Econometric Theory, **30**, 2014, 1207–1246.
doi:10.1017/S0266466614000140

MEASUREMENT ERROR AND DECONVOLUTION IN SPACES OF GENERALIZED FUNCTIONS

VICTORIA ZINDE-WALSH
McGill University
and
CIREQ

This paper considers convolution equations that arise from problems such as measurement error and nonparametric regression with errors in variables with independence conditions. The equations are examined in spaces of generalized functions to account for possible singularities; this makes it possible to consider densities for arbitrary and not only absolutely continuous distributions, and to operate with Fourier transforms for polynomially growing regression functions. Results are derived for identification and well-posedness in the topology of generalized functions for the deconvolution problem and for some regression models. Conditions for consistency of plug-in estimation for these models are provided.

1. INTRODUCTION

The focus of this paper is on convolution equations arising in models with measurement error. Reviews of measurement error models are in Carroll, Ruppert, Stefanski, and Crainiceanu (2006), Chen, Hong, and Nekipelov (2011), and Meister (2009). The convolution equations that are examined here also arise in other contexts; various models that go beyond measurement error models are enumerated in Zinde-Walsh (2014). This paper is devoted to the mathematical treatment of such equations.

Start with the classical measurement error where the variable of interest x^* is observed with error, u :

$$z = x^* + u. \tag{1}$$

Here the density of x^* is the function of interest, denoted g; the observed z has density w; suppose that the measurement/contamination error u is independent of

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 support and comments. The Associate Editor, Y. Kitamura, and anonymous referees provided corrections, comments, and suggestions that were extremely helpful; I am very grateful to them. Address correspondence to Victoria Zinde-Walsh, Department of Economics, McGill University, 855 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2T7; e-mail: victoria.zinde-walsh@mcgill.ca.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

x^* and has density f. Then the convolution equation

$$g * f = w \tag{2}$$

holds. When densities exist, $g * f$ denotes $\int g(z-u)f(u)du$. In problems that often arise in image processing, epidemiology, and medicine, the error density f could be assumed known, e.g., the error is Gaussian noise.

The assumption of known error distribution may not be realistic; if f is not known additional conditions are needed to identify g. If another observation, x, on x^* is available: $x = x^* + u_x$, conditions under which a unique solution exists for the density of interest, g, were given by Kotlyarski (1967) in the case of independence between x^* and u_x; the approach there was to consider the joint characteristic function of the two measurements.

The assumption of independence of the error in the second measurement may be too strong for some applications, for example, in Cunha, Heckman, and Schennach (2010) one measurement for the latent variable representing a skill of a child was constructed from test scores where one could plausibly assume independence for measurement error, but the extra measurements came from reports by teachers and parents where such an assumption could be unrealistic. Assume that for the error of the second measurement, u_x, the conditional expectation is zero: $E(u_x|x^*, u) = 0$ (but heteroscedasticity is not ruled out). Denote by x_k the $k-th$ component of $x = (x_1, \ldots, x_d)$ and by $h_k(x)$ the function $x_k g(x)$, and assume existence of density weighted conditional moments $w_{2k}(z) = E\left(w(z)x_k|z\right)$. Then in addition to equation (2) more convolution equations can be written:

$$h_k * f = w_{2k},\ k = 1, \ldots, d.$$

Indeed (with integration here and everywhere in this paper over the whole space),

$$E\left(w(z)x_k|z\right) = E\left(w(z)x_k^*|z\right) = \int (z_k - u_k)\, g(z-u) f(u)du.$$

Therefore, a system of convolution equations arises for the unknown function g and functions h_k, where $h_k(x) = x_k g(x)$,

$$\begin{aligned} &g * f = w_1; \\ &(h_k) * f = w_{2k}, \quad k = 1, \ldots d. \end{aligned} \tag{3}$$

Another model that leads to a system of equations is nonparametric regression with Berkson error (see, e.g., Meister, 2009 for review). Consider a nonparametric regression model

$$y = g(x) + u_y,$$

where x may be correlated with the error, u_y, but where some instruments, z, are available such that

$$z = x + u, \tag{4}$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

with z independent of u (Berkson error) and $E(u_y|z) = 0$. Denote the density of x by f_x, density of z by f_z, density of u by f_u, and correspondingly that of $-u$ by f_{-u}. Then equation (4) gives

$$f_x = f_z * f_{-u}. \tag{5}$$

Additionally if expectation conditional on z exists, using independence between z and u,

$$w(z) = E(y|z) = E(g(x)|z) = \int g(z-u) f_u(u) du;$$

then

$$g * f_u = w. \tag{6}$$

The system of equations (5) and (6) involves two unknown functions, f_u and g, where usually the interest is in the regression function.

The regression function could have instead of an observable argument, x, a mismeasured or latent argument, x^*. The model could provide more equations if another measurement, x, on x^* was available. Consider

$$y = g(x^*) + u_y; \tag{7}$$
$$x = x^* + u_x; \tag{8}$$
$$z = x^* + u. \tag{9}$$

Here x, y, z are observed; assume that u is a Berkson type measurement error independent of z with (unknown) density f and assume that u_y, u_x have zero conditional (on z and the other errors) expectations. For this model Newey (2001) proposed to consider an additional equation assuming that the conditional moment $E(g(x)x|z)$ exists. In the univariate case this gives a system of two equations with two unknown functions as discussed in Schennach (2007) and Zinde-Walsh (2009). Define $h(x) = xg(x)$, then

$$\begin{aligned} g * f &= w_1, \\ h * f &= w_2, \end{aligned} \tag{10}$$

with $w_1 = E(y|z)$, $w_2 = E(xy|z)$ known.

In the multivariate case for $x = (x_1, \ldots, x_d) \in R^d$, $w_{2k} = E(x_k y|z)$, $k = 1, \ldots, d$, $h_k(x) = x_k g(x)$ this can be generalized to equations (3). If the density f_{x^*} of x^* were of interest, equation $f_{x^*} = f_z * f_{-u}$ could be added.

If the independence between the mismeasured variable and measurement error holds conditionally then equation (2) can be written for densities conditional on some x_c, where if such densities exist this equation is

$$\int g((z-u)|x_c) f(u|x_c) du = w(z|x_c).$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Then equation (2) is defined for functions in spaces where the dimension or the argument is augmented by the dimension of the conditioning variable, x_c.

A common way of providing solutions to (2) and other equations is to consider them in some normed function spaces, e.g., of integrable functions such as L_p or weighted L_p spaces, e.g., Carrasco, Florens, and Renault (2007). Solving the equations is often done by employing Fourier transforms. Since convolutions and Fourier transforms can be defined in different spaces, the question is which spaces are best suited for the problems.

Devroye and Győrfi (1985) view density from the perspective of L_1 space, since the density (when it exists as a function) is absolutely integrable. However, in various problems of interest density may not exist, as in cases of measurement error for individuals answering survey questions (say, about income or consumption) where the probability of truthful reporting is nonzero and a mass point can arise (e.g., Hu, 2008). Density functions in L_1 do not necessarily converge even if the corresponding distribution functions converge uniformly. The way to overcome both the nonexistence of density and convergence problems is to consider density as a generalized derivative of the distribution function as proposed in Zinde-Walsh (2008); this is done by defining the distribution function as a functional on a suitable space of well-behaved differentiable functions so that with this definition the distribution function inherits the good properties of the well-behaved functions and becomes differentiable (details in Zinde-Walsh, 2008, also see Sect. 2.1); moreover, in the space of generalized functions the generalized densities converge if the distribution functions converge; thus the problem of defining the density for a distribution is well-posed there.

Working in spaces of generalized functions also extends the classes of regression functions for which solutions can be obtained. In various applications the object of interest may be represented by "sum of peaks" function, such as a sum of delta functions, e.g., in Klann, Kuhn, Lorenz, Maass, and Thiele (2007) where applications in astrophysics and mass spectroscopy are discussed; functions with sparse support or support that includes isolated points can arise in various applications. If g represents a regression function, applying Fourier transform to the convolution equations in spaces of functions may require severe restrictions on the function. For example, in spaces of integrable functions (such as L_1, L_2) linear and polynomial regression functions as well as distribution functions in binary choice models would be excluded. Again, a natural extension is to consider spaces of generalized functions where Fourier transforms are defined for functions that can grow at polynomial rates as well as for objects with sparse support. Spaces of generalized functions were utilized by Klann et al. (2007) for sum of peaks regression, by Zinde-Walsh (2008) for generalized density functions, and by Schennach (2007) and Zinde-Walsh (2009) for the problem in errors in variables univariate regression model with possible polynomial growth in the function.

This paper examines convolution equations in generalized function spaces. The interest here focuses on equation (2) where only the function g is unknown (deconvolution) and the system of equations (3) with two unknown functions,

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

namely g and f. The generalized functions spaces considered here are described in detail in Schwartz (1966) and Gelfand and Shilov (1964a, 1964b).

Much of the paper is devoted to a theoretical development of the problem of solving these convolution equations in spaces of generalized functions: existence of a unique solution (identification) and continuity of the mapping from the known functions to the solution (well-posedness).

The usual blueprint for deconvolution in function spaces works as follows. Assume, for example, that the functions g, f are in L_1, e.g., densities; then convolution exists: $g * f = w$. Fourier transform (Ft) is defined: for the function g, $Ft(g) = \int g(x) \exp\left(ix^T\zeta\right)dx$; this is a characteristic function if g is a density. For Fourier transforms $\gamma = Ft(g); \phi = Ft(f); \varepsilon = Ft(w)$ a convolution is transformed into the product of transforms:

$$g * f = \omega \Longrightarrow \gamma\phi = \varepsilon. \tag{11}$$

Also, if additionally Fourier transform exists for $h_k(x) = x_k g(x)$, then $Ft(h_k) = -i\frac{\partial}{\partial\xi_k}\gamma(\xi)$; denote the derivative $\frac{\partial}{\partial\xi_k}\gamma(\xi)$ by γ_k'. If w and f in equation (2) (thus also ε and ϕ) are known and $\phi \neq 0$, then solve the algebraic equation $\gamma = \phi^{-1}\varepsilon$. Then apply the inverse Fourier transform, $(Ft)^{inv}$, to obtain g,

$$g = (Ft)^{inv}(\gamma).$$

When the functions that enter the equations are estimated based on available data on the observables, the solutions will be stochastic and for establishing consistency well-posedness of the solutions becomes crucial; Carrasco et al. (2007) and An and Hu (2012) discuss well-posedness that applies to similar problems in various normed spaces, mostly in spaces of integrable functions, L_p.

Thus for pursuing a similar approach in spaces of generalized functions the following questions are addressed here: 1. Under what assumptions do the convolution equations hold? 2. When is Fourier transform and its inverse defined? 3. When does the exchange formula (11) hold (and when is a multiplicative product defined)? 4. When do transformed equations have a unique solution? 5. When is the problem well-posed, that is the solution continuously depends on the known generalized functions in the equation?

Once the conditions for identification and well-posedness are established in the generalized functions space, the question of consistent plug-in estimation can be examined. Suppose that the known functions, such as the densities (correspondingly, characteristic functions) of the observables and conditional expectations of the observables (and their Fourier transforms) in the models considered are consistently estimated; the solutions to the convolution equations based on these estimated functions are now random generalized functions; do they converge in some stochastic sense to the true function in the topology of generalized functions?

To answer this one needs to consider stochastic generalized functions, which represent stochastic functionals on the spaces of well-behaved (differentiable, etc.) functions. These are described in Gelfand and Vilenkin (1964c) and

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Koralov and Sinai (2007). The question of consistency requires providing conditions for stochastic convergence of Fourier transforms and inverse Fourier transforms, of derivatives and of products for random generalized functions—all the operations that are involved in solving for the unknown functions.

Section 2 of this paper introduces the spaces of generalized functions considered here, 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") is provided making it possible to transform convolutions into products of Fourier transforms. Section 3 gives results on existence and uniqueness of solutions of the transformed equation or system of equations. These results give conditions for identification. General results on well-posedness of the solutions are proved here in Section 3. Section 4 provides stochastic properties of generalized functions. Section 5 gives conditions for consistent (in the topology of generalized functions) deconvolution and also for consistent nonparametric estimation of a regression function in the model ((7)–(9)).

2. CONVOLUTION EQUATIONS IN GENERALIZED FUNCTIONS

2.1. Spaces of Generalized Functions

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 discussed in the books by Schwartz (1966) and Gelfand and Shilov (1964a, 1964b); reference is also made to related spaces as presented, e.g., by Sobolev (1992).

2.1.1. Spaces of Well-Behaved Functions. To introduce generalized functions first a space of well-behaved functions is defined; denote a generic such space by G. The functions in G are defined on the real space, R^d, but may take values in the complex space, since we consider characteristic functions and Fourier transforms that can take complex values. Characterization of functions as well-behaved refers to smoothness and decline at infinity. Two widely used spaces are $G = D$ and $G = S$. Both spaces D and S contain infinitely differentiable functions from the space $C^\infty(R^d)$ of all infinitely differentiable functions: $D \subset C^\infty(R^d); S \subset C^\infty(R^d)$. They differ in that D contains functions with bounded support, while support of functions in S may be unbounded, however, they decline at infinity faster than any polynomial as do their derivatives.

Recall that for a function $\psi(x)$ on R^d support is the closure of the set $W = \text{supp}(\psi)$, such that

$$\psi(x) = \begin{cases} \neq 0 & \text{for } x \in W \\ 0 & \text{for } x \in R^d \backslash W. \end{cases}$$

The set W is an open set.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

For any vector of nonnegative integers $m = (m_1, \dots m_d)$ and vector $t \in R^d$ denote by t^m the product $t_1^{m_1} \dots t_d^{m_d}$ and by ∂^m the differentiation operator $\frac{\partial^{m_1}}{\partial x_1^{m_1}} \dots \frac{\partial^{m_d}}{\partial x_d^{m_d}}$. Note that $\partial^m \psi = \psi$ for $m = (0, \dots 0)$, so that for such an m we get the function itself.

DEFINITION D. *Topological space D is the set of functions*

$$D = \left\{ \psi(t) \in C^\infty(R^d) : supp(\psi) \text{ is a bounded set in } R^d \right\};$$

where convergence is defined: a sequence $\psi_n(t) \underset{n\to\infty}{\to} \psi(t)$ in D if $W = \cup supp(\psi_n)$ is a bounded set and for any m the sequence

$$\sup_{t\in W} \left| \partial^m \psi_n(t) - \partial^m \psi(t) \right| \underset{n\to\infty}{\to} 0.$$

Thus the space D is the linear topological space of infinitely differentiable functions each defined on a compact support; to converge in D the sequence of functions should be supported on a common bounded set and converge uniformly as well as have uniformly converging derivatives of all orders.

Denote for $t \in R^d$ by $\|t\|$ the vector norm of t.

DEFINITION S. *Topological space S is the set of functions*

$$S = \left\{ \psi(t) \in C^\infty\left(R^d\right) : \text{for any vectors of nonnegative integers } m, l, \right.$$
$$\left. \left| t^m \partial^l \psi(t) \right| \underset{\|t\|\to\infty}{\to} 0 \right\};$$

where convergence is defined: a sequence $\psi_n(t) \underset{n\to\infty}{\to} \psi(t)$ in S if for any m and l the sequence $\sup_{t\in R^d} \left| t^m \left(\partial^l \psi_n(t) - \partial^l \psi(t) \right) \right| \underset{n\to\infty}{\to} 0$.

The functions in S decline fast at infinity; convergence in this space requires uniform convergence that is faster at infinity than any polynomial rate.

2.1.2. Generalized Functions. The generalized functions space G^* is the dual space, the space of linear continuous functionals on G. For any $b \in G^*$ and any $\psi \in G$ denote the value of the functional b applied to ψ by (b, ψ). The topology is defined by weak convergence: a sequence $b_n \in G^*$ converges to $b \in G^*$ if for any $\psi \in G$ the sequence of the values of the functionals converges: $(b_n, \psi) \to (b, \psi)$.

There are several alternative and equivalent definitions of generalized functions.

For example, any generalized function, $b \in G^*$, can be defined by an equivalence class $\{b_n\}$ of weakly converging sequences of functions $b_n \in G$:

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

$$b = \Bigg\{ \{b_n\} : b_n \in G, \text{ such that for any } \psi \in G, \lim_{n\to\infty} \int b_n(t)\overline{\psi(t)}dt = (b,\psi) < \infty \Bigg\}, \tag{12}$$

where $\int \cdot dt$ denotes the multivariate integral over R^d, over-bar indicates complex conjugate for complex-valued functions, and (b, ψ) provides the value of the functional $b \in G^*$ for $\psi \in G$.

Sobolev (1992) points out a subtle distinction between the functional and a generalized function as defined in (12). The same functional can be represented by sequences in (12) 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 $(\delta, \psi) = \psi(0)$; it can be represented as an equivalence class of $\delta-$convergent sequences of continuous functions as well as of functions from D or S.

The advantage of the definition as a functional is that a generalized function considered as a functional can sometimes be extended to a linear continuous functional on a wider space and such extensions will be used to facilitate proofs. In Sobolev (1992, 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 compact support and Sobolev spaces.

Note that $D \subset S$ and thus there is the inclusion of the dual spaces: $S^* \subset D^*$; convergence in S^* of linear continuous functionals implies their convergence in D^*, however, a sequence of elements of S^* that converges in D^* may not converge in the topology of S^* (see the example in Sect. 3.2).

In the terminology of Schwartz (1966) generalized functions are sometimes called "distributions" and elements of S^* "tempered distributions"; here we shall call them generalized functions indicating the specific space considered.

2.1.3. Regular and Singular Generalized Functions. A regular element in G^* corresponds to a pointwise defined function $b(t)$ such that the value of the functional (b, ψ) for any $\psi \in G$ can be obtained as

$$(b,\psi) = \int b(t)\psi(t)dt \tag{13}$$

on the space of real-valued functions or by

$$(b,\psi) = \int b(t)\overline{\psi(t)}dt; \tag{14}$$

for complex-valued functions.

Suppose that a function $b(t)$ is integrable on any bounded set; such functions are called locally summable. Any locally summable function $b(t)$ defines a regular

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

generalized function b in D^* by (14); this implies that any continuous function represents a regular element of D^*.

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

$$\int \left(1+t^2\right)^{-m} |b(t)|\, dt < \infty \tag{15}$$

for some nonnegative integer-valued vector $m = \left(m_1, \ldots, m_d\right)$ with $(1+t^2)^{-m}$ denoting the corresponding product $\Pi_{i=1}^{d}\left(1+t_i^2\right)^{-m_i}$ similarly by (14) defines a generalized function b in S^* and thus is a regular element in S^*. For a complex-valued function $|\cdot|$ denotes the modulus.

Thus a regular function in D^* is any locally summable function, but in S^* a regular function cannot grow too fast at infinity.

Any generalized function in G^* that is not regular is a singular generalized function. For example, the $\delta-$function is a singular generalized function; the value of the functional $(\delta, \psi) = \psi(0)$ for any $\psi \in G$ and cannot be represented as (13).

Although generalized functions are in general not defined pointwise support of a generalized function can be defined (see, e.g., Schwartz, 1966, p. 28).

Definition: Support of a generalized function. For a generalized function $b \in G^*$ define the null set $\Omega = \cup W$, where the open set $W \subset R^d$ is such that for any $\psi \in G$ with $\text{supp}\psi = W$ the value $(g, \psi) = 0$. Then $\text{supp}(b) = R^d \backslash \Omega$.

Note that a generalized function has support in a closed set. For example, support of the $\delta-$function is just one point 0.

2.1.4. Generalized Derivatives. An advantage of generalized functions is that despite their possible singularity they can be differentiated. Generalized derivative $\partial^l b$, for any nonnegative vector of integers $l = (l_1, \ldots l_d)$ exists for any generalized function b in $G^* = D^*$ or S^* and is defined as a generalized function that as a functional for any function, $\psi \in G$ provides

$$\left(\partial^l b, \psi\right) = (-1)^{l_1+\ldots+l_d}\left(b, \partial^l \psi\right).$$

The differentiation operator is continuous in $S^*$and in D^*.

An example of a generalized derivative is provided by a generalized probability density function. A probability distribution function, $F(x)$, is a regular function in G^* (either D^*, or S^*). To define the probability density, f, as a generalized function we need to indicate for any $\psi \in G$ the value of the functional (f, ψ); this is done by defining f as a generalized derivative of F.

For example, in the univariate case first consider an absolutely continuous distribution, F, then the density, f, is a regular function and provides for any $\psi \in G$

$$(f, \psi) = \int f(x)\, \psi(x)\, dx,$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

integrating by parts this is

$$F(x)\,\psi(x)\,|_{-\infty}^{+\infty} - \int F(x)\,\frac{d}{dx}\psi(x)\,dx,$$

where $F(x)\,\psi(x)\,|_{-\infty}^{+\infty} = \lim_{x\to+\infty} F(x)\,\psi(x) - \lim_{x\to-\infty} F(x)\,\psi(x) = 0$ by the properties of $\psi \in G$. Thus

$$(f,\psi) = -\int F(x)\frac{d\psi}{dx}(x)dx. \tag{16}$$

The right-hand side is defined as $-(F,\partial\psi)$ regardless of the existence of density and this expression is used to define the generalized density function for any distribution.

The generalized density is an element of D^*, or of S^*, moreover, it continuously depends on the distribution function in the topology of either S^* or D^* by continuity of the differentiation operator. For example, if $F(x) = I(x \geq 0)$ (the indicator function $I(\theta) = 1$ if θ is true, zero otherwise), by substituting into (16) we obtain the generalized derivative as the Dirac $\delta-$function that provides $(\delta,\psi) = \psi(0)$.

2.1.5. Generalized Functions as Generalized Derivatives of Continuous Functions; Bounded Set in S^*. An important result for generalized functions states that any generalized function can be represented as a generalized derivative of a continuous function. Thus by Schwartz (1966, p. 239, Thm. VI) any generalized function $b \in S^*$ can be represented as follows: there exists some fixed vector of nonnegative integers, l, an integer m and a bound V such that

$$b = \partial^l c, \text{ where the continuous function } c(x) \text{ is}$$
$$c(x) = \left(1+\|x\|^2\right)^{\frac{m}{2}} \tilde{c}(x) \text{ with } |\tilde{c}(x)| < V < \infty. \tag{17}$$

Of course, any $c(x)$ in (17) is a regular element in S^*.

For example, on R^1 the singular $\delta-$function can be obtained as the second derivative of the continuous function, $c(x) = xI(x \geq 0)$. Note that $c(x)$ is $\left(1+x^2\right)^{\frac{1}{2}}\tilde{c}(x)$ with $\tilde{c}(x) = \frac{xI(x\geq 0)}{\sqrt{1+x^2}}$, and $|\tilde{c}(x)| < 1$. Subsequent derivatives of the $\delta-$function can be represented as $\partial^l c$ with $l > 2$.

Denote by $S^*_{l,m}(V)$ the class of generalized functions, b, that have representation (17) for l, m, and V. The set $S^*_{l,m}(V)$ is **bounded** in S^*, moreover, any set of generalized functions bounded in the topology of S^* has the representation $S^*_{l,m}(V)$ (see Schwartz, 1966, pp. 246–247, VII,5;5). For example, the univariate $\delta-$function is in $S^*_{2,1}(1)$.

2.1.6. Fourier Transforms of Generalized Functions. The Fourier transform (Ft) is defined for functions in D, and more generally in S. The value of the

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Fourier transform for the function $\psi \in S$, $Ft(\psi)$ for any $s \in R^d$ is

$$Ft(\psi)(s) = \int \psi(x) e^{ix^T s} dx.$$

The function $Ft(\psi) \in S$. Fourier transform is an isomorphism of the space S.

In the dual spaces of generalized functions D^* and S^* the Fourier transform is given by

$$(Ft(b), \psi) = (b, Ft(\psi)).$$

Generally, the properties of Fourier transforms are related to smoothness and behavior at infinity. The more restricted the behavior of a generalized function at infinity, the smoother its Fourier transform and the smoother a function, the faster its Fourier transform converges to zero at infinity. The $\delta-$function has one point of support; this may be interpreted as the ultimate restriction on behavior at infinity. Fourier transform of the $\delta-$function is a constant, the ultimate smooth function. Similarly, Fourier transform of a constant is proportionate to a $\delta-$function.

The Fourier transform is an isomorphism of the space S^*. For the analysis of this paper we mostly consider the functions of interest in the generalized functions space S^*, because of the fact that Fourier transform represents an isomorphism, which permits to apply it or its inverse to any element, and the operator Ft and the inverse operator, $(Ft)^{inv}$, are continuous in S^*.

Assumption 1. The generalized functions in the equations (2) or (3) are elements of the space S^*.

Assumption 1 allows for any probability distribution function on R^d in the errors in variables models represented by (2) and (3) and does not require existence of density functions, for regression functions in (3) the assumption allows growth at infinity, but limits it by (15), thus binary choice or polynomially growing regression functions are included, but not functions of exponential growth. Note that exponentially growing functions are included in the space D^*.

2.2. Existence of Convolutions; Convolution Pairs

The convolution of generalized functions can be defined in different ways: see, e.g., in Schwartz (1966, p. 154), in Sobolev (1992, p. 63), and in Gelfand and Shilov (1964a, vol. I, pp. 103–104); convolution does not always have meaning and exists for specific pairs of mutual convolutors.

In addition to the spaces $C^\infty = C^\infty(R^d)$, D, S, D^*, and S^* define other spaces that are characterized by different conditions on behavior at infinity and smoothness.

The space $\mathcal{O}_M \subset C^\infty$ is the subspace of infinitely differentiable functions with every derivative growing no faster than a polynomial at infinity; it has the property that for any $\phi \in \mathcal{O}_M$ and $\psi \in S$ the product of these smooth functions, $\phi\psi$,

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

TABLE 1. The convolution table

$g \setminus f$	D	S	E^*	O_C^*	S^*	D^*
D	D	S	D	S	O_M	C_∞
S		S	S	S	O_M	X
E^*			E^*	O_C^*	S^*	D^*
O_C^*				O_C^*	S^*	X
S^*					X	X
D^*						X

belongs in S. The space E^* is the subspace of generalized functions with compact support (but no smoothness conditions). The space $\mathcal{O}_C^*$ defined in Schwartz (1966, p. 244) is the subspace of rapidly decreasing (faster than any polynomial) generalized functions without any smoothness conditions. 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) and summarizes the well-established results in the literature. Convolution is a symmetric operation and the empty entries can be filled by symmetry.

Convolution of a pair of arbitrary generalized functions is not always defined. Bounded support or at least rapid decline at infinity is needed for a convolution to exist. For example, convolution of a constant function with another constant function on R^1 is not defined. Nevertheless, there are pairs of subspaces of generalized functions beyond those in the table, for which convolution defines a generalized function; spaces where convolution is defined can be combined; the following properties are in Schwartz (1966, pp. 157–160). The convolution is a bilinear operation; 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. 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; 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.

Convolution is a bilinear operation. It is continuous in each argument (holding the other constant) but may not be continuous in both arguments in the space G^*. The problem is that unlike convergence in normed spaces there may not exist a bounded set that eventually contains a convergent sequence. **Hypocontinuity** of a bilinear operation as defined in Schwartz (1966, p. 72) is the property 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.

As shown in Schwartz (1966, pp. 247–257) the convolution pairs in the table where convolution is defined all possess the hypocontinuity property.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Definition: Convolution pair. 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^*; it is a hypocontinuous operation in the topology of G^*.

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

Assumption 2. The functions g and f in (2) are such that $g \in A \subset S^*$ and $f \in B \subset S^*$; the subspaces (A, B) form a convolution pair.

This assumption implies that (2) holds in S^*; it is often satisfied for functions arising in various models. Convolution of generalized density functions exists, thus (1) leads to (2) for all distributions 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^*(respectively, S^*). Schennach (2007) considered univariate errors in variables model with instrumental variables and a regression function, g, bounded by polynomials; these regression functions are in S^*, convolution with any generalized density function from $\mathcal{O}_C^*$ exists in S^*. For regression functions g in subspaces of S^* where growth is more restricted, convolution with less rapidly declining f may exist.

For a generalized function g denote by h_k the product of g by the kth component of $x \in R^d$, x_k, for a regular function $g(x)$ this is $h_k(x) = x_k g(x)$. In many cases when the convolution $g * f$ is defined, the convolution $h_k * f$ is also defined. Indeed, this is so in all the examples above: if g has compact support, so does any h_k; if $g \in C^\infty$, it is true of h_k as well, etc. Thus some models accommodate not only (2), but also other equations, e.g., providing (3).

Assumption 3. In addition to Assumption 2, the functions $h_k, k = 1, \ldots, d$ in (3) are such that $h_k \in A$.

2.3. Exchange Formula for Fourier Transforms and Some Special Convolution Pairs

For the generalized functions in equations (3) denote Fourier transforms as $\gamma = Ft(g)$, $\phi = Ft(f)$, and $\varepsilon_i = Ft(w_i)$. Recall that the Fourier transform always exists in S^* and is a continuous isomorphism. If Assumptions 1 and 2 are satisfied and $(g, f) \in (A, B)$, with (A, B) a convolution pair, we can consider the spaces of Fourier transform of the elements of A and B, $Ft(A)$ and $Ft(B)$ in S^*.

The classical case of the convolution pair $(S^*, \mathcal{O}_C^*)$ is examined in Schwartz (1966, pp. 281–282). If (g, f) belong to $(S^*, \mathcal{O}_C^*)$ the convolution exists and $\gamma \in S^*, \phi \in \mathcal{O}_M$.

The product between a generalized function in S^* and a function from $\mathcal{O}_M$ always exists in S^*, indeed for any $\psi \in S$ the value of the functional $(\gamma\phi, \psi)$ is defined as $(\gamma, \phi\psi)$; the definition is valid, since the product of smooth functions, $\phi\psi \in S$. This multiplication is a hypocontinuous operation (pp. 243–246).

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Then the convolution equation (2) transforms into the multiplicative equation $\gamma \phi = \varepsilon$, where $\varepsilon \in S^*$. Thus there is a dichotomous relation between the classical convolution pair $\left(S^*, \mathcal{O}_C^*\right)$ and the product pair of generalized functions spaces, $\left(Ft(S^*), Ft(\mathcal{O}_C^*)\right)$. Below we show that this dichotomy extends to other convolution pairs of spaces.

DEFINITION. *A product pair of spaces is a pair* (Γ, Φ) *of subspaces of* S^* *such that for any* $\gamma \in \Gamma, \phi \in \Phi$ *the product* $\gamma \phi$ *defines an element* ε *in* S^*; *the operation of multiplication for* (Γ, Φ) *is hypocontinuous in the topology of* S^*.

Next, we provide a result about equivalence of convolution and of products of Fourier transforms.

THEOREM 1. *If Assumptions 1 and 2 are satisfied, then the pair of spaces* $(Ft(A), Ft(B))$ *is a product pair in* S^* *and for any* $g \in A$, $f \in B$ *the exchange formula*

$$g * f = w \text{ implies and is implied by } Ft(g) \times Ft(f) = Ft(w) \tag{18}$$

holds.

Proof. Consider the special sequences studied by Antosik, Mikusinski, and Sikorski (1973) and Hirata and Ogata (1958). First define the sequence δ_n: for a number sequence: $\alpha_n > 0$ and $\alpha_n \to 0$, the regular function $\delta_n(x)$ is nonnegative with support in $|x| < \alpha_n$ and $\int \delta_n(x)dx = 1$. Then the convolution with $\delta_n \in E^*$ is always defined. Next, for a generalized function b introduce the sequence $\tilde{b}_n = b * \delta_n$. Note that $\tilde{b}_n \to b$ in S^* : indeed, $(b * \delta_n, \psi) = (b, \psi * \delta_n)$, but $\psi * \delta_n$ is in S and converges there to ψ, so $(b * \delta_n, \psi)$ converges to (b, ψ).

Since g, f belong to a convolution pair of spaces (A, B), we can enlarge the convolution pair (A, B) to include with every $b \in A$ or B the corresponding sequences $\tilde{b}_n$. Denote the resulting pair of spaces by $\left(\tilde{A}, \tilde{B}\right)$. Show that this is also a convolution pair. Note that the convolution $\tilde{g}_n * \tilde{f}_n$ is defined as support of δ_n is bounded. All that is needed is to show hypocontinuity; it follows from the fact that if a set T from A or B is bounded in S^* so is the corresponding set $\tilde{T}$ that contains every element $b \in T$ as well as all $\tilde{b}_n$. If a sequence b_m converges to zero in S^* so does the corresponding $\tilde{b}_{mn}$, and hypocontinuity extends to the enlarged convolution pair.

To show the exchange formula we need to establish that for any $(g, f) \in (A, B)$ we get that $Ft(g)\, Ft(f) = Ft(g * f)$.

Start by the exchange formula of Hirata and Ogata for the sequences:

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

Consider first the left-hand side; by the continuity of Fourier transform in S^*

$$\lim Ft\left(\tilde{g}_n * \tilde{f}_n\right) = Ft\left(\lim\left(\tilde{g}_n * \tilde{f}_n\right)\right),$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

then by hypocontinuity of the convolution and because $\tilde{f}_n \to f$, $\tilde{g}_n \to g$ we get that this is $Ft(g * f)$.

On the right-hand side by continuity of Fourier transform and convergence

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

Thus $(Ft(A), Ft(B))$ is a product pair; hypocontinuity of the product follows immediately from the hypocontinuity of the convolution and continuity of the Fourier transform and its inverse. ■

COROLLARY 1. *Given a product pair of spaces* (Γ, Φ) *in* S^* *the exchange formula applies to any* $\gamma, \phi \in (\Gamma, \Phi)$:

$$(Ft)^{inv}(\gamma\, \phi) = (Ft)^{inv}(\gamma) * (Ft)^{inv}(\phi).$$

The proof follows the proof of Theorem 1 above by replacing convolution with product and Fourier transform by the inverse Fourier transform.

In examining problems such as deconvolution much of the literature focuses on Fourier transforms, e.g., characteristic functions. If ϕ is the characteristic function of a measurement or contamination error, the condition $\phi \in \mathcal{O}_M$ for the classical product pair 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); relaxing the smoothness of ϕ will require restricting singularity of γ.

For a continuous function $\phi \in C^{(0)}$ define a space of 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). Consider $\gamma \in G^*$ that can be extended to a continuous linear functional on $G \oplus \phi G$, denote the linear space generated by such γ by $G(\phi)^*$. If $\phi \in G$, then $G(\phi)^* = G^*$. 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 $\phi\gamma$: $(\phi\gamma, \psi) = (\gamma, \phi\psi)$. Then $\left(G(\phi)^*, G \oplus \phi G\right)$ is a product pair; the corresponding spaces of inverse Fourier transforms form a convolution pair.

For example, the derivative, δ', of a univariate 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 $(\delta', \phi\psi) = \phi'(0)\psi(0) + \phi(0)\psi'(0)$. More generally, for a product between a continuous function and a generalized function to be defined, there is a trade-off between differentiability of the continuous function and singularity of the generalized function.

Under Assumptions 1–3 the convolution equations (2) and (3) lead to corresponding equations for Fourier transforms:

$$\gamma \cdot \phi = \varepsilon \tag{19}$$

or to the system of equations

$$\begin{aligned} &\gamma \cdot \phi = \varepsilon_1 \\ &\gamma_k' \cdot \phi = i\varepsilon_{2k}, \quad k = 1, \ldots, d. \end{aligned} \tag{20}$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

3. SOLUTIONS TO THE CONVOLUTION EQUATIONS: IDENTIFICATION AND WELL-POSEDNESS

3.1. Identification

For solving the equations supports of the functions play an important role. It is useful to consider generalized functions over some domain $W \subset R^d$. Denote by $D(W)$ the space of all the functions $\psi \in D$ such that closure of $\text{supp}\psi$ belongs to W; the space $D^*(W)$ is the dual space for $D(W)$.

We start with deconvolution for (2). In the deconvolution problem f, or equivalently its Fourier transform, ϕ, is assumed to be given. Typically, ϕ is a characteristic function and thus is continuous and bounded. The convolution equation (2) uniquely identifies g for a known f if it can be shown that the corresponding equation (19) has meaning and can be uniquely solved for γ.

Next a useful Lemma is proved. This Lemma shows that if a product between a generalized function γ and a continuous function ϕ is defined: $\varepsilon = \gamma\phi$, then on the domain W that represents $\text{supp}(\phi)$ division of ε by ϕ is uniquely defined in $D^*(W)$.

LEMMA 1. *Suppose that $\varepsilon = \gamma\phi$ in D^*; ϕ is a continuous function with $\text{supp}\phi = W \subseteq R^d$. Then the generalized function $\phi^{-1}\varepsilon$ is uniquely defined in $D^*(W)$.*

Proof. Consider a covering of the open set W by bounded sets: $W = \cup W_\nu$ where each W_ν is an open bounded set. Similarly, consider $D(W_\nu)$. Then any generalized function in D^* can be restricted to the dual space $D^*(W_\nu)$.

In $D^*(W_\nu)$ the generalized function γ solves $\varepsilon = \gamma\phi$. Suppose that the solution is not unique and there is $\tilde{\gamma} \neq \gamma$ such that $\tilde{\gamma}\phi = \varepsilon$. Then $(\gamma - \tilde{\gamma})\phi$ is defined and represents a zero element in $D^*(W_\nu)$. A zero functional can be extended from the space $D(W_\nu)$ to a zero functional on the space of continuous functions on W_ν, $C^{(0)}(W_\nu) \subset C^{(0)}\left(R^d\right)$. Then $((\gamma - \tilde{\gamma})\phi, \psi) = 0$ for any $\psi \in C^{(0)}(W_\nu)$, but since ϕ is invertible as a continuous function on W_ν any continuous function ψ in $C^{(0)}(W_\nu)$ has the representation $\phi\tilde{\psi}$ for some $\tilde{\psi} \in C^{(0)}(W_\nu)$. This implies that $(\gamma - \tilde{\gamma}, \psi)$ is defined (equals zero) for any ψ thus the functional $\gamma - \tilde{\gamma}$ extends to $C^{(0)}(W_\nu)$ as a zero functional. This is only possible if $\gamma - \tilde{\gamma}$ is a zero functional in $D^*(W_\nu)$. Then $\gamma - \tilde{\gamma}$ is a zero generalized function in $D^*(W)$. ■

Recall that the difference between D^* and S^* is in the "tail behavior" only; on bounded sets the two spaces coincide. The following theorem provides deconvolution in S^*.

THEOREM 2. *Under Assumptions 1 and 2 assume that $\phi = Ft(f)$ is a known continuous function; then for any W where $\text{supp}(\phi) \supset W$, the Fourier transform of g, γ, is uniquely defined on W; if it is further known that $\text{supp}(\gamma) = W$, then g is uniquely defined in S^*. Uniqueness holds automatically if $\text{supp}(\phi) = R^d$.*

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Proof. By Theorem 1, (19) holds in S^*. Since $\text{supp}(\phi) \supset W$ by Lemma 1 a unique solution γ to (19) exists in $D^*(W)$. Since γ is the Fourier transform of $g \in S^*$, $\gamma \in S^*$ and defines an element in $S^*(W) \subset D^*(W)$ uniquely. If support of γ is restricted to W a priori, then the solution is unique, thus when $\text{supp}(\phi) = R^d$, the generalized function γ is defined uniquely in S^*. An inverse Fourier transform exists in S^* for any γ. It is then possible to recover g by the inverse Fourier transform $g = (Ft)^{inv}(\gamma)$, thus g is uniquely defined whenever γ is. ■

Generally by this Theorem identification in deconvolution holds whenever $\text{supp}(\phi) = R^d$. However, this excludes some error distributions such as the uniform or the triangular, where the characteristic function has isolated zeros. A more general result in Schwartz (1966, pp. 123–125) establishes the possibility of division by ϕ even when ϕ has zeros. In the one-dimensional case as long as ϕ is infinitely differentiable, the zeros are isolated and there exists a finite order derivative that is nonzero at every zero point of ϕ, any generalized function can be divided by ϕ, the result is unique up to a generalized function with support restricted to the zeros of ϕ. Since the uniform and triangular distributions have the isolated zeros property, deconvolution with these error distributions is also identified in S^*, as long as local singularity of γ at the zero of ϕ is excluded; this would be the case if, e.g., γ is a (continuous) characteristic function.

The next Theorem examines identification in the case when there are two unknown functions in the system (3); under Assumptions 1–3 it leads to (20). The main conditions are continuous differentiability of one of γ, or ϕ, assumptions about support and knowledge of the value of the differentiable function at an interior point. If that function is a characteristic function, its value at 0 is always 1.

THEOREM 3. *Under Assumptions 1–3 for the system of equations* (20) *if* $\text{supp}(\phi) \supset \text{supp}(\gamma) = W$, *where* W *is a closed connected set in* R^d *that includes* 0 *as an interior point, and*

(a) *if* γ *is continuously differentiable in* W, $\gamma(0) = c$, *then* γ *is uniquely defined on* W *by*

$$\gamma(s) = c \exp \int_0^s \sum_{k=1}^d \varkappa_k(t) dt_k, \tag{21}$$

with the uniquely defined continuous functions $\varkappa_k$ *that solve*

$$\varkappa_k \varepsilon_1 - i\varepsilon_{2k} = 0, \quad k = 1, \ldots, d;$$

or

(b) *if* ϕ *is continuously differentiable in* W, $\phi(0) = 1$, *then* γ *is uniquely defined on* W *by*

$$\gamma = \tilde{\phi}^{-1} \varepsilon_1, \tag{22}$$

where

$$\tilde{\phi}(s) = \exp \int_0^s \sum_{k=1}^d \tilde{\varkappa}_k(t) dt_k, \tag{23}$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

with the uniquely defined on W continuous functions $\tilde{\varkappa}_k$ that solve

$$\varepsilon_1 \tilde{\varkappa}_k - \left((\varepsilon_1)'_k - i\varepsilon_{2k} \right) = 0, \quad k = 1, \ldots, d.$$

Then $g = (Ft)^{inv}(\gamma)$. If $supp(\gamma) = W$, then g is uniquely defined. Uniqueness holds automatically if $W = R^d$. If $supp(\phi) = W$, then ϕ is also uniquely defined; and so is $f = (Ft)^{inv}(\phi)$.

Proof.

(a) Consider the space of generalized functions $D^*(W)$. Since continuous γ is nonzero on W by Lemma 1 (reversing the roles of ϕ and γ there) in $D^*(W)$ the generalized function ϕ is uniquely expressed as $\gamma^{-1}\varepsilon_1$. Substitute this expression for ϕ into the differential equations in (20), and denote the continuous functions $\gamma'_k \gamma^{-1}$ by $\varkappa_k(\xi)$ to obtain equations

$$\varkappa_k \varepsilon_1 - i\varepsilon_{2k} = 0, \quad k = 1, \ldots, d, \tag{24}$$

where the left-hand side is defined and equals zero in S^*.

We can show that the function $\varkappa_k$ is uniquely determined in the class of continuous functions on W by equation (24). Proof is by contradiction. Suppose that there are two distinct continuous functions on $\text{supp}(\gamma)$, $\varkappa_{k1} \neq \varkappa_{k2}$ that satisfy (24). 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 convex domain $U \subset W$. 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 nonzero by assumption.

Next we show that γ is then uniquely determined on W. Indeed, for any $t, s \in \text{supp}(\gamma)$ with the kth coordinate denoted t_k, write the continuous function

$$\gamma(s) = c \exp \int_0^s \sum_{k=1}^d \varkappa_k(t) dt_k,$$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

where integration is along any arc joining 0 and s in W. This is the unique solution to $\gamma(0) = c$, $\gamma^{-1}\gamma_k' = \varkappa_k$ (see, e.g., in Schwartz, 1966, p. 61). Then g is defined as $g = (Ft)^{inv}(\gamma)$.

(b) In view of the result in Theorem 2 it is sufficient to show that ϕ is uniquely determined on W. Consider the space of generalized functions $D^*(W)$. Since ϕ is nonzero on W and continuously differentiable, then by differentiating the first equation in (20), substituting from the second equation and multiplying by ϕ^{-1} in $D^*(W)$ (where the product exists as shown in Lemma 1) we get that the generalized function

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

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

$$\varepsilon_1\tilde{\varkappa}_k - ((\varepsilon_1)_k' - i\varepsilon_{2k}) = 0. \tag{25}$$

We can show that the function $\tilde{\varkappa}_k$ is uniquely determined in the class of continuous functions on W by equation (25). Proof is identical to the proof in part (a) for equation (24) Next we show that ϕ is then uniquely determined on W. Indeed, the continuous function

$$\tilde{\phi}(s) = \exp\int_0^s \sum\nolimits_{k=1}^{d} \tilde{\varkappa}_k(t)dt_k$$

is the unique solution to $\tilde{\phi}(0) = 1$, $\tilde{\phi}_k^{-1}\tilde{\phi}' = \tilde{\varkappa}_k$; then since $\tilde{\varkappa}_k(= \phi_k'\phi^{-1})$ is uniquely determined on W, so is ϕ on W where it coincides with $\tilde{\phi}$.

By Theorem 2, then γ is uniquely defined on W and $g = (Ft)^{inv}(\gamma)$ is uniquely defined.

■

This Theorem extends the identification discussed in Schennach (2007) and Zinde-Walsh (2009) to the multivariate case and to the case when γ rather than ϕ is continuously differentiable (part (a) of this Theorem) and extends the identification result of Cunha et al. (2011) by showing that in the model with several measurements considered there identification additionally holds with the requirement of continuous differentiability of ϕ replacing the requirement that γ be continuously differentiable (part (b)).

3.2. Well-Posedness of the Deconvolution in S^*

Well-posedness requires that a unique solution to the problem exists and that this solution be continuous in some "reasonable topology" (Hadamard, 1923).

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

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, thus well-posedness may hold in this topology when it does not hold in the usual norms. But well-posedness may not always obtain even in this weak topology. An example is provided that involves deconvolution with the normal distribution.

Consider a convolution pair of spaces (A, B) in S^* and denote by C the space of elements in S^* represented by convolutions of generalized functions from (A, B). Well-posedness of the deconvolution would require that for the given $f \in B$ and any sequence of $w_n \in C$ that converges to some $w \in C$ in the topology of S^* : $w_n \to w$, the sequence of $g_n \in A$ such that $g_n * f = w_n$ would converge to $g \in A$, with $g * f = w$. By Theorem 1 this convergence can be restated for the product pair of spaces (Γ, Φ) where the product pair is the image of the convolution pair (A, B) under Fourier transform; denote by $\Pi \in S^*$ the space of products of elements from the pair (Γ, Φ). Then well-posedness can be restated in terms of the Fourier transforms: the problem is well-posed in S^* if for any $\varepsilon_n \to \varepsilon$ where $\varepsilon_n, \varepsilon \in \Pi$ the corresponding sequence $\gamma_n = \phi^{-1}\varepsilon_n$ converges to $\gamma = \phi^{-1}\varepsilon$.

If $\phi^{-1} \in \mathcal{O}_M$, then by hypocontinuity of the product, for any sequence $\varepsilon_n - \varepsilon$ that converges to zero in S^* the corresponding sequence $\gamma_n - \gamma$ also converges to zero and well-posedness obtains without any restrictions on γ. This result could also extend to infinitely differentiable ϕ with some zeros and applies in the univariate deconvolution to error distributions such as the uniform and triangular.

Below we provide cases of well-posed deconvolution where less restrictive differentiability conditions are imposed on ϕ; this would require additional conditions on the γ's. The nature of the conditions is to ensure that the product pair is defined for the Fourier transforms; for a continuous ϕ this requires a trade-off between possible singularity of γ (and correspondingly, ε) and the differentiability of ϕ; these trade-offs can occur locally as in the example where γ is the derivative of a $\delta-$function and ϕ is continuously differentiable at 0, but to streamline the proofs we consider global restrictions in the product pairs.

Here deconvolution is proved for ϕ supported on the whole space R^d. This assumption is common in the literature; various ways of relaxing this assumption are possible; they would require additional assumptions on the zeros of ϕ or on the relative supports of γ and ϕ.

First, suppose that both ϕ and γ are continuous functions, then ε is continuous as well. Let Γ be a subspace of all continuous functions on R^d such that all $\gamma \in \Gamma$ belong in a bounded set in S^*; then for some $\bar{m} \geq 0$ we have $\Gamma = S^*_{0,\bar{m}}(V)$ (implying $\left(1 + \|t\|^2\right)^{\frac{\bar{m}}{2}} |\gamma| < V < \infty$). Then for any bounded continuous ϕ the product $\varepsilon = \gamma\phi \in S^*_{0,\bar{m}}(V)$.

LEMMA 2. *Suppose that $\gamma, \gamma_n \in S^*_{0,\bar{m}}(V)$, ϕ is a bounded continuous function with support R^d; $\varepsilon_n = \gamma_n\phi, \varepsilon = \varepsilon_0 = \gamma\phi, 0 \leq n < \infty$ and ϕ^{-1} is a continuous*

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

function that satisfies (15). Then if $\varepsilon_n \to \varepsilon_0$ *as* $n \to \infty$ *in* S^* *we get that* $\gamma_n = \phi^{-1}\varepsilon_n$ *converges to* $\gamma = \phi^{-1}\varepsilon_0$.

Proof. First note that for any number $\xi > 0$ for any $\psi \in S$ there exists some bounded set $\Lambda \in R^d$ such that

$$\left| \int_{R^d \setminus \Lambda} \varepsilon_n(t)\phi^{-1}(t)\psi(t)dt \right| < \xi.$$

Indeed, using the fact that $\varepsilon \in S^*_{0,\bar{m}}(V)$ and *(15) for* ϕ^{-1} we can bound $\left| \int_{R^d \setminus \Lambda} \varepsilon_n(t)\phi^{-1}(t)\psi(t)dt \right| \leq$

$$V \int_{R^d \setminus \Lambda} \left(1 + \|t\|^2\right)^{-\frac{\bar{m}}{2}} \left(1 + t^2\right)^{-m} \left|\phi^{-1}(t)\right| \left(1 + t^2\right)^{m} \left|\psi(t)\right| dt$$

$$\leq V \sup_{R^d \setminus \Lambda} \left(1 + \|t\|^2\right)^{-\frac{\bar{m}}{2}} \left(1 + t^2\right)^{m} |\psi(t)| \left| \int_{R^d} \left(1 + t^2\right)^{-m} \left|\phi^{-1}(t)\right| dt.$$

Since by (15) $\int_{R^d} \left(\left(1 + t^2\right)^{-1}\right)^{-m} |\phi^{-1}(t)|dt$ is bounded, say, by some V_ϕ and for $\psi \in S$ as $t \to \infty$ the value $\left(1 + \|t\|^2\right)^{-\frac{\bar{m}}{2}} \left(1 + t^2\right)^{m} |\psi(t)|$ converges to zero by the properties of $\psi \in S$, the set Λ can be selected such that

$$\sup_{R^d \setminus \Lambda} \left(1 + \|t\|^2\right)^{-\frac{\bar{m}}{2}} \left(1 + t^2\right)^{m} |\psi(t)| < \xi V_\phi^{-1} V^{-1}.$$

Define $\zeta = 2\xi V_\phi^{-1} V^{-1}$. Continuous function ϕ^{-1} is bounded on compact set Λ; denoting by ϕ_Λ^{-1} the function ϕ^{-1} restricted to Λ we have that by hypocontinuity $\left|\left((\varepsilon_n - \varepsilon)\phi_\Lambda^{-1}, \psi\right)\right|$ converges to zero, so that for $\zeta/2$ there is an N, such that for any $n > N$

$$\left|\left((\varepsilon_n - \varepsilon)\phi_\Lambda^{-1}, \psi\right)\right| < \zeta/2.$$

It follows that then $\left|\left(\phi^{-1}\varepsilon_n - \phi^{-1}\varepsilon, \psi\right)\right| < \zeta$.

Thus $|\gamma_n - \gamma| \to 0$ in S^*. ■

In the measurement error problem the function ε is a characteristic function; if all ε_n in the equation (19) are characteristic functions or at least some continuous and uniformly bounded functions, then they satisfy the conditions of Lemma 2 with $\bar{m} = 0$.

Usually well-posedness in deconvolution is examined in terms of ordinary smooth and supersmooth probability distributions. An ordinary smooth distribution with the characteristic function $\phi(t)$ has the property that $\phi(t)\|t\|^m$, for some $m > 0$ as $t \to \infty$ is bounded from below by a nonzero constant for a supersmooth distribution $\phi(t)\|t\|^{-m}\exp\left(\|t\|^{\frac{\lambda}{\mu}}\right)$ for any $m, \lambda > 1, \mu > 0$ is similarly

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

bounded from below (e.g., the Gaussian distribution is supersmooth, but not ordinary smooth). When the distribution is ordinary smooth the characteristic function declines no faster than at some polynomial rate at infinity and condition (15) holds; it is not satisfied for supersmooth distributions making deconvolution for that case more difficult.

An and Hu (2011) demonstrate that well-posedness holds automatically in a measurement error problem, provided there is a mass at zero. This may occur if there is a reporting error but with some nonzero probability some values are truthfully reported. Indeed, if there is a mass at zero the error distribution is a mixture with the delta-function that has Fourier transform equal to 1, then ϕ is separated from zero and condition (15) for ϕ^{-1} holds.

The Lemma demonstrates that unlike the behavior in function spaces where the deconvolution is usually ill-posed, in the weaker topology of the space S^* deconvolution is well-posed when the error distribution is such that the condition (15) is satisfied (e.g., for any ordinary smooth distributions as well as for distributions with mass points).

A further extension provides deconvolution when the generalized function γ is not necessarily a continuous function and can be singular. Then additional differentiability conditions are needed to ensure that γ and ϕ are in a product pair. Let $\Gamma = S^*_{l,m}(V)$ with $l = (l_1, \ldots, l_d)$; ϕ be a bounded continuous function that has continuous derivatives of any order $\leq l_1 + \ldots + l_d$; then a product $\varepsilon = \gamma\phi$ exists in S^* for any $\gamma \in \Gamma$.

THEOREM 4. *Assume $\gamma, \gamma_n \in S^*_{l,m}(V)$, and ϕ is a continuous regular function with support R^d that has continuous derivatives ∂^{l_i} for all subvectors l_i of l; $\varepsilon = \gamma\phi$, $\varepsilon_n = \gamma_n\phi$ for $n = 1, 2, \ldots$; and $\varepsilon_n - \varepsilon \to 0$ in S^*. If ϕ^{-1} and all $\partial^{l_i}\left(\phi^{-1}\right)$ for any subvector l_i of each l satisfy (15), then $\varepsilon_n\phi^{-1}$ exists in S^* and $(Ft)^{inv}\left(\varepsilon_n\phi^{-1}\right) \to g$ in S^*.*

Proof. Denote by c_n and c the continuous functions for which ε_n and ε can be defined via the operator ∂^l. Consider the functional $\left(\phi^{-1}\varepsilon_n - \phi^{-1}\varepsilon, \psi\right) = \left(\phi^{-1}\partial^l\left(c_n - c\right), \psi\right)$; it can be extended to the functional

$$\left(c_n - c, (-1)^{|l|}\partial^l\left(\phi^{-1}\psi\right)\right) = (-1)^{|l|}\sum_{(l_1;l_2)=l}\kappa\left(l_1, l_2\right)\left(c_n - c, \partial^{l_1}\left(\phi^{-1}\right)\partial^{l_2}\psi\right),$$

where $\kappa\left(l_1, l_2\right)$ is the corresponding integer coefficient arising from differentiation of the product. All the functionals $\left(c_n - c, \partial^{l_1}\left(\phi^{-1}\right)\partial^{l_2}\psi\right)$ that enter into the linear combinations on the right-hand side satisfy the conditions of Lemma 2 and thus converge to zero. Therefore, $\gamma_n - \gamma = (\varepsilon_n - \varepsilon)\phi^{-1} \to 0$ in S^*.

By continuity in S^* of the inverse Fourier transform the limit

$$(Ft)^{inv}\left(\varepsilon_n\phi^{-1}\right) \to g$$

in S^* follows. ■

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

If the measurement error model holds conditionally on some d_c-dimensional x_c, assume that all the distribution functions are defined on $R^d \times R^{d_c}$ as $F(\cdot, x_c)$. For any generalized function b on $R^{d_x} \times R^{d_c}$ the partial Fourier transform $Ft_{|x_c}$ is defined as follows: for $\psi(x, x_c) \in S$ with $x \in R^{d_x}, x_c \in R^{d_c}$ define $Ft_{|x_c}(\psi)(s, x_c) \in S$ by $\int e^{ix^T s}\psi(x, x_c)dx$, then $\left(Ft_{|x_c}b, \psi\right) = \left(b, Ft_{|x_c}(\psi)\right)$. Consider the following possibilities: 1. the generalized functions $\gamma, \phi, \varepsilon$ denote now the partial Fourier transforms of the distribution functions; 2. instead of distribution functions consider generalized density functions and denote by $\gamma, \phi, \varepsilon$ the corresponding partial Fourier transforms; 3. assume continuous differentiability of the probability distribution functions with respect to x_c and consider conditional probability or, correspondingly, conditional generalized densities; additional conditions will be needed to make sure that dividing by the density of x_c is possible in S^*. In all these cases (19) holds. To avoid imposing extra constraints here examine case 2. For deconvolution ϕ is assumed known and often ϕ does not vary with x_c. The conditions of Lemma 2 may not apply as the functions $\varepsilon, \varepsilon_n$ may no longer be continuous, however, since a probability distribution is a bounded monotone function, ε is in some set $S^*_{1,d_c}(V)$ and thus Theorem 4 requires only that all ε_n belong to $S^*_{1,d_c}(V)$, that (15) is satisfied for ϕ^{-1} and that ϕ be differentiable in components of x.

The condition (15) for ϕ^{-1} is necessary for well-posedness: the example below shows that even in the weak topology of generalized functions well-posedness does not hold if it is not satisfied, e.g., for deconvolution with supersmooth distributions such as the normal.

Example. Consider the function $\phi(x) = e^{-x^2}$, $x \in R$, then ϕ^{-1} does not satisfy (15). Then there exists a sequence of functions γ_n and a γ in S^* such that $\gamma_n\phi \to \gamma\phi$ in S^*, but γ_n does not converge to γ in S^*.

Define the function $b_n(x) = \phi^{-1}(x)I\left(n - \frac{1}{n} < x \le n + \frac{1}{n}\right)$; then $b_n \in S^*$. Then define $\gamma_n = \gamma + b_n$ for some fixed $\gamma \in S^*$. Define $\varepsilon_n = \gamma_n\phi$ and $\varepsilon = \gamma\phi$; since $\phi \in S$, the products are always defined. Consider the difference $\varepsilon_n - \varepsilon$; it equals $I\left(n - \frac{1}{n} < x \le n + \frac{1}{n}\right)$. In S^* this sequence converges to zero. On the other hand, $(\gamma_n - \gamma, \psi)$ for any $\psi \in S$ equals

$$\int_{n-2/n}^{n+2/n} e^{x^2}\psi(x)dx.$$

Select $\psi \in S$ given by $\psi(x) = \exp(-|x|)$; then $(\gamma_n - \gamma, \psi) =$

$$\int_{n-2/n}^{n+2/n} e^{x^2}\psi(x)dx \ge \int_{n-1/n}^{n+1/n} e^{x^2-x}dx \ge \frac{1}{2n}e^{-(n+\frac{1}{n})+\left(n-\frac{1}{n}\right)^2}.$$

Thus for this ψ the values of the functional, $(\gamma_n - \gamma, \psi)$ diverge as $n \to \infty$, and so $\gamma_n - \gamma$ does not converge to zero in S^*.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

3.3. Well-Posedness in S^* of the Solution to the System of Equations (20)

In the system of equations (20) when the solution exists according to Theorem 3, say Theorem 3 (b) with continuous ϕ, well-posedness can be established over a set of functions ϕ that is a bounded set in S^*. Additionally for well-posedness, the set of the inverses, ϕ^{-1}, needs to be bounded in S^*. Compared to just deconvolution, this requirement will imply that (15) applies in the set of unknown ϕ uniformly. The next Theorem considers well-posedness of the problem in (3) (and correspondingly, (20)) under the identification conditions of Theorem 3(b). The conditions of Theorem 3(a) can be similarly considered.

THEOREM 5. *Suppose that* $(\gamma_n, \phi_n), n = 1, 2, \ldots$ *and* (γ, ϕ) *belong to a product pair* (Γ, Φ). *Additionally, let the conditions of Theorem 3(b) apply to each pair* (γ_n, ϕ_n). *Suppose that* $\varepsilon_{1n} - \varepsilon_1 \to 0$ *and* $\varepsilon_{2kn} - \varepsilon_{2k} \to 0$ *in* S^*; *the continuous functions* ϕ, ϕ_n *as well as* ϕ^{-1}, ϕ_n^{-1} *restricted to* W *all belong to some* $S^*_{0,m}(V)$. *Then if* $(\Gamma, \Phi) \equiv (S^*, \mathcal{O}_M)$, *the products* $\varepsilon_n \phi_n^{-1}$ *exist in* S^* *and* $(Ft)^{inv}\left(\varepsilon_{1n}\phi_n^{-1}\right) \to g$ *in* S^*.

Proof. We have that $\phi, \phi_n \in \mathcal{O}_M$. It follows that $(\phi)'_k, (\phi_n)'_k \in \mathcal{O}_M$. Also $(\phi)'_k, (\phi_n)'_k \in S^*_{0,m'}(V)$, where $m' = m + 1$. 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 and function spaces restricted to W, even when W does not coincide with R^d, but suppress dependence on W in the notation. The functions belong also to $S^*_{0,\tilde{m}}(V)$, where $\tilde{m} = m + m'$. Without loss of generality assume that each $\varkappa_k$ is also in the same $S^*_{0,\tilde{m}}(V)$, and so all $\varkappa_{kn}, \varkappa_k$ are in a bounded set in S^*. Since from condition (a) it follows that $\varkappa_{kn} = (\phi_n)'_k \phi_n^{-1} \in \mathcal{O}_M$, products are defined and 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. 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. Recall that all ϕ_n, ϕ satisfy (23) for the appropriate $\varkappa$'s. It then follows that $\phi_n - \phi \to 0$ in S^*. Moreover,

$$\phi_\cdot^{-1}(s) = \exp\left(-\int_0^s \sum\nolimits_{k=1}^d \varkappa_{k\cdot}(t)dt_k\right),$$

thus similarly, $\phi_n^{-1} - \phi^{-1}$ converges to zero in S^*.

Consider $\varepsilon_{1n}\phi_n^{-1} - \varepsilon_1\phi^{-1} = \varepsilon_{1n}\left(\phi_n^{-1} - \phi^{-1}\right) + \left(\varepsilon_{1n} - \varepsilon_1\right)\phi^{-1}$; 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^*, $g_n = (Ft)^{inv}(\varepsilon_{1n}\phi_n^{-1}) \to g$ in S^*. ■

Thus well-posedness obtains for the solution to the errors in variables regression with Berkson type instruments that provides (3) as long as ϕ^{-1} and all the

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

ϕ_n^{-1} are all in bounded sets. E.g., a set of the characteristic functions would satisfy this if the growth of all ϕ_n^{-1} is bounded by the same order polynomial. Without restricting $\gamma \in S^*$ the Theorem holds if the supports of ϕ, ϕ_n are bounded or if ϕ's are Fourier transforms of functions with some singularity points (e.g., from distributions with mass points), when ϕ would include a constant. Imposing restrictions on degree of singularity of γ and relaxing differentiability conditions on ϕ and ϕ_n to those imposed in Theorem 4 to ensure that all the products exist would extend the well-posedness result. This result did not require any restrictions on γ and thus on the regression function in the model beyond belonging to S^*. Under Theorem 3(a) a similar result holds after imposing the conditions on γ, γ_n rather than ϕ, ϕ_n to ensure that generalized functions $\varkappa_{kn} = (\gamma_n)_k' \gamma_n^{-1}$ belong to $\mathcal{O}_M$ and to some $S^*_{0,\tilde{m}}(V)$.

For example, if as in Cunha et al. (2010) both γ and ϕ are characteristic functions and thus continuous and bounded, products exist and Theorem 5(b) applies, as long as the conditions of Theorem 3 are satisfied and under either Theorem 3(a) for γ_n^{-1}, or under 3(b) for ϕ_n^{-1} the functions are in some bounded set $S^*_{0,m}(V)$, well-posedness holds.

The implications that well-posedness has for estimation are two-fold. One is that unless well-posedness holds, that is the inverse mapping from the class of the known functions into the class of identified solutions is continuous, the solutions corresponding to the consistently estimated known functions will not in general provide consistent estimators for the solutions. The other is that in a well-posed problem consistent estimation of the known functions automatically gives rise to consistency of plug-in estimators of solutions; the next section examines consistent estimation of the deconvolution and the solution to the system (3) in the space S^*.

4. RANDOM GENERALIZED FUNCTIONS AND STOCHASTIC CONVERGENCE

This section examines stochastic convergence of the solutions to the convolution equation (2) and to (3), equivalently to (19) and (20). If some consistent 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 for plug-in estimators of g as long as well-posedness holds. Since generalized functions in any space G^* are represented as linear continuous functionals on G, results for random functionals are applicable here, the specific feature is that for any generalized function $b \in G^*$ one needs to consider the collection of random functionals indexed by the functions from the space G.

4.1. Random Generalized Functions

Following Gelfand and Vilenkin (1964c) or Koralov and Sinai (2007) who consider specifically S^* define random generalized functions as random linear

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

continuous functionals on the space of test functions. In particular, any random generalized function $\tilde{b}$ on G is represented by a collection of (complex-valued) random variables on a common probability space that are indexed by $\psi \in G$, denoted $(\tilde{b}, \psi)$, such that

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

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

$$\left(\left(\tilde{b}, \psi_{1n}\right) \ldots \left(\tilde{b}, \psi_{mn}\right)\right) \to_d \left(\left(\tilde{b}, \psi_1\right) \ldots \left(\tilde{b}, \psi_m\right)\right),$$

where $\to_d$ denotes convergence in distribution.

As shown in Gelfand and Vilenkin (1964c), equivalently, there exists a probability measure on G^* such that for any set $\psi_1, \ldots \psi_m \in G$ the random vectors $\left((b, \psi_1), \ldots (b, \psi_m)\right)$ have the same distribution as for some random functional $\tilde{b}$, $\left((\tilde{b}, \psi_1) \ldots (\tilde{b}, \psi_m)\right)$. An example is a generalized Gaussian process b, so defined if for any $\psi_1, \ldots, \psi_m$ the joint distribution of $(b, \psi_1), \ldots, (b, \psi_m)$ is Gaussian. A generalized Gaussian process is uniquely determined by its mean functional, $\mu_b : \ (\mu_b, \psi) = E(b, \psi)$, and the covariance bilinear functional, $B_b(\psi_i, \psi_j) = E\left((b, \psi_i)\left(b, \psi_j\right)\right)$.

Gelfand and Vilenkin (1964c, p. 260) give the covariance functional of the generalized derivative, W', of the Wiener process as

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

The Fourier transform of a Gaussian random process b is also a Gaussian random process with covariance functional

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

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

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

The mean functional is zero for W' and $Ft(W')$.

Gelfand and Vilenkin (1964c) provide definitions and results for generalized random functions in $G^* = D^*$, 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 bounded support, leading to space D_k^* of random linear continuous functionals on 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: $b_n \to_p 0$ in G^*, or almost surely: $b_n \to_{a.s.} 0$ in G^*

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

if for any finite set $\psi_1, \dots \psi_v \in G$ the random vectors $((b_n, \psi_1), \dots (b_n, \psi_v)) \to_p (0, \dots, 0)$ or correspondingly, $\Pr(((b_n, \psi_1), \dots (b_n, \psi_v)) \to (0, \dots, 0)) = 1$.

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 $((b_n, \psi_1), \dots (b_n, \psi_v)) \to_d ((b, \psi_1) \dots (b, \psi_v))$ for any set $\psi_1, \dots, \psi_v \in G$.

Remark 1. (a) If $b_n - b \to_p 0$ in S^*, then $Ft(b_n) - Ft(b) \to_p 0$ in S^* and $(Ft)^{inv}(b_n) - (Ft)^{inv}(b) \to_p 0$ in S^*. Indeed, for any set $\psi_1, \dots \psi_v \in S$

$$((Ft(b_n) - Ft(b), \psi_1), \dots, (Ft(b_n) - Ft(b), \psi_v)) = ((b_n - b, Ft(\psi_1)), \dots, (b_n - b, Ft(\psi_v)))$$

and since the set $Ft(\psi_1), \dots, Ft(\psi_v) \in S$, then $((b_n - b, Ft(\psi_1)), \dots, (b_n - b, Ft(\psi_v))) \to_p 0$. Similarly for $(Ft)^{inv}$.

(b) If $\mu \in G$, then the product with any element $b \in G^*$ is defined: for any $\psi \in G$ define $(b\mu, \psi)$ by $(b, \mu\psi)$. Then if $b_n - b \to_p 0$ in G^*, for the product convergence $b_n\mu - b\mu \to_p 0$ holds 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.

5. CONSISTENT ESTIMATION OF SOLUTIONS TO STOCHASTIC EQUATIONS

Suppose that the known functions, w or w_1, w_{2k} (equivalently, ε or $\varepsilon_1, \varepsilon_{2k}$) are consistently estimated in S^*. In the models discussed here these functions are density functions or conditional mean functions; typically these are ordinary functions for which commonly used nonparametric estimators are shown to be consistent pointwise or in some norm (uniform or L_1, say) under some sufficient conditions. If the functions and the estimators can be represented as elements in S^*, then convergence in many common norms implies convergence in the weaker topology of S^*, so that such consistent estimators are consistent in S^* and the following discussion of consistency of the solutions to convolution equations considered here applies. However, consistency in S^* applies more widely than the usual consistency results. For example, as shown in Zinde-Walsh (2008), kernel estimators of density converge as stochastic generalized functions even when they diverge pointwise (e.g., at mass points, or for fractal measures).

For the equations considered here, when identification and well-posedness holds consistent estimation of the known functions can provide consistent plug-in estimators for the solutions as long as the estimators all be in the classes over which well-posedness obtains.

The fact that well-posedness implies consistency of some plug-in estimators holds generally. Denote (generically) the known functions by $w_1, \dots w_M$, their consistent in topology of S^* estimators by $\hat{w}_i$; denote the solutions by $g_1, \dots, g_K$ and the plug-in estimators from solving the equations using $\hat{w}_i$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

by $\hat{g}_j$. If the estimators, $\hat{w}_i$, are in classes where well-posedness holds, then for any set of $(\psi_1, \ldots \psi_v)$ from S and any neighborhood $N_{vK}(0)$ of zero in real or complex $vK-$ dimensional Euclidean space there is a neighborhood of zero, $\tilde{N}_{vM}(0)$ in the corresponding $vM-$dimensional space such that event $E_w = \left\{\left(\left(\hat{w}_1, \psi_1\right), \ldots, \left(\hat{w}_1, \psi_v\right), \left(\hat{w}_2, \psi_1\right), \ldots, \left(\hat{w}_M, \psi_v\right)\right)' \in \tilde{N}_{vM}(0)\right\}$ by well-posedness implies the event

$$E_g = \left\{\left(\left(\hat{g}_1, \psi_1\right), \ldots, \left(\hat{g}_1, \psi_v\right), \left(\hat{g}_2, \psi_1\right), \ldots, \left(\hat{g}_K, \psi_v\right)\right)' \in N_{vK}(0)\right\}.$$

Consistency of $\hat{w}_i$ as $n \to \infty$ means that for large enough n probability of E_w can be arbitrarily close to 1, thus probability of E_g is as close or closer to 1. Thus the condition for consistent plug-in estimation is that the estimators are in the classes of generalized functions that provide well-posedness of solutions with probability approaching 1.

5.1. Consistency of Deconvolution: Solving (19)

The well-posedness condition directly applies to the measurement error deconvolution problem where ϕ is known and the conditions apply only to estimators of observables. The requirement of boundedness of the sequence of the estimators in the space S^* would follow if the norms of the functions were bounded in probability; L_p are the spaces typically used in the literature (e.g., Fan, 1991; Carrasco and Florens, 2010). But here more generally consistency of plug-in deconvolution holds for any probability distribution; the only important restriction is on ϕ^{-1} to satisfy (15); that excludes supersmooth measurement error.

Moreover, the functions in (2) and the corresponding (19) need not be generalized densities and characteristic functions; the conditions for consistent estimation follow from those for well-posedness and are essentially that the two functions belong to a convolution pair (equivalently, Fourier transforms to a product pair) and that the known function be continuous, supported on R^d and ϕ^{-1} satisfy (15).

For example, consider in the univariate case a generalized density function, w. One can use empirical distribution functions to estimate the corresponding distribution function, then the estimator of generalized density by generalized derivatives provides as an estimator the sum of delta-functions: $\hat{w}_n = \frac{1}{n}\sum_{j=1}^{n}\delta(x_j)$, where $\left(\delta(x_j), \psi\right) = \psi(x_j)$. Then the corresponding Fourier transform $\hat{\varepsilon}_n = Ft(\hat{w}_n)$ is given by a continuous function $\hat{\varepsilon}_n(s) = \frac{1}{n}\sum_{j=1}^{n} e^{is^T x_j}$. Deconvolution using the deconvolution kernel as, e.g., in An and Hu (2012) is applied when the density of the mismeasured variable is assumed to exist and belong in some L_p, the deconvolution kernel incorporates spectral cut-off by employing estimators of the form $\tilde{\varepsilon}_n = \hat{\varepsilon}_n(s) I(|s| < T_n)$, where T_n goes to ∞ as $n \to \infty$ at some rate; the indicator function allows for smoothing in the inverse Fourier transform.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

By the results here consistency in S^* is ensured for both estimators. Indeed, since empirical distribution converges uniformly in probability to the distribution function, it thus converges in S^*, then the generalized derivatives also converge in S^* and by continuity of the Fourier transform estimators $\hat{\varepsilon}_n$ and for $T_n \to \infty$ also $\tilde{\varepsilon}_n$ converge to ε in probability. All these functions are continuous and bounded and only (15) for ϕ^{-1} is needed for consistency of the deconvolution solution in S^* and the spectral cut-off is not required. The following remark summarizes this result.

Remark 2. If the functions in equation (2) are all generalized density functions, the known characteristic function ϕ is such that ϕ^{-1} satisfies (15), $\{z_i\}_{i=1}^n$ is a random sample from the distribution with the generalized density w, the deconvolution estimator

$$\hat{g}_n = (Ft)^{inv}\left(\phi^{-1}\hat{\varepsilon}_n\right) \text{ where } \hat{\varepsilon}_n(s) = \frac{1}{n}\sum_{j=1}^{n} e^{is^T z_j}$$

is consistent in S^* : for any finite set $\psi_1, \ldots \psi_v \in S$ the random vector $\left((\hat{g}_n - g, \psi_1), \ldots, (\hat{g}_n - g, \psi_v)\right)^T \to_p 0$.

Regularization with suitable spectral cut-off in the case of supersmooth error distributions typically provides a sequence of estimators that will converge in norm, albeit slowly for supersmooth distributions as demonstrated by Fan (1991); convergence to the limit in the normed space implies convergence in the topology of S^*. However, if ϕ is supersmooth the class of generalized functions where consistent estimation holds does not include all probability distributions. Indeed, spectral cut-off estimation provides generalized densities that are a limit of inverse Fourier transforms of truncated generalized functions. Schwartz (1964, pp. 271–273) gives a characterization of any function in S^* with Fourier transform that has compact support (in a cube $|z_k| < C, k = 1, \ldots d$) based on Wiener–Paley theorem. Such a function is a continuous function g that can be extended to a entire analytic function G of a complex argument and is of exponential type $\leq 2C$, meaning

$$\lim_{|z|\to\infty} \frac{log|G(z)|}{|z_1| + \ldots + |z_d|} \leq 2C.$$

Thus as long as g is such a function or a limit of such functions it can be expressed via the regularized solution. As in Schwartz the subspace of all functions of exponential type (for any finite C) can also be considered. However, if g does not belong to this subspace regularized solutions may not converge to it even in S^*.

5.2. Consistency of Estimated Solutions to the System of Equations (3) and (20)

Establishing consistency for the system of equations is more complicated, since consistency requires not only conditions on estimators of known functions of the

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

observables, $\varepsilon.$, but also verification that the resulting estimators of ϕ satisfy the conditions of well-posedness in Theorem 5.

The next Theorem gives conditions for a consistent plug-in nonparametric estimator for the model ((7)–(9)) that leads to the system (3) and thus (20) with continuously differentiable ϕ.

Semiparametric generalized method of moments estimation of this model for a class of regression functions that includes functions in $L_1(R^d)$ was proposed by Wang and Hsiao (2011); semiparametric estimation was also discussed for somewhat different classes of univariate parametric regression functions in Schennach (2007) and Zinde-Walsh (2009). Semiparametric estimation of polynomial regression is in Hausman, Newey, Ichimura, and Powell (1991).

Start by formally stating the assumptions.

Assumption 4 (model)**.** In the model $(7-9)$ in R^d the moments of the errors $E(u_y|z,u,u_x)=0$; $E(u_x|z,v)=0$; z is independent of u; g is a generalized function in S^*.

This assumption implies that if the function g is a regular locally summable function, it cannot grow at infinity at a rate faster than some polynomial rate. This assumption does not exclude the possibility that g has some singularities, for example, it could be a sum of $\delta-$functions or a mix of a regular function and some peaks represented by δ-functions.

Denote by f_z the density of z. Recall that f denoted the generalized density of u and $\gamma = Ft(g); \phi = Ft(f), \varepsilon. = Ft(w.)$.

Assumption 5 (support)**.**

(a) supp(γ) is a connected set that includes 0 as an interior point;
(b) supp$(\phi) = R^d$.
(c) supp$(f_z) = R^d$.

Assumption 5(a) would not be satisfied by a polynomial regression function when support of γ consists of one isolated point 0; semiparametric estimation for that case was provided in Hausman et al. (1991). Support assumptions (b) and (c) are standard; they could be relaxed.

Recall that here ϕ is a characteristic function, thus $\phi(0)=1$.

Assumption 6 (generalized functions)**.**

(a) The function f_z is continuous; ϕ belongs to $\mathcal{O}_M$.
(b) The continuous functions $w.$ belong to a bounded set in S^*, $S^*_{0,m}(V)$ (as do their Fourier transforms, $\varepsilon.$); also, ϕ'_k for $k=1,\ldots,d$, as well as ϕ^{-1} and f_z^{-1} all belong to $S^*_{0,m}(V)$.
(c) The regression function g is absolutely integrable.
(c') The regression function g can be represented as a sum $g = g_{L1} + g_g$,

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

where $g_{L1}$is absolutely integrable and g_g is such that the Fourier transform of g_g in S^*, $\gamma_g = Ft(g_g)$, is singular and has support set Λ_g that is compact and of zero Lebesgue measure; Λ_g is a proper subset of supp(γ). The generalized function γ_g is such that there exists a deterministic sequence of regular functions, $(\gamma_g)_n$ that converges to γ_g in S^* and such that support of $(\gamma_g)_n$ is in a compact set Λ_{gn}; there exists $\zeta_0 > 0$ such that $|(\gamma_g)_n| > 2\zeta_0$ and for $\gamma_{L1} = Ft(g_{L1})$ on Λ_g $|\gamma_{L1} - (\gamma_g)_n| > \zeta_0$.

Assumptions 6(a) and 6(b) imply that $w.$ can be divided by f_z and any generalized function can be divided by ϕ. Requiring that ϕ be in $\mathcal{O}_M$ is sufficient to ensure that the model leads to equations (3) in S^*. More detailed conditions similar to those employed in Theorem 2 would allow relaxing the infinite differentiability Assumption 6(a). In particular, if γ is a characteristic function, Assumption 6(a) for ϕ is not needed.

Assumption 6(b) on $w.$ implies that the Fourier transform is in the bounded set as well. Continuity of $w.$ in 6(b) would follow by properties of convolution if either g were continuously differentiable, or f were continuous.

In (b) using the same bound on growth, m, and the same V for all the functions simplifies exposition without loss of generality. Existence of bounds is all that is required if supports are bounded. In construction of estimators on unbounded support the upper bounds could be liberal but are assumed known. The constraint on the ϕ^{-1} restricts the measurement error from being supersmooth and the constraint on f_z^{-1} does not permit fast decline to zero at infinity for the density of conditioning z; these would be automatically satisfied if supports were bounded.

Assumption (c) implies that γ and therefore ε_1 are continuous functions. Indeed, an integrable function has a continuous Fourier transform, γ, and $\varepsilon_1 = \gamma\phi$ is continuous, since ϕ is a characteristic function and thus continuous.

Assumption 6(c') holds more generally, e.g., if g is a sum of an integrable function and a polynomial or a sine or cosine function; in such cases γ is a sum of a continuous function (that is separated from zero on bounded sets within its support) with singular functions such as the $\delta-$function, shifted $\delta-$function and its derivatives. The support conditions in the assumption imply that the open support set of the continuous γ_{L1} contains Λ_g. The existence of a function sequence converging in S^* to γ_g with the support properties stated follows from the general properties of generalized functions; the only substantive condition there is that the values of the approximating functions $(\gamma_g)_n$ be separated from zero and from the values of γ_{L1}. For example, if g included an additive constant, then γ_g is a $\delta-$function, and $(\gamma_g)_n$ could be selected as a sequence of step functions. These approximating functions need not be specified, only their existence is required for the proof.

The next assumption is on the stochastic properties of the data generating process, the sampling and on the kernel and bandwidth.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Assumption 7. Moments of order $q > d+1$ of $\frac{1}{(1+\|z\|^2)^{m/2}}y$, $\frac{1}{(1+\|z\|^2)^{m}/2}xy$ conditional on z are bounded; $\{x_i, y_i, z_i\}_{i=1}^n$ is a random sample from $\{x, y, z\}$; the kernel K is the indicator function of the unit sphere; the bandwidth h is such that $h \to 0$ and satisfies $n^{1-\frac{d}{q-1}}h^d \to \infty$.

Note that Assumption 6(b) implies boundedness of first conditional moments (functions $w.$) of $\frac{1}{(1+\|z\|^2)^{m/2}}y$, $\frac{1}{(1+\|z\|^2)^{m/2}}xy$.

Define the estimators (denoting by $(x_k)_i$ the ith observation on the kth component of vector x)

$$\hat{w}_1(z)_n = \frac{\sum_{i=1}^n y_i K\left(\frac{z_i - z}{h}\right)}{\sum_{i=1}^n K\left(\frac{z_i - z}{h}\right)}; \tag{26}$$

$$\hat{w}_{2k}(z)_n = \frac{\sum_{i=1}^n y_i (x_k)_i K\left(\frac{z_i - z}{h}\right)}{\sum_{i=1}^n K\left(\frac{z_i - z}{h}\right)}, \quad k = 1, \ldots d.$$

By Assumption 6(b) the functions $\frac{1}{(1+\|z\|^2)^{m/2}}w_1(z)$, $\frac{1}{(1+\|z\|^2)^{m/2}}w_{2k}(z)$ are bounded in absolute value by V; by 6(a) the density of z, $f_z(z)$, exists and is continuous, then that for z in any closed sphere $S(x,r) \in \text{supp}(f_z)$ the essinf $f_z(z) > 0$. Together with Assumption 7 this is sufficient to ensure that estimators $\frac{1}{(1+\|z\|^2)^{m/2}}\hat{w}_1(z)_n$, $\frac{1}{(1+\|z\|^2)^{m/2}}\hat{w}_{2k}(z)_n$, of $\frac{1}{(1+\|z\|^2)^{m/2}}w_1(z)$, $\frac{1}{(1+\|z\|^2)^{m/2}}w_{2k}(z)$, with $\hat{w}.(z)_n$ computed as (26), with K and h that satisfy Assumption 7 converge in probability uniformly over any compact set (see Devroye, 1978). If $m = 0$, the moment condition is a usual condition made for the Nadaraya–Watson estimator of $w.$; here essentially just the growth of the conditional moment functions has to be restricted; this provides estimators that converge in the topology of S^*.

Uniform convergence in probability implies that the estimators converge in probability in the topology of D^*. If the functions have compact support, this implies convergence in S^*, since on compact support it coincides with convergence in D^*, but on unbounded support growth at infinity needs to be controlled for convergence in S^*. Thus, for any generic estimator $\hat{w}_n$ (represented by a regular function, $\hat{w}(z)_n$) of a regular generalized function $w = w(z)$ in $S^*_{1,m}(V)$ define a corresponding estimator $\tilde{w}_n = \tilde{w}(z)_n$ by setting it to $\hat{w}(z)_n$ if $|\hat{w}(z)_n| < V(1+\|z\|^2)^{m/2}$ and $V(1+\|z\|^2)^{m/2}$ otherwise.

LEMMA 3. *Suppose that the estimator $\hat{w}_n$ converges to w in probability uniformly on bounded sets; then $\tilde{w}_n$ converges to w in probability in S^* and the corresponding Fourier transform $\hat{\varepsilon}_n = Ft(\tilde{w}_n)$ converges in probability in S^* to $\varepsilon = Ft(w)$.*

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Proof. Consider a set $\psi_1, \dots \psi_v \in S$. For any $\zeta > 0$ find a compact set Λ such that for $\tilde{\psi}$ representing any of $\psi_1, \dots, \psi_v, Ft(\psi_1), \dots, Ft(\psi_v)$

$$\int_{R^d \setminus \Lambda} V(1 + \|t\|^2)^{m/2} |\tilde{\psi}(t)| dt < \zeta. \tag{27}$$

Consider

$$|\tilde{w}_n - w| \le \left|\hat{w}_n - w\right| I(z \in \Lambda) + \left|\tilde{w}_n - \hat{w}_n\right| I(z \in \Lambda) + |\tilde{w}_n - w| I\left(z \in R^d \setminus \Lambda\right).$$

Then for any ζ_1 there exists N such that

$$\Pr\left(\left|\left(\tilde{w}_n - w, \tilde{\psi}\right)\right| > \zeta_1\right) \le \Pr\left(\sup_{\Lambda} \left|\hat{w}_n - w\right| \left| \int \tilde{\psi}(t) dt \right| > \zeta_1\right)$$
$$+ \Pr\left(z \in \Lambda : \left|\tilde{w}_n - \hat{w}_n\right| \left| \int \tilde{\psi}(t) dt \right| > \zeta_1 - 2\zeta\right) \le \zeta_1,$$

since the value $\left|\left(\tilde{w}_n - w, \tilde{\psi}\right)\right|$ outside Λ is bonded with certainty by 2ζ and the two probabilities on bounded Λ can be bounded because of uniform convergence in probability of the estimators. Recall that for Fourier transforms $(Ft(\tilde{w}_n - w), \psi) = \left(\hat{w}_n - \hat{w}, Ft(\psi)\right)$. Thus convergence in probability in S^* of the estimators of the functions $\hat{w}$, and also of their Fourier transforms $\hat{\varepsilon}$ is established. ■

Denote $Ft(\tilde{w}_{\cdot n})$ by $\hat{\varepsilon}_{\cdot n}$, note that by construction these random generalized functions are infinitely differentiable, since by the assumption on the kernel K the support of $\hat{w}_{\cdot n}$ is bounded and thus the Fourier transform is differentiable. Define

$$\hat{\phi}_n^{-1} = \exp\left(- \int_0^s \sum_{k=1}^d \hat{\varepsilon}_{1n}^{-1}(t)\left(\left(\hat{\varepsilon}_{1n}(t)\right)_k' - i\hat{\varepsilon}_{2kn}(t)\right) dt_k \right);$$

where the right-hand side may diverge and introduce a trimmed estimator

$$\tilde{\phi}_n^{-1} = \begin{cases} \hat{\phi}_n^{-1}, & \text{if } \left|\hat{\phi}_n^{-1}\right| < V(1 + \|s\|^2)^{m/2} \\ V(1 + \|s\|^2)^{m/2}, & \text{otherwise,} \end{cases} \tag{28}$$

The Theorem below establishes consistency of the estimator of the regression function, g.

THEOREM 6. *If Assumptions 4–7 are satisfied, then the estimator* $\hat{\gamma}_n = \tilde{\phi}_n^{-1} \hat{\varepsilon}_{1n}$ *with* $\tilde{\phi}^{-1}$ *given by* (28) *is such that it exists with probability going to 1 and* $(Ft)^{inv}(\hat{\gamma}_n) - g \to_p 0$ *in* S^*.

Proof. The proof consists of the following steps.

Step 1 considers some compact set Λ inside the support of γ and shows that restricted to such a set $\hat{\phi}^{-1}$ converges to ϕ^{-1} in S^* in probability.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Step 2 examines the continuous part of γ, γ_{L1}, and convergence is probability of estimators defined on compact sets that exclude a set containing the support of the singular part.

Step 3 considers the (possibly unbounded) support set. On any compact set the results in Steps 1 and 2 hold. Consistency on the support of γ (possibly unbounded) is shown in topology of S^* by selecting for any given set of $\psi_1, \dots \psi_v$ from S the corresponding compact set Λ defined in Lemma 3 to bound all the functionals in probability outside of the compact set.

Next, the details are provided.

Step 1. First, $\hat{\varepsilon}_{1n}$; this is a sequence of continuous functions that converge in probability in S^* by Lemma 3. Fourier transforms of functions in L_1 are uniformly continuous on compact sets. Consider a compact set Λ that is inside the support of γ (and thus of ε_1). If Assumption 6(c) holds, the functions w_{1n} are in L_1; $\hat{\varepsilon}_{1n}$ converge in probability pointwise and uniformly to ε_1 on a compact set Λ. Then for any $0 < \zeta_1, \zeta_2 < 1$ we can find $N_1 \equiv N(\Lambda, \zeta_1, \zeta_2)$ such that for $n > N_1$

$$\Pr(\sup_{\Lambda} \left|\hat{\varepsilon}_{1n} - \varepsilon_1\right| > \zeta_1^3) < \zeta_2.$$

Consider Λ that satisfies $\inf_{\Lambda} |\varepsilon_1| > 0$. Set $\zeta_1 \leq \inf_{\Lambda} |\varepsilon_1|$, then

$$\Pr\left(\inf_{\Lambda} \left|\hat{\varepsilon}_{1n}\right| < \zeta_1\right) < \zeta_2; \quad \text{and} \quad \Pr\left(\sup_{\Lambda} \left|\hat{\varepsilon}_{1n}^{-1}\right| > \zeta_1^{-1}\right) < \zeta_2. \tag{29}$$

Under 6(c') for a singular γ_g there exists a deterministic sequence of regular functions, $\left(\gamma_g\right)_n$ that converges to γ_g in S^* and such that support of $\left(\gamma_g\right)_n$ is in a compact set Λ_{gn} that is a proper subset of support of γ; it can be selected such that for some sequence $\zeta_{gn} \to 0$ the Lebesgue measure of Λ_{gn}, $\lambda\left(\Lambda_{gn}\right) < \zeta_{gn}$; $\Lambda_{gn_1} \subset \Lambda_{gn_2}$ for $n_1 > n_2$ and on any Λ_{gn} for some fixed ζ_0 we get $\inf_{\Lambda_{gn}} \left|\left(\gamma_g\right)_n\right| > \zeta_0$. For the corresponding sequence $\left(\varepsilon_g\right)_n = \left(\gamma_g\right)_n \phi$ on Λ_{gn} positive lower bounds on modulus exist and are no less than $\zeta_0 \inf_{\Lambda_{gn}} |\phi|$.

Under 6(c') consider the sequence of deterministic functions $\left(\varepsilon_g\right)_n = \left(\gamma_g\right)_n \phi$ and the difference $(\varepsilon_{L1})_n = \varepsilon_1 - \left(\varepsilon_g\right)_n$. This is a sequence of piecewise deterministic continuous functions that converges in S^* to the continuous function $\varepsilon_{1L1} = \gamma_{L1}\phi$. Then on a compact set Λ for $\zeta_1 \leq \min\left(\inf_{\Lambda} |\varepsilon_{1L1}|, \zeta_0 \inf_{\Lambda_{gn}} |\phi|\right)$ we can find the corresponding N so that for $\hat{\varepsilon}_{1n} = \hat{\varepsilon}_{1n} - \left(\varepsilon_g\right)_n + \left(\varepsilon_g\right)_n$

$$\Pr(\sup_{\Lambda} \left|\hat{\varepsilon}_{1n} - \left(\varepsilon_g\right)_n - \varepsilon_{1L1}\right| < \zeta_1) < \zeta_2.$$

Then

$$\Pr(\inf_{\Lambda} \left|\hat{\varepsilon}_{1n}\right| < \zeta_1) < \Pr(\inf_{\Lambda} \left|\hat{\varepsilon}_{1n} - \left(\varepsilon_g\right)_n\right| < \zeta_1) < \zeta_2;$$

and (29) also holds.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Then with probability approaching 1 on compact set Λ the continuous random functions $\hat{\varkappa}_{kn} = \hat{\varepsilon}_{1n}^{-1}\left(\left(\hat{\varepsilon}_{1n}\right)_k' - i\hat{\varepsilon}_{2kn}\right)$, $k = 1, \ldots, d$ are in a bounded set in S^*, and thus $\hat{\phi}_n^{-1}(s) = \exp\left(-\int_0^s \sum_{k=1}^d \hat{\varkappa}_{kn}(t)dt_k\right)$ is also in a bounded set, and thus with probability approaching 1 satisfies the condition for well-posedness of Theorem 5.

Then the estimators $\hat{\gamma}_n = \hat{\phi}_n^{-1}\varepsilon_1$ are consistent for γ in S^* on compact set Λ where $\inf_{\Lambda}|\varepsilon_{1L1}| > 0$.

Since γ_{L1} is a continuous function, its support is an open set. By Assumption 6(c') the compact support of γ_g is contained inside the open support of γ_{L1}. If $\inf_{\text{supp}(\gamma_{l1})}|\varepsilon_{1L1}| > \zeta_1 > 0$ this would conclude the proof of the Theorem.

Step 2. Otherwise consider the set $\bar{\Lambda} = \cup\Lambda_{gn}$; and a bounded open set $\Omega \subset \text{supp}(\gamma) \setminus \bar{\Lambda}$; then $\hat{\varepsilon}_{1n}$ converge to ε_1 that is continuous on that set. Generally Ω is a union of open connected sets and we can proceed by considering each component in Ω. Without loss of generality such a component can be assumed to be a bounded connected open set containing zero as an interior point, since if it does not contain zero by a shift (which is a continuous operation in S^*) of an arbitrary interior point into zero this can be attained.

All the proofs that follow apply to a bounded open set that will be denoted Ω that is connected and contains zero as an interior point.

Consider a compact set $\Lambda \subset \Omega$ where $\inf_{\Lambda}|\varepsilon_1| > 0$. It follows from the proof in step 1 under continuity of ε_1 that

$$\begin{aligned}
\Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1} - \varepsilon_1^{-1}\right| > \zeta_1\right) &= \Pr\left(\sup_{\Lambda}\left|\left(\hat{\varepsilon}_{1n}^{-1}(\varepsilon_1 - \hat{\varepsilon}_{1n})\right)\varepsilon_1^{-1}\right| > \zeta_1\right) \\
&\le \Pr\left(\sup_{\Lambda}\left|\left(\hat{\varepsilon}_{1n}^{-1}(\varepsilon_1 - \hat{\varepsilon}_{1n})\right)\right| > \zeta_1^2\right) \\
&\le \Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}\right| > \zeta_1^{-1}, \sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}(\varepsilon_1 - \hat{\varepsilon}_{1n})\right| > \zeta_1^2\right) \\
&\quad + \Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}\right| \le \zeta_1^{-1}, \sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}(\varepsilon_1 - \hat{\varepsilon}_{1n})\right| > \zeta_1^2\right) \\
&\le \Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}\right| > \zeta_1^{-1}\right) + \Pr\left(\sup_{\Lambda}\left|\varepsilon_1 - \hat{\varepsilon}_{1n}\right| > \zeta_1^3\right) \\
&\le 2\zeta_2.
\end{aligned}$$

Theorem 3(b) implies that $\phi^{-1}(s) = \exp\left(-\int_0^s \sum_{k=1}^d \varkappa_k(t)dt_k\right)$, where $\varkappa_k$ is the unique continuous function that solves $\varepsilon_1\varkappa - ((\varepsilon_1)_k' - i\varepsilon_{2k}) = 0$ in S^* on Ω. By Assumptions 6(a) and (b) for ϕ, $\varepsilon_1\phi^{-1}$ exists in $S^*$and $g = (Ft)^{inv}\left(\varepsilon_1\phi^{-1}\right)$ in S^*. By Assumption 5a $\varepsilon_1 = \gamma\,\phi$ is nonzero on supp(γ).

Consider the estimator function $\hat{\varkappa}_{kn} = \hat{\varepsilon}_{1n}^{-1}\left(\left(\hat{\varepsilon}_{1n}\right)_k' - i\hat{\varepsilon}_{2kn}\right)$ on compact $\Lambda \subset \Omega$; the function ε_1 is continuous there and it follows that $(\varepsilon_1)_k' - i\varepsilon_{2k} = \varepsilon_1\varkappa_k$ is continuous.

The sequences of random functions $\left(\left(\hat{\varepsilon}_{1n}\right)_k' - i\hat{\varepsilon}_{2kn}\right)$ converge in probability uniformly on Λ in S^* to the continuous function $((\varepsilon_1)_k' - i\varepsilon_{k2})$. Define

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

$\bar{B} = \sup_{\Lambda,k} \left|(\varepsilon_1)'_k - i\varepsilon_{k2}\right|$. For $0 < \zeta_4$ find N_2 such that

$$\Pr\left(\sup_{\Lambda}\left|\left(\left(\hat{\varepsilon}_{1n}\right)'_k - i\hat{\varepsilon}_{2kn}\right) - \left((\varepsilon_1)'_k - i\varepsilon_{k2}\right)\right| > \zeta_4\right) < \zeta_2 \text{ for } n > N_2.$$

Bound $\Pr(\sup_{\Lambda}\left|\hat{\varkappa}_{kn} - \varkappa_k\right| > \zeta_5) \leq$

$$\Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}\right|\left|\left(\left(\hat{\varepsilon}_{1n}\right)'_k - i\hat{\varepsilon}_{2n} - (\varepsilon_1)'_k + i\varepsilon_2\right| + \sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1} - \varepsilon_1^{-1}\right|\left|(\varepsilon_1)'_k - i\varepsilon_{k2}\right| > \zeta_5\right)$$

$$\leq \Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1}\right| > \zeta_1^{-1}\right) + \Pr\left(\sup_{\Lambda}\left|\left(\left(\hat{\varepsilon}_{1n}\right)'_k - i\hat{\varepsilon}_{2kn} - (\varepsilon_1)'_k + i\varepsilon_{k2}\right| > \zeta_5\zeta_1\right)$$

$$+ \Pr\left(\sup_{\Lambda}\left|\hat{\varepsilon}_{1n}^{-1} - \varepsilon_1^{-1}\right| > \zeta_5/\bar{B}\right).$$

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 $\Pr(\sup_{\Lambda}\left|\int_0^s \sum_{k=1}^d \hat{\varkappa}_{kn}(t)dt_k - \int_0^s \sum_{k=1}^d \varkappa_k(t)dt_k\right| > \zeta_6) \leq$

$$\Pr\left(\sup_{\Lambda}\int_0^s \sum_{k=1}^d |\hat{\varkappa}_{kn}(t) - \varkappa_k(t)|dt_k > \zeta_6\right)$$

$$\leq \Pr\left(\sup_{\Lambda}\left|\hat{\varkappa}_{kn}(t) - \varkappa_k(t)\right| > \zeta_6/\mu(\Lambda)\right),$$

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

Consider now on Λ the function $\hat{\phi}_n^{-1}(s) = \exp\left(-\int_0^s \sum_{k=1}^d \hat{\varkappa}_{kn}(t)dt_k\right)$. Define $\tilde{B} = \sup_{\Lambda} V\left(1 + \|s\|^2\right)^{m/2}$; then $\sup_{\Lambda}\left|\phi^{-1}(s)\right| < \tilde{B}$ by Assumption 6(b).

Then $\Pr(\sup_{\Lambda}\left|\hat{\phi}_n^{-1} - \phi^{-1}\right| > \zeta_7) \leq$

$$\Pr\left(\sup_{\Lambda}\left|\int_0^s \sum_{k=1}^d \hat{\varkappa}_{kn}(t)dt_k - \int_0^s \sum_{k=1}^d \varkappa_k(t)dt_k\right| > \ln\left(1 + \tilde{B}^{-1}\zeta_7\right)\right),$$

and is smaller than $4\zeta_2$ for $\zeta_7 = \ln(1 + \tilde{B}^{-1}\zeta_6)$.

Since for the continuous functions $\sup_{\Lambda}\left|\phi_n^{-1}\hat{\varepsilon}_{1n} - \phi^{-1}\varepsilon_1\right| <$

$$\tilde{B}\sup_{\Lambda}\left|\varepsilon_1 - \hat{\varepsilon}_{1n}\right| + \sup_{\Lambda}|\varepsilon_1|\sup_{\Lambda}\left|\phi_n^{-1} - \phi^{-1}\right|$$

by similar derivations

$$\Pr\left(\sup_{\Lambda}\left|\hat{\phi}_n^{-1}\hat{\varepsilon}_{1n} - \phi^{-1}\varepsilon_1\right| > \zeta_8\right) < 5\zeta_2$$

if $\zeta_8 < \min\left\{\tilde{B}^{-1}\zeta_1, \zeta_7(\sup_{\Lambda}|\varepsilon_1|)^{-1}\right\}$.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Step 3. Consider an arbitrary set $\psi_1, \ldots, \psi_v \in S$; we need to show that for any $\tilde{\psi}$ from that set $\Pr\left(\left|(\hat{\gamma}_n - \gamma, \tilde{\psi})\right| > \zeta\right)$ goes to zero. Let the corresponding compact set Λ be defined by (27). Note that once the set of functions $\psi_1, \ldots, \psi_v \in S$ is given, the compact set Λ is determined by those functions and fixed.

Consider $\tilde{\phi}_n^{-1}$. Define a compact set $\Lambda_1 \subset \Lambda$ such that Λ_1 includes $\cup \Lambda_{gn}$. From the proof in Step 1 it follows that for large enough N the estimator $\tilde{\phi}_n^{-1} = \hat{\phi}_n^{-1}$ on any compact Λ_1 with arbitrarily high probability and then $\hat{\varepsilon}_{1n}\tilde{\phi}_n^{-1}$ converges to γ on Λ_1 in probability in S^*.

Next consider Ω defined in Step 2. Since $\hat{\varepsilon}_{1n} - \varepsilon_1$ converges to zero in probability in S^* by Lemma 3 and since on Ω this difference is a continuous function, then also the continuous function $|\hat{\varepsilon}_{1n}(t) - \varepsilon_1(t)|$ converges to zero in probability in S^* on Ω. Thus

$$\int_{\Omega \setminus \Lambda} V\left(1 + \|t\|^2\right)^{m/2} |\hat{\varepsilon}_{1n}(t) - \varepsilon_1(t)| \left|\tilde{\psi}(t)\right| dt$$

converges in probability to zero. Then since

$$\begin{aligned}\left|\left(\tilde{\gamma}_n - \gamma, \tilde{\psi}\right)\right| &\le \sup_{\Lambda} \left|\tilde{\phi}_n^{-1}\hat{\varepsilon}_{1n} - \phi^{-1}\varepsilon_1\right| \int_{\Lambda} \left|\tilde{\psi}(t)\right| dt \\ &+ \int_{\Omega\setminus\Lambda} V\left(1 + \|t\|^2\right)^{m/2} |\varepsilon_1(t)| \left|\tilde{\psi}(t)\right| dt \\ &+ \int_{\Omega\setminus\Lambda} V\left(1 + \|t\|^2\right)^{m/2} |\hat{\varepsilon}_{1n}(t) - \varepsilon_1(t)| \left|\tilde{\psi}(t)\right| dt,\end{aligned}$$

it follows that

$$\begin{aligned}\Pr\left(\left|(\tilde{\gamma}_n - \gamma, \tilde{\psi})\right| > \zeta\right) &\le \Pr\left(\sup_{\Lambda} \left|\hat{\phi}_n^{-1}\hat{\varepsilon}_{1n} - \phi^{-1}\varepsilon_1\right| > \frac{\zeta}{\int_{\Lambda} \left|\tilde{\psi}(t)\right| dt}\right) \\ &+ \Pr\left(\int_{\Omega\setminus\Lambda} V(1 + \|t\|^2)^{m/2} |\varepsilon_1(t)| \left|\tilde{\psi}(t)\right| dt > \zeta\right) \\ &+ \Pr\left(\int_{\Omega\setminus\Lambda} V(1 + \|t\|^2)^{m/2} |\hat{\varepsilon}_{1n}(t) - \varepsilon_1(t)| \left|\tilde{\psi}(t)\right| dt > \zeta\right).\end{aligned}$$

It follows from the proof in Step 2 that the first probability converges to zero. The second converges to zero, since by Assumption 6(b) ε_1 is in a bounded set in S^* and by definition of the set Λ and Lemma 3. The third converges to zero in S^* by convergence of $\hat{\varepsilon}_{1n}$. Combining this with the convergence restricted to Λ_1 we obtain that $\hat{\gamma}_n$ converges in probability to $\gamma = \phi^{-1}\varepsilon_1$ in S^*. Taking inverse Fourier transforms in S^* concludes the proof. ■

For the proof of consistency trimming is used in the estimator; this requires knowing the bounds on the values of m and V. The bounds could be as liberal as desired but are assumed known in the proof. In practice in finite sample $\hat{\phi}_n$

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

may never take zero values and then the bound could be considered to be liberal enough to allow $\tilde{\phi}_n = \hat{\phi}_n$; if there are zero values some sample size dependent cut-off could apply, e.g., based on $V = n$ (this would be analogous to using n^{-1} in the denominator of nonparametric estimators as is sometimes done). While asymptotic properties of such sample-based trimming would need to be established and this task is not attempted here, these alternatives to imposing bounds could be considered.

The Theorem provides consistency of plug-in estimators for solutions to the system of equations (20) and consequently (3) in a fairly general setup. Nevertheless, some assumptions can be further relaxed. Of course, establishing results for a compact support of the Fourier transforms is much easier and thus using spectral cut-off can be advantageous especially when high frequency components of the regression function may be commensurate with the magnitude of the error components.

Computation of the estimators requires applying Fourier transforms and inverse Fourier transforms. This can be accomplished with numerical algorithms. However, it is possible to simplify the estimated $\hat{\varepsilon}$. Consider instead of the estimator in (26), $\hat{w}_1(z)_n$, an estimator computed as $\sum_{i=1}^{n} \alpha_i^{-1} y_i K\left(\frac{z_i - z}{h}\right)$ with the weight $\alpha_i = \sum_{j \neq i}^{n} K\left(\frac{z_j - z_i}{h}\right)$ replacing the $z-$dependent weight in $\hat{w}_1(z)_n$ and similar estimators for $\hat{w}_{2k}(z)_n$. Then the corresponding Fourier transform $\hat{\varepsilon}_{1n}(s)$ can be expressed as $\sum_{i=1}^{n} \alpha_i^{-1} y_i e^{is^T z_i} sinc\left(\frac{s^T \bar{h}}{\pi}\right)$, where $\bar{h}$ is the vector of bandwidths and by definition $sinc(x) = \frac{\sin \pi x}{\pi x}$, and similar expressions for $\hat{\varepsilon}_{2kn}$. Further computation for the estimators would have to be done numerically.

A referee pointed out that the residuals in the estimated model (3) could be employed for diagnostic testing of the model assumptions. Consider the case when the model is fully identified; this could hold if both γ and ϕ are supported on R^d. Then under the conditions of Theorem 6 the distribution of the error, u, is consistently estimated by $Ft^{inv}(\tilde{\phi}_n)$. Since under the assumptions, u is independent of z, by sampling from its estimated distribution one could obtain $\hat{u}^*$ (the superscript * refers here to the fact that the estimate is obtained by resampling and the process can be repeated). Estimate $\hat{x}_i^* = z_i - \hat{u}_i^*$. Then obtain $\hat{u}_{xi}^* = x_i - \hat{x}_i^*$. The residuals $\hat{u}_{yi}^* = y_i - \hat{g}(\hat{x}_i^*)$, where $\hat{g} = Ft^{inv}(\tilde{\gamma}_n)$. Resampling provides draws from the estimated distribution of the model residuals. The conditions of the model as well as other hypotheses can be tested based on this distribution.

6. CONCLUSION

This paper was devoted to treating a single convolution equation and a specific system of convolution equations; many statistical models with various independence conditions give rise to such equations; measurement error is emphasized

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

here, but equations of this type are also applicable in other models, such as factor models and panel data models; many examples are presented in Zinde-Walsh (2014). The results of this paper indicate conditions for identification and well-posedness when casting these equations in terms of generalized functions; the generalized functions approach enlarges the area of applicability of the models. Consistent estimation and inference can be extended to wider nonparametric classes of model.

REFERENCES

An, Y. & Y. Hu (2012) Well-posedness of measurement error models for self-reported data. *Journal of Econometrics* 168, 259–269.

Antosik, P., J. Mikusinski, & R. Sikorski (1973) *Theory of Distributions. The Sequential Approach*. Elsevier/PWN.

Carrasco, M. & J.-P. Florens (2010) A spectral method for deconvolving a density. *Econometric Theory* 27, 546–581.

Carrasco, M., J.-P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 5633–5751. North-Holland.

Carroll, R.J., D. Ruppert, L.A. Stefanski, & C.M. Crainiceanu (2006) *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman & Hall.

Chen, X., H. Hong, & D. Nekipelov (2011) Nonlinear models of measurement errors. *Journal of Economic Literature* 49, 901–937.

Cunha, F., J.J. Heckman, & S.M. Schennach (2010) Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78, 883–933.

Devroye, L. (1978) The uniform convergence of the Nadaraya-Watson regression function estimate. *The Canadian Journal of Statistics* 6, 179–191.

Devroye, L. & L. Győrfi (1985) *Nonparametric Density Estimation: The L_1 View*. Wiley.

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

Gelfand, I.M. & G.E. Shilov (1964a) *Generalized Functions, vol. 1, Properties and Operations*. Academic Press.

Gelfand, I.M. & G.E. Shilov (1964b) *Generalized Functions, vol. 2, Spaces of Test Functions and Generalized Functions*. Academic Press.

Gelfand, I.M. & N. Ya Vilenkin (1964c) *Generalized Functions, vol. 4, Applications of Harmonic Analysis*. Academic Press.

Hadamard, J. (1923) *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press.

Hausman, J.A., W.K. Newey, H. Ichimura, & J. Powell (1991) Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics* 50(3), 273–295.

Hirata, Y. & H. Ogata (1958) On the exchange formula for distributions. *Journal of Science of the Hiroshima University, Series A* 22, 147–152.

Hu, Y. (2008) Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* 144, 27–61.

Kirillov, A.A. & A.D. Gvishiani (1982) *Theorems and Problems in Functional Analysis*. Springer-Verlag.

Klann, E., M. Kuhn, D.A. Lorenz, P. Maass, & H. Thiele (2007) Shrinkage versus deconvolution. *Inverse Problems* 23, 2231–2248.

Koralov, L. & Ya. Sinai (2007) *Theory of Probability and Random Processes*. Springer.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press

Kotlyarski, I. (1967) On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics* 20, 69–76.

Meister, A. (2009) *Deconvolution Problems in Nonparametric Statistics*. Lecture Notes in Statistics. Springer-Verlag.

Newey, W. (2001) Flexible simulated moment estimation of nonlinear errors-in-variables models. *Review of Economics and Statistics* 83, 616–627.

Schennach, S. (2007) Instrumental variable estimation in nonlinear errors-in-variables models. *Econometrica* 75, 201–239.

Schwartz, L. (1966) *Théorie des distributions*. Hermann.

Sobolev, S.L. (1992) *Cubature Formulas and Modern Analysis*. Gordon and Breach Science Publishers.

Wang, L. & Ch. Hsiao (2011) Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics*, 30–44.

Zinde-Walsh, V. (2008) Kernel estimation when density may not exist. *Econometric Theory* 24, 696–725.

Zinde-Walsh, V. (2009) Errors-In-Variables Models: A Generalized Functions Approach. McGill University Working paper, arXiv:0909.5390v1 [stat.ME].

Zinde-Walsh, V. (2011) Presidential address: Mathematics in economics and econometrics. *Canadian Journal of Economics* 44, 1052–1068.

Zinde-Walsh (2014) Identification and well-posedness in nonparametric models with independence conditions. In J. Racine, L. Su, & A. Ullah (eds.), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford University Press, 97–128.

https://doi.org/10.1017/S0266466614000140 Published online by Cambridge University Press