Measurement error and convolution in generalized functions spaces

Victoria Zinde-Walsh*

McGill University and CIREQ

victoria.zinde-walsh@mcgill.ca

(514) 398 4834

September 23, 2010

*The support of the Social Sciences and Humanities Research Council of Canada (SSHRC), the Fonds québecois de la recherche sur la société et la culture (FRQSC) is gratefully acknowledged. P.C.B.Phillips provided invaluable encouragement, comments and advice; I also thank J. W. Galbraith for his comments.

Running head: Measurement error and convolution

Victoria Zinde-Walsh

Department of Economics, McGill University

855 Sherbrooke Street West,

Montreal, Quebec, Canada

H3A 2T7

Abstract

This paper considers convolution equations that arise from problems such as deconvolution and non-parametric regression with errors in variables. The equations are examined in spaces of generalized functions to account for possible singularities; identification is proved for a wide class of problems. Conditions for well-posedness in the topology of generalized functions are derived for the deconvolution problem and some regressions; an example shows that even in this weak topology well-posedness may not hold. Well-posedness is crucial for consistency of non-parametric deconvolution and important in cases when a non-parametric model is mis-specified as parametric. Stochastic properties and convergence for generalized random processes are derived for solutions of convolution equations. This paper focuses on independent data. Deconvolution estimation for a generalized function on a bounded support in a fixed design case is examined; the rate for a shrinkage deconvolution estimator useful for a sparse support is derived. A consistent non-parametric

estimator in an errors in variables regression model with the regression function in L1 is constructed.

1 Introduction

Various statistical models lead to equations involving convolutions of unknown functions such as

$$g * f = w, \tag{1}$$

where f could be a known probability density (e.g. Gaussian) and w some known function; one wishes to solve for g. For example, g is the density of a mismeasured variable, x^*, observed with error, u :

$$z = x^* + u; \tag{2}$$

observed z has density w; u is measurement/contamination error independent of x^* with a known density f. If f is not known additional conditions are needed to identify g. Thus if another observation, x, on x^* is available: $x = x^* + u_x$, where u_x is not necessarily independent but $E(u_x|x^*, u) = 0$ one can add more convolution equations, e.g. with x_k denoting the $k-th$ coordinate of $x = (x_1, ..., x_d)$:

$$(x_k^* g) * f = w_{2k},$$

where $w_{2k} = E\left(w(z)x_k|z\right).$ Indeed,

$$E(x_k|z) = E(x_k^*|z) = \int (z_k - u_k)\frac{g(z-u)f(u)}{w(z)}du.$$

Errors in variables regression (EIV, see Chen et al, 2009 for review) models often lead to convolution equations. An example is

$$\begin{aligned} y &= g(x^*) + u_y; & (3)\\ x &= x^* + u_x; & (4)\\ z &= x^* + u, & (5) \end{aligned}$$

where (3),(5) represent a classical EIV model for independent identically distributed (i.i.d.) data and (3)-(5) provide a EIV regression with z representing a second measurement or possibly a given projection onto a set of instruments. Here y, z or x, y, z are observed; u is a Berkson type measurement error independent of z; u_y, u_x have zero conditional (on z and the other errors) expectations. The model (3,5) leads to (1) with $w = E(y|z)$, the unknown regression function, g, and measurement error density $f(u)$. For (3-5) with an unknown measurement error distribution Newey (2001) proposed to consider an additional equation assuming that the conditional moment of the product $E(gx|z)$ exists; that leads to a system of two equations with two unknown functions in the univariate case (as in Schennach, 2007, Zinde-Walsh, 2009); this provides

$$\begin{aligned} g * f &= w_1, & (6)\\ (x^* g) * f &= w_2, \end{aligned}$$

with $w_1 = E(y|z)$, $w_2 = E(xy|z)$ known. In the multivariate case for $w_{2k} =$

$E(x_k y|z)$, $k = 1, ..., d$ we obtain equations

$$
\begin{aligned}
g * f &= w_1 \\
(x_k^* g) * f &= w_{2k}, \; k = 1, ...d. \qquad (7)
\end{aligned}
$$

Another way in which additional equations may arise is if there are observations on derivatives, e.g. $\frac{\partial y}{\partial x_k}$; examples are discussed in Hall and Yatchew (2007). For such a case, if variables are measured with error and instruments can be found, it is natural to assume in addition to $w_1 = E(y|z)$, that $w_2 = E(g_k' x_k|z)$ is known as well, where $g_k' = \frac{\partial g}{\partial x_k}$. This leads to:

$$
\begin{aligned}
g * f &= w_1, \qquad (8) \\
x_k^* g_k' * f &= w_{2k}.
\end{aligned}
$$

If there are additional controls, in particular discrete ones, on which one could condition, further systems of convolution equations could arise. For example, suppose conditioning on a binary covariate; with the same distribution for measurement or contamination error one could have more equations:

$$
\begin{aligned}
g_1 * f &= w_1, \qquad (9) \\
g_2 * f &= w_2.
\end{aligned}
$$

A common way of providing solutions is to consider these equations in some function spaces, such as L_2 and its sub-spaces, use Fourier transform to obtain corresponding algebraic equations (since Fourier transform of a convolution in L_2 provides a product of Fourier transforms) and solve those.

Working in function spaces limits the classes of models for which solutions can be obtained, for example, the distribution of the variables in (2) may not possess a density. In astrophysics and mass spectroscopy the object of interest may be represented by "sum of peaks" function, such as a sum of delta functions (Klann et al, 2007). If g represents a regression function, restricting it to have a usual Fourier transform means that important cases of linear and polynomial regressions as well as binary choice would be excluded. A natural extension is to consider spaces of generalized functions, where these restrictions can be overcome.

Thus in Cunha et al (2010) the possibility that the mismeasured variable may not have a density is mentioned in Theorem 1 which proposes that it be interpreted as a generalized function ("distribution"), although the proof there does not extend to that case; a complete proof is in Section 3 here. Klann et al (2007) worked with Besov space of generalized functions, $B^s_{p,p}$ with negative s; Zinde-Walsh (2008) utilized generalized functions spaces denoted K' or D' (Gelfand, Shilov, 1964 or Schwartz, 1966) to represent derivatives of distribution functions; Shennach (2007) made use of the space S' of tempered distributions (see the same books) for the problem in errors in variables regression model, and Zinde-Walsh (2009) provided extensions and corrections for that case.

Using generalized functions easily allows one to consider in a unified framework, a multivariate set-up where only some of the variables are measured with error, or the variables are only partially contaminated (e.g. in a survey where some proportion of the responses is truthful). Indeed, the multivariate density in the convolution can be represented as a generalized function, and if there is no measurement error the corresponding generalized density is just the $\delta-$function and the convolution $g * \delta = g$. Existence of derivatives and moments in generalized functions spaces often holds without extra assumptions.

This paper examines convolution equations in classical spaces of generalized functions, D' and S', and related spaces. Section 2 gives conditions when convolutions are defined for generalized functions of interest, and when products of Fourier transforms are defined. Then the convolution theorem ("exchange formula") permits to transform convolutions into products of Fourier transforms. Section 3 provides results on existence and uniqueness of solutions of the transformed equation or system of equations; these results provide identification of the econometric and statistical models. Well-posedness of the problem means that the solution continuously depends on the known functions; it is crucial for establishing consistency of nonparametric estimation and for justifying use of parametric models in place of less tractable general ones. General results on well-posedness of the solutions in the models considered are presented here in Section 3 for the first time in this literature (some were also given in the working paper Zinde-Walsh, 2009).

Section 4 provides stochastic properties of generalized function spaces and stochastic convergence of solutions to convolution equations in some classical spaces of generalized random functions. In Section 5 more specific results are provided. In the first part of that section the limit generalized Gaussian process is derived for the solution to the deconvolution problem on a bounded support in the fixed design case; then rates are provided for the deconvolution shrinkage estimator that is useful when the support is sparse (e.g. the function is a finite sum of delta-functions). The second part of Section 5 examines nonparametric estimation of a regression function in L_1 for an errors in variables regression; the estimator is shown to be consistent in the topology of generalized random functions.

2 Convolution equations in generalized functions

2.1 Generalized functions spaces

Many different spaces of generalized functions can be defined; each may be best suited to some particular class of problems. This paper focuses on well known classical spaces of generalized functions, D' and S', discussed in the books by Schwartz (1960) (Sz) and Gel'fand and Shilov (1964)(GS); reference is also made to related spaces as presented e.g. by Sobolev (1992) (Sob).

Consider a space of test functions, G. Two widely used spaces are $G =$

D and $G = S$. The space D is the linear topological space of infinitely differentiable functions with finite support: $D \subset C_\infty(R^d)$, where $C_\infty(R^d)$ is the space of all infinitely differentiable functions; convergence is defined for a sequence of functions that are zero outside a common bounded set and converge uniformly together with derivatives of all orders. For any vector of non-negative integers $m = (m_1, ...m_d)$ and vector $t \in R^d$ denote by t^m the product $t_1^{m_1}...t_d^{m_d}$ and by ∂^m the differentiation operator $\frac{\partial^{m_1}}{\partial x_1^{m_1}}...\frac{\partial^{m_d}}{\partial x_d^{m_d}}$. The space $S \subset C_\infty(R^d)$ of test functions is defined as:

$$S = \left\{ s \in C_\infty(R^d) : \left| t \right|^l \left| \partial^k s(t) \right| = o(1) \text{ as } t \to \infty \right\},$$

for any k, l, where $k = (0, ...0)$ corresponds to the function itself, $|t|$ is the vector of absolute values of vector t, $t \to \infty$ coordinatewise; thus the functions in S go to zero faster than any power as do their derivatives. A sequence in S converges if in every bounded region each $\left| t \right|^l \left| \partial^k s(t) \right|$ converges uniformly. The generalized functions space G' is the space of linear continuous functionals on G with the corresponding weak convergence (see, e.g. GS, v.1).

Sobolev gives a general definition (Sob, 1.8) where he points out a subtle distinction between the functional and a generalized function. Any generalized function, $b \in G'$, can be defined by an equivalence class $\{b_n\}$ of weakly converging sequences of test functions $b_n \in G$:

$$b = \left\{ \{b_n\} : b_n \in G, \text{ such that for any } s \in G, \lim_{n\to\infty} \int b_n(t)\overline{s(t)}dt = (b,s) < \infty \right\},$$

where $\int \cdot dt$ denotes the multivariate integral over R^d, over-bar indicates complex conjugate for complex-valued functions and (b,s) provides the value of the functional $b \in G'$ for $s \in G$. However, the same functional can be represented by different generalized functions corresponding to different spaces G. For example, consider the $\delta-$function. This is a linear continuous functional on the space $C^{(0)}$ of continuous functions as well as on D or S and provides $(\delta, s) = s(0)$; on $C^{(0)}$ it can be represented as an equivalence class of $\delta-$convergent sequences of continuous functions. This implies that a generalized function considered as a functional can sometimes be extended to a linear continuous functional on a wider space.

Note that $D \subset S$ and thus $S' \subset D'$ as a linear topological subspace, however, a sequence of elements of S' that converges in D' may not converge in S'. In the terminology of (Sz) generalized functions are sometimes called "distributions" and elements of S' "tempered distributions"; here we shall call them generalized functions indicating the specific space considered. In (Sob, p.59) a diagram shows various chains of generalized functions spaces embedded in each other; these are spaces of functionals on spaces of continuously differentiable (of different orders) functions, continuously differentiable functions with finite support and Sobolev spaces.

Any locally summable (integrable on any bounded set) function $b(t)$ defines a generalized function b in D' by

$$(b, s) = \int b(t)\overline{s(t)}dt; \tag{10}$$

any locally summable function $b(t)$ that additionally satisfies

$$\int \left((1+t^2)^{-1}\right)^m |b(t)|\, dt < \infty \tag{11}$$

for some non-negative integer-valued vector $m = (m_1, ..., m_d)$ with $\left((1+t^2)^{-1}\right)^m$ denoting the corresponding product $\Pi_{i=1}^d \left((1+t_i^2)^{-1}\right)^{m_i}$similarly by (10) defines a generalized function b in S'; such functions are called regular.

Generalized derivatives are defined for all generalized function in D' and S' by $(\partial^m b, s) = (-1)^m (b, \partial^m s)$.

The Fourier transform (Ft) is defined for functions in D, and more generally in S, and is an isomorphism of $S : Ft\psi(\xi) = \int \psi(x)e^{ix'\xi}dx$ (x' denotes transpose); in the spaces D' and S' the Fourier transform is given by $(Ft(b), s) = (b, Ft(s))$. The Fourier transform is an isomorphism of the space S'.

2.2 Existence of convolutions; convolution pairs

The convolution of generalized functions can be defined in different ways (see, e.g. Sz, p.154 or Sob.,p. 63; GS, v. I, p.103-104); it does not always have meaning and exists for specific pairs of mutual convolutors.

Consider the following spaces of test functions and of generalized functions on R^d : $D, S, C_\infty, \mathcal{O}_M, D', S', E', E'_+, \mathcal{O}'_C, \mathcal{O}_M$, where $\mathcal{O}_M \subset C_\infty$ is the subspace of functions with every derivative growing no faster than a polynomial, E' is the subspace of generalized functions with compact support, and $\mathcal{O}'_C$ is the subspace of rapidly decreasing (faster than any polynomial) generalized functions (Sz, p.244). Table 1 shows pairs of spaces for elements of which convolution is defined (X indicates that convolution cannot be defined for some pairs of elements of the spaces); the table entries indicate to which space the element resulting from the convolution operation belongs. The table is an extended version of the one in the textbook by Kirillov and Gvishiani (1982, p.102) and summarizes the well-established results in the literature (see, e.g., Sz).

The convolution pairs in the table where convolution is defined all possess the hypocontinuity property (Sz, p.167, p.247-257). Hypocontinuity of a bilinear operation means that if one component of a pair is in a bounded set in G' and the other converges to zero in G', the result of the bilinear operation converges to zero (Sz, pp.72,73).

Although the Table implies that convolution of generalized functions, say, both from S', cannot be always defined and there are examples of even regular generalized functions in S' such that their convolution does not provide an element in S' (see, e.g. Kaminski, 1991), nevertheless there are pairs of generalized functions spaces beyond those in the Table for which convolution

Table 1: The convolution table

$g \setminus f$	D	Sz	E'	O'_C	S'	D'
D	D	S	D	S	O_M	C_∞
S	S	S	S	S	O_M	X
E'	D	S	E'	O'_C	S'	D'
O'_C	S	S	O'_C	O'_C	S'	X
S'	O_M	O_M	S'	S'	X	X
D'	C_∞	X	D'	X	X	X

defines a generalized function; spaces where convolution is defined can be combined. The convolution is a bilinear operation (Sz, p.157); convolution of a tensor product of generalized functions on two vector spaces, R^{d_1}, R^{d_2} is the tensor product of the convolutions of functions in each space (Sz, p.158). Moreover convolution of any number of generalized functions can be defined in D' as long as all except possibly one have compact supports and this operation is associative and commutative (Sz, p.158); a variable shift or derivative of a convolution exists and is obtained by a shift or differentiation of any of the generalized functions entering the convolution (Sz, p.160).

Definition 1. *Call a pair of subspaces of generalized functions, $A \subset G'$ and $B \subset G'$ a* convolution pair (A, B) *if for any $a \in A, b \in B$ convolution $a * b$ is defined in G'; is a hypocontinuous operation in topology of G' for (A, B) and if $G \in A$ and $G \in B$.*

All the pairs of spaces in the Table satisfy this definition.

For a pair of spaces of generalized functions (in D') that have support bounded on the same side convolution is defined (Sz, p.177); the result is in D' but such a pair does not satisfy the definition because D does not belong to the subspaces. The results on identification could be extended to this case, but at the cost of working out a somewhat specialized approach; this case is not considered here.

Assumption 1. *The statistical model defines two functions, g and f in G' such that $g \in A \subset G'$ and $f \in B \subset G'$; the subspaces (A, B) form a convolution pair.*

This assumption implies that (1) holds; it is often satisfied in statistical problems. Convolution of generalized density functions exists, thus (2) leads to (1) even when the density functions do not exist in the ordinary sense. The finite sum of $\delta-$functions considered by Klann et al (2007) is in E', thus convolution with any element of D' (or S') exists in $D'(S')$. Schennach (2007) considered univariate errors in variables models with instrumental variables and a regression function, g, bounded by polynomials whereas Zinde-Walsh (2009) required more generally that g satisfy (11); these regression functions are in S', convolution with any generalized density functions from $\mathcal{O}'_C$ exists in S'. For regression functions g in subspaces of S' where growth is more specifically restricted, convolution with less rapidly declining f may exist.

In many cases when the convolution $g * f$ is defined, the convolution $x_k g(x) * f$, where x_k is a coordinate of the $x \in R^d$ vector, is also defined. Indeed, this is so in all the examples above: if g has bounded support, so does any $x_k g$; if $g \in C_\infty$, it is true of $x_k g$ as well, etc. Thus some models accommodate not only (1), but also other equations, e.g. providing (7).

Also, it is often the case that equations (8) hold; in fact (7) implies (8).

To demonstrate this, denote the right-hand side of the second equation in (8) by $\bar{w}_{2k}$ and show that $\bar{w}_{2k}$ can be obtained from w_1 and w_{2k} of (7). Since generalized functions are differentiable and the derivative of the convolution is defined, then if w_{2k} is defined as a generalized function it possesses all derivatives and $(w_{2k})'_k = (x_k g)'_k * f = w_1 + \bar{w}_{2k}$, leading to $\bar{w}_{2k} = (w_{2k})'_k - w_1$.

Assumption 2. *(a) The statistical model is such that in addition to Assumption 1,* $x_k g \in A$, $k = 1, ..., d$.

(b) The statistical model is such that in addition to Assumption 1, $x_k g'_k \in A$, $k = 1, ..., d$.

If Assumption 1 and 2(a) hold then equations (7) define generalized functions w_1, w_{2k} for the generalized functions g and f in S'(or D'); similarly for Assumption 1 and 2(b), (8) holds in S'(or D').

2.3 Products of Fourier transforms and transformed convolution equations

For the generalized functions in equations (7) and (8) denote Fourier transforms as $\gamma = Ft(g)$, $\phi = Ft(f)$ and $\varepsilon_i = Ft(w_i)$.

The classical case of the convolution pair $(S', \mathcal{O}'_C)$ is examined in (Sz). If (g, f) belong to $(S', \mathcal{O}'_C)$ the convolution equation transforms into the equation $\gamma\phi = \varepsilon$, where $\gamma \in S'$, $\phi \in \mathcal{O}_M$ and $\varepsilon \in S'$ (Sz, p.281-282). The product between a generalized function in S' and a function from $\mathcal{O}_M$ always exists in S'; multiplication is a hypocontinuous operation (Sz, p.243-246).

In a statistical problem ϕ is often the characteristic function of a measurement or contamination error; the condition $\phi \in \mathcal{O}_M$ would require existence of all moments. It may be of interest to consider pairs where products of Fourier transforms of generalized functions exist for less smooth functions (e.g. with relaxed moments requirements on measurement error). To provide a more general treatment, less restrictive multiplication pairs are introduced here.

Define the space of test functions, denoted $G \oplus \phi G$, that consists of functions that can be represented as $\psi_1 + \phi\psi_2$, with $\psi_i \in G$ ($G = D$ or S) and $\phi \in C^{(0)}$. With ϕ fixed, topology is defined by convergence in G. Denote by $(G \oplus \phi G)'$ the space of linear continuous functionals on $G \oplus \phi G$; consider $\gamma \in G'$ that can be extended to a continuous linear functional on $G \oplus \phi G$, denote the linear space of such γ by $G(\phi)' \subset G'$. If $\phi \in G$ then $G(\phi)' = G'$, leading to the classical case, so that case is included. For any $\psi \in G$ the value $(\gamma, \phi\psi)$ is defined and is a continuous functional with respect to ψ; this defines the generalized function $\phi\gamma$: $(\phi\gamma, \psi) = (\gamma, \phi\psi)$. For example, the derivative of a Dirac $\delta-$function in G' can be multiplied by any continuous function, ϕ, that is differentiable at 0, since then $(\delta'\phi, \psi) = (\delta', \phi\psi)$ with $(\delta', \phi\psi) = \phi'(0)\psi(0) + \phi(0)\psi'(0)$.

Denote by 0_n any sequence in the equivalence class of sequences for the zero functional $0 \in G'$; call 0_n a zero-convergent sequence.

Remark 1. *(a) $G \subset G(\phi)'$.*

(b) If $G' = D'$ then the generalized function identically equal to zero,

$0 \in D'$, belongs to $D(\phi)'$ and any zero-convergent sequence, 0_n, converges in $D(\phi)'$ to zero. Indeed, any sequence 0_n converges to zero uniformly on any bounded set, and so $\phi 0_n$ also converges to zero uniformly on any bounded set. Any $\psi \in D$ has bounded support and $(\phi 0_n, \psi) \to 0$.

(c) If $G' = S'$, and ϕ satisfies (11) then the generalized function identically equal to zero, $0 \in S'$, belongs to $S(\phi)'$ and any zero-convergent sequence, 0_n, in the equivalence class for $0 \in S'$ converges in $S(\phi)'$ to zero. By (11) for some l the function $(1+x^2)^{-l}\phi$ is absolutely integrable. Denote the generalized function that is its integral by ϕ_I; this continuous function is bounded. For any 0_n the sequence $0_n(1+x^2)^l$ is also zero convergent in S'; the product $\phi_I 0_n \to 0$ in S' by hypocontinuity; the sequence of derivatives of any sequence 0_n is also zero convergent and thus $\phi 0_n \to 0$ in S'.

Theorem 1. *If Assumption 1 is satisfied, and (a) if also $\phi \in C^{(0)}$, then $\gamma \in G(\phi)'$ for $G = D$; $\gamma\phi = \varepsilon$; and (b) if further $\phi \in C^{(0)}$ satisfies (11), then $\gamma \in G(\phi)'$ for $G = S$; $\gamma\phi = \varepsilon$.*

Proof. If $g * f \in G'$, then also the convolution exists in the sense of Hirata and Ogata (1958) (see also Kaminski, 1982). Consider the special sequences studied by Mikusinski (Antosik et al, 1973) and Hirata and Ogata (1958) that are defined for a generalized function b as $\tilde{b}_n(x) = b * \delta_n(x)$ with δ_n representing the following delta-convergent sequence: for $\alpha_n > 0$ and $\alpha_n \to 0$ the function $\delta_n(x)$ is non-negative with support in $|x| < \alpha_n$ and $\int \delta_n(x)dx = 1$; the convolution with $\delta_n \in E'$ is always defined. By Assumption 1 convolution

$\tilde{g}_n * \tilde{f}_n$ is defined and by hypocontinuity

$$g * f = \lim \tilde{g}_n * \lim \tilde{f}_n.$$

The exchange formula of Hirata and Ogata applies:

$\lim Ft(\tilde{g}_n * \tilde{f}_n) = \lim Ft(\tilde{g}_n) \lim Ft(\tilde{f}_n)$.

Since $\phi_n = Ft(\tilde{f}_n)$ is continuous and converges in G' to the continuous function ϕ, the convergence is uniform on bounded sets and thus also for $G = D$ in $G \oplus \phi G \subset C^{(0)}$. Then $\gamma_n = Ft(\tilde{g}_n) \in G(\phi)'$. Next, show that $\gamma \in G(\phi)'$. By continuity of the Fourier transform γ_n is a sequence from the equivalence class of γ. We show that any sequence from the equivalence class of γ is in $G(\phi)'$; therefore γ is in $G(\phi)'$ also. Indeed, any sequence from the equivalence class of γ differs from γ_n by a zero-convergent in G' sequence 0_n. But by Remark (b) or (c) such a sequence is in $G(\phi)'$ under conditions (a) or (b), correspondingly. ■

Under the conditions of Theorem 1 and the appropriate part of Assumption 2 the convolution equations (1,7,8) lead to corresponding equations for Fourier transforms:

$$\gamma \cdot \phi = \varepsilon \tag{12}$$

or the system

$$\begin{aligned} \gamma \cdot \phi &= \varepsilon_1 \\ \gamma'_k \cdot \phi &= i\varepsilon_{2k},\ k = 1, ..., d. \end{aligned} \tag{13}$$

or

$$\begin{aligned} \gamma \cdot \phi &= \varepsilon_1 \\ \zeta_k \gamma_k' \cdot \phi &= -i\varepsilon_{2k},\ k = 1, ..., d. \end{aligned} \tag{14}$$

3 Solving the convolution equations

3.1 Identification

The convolution equation (1) uniquely identifies g for a known f if it can be shown that the corresponding equation (12) has meaning and can be uniquely solved for γ.

In the deconvolution problem f, or equivalently ϕ, is assumed to be given.

Theorem 2. *Under the conditions of Theorem 1 assume that ϕ is a known function and supp(ϕ) $\supset$supp(γ); then g is uniquely defined.*

Proof. By Theorem 1, (12) holds in D'.

Define the space $\tilde{D}' = D'(\text{supp}(\gamma))$, which is D' restricted to generalized functions whose support belongs to supp(γ). If supp(γ) $= R^d$, then $\tilde{D}' = D'$.

We show that the product of ε (which equals $\gamma\phi$) with the continuous non-zero function ϕ^{-1} in the space $\tilde{D}'$ exists and equals γ. Consider a sequence $(\gamma\phi)_n$ defined as follows: select some sequence $\tilde{\gamma}_n$ for γ from $\tilde{D} \subset D$. Then each $\tilde{\gamma}_n$ has bounded support. For a sequence of numbers $v_n \to 0$ select $\tilde{\phi}_n$ in $\tilde{D}$ such that $\left|\tilde{\phi}_n - \phi\right| < \frac{v_n}{\sup\left|\tilde{\gamma}_n\phi^{-1}\right|}$ on compact support of γ_n. Then for the sequence $(\gamma\phi)_n = \tilde{\gamma}_n\tilde{\phi}_n$ (this sequence is in the equivalence class of ε) and any $\psi \in \tilde{D}$

$$\begin{aligned} & \int \tilde{\gamma}_n(t)\tilde{\phi}_n(t)\phi^{-1}(t)\overline{\psi(t)}dt \\ = & \int \tilde{\gamma}_n(t)\overline{\psi(t)}dt + \int \tilde{\gamma}_n(t)(\tilde{\phi}_n(t) - \phi(t))\phi^{-1}(t)\overline{\psi(t)}dt \\ \to & \ (\gamma, \psi). \end{aligned}$$

Now we check that for any sequence ε_n from the equivalence class of ε the sequence $\varepsilon_n\phi^{-1}$ converges to γ in $\tilde{D}'$. Indeed, this is so since $0_n\phi^{-1}$ converges to zero in $\tilde{D}'$. Thus $\varepsilon \in \tilde{D}'(\phi^{-1})$ and γ is defined in D'.

Consider now $S' \subset D'$; via the multiplication $\varepsilon\phi^{-1}$ in $\tilde{D}'$ we obtain the function $\gamma \in D'$. Since γ is the Fourier transform of $g \in S'$, γ also defines an element in S' uniquely and an inverse Fourier transform exists in S' for γ. It is then possible to recover g by the inverse Fourier transform $g = Ft^{-1}(\gamma)$.

■

The next Theorem examines the case when there are two unknown functions.

Theorem 3. *(a) If Assumptions 1 and 2(a) hold in S', supp(ϕ) $\supseteq$supp(γ) $=$ W, where W is a convex set in R^d that includes 0 as an interior point, ϕ is continuously differentiable in W with $\phi(0) = 1$, then g is uniquely defined; if supp(ϕ) $=$supp(γ), then ϕ is also uniquely defined; (b) the same conclusion as in (a) obtains when the condition of (a) holds with Assumption 2(b) replacing 2(a).*

Proof. (a). The proof makes use of different spaces of generalized functions and exploits relations between them. In view of the result in Theorem 1 it is sufficient to show that ϕ is uniquely determined on supp(γ).

Consider the space of generalized functions $\tilde{D}'$ (defined in proof of Theorem 1). Since ϕ is non-zero on supp(γ) and continuously differentiable, then by differentiating the first equation in (13), substituting from the second equation and multiplying by ϕ^{-1} in $\tilde{D}'$ (where the product exists as shown in Theorem 1) we get that the generalized function

$$\varepsilon_1 \phi^{-1} \phi'_k - ((\varepsilon_1)'_k - i\varepsilon_{2k})$$

equals zero in the sense of generalized functions, in $\tilde{D}'$. Note that by assumption ε_1 cannot be zero on supp(γ) and both ε_1 and ε_{2k} are zero outside of supp(γ). Define $\varkappa_k = \phi'_k \phi^{-1}$; then $\varkappa_k$ is continuous on supp(γ) and is a regular function in $\tilde{D}'$ that satisfies the equation

$$\varepsilon_1 \varkappa_k - ((\varepsilon_1)'_k - i\varepsilon_{2k}) = 0. \tag{15}$$

We can show that the function $\varkappa_k$ is uniquely determined in the class of continuous functions on supp(γ) by the equation (15). Proof is by contradiction. Suppose that there are two distinct continuous functions on supp(γ), $\varkappa_{k1} \neq \varkappa_{k2}$ that satisfy (15). Then $\varkappa_{k1}(\bar{x}) \neq \varkappa_{k2}(\bar{x})$ for some $\bar{x} \in supp(\gamma)$. Without loss of generality assume that $\bar{x}$ is in the interior of W; by continuity $\varkappa_{k1} \neq \varkappa_{k2}$ everywhere for some closed convex $U \subset W$. Consider now $D(U)'$;

we can write

$$(\varepsilon_1(\varkappa_{k1} - \varkappa_{k2}), \psi) = 0$$

for any $\psi \in D(U)$. A generalized function that is zero for all $\psi \in D(U)$ coincides with the ordinary zero function on U and is also zero for all $\psi \in D_0(U)$, where $D_0(U)$ denotes the space of continuous test functions on U. For the space of test functions $D_0(U)$ multiplication by continuous $(\varkappa_{k1} - \varkappa_{k2}) \neq 0$ is an isomorphism. Then we can write

$$0 = ([\varepsilon_1(\varkappa_{k1} - \varkappa_{k2})], \psi) = (\varepsilon_1, (\varkappa_{k1} - \varkappa_{k2})\psi)$$

implying that ε_1 is defined and is a zero generalized function in $D_0(U)'$. If that were so ε_1 would be a zero generalized function in $D(U)'$ since $D(U) \subset D_0(U)$ but this is not possible since $\varepsilon_1 = \gamma\phi$, which is non-zero by assumption.

Next we show that ϕ is then uniquely determined on supp(γ). Indeed, for any $\zeta \in$supp(γ) with the kth coordinate denoted ζ_k, write the continuous function

$$\tilde{\phi}(\zeta) = \exp \int_0^{\zeta} \Sigma_{k=1}^{d} \varkappa_k(\xi) d\xi_k,$$

where integration is along any arc joining 0 and ζ. This is the unique solution to $\tilde{\phi}(0) = 1$, $\tilde{\phi}_k^{-1} \tilde{\phi}' = \varkappa_k$ (see, e.g., Sz, p.61); then since $\varkappa_k (= \phi_k' \phi^{-1})$ is uniquely determined on supp(γ), so is ϕ on supp(γ) where it coincides with $\tilde{\phi}$.

By Theorem 2 g is then uniquely defined.

(b) Multiplication by ζ_k is defined in D'. Similarly to the proof in (a) for $\varkappa_k = \phi'_k \phi^{-1}$ with the equation

$$\zeta_k \varepsilon_1 \varkappa_k - (\zeta_k (\varepsilon_1)'_k + i\varepsilon_{2k}) = 0$$

replacing (15), the same reasoning as in (a) establishes uniqueness of $\varkappa_k$ for every ζ with $\zeta_k \neq 0$, but by continuity uniqueness holds everywhere. ■

Corollary. *If under Assumptions 1 and 2(a) and the support assumption of Theorem 3(a)* γ *is continuously differentiable in* W *with* $\gamma(0) = 1$, *then* g *is uniquely defined by* $Ft^{-1}(\gamma)$, *where*

$$\gamma(\zeta) = \exp \int_0^{\zeta} \Sigma_{k=1}^{d} \varkappa_k(\xi) d\xi_k,$$

with unique continuous $\varkappa_k(\xi)$ *solves*

$$\varkappa_k(\xi)\varepsilon_1 - i\varepsilon_{2k} = 0$$

in generalized functions.

Proof. The proof is analogous to the proof of the theorem after replacing the function $\varepsilon_1 \phi^{-1} \phi'_k - ((\varepsilon_1)'_k - i\varepsilon_{2k})$ of the theorem by the generalized function

$$\varepsilon_1 \gamma^{-1} \gamma'_k - i\varepsilon_{2k}.$$

■

This corollary provides the proof for Theorem 1 of Cunha et al. (2010).

3.2 Well-posedness of the deconvolution in S'

Well-posedness requires that a unique solution to the problem exist and that this solution be continuous in some "reasonable topology" (Hadamard, 1923). Here most of the results consider the topology of generalized functions which is weaker than, say, the uniform or L_1 norm for corresponding subspaces. Under some conditions it is possible to indicate when continuity in this topology holds; some such conditions are presented here for the deconvolution problem. An example that illustrates that well-posedness does not always obtain even in this weak topology is also given.

Theorem 4. *Assume that for g, f as well as for each g_n, f , $n = 1, 2, ...,$ Assumption 1 holds with the same (A, B), f is a known function such that $\phi = Ft(f)$ satisfies the condition of Theorem 2 and that the sequence $\varepsilon_n = Ft(w_n)$ for $\varepsilon_n = g_n * f$ is such that $\varepsilon_n - \varepsilon \to 0$ in S'. Then if ϕ^{-1} satisfies (11) $\varepsilon_n \phi^{-1}$ exists in S' and $Ft^{-1}(\varepsilon_n \phi^{-1}) \to g$ in S'.*

Proof. By Theorem 2 $\gamma_n = \varepsilon_n \phi^{-1}$ is uniquely defined in S' for every n. Consider $\gamma_n - \gamma = (\varepsilon_n - \varepsilon)\phi^{-1}$ in S'.

Since ϕ^{-1} satisfies (11) it is a regular generalized function in S'. For every $\varepsilon_n - \varepsilon$ consider equivalence classes $\left\{(\varepsilon_n - \varepsilon)_{n_1}\right\}$. From convergence to zero in S' it follows that

$$\lim_{n\to\infty}\lim_{n_1\to\infty}(\varepsilon_n-\varepsilon)_{n_1}=0$$

for any sequence $n\to\infty, n_1\to\infty$, and each sequence $(\varepsilon_n-\varepsilon)_{n_1}$ represents some zero-convergent sequence, 0_{n_2}. By Remark 1 (c) then

$$\lim_{n\to\infty}(\varepsilon_n-\varepsilon)\phi^{-1}=\lim 0_{n_2}\phi^{-1}=0.$$

By continuity in $S'$of the inverse Fourier transform operator the limit

$$Ft^{-1}(w_n\phi^{-1})\to g$$

in S' follows. ■

It is well known that the deconvolution problem may not be well-posed in function spaces, but in the weaker topology of generalized functions well-posedness obtains under more general conditions. Indeed, as shown in Zinde-Walsh (2008), the biases of kernel density estimators converge to zero as generalized functions even though they may diverge point-wise and in L_1.

However, even in the generalized functions topology well-posedness may not hold in some cases, in particular for super-smooth densities. Suppose that $\varepsilon_n, n=1,2,...$ and ε satisfy Theorem 2, that $\varepsilon_n\to\varepsilon$ in $S^{'}$, but ϕ^{-1} does not satisfy (11). As the following example shows $\phi^{-1}\varepsilon_n$ may then not converge in S' to $\phi^{-1}\varepsilon$.

Example. *Consider the function $\phi(x)=e^{-x^2}$, $x\in R$. Define in S a* *function* $b_n(x) =$

$$\begin{cases} e^{-n} & \text{if } n-\frac{1}{n}< x < n+\frac{1}{n}; \\ 0 < b_n(x) \le e^{-n} & \text{if } n-\frac{2}{n}< x < n+\frac{2}{n}; \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

This $b_n(x)$ *converges to* $b(x) \equiv 0$ *in* S'. *Indeed for any* $\psi \in S$

$$\int_{-\infty}^{\infty} b_n(x)\psi(x)dx = \int_{n-2/n}^{n+2/n} b_n(x)\psi(x)dx \to 0.$$

Now consider $\varepsilon_n = \varepsilon + b_n \to \varepsilon$. *We show that* $\varepsilon_n \phi^{-1}$ *does not converge in* S' *to* $\varepsilon\phi^{-1}$. *Such convergence would imply that* $(\varepsilon_n - \varepsilon)\phi^{-1} = b_n\phi^{-1} \to 0$ *in* S'.

But the sequence $b_n(x)\phi(x)^{-1}$ *does not converge. Indeed if it did then* $\int b_n(x)\phi^{-1}(x)\psi(x)dx$ *would converge for any* $\psi \in S$. *But for* $\psi \in S$ *such that* $\psi(x) = \exp(-|x|)$

$$\begin{aligned} \int b_n(x)e^{x^2}\psi(x)dx &\ge \int_{n-2/n}^{n+2/n} b_n(x)e^{x^2}\psi(x)dx \ge e^{-n}\int_{n-1/n}^{n+1/n} e^{x^2-x}dx \\ &\ge \frac{2}{n}e^{-2n+(n-1/n)^2}. \end{aligned}$$

This diverges. ■

3.3 Well-posedness in S' of the solution to the system of equations (7)

Consider the classical case of convolution pairs $(S', \mathcal{O}'_C)$ providing Fourier transforms in S' and $\mathcal{O}'_M$.

Recall (11) and define a subclass of functions on R^d, $\Phi(m,V)$, where $b \in \Phi(m,V)$ if it satisfies the condition

$$\int \left((1+t^2)^{-1}\right)^m |b(t)|\, dt < V < \infty \tag{17}$$

Theorem 5. *Suppose that* $(g_n, f_n), n = 1, 2, ...$ *and* (g, f) *belong to the convolution pair* (S', O'_C). *Additionally, let supp*$(\gamma) = W$, *with* W *a convex set in* R^d *with* 0 *as an interior point, and supp*$(\phi_n) \supseteq W$, *supp*$(\phi) \supseteq W$; $\phi_n(0) = \phi(0) = 1$. *Suppose that the sequences* $\varepsilon_{1n} = Ft(w_{1n})$ *for* $w_n = g_n * f_n$ *and* $\varepsilon_{2kn} = Ft(w_{2kn})$ *for* $w_{2kn} = (g_n)'_k * f_n$ *are such that* $\varepsilon_{1n} - \varepsilon_1 \to 0$ *and* $\varepsilon_{2kn} - \varepsilon_{2k} \to 0$ *in* S'. *Then if the functions* ϕ_n, ϕ_n^{-1} *restricted to* W *all belong to some* $\Phi(m,V)$ *the products* $\varepsilon_n \phi_n^{-1}$ *exist in* S' *and* $Ft^{-1}(\varepsilon_{1n}\phi_n^{-1}) \to g$ *in* S'.

Proof. Recall that all $\phi, \phi_n \in \mathcal{O}_M$. It follows that $(\phi)'_k, (\phi_n)'_k \in \mathcal{O}_M$. Also $(\phi)'_k, (\phi_n)'_k \in \Phi(m', V)$, where $m' = m + \iota$, with ι a vector of ones. From Theorem 3 it follows that for every n the functions γ_n and ϕ_n are uniquely identified on W. From now on we consider all functions restricted to W, but keep the same notation. The functions $\varkappa_{kn} = (\phi_n)'_k \phi_n^{-1} \in \mathcal{O}_M$ belong also to $\Phi(\tilde{m}, V)$ where $\tilde{m} = m + m'$. Without loss of generality assume that each $\varkappa_k$ is also in the same $\Phi(\tilde{m}, V)$, and so all $\varkappa_{kn}, \varkappa_k$ are in a bounded set in S' (Sz, p.240). From equations $\varepsilon_{1n}\varkappa_{kn} - ((\varepsilon_{1n})'_k - i\varepsilon_{2kn}) = 0$ and convergence of ε_{in} to ε_i we get that $\varepsilon_{1n}\varkappa_{kn} - \varepsilon_1\varkappa_k$ converges to zero in S'. For functions in $\mathcal{O}_M$ products with any elements from S' exist, thus $\varepsilon_{1n}\varkappa_{kn} - \varepsilon_1\varkappa_{kn}$ exists; moreover $(\varepsilon_{1n} - \varepsilon_1)\,\varkappa_{kn}$ converges to zero in S' by the hypocontinuity property (S, p.246). It follows that $\varepsilon_1(\varkappa_{kn} - \varkappa_k)$ converges to zero in S'. Since ε_1 is supported on W and $(\varkappa_{kn} - \varkappa_k) \in \mathcal{O}_M$ by continuity of the functional ε_1 it follows that $\varkappa_{kn} - \varkappa_k$ converges to zero on W. It then follows that $\phi_n - \phi \to 0$ in S' as well as pointwise and uniformly on bounded sets in W, thus $\phi_n^{-1} \to \phi^{-1}$ on W pointwise and uniformly on bounded sets and thus in S'.

Consider $\varepsilon_{1n}\phi_n^{-1} - \varepsilon_1\phi^{-1} = \varepsilon_{1n}(\phi_n^{-1} - \phi^{-1}) + (\varepsilon_{1n} - \varepsilon_1)\phi^{-1}$; this difference converges to zero in S', thus γ_n converges to γ in S' and since the Fourier transform is continuously invertible in S', $g_n = Ft^{-1}(\varepsilon_{1n}\phi_n^{-1}) \to g$ in S'. ■

Thus in the classical case well-posedness obtains for the solution of (7) as long as ϕ_n^{-1} satisfy (11) uniformly (all in the same class $\Phi(m, V)$).

4 Random generalized functions and stochastic convergence; consistency of solutions

This section examines stochastic convergence of the solutions to the deconvolution equation (1) and to (7) for stochastic sequences. If some estimators are available for either the function w (w_1 and w_{2k}) or, equivalently, for the Fourier transform, ε (ε_1 and ε_{2k}), stochastic convergence of the solutions provides consistency results.

4.1 Random generalized functions

Following Gel'fand and Vilenkin (1964) define random generalized functions as random linear continuous functionals on the space of test functions (see e.g. Koralov and Sinai, 2007 who consider specifically $S'-$ ch.17). In particular, any random generalized function $\tilde{b}$ on S is represented by a collection of (complex-valued) random variables on a common probability space that are indexed by $\psi \in S$, denoted $(\tilde{b}, \psi)$, such that

(a) $(\tilde{b}, (a_1\psi_1 + a_2\psi_2)) = a_1(\tilde{b}, \psi_1) + a_2(\tilde{b}, \psi_2)$ a.s.;

(b) if $\psi_{kn} \to \psi_k$ in S as $n \to \infty, k = 1, 2.., m$, then vectors

$$((\tilde{b}, \psi_{1n})...(\tilde{b}, \phi_{mn})) \to_d ((\tilde{b}, \psi_1)...(\tilde{b}, \phi_m))$$

where $\to_d$ denotes convergence in distribution.

As shown in Gel'fand and Vilenkin, equivalently, there exists a probability measure on S' such that for any set $\psi_1, ...\psi_m \in S$ the random vectors $((b, \psi_1), ...(b, \psi_m))$ have the same distribution as for some random functional $\tilde{b}$, $((\tilde{b}, \psi_1)...(\tilde{b}, \phi_m))$. An example is a generalized Gaussian process b, so defined if for any $\psi_1, ..., \psi_l$ the joint distribution of $(b, \psi_1), ..., (b, \psi_l)$ is Gaussian. A generalized Gaussian process is uniquely determined by its mean functional, μ_b : $(\mu_b, \psi) = E(b, \psi)$, and the covariance bilinear functional, $B_b(\psi_i, \psi_j) = E((b, \psi_i)\left(b, \psi_j\right))$.

Gelfand, Vilenkin (v.4, p. 260) give the covariance functional of the generalized derivative, W', of the Wiener process as

$$B_{W'}(\psi_1, \psi_2) = \int_0^\infty \psi_1(t)\overline{\psi_2(t)}dt$$

where the overbar represents complex conjugation; for real-valued processes it is not needed. The Fourier transform of a Gaussian random process b is also a Gaussian random process with covariance functional

$$B_{Ft(b)}(\psi_i, \psi_j) = E((b, Ft(\psi_i))\left(b, Ft(\psi_j)\right)).$$

So for $Ft(W')$ the covariance functional is

$$B_{Ft(W')}(\psi_1, \psi_2) = \int_0^\infty Ft(\psi_1)(\zeta)\overline{Ft(\psi_2)(\zeta)}d\zeta.$$

Of course, the mean functional is zero for W' and $Ft(W')$.

Gelfand and Vilenkin (1964) provide definitions and results for generalized random functions in D', rather than S'. One can similarly define random generalized functions on other spaces of test functions, not necessarily infinitely differentiable, e.g. on D_k of k times continuously differentiable functions with finite support, leading to space D'_k.

4.2 Stochastic convergence of random generalized functions

A random sequence b_n of elements of a space of generalized functions G' converges to zero in probability: $b_n - b \to_p 0$ in G', or almost surely: $b_n - b \to_{a.s.} 0$ in G' if for any set $\psi_1, ...\psi_m \in G$ the random vectors $((b_n, \psi_1), ...(b_n, \psi_m)) \to_p (0, ..., 0)$ or correspondingly, $\Pr(((b_n, \psi_1), ...(b_n, \psi_m)) \to_{S'} (0, ..., 0)) = 1$.

Similarly, convergence in distribution of generalized random processes $b_n \Rightarrow_d b$ is defined by the convergence of all multivariate distributions for random vectors $((b_n, \psi_1), ...(b_n, \psi_m)) \to_d ((b, \psi_1)...(b, \phi_m))$.

Remark 2. *(a) If* $b_n - b \to_p 0$ *in* S' *then* $Ft(b_n) - Ft(b) \to_p 0$ *in* S' *and* $Ft^{-1}(b_n) - Ft^{-1}(b) \to_p 0$ *in* S'*. Indeed, for any integer* m *and* $\psi_1, ...\psi_m \in S$

$$\begin{aligned} & ((Ft(b_n) - Ft(b), \psi_1), ..., (Ft(b_n) - Ft(b), \psi_m)) \\ = \; & ((b_n - b, Ft(\psi_1)), ..., (b_n - b, Ft(\psi_m))) \end{aligned}$$

and since if $\psi \in S$*, so does its Fourier transform, the set* $Ft(\psi_1), ..., Ft(\psi_m) \in S$ *and* $((b_n - b, Ft(\psi_1)), ..., (b_n - b, Ft(\psi_m))) \to_p 0$.

(b) If $\mu \in G$ *and* $b_n - b \to_p 0$ *in* G'*, then* $b_n\mu - b\mu \to_p 0$ *in* G'*. This follows similarly from the fact that* $\mu\psi \in G$ *for* $\psi \in G$.

(c) Parts (a) and (b) of this Remark also hold with $\to_{a.s.}$ *replacing* $\to_p$ *and with convergence to zero replaced by convergence in distribution to a limit generalized random process.*

4.3 Consistent estimation of solutions to stochastic convolution equations

Suppose that the known functions, w or w_1, w_{2k} (equivalently, ε or $\varepsilon_1, \varepsilon_{2k}$) are consistently estimated in S'. First we give conditions for consistency of deconvolution in S'. Denote by CC the subspace $CC \in S'$ that consists of all $c \in S'$ such that $c = a * b$ for some (a, b) in the convolution pair $(S', \mathcal{O}'_C)$. Then the corresponding Fourier transforms, $Ft(c)$, are in the subspace of products, $Ft(a)Ft(b)$ with $Ft(b) \in \mathcal{O}_M$.

Theorem 6. *(a) Suppose that the generalized functions (g, f) belong to the convolution pair (S', O'_C) and that ϕ is a known function that satisfies the condition of Theorem 2, and that ϕ^{-1} satisfies (11). Suppose further that w_n is a random sequence of generalized functions from CC such that for $\varepsilon_n = Ft(w_n)$ the difference $\varepsilon_n - \varepsilon \to_p 0$ in S'. Then the products $\varepsilon_n \phi^{-1}$ exist in S' and $Ft^{-1}(\varepsilon_n \phi^{-1}) - g \to_p 0$ in S'.*

(b) Suppose that the generalized function g and random generalized functions g_n for integer n are such that the pairs (g, f) and (g_n, f) satisfy the conditions of Theorem 4 with the same (A, B) and that $\varepsilon_n - \varepsilon \to_{a.s.} 0$ in S'. Then $Ft^{-1}(\varepsilon_n \phi^{-1}) - g \to_{a.s.} 0$ in S'.

Proof. (a) By Theorem 2, for every ε_n the product $\varepsilon_n \phi^{-1}$ exists in S'. We need to show that $(\varepsilon_n - \varepsilon)\, \phi^{-1}$ converges to zero in probability in S'. Since ϕ^{-1} satisfies (11) and is in $\mathcal{O}_M$, for any $\psi \in S$ the product $\phi^{-1}\psi \in S$. Then for any integer m and $\psi_1, ...\psi_m \in S$, the vector

$$\begin{aligned} & (((\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_1), ..., ((\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_m)) \\ = \; & (((\varepsilon_n - \varepsilon)\,, \phi^{-1}\psi_1), ..., ((\varepsilon_n - \varepsilon)\,, \phi^{-1}\psi_m)) \to_p 0. \end{aligned}$$

By assumption $\varepsilon_n - \varepsilon \to_p 0$ in S'. By Remark 2 (a) the rest follows.

(b) By Theorem 4 the event

$$(((\varepsilon_n - \varepsilon)\,, \psi_1), ..., ((\varepsilon_n - \varepsilon)\,, \psi_m)) \to_{a.s.} 0$$

implies $(((\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_1), ..., ((\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_m)) \to_{a.s.} 0$ and then

$$((Ft^{-1}\,(\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_1), ..., (Ft^{-1}\,(\varepsilon_n - \varepsilon)\,\phi^{-1}, \psi_m)) \to_{a.s.} 0.$$

■

The next Theorem gives conditions for a consistent plug-in estimator for the model that leads to the system (7).

Consider the space $\Phi^{1,-1}(m,V)$ of functions b, such that $b \in \Phi(m,V)$ and $b^{-1} \in \Phi(m,V)$. Denote by $\mathcal{O}_{M,m,V}$ the intersection $\mathcal{O}_M \cap \Phi^{1,-1}(m,V)$ and consider the subspace $CC(m,V)$ of all $c = a * b$ where $Ft(b) \in \mathcal{O}_M \cap \Phi^{1,-1}(m,V)$.

Theorem 7. *Suppose that the generalized functions (g,f) belong to the convolution pair (S', O'_C); supp(γ) $= W$, with W a convex set in R^d with 0 as an interior point and supp(ϕ) $\supseteq W$; $\phi(0) = 1$ and $\phi \in \Phi^{1,-1}(m,V)$. If*

w_{1n}, w_{2n} are random sequences with each $w_{1n}, w_{2n} \in CC(m, V)$ and the corresponding random sequences $\varepsilon_{1n} = Ft(w_{1n})$ and $\varepsilon_{2n} = Ft(w_{2n})$ are such that $\varepsilon_{1n} - \varepsilon_1 \to_{a.s.} 0$ and $\varepsilon_{2n} - \varepsilon_2 \to_{a.s.} 0$ in S', then there are $\phi_n \in \Phi^{1,-1}(m, V)$ such the products $\varepsilon_n \phi_n^{-1}$ exist in S' and $Ft^{-1}(\varepsilon_{1n}\phi_n^{-1}) - g \to_{a.s.} 0$ in S'.

Proof. For every pair w_{1n}, w_{2kn} the product $\gamma_n = \varepsilon_n \phi_n^{-1}$ is uniquely defined by Theorem 3; by Theorem 5 for any sequence for which $\varepsilon_{1n} - \varepsilon_1 \to 0$ and $\varepsilon_{2kn} - \varepsilon_{2k} \to 0$ in S' also $\gamma_n - \gamma \to 0$ in S'. Then $\Pr((\gamma_n - \gamma, \psi_1), ..., (\gamma_n - \gamma, \psi_m)) \to 0) \geq$

$$\begin{aligned} & \Pr((\varepsilon_{1n} - \varepsilon_1, \psi_1), ..., (\varepsilon_{1n} - \varepsilon_1, \psi_m), \\ & (\varepsilon_{2kn} - \varepsilon_{2k}, \psi_1), ..., (\varepsilon_{2kn} - \varepsilon_{2k}, \psi_m)) \\ \to \ & 0) \end{aligned}$$

equals 1. ■

The conditions of Theorems 6 and 7 require considering the classical pair of spaces (S', O'_C), but other pairs that relax the requirement of $f \in O'_C$ by restricting g can be examined; this is done for some specific applications in the next section.

5 Applications to statistical and econometrics problems

The approach developed here can find applications in various problems of deconvolution and regression estimation for errors in variables. Since convergence results here are in the weak topology of generalized functions, there is no implication that convergence in commonly used topologies holds; if the problem is amenable to conventional approaches, e.g. involves functions in normed spaces, it may be advantageous to exploit these specific properties to obtain more precise results. On the other hand, there are problems of interest that do not fit into the more conventional framework and that can benefit from the generalized functions approach. The first subsection develops properties of a deconvolution estimator of a function with bounded support in a fixed design set-up; a generalized Gaussian limit process is derived; further if the function is defined on a sparse support, e.g. is a sum of delta-functions, it can benefit from shrinkage. The rate of a deconvolution shrinkage estimator is derived here in the multivariate generalized function case, extending the results of Klann et al, 2007. The second subsection studies the errors in variables regression that leads to (7), where the regression function is in $L_1(R^d)$. A semiparametric estimator in L_1 was provided recently by Wang and Hsiao (2010); results for nonparametric estimation are given here.

5.1 Deconvolution on a compact support with a fixed design; rate of the deconvolution shrinkage estimator

Consider the function of interest g to have a compact support. The support of g is restricted, thus any $f \in S'$ can be considered for deconvolution, since $g \in \mathcal{O}'_C$ and the convolution is defined.

Theorem 2 of this paper asserts that as long as the support of the $\phi = Ft(f)$ contains support of $\gamma = Ft(g)$, identification holds in the deconvolution problem. For g equal to a sum of delta-functions the support of γ will be the whole space R^d. If only delta-functions enter then γ may be a periodic function. For identification of a periodic function it is sufficient that support of ϕ contain an interval of period length. If further ϕ^{-1} satisfies (11) by Theorem 4 the deconvolution problem is well-posed in S'; if g is a sum of delta-functions or other functions that lead to periodic γ and support of ϕ contains an interval of period length, then we can restrict ϕ to this finite interval, and there ϕ^{-1} automatically satisfies (11) since it is bounded and integrable.

Assumption 3. *(i) The function f is a known function with a Fourier transform, ϕ, that is continuous and non-zero on R^d; ϕ^{-1} satisfies (11); the generalized function g has bounded support. (ii) For design points $\{x_1, ...x_n\}$, $x_i \in R^d$ the data is $y = w + u$, where $w = g * f$ and u is white noise with variance $\sigma^2 = n^{-1}$.*

Assumptions 3 (i) and (ii) are satisfied in (K) where univariate g is either a sum of delta functions or some other sum of peaks function and where f is defined to be a second-order $B-$spline. The results here apply to any nonparametric multivariate function g with bounded support and any density f that satisfies Assumption 3 (i).

Consider a linear estimator, $\hat{w}$ for w. Here we employ a kernel estimator to directly use the results in Zinde-Walsh, 2008 (ZW). Suppose that assumption A of that paper holds; according to it the kernel function $K(w)$ is a bounded $l-$th order kernel with support on a unit cube in R^d. Here to simplify derivations we consider a bandwidth parameter, h, the same for all components of the vector in R^d.

Define the estimator

$$\hat{w}(x) = \frac{1}{n}\Sigma\frac{1}{h}y_j K(\frac{x_j - x}{h}) = \int y_j \frac{1}{h} K(\frac{x_j - x}{h}) dx_j.$$

The measure dx_j here represents the uniform measure on the design points. Under Assumption 3 and for a bandwidth, h, that satisfies $h \to 0$, $h^{2l+d}n \to 0$ and $h^d n \to \infty$ as $n \to \infty$ conditions of Theorem 3 of (ZW) are satisfied. The estimator $\hat{w}(x)$ of the generalized function $w(x)$ differs from the density estimator only by the weights y_j thus is examined similarly to the density estimator of Theorem 3. The bias is $O(h^l)$, indeed, by expanding the smooth test function and utilizing the property of the $l-$th order kernel: $(E(\hat{w}(x) - w(x)), \psi) =$

$$\begin{aligned} & \int \left[\int \left[w(x+ht) - w(x) \right] K(t) dt \right] \psi(x) dx \qquad (18) \\ = & \int \int w(x) \psi(x - ht) dx K(t) dt - \int \int w(x) \psi(x) dx \\ = & (-1)^l \frac{1}{l!} h^l B(\psi) + O(h^{l+1}) \end{aligned}$$

with $B(\psi) = \Sigma_{m(l)} \int w(x) \partial^{m(l)} \psi(x) dx \int t^{m(l)} K(t) dt$, where $m(l) = (m_1, ...m_d)$ with $\Sigma m_i = l$ and $t^{m(l)} = t_1^{m_1} ... t_d^{m_d}$, $\partial^{m(l)}$ is the differentiation operator $\frac{\partial^{m_1}}{\partial x_1^{m_1}} ... \frac{\partial^{m_d}}{\partial x_d^{m_d}}$. The variance computation similarly to the one in (ZW) provides for $(nh^k)^{\frac{1}{2}} (\hat{w}(x) - w(x))$ the limit covariance functional

$$\begin{aligned} & (C_{w_\delta}, (\psi_1, \psi_2)) \\ = & \int \psi_1(x) \psi_2(x) dx \int K(t)^2 dt + O(h). \end{aligned}$$

The proof in (ZW) can be applied to show that $(nh^d)^{\frac{1}{2}} (\hat{w}(x) - w(x))$ converges to a generalized Gaussian process. The following proposition gives the result for the deconvolution estimator.

Theorem 8. *Under Assumption 3 for* $h \to 0$, $h^{2l+d} n = 0(1)$ and $h^d n \to \infty$ as $n \to \infty$ *the estimator of* g, $\hat{g} = Ft^{-1}(\phi^{-1} Ft(w))$, *is such that the process* $(nh^d)^{\frac{1}{2}} \left[\left(\hat{g} - g, \psi \right) - \tilde{B}(\psi) \right]$, *with* $\tilde{B}(\psi) = (-1)^l \frac{1}{l!} h^l B(Ft \left[\phi^{-1} Ft(\psi) \right])$, *converges to a generalized Gaussian process with mean functional zero and co-*

variance functional $(C_{\hat{g}}, (\psi_1, \psi_2))$; *for any* $\psi_1, \psi_2 \in S$ *the functional* $C_{\hat{g}}(\psi_1, \psi_2) =$

$$\int Ft\left[\phi(x)^{-1}Ft^{-1}\psi_1(x)\right]\overline{Ft\left[\phi(x)^{-1}Ft^{-1}\psi_2(x)\right]}dx \int K(t)^2 dt \tag{19}$$

Proof. Apply deconvolution to the process $\hat{w}_\delta(\cdot) - w(\cdot)$. This requires to apply first the Fourier transform providing

$$(Ft(\hat{w}(\cdot) - w(\cdot), \psi) = (\hat{w}(\cdot) - w(\cdot), Ft(\psi)).$$

Then divide by the function ϕ to get

$$(\phi^{-1}Ft\left[\hat{w}_\delta(\cdot) - w(\cdot)\right], \psi) = (\left[\hat{w}_\delta(\cdot) - w(\cdot)\right], \phi^{-1}Ft(\psi)).$$

If ϕ^{-1} belongs to $\mathcal{O}_M$ then $\phi^{-1}Ft(\psi) \in S$ and the equality is obvious. If ϕ^{-1} is continuous and satisfies (11) the generalized process needs to be considered on the subspace of continuous test functions, $S \oplus \phi^{-1}S$, where a Gaussian process is similarly defined. The covariance for the functional $\left(nh^d\right)^{\frac{1}{2}}\left[\hat{w}(\cdot) - w(\cdot)\right]$ on $S \oplus \phi^{-1}S$ for $\tilde{\psi}_1, \tilde{\psi}_2$ where $\tilde{\psi}_i = \phi^{-1}Ft(\psi_i)$, $\psi_i \in S$ is given by the expression $C_{\hat{g}}(\tilde{\psi}_1, \tilde{\psi}_2) =$

$$\int \tilde{\psi}_1(x)\tilde{\psi}_2(x)dx \int K(t)^2 dt.$$

The limit process for $\hat{g} - g$ requires examining

$$\begin{aligned}(\hat{g} - g, \psi) &= (Ft^{-1}(\phi^{-1} Ft(\hat{w}_\delta - w)), \psi) \\ &= (\phi^{-1} Ft(\hat{w} - w), Ft^{-1}\psi) \\ &= (Ft(\hat{w} - w), \phi^{-1} Ft^{-1}(\psi)) \\ &= (\hat{w} - w, Ft(\phi^{-1} Ft^{-1}(\psi)).\end{aligned}$$

Consider $\Sigma_{m(l)} \partial^{m(l)} Ft \left[\phi^{-1} Ft^{-1}(\psi)\right]$; by (11) $t^{m(l)}\phi(t)^{-1} Ft^{-1}(\psi)$ is integrable and continuous, thus $\Sigma_{m(l)} \partial^{m(l)} Ft \left[\phi^{-1} Ft^{-1}(\psi)\right]$ is well defined and $B(Ft \left[\phi^{-1} Ft^{-1}(\psi)\right])$ can be defined. Thus $\left(nh^d\right)^{\frac{1}{2}} (\hat{g} - g, \psi)$ has mean functional

$$\frac{1}{2} \left(nh^d\right)^{\frac{1}{2}} h^l \left\{B(Ft \left[\phi^{-1} Ft^{-1}(\psi)\right]) + o(h)\right\}$$

and the covariance functional given by $(C_{g_\delta}, (\psi_1, \psi_2)) =$

$$\int Ft \left[\phi(x)^{-1} Ft^{-1} \psi_1(x)\right] \overline{Ft \left[\phi(x)^{-1} Ft^{-1} \psi_2(x)\right]} dx \int K(t)^2 dt + O(h).$$

■

Despite the possible singularity in g the rates here are the usual non-parametric rates; as discussed in ZW this reflects the weaker topology of generalized functions and the estimators may not converge at those rates pointwise or in any usual norm. However, these results may be used for inference. The structure of the generalized function g was not important for the results that only used the bounded support assumption.

To sharpen the estimator for the sum of peaks scarcity could be taken into account via shrinkage or thresholding; as in (K) this is done by truncating the linear estimator, $\hat{w}$. Consider the estimator $\hat{w}_\lambda = \hat{w} I(\hat{w} \geq \lambda)$.

Assumption 4. *Assumption 3 holds and additionally w is Lipschitz continuous and u is Gaussian white noise.*

Lipschitz continuity of w would follow if either f or g had that property.

Since here $\hat{w}(x)$ has a Gaussian distribution, the truncated estimator has a truncated Gaussian distribution. We next derive its moments and find suitable rates for λ and h and for the resulting estimator, $\hat{w}_\lambda$. Define the deconvoluition shrinkage estimator for g by $\hat{g}_\lambda = Ft^{-1}(\phi^{-1}Ft(\hat{w}_\lambda))$.

Theorem 9. *Under Assumptions 3 and 4 for* $h = n^{-\frac{1}{2l+d}} \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{4l+2d}}$; and $\lambda = n^{-\frac{1}{2}+\frac{d}{2l+d}} \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{2}-\frac{d}{4l+2d}}$ *as $n \to \infty$ the deconvolution shrinkage estimator as a generalized random process has the rate* $n^{-\frac{l}{2l+d}} \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}}$ *and for any $\psi_1, ...\psi_l$ satisfies*

$$n^{-\frac{l}{2l+d}} \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}} (((\hat{g}_\lambda - w), \psi_1), ...((\hat{g}_\lambda - w), \psi_l)) = O_p(1).$$

Proof. First, the expected value for the estimator $\hat{w}$ was

$$E\hat{w}(x) = \int w(x_j) \frac{1}{h} K(\frac{x_j - x}{h}) dx_j,$$

and under Assumption 4 $E(\hat{w}_\delta - w) = O(h)$; the variance

$$var(\hat{w}) = E\left[\hat{w}(x) - E(\hat{w}(x))\right]^2 = \frac{1}{nh^d}\left[\int K(t)^2 dt + O(h)\right].$$

Define (for each x) $a_w = \frac{\lambda - E(\hat{w}-w)}{(var\hat{w})^{\frac{1}{2}}}$; then $a_w = \sqrt{nh^k}(\lambda + O(h))$. Denote by $b(a)$ the value of the inverse Mills ratio at a : $b(a) = \frac{\phi(a)}{1-\Phi(a)}$, where ϕ, Φ are the density and distribution functions for the standard normal; for $a = \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{2}}$, $b(a)$ is bounded. Using the formulae for moments for truncated normal (Cohen, 1951) we obtain, assuming that $a_w = \sqrt{nh^d}(\lambda + O(h)) = \left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{2}}(1 + o(1))$ as $n \to \infty$,

$$E(\hat{w}_\lambda - w) = E(\hat{w} - w) + (var\hat{w})^{\frac{1}{2}} b(\frac{\lambda - E(\hat{w}-w)}{(var\hat{w})^{\frac{1}{2}}});$$

and $var(\hat{w}_\lambda) =$

$$\begin{aligned} & var\hat{w}\left[1 - b(\frac{\lambda - E(\hat{w}-w)}{(var\hat{w})^{\frac{1}{2}}})\left\{b\left(\frac{\lambda - E(\hat{w}-w)}{(var\hat{w})^{\frac{1}{2}}}\right) - \frac{\lambda - E(\hat{w}-w)}{(var\hat{w})^{\frac{1}{2}}}\right\}\right] \\ = \; & O(\frac{1}{nh^d}\left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{2}}). \end{aligned}$$

For $\psi \in S$ recalling (18) we have $(E(\hat{w}_\lambda - w), \psi) = O(h^l) + O(\left(\frac{1}{nh^d}\right)^{\frac{1}{2}})$.

Set $h = n^{-\frac{1}{2l+d}}\left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{4l+2d}}$; and $\lambda = n^{-\frac{1}{2}+\frac{d}{2l+d}}\left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{1}{2}-\frac{d}{4l+2d}}$.

Then

$$\begin{aligned}
E(\hat{w}_\lambda - w, \psi) &= O(n^{-\frac{l}{2l+d}} \left(2 \left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}});\\
E[(\hat{w}_\lambda - w, \psi_1)(\hat{w}_\lambda - w, \psi_2)] &= O\left(n^{-\frac{2l}{2l+d}} \left(2 \left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{2l}{4l+2d}}\right).
\end{aligned}$$

This determines the rates for the generalized random process $\hat{w}_\lambda - w$ (given here by the ordinary random functions), so that for any $\psi_1, ...\psi_v \in S$

$$n^{-\frac{l}{2l+d}} \left(2 \left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}} (((\hat{w}_\lambda - w), \psi_1), ...((\hat{w}_\lambda - w), \psi_v)) = O_p(1).$$

Then similarly to the Proof of Theorem 7 we have for the deconvolution estimator, $\hat{g}_\lambda$,

$$\begin{aligned}
(\hat{g}_\lambda - g, \psi) &= (Ft^{-1}(\phi^{-1}Ft(\hat{w}_\lambda - w)), \psi)\\
&= (\phi^{-1}Ft(\hat{w}_\lambda - w), Ft^{-1}\psi)\\
&= (Ft(\hat{w}_\lambda - w), \phi^{-1}Ft^{-1}(\psi))\\
&= (\hat{w}_\lambda - w, Ft(\phi^{-1}Ft^{-1}(\psi)).
\end{aligned}$$

It follows that the deconvolution kernel shrinkage estimator has the rate $O(n^{-\frac{l}{2l+d}} \left(2 \left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}})$ as a generalized random process:

$$n^{-\frac{l}{2l+d}} \left(2 \left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2d}} (((\hat{g}_\lambda - w), \psi_1), ...((\hat{g}_\lambda - w), \psi_v)) = O_p(1).$$

■

In the univariate case the rate is $n^{-\frac{l}{2l+1}}\left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+2}}$; selecting a higher order kernel can bring the rate for the generalized random function estimators closer to parametric; the rate is comparable to the best in (K) that could be achieved by weakening the Besov norm topology, but the approach here does not require the extra assumptions made there on the function to be estimated. The results here apply to the multivariate case, thus in the bivariate case that may be of interest e.g. in astrophysics, the rate becomes $n^{-\frac{l}{2l+2}}\left(2\left|\log n^{-\frac{1}{2}}\right|\right)^{\frac{l}{4l+4}}$.

5.2 Consistent nonparametric estimation in L_1 space for the errors in variables model

Consider the model in (3)-(5) that provides a EIV regression where z represents a second measurement or possibly a given projection onto a set of instruments.

Assumption 4. *For x, y, z the model (3)-(5) holds with $E(u_y|z,u) = 0; E(u_x|z,u,u_y) = 0$ and u a zero-mean Berkson type measurement error independent of z. The observations $\{x_i, y_i, z_i\}_{i=1}^n$ are independently sampled for any integer n.*

These model assumptions combined with Assumption 2(a) lead to the system of equations (7) that generalizes (6); the univariate (6) case was examined by Schennach (2007) and (in particular for L_1 class) by Zinde-

Walsh (2009).

Semiparametric GMM estimation of this model for a class of regression functions that includes functions in $L_1(R^d)$ was proposed by Wand and Hsiao (2010); semiparametric GMM estimation was also discussed for somewhat different classes of univariate parametric regression functions in Schennach (2007); Zinde-Walsh (2009) gave a specific description of the class of function to which the semiparametric GMM method proposed by Schennach applied and corrected the estimator by providing appropriate weighting functions.

Here a method of consistent nonparametric estimation of a regression function in $L_1(R^d)$ for the model (3)-(5) is developed.

Assumption 5. *The function $g \in L_1(R^d)$ and f is a density function with continuous and continuously differentiable characteristic function ϕ such that either (i) supp(γ) is compact, or (ii) $|\ln|\phi(\xi)|| \leq \Phi(\xi)$ and $|\ln|\phi'_k(\xi)|| \leq \Phi(\xi)$ with $\Phi(\xi) \leq a + b \ln \Pi_{i=1}^d \left(1+\xi_i^2\right)$ for some $a, b \geq 0, k = 1, ..., d$.*

Assumption 5 implies that $w_1 \in L_1$ and that ϕ, ϕ'_k and ϕ^{-1} satisfy (11) on supp(γ). This assumption is less restrictive than requiring $\phi \in \mathcal{O}_M$ since only existence of first continuous derivatives is required, on the other hand, it calls for the existence of a bounding function in the case of unbounded support of γ.

Recall that $w_i(z)$, $i = 1, 2$ represent conditional mean functions.

Assumption 6. *For some $m \geq 0$ moments of order $q > d+1$ of $\frac{1}{(1+z^2)^m} y$, $\frac{1}{(1+z^2)^m} xy$ conditional on z are bounded; the functions $\frac{1}{(1+z^2)^m} w_1(z)$, $\frac{1}{(1+z^2)^m} w_{2k}(z)$, $k = 1, ...d$, are bounded in absolute value by some $V < \infty$ and* *continuous; the density of* z, $f_z(z)$, *exists and is such that for* z *in any closed sphere* $S(x,r) \in$*supp*(f_z) *the* ess$\inf f_z(z) > 0$; *the kernel* K *is the indicator function of the unit sphere; the bandwidth* h *is such that* $h \to 0$ *and satisfies* $n^{1-\frac{d}{q-1}}h^d \to \infty$.

Assumption 6 is sufficient to ensure that Nadaraya-Watson estimators, $\left(\frac{1}{(1+z^2)^m}w_1(z)\right)_n$, $\left(\frac{1}{(1+z^2)^m}w_{2k}(z)\right)_n$, of $\frac{1}{(1+z^2)^m}w_1(z)$, $\frac{1}{(1+z^2)^m}w_{2k}(z)$, respectively, with K and h that satisfy the assumption converge in probability uniformly over any compact set (Devroye, 1978). This implies convergence in probability in D' for $\tilde{w}_{in} = (1+z^2)^m \left(\frac{1}{(1+z^2)^m}w_{in}(z)\right)$, subscript $i = 1, 2k$ $(k = 1, ..., d)$:

$$\begin{aligned} (\tilde{w}_{1n}(z) - w_1(z)) &= o_p(1) \text{ in } D'; \\ (\tilde{w}_{2kn}(z) - w_{2k}(z)) &= o_p(1) \text{ in } D'. \end{aligned} \tag{20}$$

The continuity of the functions w_i required in Assumption 6 would follow if either one of g or f were continuous. Some of the conditions on z such as existence of density (note that continuity is not assumed) and on the kernel can be relaxed somewhat as in Devroye (1978). Note that $\tilde{w}_{in}$ for $i = 1, 2$ are integrable functions since they are defined on a finite support.

If support of the conditional moment functions, w_i, is bounded, convergence in S' is implied; then denote $\tilde{w}_{in}$ by w_{in}. Otherwise define estimators $w_{in}(z)$ that equal $\tilde{w}_{in}(z)$ if $|w_{in}(z)| < V(1+z^2)^m$ and $V(1+z^2)^m$, otherwise. For any $\zeta > 0$ and any $\psi \in S$ there is a compact subset $A \subset R^d$ such that

$\int_{R^d \backslash A} V(1+z^2)^m |\psi(z)| \, dz < \zeta$. It follows that $w_{in}(z) - w_i(z) = o_p(1)$ in S'. If $m = 0$ the moment condition is a usual condition made for the Nadaraya-Watson estimator of w_i; here essentially just the growth of the conditional moment functions has to be restricted; this provides estimators that converge in the topology of S'. The bound V is assumed known here; more restrictive assumptions, including in particular differentiability of w_i could provide a uniform over compact sets rate of convergence for e.g. asymptotically optimal estimators (e.g. Stone, 1982); then V that defines w_{in} could grow with sample size.

For Fourier transforms $\varepsilon_{in} = Ft(w_{in})$ we obtain

$$\begin{aligned} \varepsilon_{1n}(z) &\rightarrow {}_p\varepsilon(z) \text{ in } S'; \\ \varepsilon_{2kn}(z) &\rightarrow {}_p\varepsilon_{2k}(z) \text{ in } S'. \end{aligned} \tag{21}$$

Since $w_1 \in L_1$ its Fourier transform, ε_1, is continuous.

Theorem 10. *If Assumptions 4-6 and conditions of Theorem 3(a) are satisfied, then (i) if supp(γ) is compact the plug-in estimator* $\gamma_n = \phi_n^{-1}\varepsilon_1$

$$\phi_n = \left[\exp(-\int_0^\zeta \Sigma\varepsilon_{1n}^{-1}(\xi)((\varepsilon_{1n}(\xi))'_k - i\varepsilon_{2n}(\xi))d\xi_k\right]\varepsilon_{1n}$$

defined on supp(γ) is such that $Ft^{-1}(\gamma_n) - g \rightarrow_p 0$ *in* S'

(ii) generally the estimator $\gamma_n = \tilde{\phi}_n^{-1}\varepsilon_1$ *with*

$$\tilde{\phi}_n^{-1} = \left[\exp(-\int_0^\zeta \Sigma\varepsilon_{1n}^{-1}(\xi)((\varepsilon_{1n}(\xi))'_k - i\varepsilon_{2n}(\xi))d\xi_k\right],$$

if

$$\left|\left[\exp(-\int_0^\zeta \Sigma\varepsilon_{1n}^{-1}(\xi)((\varepsilon_{1n}(\xi))'_k - i\varepsilon_{2n}(\xi))d\xi_k\right]\right| < 2B(\zeta)$$

and

$$\tilde{\phi}_n^{-1} = \exp(a + b\ln\Pi_{i=1}^d\left(1+\zeta_i^2\right))$$

otherwise, where $B(\zeta) = \exp(a + b\ln\Pi_{i=1}^d\left(1+\zeta_i^2\right))$, *is such that* $Ft^{-1}(\gamma_n) - g \to_p 0$ *in* S'.

Proof. Theorem 3(a) implies that $\phi^{-1}(\zeta) = \exp(-\int_0^\zeta \Sigma\varkappa_k(\xi)d\xi_k)$, where $\varkappa_k(\xi)$ is the unique continuous function that solves $\varepsilon_1\varkappa - ((\varepsilon_1)'_k - i\varepsilon_{2k}) = 0$ in S'; note that since here ε_1 is continuous, it follows that the difference $(\varepsilon_1)'_k - i\varepsilon_{2k}$ is continuous (even though $(\varepsilon_1)'_k, \varepsilon_{2k}$ may not be continuous functions). Since ϕ^{-1} satisfies (11), $\varepsilon_1\phi^{-1}$ exists in S' (see, e.g. proof of Theorem 4); then $g = Ft^{-1}(\varepsilon_1\phi^{-1})$ in S'. By the conditions of Theorem 3 ε_1 is non-zero on convex supp(γ). Moreover, by Assumption 5 if supp(γ) is non-compact there still exist bounds $\left|\phi^{-1}\right| \leq \exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b$ and $|\varkappa_k| \leq \left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]^2$ for any k.

Start by examining the behavior of the estimators on a compact set A. In the case of compact support, set $A =$supp(γ).

Consider the estimator function $\varkappa_{kn}(\xi) = \varepsilon_{1n}^{-1}((\varepsilon_{1n})'_k - i\varepsilon_{2kn})$ on A.

First, ε_{1n}; this is a sequence of (infinitely differentiable) functions that converge in probability in S' to the continuous function ε_1. It follows that they converge in probability pointwise and uniformly on the bounded set A. Then for any $0 < \zeta_1, \zeta_2 < 1$ we can find $N_1 \equiv N(A, \zeta_1, \zeta_2)$ such that for $n > N_1$

$$\Pr(\sup_A |\varepsilon_{1n} - \varepsilon_1| > \zeta_1^3) < \zeta_2.$$

Set $\zeta_1 \leq \inf_A |\varepsilon_1|$, then

$$\begin{aligned} \Pr(\inf_A |\varepsilon_{1n}| &< 2\zeta_1) < \zeta_2; \text{ and} \\ \Pr(\sup_A \left|\varepsilon_{1n}^{-1}\right| &> \zeta_1^{-1}) < \zeta_2. \end{aligned}$$

It follows that $\Pr\left(\sup_A \left|\varepsilon_{1n}^{-1} - \varepsilon_1^{-1}\right| > \zeta_1\right) =$

$$\begin{aligned} & \Pr\left(\sup_A \left|\left(\varepsilon_{1n}^{-1}(\varepsilon_1 - \varepsilon_{1n}\right)\varepsilon_1^{-1}\right| > \zeta_1\right) \\ \leq\ & \Pr\left(\sup_A \left|\left(\varepsilon_{1n}^{-1}(\varepsilon_1 - \varepsilon_{1n}\right)\right| > \zeta_1^2\right) \\ \leq\ & \Pr\left(\sup_A \left|\varepsilon_{1n}^{-1}\right| > \zeta_1^{-1}, \sup_A \left|\varepsilon_{1n}^{-1}(\varepsilon_1 - \varepsilon_{1n})\right| > \zeta_1^2\right) \\ & + \Pr\left(\sup_A \left|\varepsilon_{1n}^{-1}\right| \leq \zeta_1^{-1}, \sup_A \left|\varepsilon_{1n}^{-1}(\varepsilon_1 - \varepsilon_{1n})\right| > \zeta_1^2\right) \\ \leq\ & \Pr\left(\sup_A \left|\varepsilon_{1n}^{-1}\right| > \zeta_1^{-1}\right) + \Pr\left(\sup_A |\varepsilon_1 - \varepsilon_{1n}| > \zeta_1^3\right) \\ \leq\ & 2\zeta_2. \end{aligned}$$

Since the sequence of continuous random functions $((\varepsilon_{1n})'_k - i\varepsilon_{2kn})$ converges in probability in S' to a continuous function $((\varepsilon_1)'_k - i\varepsilon_{k2})$ this sequence also converges uniformly on the compact set A. Define $\bar{B} = \sup_{A,k} \left|(\varepsilon_1)'_k - i\varepsilon_{k2}\right|$. For $0 < \zeta_4$ find N_2 such that $\Pr(\sup_A \left|((\varepsilon_{1n})'_k - i\varepsilon_{2kn}) - ((\varepsilon_1)'_k - i\varepsilon_{k2})\right| > \zeta_4) < \zeta_2$ for $n > N_2$. Bound $\Pr(\sup_A |\varkappa_{kn}(\xi) - \varkappa_k(\xi)| > \zeta_5) \leq$

$$\Pr(\sup_A \left|\varepsilon_{1n}^{-1}\right| \left|((\varepsilon_{1n})'_k - i\varepsilon_{2n} - (\varepsilon_1)'_k + i\varepsilon_2\right| + \sup_A \left|\varepsilon_{1n}^{-1} - \varepsilon_1^{-1}\right| \left|(\varepsilon_1)'_k - i\varepsilon_{k2}\right| > \zeta_5) \leq$$

$$\begin{aligned} \Pr(\sup_A \left|\varepsilon_{1n}^{-1}\right| &> \zeta_1^{-1}) + \Pr\left(\sup_A \left|((\varepsilon_{1n})'_k - i\varepsilon_{2kn} - (\varepsilon_1)'_k + i\varepsilon_{k2}\right| > \zeta_5\zeta_1\right) \\ + \Pr(\sup_A \left|\varepsilon_{1n}^{-1} - \varepsilon_1^{-1}\right| &> \zeta_5/\bar{B}). \end{aligned}$$

If $\zeta_5 = \min\left\{\zeta_1\bar{B}, \zeta_4/\zeta_1\right\}$ the probability as $n > \max\{N_1, N_2\}$ is less than $4\zeta_2$.

Then $\Pr(\sup_A \left|\int_0^\zeta \Sigma\varkappa_{kn}(\xi)d\xi_k - \int_0^{\zeta_k} \Sigma\varkappa_k(\xi)d\xi_k\right| > \zeta_6) \leq$

$$\Pr(\sup_A \int_0^\zeta \Sigma \left|\varkappa_{kn}(\xi) - \varkappa_k(\xi)\right| d\xi_k > \zeta_6) \leq \Pr\left(\sup_A |\varkappa_{kn}(\xi) - \varkappa_k(\xi)| > \zeta_6/\mu(A)\right),$$

where $\mu(A)$ is the measure of the compact set A. For $\zeta_6 = \mu(A)\zeta_5$ then the probability is less than $4\zeta_2$.

Consider now on A the function $\phi_n^{-1}(\zeta) = \exp(-\int_0^\zeta \Sigma\varkappa_n(\xi)d\xi_k)$. Define $\tilde{B} = \sup_A \exp(a + b\ln\Pi_{i=1}^d \left(1 + \zeta_i^2\right))$; then $\sup_A \left|\phi^{-1}(\zeta)\right| < \tilde{B}$.

Then $\Pr(\sup_A \left|\phi_n^{-1}(\zeta) - \phi^{-1}(\zeta)\right| > \zeta_7) \leq$

$$\Pr(\sup_A \left|\int_0^\zeta \Sigma\left(\varkappa_{kn}(\xi) - \varkappa_k(\xi)\right) d\xi_k\right| > \ln(1 + \tilde{B}^{-1}\zeta_7)),$$

and is smaller than $4\zeta_2$ for $\zeta_7 = \ln(1+\tilde{B}^{-1}\zeta_6)$. Since $\sup_A \left|\phi_n^{-1}(\zeta)\varepsilon_{1n} - \phi^{-1}(\zeta)\varepsilon_1\right| <$

$$\tilde{B}\sup_A |\varepsilon_1 - \varepsilon_{1n}| + \sup_A |\varepsilon_1| \sup_A \left|\phi_n^{-1}(\zeta) - \phi^{-1}(\zeta)\right|$$

by similar derivations $\Pr\left(\sup_A \left|\phi_n^{-1}(\zeta)\varepsilon_{1n} - \phi^{-1}(\zeta)\varepsilon_1\right| > \zeta_8\right) < 5\zeta_2$ if $\zeta_8 < \min\{\tilde{B}^{-1}\zeta_1, \zeta_7(\sup_A |\varepsilon_1|)^{-1}\}$.

If $\operatorname{supp}(\gamma)$ is A, then defining γ_n as the plug-in estimator $\varepsilon_{1n}\phi_n^{-1}$ we obtain the consistent in S' estimator $Ft^{-1}(\gamma_n)$ of the function g.

If $\operatorname{supp}(\gamma)$ is non-compact some trimming is needed. Define $\tilde{\phi}_n^{-1}$ to equal $\phi_n^{-1} = \exp(-\int_0^\zeta \Sigma\varkappa_{kn}(\xi)d\xi_k)$ if $\left|\exp(-\int_0^\zeta \Sigma\varkappa_{kn}(\xi)d\xi_k)\right| < 2B(\zeta)$, where $B(\zeta) = \exp(a+b\ln\Pi_{i=1}^d\left(1+\zeta_i^2\right))$, otherwise $\tilde{\phi}_n^{-1} = \exp(a+b\ln\Pi_{i=1}^d\left(1+\zeta_i^2\right))$. For any ζ and functions $\psi_1, ...\psi_l \in S$ consider a compact A such that

$$\int_{R^d\backslash A}\left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]|\varepsilon_1(\xi)|\left|\psi_j(\xi)\right| d\xi < \zeta.$$

Then, since $\exp(-\int_0^\zeta \Sigma\varkappa_{kn}(\xi)d\xi_k)$ converges on A uniformly in probability to ϕ^{-1} and the bound on ϕ^{-1} is $B(\zeta)$ for large enough N the estimator $\tilde{\phi}_n^{-1} = \phi_n^{-1}$ on A with arbitrarily high probability and $\varepsilon_{1n}\tilde{\phi}_n^{-1}$ converges to γ on A in probability uniformly. Since ε_{1n} converges in probability in S'

to ε_1 then also $|\varepsilon_{1n}(\xi)-\varepsilon_1(\xi)|$ converges in S' to zero in probability and $\int_{R^d\backslash A}\left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]|\varepsilon_{1n}(\xi)-\varepsilon_1(\xi)|\left|\psi_j(\xi)\right|d\xi$ converges in probability to zero. Then since

$$|(\tilde{\gamma}_n-\gamma,\psi_i)|\leq\sup_A\left|\phi_n^{-1}(\zeta)\varepsilon_{1n}-\phi^{-1}(\zeta)\varepsilon_1\right|\int_A|\psi_i|$$

$$+\int_{R^d\backslash A}\left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]|\varepsilon_1(\xi)|\left|\psi_j(\xi)\right|d\xi$$
$$+\int_{R^d\backslash A}\left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]|\varepsilon_{1n}(\xi)-\varepsilon_1(\xi)|\left|\psi_j(\xi)\right|d\xi,$$

it follows that

$$\Pr(|(\tilde{\gamma}_n-\gamma,\psi_i)|>\zeta)\leq\Pr(\sup_A\left|\phi_n^{-1}(\zeta)\varepsilon_{1n}-\phi^{-1}(\zeta)\varepsilon_1\right|>\left(\int_A|\psi_i|\right)^{-1}\zeta)$$

$$+\Pr\left(\int_{R^d\backslash A}\left[\exp(a)\Pi_{i=1}^d\left(1+\xi_i^2\right)^b\right]|\varepsilon_{1n}(\xi)-\varepsilon_1(\xi)|\left|\psi_j(\xi)\right|d\xi\right)>\zeta).$$

By similar argument this implies that $\tilde{\gamma}_n$ converges in probability to $\gamma=\phi^{-1}\varepsilon_1$ in S'. Taking inverse Fourier transforms in S' concludes the proof. ■

Note that convergence in probability in S' for the sequence of estimators $Ft^{-1}(\gamma_n)$ to g in the case when g is a continuous function and $\text{supp}(\gamma)$ is bounded implies that $Ft^{-1}(\gamma_n)$ is a continuous function and thus converges to g in probability uniformly on bounded sets.

References

[1] Antosik, P., J.Mikusinski and R. Sikorski, (1973) Theory of Distributions. The Sequential approach. Elsevier-PWN, Amsterdam-Warszawa

[2] Devroye, L. (1978), The Uniform Convergence of the Nadaraya-Watson Regression Function Estimate, The Canadian Journal of Statistics, 6, pp.179-191.

[3] Chen, X., H. Hong and D.Nekipelov (2009), Nonlinear models of measurement errors, Journal of Economic Literature, under review.

[4] Cohen, A.C. Jr (1951) Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples, Annals of Mathematical Statistics, 21(4), pp. 557-569.

[5] Cunha, F., J.J. Heckman and S.M.Schennach (2010), Estimating the Technology of Congnitive and Noncognitive Skill Formation, Econometrica, v.78, 883-933.

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

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

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

[9] Hadamard, J. (1923): Lectures on Cauchy's Problem in Linear Partial Differential Equations. Yale University Press, New Heaven.

[10] Hall, P. and A.Yatchew (2007) Nonparametric Estimation When Data on Derivatives are Available, Annals of Statistics, 35:1, 300-323.

[11] Hirata, Y. and H. Ogata (1958), On the exchange formula for distributions, J.Sci. Hiroshima Univ. Ser.A, 22, pp.147-152.

[12] Kaminski, A (1982), Convolution, product and Fourier transform of distributions, Studia Mathematica, v. 74, pp.83-96.

[13] Kaminski, A and R. Rudnicki (1991), A note on the convolution and the product D' and S', Internat. J. Math&Math Sci., v.14, 275-282.

[14] Kirillov, A.A. and A.D.Gvishiani (1982), Theorems and Problems in Functional Analysis, Springer-Verlag.

[15] Klann E., M.Kuhn, D.A.Lorenz, P.Maass and H.Thiele (2007) Srinkage versus deconvolution, Inverse problems, 23, pp.2231-2248.

[16] Koralov, L. and Ya.Sinai (2007), Theory of Probability and Random Processes, Springer.

[17] Newey, W. (2001), Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models, Review of Economics and Statistics, v. 83, 616-627.

[18] Schwartz, L. (1966) "Théorie des distributions", Hermann, Paris.

[19] Schennach, S. (2007) Instrumental variable estimation in nonlinear errors-in-variables models, Econometrica, v.75, pp. 201-239.

[20] Stone, Ch.J. (1982) Optimal global rates of convergence for nonparametric regression, The Annals of Statistics, v. 10, 1040-1053.

[21] Wang, L. and Ch.Hsiao (2010), Method of Moments estimation and identifyability of semiparametic nonlinear errors-in-variables models, Journal of Econometrics, in press.

[22] Zinde-Walsh, V. (2008), Kernel Estimation when Density May not Exist, Econometric Theory, v.24, 696-725.

[23] Zinde-Walsh, V. (2009), Errors-in-variables models: a generalized functions approach, working paper, arXiv:0909.5390v1 [stat.ME], McGill University working paper.