Identification and well-posedness in nonparametric models with independence conditions

Victoria Zinde-Walsh*

McGill University and CIREQ

victoria.zinde-walsh@mcgill.ca

September 10, 2012

*The support of the Social Sciences and Humanities Research Council of Canada (SSHRC) and the *Fonds québecois de la recherche sur la société et la culture* (FRQSC) is gratefully acknowledged.

Abstract

This paper provides a nonparametric analysis for several classes of models, with cases such as classical measurement error, regression with errors in variables, and other models that may be represented in a form involving convolution equations. The focus here is on conditions for existence of solutions, nonparametric identification and well-posedness in the space S^* of generalized functions (tempered distributions). This space provides advantages over working in function spaces by relaxing assumptions and extending the results to include a wider variety of models, for example by not requiring existence of density. Classes of (generalized) functions for which solutions exist are defined; identification conditions, partial identification and its implications are discussed. Conditions for well-posedness are given and the related issues of plug-in estimation and regularization are examined.

1 Introduction

Many statistical and econometric models involve independence (or conditional independence) conditions that can be expressed via convolution. Examples are independent errors, classical measurement error and Berkson error, regressions involving data measured with these types of errors, common factor models and models that conditionally on some variables can be represented in similar forms, such as a nonparametric panel data model with errors conditionally on observables independent of the idiosyncratic component.

Although the convolution operator is well known, this paper provides explicitly convolution equations for a wide list of models for the first time. In many cases the analysis in the literature takes Fourier transforms as the starting point, e.g. characteristic functions for distributions of random vectors (as in the famous Kotlyarski lemma, 1967). The emphasis here on convolution equations for the models provides the opportunity to explicitly state nonparametric classes of functions defined by the model for which such equations hold, in particular, for densities, conditional densities and regression functions. The statistical model may give rise to different systems of convolution equations and may be over-identified in terms of convolution equations; some choices may be better suited to different situations, for example, here in Section 2 two sets of convolution equations (4 and 4a in Table 1) are provided for the same classical measurement error model with two measurements; it turns out that one of those allows to relax some independence conditions, while the other makes it possible to relax a support assumption in identification. Many of the convolution equations derived here are based on density-weighted conditional averages of the observables.

The main distinguishing feature is that here all the functions defined by the model are considered within the space of generalized functions S^*, the space of so-called tempered distributions (they will be referred to as generalized functions). This is the dual space, the space of linear continuous functionals, on the space S of well-behaved functions: the functions in S are infinitely differentiable and all the derivatives go to zero at infinity faster than any power. An important advantage of assuming the functions are in the space of generalized functions is that in that space any distribution function has a density (generalized function) that continuously depends on the distribution function, so that distributions with mass points and fractal measures have well-defined generalized densities.

Any regular function majorized by a polynomial belongs to S^*; this includes polynomially growing regression functions and binary choice regression as well as many conditional density functions. Another advantage is that Fourier transform is an isomorphism of this space, and thus the usual approaches in the literature that employ characteristic functions are also included. Details about the space S^* are in Schwartz (1966) and are summarized in Zinde-Walsh (2012).

The model classes examined here lead to convolution equations that are similar to each other in form; the main focus of this paper is on existence, identification, partial identification and well-posedness conditions. Existence and uniqueness of solutions to some systems of convolution equations in the space S^* were established in Zinde-Walsh (2012). Those results are used here to state identification in each of the models. Identification requires examining support of the functions and generalized functions that enter into the models; if support excludes an open set then identification at least for some unknown functions in the model fails, however, some isolated points or lower-dimensional manifolds where the e.g. the characteristic function takes zero values (an example is the uniform distribution) does not preclude identification in some of the models. This point was made in e.g. Carrasco and Florens (2010), Evdokimov and White (2011) and is expressed here in the context of operating in S^*. Support restriction for the solution may imply that only partial identification will be provided. However, even in partially identified models some features of interest (see, e.g. Matzkin, 2007) could be identified thus some questions could be addressed even in the absence of full identification. A common example of incomplete identification which nevertheless provides important information is Gaussian deconvolution of a blurred image of a car obtained from a traffic camera; the filtered image is still not very good, but the licence plate number is visible for forensics.

Well-posedness conditions are emphasized here. The well-known definition by Hadamard (1923) defines well-posedness via three conditions: existence of a solution, uniqueness of the solution and continuity in some suitable topology. The first two are essentially identification. Since here we shall be defining the functions in subclasses of S^* we shall consider continuity in the topology of this generalized functions space. This topology is weaker than the topologies in functions spaces, such as the uniform or L_p topologies; thus differentiating the distribution function to obtain a density is a well-posed problem in S^*, by contrast, even in the class of absolutely continuous distributions with uniform metric where identification for density in the space L_1 holds, well-posedness however does not obtain (see discussion in Zinde-Walsh, 2011). But even though in the weaker topology of S^* well-posedness obtains more widely, for the problems considered here some additional restrictions may be required for well-posedness.

Well-posedness is important for plug-in estimation since if the estimators are in a class where the problem is well-posed they are consistent, and conversely, if well-posedness does not hold consistency will fail for some cases. Lack of well-posedness can be remedied by regularization, but the price is often more extensive requirements on the model and slower convergence. For example, in deconvolution (see e.g. Fan, 1991, and most other papers cited here) spectral cut-off regularization is utilized; it crucially depends on knowing the rate of the decay at infinity of the density.

Often non-parametric identification is used to justify parametric or semi-parametric estimation; the claim here is that well-posedness should be an important part of this justification. The reason for that is that in estimating a possibly misspecified parametric model, the misspecified functions of the observables belong in a nonparametric neighborhood of the true functions; if the model is non-parametrically identified, the unique solution to the true model exists, but without well-posedness the solution to the parametric model and to the true one may be far apart.

For deconvolution An and Hu (2012) demonstrate well-posedness in spaces of integrable density functions when the measurement error has a mass point; this may happen in surveys when probability of truthful reporting is non-zero. The conditions for well-posedness here are provided in S^*; this then additionally does not exclude mass points in the distribution of the mismeasured variable itself; there is some empirical evidence of mass points in earnings and income. The results here show that in S^* well-posedness holds more generally: as long as the error distribution is not super-smooth.

The solutions for the systems of convolution equations can be used in plug-in estimation. Properties of nonparametric plug-in estimators are based on results on stochastic convergence in S^* for the solutions that are stochastic functions expressed via the estimators of the known functions of the observables.

Section 2 of the paper enumerates the classes of models considered here. They are divided into three groups: 1. measurement error models with classical and Berkson errors and possibly an additional measurement, and common factor models that transform into those models; 2. nonparametric regression models with classical measurement and Berkson errors in variables; 3. measurement error and regression models with conditional independence. The corresponding convolution equations and systems of equations are provided and discussed. Section 3 is devoted to describing the solutions to the convolution equations of the models. The main mathematical aspect of the different models is that they require solving equations of a similar form. Section 4 provides a table of identified solutions and discusses partial identification and well-posedness. Section 5 examines plug-in estimation. A brief conclusion follows.

2 Convolution equations in classes of models with independence or conditional independence

This section derives systems of convolution equations for some important classes of models. The first class of model is measurement error models with some independence (classical or Berkson error) and possibly a second measurement; the second class is regression models with classical or Berkson type error; the third is models with conditional independence. For the first two classes the distributional assumptions for each model and the corresponding convolution equations are summarized in tables; it is indicated which of the functions are known and which unknown; a brief discussion of each model and derivation of the convolution equations follows. The last part of this section discusses convolution equations for two specific models with conditional independence; one is a panel data model studied by Evdokimov (2011), the other a regression model where independence of measurement error of some regressors obtains conditionally on a covariate.

The general assumption made here is that all the functions in the convolution equations belong to the space of generalized functions S^*.

Assumption 1. *All the functions defined by the statistical model are in the space of generalized functions* S^*.

This space of generalized function includes functions from most of the function classes that are usually considered, but allows for some useful generalizations. The next subsection provides the necessary definitions and some of the implications of working in the space S^*.

2.1 The space of generalized functions S^*.

The space S^* is the dual space, i.e. the space of continuous linear functionals on the space S of functions. The theory of generalized functions is in Schwartz (1966); relevant details are summarized in Zinde-Walsh (2012). In this subsection the main definitions and properties are reproduced.

Recall the definition of S.

For any vector of non-negative integers $m = (m_1, ...m_d)$ and vector $t \in R^d$ denote by t^m the product $t_1^{m_1}...t_d^{m_d}$ and by ∂^m the differentiation operator $\frac{\partial^{m_1}}{\partial x_1^{m_1}}...\frac{\partial^{m_d}}{\partial x_d^{m_d}}$; C_∞ is the space of infinitely differentiable (real or complex-valued) functions on R^d. The space $S \subset C_\infty$ of test functions is defined as:

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

for any $k = (k_1, ...k_d), l = (l_1, ...l_d)$, where $k = (0, ...0)$ corresponds to the function itself, $t \to \infty$ coordinate-wise; thus the functions in S go to zero at infinity faster than any power as do their derivatives; they are rapidly decreasing functions. A sequence in S converges if in every bounded region each $\left| t^l \partial^k \psi(t) \right|$ converges uniformly.

Then in the dual space S^* any $b \in S^*$ represents a linear functional on

S; the value of this functional for $\psi \in S$ is denoted by (b, ψ). When b is an ordinary (point-wise defined) real-valued function, such as a density of an absolutely continuous distribution or a regression function, the value of the functional on real-valued ψ defines it and is given by

$$(b, \psi) = \int b(x)\psi(x)dx.$$

If b is a characteristic function it may be complex-valued, then the value of the functional b applied to $\psi \in S$ where S is the space of complex-valued functions, is

$$(b, \psi) = \int b(x)\overline{\psi(x)}dx,$$

where overbar denotes complex conjugate. The integrals are taken over the whole space R^d.

The generalized functions in the space S^* are continuously differentiable and the differentiation operator is continuous; Fourier transforms and their inverses are defined for all $b \in S^*$, the operator is a (continuos) isomorphism of the space S^*. However, convolutions and products are not defined for all pairs of elements of S^*, unlike, say, the space L_1; on the other hand, in L_1 differentiation is not defined and not every distribution has a density that is an element of L_1.

Assumption 1 places no restrictions on the distributions, since in S^* any distribution function is differentiable and the differentiation operator is continuous. The advantage of not restricting distributions to be absolutely continuous is that mass points need not be excluded; distributions representing fractal measures such as the Cantor distribution are also allowed. This means that mixtures of discrete and continuous distributions e.g. such as those examined by An and Hu (2012) for measurement error in survey responses, some of which may be error-contaminated, but some may be truthful leading to a mixture with a mass point distribution are included. Moreover, in S^* the case of mass points in the distribution of the mismeasured variable is also easily handled; in the literature such mass points are documented for income or work hours distributions in the presence of rigidities such as unemployment compensation rules (e.g. Green and Riddell, 1997). Fractal distributions may arise in some situations, e.g. Karlin's (1958) example of the equilibrium price distribution in an oligopolistic game.

For regression functions the assumption $g \in S^*$ implies that growth at infinity is allowed but is somewhat restricted. In particular for any ordinary point-wise defined function $b \in S^*$ the condition

$$\int ... \int \Pi_{i=1}^{d} \left(\left(1 + t_i^2\right)^{-1} \right)^{m_i} |b(t)| \, dt_1 ... dt_d < \infty, \tag{1}$$

needs to be satisfied for some non-negative valued $m_1, ..., m_d$. If a locally integrable function g is such that its growth at infinity is majorized by a polynomial, then $b \equiv g$ satisfies this condition. While restrictive this still widens the applicability of many currently available approaches. For example in Berkson regression the common assumption is that the regression function be absolutely integrable (Meister, 2009); this excludes binary choice, linear and polynomial regression functions that belong to S^* and satisfy Assumption 1. Also, it is advantageous to allow for functions that may not belong to any ordinary function classes, such as sums of $\delta-$functions ("sum of peaks") or (mixture) cases with sparse parts of support, such as isolated points; such functions are in S^*. Distributions with mass points can arise when the response to a survey questions may be only partially contaminated; regression "sum of peaks" functions arise e.g. in spectroscopy and astrophysics where isolated point supports are common.

2.2 Measurement error and related models

Current reviews for measurement error models are in Carrol et al, (2006), Chen et al (2011), Meister (2009).

Here and everywhere below the variables x, z, x^*, u, u_x are assumed to be in R^d; y, v are in R^1; all the integrals are over the corresponding space; density of ν for any ν is denoted by f_v; independence is denoted by $\perp$; expectation of x conditional on z is denoted by $E(x|z)$.

2.2.1 List of models and corresponding equations

The table below lists various models and corresponding convolution equations. Many of the equations are derived from density weighted conditional expectations of the observables.

Recall that for two functions, f and g convolution $f * g$ is defined by

$$(f * g)(x) = \int f(w)g(x-w)dw;$$

this expression is not always defined. A similar expression (with some abuse of notation since generalized functions are not defined pointwise) may hold for generalized functions in S^*; similarly, it is not always defined. With Assumption 1 for the models considered here we show that convolution equations given in the Tables below hold in S^*.

Table 1. Measurement error models: 1. Classical measurement error; 2. Berkson measurement error; 3. Classical measurement error with additional observation (with zero conditional mean error); 4., 4a. Classical error with additional observation (full independence).

Model	Distributional assumptions	Convolution equations	Known functions	Unknown functions
1.	$z = x^* + u$ $x^* \perp u$	$f_{x^*} * f_u = f_z$	f_z, f_u	f_{x^*}
2.	$z = x^* + u$ $z \perp u$	$f_z * f_{-u} = f_{x^*}$	f_z, f_u	f_{x^*}
3.	$z = x^* + u;$ $x = x^* + u_x$ $x^* \perp u;$ $E(u_x \| x^*, u) = 0;$ $E\left\|z\right\| < \infty; E\left\|u\right\| < \infty.$	$f_{x^*} * f_u = f_z;$ $h_k * f_u = w_k,$ with $h_k(x) \equiv x_k f_{x^*}(x);$ $k = 1, 2...d$	$f_z, w_k,$ $k = 1, 2...d$	$f_{x^*};\ f_u$
4.	$z = x^* + u;$ $x = x^* + u_x; x^* \perp u;$ $x^* \perp u_x; E(u_x) = 0;$ $u \perp u_x;$ $E\left\|z\right\| < \infty; E\left\|u\right\| < \infty.$	$f_{x^*} * f_u = f_z;$ $h_k * f_u = w_k;$ $f_{x^*} * f_{u_x} = f_x;$ with $h_k(x) \equiv x_k f_{x^*}(x);$ $k = 1, 2...d$	$f_z,\ f_x; w; w_k$ $k = 1, 2...d$	$f_{x^*}; f_u,\ f_{u_x}$
4a.	Same model as 4., alternative equations:	$f_{x^*} * f_u = f_z;$ $f_{u_x} * f_{-u} = w;$ $h_k * f_{-u} = w_k,$ with $h_k(x) \equiv x_k f_{u_x}(x);$ $k = 1, 2...d$	–"–	–"–

Notation: $k = 1, 2, ..., d$; in 3. and 4, $w_k = E(x_k f_z(z)|z)$; in 4a $w = f_{z-x}; w_k = E(x_k w(z-x)|\,(z-x))$.

Theorem 1. *Under Assumption 1 for each of the models 1-4 the corresponding convolution equations of Table 1 hold in the generalized functions space S^*.*

The proof is in the derivations of the following subsection.

Assumption 1 requires considering all the functions defined by the model as elements of the space S^*, but if the functions (e.g. densities, the conditional moments) exist as regular functions, the convolutions are just the usual convolutions of functions, on the other hand, the assumption allows to consider convolutions for cases where distributions are not absolutely continuous.

2.2.2 Measurement error models and derivation of the corresponding equations.

1. The classical measurement error model.

The case of the classical measurement error is well known in the literature. The concept of error independent of the variable of interest is applicable to many problems in seismology, image processing, where it may be assumed that the source of the error is unrelated to the signal. In e.g. Cunha et al. (2010) it is assumed that some constructed measurement of ability of a child derived from test scores fits into this framework. As is well-known in regression a measurement error in the regressor can result in a biased estimator (attenuation bias).

Typically the convolution equation

$$f_{x^*} * f_u = f_z$$

is written for density functions when the distribution function is absolutely continuous. The usual approach to possible non-existence of density avoids considering the convolution and focuses on the characteristic functions. Since density always exists as a generalized function and convolution for such generalized functions is always defined it is possible to write convolution equations in S^* for any distributions in model 1. The error distribution (and thus generalized density f_u) is assumed known thus the solution can be obtained by "deconvolution" (Carrol et al (2006), Meister (2009), the review of Chen et al (2011) and papers by Fan (1991), Carrasco and Florens(2010) among others).

2. The Berkson error model.

For Berkson error the convolution equation is also well-known. Berkson error of measurement arises when the measurement is somehow controlled and the error is caused by independent factors, e.g. amount of fertilizer applied is given but the absorption into soil is partially determined by factors independent of that, or students' grade distribution in a course is given in advance, or distribution of categories for evaluation of grant proposals is determined by the granting agency. The properties of Berkson error are very different from that of classical error of measurement, e.g. it does not lead to attenuation bias in regression; also in the convolution equation the unknown function is directly expressed via the known ones when the distribution of Berkson error is known. For discussion see Carrol et al (2006), Meister (2009), and Wang (2004).

Models 3. and 4. The classical measurement error with another observation.

In 3., 4. in the classical measurement error model the error distribution is not known but another observation for the mis-measured variable is available; this case has been treated in the literature and is reviewed in Carrol et al (2006), Chen et al (2011). In econometrics such models were examined by Li and Vuong (1998), Li (2002), Schennach (2004) and subsequently others (see e.g. the review by Chen et al, 2011). In case 3 the additional observation contains an error that is not necessarily independent, just has conditional mean zero.

Note that here the multivariate case is treated where arbitrary dependence for the components of vectors is allowed. For example, it may be of interest to consider the vector of not necessarily independent latent abilities or skills as measured by different sections of an IQ test, or the GRE scores.

Extra measurements provide additional equations. Consider for any $k = 1, ...d$ the function of observables w_k defined by density weighted expectation $E(x_k f_z(z)|z)$ as a generalized function; it is then determined by the values of the functional (w_k, ψ) for every $\psi \in S$. Note that by assumption $E(x_k f_z(z)|z) = E(x_k^* f_z(z)|z)$; then for any $\psi \in S$ the value of the functional:

$$
\begin{aligned}
(E(x_k^* f_z(z)|z), \psi) &= \int [\int x_k^* f_{x^*,z}(x^*, z) dx^*] \psi(z) dz = \\
\int \int x_k^* f_{x^*,z}(x^*, z) \psi(z) dx^* dz &= \int \int x_k^* \psi(x^* + u) f_{x^*,u}(x^*, u) dx^* du = \\
\int \int x_k^* f_{x^*}(x^*) f_u(u) \psi(x^* + u) dx^* du &= (h_k * f_u, \psi).
\end{aligned}
$$

The third expression is a double integral which always exists if $E\left\|x^*\right\| < \infty$; this is a consequence of boundedness of the expectations of z and u. The fourth is a result of change of variables (x^*, z) into (x^*, u), the fifth uses independence of $x^*$and u, and the sixth expression follows from the corresponding expression for the convolution of generalized functions (Schwartz, 1967, p.246). The conditions of model 3 are not sufficient to identify the distribution of u_x; this is treated as a nuisance part in model 3.

The model in 4 with all the errors and mis-measured variable independent of each other was investigated by Kotlyarski (1967) who worked with the joint characteristic function. In 4 consider in addition to the equations written for model 3 another that uses the independence between x^* and u_x and involves f_{u_x}.

In representation 4a the convolution equations involving the density f_{u_x} are obtained by applying the derivations that were used here for the model in 3.:

$$z = x^* + u;$$

$$x = x^* + u_x,$$

to the model in 4 with $x - z$ playing the role of z, u_x playing the role of x^*, $-u$ playing the role of u, and x^* playing the role of u_x. The additional convolution equations arising from the extra independence conditions provide extra equations and involve the unknown density f_{u_x}. This representation leads to a generalization of Kotlyarski's identification result similar to that obtained by Evdokimov (2011) who used the joint characteristic function. The equations in 4a make it possible to identify f_u, f_{u_x} ahead of f_{x^*}; for identification this will require less restrictive conditions on the support of the characteristic function for x^*.

2.2.3 Some extensions

A. Common factor models.

Consider a model $\tilde{z} = AU$, with A a matrix of known constants and $\tilde{z}$ a $m \times 1$ vector of observables, U a vector of unobservable variables. Usually, A is a block matrix and AU can be represented via a combination of mutually independent vectors. Then without loss of generality consider the model

$$\tilde{z} = \tilde{A}x^* + \tilde{u}, \tag{2}$$

where $\tilde{A}$ is a $m \times d$ known matrix of constants, $\tilde{z}$ is a $m \times 1$ vector of observables, unobserved x^* is $d \times 1$ and unobserved $\tilde{u}$ is $m \times 1$. If the model (2) can be transformed to model 3 considered above, then x^* will be identified whenever identification holds for model 3. Once some components are identified identification of other factors could be considered sequentially.

Lemma 1. *If in (2) the vectors x^* and $\tilde{u}$ are independent and all the components of the vector $\tilde{u}$ are mean independent of each other and are mean zero and the matrix A can be partitioned after possibly some permutation of rows as* $\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ *with* $rank A_1 = rank A_2 = d$, *then the model (2) implies model 3.*

Proof. Define $z = T_1\tilde{z}$, where conformably to the partition of A the partitioned $T_1 = \begin{pmatrix} \tilde{T}_1 \\ 0 \end{pmatrix}$, with $\tilde{T}_1 A_1 x^* = x^*$ (such a $\tilde{T}_1$ always exists by the rank condition); then $z = x^* + u$, where $u = T_1\tilde{u}$ is independent of x^*. Next define $T_2 = \begin{pmatrix} 0 \\ \tilde{T}_2 \end{pmatrix}$ similarly with $\tilde{T}_2 A_2 x^* = x^*$.

Then $x = T_2\tilde{z}$ is such that $x = x^* + u_x$, where $u_x = T_2\tilde{u}$ and does not include any components from u. This implies $Eu_x|(x^*, u) = 0$. Model 3 holds. ■

Here dependence in components of x^* is arbitrary. A general structure with subvectors of U independent of each other but with components which may be only mean independent (as $\tilde{u}$ here) or arbitrarily dependent (as in x^*) is examined by Ben-Moshe (2012). Models of linear systems with full independence were examined by e.g. Li and Vuong (1998). These models lead to systems of first-order differential equations for the characteristic functions.

It may be that there are no independent components x^* and $\tilde{u}$ for which the conditions of Lemma 1 are satisfied. Bonhomme and Robin (2010) proposed to consider products of the observables to increase the number of equations in the system and analyzed conditions for identification; Ben-Moshe (2012) provided necessary and sufficient conditions under which this strategy leads to identification when there may be some dependence.

B. Error correlations with more observables.

The extension to non-zero $E(u_x|z)$ in model 3 is trivial if this expectation is a known function. A more interesting case results if the errors u_x and u are related, e.g.

$$u_x = \rho u + \eta; \eta \perp z.$$

With an unknown parameter (or function of observables) ρ if more observations are available more convolution equations can be written to identify all the unknown functions. Suppose that additionally a observation y is available with

$$\begin{aligned} y &= x^* + u_y; \\ u_y &= \rho u_x + \eta_1; \eta_1 \perp, \eta, z. \end{aligned}$$

Without loss of generality consider the univariate case and define $w_x = E(xf(z)|z); w_y = E(yf(z)|z)$. Then the system of convolution equations expands to

$$
\begin{cases}
f_{x^*} * f_u & & = w; \\
(1-\rho)h_{x^*} * f_u & +\rho z f(z) & = w_x; \\
(1-\rho^2)h_{x^*} * f_u & +\rho^2 z f(z) & = w_y.
\end{cases} \tag{3}
$$

The three equations have three unknown functions, f_{x^*}, f_u and ρ. Assuming that support of ρ does not include the point 1, ρ can be expressed as a solution to a linear algebraic equation derived from the two equations in (3) that include ρ :

$$
\rho = (w_x - zf(z))^{-1}(w_y - w_x).
$$

2.3 Regression models with classical and Berkson errors and the convolution equations

2.3.1 The list of models

The table below provides several regression models and the corresponding convolution equations involving density weighted conditional expectations.

Table 2. Regression models: 5. Regression with classical measurement error and an additional observation; 6. Regression with Berkson error (x, y, z are observable); 7. Regression with zero mean measurement error and Berkson instruments.

Model	Distributional assumptions	Convolution equations	Known functions	Unknown functions
5.	$y = g(x^*) + v$ $z = x^* + u;$ $x = x^* + u_x$ $x^* \perp u; E(u) = 0;$ $E(u_x \mid x^*, u) = 0;$ $E(v \mid x^*, u, u_x) = 0.$	$f_{x^*} * f_u = f_z;$ $(g f_{x^*}) * f_u = w,$ $h_k * f_u = w_k;$ with $h_k(x) \equiv x_k g(x) f_{x^*}(x);$ $k = 1, 2...d$	$f_z;\ w; w_k$	$f_{x^*};\ f_u;\ g.$
6.	$y = g(x) + v$ $z = x + u; E(v \mid z) = 0;$ $z \perp u; E(u) = 0.$	$f_x = f_{-u} * f_z;$ $g * f_{-u} = w$	$f_z; f_x, w$	$f_u;\ g.$
7.	$y = g(x^*) + v;$ $x = x^* + u_x;$ $z = x^* + u; z \perp u;$ $E(v \mid z, u, u_x) = 0;$ $E(u_x \mid z, v) = 0.$	$g * f_u = w;$ $h_k * f_u = w_k,$ with $h_k(x) \equiv x_k g(x);$ $k = 1, 2...d$	w, w_k	$f_u;\ g.$

Notes. Notation: $k = 1, 2...d$; in model 5.$w = E(y f_z(z) \mid z); w_k = E(x_k f_z(z) \mid z)$; in model 6. $w = E(y \mid z)$; in model 7. $w = E(y \mid z); w_k = E(x_k y \mid z)$.

Theorem 2. *Under Assumption 1 for each of the models 5-7 the corresponding convolution equations hold.*

The proof is in the derivations of the next subsection.

2.3.2 Discussion of the regression models and derivation of the

convolution equations.

5. The nonparametric regression model with classical measurement error and an additional observation.

This type of model was examined by Li (2002) and Li and Hsiao (2004); the convolution equations derived here provide a convenient representation. Often models of this type were considered in semiparametric settings. Butucea and Taupin (2008) (extending the earlier approach by Taupin, 2001) consider a regression function known up to a finite dimensional parameter with the mismeasured variable observed with independent error where the error distribution is known. Under the latter condition the model 5 here would reduce to the two first equations

$$f_{x^*} * f_u = f_z; \;\; (gf_{x^*}) * f_u = w,$$

where f_u is known and two unknown functions are g (here nonparametric) and f_{x^*}.

The model 5 incorporates model 3 for the regressor and thus the convolution equations from that model apply. An additional convolution equation is derived here; it is obtained from considering the value of the density weighted conditional expectation in the dual space of generalized functions, S^*, applied to arbitrary $\psi \in S$,

$$(w, \psi) = (E(f(z)y|z), \psi) = (E(f(z)g(x^*)|z), \psi);$$

this equals

$$
\begin{aligned}
& \int\int g(x^*) f_{x^*,z}(x^*,z)\psi(z)dx^*dz \\
= & \int\int g(x^*) f_{x^*,u}(x^*,u)\psi(x^*+u)dx^*du \\
= & \int g(x^*) f_{x^*}(x^*) f_u(u)dx^*\psi(x^*+u)dx^*du = ((gf_{x^*}) * f_u, \psi).
\end{aligned}
$$

Conditional moments for the regression function need not be integrable or bounded functions of z; we require them to be in the space of generalized functions S^*.

6. Regression with Berkson error.

This model may represent the situation when the regressor (observed) x is correlated with the error v, but z is a (vector) possibly representing an instrument uncorrelated with the regression error.

Then as is known in addition to the Berkson error convolution equation the equation

$$
w = E(y|z) = E(g(x)|z) = \int g(x)\frac{f_{x,z}(x,z)}{f_z(z)}dx = \int g(z-u)f_u(u)dx = g * f_u
$$

holds. This is stated in Meister (2008); however, the approach there is to consider g to be absolutely integrable so that convolution can be defined in the L_1 space. Here by working in the space of generalized functions S^* a much wider nonparametric class of functions that includes regression functions with polynomial growth is allowed.

7. Nonparametric regression with error in the regressor, where Berkson type instruments are assumed available.

This model was proposed by Newey (2001), examined in the univarite case by Schennach (2007) and Zinde-Walsh (2009), in the multivariate case in Zinde-Walsh (2012), where the convolution equations given here in Table 2 were derived.

2.4 Convolution equations in models with conditional independence conditions.

All the models 1-7 can be extended to include some additional variables where conditionally on those variables, the functions in the model (e.g. conditional distributions) are defined and all the model assumptions hold conditionally.

Evdokimov (2011) derived the conditional version of the model 4 from a very general nonparametric panel data model. Model 8 below describes the panel data set-up and how it transforms to conditional model 4 and 4a and possibly model 3 with relaxed independence condition (if the focus is on identifying the regression function).

Model 8. Panel data model with conditional independence.

Consider a two-period panel data model with an unknown regression function m and an idiosyncratic (unobserved) α :

$$\begin{aligned} Y_{i1} &= m(X_{i1}, \alpha_i) + U_{i1}; \\ Y_{i2} &= m(X_{i2}, \alpha_i) + U_{i2}. \end{aligned}$$

To be able to work with various conditional characteristic functions corresponding assumptions ensuring existence of the conditional distributions need to be made and in what follows we assume that all the conditional density functions and moments exist as generalized functions in S^*.

In Evdokimov (2011) independence (conditional on the corresponding period $X's$) of the regression error from α, and from the $X's$ and error of the other period is assumed:

$$f_t = f_{Uit}|_{X_{it},\alpha_i,X_{i(-t)},U_{i(-t)}}(u_t|x,...) = f_{Uit}|_{X_{it}}(u_t|x), t = 1, 2$$

with $f_{\cdot|\cdot}$ denoting corresponding conditional densities. Conditionally on $X_{i2} = X_{i1} = x$ the model takes the form 4

$$\begin{aligned} z &= x^* + u; \\ x &= x^* + u_x \end{aligned}$$

with z representing Y_1, x representing Y_2, x^* standing in for $m(x, \alpha)$, u for U_1 and u_x for U_2. The convolution equations derived here for 4 or 4a now apply to conditional densities.

The convolution equations in 4a are similar to Evdokimov; they allow for equations for f_u, f_{u_x} that do not rely on f_{x^*}. The advantage of those lies in the possibility of identifying the conditional error distributions without placing the usual non-zero restrictions on the characteristic function of x^* (that represents the function m for the panel model).

The panel model can be considered with relaxed independence assumptions. Here in the two-period model we look at forms of dependence that assume zero conditional mean of the second period error, rather than full independence of the first period error:

$$
\begin{aligned}
f_{Ui1|X_{i1},\alpha_i,X_{i2},Ui2}(u_t|x,...) &= f_{Ui1|Xi1}(u_t|x); \\
E(U_{i2}|X_{i1},\alpha_i,X_{i2},U_{i1}) &= 0; \\
f_{Ui2|\alpha_i,X_{i2}=X_{i1}=x}(u_t|x,...) &= f_{Ui2|Xi2}(u_t|x).
\end{aligned}
$$

Then the model maps into the model 3 with the functions in the convolution equations representing conditional densities and allows to identify distribution of x^* (function m in the model). But the conditional distribution of the second-period error in this set-up is not identified.

Evdokimov introduced parametric AR(1) or MA(1) dependence in the errors U and to accommodate that extended the model to three periods. Here this would lead in the AR case to the equations in (3) .

Model 9. Errors in variables regression with classical measurement error conditionally on covariates.

Consider the regression model

$$y = g(x^*, t) + v,$$

with a measurement of unobserved x^* given by $\tilde{z} = x^* + \tilde{u}$, with $x^* \perp \tilde{u}$ conditionally on t. Assume that $E(\tilde{u}|t) = 0$ and that $E(v|x^*, t) = 0$. Then redefining all the densities and conditional expectations to be conditional on t we get the same system of convolution equations as in Table 2 for model 5 with the unknown functions now being conditional densities and the regression function, g.

Conditioning requires assumptions that provide for existence of conditional distribution functions in S^*.

3 Solutions for the models.

3.1 Existence of solutions

To state results for nonparametric models it is important first to clearly indicate the classes of functions where the solution is sought. Assumption 1 requires that all the (generalized) functions considered are elements in the space of generalized functions S^*. This implies that in the equations the operation of convolution applied to the two functions from S^* provides an element in the space S^*. This subsection gives high level assumptions on the nonparametric classes of the unknown functions where the solutions can be sought: any functions from these classes that enter into the convolution provide a result in S^*.

No assumptions are needed for existence of convolution and full generality of identification conditions in models 1,2 where the model assumptions imply that the functions represent generalized densities. For the other models including regression models convolution is not always defined in S^*. Zinde-Walsh (2012) defines the concept of convolution pairs of classes of functions in S^* where convolution can be applied.

To solve the convolution equations a Fourier transform is usually employed, so that e.g. one transforms generalized density functions into characteristic functions. Fourier transform is an isomorphism of the space S^*. The Fourier transform of a generalized function $a \in S^*$, $Ft(a)$, is defined as follows. For any $\psi \in S$, as usual $Ft(\psi)(s) = \int \psi(x)e^{isx}dx$; then the functional $Ft(a)$ is defined by

$$(Ft(a), \psi) \equiv (a, Ft(\psi)).$$

The advantage of applying Fourier transform is that integral convolution equations transform into algebraic equations when the "exchange formula" applies:

$$a * b = c \Longleftrightarrow Ft(a) \cdot Ft(b) = Ft(c). \tag{4}$$

In the space of generalized functions S^*, the Fourier transform and inverse Fourier transform always exist. As shown in Zinde-Walsh (2012) there is a dichotomy between convolution pairs of subspaces in S^* and the corresponding product pairs of subspaces of their Fourier transforms.

The classical pairs of spaces (Schwartz, 1966) are the convolution pair (S^{*},O_{C}^{*}) and the corresponding product pair (S^{*},O_{M}), where O_{C}^{*} is the subspace of S^{*} that contains rapidly decreasing (faster than any polynomial) generalized functions and O_{M} is the space of infinitely differentiable functions with every derivative growing no faster than a polynomial at infinity. These pairs are important in that no restriction is placed on one of the generalized functions that could be any element of space S^{*}; the other belongs to a space that needs to be correspondingly restricted. A disadvantage of the classical pairs is that the restriction is fairly severe, for example, the requirement that a characteristic function be in O_{M} implies existence of all moments for the random variable. Relaxing this restriction would require placing constraints on the other space in the pair; Zinde-Walsh (2012) introduces some pairs that incorporate such trade-offs.

In some models the product of a function with a component of the vector of arguments is involved,such as $d(x)=x_{k}a(x)$, then for Fourier transforms $Ft(d)\,(s)=-i\frac{\partial}{\partial s_{k}}Ft(a)(s)$; the multiplication by a variable is transformed into $(-i)$ times the corresponding partial derivative. Since the differentiation operators are continuous in S^{*} this transformation does not present a problem.

Assumption 2. *The functions $a\in A,b\in B$, are such that (A,B) form a convolution pair in S^{*}.*

Equivalently, $Ft(a)$, $Ft(b)$ are in the corresponding product pair of spaces.

Assumption 2 is applied to model 1 for $a = f_{x^*}, b = f_u$; to model 2 with $a = f_z, b = f_u$; to model 3 with $a = f_{x^*}, b = f_u$ and with $a = h_k, b = f_u$, for all $k = 1, ..., d$; to model 4a for $a = f_{x^*}$, or f_{u_x}, or h_k for all k and $b = f_u$; to model 5 with $a = f_{x^*}$, or $g f_{x^*}$, or $h_k f_{x^*}$ and $b = f_u$; to model 6 with $a = f_z$, or g and $b = f_u$; to model 7 with $a = g$ or h_k and $b = f_u$.

Assumption 2 is a high-level assumption that is a sufficient condition for a solution to the models 1-4 and 6-7 to exist. Some additional conditions are needed for model 5 and are provided below.

Assumption 2 is automatically satisfied for generalized density functions, so is not needed for models 1 and 2. Denote by $\bar{D} \subset S^*$ the subset of generalized derivatives of distribution functions (corresponding to Borel probability measures in R^d) then in models 1 and 2 $A = B = \bar{D}$; and for the characteristic functions there are correspondingly no restrictions; denote the set of all characteristic functions, $Ft\left(\bar{D}\right) \subset S^*$, by $\bar{C}$.

Below a (non-exhaustive) list of nonparametric classes of generalized functions that provide sufficient conditions for existence of solutions to the models here is given. The classes are such that they provide minimal or often no restrictions on one of the functions and restrict the class of the other in order that the assumptions be satisfied.

In models 3 and 4 the functions h_k are transformed into derivatives of continuous characteristic functions. An assumption that either the characteristic function of x^* or the characteristic function of u be continuously differentiable is sufficient, without any restrictions on the other to ensure that Assumption 2 holds. Define the subset of all continuously differentiable characteristic functions by $\bar{C}^{(1)}$.

In model 5 equations involve a product of the regression function g with f_{x^*}. Products of generalized functions in S^* do not always exist and so additional restrictions are needed in that model. If g is an arbitrary element of S^*, then for the product to exist, f_{x^*} should be in O_M. On the other hand, if f_{x^*} is an arbitrary generalized density it is sufficient that g and h_k belong to the space of d times continuously differentiable functions with derivatives that are majorized by polynomial functions for $gf_{x^*}, h_k f_{x^*}$ to be elements of S^*. Indeed, the value of the functional $h_k f_{x^*}$ for an arbitrary $\psi \in S$ is defined by

$$(h_k f_{x^*}, \psi) = (-1)^d \int F_{x^*}(x) \partial^{(1,\dots,1)}(h_k(x)\psi(x))dx;$$

here F is the distribution (ordinary bounded) function and this integral exists because ψ and all its derivatives go to zero at infinity faster than any polynomial function. Denote by $\bar{S}^{B,1}$ the space of continuously differentiable functions $g \in S^*$ such that the functions $h_k(x) = x_k g(x)$ are also continuously differentiable with all derivatives majorized by polynomial functions. Since the products are in S^* then the Fourier transforms of the products are defined in S^*. Further restrictions requiring the Fourier transforms of the products gf_{x^*} and $h_k f_{x^*}$ to be continuously differentiable functions in S^* would remove any restrictions on f_u for the convolution to exist. Denote the space of all continuously differentiable functions in S^* by $\bar{S}^{(1)}$.

If g is an ordinary function that represents a regular element in S^* the infinite differentiability condition on f_{x^*} can be relaxed to simply requiring continuous first derivatives.

In models 6 and 7 if the generalized density function for the error, f_u, decreases faster than any polynomial (all moments need to exist for that), so that $f_u \in O_C^*$, then g could be any generalized function in S^*; this will of course hold if f_u has bounded support. Generally, the more moments the error is assumed to have, the fewer restrictions on the regression function g are needed to satisfy the convolution equations of the model and the exchange formula. The models 6, 7 satisfy the assumptions for any error u when support of generalized function g is compact (as for the "sum of peaks"), then $g \in E^* \subset S^*$, where E^* is the space of generalized functions with compact support. More generally the functions g and all the h_k could belong to the space O_C^* of generalized functions that decrease at infinity faster than any polynomial, and still no restrictions need to be placed on u.

Denote for any generalized density function $f_.$ the corresponding characteristic function, $Ft(f_.)$, by $\phi_.$. Denote Fourier transform of the (generalized) regression function g, $Ft(g)$, by γ.

The following table summarizes some fairly general sufficient conditions on the models that place restrictions on the functions themselves or on the characteristic functions of distributions in the models that will ensure that Assumption 2 is satisfied and a solution exists. The nature of these assumptions is to provide restrictions on some of the functions that allow the others to be completely unrestricted for the corresponding model.

Table 3. Some nonparametric classes of generalized functions for which the convolution equations of the models are defined in S^*.

Model	Sufficient	assumptions
1	no restrictions:	$\phi_{x^*} \in \bar{C}; \phi_u \in \bar{C}$
2	no restrictions:	$\phi_{x^*} \in \bar{C}; \phi_u \in \bar{C}$
	Assumptions A	Assumptions B
3	any $\phi_{x^*} \in \bar{C}; \phi_u \in \bar{C}^{(1)}$	any $\phi_u \in \bar{C}; \phi_{x^*} \in \bar{C}^{(1)}$
4	any $\phi_{u_x}, \phi_{x^*} \in \bar{C}; \phi_u \in \bar{C}^{(1)}$	any $\phi_u, \phi_{x^*} \in \bar{C}; \phi_{u_x} \in \bar{C}^{(1)}$
4a	any $\phi_{u_x}, \phi_{x^*} \in \bar{C}; \phi_u \in \bar{C}^{(1)}$	any $\phi_u, \phi_{u_x} \in \bar{C}; \phi_{x^*} \in \bar{C}^{(1)}$
5	any $g \in S^*; f_{x^*} \in O_M; f_u \in O_C^*$	any $f_{x^*} \in \bar{D};\ g, h_k \in \bar{S}^{B,1}; f_u \in O_C^*$
6	any $g \in S^*; f_u \in O_C^*$	$g \in O_C^*$; any $f_u : \phi_u \in \bar{C}$
7	any $g \in S^*; f_u \in O_C^*$	$g \in O_C^*$; any $f_u : \phi_u \in \bar{C}$

The next table states the equations and systems of equations for Fourier transforms that follow from the convolution equations.

Table 4. The form of the equations for the Fourier transforms:

Model	Eq's for Fourier transforms	Unknown functions
1	$\phi_{x^*}\phi_u = \phi_z;$	ϕ_{x^*}
2	$\phi_{x^*} = \phi_z\phi_{-u};$	ϕ_{x^*}
3	$\begin{cases} \phi_{x^*}\phi_u = \phi_z; \\ (\phi_{x^*})'_k \phi_u = \varepsilon_k, k = 1, ..., d. \end{cases}$	ϕ_{x^*}, ϕ_u
4	$\begin{cases} \phi_{x^*}\phi_u = \phi_z; \\ (\phi_{x^*})'_k \phi_u = \varepsilon_k, k = 1, ..., d; \\ \phi_{x^*}\phi_{u_x} = \phi_x. \end{cases}$	$\phi_{x^*}, \phi_u, \phi_{u_x}$
4a	$\begin{cases} \phi_{u_x}\phi_u = \phi_{z-x}; \\ \left(\phi_{u_x}\right)'_k \phi_u = \varepsilon_k, k = 1, ..., d. \\ \phi_{x^*}\phi_{u_x} = \phi_x. \end{cases}$	–"–
5	$\begin{cases} \phi_{x^*}\phi_u = \phi_z; \\ Ft\left(gf_{x^*}\right)\phi_u = \varepsilon \\ \left(Ft\left(gf_{x^*}\right)\right)'_k \phi_u = \varepsilon_k, k = 1, ..., d. \end{cases}$	ϕ_{x^*}, ϕ_u, g
6	$\begin{cases} \phi_x = \phi_{-u}\phi_z; \\ Ft(g)\phi_{-u} = \varepsilon. \end{cases}$	ϕ_u, g
7	$\begin{cases} Ft(g)\phi_u = \varepsilon; \\ \left(Ft\left(g\right)\right)'_k \phi_u = \varepsilon_k, k = 1, ..., d. \end{cases}$	ϕ_u, g

Notes. Notation $(\cdot)'_k$ denotes the k-th partial derivative of the function. The functions ε are Fourier transforms of the corresponding w, and $\varepsilon_k = -iFt(w_k)$ defined for the models in Tables 1 and 2.

Assumption 2 (that is fulfilled e.g. by generalized functions classes of

Table 3) ensures existence of solutions to the convolution equations for models 1-7; this does not exclude multiple solutions and the next section provides a discussion of solutions for equations in Table 4.

3.2 Classes of solutions; support and multiplicity of solutions

Typically, support assumptions are required to restrict multiplicity of solutions; here we examine the dependence of solutions on the support of the functions. The results here also give conditions under which some zeros, e.g. in the characteristic functions, are allowed. Thus in common with e.g. Carrasco and Florens (2010), Evdokimov and White (2011), distributions such as the uniform or triangular for which the characteristic function has isolated zeros are not excluded. The difference here is the extension of the consideration of the solutions to S^* and to models such as the regression model where this approach to relaxing support assumptions was not previously considered.

Recall that for a continuous function $\psi(x)$ on R^d support is defined as the set $W =$supp(ψ), such that

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

Support of a continuous function is an open set.

Generalized functions are functionals on the space S and support of a generalized function $b \in S^*$ is defined as follows (Schwartz, 1967, p. 28). Denote by (b, ψ) the value of the functional b for $\psi \in S$. Define a null set for $b \in S^*$ as the union of supports of all functions in S for which the value of the functional is zero: $\Omega = \{\cup \text{supp}(\psi)\,,\ \psi \in S, \text{ such that } (b, \psi) = 0\}$. 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.

Note that for model 2 Table 4 gives the solution for ϕ_{x^*} directly and the inverse Fourier transform can provide the (generalized) density function, f_{x^*}.

In Zinde-Walsh (2012) identification conditions in S^* were given for models 1 and 7 under assumptions that include the ones in Table 3 but could also be more flexible.

The equations in Table 3 for models 1,3, 4, 4a, 5, 6 and 7 are of two types, similar to those solved in Zinde-Walsh (2012). One is a convolution with one unknown function; the other is a system of equations with two unknown functions, each leading to the corresponding equations for their Fourier transforms.

3.2.1 Solutions to the equation $\alpha\beta = \gamma$.

Consider the equation

$$\alpha\beta = \gamma, \tag{5}$$

with one unknown function α; β is a given continuous function. By assumption 2 the non-parametric class for α is such that the equation holds in S^* on

R^d; it is also possible to consider a nonparametric class for α with restricted support, $\bar{W}$. Of course without any restrictions $\bar{W} = R^d$. Recall the differentiation operator, ∂^m, for $m = (m_1, ...m_d)$ and denote by $supp(\beta, \partial)$ the set $\cup_{\Sigma m_i=0}^{\infty} supp(\partial^m \beta)$; where $supp(\partial^m \beta)$ is an open set where a continuous derivative $\partial^m \beta$ exists. Any point where β is zero belongs to this set if some finite-order partial continuous derivative of β is not zero at that point (and in some open neighborhood); for β itself $supp(\beta) \equiv supp(\beta, 0)$.

Define the functions

$$\alpha_1 = \beta^{-1}\gamma I\left(supp(\beta, \partial)\right); \alpha_2(x) = \begin{cases} 1 & \text{for } x \in supp(\beta, \partial); \\ \tilde{\alpha} & \text{for } x \in \bar{W} \backslash (supp(\beta, \partial)) \end{cases} \tag{6}$$

with any $\tilde{\alpha}$ such that $\alpha_1\alpha_2 \in Ft\left(A\right)$.

Consider the case when α, β and thus γ are continuous. For any point x_0 if $\beta(x_0) \neq 0$, there is a neighborhood $N(x_0)$ where $\beta \neq 0$, and division by β is possible. If $\beta(x_0)$ has a zero, it could only be of finite order and in some neighborhood, $N(x_0) \in supp(\partial^m \beta)$ a representation

$$\beta = \eta(x)\Pi_{i=1}^{d}\left(x_i - x_{0i}\right)^{m_i} \tag{7}$$

holds for some continuous function η in S^*, such that $\eta > c_\eta > 0$ on $supp(\eta)$.Then $\eta^{-1}\gamma$ in $N(x_0)$ is a non-zero continuous function; division of such a function by $\Pi_{i=1}^{d}\left(x_i - x_{0i}\right)^{m_i}$ in S^* is defined (Schwartz, 1967, pp. 125-126), thus division by β is defined in this neighborhood $N(x_0)$. For the set $supp(\beta, \partial)$ consider a covering of every point by such neighborhoods, the possibility of division in each neighborhood leads to possibility of division globally on the whole $supp(\beta, \partial)$. Then a_1 as defined in (6) exists in S^*.

In the case where γ is an arbitrary generalized function, if β is infinitely differentiable then then by (Schwartz, 1967, pp.126-127) division by β is defined on $supp(\beta, \partial)$ and the solution is given by (6).

For the cases where γ is not continuous and β is not infinitely differentiable the solution is provided by

$$\alpha_1 = \beta^{-1}\gamma I\left(supp(\beta, 0)\right); \alpha_2(x) = \begin{cases} 1 & \text{for } x \in supp(\beta, 0); \\ \tilde{\alpha} & \text{for } x \in \bar{W} \backslash (supp(\beta, 0)) \end{cases}$$

with any $\tilde{\alpha}$ such that $\alpha_1\alpha_2 \in Ft\left(A\right)$.

Theorem 2 in Zinde-Walsh (2012) implies that the solution to (5) is $a = Ft^{-1}(\alpha_1\alpha_2)$; the sufficient condition for the solution to be unique is $supp(\beta, 0) \supset \bar{W}$; if additionally either γ is a continuous function or β is an infinitely continuously differentiable function it is sufficient for uniqueness that $supp(\beta, \partial) \supset \bar{W}$.

This provides solutions for models 1 and 6 where only equations of this type appear.

3.2.2 Solutions to the system of equations

For models 3,4,5 and 7 a system of equations of the form

$$\begin{aligned} \alpha\beta &= \gamma; \\ \alpha\beta_k' &= \gamma_k, \\ k &= 1,...,d. \end{aligned} \tag{8}$$

(with β continuously differentiable) arises. Theorem 3 in Zinde-Walsh (2012) provides the solution and uniqueness conditions for this system of equations. It is first established that a set of continuous functions $\varkappa_k, k = 1,...,d$, that solves the equation

$$\varkappa_k\gamma - \gamma_k = 0 \tag{9}$$

in the space S^* exists and is unique on $W = supp(\gamma)$ as long as $supp(\beta) \supset W$. Then $\beta_k'\beta^{-1} = \varkappa_k$ and substitution into (9) leads to a system of first-order differential equations in β.

Case 1. Continuous functions; W is an open set.

For the models 3 and 4 the system (8) involves continuous characteristic functions thus there W is an open set. In some cases W can be an open set under conditions of models 5 and 7, e.g. if the regression function is integrable in model 7.

For this case represent the open set W as a union of (maximal) connected components $\cup_v W_v$.

Then by the same arguments as in the proof of Theorem 3 in Zinde-

Walsh (2012) the solution can be given uniquely on W as long as at some point $\zeta_{0v} \in (W_v \cap W)$ the value $\beta\left(\zeta_{0\nu}\right)$ is known for each of the connected components . Consider then $\beta_1(\zeta) = \Sigma_\nu[\beta\left(\zeta_{0\nu}\right)\exp\int_{\zeta_0}^{\zeta}\sum_{k=1}^{d}\varkappa_k(\xi)d\xi]I(W_\nu)$, where integration is along any arc within the component that connects ζ to $\zeta_{0\nu}$. Then $\alpha_1 = \beta_1^{-1}\gamma$, and α_2, β_2 are defined as above by being 1 on $\cup_v W_v$ and arbitrary outside of this set.

When $\beta(0) = 1$ as is the case for the characteristic function, the function is uniquely determined on the connected component that includes 0.

Evdokimov and White (2012) provide a construction that permits in the univariate case to extend the solution $\beta\left(\zeta_{0\nu}\right)[\exp\int_{\zeta_0}^{\zeta}\sum_{k=1}^{d}\varkappa_k(\xi)d\xi]I(W_\nu)$ from a connected component of support where $\beta\left(\zeta_{0\nu}\right)$ is known (e.g. at 0 for a characteristic function) to a contiguous connected component when on the border between the two where $\beta = 0$, at least some finite order derivative of β is not zero. In the multivariate case this approach can be extended to the same construction along a one-dimensional arc from one connected component to the other. Thus identification is possible on a connected component of $supp(\beta, \partial)$.

Case 2. W is a closed set.

Generally for models 5 and 7, W is the support of a generalized function and is a closed set. It may intersect with several connected components of support of β. Denote by W_v here the intersection of a connected component of support of β and W. Then similarly $\beta_1(\zeta) = \sum_\nu[\beta\left(\zeta_{0\nu}\right)\exp\int_{\zeta_0}^{\zeta}\sum_{k=1}^{d}\varkappa_k(\xi)d\xi]I(W_\nu)$, where integration is along any arc within the component that connects ζ to

$\zeta_{0\nu}$. Then $\alpha_1 = \beta_1^{-1}\varepsilon$, and α_2, β_2 are defined as above by being 1 on $\cup_v W_v$ and arbitrary outside of this set. The issue of the value of β at some point within each connected component arises. In the case of β being a characteristic function if there is only one connected component, W and $0 \in W$ the solution is unique, since then $\beta(0) = 1$.

Note that for model 5 the solution to equations of the type (8) would only provide $Ft(gf_{x^*})$ and ϕ_u; then from the first equation for this model in Table 4 ϕ_{x^*} can be obtained; it is unique if suppϕ_{x^*} =suppϕ_z. To solve for g find $g = Ft^{-1}\left(Ft\left(gf_{x^*}\right)\right) \cdot \left(f_{x^*}\right)^{-1}$.

4 Identification, partial identification and well-posedness

4.1 Identified solutions for the models 1-7

As follows from the discussion of the solutions uniqueness in models 1,2,3,4,4a,5,6 holds (in a few cases up to a value of a function at a point) if all the Fourier transforms are supported over the whole R^d; in many cases it is sufficient that $supp(\beta, \partial) = R^d$.

The classes of functions could be defined with Fourier transforms supported on some known subset $\bar{W}$ of R^d, rather than on the whole space; if all the functions considered have $\bar{W}$ as their support, and the support consists of one connected component that includes 0 as an interior point then identification for the solutions holds. For the next table assume that $\bar{W}$ is a single connected component with 0 as an interior point; again $\bar{W}$ could coincide with $supp(\beta, \partial)$. For model 5 under Assumption B assume additionally that the value at zero: $Ft(gf_{x^*})(0)$ is known; similarly for model 7 under assumption B additionally assume that $Ft(g)(0)$ is known.

Table 5. The solutions for identified models on $\bar{W}$.

Model	Solution to equations
1.	$f_{x^*} = Ft^{-1}\left(\phi_u^{-1}\phi_z\right).$
2.	$f_{x^*} = Ft^{-1}\left(\phi_{-u}\phi_z\right).$
3.	Under Assumption A $f_{x^*} = Ft^{-1}(\exp\int_{\zeta_0}^{\zeta}\sum\limits_{k=1}^{d}\varkappa_k(\xi)d\xi),$ where $\varkappa_k$ solves $\varkappa_k\phi_z - [(\phi_z)'_k - \varepsilon_k] = 0;$ $f_u = Ft^{-1}(\phi_{x^*}^{-1}\varepsilon).$ Under Assumption B $f_u = Ft^{-1}(\exp\int_{\zeta_0}^{\zeta}\sum\limits_{k=1}^{d}\varkappa_k(\xi)d\xi);$ $\varkappa_k$ solves $\varkappa_k\phi_z - \varepsilon_k = 0;$ $f_{x^*} = Ft^{-1}(\phi_u^{-1}\varepsilon).$
4	f_{x^*}, f_u obtained similarly to those in 3.; $\phi_{u_x} = \phi_{x^*}^{-1}\phi_x.$
4a.	f_{u_x}, f_u obtained similarly to ϕ_{x^*}, ϕ_u in 3.; $\phi_{x^*} = \phi_{u_x}^{-1}\phi_x.$
5.	Three steps: 1. (a) Get $Ft(gf_{x^*}), \phi_u$ similarly to ϕ_{x^*}, ϕ_u in model 3 (under Assumption A use $Ft(gf_{x^*})(0)$); 2. Obtain $\phi_{x^*} = \phi_u^{-1}\phi_z;$ 3. Get $g = \left[Ft^{-1}\left(\phi_{x^*}\right)\right]^{-1} Ft^{-1}(Ft(gf_{x^*})).$
6.	$\phi_{-u} = \phi_z^{-1}\phi_x$ and $g = Ft^{-1}(\phi_x^{-1}\phi_z\varepsilon).$
7.	$\phi_{x^*}, Ft(g)$obtained similarly to ϕ_{x^*}, ϕ_uin 3 (under Assumption A use $Ft(g)(0)$).

4.2 Implications of partial identification.

Consider the case of Model 1. Essentially lack of identification, say in the case when the error distribution has characteristic function supported on a convex domain W_u around zero results in the solution for $\phi_{x^*} = \phi_1\phi_2$, with ϕ_1 non-zero and unique on W_u, and thus captures the lower-frequency components of x^*, and with ϕ_2 is a characteristic function of a distribution with arbitrary high frequency components. Transforming back to densities provides a corresponding model with independent components

$$z = x_1^* + x_2^* + u,$$

where x_1^* uniquely extracts the lower frequency part of observed z. The more important the contribution of x_1^* to x^* the less important is lack of identification.

If the feature of interest as discussed e.g. by Matzkin (2007) involves only low frequency components of x^*, it may still be fully identified even when the distribution for x^* itself is not. An example of that is a deconvolution applied to an image of a car captured by a traffic camera; although even after deconvolution the image may still appear blurry the licence plate number may be clearly visible. In nonparametric regression the polynomial growth of the regression or the expectation of the response function may be identifiable even if the regression function is not fully identified.

Features that are identified include any functional, Φ, linear or non-linear on a class of functions of interest, such that in the frequency domain Φ is supported on W_u.

4.3 Well-posedness in S^*

Conditions for well-posedness in S^* for solutions of the equations entering in models 1-7 were established in Zinde-Walsh (2012). Well-posedness is needed to ensure that if a sequence of functions converges (in the topology of S^*) to the known functions of the equations characterizing the models 1-7 in tables 1 and 2, then the corresponding sequence of solutions will converge to the solution for the limit functions. A feature of well-posedness in S^* is that the solutions are considered in a class of functions that is a bounded set in S^*.

The properties that differentiation is a continuous operation, and that the Fourier transform is an isomorphism of the topological space S^*, make conditions for convergence in this space much weaker than those in functions spaces, say, L_1, L_2. Thus for density that is given by the generalized derivative of the distribution function well-posedness holds in spaces of generalized functions by the continuity of the differentiation operator.

For the problems here however, well-posedness does not always obtain. The main sufficient condition is that the inverse of the characteristic function of the measurement error satisfy the condition (1) with $b = \phi_u^{-1}$ on the corresponding support. This holds if either the support is bounded or if the distribution is not super-smooth. If ϕ_u has some zeros but satisfies the identification conditions so that it has local representation (7) where (1) is satisfied for $b = \eta^{-1}$ well-posedness will hold.

Example in Zinde-Walsh (2012) demonstrates that well-posedness of deconvolution will not hold even in the weak topology of S^* for super-smooth (e.g. Gaussian) distributions on unbounded support. On the other hand, well-posedness of deconvolution in S^* obtains for ordinary smooth distributions and thus under less restrictive conditions than in function spaces, such as L_1 or L_2 usually considered.

In the models 3-7 with several unknown functions, more conditions are required to ensure that all the operations by which the solutions are obtained are continuous in the topology of S^*. It may not be sufficient to assume (1) for the inverses of unknown functions where the solution requires division; for continuity of the solution the condition may need to apply uniformly.

Define a class of ordinary functions on R^d, $\Phi(m, V)$ (with m a vector of integers, V a positive constant) where $b \in \Phi(m, V)$ if

$$\int \Pi \left((1 + t_i^2)^{-1} \right)^{m_i} |b(t)| \, dt < V < \infty. \tag{10}$$

Then in Zinde-Walsh (2012) well-posedness is proved for model 7 as long as in addition to Assumption A or B, for some $\Phi(m, V)$ both ϕ_u and ϕ_u^{-1} belong to the class $\Phi(m, V)$. This condition is fulfilled by non-supersmooth ϕ_u; this could be an ordinary smooth distribution or a mixture with some mass point.

A convenient way of imposing well-posedness is to restrict the support of functions considered to a bounded $\bar{W}$. If the features of interest are associated with low-frequency components only, then if the functions are restricted to a bounded space the low-frequency part can be identified and is well-posed.

5 Implications for estimation

5.1 Plug-in non-parametric estimation

Solutions in Table 5 for the equations that express the unknown functions via known functions of observables give scope for plug-in estimation. As seen e.g. in the example of Model 4, 4 and 4a are different expressions that will provide different plug-in estimators for the same functions.

The functions of the observables here are characteristic functions and Fourier transforms of density-weighted conditional expectations and in some cases their derivatives, that can be estimated by non-parametric methods. There are some direct estimators, e.g. for characteristic functions. In the space S^* the Fourier transform and inverse Fourier transform are continuous operations thus using standard estimators of density weighted expectations and applying the Fourier transform would provide consistency in S^*; the details are provided in Zinde-Walsh (2012). Then the solutions can be expressed via those estimators by the operations from Table 5 and, as long as the problem is well-posed, the estimators will be consistent and the convergence will obtain at the appropriate rate. As in An and Hu (2012), the convergence rate may be even faster for well-posed problems in S^* than the usual nonparametric rate in (ordinary) function spaces. For example, as demonstrated in Zinde-Walsh (2008) kernel estimators of density that may diverge if the distribution function is not absolutely continuous, are always (under the usual assumptions on kernel/bandwidth) consistent in the weak topology of the space of generalized functions, where the density problem is well-posed. Here, well-posedness holds for deconvolution as long as the error density is not super-smooth.

5.2 Regularization in plug-in estimation

When well-posedness cannot be ensured, plug-in estimation will not provide consistent results and some regularization is required; usually spectral cut-off is employed for the problems considered here. In the context of these non-parametric models regularization requires extra information: the knowledge of the rate of decay of the Fourier transform of some of the functions.

For model 1 this is not a problem since ϕ_u is assumed known; the regularization uses the information about the decay of this characteristic function to construct a sequence of compactly supported solutions with support increasing at a corresponding rate. In S^* no regularization is required for plug-in estimation unless the error distribution is super-smooth. Exponential growth in ϕ_u^{-1} provides a logarithmic rate of convergence in function classes for the estimator (Fan, 1991). Below we examine spectral cut-off regularization for the deconvolution in S^* when the error density is super-smooth.

With super-smooth error in S^* define a class of generalized functions

$\Phi(\Lambda, m, V)$ for some non-negative-valued function Λ; a generalized function $b \in \Phi(\Lambda, m, V)$ if there exists a function $\bar{b}(\zeta) \in \Phi(m, V)$ such that also $\bar{b}(\zeta)^{-1} \in \Phi(m, V)$ and $b = \bar{b}(\zeta) \exp(-\Lambda(\zeta))$. Note that a linear combination of functions in $\Phi(\Lambda, m, V)$ belongs to the same class. Define convergence: a sequence of $b_n \in \Phi(\Lambda, m, V)$ converges to zero if the corresponding sequence $\bar{b}_n$ converges to zero in S^*.

Convergence in probability for a sequence of random functions, ε_n, in S^* is defined as follows: $(\varepsilon_n - \varepsilon) \to_p 0$ in S^* if for any set $\psi_1, ..., \psi_v \in S$ the random vector of the values of the functionals converges: $((\varepsilon_n - \varepsilon, \psi_1), ..., (\varepsilon_n - \varepsilon, \psi_v)) \to_p 0$.

Lemma 2. *If in model 1 $\phi_u = b \in \Phi(\Lambda, m, V)$, where Λ is a polynomial function of order no more than k, and ε_n is a sequence of estimators of ε that are consistent in $S^* : r_n(\varepsilon_n - \varepsilon) \to_p 0$ in S^* at some rate $r_n \to \infty$, then for any sequence of constants $\bar{B}_n : 0 < \bar{B}_n < (\ln r_n)^{\frac{1}{k}}$ and the corresponding set $B_n = \left\{\zeta : \|\zeta\| < \bar{B}_n\right\}$ the sequence of regularized estimators $\phi_u^{-1}(\varepsilon_n - \varepsilon) I(B_n)$ converges to zero in probability in S^*.*

Proof. For n the value of the random functional

$$(\phi_u^{-1}(\varepsilon_n - \varepsilon) I(B_n), \psi) = \int \bar{b}^{-1}(\zeta) r_n(\varepsilon_n - \varepsilon) r_n^{-1} I(B_n) \exp(\Lambda(\zeta)) \psi(\zeta) d\zeta.$$

Multiplication by $\bar{b}^{-1} \in \Phi(m, V)$, that corresponds to $\phi_u = b$ does not affect convergence thus $\bar{b}^{-1}(\zeta) r_n(\varepsilon_n - \varepsilon)$ converges to zero in probability in S^*. To show that $(\phi_u^{-1}(\varepsilon_n - \varepsilon) I(B_n), \psi)$ converges to zero it is sufficient to show that the function $r_n^{-1}I(B_n)\exp(\Lambda(\zeta))\,\psi(\zeta)$ is bounded. It is then sufficient to find B_n such that $r_n^{-1}I(B_n)\exp(\Lambda(\zeta))$ is bounded (by possibly a polynomial), thus it is sufficient that $\sup_{B_n}|\exp(\Lambda(\zeta))\,r_n^{-1}|$ be bounded. This will hold if $\exp\left(\bar{B}_n^k\right) < r_n$, $\bar{B}_n^k < \ln r_n$.■

Of course an even slower growth for spectral cut-off would result from Λ that grows faster than a polynomial. The consequence of the slow growth of the support is usually a correspondingly slow rate of convergence for $\phi_u^{-1}\varepsilon_n I(B_n)$. Additional conditions (as in function spaces) are needed for the regularized estimators to converge to the true γ.

It may be advantageous to focus on lower frequency components and ignore the contribution from high frequencies when the features of interest depend on the contribution at low frequency.

6 Concluding remarks

Working in spaces of generalized functions extends the results on nonparametric identification and well-posedness for a wide class of models. Here identification in deconvolution is extended to generalized densities in the class of all distributions from the usually considered classes of integrable density functions. In regression with Berkson error nonparametric identification in S^* holds for functions of polynomial growth, extending the usual results obtained in L_1; a similar extension applies to regression with measurement error and Berkson type measurement; this allows to consider binary choice and polynomial regression models. Also, identification in models with sum-of-peaks regression function that cannot be represented in function spaces is included. Well-posedness results in S^* also extend the results in the literature provided in function spaces; well-posedness of deconvolution holds as long as the characteristic function of the error distribution does not go to zero at infinity too fast (as e.g. super-smooth) and a similar condition provides well-posedness in the other models considered here.

Further investigation of the properties of estimators in spaces of generalized functions requires deriving the generalized limit process for the function being estimated and investigating when it can be described as a generalized Gaussian process. A generalized Gaussian limit process holds for kernel estimator of the generalized density function (Zinde-Walsh, 2008). Determining the properties of inference based on the limit process for generalized random functions requires both further theoretical development and simulations evidence.

References

[1] An, Y, and Y. Hu (2012) Well-posedness of measurement error models for self-reported data, Journal of Econometrics, vol. 168 (2012), pp. 259-269.

[2] Ben-Moshe, D. (2012) Identification of Dependent Multidimensional Unobserved Variables in a System of Linear Equations, working paper,

UCLA.

[3] Bonhomme, S. and J-M. Robin (2010) Generalized nonparametric deconvolution with an application to earnings dynamics, Review of Economic Studies, 77, pp. 491-533.

[4] Butucea, C. and M.L.Taupin (2008) New M-estimators in semiparametric regression with errors in variables, Annales de l'Institut Henri Poincaré - Probabilités et Statistiques,V. 44, pp. 393–421.

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

[6] Carrasco, M. and J.-P. Florens (2010), A Spectral Method for Deconvolving a Density, Econometric Theory, 27 , pp 546-581.

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

[8] Chen, X., Hong, H. and Tamer, E. (2005), Measurement Error Models with Auxiliary Data, Review of Economic Studies, 72, pp. 343–366.

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

[10] Evdokimov, K. (2011), Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity, working paper.

[11] Evdokimov, K. and H. White (2011) An Extension of a Lemma of Kotlyarski, Econometric Theory, forthcoming.

[12] Fan, J. Q. (1991), On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems, Annals of Statistics, 19, pp. 1257-1272.

[13] Green, D.A. and W.C. Riddell (1997) Qualifying for unemployment insurance: An empirical analysis, Economic Journal, 107, pp. 67-84.

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

[15] Hu, Y. (2008), Identification and Estimation of Nonlinear Models with Misclassification Error using Instrumental Variables: A General Solution, Journal of Econometrics, 144, pp. 27-61.

[16] Hu, Y. and G. Ridder (2010), On Deconvolution as a First Stage Nonparametric Estimator, Econometric Reviews, 29, pp. 1-32.

[17] Karlin, S. (1959) The Theory of Infinite Games, v.II, Addison-Wesley Publ.

[18] Kotlyarski, I. (1967), On Characterizing the Gamma and Normal Distribution, Pacific Journal of Mathematics, 20, pp. 69-76.

[19] Li., T. (2002) Robust and consistent estimation of nonlinear errors-in-variables models, Journal of Econometrics, 110, pp.1-26.

[20] Li, T. and Ch. Hsiao (2004) Robust estimation of generalized models with measurement error, Journal of Econometrics, 118, pp. 51-65.

[21] Li, T. and Vuong, Q. (1998), Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators, Journal of Multivariate Analysis, 65, 139-165.

[22] Mahajan, A. (2006) Identification and estimation of regression models with misclassification, Econometrica, 74, pp. 631-665.

[23] Matzkin, R.L. (2007) Nonparametric Identification, Chapter 73 in Handbook of Econometrics, Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

[24] A.Meister (2009) Deconvolution problems in nonparametric statistics, Lecture notes in statistics, Springer-Verlag.

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

[26] Schennach, S. (2004) Nonparametric regression in the presence of measurement error, Econometric Theory, 20, pp. 1046-1093.

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

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

[29] Taupin, M.-L. (2001) Semi-parametric estimation in the nonlinear structural errors-in-variables model. Ann. Statist., 29, pp. 66–93.

[30] Wang, L. Estimation of nonlinear models with Berkson measurement errors, Annals of Statistics, 32, pp. 2559-2579.

[31] Zinde-Walsh, V. (2008) Kernel estimation when density may not exist, Econometric Theory, v.24, 696-725.

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

[33] Zinde-Walsh, V. (2011), Presidential Address: Mathematics in economics and econometrics, Canadian Journal of Economics, v.44, pp. 1052-1068.

[34] Zinde-Walsh, V. (2012), Measurement error and deconvolution in spaces of generalized functions, arXiv:1009.4217v2 [math.ST]; earlier version (2010): arXiv:1009.4217v1[MATH.ST].