin{aligned}(C,(\psi_1,\psi_2)) &= \int \psi_1(x)\psi_2(x)f(x)dx \int K(w)^2dw \qquad (13)\\ &= E(\psi_1(x)\psi_2(x)) \int K(w)^2dw;\end{aligned}$$

Proof. See Appendix.

Note that by Gel'fand, Vilenkin (1964) derivatives of the kernel estimator have as a limit process the generalized Gaussian process with mean functional zero and covariance functional given by generalized derivatives of the limit covariance functional for the kernel estimator itself, and so the covariance functional for the limit process for the derivatives of the kernel density estimator can be derived from C.

4 Relaxing asssumptions on smoothness of density in nonparametric estimation

We consider here estimators of conditional mean and its derivatives and average derivatives in the situation where density may not be smooth (or not exist). We concentrate here on the asymptotic bias of the estimators since the variance and asymptotic normality results do not rely on long expansions.

Consider the conditional expectation function

$$m(x) = E(y\,|x\,) = \int yf(y\,|x\,)dy$$

and the kernel estimator

$$\widehat{m(x)} = \frac{\Sigma y_j K(\frac{x_j - x}{h})}{\Sigma K(\frac{x_j - x}{h})} = \frac{\Sigma y_j K(\frac{x_j - x}{h})}{\widehat{f(x)}}.$$

For the derivative of conditional expectation , $\frac{\partial m(x)}{\partial x}$, consider the estimator given by $\frac{\partial \widehat{m(x)}}{\partial x}$. Denote the average derivative $E(\frac{\partial m}{\partial x}(x))$ by δ and by $\hat{\delta}$ the estimator of average derivative that is given by

$$\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \widehat{m(x_i)}}{\partial x}.$$

Consider all bandwidths to be equal: $h_i = h$.

Assumption B

B1. The joint distribution function $F(x, y)$ is well defined on R^{k+1} as an ordinary function; the marginal distribution function $F_x(x)$ is similarly defined as an ordinary function with the generalized marginal density function $\frac{\partial F_x(x)}{\partial x}$ having support in an open convex subset $\Omega \subset R^k$. The associated measure is $\mu_{(x,y)} = \mu_x \times \mu_y$.

B.2. The kernel K satisfies Assumption A for $l \geq 2$.

B.3. $K \in D_m$.

B.4. The function $m(x)$ is v times continuously differentiable on Ω and for every partial derivative of order v the expectation $E\left|\frac{\partial^v m(x)}{\partial^{l_1} x_1 \ldots \partial^{l_k} x_k}\right| < B < \infty$.

B.5. For some $b > 0$ over $x \in \Omega$

$$\sup_{\Omega} \Pr(I(\left|\widehat{f(x)}\right| > b)) \to 0.$$

B.6. Asymptotic variance of $(nh^k)^{\frac{1}{2}}\widehat{m(x)}$ exists.

B.7. Asymptotic variance of $(nh^{k+1})^{\frac{1}{2}}\widehat{\frac{\partial m(x)}{\partial x}}$ exists.

B.8. Asymptotic variance of $\sqrt{n}\hat{\delta}_n$ exists.

B.9. $h^k n \to \infty$ and $nh^{2l+k} \to 0$.

B.10. $h^{k+1}n \to \infty$ and $nh^{2l+k+1} \to 0$.

The assumptions made here differ from the ones usually made in the literature in the following ways. We require here more smoothness from the conditional mean function while making no such requirements on the marginal density which may not be an ordinary function but only exist as a generalized function. Assumption B.5. is made for the estimator of density (ordinary function) rather than density itself; if density exists as an ordinary function and is bounded away from zero B.5 follows. To focus on the asymptotic bias we assume existence of asymptotic variance without specifying conditions for that; proofs of existence of asymptotic variance do not need to rely on expansions of the density function; the proofs in the literature (see e.g. Pagan and Ullah) that use smoothness of density can be modified to use smoothness of the K function instead. The following theorem provides results for the asymptotic bias of the conditional mean estimator and estimator of its derivative.

Theorem 5 *Under assumptions B.1, B.2, B.3 for $m = l + k$, B.4 for $v = l + 1$, B.5, B.6 and B.9 for $x \in \Omega$*

(a)

$$abias\left[(nh^k)^{\frac{1}{2}}(\widehat{m(x)} - m(x))\right] = 0;$$

(b) if additionally B.3 holds for $m = l + k + 1$, B.4 for $v = l + 2$, and B.7 d B.10 hold

$$abias\left[(nh^{k+1})^{\frac{1}{2}}(\frac{\partial \hat{m}}{\partial x}(x) - \frac{\partial m}{\partial x}(x))\right] = 0.$$

Proof. See Appendix A.

For the average derivative estimator $\hat{\delta}$ the results can be extended to provide the same rate of asymptotic bias by strengthening the assumption B.5 to provide a suitable rate of decline over Ω, as e.g. in Hardle and Stoker (1989). Note however, that the estimator in that paper is not $\hat{\delta}$, and the argument by which it is obtained in integration by parts requires smoothness of density.

For density weighted average derivatives introduced in Powell, Stock and Stoker (1989) this approach will not work because where the proof of Theorem 5 required differentiability of $m(x)$ (and $m'(x)$), in the density weighted case a similar requirement will have to apply to $m'(x)f(x)$, implying smoothness of the density function as well as of the conditional mean.

Appendix A.

Proof of Theorem 1.

We have

$$\begin{aligned} & \lim_{h\to 0}\int\int F(x+hw)K(w)dw\psi(x)dx \\ = & \lim_{h\to 0}\int\int F(y)\psi(y-hw)K(w)dwdy \\ = & \int F(y)\psi(y)dy\int K(w)dw=(F,\psi) \end{aligned}$$

where the second equality follows from ψ being a continuous function. If F is continuous at x then by interchanging the integral and limit in the first line we get

$$\lim_{h\to 0}\int F(x+hw)K(w)dw=F(x).$$

■

Proof of Theorem 2.

(a) Since ψ is a continuously differentiable function ($\psi\in D_k$)

$$\begin{aligned} (f_{hK},\psi) &= (-1)^k\int\frac{\partial^k}{\partial x_1...\partial x_k}\left(\frac{1}{\Pi h_i}\int F(\tilde{x})K(\frac{\tilde{x}-x}{h})d\tilde{x}\right)\psi(x)dx \quad (14)\\ &= \int\left(\frac{1}{\Pi h_i}\int F(\tilde{x})K(\frac{\tilde{x}-x}{h})d\tilde{x}\right)\frac{\partial^k}{\partial x_1...\partial x_k}\psi(x)dx \\ &= (-1)^k\int\int F(\tilde{x})\frac{\partial^k}{\partial x_1...\partial x_k}\psi(\tilde{x}-hw)K(w)dwdy \\ &\to (f,\psi)\int K(w)dw=(f,\psi) \end{aligned}$$

by continuity of $\frac{\partial^k}{\partial x_1...\partial x_k}\psi$ and $\int K=1$.

Consider f_{hK} defined in (7). For F univariate continuously differentiable in the neighbourhood of x with ordinary density function $f(x)$

$$\begin{aligned} f_{hK}(x) &= \frac{\partial}{\partial x}\left(\frac{1}{h}\int F(\tilde{x})K(\frac{\tilde{x}-x}{h})d\tilde{x}\right) = \int f(x+hw)K(w)dw \\ &\to f(x)\int K(w)dw = f(x) \end{aligned}$$

by the assumption $\int K = 1$, similarly in the multivariate case.

(b) For the univariate case

$$\begin{aligned} (f_{hK},\psi)-(f,\psi) &= -\int F(\tilde{x})[\psi'(\tilde{x}-hw)-\psi'(\tilde{x})]K(w)dw \\ &= (-1)^l\frac{h^l}{l!}\int F(\tilde{x})\psi^{l+1}(\tilde{x})d\tilde{x}\int K(w)w^l dw + o(h^l) \end{aligned}$$

which follows from expansion of $\psi'(\tilde{x}-hw)$ and the order of kernel. Consider now the multivariate case. For simplicity assume that $\psi(x)=\psi_1(x_1)...\psi_k(x_k)$ with $\psi_i \in D$; recall that any multivariate $\psi \in D$ can be approximated by such products. Expanding the product function ψ

$$\frac{\partial^k}{\partial x_1...\partial x_k}\psi(\tilde{x}-hw) = \prod_{i=1}^{k}\left[\sum_{m=1}^{l+1}\frac{\partial^m\psi_i}{\partial x_i^m}(\tilde{x}_i)\frac{(-1)^{m-1}}{(m-1)!}(h_iw_i)^{m-1}+O(h_i^{l+1})\right],$$

substituting into the third line of (14), subtracting the limit and using the order of kernel we get that

$$\begin{aligned} &(f_{hK},\psi)-(f,\psi) \\ = \ &(-1)^l\sum_{m_1+...+m_k=l}\int F(\tilde{x})\prod_{i=1}^{k}\frac{\partial^{m_i+1}\psi_i}{\partial x_i^{m_i+1}}(\tilde{x}_i)\frac{h_i^{m_i}}{m_i!}d\tilde{x}+O(h^{l+1}) = O(h^l). \end{aligned}$$

■

Proof of Theorem 3.

We have that

$$E\widehat{f_{hK}(x)} = E_{\tilde{x}}\frac{1}{\Pi h_i}K(\frac{x-\tilde{x}}{h}) = f_{hK}(x);$$

variance is

$$E[f_{hK}(x)-f_{hK}(x)]^2 = \frac{1}{n}E_{\tilde{x}}[\frac{1}{\Pi h_i}K(\frac{x-\tilde{x}}{h})-\frac{1}{\Pi h_i}E(K(\frac{x-\tilde{x}}{h}))]^2.$$

For a fixed h the value $E_{\tilde{x}}[\frac{1}{\Pi h_i}K(\frac{x-\tilde{x}}{h}) - \frac{1}{\Pi h_i}E(K(\frac{x-\tilde{x}}{h}))]^2$ is bounded and (11) follows. If $h \to 0$ the value $E_{\tilde{x}}[\frac{1}{\Pi h_i}K(\frac{x-\tilde{x}}{h}) - \frac{1}{\Pi h_i}E(K(\frac{x-\tilde{x}}{h}))]^2 = O((\Pi h_i)^{-1})$. For (12) write for $\psi \in D$

$$\begin{aligned} & \int \left[\widehat{f_{hK}(x)} - f(x)\right] \psi(x)dx \\ = & \int [\widehat{f_{hK}(x)} - f_{hK}(x)]\psi(x)dx + \int [f_{hK}(x) - f(x)]\psi(x)dx. \end{aligned}$$

The first bracketed term as $n \to \infty, h \to 0, n\Pi h_i \to \infty$ is $O_p((n\Pi h_i)^{-\frac{1}{2}})$, the second by Theorem 2 is $o(1)$.

■

Proof of Theorem 4.

Define a (generalized) function

$$e_{nhj}(x) = \frac{1}{\Pi h_i}K(\frac{x - x_j}{h}) - f(x)$$

and consider $e_{hn}(x) = \frac{1}{n}\sum_{j=1}^{n} e_{nhj}(x)$; it equals $\widehat{f_{hK}(x)} - f(x)$.

By Theorem 2 (b) $Ee_{hni}(x) \approx O(h^l)$.

Next consider $T_{ij} = E(e_{hni}(x), \psi_1)(e_{hnj}(x), \psi_2))$.

For $i \neq j$ by independence

$$\begin{aligned} E(T_{ij}) &= E(e_{hni}(x), \psi_1)(e_{hnj}(x), \psi_2) = E(e_{hni}(x), \psi_1)E(e_{hnj}(x), \psi_2) \\ &= O(h^{2l}). \end{aligned}$$

For $i = j$

$$\begin{aligned} E(T_{ii}) &= E(e_{hni}(x), \psi_1)(e_{hni}(x), \psi_2) \\ &= \int \left[\int \frac{1}{\Pi h_i}K(\frac{x_i - x}{h})\psi_1(x)dx - \int f(x)\psi_1(x)dx\right] \times \\ & \left[\int \frac{1}{\Pi h_i}K(\frac{x_i - x}{h})\psi_2(x)dx - \int f(x)\psi_2(x)dx\right] f(x_i)dx_i. \end{aligned}$$

To express this as a bilinear functional applied to (ψ_1, ψ_2) the order of integration has to be changed. Consider now for any fixed (x, y)

$$\int \frac{1}{\Pi h_i^2}K(\frac{x_i - x}{h})K(\frac{x_i - y}{h})f(x_i)dx_i.$$

Substituting $\frac{x_i - x}{h} = w$ this becomes

$$\int \frac{1}{\Pi h_i} K(w) K(w + \frac{x-y}{h}) f(x + hw) dw;$$

if $x - y \neq 0$ for small enough h we have, e.g. $\left|\frac{x-y}{h}\right| > 2\left|support(K)\right|$ and then $K(w + \frac{x-y}{h}) = 0$. If $x = y$ this expression multiplied by Πh_i becomes

$$\int K(w)^2 f(x + hw) dw.$$

By Theorem 3 as $h \to 0$ it converges as a generalized function to $f(x) \int K(w)^2 dw$. Thus $\Pi h_i T_{ii} \to \int K(w)^2 dw E(\psi_1 \psi_2)$.

Combining we get the limit covariance matrix.

Consider now

$$\begin{aligned} \eta_{hni}(x) &= n^{\frac{1}{2}} \Pi h_i^{\frac{1}{2}} e_{hni}(x) - E(n^{\frac{1}{2}} \Pi h_i^{\frac{1}{2}} e_{hni}(x)); \\ \eta_{hn}(x) &= n^{\frac{1}{2}} \Pi h_i^{\frac{1}{2}} e_{hn}(x) - E(n^{\frac{1}{2}} \Pi h_i^{\frac{1}{2}} e_{hn}(x)) = \frac{1}{n} \sum \eta_{hni}(x). \end{aligned} \quad (15)$$

This generalized random function has expectation zero. In the covariance $E(\eta_{hni}(x), \psi_1)(e_{\eta hnj}(x), \psi_2)$ the terms where $i \neq j$ are zero and

$$E(\eta_{hni}(x), \psi_1)(e_{\eta hnj}(x), \psi_2)$$

is $O(1)$ and converges to $\int K(w)^2 dw E(\psi_1 \psi_2)$.

Next we show that for any set of linearly independent functions $\psi_1, ..., \psi_m \in D$ the joint distribution of the vector $\overrightarrow{\eta}_{hn} = ((\eta_{hn}, \psi_1)..., (\eta_{hn}, \psi_m))'$ converges to a multivariate Gaussian. Define similarly the vector $\overrightarrow{\eta}_{hni}$ with components (η_{hni}, ψ_l). Denote by Σ the $m \times m$ matrix with ts component $\{\Sigma\}_{ts} = (C, (\psi_t, \psi_s))$ where the functional C is given by (13). Denote by $\hat{\Sigma}_n$ the covariance matrix of $\overrightarrow{\eta}_{hni}$. By the convergence results for T_{ij}, $\hat{\Sigma}_n \to_p \Sigma$. Since the functions $\psi_1, ..., \psi_m$ are linearly independent the matrix Σ and thus $\hat{\Sigma}$ (in probability for large enough n) is invertible. Define ξ_{hni} to equal $\hat{\Sigma}^{-1/2} \overrightarrow{\eta}_{hni}$, then $\hat{\Sigma}^{-1/2} \overrightarrow{\eta}_{hni} - \Sigma^{-1/2} \overrightarrow{\eta}_{hni} \to_p 0$.

Next, consider an $m \times 1$ vector λ with $\lambda' \lambda = 1$. The random variables $\lambda' \xi_{hni}$ are independent with expectation 0, $var \sum \lambda' \xi_{hni} = 1$; they satisfy the Liapunov condition: $\sum E \left|\lambda' \xi_{hni}\right|^{2+\delta} \to 0$ for $\delta > 0$ since the kernel function is bounded with finite support. Thus

$$\sum \lambda' \xi_{hni} \to_d N(0, 1)$$

and by Cramer-Wold theorem convergence to a limit Gaussian process for $\hat{\Sigma}_n^{-1/2}\overrightarrow{\eta}_{hn}$ and thus for $\Sigma^{-1/2}\overrightarrow{\eta}_{hn}$ follows.■

Proof of Theorem 5.

Consider

$$\begin{aligned}\hat{m}(x) &= \hat{m}(x)I(\left|\hat{f}(x)\right| > b) + \hat{m}(x)I(\left|\hat{f}(x)\right| \leq b) = \\ &= \hat{m}_1(x) + \hat{m}_2(x).\end{aligned}$$

By B.5 the limit process for $\hat{m}(x)$ coincides with that for $\hat{m}_1(x)$. For the expectation functional write

$$\begin{aligned}E(\hat{m}(x)\hat{f}(x) - m(x)\hat{f}(x)) &= \frac{1}{h^k}E[E_{x_i}(y_i K(\frac{x_i - x}{h}) - m(x)K(\frac{x_i - x}{h})] \\ &= \frac{1}{h^k}E[m(x_i)K(\frac{x_i - x}{h}) - m(x)K(\frac{x_i - x}{h})].\end{aligned}$$

For any test function $\psi(x) \in D$ the value of the functional

$$\begin{aligned}&\left(E(\hat{m}(x)\hat{f}(x) - m(x)\hat{f}(x)), \psi(x)\right) \\ =& \int\int (m(x + hw) - m(x))K(w)f(x + hw)\psi(x)dwdx \\ =& \int\int (m(x) - m(x - hw))K(w)f(x)\psi(x - hw)dwdx \\ =& \int\int \sum \frac{\partial^{l+1}m(x + h\tilde{w})}{\partial x_1^{l_1}...\partial x_k^{l_k}} h^l w^l K(w)f(x)\psi(x - hw)dwdx \\ =& h^l \int \sum \frac{\partial^{l+1}m(x)}{\partial x_1^{l_1}...\partial x_k^{l_k}}\psi(x)f(x)dx \int w^l K(w)dw + o(h^l)\end{aligned}$$

by continuity of $\psi(x)$ and B.4. Note that since expectation of $\frac{\partial^{l+1}m(x)}{\partial x_1^{l_1}...\partial x_k^{l_k}}$ exists in an ordinary sense by B.4 the $O(h^l)$ term is an ordinary function. Since $\hat{m}_1(x)\hat{f}(x) - m_1(x)\hat{f}(x)$ has the same limit process as $\hat{m}(x)\hat{f}(x) - m(x)\hat{f}(x)$ the expectation is also $O(h^l)$ and also exists as an ordinary function. The same expansion can be applied to $E\left|\hat{m}_1(x)\hat{f}(x) - m_1(x)\hat{f}(x)\right|$. Since by B.5

$$E\left|\hat{m}_1(x)\hat{f}(x) - m_1(x)\hat{f}(x)\right| b^{-1} > E\left|\hat{m}_1(x) - m_1(x)\right|$$

we get that the expectation of $\hat{m}_1(x) - m_1(x)$ and thus of $\hat{m}(x) - m(x)$ is also $O(h^l)$. By assumption B.9 (a) follows.

Proof for (b) is similar.

■

Appendix B.

I. Generalized functions: Definitions and examples.

Here we summarize some of the definitions and results from Gel'fand and Shilov (1964, v.1 and v.2) and Ch. 1 of Sobolev (1992).

B.I.1. Spaces of test functions. The space K (v.1,1.2). Consider all infinitely differentiable real functions on R^k with finite support; this is a linear space. Convergence is defined for a sequence $\psi_1, ...\psi_n, ...$ if all ψ_n are zero outside a bounded interval and on it converge (uniformly) as well as each of the derivatives. K is a non-metrizable topological space.

The space $K(a) \subset K$ consists of $\psi \in K$ such that $\psi(x) = 0$ for $\|x\| > a$.

The space S (v.1,1.10) is defined as that of all infinitely differentiable real functions on R^k that go to zero at infinity faster than any power; this is a linear space and topology can be defined similarly.

Spaces D_m, $m = 0, 1, ...$ are those of functions with finite support with m continuous derivatives. On R^k the space D_m contains all ψ with continuous derivatives $\frac{\partial^l \psi}{\partial^{l_1} x_1 ... \partial^{l_k} x_k}$ with $l_1 + ... + l_k \leq m$. Any function in D can be approximated by a product of univariate functions.

Properties (from embedding diagrams, Sobolev, p. 56; notation: our D_m is $\overset{o}{C}{}^{(m)}$ there):

(i) $K \subset D_m$ and $D_m \subset D_{m'}$ for any m, where $m' < m$; also $K \subset S$;

(ii) each of the subspaces is dense in the larger space in the topology of that space.

We denote a generic space of test functions by D. The space $D(a) \subset D$ consists of $\psi \in D$ such that $\psi(x) = 0$ for $\|x\| > a$.

B.I.2. Generalized functions.

2.a. Ordinary functions (locally summable). A real function defined on R^k and Lebesgue-integrable on any bounded set (locally summable) is an ordinary function (v.1, 1.3). Each ordinary function f defines a functional on D: for $\psi \in D$ the value of the functional (f, ψ) can be defined by

$$(f, \psi) = \int_{-\infty}^{+\infty} f(x)\psi(x)dx. \qquad (16)$$

It is easy to see that this is a linear continuous functional. Note that if two ordinary functions differ they define different functionals (v.2, 1.5).

2.b. Generalized functions. Denote the space of linear continuous functionals on D by D'. For any D the linear continuous functionals form a linear space D' with the weak topology: a sequence of functionals in D', g_n, converges to g if for any $\psi \in D$ $(g_n, \psi) \to (g, \psi)$. The space D' is complete in the weak topology. The subspace of functionals given by (16) is dense in D' (v.1, 1.5 for the space K).

Define a generalized function as a functional from the space D'. If it is given by (16) it is a regular functional (function); if it cannot be represented in the form (16) it is a singular functional (function) (v.1, 1.3). Usually the same notation (16) is used for any functional even though for singular functionals it does not have the ordinary interpretation. An example of a singular function is the $\delta-$function defined by $(\delta, \psi) = \psi(0)$. If a generalized function can be represented as

$$(f, \psi) = \sum_{k=0}^{p} \int_{-\infty}^{\infty} f_k(x)\psi^{(k)}(x)dx \qquad (17)$$

where $f_0, ...f_k$ are ordinary functions, it is said that f has an order of singularity $\leq p$. E.g. an ordinary function has order of singularity zero, the δ-function has order of singularity ≤ 1, and since it is not an ordinary function (with order of singularity zero), its order of singularity is exactly 1.

Support of a generalized function f is defined (v.1,1.4) as the set of its essential points, where an essential point x is such that for any neighbourhood $U(x)$ there is a test function $\psi \in D$ with support contained in $U(x)$ for which $(f, \psi) \neq 0$.

Convergence of generalized functions is defined as weak convergence of functionals: $f_n \Rightarrow f$ iff for any $\psi \in D$ the sequence of values of the functionals converges: $(f_n, \psi) \to (f, \psi)$. Generalized functions form a complete linear space D' (v.1, 1.8)

Properties (Sobolev, 1992, p.59; notation: our D'_m is $\overset{o}{C}^{(m)\#}$):
(i) $D'_m \subset K'$; $D'_{m'} \subset D'_m$ for any m where $m' < m$; also $S' \subset K'$;
(ii) for any of the spaces $D \subset D'$ and is dense there in the weak topology.

2.c Derivatives of generalized functions. For any generalized function f on R its derivative f' is defined as long as $D \subseteq D_1$by (v.1, 2.1)

$$(f', \psi) = (f, -\psi') \tag{18}$$

It is easy to check that this definition provides an ordinary derivative if the functional was differentiable as an ordinary function. Differentiation is a continous operation (v.1, 2.4).

Any continuous functional on D of degree of singularity less than m can be extended to a continuous functional on D_m.

One can similarly consider multivariate generalized functions (examples in v.1, 2.3). For $x = (x_1, ..., x_k) \in R^k$ and $D(R^k)$ a continuous linear functional $F \in D'$ and its value on $\psi \in D(R^k)$, (F, ψ), is similarly defined. If $F(x_1, ..., x_k)$ is an ordinary locally summable function then

$$(F, \psi) = \int .. \int F(x_1, ..., x_k)\psi(x_1, ..., x_k)dx_1...dx_k,$$

which we shall often write if there is no ambiguity as $\int F(x)\psi(x)dx$. For any $\in D(R^k)$ the derivative

$$(\frac{\partial^k F(x)}{\partial x_1 ... \partial x_k}, K) = (-1)^k (F, \frac{\partial^k \psi(x)}{\partial x_1 ... \partial x_k})$$

can be defined as long as $D \subseteq D_k$.

II. Generalized random processes: Definitions and examples.

Here we summarize some definitions and results from Ge'fand and Vilenkin (1964), v.4.

If f is a continuous linear functional on the space D (note that Gelfand and Vilenkin consider only D coinciding with the space K) and additionally (f, ψ) is a random variable for any $\psi \in D$ which implies that for any number l of $\psi_1, ...\psi_l \in D$ the set $(f, \psi_1), ..., (f, \psi_l)$ has a joint probability distribution, then f defines a generalized random function on D. (See pp. 241-243; the notation is different there from ours). Gel'fand and Vilenkin distinguish between a generalized random process defined by f when the functions in D are univariate and a generalized random field for the case of multivariate functions. We shall not make that distinction and refer to a generalized random process (univariate or multivariate) in all cases.

An expectation or mean functional is defined by (again changing notation, p.246)

$$m(\psi) = (E(f), \psi) = E(f, \psi)$$

if $E(f, \psi)$ defines a continuous linear functional on D.

If for any ψ_1, ψ_2 the expectation $E((f, \psi_1), (f, \psi_2))$ exists then the "correlation" functional of the process is given by

$$B(\psi_1, \psi_2) = E((f, \psi_1), (f, \psi_2)),$$

and the covariance functional by

$$C(\psi_1, \psi_2) = B(\psi_1, \psi_2) - m(\psi_1)m(\psi_2)$$

provided these functionals exist. These functionals are bilinear in ψ_1, ψ_2. The covariance functional is positive definite for non-zero f (to check consider $\psi_1 = \psi_2 \neq 0$) (p. 247). Higher order moments are similarly defined.

A generalized random process is a generalized Gaussian process if for any linearly independent $\psi_1, ..., \psi_l$ from D the joint distribution of $(f, \psi_1), ..., (f, \psi_l)$ is Gaussian. (p.248). A generalized Gaussian process is uniquely determined by its mean functional and "correlation" (or covariance) bilinear functional (Theorem 1, p.250-251).

Generalized Gaussian processes are differentiable; the derivative of a generalized Gaussian process with correlation functional $B(\psi_1, \psi_2)$ is a generalized Gaussian process with correlation functional B' given by $B'(\psi_1, \psi_2) \equiv B(\psi_1', \psi_2')$ (p. 257). On pp. 258-259 the correlation functional for a Wiener process is derived; and on p.260 it is shown that the correlation functional for the generalized derivative of a Wiener process is

$$B'(\psi_1, \psi_2) = \int_0^\infty \psi_1(w)\psi_2(w)dw$$

which implies that the correlation functional is a δ-function:

$$B'(\psi_1, \psi_2) = \int_0^\infty \int_0^\infty \delta(w-t)\psi_1(w)\psi_2(t)dwdt.$$

Recall that the derivative of a Wiener process cannot be an ordinary random process, but here we see that it is well defined as a generalized random process.

Zinde-Walsh and P.C.B.Phillips (2003) similarly derive the derivative of the fractional Wiener process.

References

[1] Ahmad I.A. and P.B. Cerrito (1994) Nonparametric Estimation of joint Discrete-continuous Probability Densities with Applications, Journal of Statistical Planning and Inference 41, 349-364.

[2] Gel'fand I.M. and G.E.Shilov (1964) Generalized Functions, Vol.1, Properties and Operations, Academic Press, San Diego.

[3] Gel'fand I.M. and G.E.Shilov (1964) Generalized Functions, Vol.2, Spaces of Test functions and Generalized Functions, Academic Press, San Diego.

[4] Gel'fand I.M. and N.Ya.Vilenkin (1964) Generalized Functions, Vol.4, Applications of Harmonic Analysis, Academic Press, San Diego.

[5] Halperin I. (1952) Introduction to the Theory of Distributions (based on Lectures by Laurent Schwartz), University of Toronto Press.

[6] Haerdle W. and T. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives, Journal of the American Statistical Association, vol.84, pp. 986-995.

[7] Pagan, A. and A. Ullah (1999) Nonparametric Econometrics, Cambridge University Press.

[8] Phillips P.C.B. (1991) A Shortcut to LAD Estimator Asymptotics, Econometric Theory, 7,450-463

[9] Phillips, P.C.B. (1995) Robust Non-Stationary Regression, Econometric Theory, 11, 912-951.

[10] Powell J.,J.Stock and T.Stoker (1989) Semiparametric estimation of index coefficients, Econometrica, v.57, pp. 1403-1430.

[11] Schwartz L. (1950) Théorie des Distributions, v.1,2.. Hermann, Paris.

[12] Sobolev, S. (1992) Cubature formulas and Modern Analysis, Gordon and Breach Science Publishers, S.A...

[13] Zinde-Walsh, V. (2002) Asymptotic Theory for some High Breakdown Point Estimators. *Econometric Theory* 18, 1172-1196.

[14] Zinde-Walsh, V. and Peter C. B. Phillips (2003) Fractional Brownian motion as a differentiable generalized Gaussian process, Chapter 17, pp. 285-292 in K. Athreya, M. Majumdar, M. Puri and W. Waymire (Eds.) Probability, Statistics and their Applications: Papers in Honor of Rabi Bhattacharya. Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 41, 2003. Beachwood Ohio: Institute of Mathematical Statistics.

[15] Zinde-Walsh, V. and P.C.B.Phillips (2004), Smoothing Extremum Estimators: a Generalized Functions Approach, unpublished manuscript.