Errors-in-variables models: a generalized functions approach

Victoria Zinde-Walsh*
McGill University and CIREQ

Revised version of September 29, 2009

Abstract

Identification in errors-in-variables regression models was recently extended to wide models classes by S. Schennach (Econometrica, 2007) (S) via use of generalized functions. In this paper the problems of non- and semi- parametric identification in such models are re-examined. Nonparametric identification holds under weaker assumptions than in (S); the proof here does not rely on decomposition of generalized functions into ordinary and singular parts, which may not hold. Conditions for continuity of the identification mapping are provided and a consistent nonparametric plug-in estimator for regression functions in the L_1 space constructed. Semiparametric identification via a finite set of moments is shown to hold for classes of functions that are

*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. The author thanks participants in the Cowles Foundation conference, the UK ESG, CEA 2009, Stats in the Chateau and CESG meetings and P.C.B.Phillips, X. Chen, D. Nekipelov, L.Wang and T.Wootersen for illuminating discussions and suggestions. An anonymous referee provided a very thorough report containing many important points and suggestions. The remaining errors are all mine.

explicitly characterized; unlike (S) existence of a moment generating function for the measurement error is not required.

Keywords: errors-in-variables model, generalized functions

1 Introduction

The familiar errors in variables model with an unknown regression function, g, and measurement error in the scalar variable has the form

$$\begin{aligned} Y &= g(X^*) + \Delta Y; \\ X &= X^* + \Delta X, \end{aligned}$$

where variables X and Y are observable; X^* and $\Delta X, \Delta Y$ are not observable. A widely used approach makes use of instrumental variables. Suppose that instruments are available and Z represents an identified projection of X on the instruments so that additionally $X^* = Z - U$; assume that U is independent of Z (Berkson-type error from using the instruments) and that

$$\begin{aligned} E[\Delta Y \,|Z, U] &= 0; \\ E[\Delta X \,|Z, U, \Delta Y] &= 0; \\ E(U) &= 0. \end{aligned}$$

These assumptions were made by e.g. Hausman et al. (1991) who examined polynomial regression. Newey (2001) added another moment condition for estimation in semiparametric regression leading to two equations for unknown g and F, the measurement error distribution (all integrals over $(-\infty, \infty)$) :

$$\begin{aligned} E(Y|Z &= z) = \int g(z-u)dF(u); \\ E(YX|Z &= z) = \int (z-u)g(z-u)dF(u). \end{aligned}$$

Define $W_y(z) \equiv E(Y|Z=z); W_{xy}(z) \equiv E(Y(Z-X)|Z=z)$. The model assumptions then can be considered in terms of classes of functions W_y, W_{xy}, g and distribution F that satisfy the equations:

$$\begin{aligned} W_y(z) &= \int g(z-u)dF(u); \\ W_{xy}(z) &= \int (z-u)g(z-u)dF(u); \end{aligned} \tag{1}$$

we say that these functions W_y, W_{xy}, g , F satisfy model assumptions. The functions g and F enter in convolutions; this motivates using Fourier transforms (Ft). Fourier transforms:

$$\begin{aligned} \varepsilon_y(\zeta) &= Ft(W_y(\cdot)); \\ \varepsilon_{xy}(\zeta) &= Ft(W_{xy}(\cdot)); \\ \gamma(\zeta) &= Ft(g(\cdot)); \end{aligned} \tag{2}$$

the characteristic function is obtained as $\phi(\zeta) = \int e^{i\zeta u}dF(u)$.

Provided that for some subclass of functions Fourier transforms are well defined, derivatives exist and the convolution theorem applies, (1) is equivalent to a system with two unknown functions, γ, ϕ :

$$\varepsilon_y(\zeta) = \gamma(\zeta)\phi(\zeta); \tag{3}$$

$$i\varepsilon_{xy}(\zeta) = \dot{\gamma}(\zeta)\phi(\zeta), \tag{4}$$

where $\dot{\gamma} = \frac{d\gamma}{d\zeta}$. S. Schennach (2007) (S) suggested that these equations can be justified for a wide class of functions if one uses generalized functions, specifically, those in the space of tempered distributions, T' (defined below in section 2.1)[1].

Assumption 1. The functions g, W_y, W_{xy} that satisfy the model as-

[1] A referee pointed out that the usual notation for the space of tempered distributions is S', but here we follow the notation in (S).

sumptions are such that each represents an element in the space of tempered distributions, T'.

Some examples of such functions are the class considered in (S, Assumption 1): functions such that $|g(x^*)|, |W_y(z)|, |W_{xy}(z)|$ are defined and bounded by polynomials on R. However, the assumption here allows for very wide classes of functions. This class may be difficult to characterize explicitly; the Assumption 1' below provides an important subclass of locally integrable functions in T'.

Consider functions $b(t)$ for $t \in R$ that satisfy

$$\int (1+t^2)^{-l} |b(t)| \, dt < \infty \textit{ for some } l \geq 0. \tag{5}$$

Assumption 1'. The functions g, W_y, W_{xy} that satisfy the model assumptions are such that each satisfies (5).

The functions g, W_y, W_{xy} that satisfy (5) satisfy Assumption 1. Any function in the space L_1 of absolutely integrable functions satisfies Assumption 1' here but not the Assumption 1 in (S) unless the function is everywhere bounded. While the assumption in (S) extends to polynomial regression functions or distribution functions for binary choice models (where Ft do not exist in the ordinary sense), a regression function that is unbounded at some points is not allowed. There are cases where such properties may arise, e.g. for some hazard functions, for liquidity trap; the more general assumption here accommodates such cases.

Fourier transform is a continuous invertible operator in T', all tempered distributions are differentiable in T' (thus $\dot{\gamma}$ is defined). Fourier transform of an ordinary function of the type considered here may no longer be an ordinary function (e.g. $Ft(const) = \delta$, the Dirac delta-function that cannot be represented as an ordinary function), and thus is not defined point-wise; thus the notation $\gamma(\zeta)$, etc. for the Ft in e.g. (3,4) which we keep here for convenience refers just to the generalized function γ without necessarily giving meaning to values at a point.

In the class of functions that satisfies the model assumptions denote by A the subclass of functions (g, F), by A^* the subclass of functions (g); the mapping $P : A \to A^*$ is given by $P(g, F) = g$. Denote by B the class of functions (W_y, W_{xy}). Equations (1) of model assumptions map A into B (mapping $M : A \to B$); Fourier transforms map B into $Ft(B)$, the class of functions that are Fourier transforms of functions from B; if equations (3,4) could be solved they would provide solutions $\phi^* = \phi I(\gamma \neq 0)$ (where $I(A) = 1$ if A is true, zero otherwise) and γ^* if $\phi \neq 0$; applying inverse Fourier transform would give $g^* = Ft^{-1}(\gamma^*)$. This sequence of mappings can be represented as follows:

$$\underset{(g,F)}{A} \overset{M}{\to} \underset{(W_y,W_{xy})}{B} \overset{Ft}{\to} \underset{(\varepsilon_y,\varepsilon_{xy})}{Ft(B)} \overset{S}{\to} \underset{(\phi^*,\gamma^*)}{Ft(\tilde{A})} \to \underset{(\gamma^*)}{Ft(A^*)} \overset{Ft^{-1}}{\to} \underset{(g^*)}{A^*}. \tag{6}$$

If (6) provides the same result as P so that $g^* \equiv g$ (and $\gamma^* \equiv \gamma$) then g can be identified from the functions (W_y, W_{xy}) with the identification mapping

$$M^* : B \to A^* \tag{7}$$

given by composition of the last five mappings in (6). The most challenging part is in solving the equations to establish the mapping (for $\gamma^* \equiv \gamma$)

$$S : \underset{(\varepsilon_y,\varepsilon_{xy})}{Ft(B)} \to \underset{(\phi^*,\gamma)}{Ft(\tilde{A})} \tag{8}$$

Two additional assumptions are similar to those in (S) and are standard.

Assumption 2. The function $\phi(\zeta)$ is continuous, continuously differentiable on R; and $\phi(\zeta) \neq 0$.

In terms of the model this implies a further condition that absolute moment of U exist.

Assumption 3. Support of generalized function γ coincides with $|\zeta| \leq \bar{\zeta}$ where $\bar{\zeta} > 0$ and could be infinite.

Under Assumptions 1-3 identification is possible as shown in Theorem 1 of this paper; the theorem in (S) asserts an analytic formula (S, (13)) that relies on a decomposition that may not hold.

When the errors-in-variables problem is examined in the space of tempered distributions the corresponding (weak) topology is that of the space T'; in that topology the mappings Ft, Ft^{-1} are known to be continuous, however, the mapping (8) may be discontinuous, rendering the identification mapping (7) discontinuous as well thus implying ill-posedness of the problem. One reason for this is that a too thin-tailed characteristic function may mask high-frequency components in the Fourier transform of the regression function. Theorem 1 here provides a condition under which continuity obtains. When identification is provided by a continuous mapping nonparametric plug-in estimation is possible as long as Fourier transforms of the conditional moment functions can be consistently estimated in T'; this applies e.g. if the regression function is in the L_1 space; Proposition 2 establishes this result.

Identification in classes of parametric functions requires that the mapping from the parameter space to the function space be (at least locally) invertible. (S) uses generalized functions to widen classes of parametric functions for which identification is provided by a finite number of moment conditions; in particular she expands classes of L_1 functions to which the results of Wang and Hsiao (1995, 2009) apply and also allows sums of such functions with polynomial functions, where before polynomial functions were considered only by themselves in Hausman et al. (1991). Her results rely on existence of a moment generating function for the measurement error and use special weighting functions (some of which are improperly defined). Here general classes of functions where such identification is achievable are explicitly characterized rather than via existence of moments conditions (as in S, Assumption 6), the requirement of a moment generating function for measurement error is avoided; appropriate weighting functions are given.

Section 2 deals with identification and well-posedness in the non-parametric case. Section 3 examines identification for the semiparametric model. Proofs are in Appendix A. Appendix B provides an explanation of the claims about the main errors in (S).

2 Non-parametric identification

In the first part of this section known results on generalized functions that confirm the existence and continuity of some of the mappings in (6) are provided, in particular, for the Fourier transform and its inverse. Other mappings, such as (8) require special treatment because they involve multiplication of generalized functions. Multiplication in spaces of generalized functions cannot be defined (Schwartz's impossibility result, 1954, see also Kaminski and Rudnicki (1991) for examples) although there are cases when specific products are known to exist. Here conditions under which some generalized functions can be multiplied by some continuous functions to obtain generalized functions are provided. With this additional insight the existence and continuity of the mappings can be examined. In the second part of this section the identification result is proved and sufficient conditions for the identification mapping to be continuous are provided. A proposition about consistent (in topology of T') nonparametric estimation that in particular applies to functions in space L_1 completes this section.

2.1 Results about generalized functions and existence and continuity of mappings

All the known results in this section are in Schwartz (1966), Gel'fand and Shilov's monograph (vol.1 and 2, 1964) - (GS) and in Lighthill (1959)- (L); they are listed for the convenience of the reader. The sequential approach of Mikusinski in Antonisek et al (1973) is also referred to here. A somewhat distinct approach to multiplication by a continuous function in the space of generalized functions is developed at the end of this section to explain the validity of some of the mappings in (6).

Definitions of generalized function spaces usually start with a topological linear space of well-behaved "test functions", G. Two most widely used spaces are $G = D$ and $G = T$ (usually denoted S in the literature). The linear topological space of infinitely differentiable functions with finite support $D \subset C_\infty(R)$, where $C_\infty(R)$ is the space of all infinitely differentiable functions; convergence is defined for a sequence of functions that are zero outside a common bounded set and converge uniformly together with derivatives of all orders. The space $T \subset C_\infty(R)$ of test functions is defined as:

$$T = \left\{ s \in C_\infty(R) : \left| \frac{d^k s(t)}{dt^k} \right| = O(|t|^{-l}) \text{ as } t \to \infty, \text{ for integer } k \geq 0, l > 0 \right\},$$

$k = 0$ corresponds to the function itself; $|\cdot|$ is the absolute value; these functions go to zero faster than any power. A sequence in T converges if in every bounded region the product of $|t|^l$ (for any l) with any order derivative converges uniformly.

A generalized function, b, is defined by an equivalence class of weakly converging sequences of test functions in G :

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

An alternative equivalent definition is that b is a linear continuous functional on G with values defined by (b, s)[2]. The linear topological space of generalized functions is denoted G'; the topology is that of convergence of values of functionals for any converging sequence of test functions (weak topology); G' is complete in that topology. For $G = D$ or T the spaces are D' and T',

[2] As a referee pointed out this is the more commonly used definition of generalized function; the one above (used by S) is a necessary and sufficient condition and thus represents an equivalent characterization.

correspondingly. It is easily established that $D \subset T$; $T' \subset D'$ and that D' has a weaker topology than T', meaning that any sequence that converges in T' converges in D', but there are sequences that converge in D', but not in T'. The space T' is also called the space of tempered distributions.

Any generalized function b in T' or D' is (weakly) infinitely differentiable: the generalized function $b^{(k)}$ is the $k-th$ order generalized derivative defined by $(b^{(k)}, s) = (-1)^k (b, s^{(k)})$. The differentiation operator is continuous in these spaces. For any probability distribution function F on R^k the density function exists as a generalized function (see e.g., Zinde-Walsh, 2008) and continuously depends on the distribution function, thus the generalized derivative of F, generalized density function f, is in T'.

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

$$(b, s) = \int b(t)s(t)dt; \tag{9}$$

any such function that additionally satisfies (5) similarly by (9) defines a generalized function b in T'. A distinction between functions in the ordinary sense (a pointwise mapping from the domain of definition into the reals or complex numbers) and generalized functions is that generalized functions are not defined pointwise. Generalized functions defined via (9) by ordinary functions $b(t)$ are called regular functions; we can refer to them as ordinary regular functions in G'. The functions F, g, W_y, W_{xy} are ordinary regular functions in T' (and thus in D') if they satisfy Assumption 1'.

If a generalized function b is such that a representation (9) does not hold, it is said that b is singular, so any $b \in G'$ is either regular or singular. A well-known singular generalized function is the $\delta-$function: $\delta : (\delta, s) = s(0)$. Any generalized function in D' or T' is a generalized finite order derivative of a continuous function. An ordinary function that defines a generalized function is regular if it integrates to a continuous function and singular otherwise.

For example, the ordinary function $b(t) = |t|^{-\frac{3}{2}}$ defines a singular generalized function; it does not integrate to a continuous function; it does not satisfy (9), in fact (see GS, v.1, p.51) it defines a generalized function by

$$(b, s) = \int_0^\infty t^{-\frac{3}{2}} \{s(t) + s(-t) - 2s(0)\} \, dt. \tag{10}$$

No special treatment is needed to consider complex-valued generalized functions; all the same properties hold. For $s \in T$ or D Fourier transform $Ft(s) = \int s(t)e^{it\zeta}dt$ exists and is in T. For $b \in T' : (Ft(b), s) = (b, Ft(s))$, so $Ft(b) \in T'$. Fourier transform defines a continuous and continuously invertible linear operator in T' (but not for D'). Thus Fourier transforms of W_y, W_{xy}, g, and of the generalized derivative, f, of F exist in T' and their inverse Fourier transforms coincide with the original functions. Since all the functions are differentiable as generalized functions $\dot{\gamma}$ exists in T'. By Assumption 2 the characteristic function ϕ is continuous, as is its derivative, $\dot{\phi}$; they are regular ordinary functions in T'.

Since G' does not have a multiplicative structure, products and convolutions can be defined for specific pairs only and generally exist only for special classes. The product between a generalized function in T' and a function from C_∞ with all derivatives bounded by polynomial functions exists. This class of multipliers is denoted by $\mathcal{O}_M$. Convolution of Fourier transforms of generalized functions with Fourier transforms of functions from $\mathcal{O}_M$ exists and the convolution theorem applies. Products and convolutions may exist for other specific pairs of generalized functions. When such convolutions and products of their Fourier transforms exist as generalized functions the convolution theorem similarly applies to such pairs.

Convolution Theorem. If for $b_1, b_2 \in T'$, convolution $b_1 * b_2 \in T'$, product $Ft(b_1) \cdot Ft(b_2) \in T'$, then $Ft(b_1 * b_2) = Ft(b_1) \cdot Ft(b_2)$.

The proof of this theorem uses exactly the same sequential argument as in Antonisek et al (1973), where it utilized the specific delta-convergent sequences; the only difference here is that the argument can be applied to any sequence in the equivalence class that defines every given generalized function.

To consider the product of a generalized function with a continuous function that may not be infinitely differentiable, the property that the product does not depend on the sequence that defines the generalized function has to be made a requirement. We thus say that ab for $b \in G'$ and continuous a is defined in G' if for any sequence b_n from the equivalence class of b there exists a sequence $(ab)_n$ in G such that for any $\psi \in G$

$$\lim \int a(x) b_n(x) \psi(x) dx \text{ exists and equals } \lim \int (ab)_n (x) \psi(x) dx. \qquad (11)$$

Denote by 0_n a zero-convergent sequence that belongs to the equivalence class defining the function that is identically zero in G'.

Proposition 1 *For the product ab between a continuous function a* and $b \in G'$ *to be defined in* G' *it is necessary and sufficient that (i)* (11) *hold for some sequence* $\tilde{b}_n$ *in the class that defines b and (ii) for any zero-convergent sequence,* $0_n(x)$,

$$\lim \int a(x) 0_n(x) \psi(x) dx = 0. \qquad (12)$$

Proof.

Any sequence b_n differs from a specific $\tilde{b}_n$ by a zero-convergent sequence.■

Here we consider functions that stem from the model assumptions. Additionally, we distinguish the following cases.

Case 1. Support of γ is a bounded set: $\bar{\zeta} < \infty$.

Case 2. The function ϕ^{-1} *satisfies (5).*

Case 3. The function $\phi \in \mathcal{O}_M$.

Lemma 1 *Under Assumptions 1-3*

(i) the products $\gamma\phi$ *and* $\gamma\dot{\phi}$ *are defined in* T' *and in* D';

(ii) for $\tilde{\phi}^{-1} = \phi^{-1}(\zeta) I(|\zeta| < \bar{\zeta})$ *the product* $(\gamma\phi) \cdot \tilde{\phi}^{-1}$ *is always defined in* D';

(iii) if either case 1 applies or both cases 2 and 3 apply the product $(\gamma\phi)\cdot\tilde{\phi}^{-1}$ *is defined in* T'*;*

(iv) *if neither case 1 nor case 2 applies the product may not be defined in* T'.

Proof. See Appendix.

From Lemma 1 existence of products to justify the convolution theorem and thus equations (3,4) follows. The mapping (8) involves solving equations (3,4) for the unknown functions and requires multiplication by ϕ^{-1}; as one can see from Lemma 1 existence of such products in T' is not guaranteed.

2.2 The nonparametric identification theorem

This section contains two results. The first is Theorem 1 that proves the existence of the identification mapping M^* under Assumptions 1-3. It differs from the statement in (S, Theorem 1) in three ways: first, Assumption 1 (and even the more restrictive Assumption 1') of this paper is more general; second, it does not rely on decomposition of generalized functions[3]; third, it provides the condition under which the mapping is obtained via operations in the space T' and discusses the continuity of the identification mapping. The second result is Proposition 2 that shows that when continuity holds, consistent (in the topology of T') plug-in non-parametric estimation of the regression function is possible, e.g. for functions in space L_1.

Theorem 1 *For functions satisfying model assumptions and Assumptions 1-3 the mapping* M^* *in (7) exists and provides identification for* g*; if conditions of (iii) of Lemma 1 are satisfied the mapping is defined via operations in* T'*; it can be discontinuous under condition (iv) of Lemma 1.*

Proof. See Appendix.

[3] There is no known decomposition in the space of generalized functions into generalized functions corresponding to ordinary functions and to singular functions, claimed in (S); the pointwise argument provided in (S, 2007, Supplementary Material) is incorrect (see Appendix B).

The implication of this Theorem is that the identification result holds under the general assumptions 1-3. If ϕ is too thin-tailed, however, the mapping whereby the identification is achieved may not be continuous: this point is illustrated by the example in the proof of Theorem 1 where high frequency components b_n are magnified by multiplication with ϕ^{-1} from a thin-tailed distribution; this produces inverse Fourier Transforms that diverge.

Continuity requires that the mapping M given by the model assumptions be continuous in T'. Continuity in T' allows for a consistent (in T') plug-in estimator; the following Proposition provides sufficient conditions. Denote by $\rightarrow_{T'}$ convergence in topology of T'. Following Gel'fand and Vilenkin (1964) we define a random generalized function as the random continuous functional on the space of test functions.

Proposition 2 *(a) Under the conditions of Theorem 1 suppose that* W_{yn}, W_{xyn} *are random generalized functions (estimators) that satisfy model assumptions and Assumptions 1-3 together with some (unknown) functions* g_n, F_n*; condition (iii) of Lemma 1 is satisfied;* $\phi_n \in \mathcal{O}_M$*; for* $\varepsilon_{yn} = Ft(W_{yn})$, $\varepsilon_{xyn} = Ft(W_{xyn})$ *assume that the Fourier transforms satisfy:* $\varepsilon_{yn}(\zeta)$ *is continuous and non-zero a.e. on* $\text{supp}(\gamma)$ *and* $i\varepsilon_{xyn}(\zeta) - \dot{\varepsilon}_{yn}(\zeta)$ *is continuous and that*

$$\begin{aligned} \Pr(\varepsilon_{yn}(\zeta) &\rightarrow {}_{T'}\varepsilon_y(\zeta)) \rightarrow 1; \\ \Pr(\varepsilon_{xyn}(\zeta) &\rightarrow {}_{T'}\varepsilon_{xy}(\zeta)) \rightarrow 1, \end{aligned} \tag{13}$$

then it is possible to find a sequence $g_n(x)$ *such that* $\Pr(g_n(x) \rightarrow_{T'} g(x)) \rightarrow 1$.

(b) Suppose that the function $g(x) \in L_1$. Then there exists a sequence of step function estimators, g_n, such that

$$\Pr(g_n(x) \rightarrow_{T'} g(x)) \rightarrow 1.$$

Proof. See Appendix.

Convergence of the estimators is in the weak topology of space T' of generalized functions, not in L_1. If $g_n(x) \to_{T'} g(x)$ and $g(x)$ is a continuous function then there is pointwise convergence and uniform convergence on bounded sets.

3 Semiparametric specification and identification

Semiparametric models with measurement error were examined for polynomial regression functions by Hausman et al (1991), for regression function in the L_1 space by Wang and Hsiao (1995, 2009). (S) significantly widened the classes of semiparametric models with errors-in-variables where identification can be achieved via moment conditions by utilizing generalized functions, but did not explicitly characterize the class of functions which she considered: verifying moment conditions of (S, Assumption 6) is needed. In contrast, the class of parametric functions is characterized directly in Assumptions 5 and 6 of this paper; the assumptions give sufficient conditions for identification via moments. The results in (S) rely on existence of a moment generating function for the measurement error; this restriction in not imposed here. In this paper as well as in (S) some moment conditions involve limits for sequences of weighting functions; the limits are explicitly given here.

Assumption 4. The function $g(x^*)$ is in a parametric class of locally integrable functions $g(x^*, \theta)$ where $\theta \in \Theta$; Θ is an open set in R^m; for some $\theta^* \in \Theta$ model assumptions and (1) hold.

Denote all the Fourier transforms of the parametric functions in the model assumptions as $\gamma(\theta); \varepsilon_y(\theta); \varepsilon_{xy}(\theta)$. The following assumption restricts the generalized function $\gamma(\theta)$ to have no more than a finite number of special points: Δ points of singularity and J of "jump" discontinuity in some region $|\zeta| < \bar{\zeta} < \infty$. Notation $[x]$ is for integer part of x; $\delta(\zeta - a)$ denotes a shilted $\delta - function : (\delta(\zeta - a), \psi) = \psi(a)$ for $\psi \in G$.

Assumption 5. The Fourier transform, $\gamma(\theta)$, of the real function $g(x^*, \theta)$

in the region $|\zeta| < \bar{\zeta} < \infty$ that belongs to its support (and may coincide with it) can be represented as

$$\gamma(\theta) = \gamma_o(\theta) + \gamma_s(\theta), \tag{14}$$

where

(i) if $\Delta = 0$, $\gamma_s(\theta) \equiv 0$; if $\Delta \geq 1$

$$\begin{aligned} \gamma_s(\theta) &= 2\pi \sum_{l=0}^{L} \gamma_{s_l}(\theta), \text{ where } L = \left[\frac{\Delta}{2}\right] \text{ and} \qquad (15) \\ \gamma_{s_l}(\theta) &= \sum_{k=0}^{\bar{k}_l} \left(\gamma_k(s_l,\theta)\delta^{(k)}(\zeta - s_l) + \overline{\gamma_k(s_l,\theta)}\delta^{(k)}(\zeta + s_l)\right); \text{ for } l = 0, 1, ...L; \end{aligned}$$

(ii) $\gamma_o(\theta) \equiv \gamma_o(\zeta,\theta)$ is defined by a locally integrable function of ζ continuous except possibly in a finite number of points and such that its generalized derivative, $\dot{\gamma}_o(\theta)$, is of the form

$$\dot{\gamma}_o(\theta) = \dot{\gamma}_{oo}(\theta) + \dot{\gamma}_{os}(\theta),$$

where if $J = 0$, then $\dot{\gamma}_{os}(\theta) = 0$, and if $J > 0$, then for points b_j, $j = 1, ... \left[\frac{J}{2}\right]$, $\dot{\gamma}_{os}(b_j,\theta) =$

$$\gamma_{os0}(0,\theta)\delta(\zeta)I(J \text{ is odd}) + \Sigma_{j=1}^{\left[\frac{J}{2}\right]}\left(\gamma_{osj}(b_j,\theta)\delta(\zeta - b_j) + \overline{\gamma_{osj}(b_j,\theta)}\delta(\zeta + b_j)\right),$$

$\dot{\gamma}_{oo}(\theta) \equiv \dot{\gamma}_{oo}(\zeta,\theta)$ is an ordinary locally integrable function continuous except possibly in a finite number of points;

(iii) $\gamma_o(\zeta,\theta) \neq 0$ except possibly for a finite number of points in $(-\bar{\zeta},\bar{\zeta})$;

(iv) At any non-zero singularity point: $s_l \neq 0$, $\gamma_o(\zeta,\theta)$ is continuous and non-zero.

Under Assumptions 1 and 5 g could be in L_1, or a sum of a function from L_1 and a polynomial (singularity point $\zeta_0 = 0$) and also possibly a periodic function, e.g. $\sin(\cdot)$ or $\cos(\cdot)$ with singularities at some points $\pm s$, $s \neq 0$. Here the parameters, $\gamma_{\cdot}(\cdot, \theta)$, are allowed to take complex values, otherwise one would need to be more specific about the functions with singular Fourier transforms; since the functions are assumed known it is easy in each specific case to separate out the imaginary parts as in the case of polynomials.

Assumption 5 permits to write moment conditions; however, to get a sufficient condition for identification of all parameters additionally the following Assumption 6 is made.

If $\Delta > 0$ define the matrices $\Gamma_y(s_l, \theta)$ and $\Gamma_{xy}(s_l, \theta)$ for each $s_l \geq 0$ (similarly to (S) for the case $s_l = 0$) by their elements:

$$\begin{aligned}
\Gamma_{y,i+1,k+1}(s_l, \theta) &= \begin{pmatrix} k+i \\ i \end{pmatrix} \gamma_{k+i}(s_l, \theta) I(k+i \leq \bar{k}_l); \\
\Gamma_{xy,i+1,k+1}(s_l, \theta) &= \begin{pmatrix} k+i+1 \\ i+1 \end{pmatrix} \gamma_{k+i}(s_l, \theta) I(k+i \leq \bar{k}_l), \\
i, k &= 0, 1, ...\bar{k}_l.
\end{aligned}$$

Denote by $\{A\}_{11}$ the first matrix element of a matrix A.

Assumption 6. The function γ satisfies Assumption 5. Additionally all $\gamma_o(\zeta, \theta), \dot{\gamma}_{oo}(\zeta, \theta)$, $\gamma_{s_l}(\theta)$ are continuously differentiable with respect to the parameter, θ, in some neighborhood of θ^*. The $m \times 1$ parameter vector can be partitioned as $\theta^T = [\theta_I^T; \theta_{II}^T]$. For any component, θ_i, of $m_I \times 1$ vector θ_I (where $m \geq m_I \geq 0$) either

$$\gamma_o(\zeta, \theta^*) \frac{\partial}{\partial \theta_i} \dot{\gamma}_{oo}(-\zeta, \theta)|_{\theta^*} + \dot{\gamma}_{oo}(\zeta, \theta^*) \frac{\partial}{\partial \theta_i} \gamma_o(-\zeta, \theta)|_{\theta^*} \neq 0 \qquad (16)$$

a.e., or if (16) does not hold for some i^*, then $\frac{\partial}{\partial \theta_{i^*}} \gamma_o(-\zeta, \theta)|_{\theta^*} \neq 0$. If $m_{II} > 0$ the matrix that stacks for all s_l, $l \geq 0$ matrices

$$\begin{pmatrix} \frac{\partial}{\partial \theta_{II}^T} \left[\Gamma_y(s_l, \theta)\right]^{-1} |_{\theta^*} \Gamma_y(s_l, \theta^*) + \frac{\partial}{\partial \theta_{II}^T} \left[\Gamma_{xy}(s_l, \theta)\right]^{-1} |_{\theta^*} \Gamma_{xy}(s_l, \theta^*) \\ \frac{\partial}{\partial \theta_{II}^T} \left\{\left[\Gamma_y(s_l, \theta)\right]^{-1} |_{\theta^*} \Gamma_y(s_l, \theta^*)\right\}_{11} |_{\theta^*} \end{pmatrix}$$

is of rank m_{II}.

By checking we can see that all the examples provided in (S) satisfy assumptions 5 and 6 here and thus sufficient conditions for identification hold. If the same parameters enter into both the ordinary and singular parts (S, assumption 6) may be violated, even though identification is possible and the results of this paper hold.

Additional assumptions 7 and 8 below are needed.[4]

Assumption 7. The density function $p(z)$ exists and is positive.

Assumption 8. The characteristic function of measurement error, $\phi(\zeta)$, is such that (i) $\phi(\zeta) \neq 0$ for $|\zeta| < \bar{\zeta}$ where it is continuously differentiable; (ii) it is $\bar{k}_l$ times continuously differentiable at every s_l.

Theorem 2 below establishes that moment conditions for the parameters θ of $\gamma(\zeta, \theta)$ hold and Theorem 3 that the assumptions are sufficient for identification. The notation $\mathrm{Re}(x)$ refers to the real part of a complex x.

Theorem 2. *Under model assumption and Assumptions 1', 4, 7, 8*

(i) if Assumption 5 (i,ii) holds there exist real functions $r_y(z,\theta), r_{xy}(z,\theta)$ *such that the moment*

$$E\left(\frac{Yr_{xy}(z,\theta) + XYr_y(z,\theta)}{p(z)}\right) \tag{17}$$

exists for θ *in some neighborhood of* θ^* *and equals zero for* $\theta = \theta^*$;

(ii) if 5(i-iii) holds there are functions $r_{y1n}(z,\theta)$ *such that*

$$\lim_{n\to\infty} E\left(\frac{Yr_{y1n}(z,\theta)}{p(z)} - 1\right) \tag{18}$$

exists for θ *in some neighborhood of* θ^* *and equals zero for* $\theta = \theta^*$;

(iii) If $\Delta > 0$ *and 5(i-ii) hold then for each* $s_l \geq 0$ *there exist vector functions* $r_{ysl}(z,\theta)$, $r_{ysl,n}(z,\theta), r_{xysl}(z,\theta), r_{xysl,n}(z,\theta)$, *and a diagonal invert-*

[4] Assumptions 7 and 8(ii) are also implicit in the proofs in (S).

ible matrix M_l such that

$$\lim_{n\to\infty} \mathrm{Re}[\Gamma_y^{-1}(s_l,\theta)M_l^{-1}E\left(\frac{Yr_{ys,l,n}(z,\theta)}{p(z)}\right) \tag{19}$$
$$+\Gamma_{xy}^{-1}(s_l,\theta)M_l^{-1}E\left(\frac{XYr_{xys,l,n}(z,\theta)}{p(z)}\right)]$$

exists for θ in some neighborhood of θ^* and equals zero for $\theta=\theta^*$;

(iv) If $\Delta>0$ and 5(i-iv) hold then for each $s_l\geq 0$ there exist functions $r_{ysl,1,n}(z,\theta)$, $r_{yslo,1,n}(z,\theta)$ such that for $s_0=0$

$$\lim_{n\to\infty} E\left(\frac{Yr_{ys0,1,n}(z,\theta)}{p(z)}-1\right) \tag{20}$$

and

$$\lim_{n\to\infty} \mathrm{Re}\, E\left(\frac{Y\left(r_{ysl,1,n}(z,\theta)-r_{yslo,1,n}(z,\theta)\right)}{p(z)}\right) \tag{21}$$

exist for θ in some neighborhood of θ^* and equal zero for $\theta=\theta^*$.

Proof. See Appendix. The functions $r.(z,\theta)$ and matrices M_l are provided there.

Some of the moment conditions can be redundant. Different sets of weighting functions could be appropriate; similarly to reasoning in (S) the weighting functions are designed in a way that isolates different components of the γ function: the ones in (i) are for the ordinary function component and are supplemented by moments in (ii) for the case of a scale multiple for the ordinary component, the ones in (iii) are for the coefficients of the singular part with (iv) for the possible scale factor at each singularity. If only (17) applies then the weighting functions proposed in (S) can be used, but for the other components the weights proposed here solve the problem without additional requirements that moment generating function for errors exist and avoid the problematic function μ in (S, Definition 2): $\mu(0)$ and any derivatives of μ at 0 are zero (see Appendix B).

Define by $EQ(\theta)$ the vector with components provided by the stacked expressions (whichever are defined) from (17, 18, 19, 21).

Theorem 3. *Under the conditions of Theorem 2 and Assumption 6 the functions $r.(z,\theta)$ can be selected in such a way that the matrix $\frac{\partial}{\partial\theta}EQ(\theta^*)$ exists and has rank m.*

Proof. See Appendix.

Theorem 3 provides sufficient conditions under which the equations $EQ(\theta) = 0$ fully identify the parameter vector θ^*.

4 Appendix A

4.1 Proofs

Proof of Lemma 1.

(i) By model assumption and Assumption 1 the convolutions in (1) provide elements in T'; this implies that any of the sequential definitions of convolution in Kamiski (1982) hold, therefore by his theorem 9, the "exchange formula" implies that the products for some sequences in the equivalence classes defining the $Ft's$ exist. By Proposition 1 it follows that since $\phi, \dot{\phi} \in T'$, (12) holds for continuous functions ϕ and $\dot{\phi}$ then $\gamma\phi \in T'$, $\dot{\gamma\phi} \in T'$ and additionally (by applying the product rule to (3,4)) $\gamma\dot{\phi} \in T'$. Since $T' \subset D'$, the products are defined in D' as well.

(ii) Now consider a sequence $(\gamma\phi)_n$ defined as follows: select some sequence $\tilde{\gamma}_n$ for γ from D; then each $\tilde{\gamma}_n$ has finite support; for a sequence of numbers $\varepsilon_n \to 0$ select $\tilde{\phi}_n$ in D such that $\left|\tilde{\phi}_n - \phi\right| < \frac{\varepsilon_n}{\sup\left|\tilde{\gamma}_n\phi^{-1}\right|}$ on compact support of γ_n. Then for the sequence $(\gamma\phi)_n = \tilde{\gamma}_n\tilde{\phi}_n$ and any $\psi \in D$

$$\int \tilde{\gamma}_n\tilde{\phi}_n\phi^{-1}\psi = \int \tilde{\gamma}_n\psi + \int \tilde{\gamma}_n(\tilde{\phi}_n - \phi)\phi^{-1}\psi \to \int \gamma\psi.$$

Now we check that (12) holds for $a = \phi^{-1}$. In D support of any ψ is bounded, on that compact set ϕ^{-1} is bounded thus (12) will hold and the product is defined in D'.

(iii) For Case 1 the product with $\phi^{-1}(\zeta)I(|\zeta| < \bar{\zeta})$ is similarly to (ii) defined in T' since it is sufficient to consider $\psi \in T$ with bounded support (containing support of γ). If cases 2 and 3 hold it is straightforward to verify that the function ϕ^{-1} is in $\mathcal{O}_M$, thus the product is defined (continuously) in T'.

(iv) We construct a counterexample. The function $\phi(x) = e^{-x^2}$ does not belong to either case 1 or case 2. The product of function $b(x) \equiv 0$ and $\phi(x)^{-1}$ does not exist in T'. Consider $b_n(x) =$

$$\begin{cases} e^{-n} & \text{if } n - \frac{1}{n} < x < n + \frac{1}{n}; \\ 0 \leq b_n(x) \leq e^{-n} & \text{if } n - \frac{2}{n} < x < n + \frac{2}{n}; \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

This $b_n(x)$ converges to $b(x) \equiv 0$ in T'. Indeed for any $\psi \in T$

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

But the sequence $b_n(x)\tilde{\phi}(x)^{-1}$ does not converge in the space T' of tempered distributions. Indeed if it did then $\int b_n\tilde{\phi}\psi$ would converge for any $\psi \in T$. But for $\psi \in T$ such that $\psi(x) = \exp(-|x|)$ for, e.g. $|x| > 1$

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

This diverges.■

Proof of Theorem 1.

The proof makes use of different spaces of generalized functions and exploits relations between them. It proceeds in two parts.

First in part one, it is shown that from equations (3,4) the continuous function $\varkappa = \dot{\phi}\phi^{-1}$ can be uniquely pointwise determined on the interval $[-\bar{\zeta}, \bar{\zeta}]$ (which is in the support of γ and consequently of ε_1); this requires additionally considering the generalized functions spaces, D' and also $D_0(U)'$ which is defined on the space of test functions that are continuous with support contained in U. The function ϕ is uniquely defined on the interval $(-\bar{\zeta}, \bar{\zeta})$ as the solution of the corresponding differential equation that satisfies the condition $\phi(0) = 1$. Define $\tilde{\phi} = \phi I(|\zeta| < \bar{\zeta})$; define $\tilde{\phi}^{-1}$ to equal $\phi^{-1} I(|\zeta| < \bar{\zeta})$. Of course, when $\bar{\zeta} = \infty$, $\tilde{\phi} = \phi$ and $\tilde{\phi}^{-1} = \phi^{-1}$ on R.

Next in part two, γ is defined as $\varepsilon_y \tilde{\phi}^{-1}$. By Lemma 1 this product can always be uniquely defined as a generalized function in D'; by construction γ defines a generalized function in $T' \subset D'$; this provides the required mapping M^* by applying inverse Fourier Transform to γ. If condition (iii) of Lemma 1 applies the product $\varepsilon_y \tilde{\phi}^{-1}$ is defined in T'; in this case all the mappings that define the mapping M^* are defined in T'. The proof concludes with an example that demonstrates that the mapping can be discontinuous if (iv) of Lemma 1 applies.

Part one. Consider the space of generalized functions D'. By Assumption 2 ϕ is non-zero and continuously differentiable, then by differentiating (3), substituting (4) and making use of Lemma 1 to multiply by ϕ^{-1} in D' we get that the generalized function

$$\varepsilon_y \phi^{-1} \dot{\phi} - (\dot{\varepsilon}_y - i\varepsilon_{xy})$$

equals zero in the sense of generalized function in D'. Denoting $\varkappa = \dot{\phi}\phi^{-1}$ we can characterize $\varkappa$ as a continuous at every point (by Assumption 2) function in D' that satisfies the equation

$$\varepsilon_y \varkappa - (\dot{\varepsilon}_y - i\varepsilon_{xy}) = 0. \tag{23}$$

If (23) holds in D', it holds also for any test functions with support limited to $U : \psi \in D(U) \subset D$, and thus holds in any $D(U)'$.

We show that the function $\varkappa$ is uniquely determined in the class of continuous functions on on $[-\bar{\zeta}, \bar{\zeta}]$ by (23) holding in $D(U)'$ for any interval $U \subset (-\bar{\zeta}, \bar{\zeta})$. Proof is by contradiction. Suppose that there are two distinct continuous functions on $[-\bar{\zeta}, \bar{\zeta}]$, $\varkappa_1 \neq \varkappa_2$ that satisfy (23), then $\varkappa_1(\bar{x}) \neq \varkappa_2(\bar{x})$ for some $\bar{x} \in (-\bar{\zeta}, \bar{\zeta})$; by continuity $\varkappa_1 \neq \varkappa_2$ everywhere for some interval $U \subset (-\bar{\zeta}, \bar{\zeta})$. Consider now $D(U)'$; we can write

$$(\varepsilon_y(\varkappa_1 - \varkappa_2), \psi) = 0$$

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 denotes the space of continuous test functions. For the space of test function $D_0(U)$ multiplication by continuous $(\varkappa_1 - \varkappa_2) \neq 0$ is an isomorphism. Then from (23) we can write

$$0 = ([\varepsilon_y(\varkappa_1 - \varkappa_2)], \psi) = (\varepsilon_y, (\varkappa_1 - \varkappa_2)\psi)$$

implying that ε_y is defined and is a zero generalized function in $D_0(U)'$. If that were so ε_y would be a zero generalized function in $D(U)'$ since $D(U) \subset D_0(U)$; this contradicts Assumption 2. This concludes the first part of the proof since from $\varkappa$ the function

$$\phi(\zeta) = \exp \int_0^\zeta \varkappa(\xi) d\xi$$

that solves on $[-\bar{\zeta}, \bar{\zeta}]$

$$\dot{\phi}\phi^{-1} = \varkappa; \phi(0) = 1$$

is uniquely determined on $[-\bar{\zeta}, \bar{\zeta}]$ and $\tilde{\phi}$ (and $\tilde{\phi}^{-1}$) defined above are uniquely determined.

Part two.

Consider two cases.

Case 1.Part (iii) of Lemma 1 applies. Multiplying $\varepsilon_y(=\gamma\tilde{\phi})$ by $\tilde{\phi}^{-1}$ provides a tempered distribution by Lemma 1; it is equal to γ. The inverse Fourier Transform provides g. The theorem holds and moreover, all the operations by which the solution was obtained were defined in T'.

Case 2. The condition (iii) of Lemma 1 may not hold, so multiplication by $\tilde{\phi}^{-1}=\phi^{-1}$ may not lead to a tempered distribution. Consider now D'; $T'\subset D'$. Multiplication by ϕ^{-1} is a continuous operation in D'; define the same differential equations, solve to obtain ϕ and get via multiplication $(\gamma\phi)\cdot\phi^{-1}$ in D' the function $\gamma\in D'$. Since γ is the Fourier transform of g (a tempered distribution) it also belongs to T', and it is possible to recover g by an inverse Fourier Transform.

In the following example the mapping M^* in (7) is not continuous. Define $\beta_n=Ft^{-1}(b_n)$ where b_n is defined by (22); $b_n\in T$, thus $\beta_n\in T$. In was shown in proof of Lemma 1 that $b_n(x)$ converges to $b(x)\equiv 0$ in T'.

Suppose that the model mapping M in (6) is defined for functions in L_1 and is continuous in L_1 (and thus in T'). Suppose that $W_{yn}=W_y+\beta_n$; from $b_n\to 0$ in T' and the continuity of the Fourier Transform mapping in T', it follows that $\beta_n\to 0$ and thus $W_{yn}\to W_y$ in T'. Then $\varepsilon_{yn}=\varepsilon_y+b_n$. Suppose that ϕ is proportionate to e^{-x^2}. Then by the proof in part 1 each $\gamma_n=\varepsilon_{yn}\phi^{-1}\in D'$, but as a Ft of a function in T' (even in L_1 here) is defined in T'; the inverse Fourier transform, $\tilde{g}_n=Ft^{-1}(\gamma_n)$, exists. However, $\tilde{g}_n$ does not converge to g in T'. Indeed, if it did so converge, then that would imply convergence $\gamma_n\to\gamma$ in T', but $b_n(x)e^{x^2}$ does not converge in the space T' of tempered distributions as was shown in the proof of Lemma 1. ■

Proof of Proposition 2.

(a) We establish that the mapping from (W_{yn},W_{xyn}) to g_n is continuous. Consider $\zeta\in$supp(γ). Similarly to proof in Theorem 1 applied to every n a continuous function $\varkappa_n(\zeta)$, that satisfies the equation

$$\varkappa_n(\zeta)\varepsilon_{yn}(\zeta)+(i\varepsilon_{xyn}(\zeta)-\dot{\varepsilon}_{yn}(\zeta))=0 \qquad (24)$$

in generalized functions, exists (defined as $\dot{\phi}_n\phi_n^{-1}$) and is unique. Moreover, from Lemma 1 it follows that the product with $\varkappa_n = \dot{\phi}_n\phi_n^{-1} \in \mathcal{O}_M$ always exists. Since all functions in (24) are continuous it represents an equality of continuous functions and since ε_{yn} is non-zero a.e. we have

$$\varkappa_n(\zeta) = (i\varepsilon_{xyn}(\zeta) - \dot{\varepsilon}_{yn}(\zeta))(\varepsilon_{yn}(\zeta))^{-1}.$$

The generalized functions

$$\varkappa_n\varepsilon_{yn} - \varkappa\varepsilon_y = i(\varepsilon_{xyn} - \varepsilon_{xy}) + (\dot{\varepsilon}_{yn} - \dot{\varepsilon}_y) \text{ and } \varkappa_n(\varepsilon_y - \varepsilon_{yn})$$

converge to zero as generalized functions in T'; as a result, so does $(\varkappa_n - \varkappa)\varepsilon_{yn}$, but since this is a continuous function this implies pointwise convergence. Suppose that on some bounded interval $\varkappa_n - \varkappa$ is separated away from zero for some subsequence $\{n_i\}$, this implies then that on that set ε_{yn_i} converges to zero pointwise, thus the limit in T' ($=\varepsilon_y$) is zero on this interval which belongs to support of γ, and thus of ε_y. This contradiction establishes that $\varkappa_n \to \varkappa$ pointwise and uniformly on any bounded set. From the differential equation $\phi_n^{-1}\dot{\phi}_n = \varkappa_n$ with the condition $\phi_n(0) = 1$ the function ϕ_n is uniquely determined; and $\phi_n \to \phi$ where $\phi^{-1}\dot{\phi} = \varkappa, \phi(0) = 1$. Then also since ϕ is non-zero, $\phi_n^{-1} \to \phi^{-1}$ pointwise; ϕ_n^{-1} satisfies *(5)* so that $\varepsilon_{1n}\phi_n^{-1}$ can be defined as a tempered distribution. Finally consider

$$\varepsilon_{yn}\phi_n^{-1} - \varepsilon_y\phi^{-1} = \varepsilon_{yn}(\phi_n^{-1} - \phi^{-1}) + (\varepsilon_{yn} - \varepsilon_y)\phi^{-1}.$$

The continuous function $\varepsilon_{yn}(\phi_n^{-1} - \phi^{-1}) \to 0$; the generalized function $(\varepsilon_{yn} - \varepsilon_y)\phi^{-1} \to 0$ in T' (as a tempered distribution) $\varepsilon_{yn}\phi_n^{-1}$ converges to γ in T', and its inverse Fourier Transform converges to g as a tempered distribution (by continuity of inverse Fourier Transform in T').

From continuity of the mapping the result follows.

(b) For any $g \in L_1$ there exists a sequence of step-functions $g_n \in L_1$ such that $\|g_n - g\|_{L_1} \to 0$ (implying $g_n \to_{T'} g$); for F there is a sequence of step-functions F_n such that $\sup|F_n - F| \to 0$ (implying $F_n \to_{T'} F$).

Specifically,

$$\begin{aligned} g_n(x) &= \sum_{k=1}^{N} a_k I(b_k \le x < b_{k+1}) \text{ for } b_1 < ... < b_N; \\ F_n(x) &= \sum_{j=1}^{N} c_j I(d_j \le x) \text{ with } c_j > 0; \Sigma c_j = 1; d_1 < ... < d_N, \end{aligned}$$

where all the parameters depend on n. The generalized derivative of F_n is $f_n(x) = \Sigma c_j \delta(x - d_j)$ where $\delta(x - d_j)$ is a shifted $\delta-$function: $\int \delta(x - d_j)\psi(x)dx = \psi(d_i)$ for $\psi \in T$. Then $\phi_n(\zeta) = \Sigma c_j e^{i\zeta d_j}$. The function ϕ_n is not integrable (otherwise f would be continuous), thus ϕ_n^{-1} satisfies (5). All the parameters depend on n.

Then

$$\begin{aligned} W_{yn}(v) &= \sum_{m=1}^{N^2} \alpha_m I(|v - t_m| < \tau_m); \\ W_{xyn}(v) &= \sum_{m=1}^{N^2} \alpha_m (v - \varepsilon_m) I(|v - t_m| < \tau_m), \end{aligned}$$

where m corresponds to a pair (k, j) and $\alpha_m = a_k c_j; t_m = d_j + \frac{b_k + b_{k+1}}{2}, \varepsilon_m = d_j, \tau_m = \frac{b_{k+1} - b_k}{2}$. This represents W_{yn} as a step-function and W_{xyn} as a piece-wise linear function. The conditional mean function W_y can be consistently estimated in L_1 by step functions implying existence of a sequence $W_{yn}(v)$ such that $\Pr(W_{yn} \to_{T'} W_y) \to 1$, similarly, for some piece-wise linear $W_{xyn}(v)$ $\Pr(W_{xyn} \to_{T'} W_{xy}) \to 1$ implying (13), moreover, we can write (using known

Fourier transforms)

$$\begin{aligned}
\varepsilon_{yn}(\zeta) &= \sum_{k=1}^{N} 2\tau_k \alpha_k \chi_k(\zeta) sinc(\frac{\tau_k \zeta}{\pi}); \\
\varepsilon_{xyn}(\zeta) &= -i\sum_{k=1}^{N} 2\tau_k \frac{d}{d\zeta}\left[\alpha_k \chi_k(\zeta) sinc(\frac{\tau_k \zeta}{\pi})\right] - \sum_{k=1}^{N} 2\tau_k \alpha_k \varepsilon_k \chi_k(\zeta) sinc(\frac{\tau_k \zeta}{\pi}),
\end{aligned}$$

where the $sinc(x)$ function is defined as $\frac{\sin \pi x}{\pi x}$ and $\chi_k(\zeta) = e^{it_k \zeta}$. The conditions about continuity and $\varepsilon_{yn}(\zeta)$ non-zero a.e., required in (a) are satisfied; *(13)* follows from the continuity of the Fourier transform operator in T'.■

Prior to proof of Theorem 2 we make two preliminary observations.

Firstly, under Assumption 5 and 8 (that justifies products of γ_s and $\dot{\gamma}_s$ with ϕ) and by Lemma 1 equations (3,4) in T' lead to ($\mathbf{i}^2 = -1$):

$$\begin{aligned}
\varepsilon_y &= \varepsilon_{yo} + \varepsilon_{ys} && (25) \\
\text{with } \varepsilon_{yo}(\zeta) &= \gamma_o(\zeta, \theta^*)\phi(\zeta); \varepsilon_{ys} = \gamma_s(\theta^*)\phi(\zeta); \\
\mathbf{i}\varepsilon_{xy} &= \mathbf{i}\varepsilon_{xyo} + \mathbf{i}\varepsilon_{xyos} + \mathbf{i}\varepsilon_{xys} && (26) \\
\text{with } \mathbf{i}\varepsilon_{xyo}(\zeta) &= \dot{\gamma}_{oo}(\zeta, \theta^*)\phi(\zeta), \\
\mathbf{i}\varepsilon_{xyos}(\zeta) &= \dot{\gamma}_{os}(\zeta, \theta^*)\phi(\zeta), \\
\text{and } \mathbf{i}\varepsilon_{xys} &= \dot{\gamma}_s(\theta^*)\phi, \text{ where } \dot{\gamma}_s(\theta) \text{ is the generalized derivative of } \gamma_s(\theta).
\end{aligned}$$

Second, to construct weighting functions some well-known functions are used. Denote by $T_R \subset T$ the space of real test functions that are Ft of real-valued functions from T; they satisfy $\psi(-\zeta) = \psi(\zeta)$. A smooth cut-off (or "smudge") function is defined (e.g. in GS or L) as

$$f_{cut}(\zeta) = \exp(-\frac{1}{1-\zeta^2})I(|\zeta| < 1);$$

"bump function" is

$$f_{bump}(\zeta) = \frac{f_{cut}(\zeta)}{\int_{-1}^{1} f_{cut}(\zeta)d\zeta}.$$

Consider sets V, U defined as

$$V = \cup\left([a_i, b_i] \cup [-b_i, -a_i]\right) \subset \cup(a_i - \varepsilon, b_i + \varepsilon) \cup (-b_i - \varepsilon, -a_i + \varepsilon) = U, \quad (27)$$

where $a_i \neq b_i$ and the intervals and ε are such that the only two intervals in U that could intersect would correspond to some i with $b_i = -a_i$; define the function

$$f_V(\zeta) = I(|\zeta| \in V) * f_{bump}(\frac{2\zeta}{\varepsilon})\frac{2}{\varepsilon}.$$

This function has the property that it equals 1 on V, 0 outside of U and takes values between 0 and 1.

For any $\xi \in R, p \geq 0, \varepsilon > 0$ consider a closed set $V_\xi = [\xi - \alpha, \xi + \alpha] \cup [-\xi - \alpha, -\xi + \alpha]$ and the function $f_{V_\xi}(\zeta)$, defined above. Define $f_{\xi,p}(\zeta) = (\zeta - \xi)^p f_{U_\xi, V_\xi}(\zeta)$. This function has the property that

$$\frac{d^l f_{\xi,p,\varepsilon}}{d\zeta^l}(\xi) = (-1)^l \frac{d^l f_{\xi,p,\varepsilon}}{d\zeta^l}(-\xi) = \begin{cases} p! & \text{if } l = p; \\ 0 & \text{otherwise.} \end{cases}$$

All the functions, $f_{bump}, f_V, f_{\xi,p,\varepsilon}$ are in T_R.

Proof of Theorem 2.

(i) Let e be small enough that closed $e-$neighborhoods of all the points of singularity and discontinuity of γ_o and $\dot{\gamma}_o$ do not intersect in $(-\bar{\zeta}, \bar{\zeta})$. Define the union of open intervals that is the compliment to this set in $(-\bar{\zeta}, \bar{\zeta})$ by U. Construct for a small enough ε a corresponding union of closed intervals, $V \subset U$ that can be defined by (27). Define $\mu(\zeta) = f_V(\zeta)$. Then $\frac{d^p \mu}{d\zeta^p}(s_l) = 0$ for all s_l, p; integrals $\int \dot{\gamma}_o(\zeta, \theta)\mu(\zeta)d\zeta$ and $\int \gamma_o(\zeta, \theta)\mu(\zeta)d\zeta$ are defined for any θ. The inverse Ft's $r_y(z, \theta) = Ft^{-1}(\gamma_o(-\zeta, \theta)\mu(-\zeta))$ and $r_{xy}(z, \theta) = Ft^{-1}(\mathbf{i}\dot{\gamma}_o(-\zeta, \theta)\mu(-\zeta)$ exist. Since $\varepsilon_{yo}(\zeta) = \gamma_o(\zeta, \theta^*)\phi(\zeta)$, $\varepsilon_{xyo}(\zeta) = -\mathbf{i}\dot{\gamma}_o(\zeta, \theta^*)\phi(\zeta)$, $\dot{\gamma}_o(\zeta, \theta)$ and $\varepsilon_{xyo}(\zeta)$ are ordinary locally integrable functions in T' and $\varepsilon_{yo}(\zeta)$ and $\gamma_o(\zeta,\theta)$ are continuous and satisfy (5), the products $\varepsilon_{yo}(\zeta)\dot{\gamma}_o(-\zeta,\theta)$ and $\varepsilon_{xyo}(\zeta)\gamma_o(-\zeta,\theta)$ are well defined in T' . Thus the integral (where $\mu \in T_R$)

$$\int \left[\varepsilon_{yo}(\zeta)\mathbf{i}\dot{\gamma}_o(-\zeta,\theta)\mu(-\zeta) + \varepsilon_{xyo}(\zeta)\gamma_o(-\zeta,\theta)\mu(-\zeta)\right] d\zeta$$

exists. Since $\varepsilon_{yo}(\zeta) = \gamma_o(\zeta,\theta^*)\phi(\zeta)$, $\varepsilon_{xyo}(\zeta) = -\mathbf{i}\dot{\gamma}_o(\zeta,\theta^*)\phi(\zeta)$ the value of the integral is zero for $\theta = \theta^*$. Moreover, because the functions μ are zero together with all the derivatives at singularity points, ε_{yo} can be replaced by ε_y providing:

$$\begin{aligned} & \int \left[\varepsilon_{yo}(\zeta)\mathbf{i}\dot{\gamma}_o(-\zeta,\theta)\mu(-\zeta) + \varepsilon_{xyo}(\zeta)\gamma_o(-\zeta,\theta)\mu(-\zeta)\right] d\zeta \\ = \; & (\mathbf{i}\varepsilon_y\dot{\gamma}_o(-\zeta,\theta),\mu(-\zeta)) + (\varepsilon_{xy}\gamma_o(-\zeta,\theta),\mu(-\zeta)) \end{aligned}$$

By applying Parseval identity to generalized functions this leads to

$$\begin{aligned} & (W_y, r_{xy}(\theta)) + (W_{xy}, r_y(\theta)) \\ = \; & \int \left[W_y(z)r_{xy}(z,\theta) + W_{xy}(z)r_y(z,\theta)\right] dz \end{aligned}$$

where the functionals are expressed via integrals for ordinary locally integrable functions.

Multiplying and dividing by the non-zero function $p(z)$ does not change the integral. Then by law of iterated expectations

$$\int \frac{1}{p(z)} E_{|z}\left(Yr_{xy}(z,\theta) + XYr_y(z,\theta)\right) p(z)dz = E\left(\frac{Yr_{xy}(z,\theta) + XYr_y(z,\theta)}{p(z)}\right).$$

This concludes the proof of (i).

(ii) By Assumption 5(iii) there exists a sequence $\xi_n \to 0$ such that $\gamma_o(\zeta,\theta) \neq 0$ for $\zeta : |\zeta - \xi_n| < \varepsilon_n < |\xi_n|$ and is continuous in those intervals; without loss of generality assume that $\zeta \in U$ defined in (i). Consider the function

$$\mu_n(\zeta) = \frac{1}{2}\{f_{bump}(\frac{\zeta - \xi_n}{\varepsilon_n}) + f_{bump}(\frac{\zeta + \xi_n}{\varepsilon_n})\}$$

The function $\frac{\mu_n(-\zeta)}{\gamma_o(-\zeta,\theta)}$ is a continuous function with bounded support. Set $r_{1yn}(z,\theta) = Ft^{-1}(\frac{\mu_n(-\zeta)}{\gamma_o(-\zeta,\theta)})$.Then for any n we get

$$\begin{aligned} \int \varepsilon_{yo}(\zeta)\overline{\gamma_o(-\zeta,\theta)}^{-1}\mu_n(-\zeta)d\zeta &= (\varepsilon_y \cdot \overline{\gamma_o(-\zeta,\theta)}^{-1}, \mu_n(-\zeta)) \\ &= \int E(Y|z)Ft^{-1}(\overline{\gamma_o(-\zeta,\theta)})^{-1}\mu_n(-\zeta))dz \\ &= E\left(\frac{Yr_{y1n}(z,\theta)}{p(z)}\right), \end{aligned}$$

where the first equality follows from the fact that $(\varepsilon_{ys}\overline{\gamma_o(-\zeta,\theta)}^{-1}, \mu_n(-\zeta)) = 0$ (since $\zeta \in U$), the second by Parseval identity and the third by multiplying and dividing by $p(z) > 0$ and iterated expectation; the integral exists for each n. For θ^* we get $(\gamma_o(\zeta) = \overline{\gamma_o(-\zeta)})$

$$\begin{aligned} E\left(\frac{Yr_{y1n}(z,\theta^*)}{p(z)}\right) &= \int \varepsilon_{yo}(\zeta)\gamma_o(\zeta,\theta^*)^{-1}\mu_n(-\zeta)d\zeta = \int \phi(\zeta)\mu_n(-\zeta)d\zeta \\ &= \frac{1}{2}\left[\phi(\xi_n) + \phi(-\xi_n)\right] + O(\varepsilon_n) \end{aligned}$$

This converges to $\phi(0) = 1$.

(iii) Consider any $s_l \geq 0$. Below all relevant functions are subscripted by l.

For ε as defined in (i) define the function $\mu_{l,i}(\zeta) = f_{s_l,i,\varepsilon}(\zeta) \in T_R$, then $\mu_{l,i}^{(i)}(0) \neq 0$, but $\mu_{l,i}^{(k)}(0) = 0$, $k = 0, ...i-1, i+1, ...\bar{k}+1$ and support of $\mu_{l,i}$ is given by $I(|\zeta - s_l| < \varepsilon) + I(|\zeta + s_l| < \varepsilon)$; denote the derivative of $\mu_{l,i}$ by $\mu_{l,i}'$. For a sequence $\varepsilon_n \to 0$ consider $f_{V_n}(\zeta)$ for $U_n = \{\zeta : |\zeta - s_l| < \varepsilon_n\} \cup \{\zeta : |\zeta + s_l| < \varepsilon_n\}; V_n = \{\zeta : |\zeta - s_l| \leq \frac{\varepsilon_n}{2}\} \cup \{\zeta : |\zeta + s_l| < \frac{\varepsilon_n}{2}\}$

and define $\mu_{l,i,n}(\zeta) = \mu_{li}(\zeta) f_{U_n,V_n}(\zeta))$. The functions $\mu_{.}$ are in T_R. Denote by $r_{xys,l,i,n}(z)$ the inverse Ft: $Ft^{-1}(\mu_{l,i,n}(-\zeta))$ and by $r_{ys,l,i,n}(z)$ the inverse Ft: $Ft^{-1}(\mathbf{i}\mu'_{l,i,n}(-\zeta))$; they exist in T. The vector $r_{xys,l,n}(z)$ is defined to have $r_{xys,l,i,n}(z)$ as its $i-$th component; vector $r_{ys,l,n}(z)$ is defined similarly. Define by M_l the diagonal matrix with non-zero diagonal entries $\{M_l\}_{ii} = \mu^{(i)}_{l,i,n}(0) \equiv \mu^{(i)}_{l,i}(0)$, $i = 0, ...\bar{k}$.

Consider now the vector $(\varepsilon_y, \mu'_{l,n})$ with components $(\varepsilon_y, \mu'_{l,i,n}(-\zeta))$ and $(\varepsilon_{xy}, \mu_{l,n})$ with $(\varepsilon_{xy}, \mu_{l,i,n}(-\zeta))$. Since the matrices $\Gamma_y(s_l, \theta)$, $\Gamma_{xy}(s_l, \theta)$ and M_l are invertible the expression

$$\Gamma_y(s_l, \theta)^{-1} M_l^{-1}(\varepsilon_y, \mu'_{l,n}) + \Gamma_{xy}(s_l, \theta)^{-1} M_l^{-1}(\varepsilon_{xy}, \mu_{l,n}) \tag{28}$$

is finite for every n. By Parseval identity

$$\begin{aligned} (\varepsilon_y(\zeta), \mu'_{l,i,n}(-\zeta)) &= (W_y(z), r_{y,l,i,n}(z)) \\ &= \int W_y(z) r_{y,l,i,n}(z) dz, \end{aligned}$$

the last equality follows since W_y is locally integrable. Thus by arguments similar to those in (i) and (ii) this integral is $E\frac{Y r_{y,l,i,n}(z)}{p(z)}$ so that $(\varepsilon_y, \mu'_{l,n}) = E\frac{Y r_{y,l,n}(z)}{p(z)}$ and analogously $(\varepsilon_{xy}, \mu_{l,n}) = E\frac{XY r_{xy,l,n}(z)}{p(z)}$. We need to establish that limits as $n \to \infty$ exist. First, note that

$$\begin{aligned} \left| \int \varepsilon_{yo}(\zeta) \mu'_{l,i,n}(-\zeta) d\zeta \right| &= \left| \int \varepsilon_{yo}(\zeta) f_{\zeta_l,i,\varepsilon}(-\zeta) f_{U_n,V_n}(-\zeta) d\zeta \right| \\ &\leq \max_U \left| \varepsilon_{yo}(\zeta) f_{\zeta_l,i,\varepsilon}(-\zeta) \right| 2\varepsilon_n \end{aligned}$$

and goes to zero;

$$(\varepsilon_{ys}, \mu'_{l,i,n}(-\zeta)) = \Sigma^{\bar{k}}_{k \geq i} \gamma(s_l, \theta^*)(-1)^i \begin{pmatrix} k+i-1 \\ i-1 \end{pmatrix} \mu^{(i)}_{l,i}(s_l) \phi^{(k-i+1)}(s_l) \tag{29}$$

and does not depend on n, finally, $\varepsilon_y = \varepsilon_{yo} + \varepsilon_{ys}$, so the limit exists. By a similar representation for $-(\varepsilon_{xys}, \mu_{l,i,n}(-\zeta))$ existence of (19) is established. For $\theta = \theta^*$ using (25,26) for ε_{ys} and ε_{xys} in (28) leads to

$$\Gamma_y(s_l, \theta^*)^{-1} M_l^{-1}(\varepsilon_{ys}, \mu'_{l,n}) + \Gamma_{xy}(s_l, \theta^*)^{-1} M_l^{-1}(\varepsilon_{xys}, \mu_{l,n}) = 0.$$

Note that the same considerations apply to singularity at $-s_l$ with the difference that the $\Gamma_.(-s_l, \theta)$ matrices now are complex conjugate to $\Gamma_.(s_l, \theta)$. Combining provides the real part in (19).

(iv) Consider the first component of $\Gamma_y(s_l, \theta)^{-1} M_l^{-1}(\varepsilon_{ys}, \mu'_{l,n})$ with $\mu'_{l,n}, M_l$ defined in (iii); this component is $\{\Gamma_y(s_l, \theta)^{-1} M_l^{-1}\}_{11}(\varepsilon_{ys}, \mu'_{l,1,n})$. Note that $\mu'_{l,1,n} = \mu_{l,0,n}$, recall that $\mu \in T_R$. We see that $\phi(s_l)$ equals $\lim\{\Gamma_y(s_l, \theta^*)^{-1} M_l^{-1}\}_{11} \int \varepsilon_{ys}\mu_{l,0,n} d\zeta$. Define

$$r_{ysl,1,n}(z, \theta) = \{\Gamma_y(s_l, \theta)^{-1} M_l^{-1}\}_{11} Ft^{-1}(\mu_{l,0,n}(-\zeta)).$$

Similarly to above by Parseval identity $\{\Gamma_y(s_l, \theta)^{-1} M_l^{-1}\}_{11}(\varepsilon_{ys}, \mu_{l,0,n}) = E(\frac{Y r_{ysl,1,n}(z,\theta)}{p(z)})$ and $\phi(s_l) = \lim E(\frac{Y r_{ysl,1,n}(z,\theta)}{p(z)})$. For $s_0 = 0$ we have $\phi(0) = 1$. Thus (20) follows.

Consider now for $s_l \neq 0$ the function

$$\mu_{l,n}(\zeta) = \frac{1}{2}\{f_{bump}(\frac{\zeta - s_l - \xi_n}{\varepsilon_n}) + f_{bump}(\frac{\zeta - s_l + \xi_n}{\varepsilon_n})\}$$

similar to the one in (ii) and define $r_{slo,n}(z) = Ft^{-1}(\frac{\mu_{l,n}(-\zeta)}{\gamma_o(-\zeta,\theta)})$. For this function $E(\frac{Y r_{slo,n}(z,\theta)}{p(z)}) = \int \varepsilon_y(\zeta)\frac{\mu_{l,n}(-\zeta)}{\gamma_o(-\zeta,\theta)} d\zeta$ exists and at θ^* converges to $\phi(s_l)$. Thus (21) follows.■

Proof of Theorem 3.

Let the vector $Q(z, \theta)$ denote the vector of functions for which expectations are taken in $E(Q)$; partition $Q(z, \theta)$ into $Q_I(z, \theta)$ corresponding to expressions in (17, 18) and $Q_{II}(z, \theta)$ for (19, 21). Then the matrix $\frac{\partial}{\partial \theta^T} E(Q(z, \theta))$

is a block matrix

$$\begin{pmatrix} \frac{\partial}{\partial\theta_I^T}E(Q_I(z,\theta)) & \cdot \\ \cdot & \frac{\partial}{\partial\theta_{II}^T}E(Q_{II}(z,\theta)) \end{pmatrix}$$

and it is sufficient to show that $\frac{\partial}{\partial\theta_I^T}E(Q_I(z,\theta))$ has rank m_I and $\frac{\partial}{\partial\theta_{II}^T}E(Q_{II}(z,\theta))$ has rank m_{II}.

For θ_I first note that interchange of differentiation with respect to the parameter and integration (taking expected value) for $\frac{\partial}{\partial\theta_I^T}EQ_I(z,\theta)$ follows from continuity in ζ of all the functions in the integrals and their continuous differentiability with respect to θ, so that $\frac{\partial}{\partial\theta_I}E(Q_I(z,\theta)) = E(\frac{\partial}{\partial\theta_I}Q_I(z,\theta))$. One can choose m_I functions μ defined in proofs of Theorem 2(i,ii) that are functionally independent and under Assumption 6 the corresponding m_I conditions of type $E(\frac{\partial}{\partial\theta_I}Q_I(z,\theta))$ will provide a rank m_I submatrix.

If for the functions μ in expressions $Q_{II}(z,\theta)$ in (iii,iv) of Proof of Theorem 2 the matrix $\frac{\partial}{\partial\theta_{II}}E(Q_s(z,\theta))$ has rank less than m_{II} consider varying the functions $\mu_.$ for all possible values of non-zero derivatives at the points s_l; the rank cannot be deficient over all such choices without violation of Assumption 6.■

Appendix B

Three main problems with (S) are listed below.

1. The issue of decomposition.

(S) claims that any generalized function in T' can be decomposed into a sum of an ordinary function and a singular function; this decomposition is used in formula (13) of S., Theorem 1.

There is no proof of existence of such a decomposition in the literature, as the example of the function in Section 2.1 in (10) shows no unique decomposition into a singular and regular ordinary function exists in T' (nor in D'). The attempted proof in (S, Supplementary material, p.3) is incorrect.

Indeed it states: "The result directly follows from the fact that every generalized function can be written as the derivative of order $k \in N$ of some continuous function $c(t)$ (Theorem III in Temple (1963) establishes this for a class of generalized functions including those considered here as a particular case). *[-my comment: this is correct]* At every point t where $c(t)$ is k times differentiable in the usual sense, the generalized function can be written as an ordinary function, while at every point where $c(t)$ is not k times differentiable, a delta function derivative is created in the differentiation process.*[my comment: this is incorrect, see (a) below]* The fact that the two pieces are additively separable follows from the linear nature of the space of generalized functions. *[my comment: this is incorrectly applied: see (b) below]"*

(a). Consider the function $b(\zeta) = |\zeta|^{-\frac{3}{2}}$; it is a "weak" (or generalized) second derivative of the continuous function $c(\zeta) = -4\,|\zeta|^{\frac{1}{2}}$. The first weak derivative, $-2\,|\zeta|^{-\frac{1}{2}}\,sign(\zeta)$, is an ordinary function that is summable and so gives a regular functional and is an ordinary function that is at the same time a generalized function. At point $\zeta = 0$ it is not differentiable in the ordinary sense; yet *no delta-function or its derivative appears.*

(b) Any generalized function is either regular or singular; the "ordinary" function $b(\zeta)$ above is a singular generalized function (see (10)). Since generalized functions are generally not defined pointwise a pointwise argument cannot be helpful.

An additional assumption would have to be used to establish formula (13) in Theorem 1 of (S).

2. Validity of products.

Validity of products in the space of generalized functions needs to be established to provide a correct proof of the identification result. Neither the paper nor the supplementary material in (S) provides a complete correct proof indicating in which space of generalized functions the multiplication operations are valid; in fact as Lemma 1 here shows multiplication may not be valid in T' under the Assumptions (despite the claim in (S)).

3. Definition 2 in (S) leads to inappropriate weights.

Def. 2 proposes the function

$$\mu(\zeta) = \sum_{k=0}^{\infty} \frac{1}{k!} d^k(\zeta^k \lambda(\zeta))/d\zeta^k$$

for λ that satisfies S, Def. 1 (λ is an analytic function). We have

$$\begin{aligned}
& \sum_{k=0}^{\infty} \frac{1}{k!} d^k(\zeta^k \lambda(\zeta))/d\zeta^k \\
= & \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=0}^{k} C_i^k \left(d^i \zeta^k/d\zeta^i\right) \left(d^{k-i}\lambda(\zeta)/d\zeta^{k-i}\right) \\
= & \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=0}^{k} C_i^k \frac{k!}{(k-i)!} \zeta^{k-i} \left(d^{k-i}\lambda(\zeta)/d\zeta^{k-i}\right) \\
= & \sum_{k=0}^{\infty} \sum_{i=0}^{k} C_i^k \frac{1}{i!} \zeta^i \left(d^i \lambda(\zeta)/d\zeta^i\right).
\end{aligned}$$

Consider the values and derivatives of the function μ at zero; they are zero and thus are not suitable for the weighting functions. Indeed

$$\begin{aligned}
\mu(0) &= \sum_{k=0}^{\infty} \lambda(0), \text{ so } \lambda(0) = 0 \text{ and } \mu(0) = 0 \\
\mu'(0) &= \sum_{k=0}^{\infty} \sum_{i=0}^{k} C_i^k \frac{1}{i!} [\zeta^{i-1} \left(d^i \lambda(\zeta)/d\zeta^i\right) + \zeta^i \left(d^{i+1}\lambda(\zeta)/d\zeta^{i+1}\right) |_{\zeta=0} \\
&= \sum_{k=0}^{\infty} [\lambda'(0) + k\lambda'(0)], \text{ so } \lambda'(0) = 0, \text{ and } \mu'(0) = 0,
\end{aligned}$$

etc., implying $\mu^{(k)}(0) = 0$ for any k.

References

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

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

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

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

[5] Hausman J., W. Newey, H. Ichimura and J. Powell (1991), Measurement error in Polynomial Regression Models, Journal of Econometrics, v. 50, 273-295.

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

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

[8] Lighthill, M.J. (1959) Introduction to Fourier Analysis and Generalized Functions, Cambridge University Press.

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

[10] Schwartz, L. (1954) "Sur l'impossibilité de la multiplication des distributions"C. R. Acad. Sci.,239, No. 15, 847–848.

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

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

[13] Wang, L. and C. Hsiao (1995), Simulation-Based Semiparametric Estimation of Nonlinear Errors-in-Variables Models, working paper, University of Southern California.

[14] Wang, L. and C. Hsiao (2009) Method of Moments Estimation and Identifiability of Semiparametric Nonlinear Errors-in-Variables Models, Journal of Econometrics, in press.

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