On the Completeness Condition in Nonparametric Instrumental Problems

Xavier d'Haultfoeuille*
ENSAE, CREST - INSEE and
Université de Paris I, Panthéon-Sorbonne

Revised version: July 2008

Abstract

The notion of completeness between two random elements has been considered recently to provide identification in nonparametric instrumental problems. This condition is quite abstract, however, and characterizations have been obtained only in special cases. This paper considers a nonparametric model between the two variables with an additive separability and a large support condition. In this framework, different versions of completeness are obtained, depending on which regularity conditions are imposed. This result enables to establish identification in an instrumental nonparametric regression with limited endogenous regressor, a case where the control variate approach breaks down.

Keywords: completeness, nonparametric identification, instrumental variables.

JEL classification numbers: C14, C31, C35.

*ENSAE, 18 boulevard Adolphe Pinard, 75 675 Paris Cedex, France. E-mail: xavier.d'haultfoeuille@ensae.fr. I am very grateful to the editor Jean-Pierre Florens, two referees, Philippe Février, Nicolas Lerner, Xavier Mary and Jean-Marc Robin for their useful comments.

1 Introduction

Let X and Z denote two random elements. X will be said to be complete for Z if, for all measurable real functions h such that $\mathbb{E}[|h(X)|] < +\infty$,

$$\Big(\mathbb{E}[h(X)|Z] = 0 \quad \text{a.s.}\Big) \Longrightarrow \Big(h(X) = 0 \quad \text{a.s.}\Big). \tag{1.1}$$

X will be bounded complete (resp. P-complete) for Z if the same holds for any bounded h (resp. for any h bounded by a polynomial).[1] Completeness is equivalent to the injectivity of the conditional expectation operator. Thus, not surprisingly, it has appeared to be a key identifying condition in nonparametric instrumental problems. Applications include nonparametric instrumental regression under additive separability (see Newey and Powell, 2003, Darolles et al., 2002 and Blundell et al., 2007),[2] local instrumental variables (see Florens et al., 2003) and nonclassical measurement error problems (see Chen and Hu, 2006 and Hu and Schennach, 2008).[3]

This dependence condition is quite abstract though, and a characterization or at least sufficient conditions on the joint distribution of (X, Z) are desirable. Newey and Powell (2003) address the finite support and exponential families cases, but results are still lacking to properly define completeness in terms of dependence between the two variables. The aim of this paper is to go one step in this direction by considering a nonparametric model on (X, Z) for which an additive separability and a large support condition hold. Building on the results of Mattner (1992, 1993) on the completeness of location families, I show that different versions of completeness can be obtained, depending on which regularity conditions are imposed on the error term. Bounded and P-completeness only require

[1] This terminology is in analogy with the notion of complete statistic (see e. g. Lehmann and Scheffé, 1947). Recall that a statistic T is said to be complete (resp. bounded complete) for a statistical model $(P_\theta)_{\theta\in\Theta}$ if for all h (resp. all bounded h), $E_\theta[h(T)] = 0$ for all $\theta \in \Theta$ implies that $h(T) = 0$ a.s. Thus, Z plays the role of θ in equation (1.1). Note also that (1.1) is sometimes referred to as a strong identification condition (see e. g. Florens et al., 1990).

[2] Chernozhukov and Hansen (2005) also rely on a condition which is close to bounded completeness(see their assumption L_1^*) for the identification of quantile treatment effects with instrumental variables.

[3] Indeed, assumption 2.4 of Chen and Hu (2006) and assumption 2 of Hu and Schennach (2008) are equivalent, under technical conditions, to a completeness condition.

mild assumptions, whereas completeness is restrictive. This contrast between the different kinds of completeness is in line with previous results of the statistical literature (see e.g. Hoeffding, 1977, Lehmann, 1986, and Mattner, 1993) and has also been acknowledged by Chernozhukov and Hansen (2005) and Blundell et al. (2007).

Implications for the nonparametric instrumental regressions are also examined. Recent analyses of such models (see e.g. Imbens and Newey, 2006 and Florens et al., 2007) have relied on a control variate approach rather than on a completeness assumption. Conditions for identification are indeed easier to obtain, and the additivity structure of the model can be relaxed. On the other hand, a strict monotonicity assumption is required, which rules out usual models with limited endogenous regressors. The previous result enables to prove the identification of the structural function in a triangular system of simultaneous equations under, roughly, an additive decomposition and a large support condition on the instrumental equation, but without any strict monotonicity condition. This shows that actually, the completeness approach may be more fruitful than the control variate one in some circumstances. Since different versions of completeness provides different identification results, there also appears to be a trade-off in the identification of such models between the regularity condition imposed on the error term of the instrumental equation and the hypothesis on the structural function.

The paper is organized as follows. The main results are given in section two. Section three examines the consequence of these results on the identification of nonparametric instrumental regression. Section four concludes, and the proofs are deferred to section five.

2 Main results

In the sequel, X and Z belong to $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively, with $p \leq q$. X and Z may share elements in common, and we let W denote these common elements, $W \in \mathbb{R}^r$. For instance, in an instrumental nonparametric regression (see e.g. Newey and Powell, 2003), W corresponds to the exogenous components of X. The remaining elements of X and Z are called respectively X_0 and Z_0, so that $X = (X_0, W)$ and $Z = (Z_0, W)$. In this framework, we will say that X is complete (resp. bounded, P-complete) for Z if (1.1) holds for all h such that, for $\mathbb{P}^W-$ almost all w, $h(., w)$ is integrable with respect to $\mathbb{P}^{X_0}$ (resp. bounded, bounded by a polynomial). In the sequel, we suppose that there exists maps μ_1 and ν_1, from respectively $\mathbb{R}^{p-r}$ and $\mathbb{R}^q$ to $\mathbb{R}^{p-r}$, such that

$$X_0 = \mu_1(\nu_1(Z) + \varepsilon_1), \qquad (2.1)$$

and we consider the following assumptions.

A1. $Z_0 \perp\!\!\!\perp \varepsilon_1 \,|\, W$.

A2. For $\mathbb{P}^W$-almost all w, the measure of $\nu_1(Z_0, w)$ is continuous with respect to the Lebesgue measure and its support is $\mathbb{R}^{p-r}$.

A3. For $\mathbb{P}^W$-almost all w, ε_1 admits a continuous density $f_{\varepsilon_1|W}(., w)$.

Assumption A1 is a conditional independence hypothesis. Because mean-independence can always be achieved by a proper normalization,[4] A1 actually strengthens this mean-independence into independence. Note that if μ_1 is known, this assumption is testable in the data in general.

A2 is a continuity and large support condition. It may hold as soon as Z has one continuous component. The large support condition is restrictive but widespread in the literature (see e.g. Manski, 1988, or Lewbel, 2000). Moreover, only $\nu_1(Z)$, not necessarily Z, should satisfy this condition. This means that $p - r$ regressors with large support may be sufficient. This assumption, however, may be too strong, and we consider below alternative assumptions (see proposition 2.3). Lastly, A3 restricts the analysis to the case of a continuous residual. The continuity condition on its density is satisfied by all usual densities with infinite support.[5]

Despite the apparently strong assumption of an additive decomposition into independent terms, the function μ_1 in (2.1) enables to encompass many nonlinear models, beyond the

[4]Indeed, if we let $\widetilde{\nu}_1(Z) = \nu_1(Z) + \mathbb{E}(\varepsilon_1|Z)$ and $\widetilde{\varepsilon}_1 = \varepsilon_1 - \mathbb{E}(\varepsilon_1|Z)$, then $\mathbb{E}(\widetilde{\varepsilon}_1|Z_0, W) = 0 = \mathbb{E}(\widetilde{\varepsilon}_1|W)$.
[5]It fails for the uniform density but this case is ruled out anyway by assumption A4 below.

nonparametric additive models with independent errors (for which $\mu_1(x) = x$). Usual ordered choice models correspond to $\mu_1(x) = \sum_{k=1}^{K} k\mathbb{1}_{]\alpha_{k-1};\alpha_k]}(x)$ (where $\mathbb{1}_A(x) = 1$ if $x \in A$, 0 otherwise) for some given thresholds $\alpha_0 = -\infty < \alpha_1 < ... < \alpha_K = +\infty$.[6] Count data models can also be handled by taking $\mu_1(x) = [\exp(x)]$ (where $[a]$ denotes the integer part of a). Simple tobit models correspond to $\mu_1(x) = \max(0, x)$. These three examples underline the fact that X may not be strictly monotonous in ε_1. Lastly, duration models like the accelerated failure time model or the proportional hazard model also fit in this framework. The first corresponds to $\mu_1(x) = \exp(x)$, while in the second, μ_1 is an unknown increasing function and $-\varepsilon_1$ is distributed according to a Gompertz distribution.

To achieve completeness, further restrictions are required.

A4. $\mathbb{P}^W-$almost surely, the conditional characteristic function $\psi_{\varepsilon_1|W}(., w)$ of ε_1 is infinitely often differentiable in $\mathbb{R}^p \backslash A(w)$ for some finite set $A(w)$ and does not vanish on the real line.

A5. All the moments of $||\varepsilon_1||$ are finite and there exists B and j such that $||\mu_1(t)|| \leq B||t||^j$ (where $||.||$ is the euclidian norm).

A6. ε_1 is gaussian or satisfies, $\mathbb{P}^W$-almost surely on w and for all $x, y \in \mathbb{R}^{p-r}$, there exists $C(.)$ and $k(.)$ such that

$$f_{\varepsilon_1|W}(x + y, w) \leq C(w)(1 + ||x||^2)^{k(w)} f_{\varepsilon_1|W}(y, w).$$

Zero-freeness of the characteristic function is a usual assumption in deconvolution problems (see e.g. Devroye, 1989, Fan and Truong, 1993, Li and Vuong, 1998, Schennach, 2004 and 2007) and is satisfied, among others, by gaussian, Student, Laplace and $\alpha-$stable distributions. The only common continuous distributions that fail to satisfy it are the uniform

[6]Binary choice models are obviously included. Note however that for binary variables X_0, model (2.1) is unnecessary since completeness is simply equivalent to non independence between X_0 and Z_0, conditional on W. When X_0 takes more than two values, it can be shown that the completeness condition is equivalent to the positivity of a variance matrix (see Das, 2005, theorem 2.1). However, it is not obvious to check this condition for a given theoretical model.

and triangular ones. All standard characteristic functions also satisfy the differentiability condition.

Assumption A5 rules out thick tails on the density of ε_1 and restricts the range of nonlinear models between X_0 and Z. It fails for instance with the previous examples of count data and accelerated failure time models, but holds for all the others aforementioned cases. A similar polynomial growth condition is imposed by Schennach (2007) to identify a nonlinear errors-in-variables model with instruments (on this issue, see also Zinde-Walsh, 2007).

Lastly, assumption A6 is rather restrictive. It imposes in particular that $f_{\varepsilon_1|W}(.,w)$ is either gaussian or has heavy tails.[7] The condition holds for instance for Student and $\alpha-$ stable distributions (see Mattner, 1992).

Theorem 2.1 *Suppose that (2.1) and A1-A3 hold. Then*
1) if A4 holds, X is bounded complete for Z.
2) If A4 and A5 hold, X is P-complete for Z.
3) If A4 and A6 hold, X is complete for Z.

Theorem 2.1 gives conditions under which different versions of completeness hold. The intuition of its proof can be explained as follows. First, one can show that completeness is equivalent to the unicity of the following convolution equation in $g(.,w)$ (for almost all w):

$$\int g(t,w)f_{-\varepsilon_1|W}(u-t,w)dt = 0. \tag{2.2}$$

If $g(.,w)$ was integrable, this would imply, by the convolution theorem,

$$\mathcal{F}(g(.,w)) \times \mathcal{F}(f_{-\varepsilon_1|W}(.,w)) = 0. \tag{2.3}$$

where $\mathcal{F}$ denotes the Fourier transform. Then, by assumption A4, $\mathcal{F}(g(.,w)) = 0$, and since the Fourier transform is injective, $g(.,w) = 0$. Actually, the problem is more involved because a priori, $g(.,w)$ is not integrable, so that its usual Fourier transform may not exist.

[7] Put $x = -y$ to see that $1/f_{\varepsilon_1|W}$ must be at most of polynomial order. It can also be shown (see Mattner, 1992) that A6 is implied by the condition $0 < c(w) \leq f_{\varepsilon_1|W}(x,w)(1+||x||)^{\gamma(w)} \leq C(w) < \infty$ for all $x \in \mathbb{R}^{p-r}$ and some real $c(w), C(w)$ and $\gamma(w) > 0$.

To circumvent this issue, I rely on the techniques developed by Ghosh and Singh (1966) and Mattner (1992) to show completeness of location families.

Theorem 2.1 shows that in model (2.1), bounded completeness holds under rather weak conditions. Many of the usual densities also satisfy the moment condition which ensures P-completeness.[8] Completeness, on the other hand, is obtained under the restrictive hypothesis A6. As theorem 2.1 only provides sufficient conditions, one may wonder whether completeness actually holds under milder conditions. If it seems difficult to provide a full characterization of completeness, the following proposition shows that it really imposes stringent condition on the distribution of ε_1.

Proposition 2.2 *Suppose that (2.1), A1-A3 hold and $\mu_1(t) = t$. Assume also that, for $\mathbb{P}^W-$almost all w, ε_1 is not normal conditional on $W = w$ and there exists $\delta_1, \delta_2 > 0$ such that $\mathbb{E}\left(\exp\left(\delta_1 ||\varepsilon_1||^{1+\delta_2}\right)|W = w\right) < +\infty$. Then X is not complete for Z.*

Hence, if $f_{\varepsilon_1|W}$ has light tails, X cannot be complete for Z. On the other hand, X can still be bounded or P-complete for Z in such situations.

As mentioned above, the large support assumption A2 is rather strong. It is possible, though, to relax it, at the cost of imposing regularity on the distribution of ε_1. For the sake of simplicity, we restrict here to the case where X_0 is real ($p - r = 1$).

A2'. For $\mathbb{P}^W$-almost all w, the measure of $\nu_1(Z_0, w)$ is continuous with respect to the Lebesgue measure.

A7. There exists $(a_k(w))_{k\in\mathbb{N}}$ such that, for all $x \in \mathbb{R}$ and for $\mathbb{P}^W$-almost all w,

$$f_{\varepsilon_1|W}(x, w) = \sum_{k=0}^{\infty} a_k(w) x^k. \tag{2.4}$$

Moreover, there exists $r_0(w) > 0$ such that $f_{\varepsilon_1|W}(., w)$, as a function on $\mathbb{C}$ defined by (2.4), is bounded on $\{z / |\text{Im}(z)| < r_0(w)\}$.

[8] Moreover, a density which does not fulfill condition A5 has heavy tails and thus is likely to satisfy A6.

Assumption A2' will generally hold if Z_0 contains a continuous regressor. The first part of assumption A7 states that $f_{\varepsilon_1|W}(.,w)$ is entire. Examples of entire functions include the polynomials, the exponential function and all compositions of these functions (including gaussian densities). On the other hand, all densities with support different from $\mathbb{R}$ are not entire. Other counterexamples include the Cauchy and Student distributions. The second part of A7 is a technical condition which is satisfied for instance by gaussian densities.

Proposition 2.3 *Suppose that (2.1), A1, A2', A3, A4 and A7 hold. Then X is bounded complete for Z.*

Proposition 2.3 shows that the large support condition can be dropped, but at the price of restricting the range of the densities of ε_1.

The easiest way to interpret (2.1) is that Z causes X. However, it may be convenient sometimes to suppose instead that X causes Z. In the measurement error models of Chen and Hu (2006) and Hu and Schennach (2008) for instance, their condition on the injectivity of operators can be restated into completeness of the unobserved variable X_0 for the measure Z_0. In this case, the model (2.1) is unnatural since one would prefer to write the measure as a function of the unobserved variable and an independent error, i.e. a model of the form

$$\mu_2(Z) = \nu_2(X) + \varepsilon_2, \tag{2.5}$$

where μ_2 and ν_2 are maps from $\mathbb{R}^q$ (resp. $\mathbb{R}^p$) to $\mathbb{R}^{q-r}$. The standard measurement error model, for instance, corresponds to $\mu_2(z_0, w) = z_0$ and $\nu_2(x_0, w) = x_0$. When $\mu_2(., w)$ is one-to-one, the model writes $Z_0 = \mu_2^{-1}(\nu_2(X) + \varepsilon_2)$ and is similar to (2.1). However, in general we cannot simply switch X_0 and Z_0 in (2.1) to obtain completeness, as the simple example $Z_0 = 1$ shows. In such a model, indeed, Z_0 would not necessarily be informative enough on X_0 for completeness to hold.

We also assume the following hypotheses, which are close to A1, A2 and A4.

A8. $X_0 \perp\!\!\!\perp \varepsilon_2 \,|\, W$.

A9. For $\mathbb{P}^W$-almost all w, $\nu_2(., w)$ is a one-to-one mapping on $\mathbb{R}^{q-r}$.

A10. $\mathbb{P}^W$−almost surely, the characteristic function $\psi_{\varepsilon_2|W}$ of ε_2 conditional on W has isolated zeros.

Assumption A8 is exactly equivalent to A1. Assumption A9 is similar but stronger than the large support condition A2. Indeed, $\nu_2(., w)$ is imposed to be one-to-one, so that here $q = p$. A10, on the other hand, is weaker than A4 and holds for all usual distributions, including the uniform and triangular ones. Actually, it holds for all distribution with exponential tails, because then the corresponding characteristic function is holomorphic on a strip of the complex plane and thus has isolated zeros (see Rudin, 1987, p. 208). The Fejer - de la Vallee Poussin density $x \mapsto (\pi x^2)^{-1}(1 - \cos(x))$ is a counterexample of a distribution which violates A10, as its characteristic function is equal to $t \mapsto \max(1 - |t|, 0)$.

Proposition 2.4 *Suppose that (2.5) and A8-A10 hold. Then X is complete for Z.*

Thus, even if the completeness condition is asymmetric in X and Z, to a certain extent the roles of X and Z in model 2.1 can be exchanged. The conditions on ν_2 are stronger than the one required for theorem 2.1 to hold, but completeness and not only bounded or P-completeness is achieved under weak restrictions on the distribution of ε_2.

3 Implications for the nonparametric instrumental regression

In this section, we apply theorem 2.1 to the identification of nonparametric instrumental regressions. Let us consider the following triangular system:

$$\begin{cases} Y &= \varphi(X) + \eta \\ X_0 &= \mu_1(\nu_1(Z) + \varepsilon_1) \end{cases} \quad \mathbb{E}(\eta|Z) = 0 \tag{3.1}$$

In this model, X_0 are the endogenous regressors, W are exogenous covariates and Z_0 denote the instruments. The aim is to recover the structural function φ. This system is close to the one studied by Newey et al. (1999), although we allow for nonlinearity in the instrumental equation. A main restriction is the additive separability assumption of the first equation. This is the price to pay for a rather weak exogenous condition $\mathbb{E}(\eta|Z) = 0$. In particular, heteroscedasticity is permitted in this framework.

Note that it is possible, through a control variate approach, to relax additive separability under full independence between Z and (ε, η). Recent contributions include Chesher (2003), Imbens and Newey (2006) and Florens et al. (2008) (see Chesher, 2007, for a survey). However, strict monotonicity in the error term of the instrumental equation is required to identify this error. Hence, this approach generally rules out limited endogenous regressor and cannot be applied to model (3.1) unless μ_1 is one-to-one.[9] The completeness approach, one the other hand, can be applied with virtually no assumption on this function.

Proposition 3.1 *Suppose that (3.1) and A1-A3 hold. Then φ is identified if one of the following conditions is satisfied:*
1) A4 holds and $\varphi(., w)$ is bounded for $\mathbb{P}^W$-almost all w ;
2) A4-A5 hold and $\varphi(., w)$ is bounded by a polynomial for $\mathbb{P}^W$-almost all w ;
3) A4 and A6 hold.

Proposition 3.1 shows that to recover φ, there is a trade-off between the regularity conditions imposed on model (3.1) and the assumptions on the function φ itself. The first condition of the proposition is useful when X_0 has a finite support, but imposes a strong restriction on $\varphi(., w)$ otherwise. Linear forms, for instance, cannot be handled by this case. The second widens considerably the range of identified models, at the price of the moment condition on ε_1 and the polynomial growth restriction on μ_1. Lastly, if one is reluctant to make any assumption on φ, identification is achieved under strong restrictions on ε_1.

[9] One could redefine the instrumental equation as $X_0 = \widetilde{\mu}(Z, \widetilde{\varepsilon}_1)$ with $\widetilde{\mu}_1$ strictly increasing in $\widetilde{\varepsilon}_1$ but then the independence condition $Z \perp\!\!\!\perp (\widetilde{\varepsilon}_1, \eta)$ would not hold anymore in general (see Florens et al., 2007, for a discussion on this point).

4 Conclusion

This paper provides general sufficient conditions to achieve varieties of completeness conditions, and apply these results to the nonparametric instrumental regression. Two questions on this topic are left for future research. Firstly, one can wonder whether the assumption of additive decomposition into independent parts could be weaken. Secondly, the adaptation of the results above to the identification condition of Chernozhukov and Hansen (2005) (see their assumption L_1^*) in the context of nonseparable models remains a challenging issue.

5 Proofs

5.1 Theorem 2.1

For all h, let $\widetilde{h}(t,w) = h(\mu_1(t), w)$. By A1,

$$\begin{aligned} \mathbb{E}[h(X)|Z] &= \mathbb{E}[\widetilde{h}(\nu_1(Z)+\varepsilon_1, W)|Z] \\ &= \int \widetilde{h}(\nu_1(Z)+u, W) f_{\varepsilon_1|W}(u,W)du \quad \text{a.s.} \\ &= \int \widetilde{h}(t, W) f_{-\varepsilon_1|W}(\nu_1(Z)-t, W)dt \qquad \text{a.s.} \end{aligned}$$

By A2 (and conditional on W), $\nu_1(Z)$ admits a continuous distribution whose support is $\mathbb{R}^{p-r}$. Thus,

$$\mathbb{E}[h(X)|Z] = 0 \text{ a.s.} \Leftrightarrow \int \widetilde{h}(t,w) f_{-\varepsilon_1|W}(u-t,w))dt = 0 \ \lambda \otimes \mathbb{P}^W - \text{a. e. in } (u,w) \tag{5.1}$$

where λ denotes the Lebesgue measure. Because $\nu_1(Z) + \varepsilon_1$ also admits a continuous distribution and its support is $\mathbb{R}^{p-r}$, it follows that

$$h(X) = 0 \text{ a.s.} \Leftrightarrow \widetilde{h}(t,w) = 0 \ \lambda \otimes \mathbb{P}^W - \text{a. e. in } (u,w). \tag{5.2}$$

Moreover, $h(X)$ integrable (resp. h bounded) implies that $\widetilde{h}(\nu_1(Z) + \varepsilon_1)$ is integrable (resp. $\widetilde{h}$ is bounded). Similarly, by A5, if $h(.,w)$ is bounded by a polynomial, $\widetilde{h}(.,w)$ is also bounded by a polynomial. Hence, to prove completeness (resp. bounded completeness, P-completeness), it suffices to prove that for all g such that $g(\nu_1(Z)+\varepsilon_1)$ is integrable (resp. g is bounded, bounded by a polynomial), $\mathbb{P}^W-$almost surely in w,

$$\int g(t,w)f_{-\varepsilon_1|W}(u-t,w)dt = 0 \quad \text{a.e. in } u \Rightarrow g(t,w)=0 \quad \text{a.e. in } t \tag{5.3}$$

This statement corresponds to the completeness of the location family with density $f_{-\varepsilon_1|W}$, except that the left part of (5.3) holds almost everywhere and not everywhere. But in theorem 1.3 of Mattner (1992) (and hence in his theorem 1.1), the statement also holds almost everywhere, so that we can apply it to obtain part 3 of the theorem.

To show part 1, we adapt the proof of theorem 2.4 of Ghosh and Singh (1966). Let L^1 (resp. L^∞) denote the space of equivalent classes of integrable (resp. essentially bounded) functions with respect to the Lebesgue measure. Let w be such that $g(.,w)\in L^\infty$, $\psi_{\varepsilon|W}(.,w)$ does not vanish anywhere and the left part of (5.3) holds (the set of such w being of probability one). Let $f_{w,u}(x)=f_{-\varepsilon_1|W}(u-x,w)$, $\mathcal{P}_w = \text{span}\left\{f_{w,u},\ u\in\mathbb{R}^{p-r} \ / \ \int g(t,w)f_{w,u}(t)dt=0\right\}$ and $\mathcal{Q}_w=\{f_{w,u}\ /\ u\in\mathbb{R}^{p-r}\}$.

Let $\mathcal{R}_w=\{u\ /\ f_{w,u}\in\mathcal{P}_w\}$. Because the Lebesgue measure of ${}^c\mathcal{R}_w$ is zero, there exists a sequence $(u_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{R}_w$ such that $u_n\to u$ for all $u\in {}^c\mathcal{R}_w$. By continuity of $f_{-\varepsilon_1|W}(.,w)$ and Scheffé's theorem (see e.g. van der Vaart, 1998, p. 22), $\int|f_{w,u_n}(t)-f_{w,u}(t)|dt\to 0$. Thus $\mathcal{Q}_w$ is included in the closure of $\mathcal{P}_w$ (for the L^1-norm).

Now, by A4 and Wiener's tauberian theorem (see e.g. Rudin, 1991, p. 229), $\mathcal{Q}_w$ is dense in L^1. Thus, $\mathcal{P}_w$ is dense in L^1. By continuity of the linear form $\phi\mapsto\int g(t,w)\phi(t)dt$ and the Riesz theorem (see e.g. Rudin, 1987, p. 130), $g(t,w)=0$ for almost every t and almost all w.

Lastly, let us turn to part 2. First, because it is integrable, $f_{-\varepsilon_1|W}(.,w)\in\mathcal{S}'$, the space of tempered distribution (see Rudin, 1991, p. 191, example d). Moreover, $g(.,w)$ is bounded by a polynomial, so that $g(.,w)\in\mathcal{S}'$ (see Rudin, 1991, p. 191, example d). Lastly, the function

$$c(.,w): u\mapsto\int g(t,w)f_{-\varepsilon_1|W}(u-t,w)dt$$

equals zero almost everywhere. Hence, it is the zero distribution and, as such, is tempered.

Now let $g_n(.,w) = g(.,w) \times \mathbb{1}_{[-n,n]}(.)$. $g_n(.,w)$ is a tempered distribution with compact support, so that it belongs to the space of quickly decreasing distributions (see Schwartz, 1973, p. 244). Let us show that $g_n(.,w)$ converges to $g(.,w)$ in $\mathcal{S}'$. We have to prove that

$$\int g_n(u,w)\phi(u)du \to \int g(u,w)\phi(u)du$$

for all $\phi \in \mathcal{S}$, the space of rapidly decreasing functions (see e.g. Rudin, 1991, p. 161). Let Φ be any bounded set in $\mathcal{S}$, the space of rapidly decreasing functions. There exists (see Schwartz, 1973, p. 235) a continuous function b with $b(x) = o(|x|^{-m})$ as $|x| \to \infty$ and for every m, such that $|\phi(x)| \leq b(x)$ for every $x \in \mathbb{R}$ and every $\phi \in \Phi$. Because $g(.,w)$ is bounded by a polynomial, $g(.,w) \times b$ is integrable. Thus, by dominated convergence,

$$\sup_{\phi\in\Phi}\left|\int \phi(u)(g_n(u,w) - g(u,w))du\right| \leq \int b(u)\mathbb{1}_{^c[-n,n]}(u)|g(u)|du \longrightarrow 0.$$

Hence, $g_n(.,w) \to g(.,w)$ in $\mathcal{S}'$.

Let us show similarly that $c_n(.,w) = \int g_n(t,w)f_{-\varepsilon_1|W}(.-t,w)dt$ converges to $c(.,w)$ in $\mathcal{S}'$. Let $D(w)$ and $l(w)$ be such that $|g(t,w)| \leq D(w)(1 + \|t\|^{l(w)})$. We get

$$\begin{aligned}\int |g(t,w)|f_{-\varepsilon_1|W}(u-t,w)dt =& \int |g(u-t,w)|f_{-\varepsilon_1|W}(t,w)dt \\ \leq& D(w)\left(1 + \int \|u-t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt\right) \\ \leq& D(w)\left[1 + 2^{l(w)-1}\left(\|u\|^{l(w)} + \int \|t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt\right)\right],\end{aligned}$$

where the second inequality follows by convexity. Moreover, by assumption A5,

$$\int \|t\|^{l(w)} f_{-\varepsilon_1|W}(t,w)dt < +\infty.$$

This, together with the previous inequality, implies that $(t,u) \mapsto b(u)g(t,w)f_{-\varepsilon_1|W}(u-t,w)$ is integrable. As a consequence, $u \mapsto b(u)(c_n(u,w) - c(u,w))$ is also integrable. Moreover, by dominated convergence,

$$\int b(u)[c_n(u,w) - c(u,w)]du = \int\int b(u)g(t,w)f_{-\varepsilon_1|W}(u-t,w)\mathbb{1}_{^c[-n,n]}(t)dtdu \longrightarrow 0.$$

As previously, this shows that $c_n(.,w) \to c(.,w)$ in $\mathcal{S}'$.

The previous results ensure that we can apply lemma 2.1 of Mattner (1992) to $f_{-\varepsilon_1|W}(.,w)$, $g(.,w)$, $c(.,w)$ and $g_n(.,w)$. As a consequence, we get, for almost all w and everywhere except on $A(w)$,

$$\mathcal{F}(g(.,w)) \times \mathcal{F}(f_{-\varepsilon_1|W}(.,w)) = 0.$$

Thus, by A4, $\mathcal{F}(g(.,w)) = 0$ everywhere except on $A(w)$. Applying the same reasoning as at the end of the proof of theorem 1.3 of Mattner (1992) finally yields $g(t,w) = 0$ almost everywhere in t. Part 2 follows and the proof is complete.

5.2 Proposition 2.2

We keep the notations of the previous proof. Because $\mu_1(t) = t$, $\widetilde{h} = h$. Hence, by (5.1) and (5.2), completeness is equivalent to

$$\int h(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0 \quad \text{a.e. in } u \Rightarrow h(t,w) = 0 \quad \text{a.e. in } t. \tag{5.4}$$

for $\mathbb{P}^W$-almost all w and all h such that $\mathbb{E}\left[|h(X_0)|\right] < \infty$. But theorem 2.4 of Mattner (1993) implies that this condition is not satisfied.[10] Hence, X is not complete for Z.

5.3 Proposition 2.3

We still keep the previous notations. Following the same lines as in the proof of theorem 2.1, we can show that bounded completeness holds if, for $\mathbb{P}^W-$almost all w and all bounded $g(.,w)$,

$$\int g(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0 \text{ for a.e. } u \in \text{Supp}(\nu_1(Z)|W=w) \quad \Rightarrow \quad g(t,w) = 0 \quad \text{a.e. in } t \tag{5.5}$$

[10] Actually, Mattner (1993) shows that $\int h(t,w) f_{-\varepsilon_1|W}(u-t,w)dt = 0$ for every u (and not almost every u) implies $h(t,w) = 0$ for almost every t. However, an inspection of the proofs of his theorem 2.4 and lemma 2.3 shows that "every u" can be replaced by "almost every u" without affecting the result.

Suppose that the left hand side holds. Then, by Assumption A2', $c(.,w)$ equals zero on an open set O_w.

Besides, by assumption A7 and the fact that $g(.,w)$ is bounded for $\mathbb{P}^W$-almost all w, the function

$$(t,u) \mapsto g(t,w) f_{-\varepsilon_1|W}(u-t,w)$$

is bounded on $\{(t,u) \in \mathbb{R} \times \mathbb{C} / |\mathrm{Im}(u)| < r_0(w)\}$. Moreover, $u \mapsto g(t,w) f_{-\varepsilon_1|W}(u-t,w)$ is holomorphic on $\{u \in \mathbb{C}/|\mathrm{Im}(u)| < r_0(w)\}$ by assumption A7. Thus (see Rudin, 1987, p. 229), $c(.,w)$ is holomorphic on this same set. An holomorphic function which vanishes on an open set actually equals zero everywhere (see e.g. Rudin, 1987, p.209). Thus, $c(.,w) = 0$ everywhere and the end of the proof of theorem 2.1, part 1, can be applied.

5.4 Proposition 2.4

Let h be such that $E[|h(X)|] < \infty$ and $\mathbb{E}[h(X)|Z] = 0$ almost surely. Let also $\nu_2^{-1}(.,w)$ denote the inverse of $\nu_2(.,w)$ and $\widetilde{h}(t,w) = h(\nu_2^{-1}(t,w),w)$. Letting $T = \nu_2(X)$, we have

$$\mathbb{E}[\widetilde{h}(T,W)|\mu_2(Z),W] = 0.$$

Hence, for all $t_1 \in \mathbb{R}^{q-r}$, almost surely,

$$\mathbb{E}[\widetilde{h}(T,W)e^{it_1'(T+\varepsilon_2)}|W] = 0.$$

Then, by assumption A8, almost surely,

$$\mathbb{E}[\widetilde{h}(T,W)e^{it_1'T}|W]\mathbb{E}[e^{it_1'\varepsilon_2}|W] = 0.$$

This implies, by assumption A10, that the function $t_1 \mapsto \mathbb{E}[\widetilde{h}(T,W)e^{it_1'T}|W]$ vanishes everywhere except perhaps on isolated points. Now, because $\mathbb{E}[|\widetilde{h}(T,W)|\,|W] < +\infty$, the function $t_1 \mapsto \mathbb{E}[\widetilde{h}(T,W)e^{it_1'T}|W]$ is continuous by dominated convergence. Thus, for all $(t_1,t_2) \in \mathbb{R}^{q-r} \times \mathbb{R}^r$,

$$\mathbb{E}[\widetilde{h}(T,W)e^{i(t_1'T+t_2'W)}] = 0.$$

This implies (see e. g. Bierens, 1982, theorem 1) that $\mathbb{E}[\widetilde{h}(T,W)|T,W] = 0$ almost surely. In other words, $h(X) = 0$ almost surely.

5.5 Proposition 3.1

$\mathbb{E}(Y|Z) = \mathbb{E}(\varphi(X)|Z)$, so that any candidate φ' for φ satisfies

$$\mathbb{E}[(\varphi' - \varphi)(X)|Z] = 0.$$

If $\varphi(., W)$ is known to be bounded, any candidate must be also bounded so that $(\varphi' - \varphi')(., W)$ is bounded. Then by theorem 2.1 (part 1), $\varphi' = \varphi$ so that $\varphi(., W)$ is identified. If $\varphi(., W)$ is bounded by a polynomial, so is $(\varphi' - \varphi')(., W)$, and the same conclusion holds by part 2 of the theorem. Lastly, if A6 holds, $\varphi' = \varphi$ by part 3 of the theorem.

References

Bierens, H. J. (1982), 'Consistent model specification tests', *Journal of Econometrics* **20**, 105–134.

Blundell, R., Chen, X. and Kristensen, D. (2007), 'Nonparametric IV estimation of shape-invariant engel curves', *Econometrica* **75**, 1613–1669.

Chen, X. and Hu, Y. (2006), Identification and inference of nonlinear models using two samples with arbitrary measurement errors. Cowles foundation discussion paper no. 1590.

Chernozhukov, V. and Hansen, C. (2005), 'An IV model of quantile treatment effects', *Econometrica* **73**, 245–261.

Chesher, A. (2003), 'Identification in nonseparable models', *Econometrica* **71**, 1405–1441.

Chesher, A. (2007), Identification of non-additive structural functions, *in* W. K. Newey, R. Blundell and T. Persson, eds, 'Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress', Vol. 3, New York: Cambridge University Press.

Darolles, S., Florens, J. P. and Renault, E. (2007), Nonparametric instrumental regression. Working paper.

Das, M. (2005), 'Instrumental variables estimators of nonparametric models with discrete endogenous regressors', *Journal of Econometrics* **124**, 335–361.

Devroye, L. (1989), 'Consistent deconvolution in density estimation', *Canadian Journal of Statistics* **17**, 235–239.

Fan, J. and Truong, Y. K. (1993), 'Nonparametric regression with errors in variables', *The Annals of Statistics* **21**, 1900–1925.

Florens, J., Heckman, J. J., Meghir, C. and Vytlacil, E. (2008), 'Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects', *Econometrica, forthcoming* .

Florens, J. P., Heckman, J. J., Meghir, C. and Vytlacil, E. (2003), Instrumental variables, local instrumental variables and control functions. IDEI Working Paper 249.

Florens, J. P., Mouchart, M. and Rolin, J. M. (1990), *Elements of Bayesian statistics*, Marcel Dekker, New York - Basel.

Ghosh, J. K. and Singh, R. (1966), 'Unbiased estimation of location and scale parameters', *The Annals of Mathematical Statistics* **37**, 1671–1675.

Hoeffding, W. (1977), 'Some incomplete and boundedly complete families of distributions', *The Annals of Statistics* **5**, 278–291.

Hu, Y. and Schennach, S. (2008), 'Identification and estimation of nonclassical errors-in-variable models with continuous distributions using instruments', *Econometrica* **76**, 195–216.

Imbens, G. W. and Newey, W. K. (2006), Identification and estimation of triangular simultaneous equations models without additivity. Working Paper.

Lehmann, E. L. (1986), *Testing Statistical Hypothesis*, 2nd ed. Wiley: New-York.

Lehmann, E. L. and Scheffé, H. (1947), 'On the problem of similar region', *Proceedings of the National Academy of Science* **33**, 382–386.

Lewbel, A. (2000), 'Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables', *Journal of Econometrics* **97**, 145–177.

Li, T. and Vuong, Q. (1998), 'Nonparametric estimation of the measurement error model using multiple indicators', *Journal of Multivariate Analysis* **65**, 139–165.

Manski, C. (1988), 'Identification of binary response models', *Journal of the American Statistical Association* **83**, 729–738.

Mattner, L. (1992), 'Completeness of location families, translated moments, and uniqueness of charges', *Probability Theory and Related Fields* **92**, 137–149.

Mattner, L. (1993), 'Some incomplete but boundedly complete location families', *The Annals of Statistics* **21**, 2158–2162.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Newey, W. K., Powell, J. L. and Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

Rudin, W. (1987), *Real and Complex Analysis, third edition*, McGraw-Hill.

Rudin, W. (1991), *Functional Analysis, second edition*, McGraw-Hill.

Schennach, S. (2004), 'Estimation of nonlinear models with measurement error', *Econometrica* **72**, 33–75.

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

Schwartz, L. (1973), *Théorie des distributions, deuxième édition*, Hemann.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Zinde-Walsh, V. (2007), Errors-in-variables models: a generalized functions approach. CIREQ - Cahier 14-2007.