

# On the Existence of $\sqrt{N}$ -consistent Estimators of Linear Functionals in Non-Parametric IV Models\*

Laurent Davezies<sup>†</sup>

September 9, 2016

## Abstract

This paper considers the issue of the existence of  $\sqrt{N}$ -consistent estimators of linear functionals of the form  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$  in a nonparametric IV model where  $\mathbb{E}(Y - m(X)|Z) = 0$ . We derive a necessary condition for the existence of  $\sqrt{N}$ -consistent estimators and we derive the semi-parametric efficiency bound under this condition. This condition ensures that  $\theta_0$  can be expressed as a moment of the form  $\mathbb{E}(q(Z)Y)$  and is not nested with the completeness condition. We then explore the implications of this necessary condition on the first stage i.e. the distribution of  $X|Z$ .

When the distribution of  $X$  conditional on  $Z$  belongs to an exponential family, or when the first stage has an additive structure, the condition that we exhibit turns out to be quite restrictive for estimating averages under counterfactual distributions. In that case, some parameters are never (respectively always)  $\sqrt{N}$  estimable. But more surprisingly in some cases, some parameters are  $\sqrt{N}$  estimable if and only if some central cross-moments of  $X$  and  $Z$  are above a critical value.

On the other hand, an additive first stage is a condition that enables the econometrician to identify average marginal effects, and ensures the finiteness of the semi-parametric efficiency bound for a large class of DGP, even if the completeness condition does not hold.

**Keywords:** semi-parametric model, efficiency bound, instrumental variables, counterfactual, average marginal effect.

**JEL classification numbers:** C01, C02 and C14.

---

\*I thank my colleagues Xavier d'Haultfoeuille and Anna Simoni for many helpful discussions. I also thank Xiaohong Chen and Arthur Lewbel for their helpful comments and encouragements. This work has also benefited from comments of participants of the 2015 Econometric Society World Congress, of the 2015 Annual Conference of the International Association for Applied Econometrics and of the CREST seminars. Last, I am indebted to Benjamin Walter for his careful reading. The usual disclaimer applies and all errors remain ours.

<sup>†</sup>CREST, laurent.davezies@ensae.fr.

# 1 Introduction

Models with endogeneity are central in econometrics. In a nonparametric setting with additively separable errors, the completeness condition (Newey & Powell, 2003) is necessary and sufficient to identify the structural function. But this condition is quite restrictive and cannot be tested without supplementary restrictions (Canay et al., 2013). However, the parameter of interest is often an expectation depending on the structural function (or on its derivatives) and not the full structural function itself. For instance, the parameter of interest could be an average marginal effect, an average effect under an exogenous change in the distribution of regressors, a linear functional considered by Severini & Tripathi (2012*b*) and Santos (2011), or a density weighted average derivative (Powell et al., 1989, Cattaneo et al., 2014).

Moreover, we can hope there exists some regular root- $N$  consistent estimators for such parameters. The motivation to focus on the existence of root- $N$  regular estimators is related to some optimality results. Indeed, in sufficiently regular models when regular estimators exist, a convolution Theorem and a local asymptotic minimax Theorem hold for any subconvex loss function (cf. van der Vaart, 2000, Theorems 25.20 and 25.21). This means that the corresponding asymptotic risk of any regular estimator would be bounded below. For the asymptotic mean square error, this lower bound is usually called the semi-parametric efficiency bound (see for instance Newey, 1990*b*). The efficient estimator that achieves this bound leads to efficient tests (Choi et al., 1996). This notion generalizes the Cramer-Rao bound originally introduced for parametric models. Not surprisingly, this notion is closely related to the Fisher information. In an iid data generating process, it measures the minimal number of observations needed to discriminate the true value of the parameter of interest from an alternative one, whatever the value of nuisance parameters. Both statisticians (Bickel et al., 1993) and econometricians (Chamberlain, 1987, Newey, 1988, Newey, 1990*a*, Severini & Tripathi, 2012*a*) have derived those bounds for various models. However, existence of such regular estimator is not always ensured (van der Vaart, 2000, Theorem 25.32 and van der Vaart, 1991) . To our best knowledge, for the case of linear functionals in a non-parametric instrumental variables model, the first paper to bring this issue to light is Ai & Chen (2007). Non-parametric instrumental variables have only been studied by econometricians (not by statisticians) and a large strand of the literature focuses on the non-parametric estimation of the regression function  $m$  (for instance, Ai & Chen, 2003, Florens, 2003, Newey & Powell, 2003, Hall & Horowitz, 2005, Chen, 2007, Florens et al., 2011, Florens & Simoni, 2012, Florens et al., 2012, Darolles et al., 2011, Chen & Pouzo, 2012). But linear functionals of the regression function, which are often the parameter of interest for the economists, have received attention more recently (Ai &

Chen, 2007, Santos, 2011, Severini & Tripathi, 2012*b*, Ai & Chen, 2012, Carrasco et al., 2014, Chen & Pouzo, 2015, Breunig & Johannes, 2016). Ai & Chen (2007), Carrasco et al. (2014), Chen & Pouzo (2015) and Breunig & Johannes (2016) propose plug-in estimators based on a non-parametric first step under the completeness condition, but the root-N consistency of such estimators is far to be ensured. For instance for the plug-in penalized sieve minimum distance estimator, the detailed discussion in Section 3 of Chen & Pouzo (2015) illustrates this negative result (see also the discussion of Assumption 4.1 in Ai & Chen (2007)). Here, we do not establish the properties of a specific estimator, but examine whether a root-N and regular consistent estimator could exist.

The next section is brief and only introduces notations and the model. If  $Y = m(X) + \varepsilon$  with  $\mathbb{E}(\varepsilon|Z) = 0$  and  $F_{Y,X,Z}$  identified by the data, we show in the third Section that parameters of the form  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$  are identifiable under weaker conditions than the completeness condition in Section 3. We generalize a result of Severini & Tripathi (2012*b*) to a larger class of parameters including for instance the average marginal effect. In the fourth Section of this paper we give a necessary and sufficient condition for the existence of a finite semi-parametric efficiency bound for such parameters. This condition is *necessary* for the existence of a regular root-N estimator. This condition states that it exists  $q(Z)$  such that  $\theta_0 = \mathbb{E}(q(Z)Y)$ . Moreover, when such a condition holds, the expression of the semi-parametric efficiency bound suggests an efficient estimator based on an inverse problem because we also need to have  $\mathbb{E}(q(Z)|X)f_X(X) = (-1)^{|j|}[\psi(X)f_X(X)]^{(j)}$ . A straightforward implication of our main result concerns the exogenous case: if there is no root-N estimator of  $\theta_0$  based only on the outcome and on the covariates, supplementary instruments won't help to reach the root-N consistency. Again, results obtained in this section generalize the previous results of Severini & Tripathi (2012*b*). Moreover, our expression of the semi-parametric efficiency bound is more explicit and shows that the estimator of Santos (2011) is not efficient. Our framework also generalizes those of Powell et al. (1989) or Newey & Stoker (1993). Ai & Chen (2012) propose an alternative expression of the semi-parametric efficiency bound for an arbitrary sequence of identifying conditional moments restrictions, but assume the existence of a solution of to a functional minimization<sup>1</sup>. The existence of finite semi-parametric efficiency bound (and next of regular root-N estimators) is related to the existence of a Riesz representation of the Hadamard derivative of  $\theta$  with respect to the square root of measure of any parametric submodel (cf. Sections 25.3 to 25.5 of van der Vaart (2000)). We emphasize the *necessary* condition under which such a linear form exists, and we derive a simple expression for the efficiency bound in which this condition appears explicitly.

A first part of our condition leads to a *necessary* condition of regularity at the boundary

---

<sup>1</sup>cf. Equations (5) and (6) in Ai & Chen (2012)

of the support of the regressors  $X$ . In the exogenous case (ie when  $X = Z$ ), regularity at the boundary of the support has already been shown to be a *sufficient* condition in many papers such as Powell et al. (1989), and we show that we cannot relax it. A second part of our condition is more restrictive, especially in the endogenous case (ie when  $X \neq Z$ ). The *sufficient* conditions <sup>2</sup> given in Ai & Chen (2007), Carrasco et al. (2007), Carrasco et al. (2014), Breunig & Johannes (2016) and Ai & Chen (2012) for the existence of a root-N consistent estimator and/or for existence of a finite semi-parametric efficiency bound are compatible with the *necessary* ones that we establish in this paper. In Section 4.3, we also show that our expression of the semiparametric efficiency bound extends the formula given in Ai & Chen (2012) to a larger class of DGP (reflecting the discrepancy between their *sufficient* conditions and our *necessary* ones). However, the second part of our necessary condition is also quite abstract and in the fifth Section, we discuss it on two special cases: the exponential first stage and the additive first stage. We also give some examples to show that naive intuition could be misleading.

To sum up, the main implications of our results are the following: 1. only "very regular" (in a sense detailed in this paper) linear functionals are root-N estimable, and in this case  $\theta_0$  could be reformulated as a moment of  $(Y, Z)$  but the completeness condition is not necessary 2. regularity conditions that we get are stronger in the endogenous case than in the exogenous case and supplementary instruments will not help reach the root-N rate when the regressors are exogenous, 3. even for some simple data generating processes usually considered in the textbooks, the mean of  $Y$  under a simple counterfactual distribution of  $X$  could be or could not be root-N estimable, depending on the counterfactual distribution chosen, 4. under an additive first stage, the average marginal effect admits a finite semi-parametric efficiency bound under easily interpretable and reasonable conditions.

## 2 DGP and parameter

Let consider the following instrumental regression:

$$Y = m(X) + \varepsilon \quad \text{and} \quad \mathbb{E}(\varepsilon|Z) = 0$$

We assume there exists a dominant measure of  $(Y, X, Z)$  and  $\mathcal{X} := \text{Supp}(X) \subset \mathbb{R}^K$ ,  $\mathcal{Z} := \text{Supp}(Z) \subset \mathbb{R}^l$ ,  $\mathcal{Y} := \text{Supp}(Y) \subset \mathbb{R}$ . The function  $m$  is a function from  $\mathbb{R}^K$  to

---

<sup>2</sup>cf. the source conditions in Carrasco et al. (2007), the Assumption 4.1 in Ai & Chen (2007), the Assumption 3.3 and 3.4 in Carrasco et al. (2014), Assumptions 1 and 2 and Remark 2.4 in Breunig & Johannes (2016), and the existence of a solution of Equation (5) in Ai & Chen (2012)

$\mathbb{R}$ . For a  $K$ -dimensional multi-index  $(j) = (j_1, \dots, j_K)$ ,  $f^{(j)}$  denotes the partial derivative  $\partial^{j_1}, \dots, \partial^{j_K} f / \partial x_1^{j_1}, \dots, \partial x_K^{j_K}$  (which is assumed to exist and to not depend on the order in which the differentiation is performed). Let  $D^{(j)}(X)$  the set of functions  $f$  from  $\mathbb{R}^K$  to  $\mathbb{R}$  such that with probability 1,  $f^{(j)}(X)$  exists. We assume that  $m \in D^{(j)}(X)$  and we are interested in estimation of the linear functional  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$  with  $\psi$  a function from  $\mathbb{R}^K$  to  $\mathbb{R}$  of the form  $\psi(x) = h(f_X(x), x)$  with  $h$  a known function differentiable with respect to the first variable and  $f_X$  the identifiable density of  $X$  with respect to a dominant measure. The average effect under a counterfactual distribution  $\widetilde{f}_X$  of regressors corresponds to  $\theta_0 = \mathbb{E}\left(\frac{\widetilde{f}_X(X)}{f_X(X)}m(X)\right)$ . The average marginal effect of  $X_k$  corresponds to  $\mathbb{E}(\partial_{x_k}m(X))$ .

To avoid functional dependence between the components of  $X$ , we assume that the product of the dominant measure for the distributions of the components of  $X$  is a dominant measure. Moreover, because differentiating with respect to the component  $x_k$  only have sense for continuous variables, for  $k$  such that  $j_k > 0$ , the dominant measure of  $X_k$  is assumed to be the Lebesgue measure.

We can now state a first minimal assumption, under which the model and the parameter considered have a meaning. In the sequel of the paper, for any random variable  $U$ ,  $L^1(U)$  denotes the set of real valued and measurable functions  $f$  on  $\text{Supp}(U)$  such that  $\mathbb{E}(|f(U)|) < \infty$ .

**Assumption 1 (Exclusion restriction and well-defined parameter)**

$$Y = m(X) + \varepsilon \text{ and } \theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$$

with

- (i)  $Y \in L^1(Y)$ ,  $m(X) \in L^1(X) \cap D^{(j)}(X)$  and  $\mathbb{E}(\varepsilon|Z) = 0$
- (ii)  $m \in \mathcal{M} \subset \{l \in L^1(X) \cap D^{(j)}(X) \text{ and } \psi l^{(j)} \in L^1(X)\}$
- (iii)  $F_{Y,X,Z}$  is identified.

The set of functions  $\mathcal{M}$  in the second condition of the previous assumption reflects the restrictions imposed by the model. For instance, Chen (2007) consider Hölder, Sobolev or Besov space and Ai & Chen (2003), Blundell et al. (2007), Florens et al. (2012) impose functional restrictions like some form of additive separability or partial linearity on  $m$ . In many contexts, economic theory also impose shape restrictions such as monotonicity and/or concavity of  $m$ . Such restrictions can be embedded in the definition of  $\mathcal{M}$ . Note that  $\mathcal{M}$  is not necessarily a linear space especially in the case of shape restrictions.

Our framework generalizes those of Severini & Tripathi (2012b) and Santos (2011) who consider the case where  $(j) = (0)$ . Such a class of parameters includes the average of  $Y$

under counterfactual distribution of  $X$ . Indeed, consider  $\psi(x) = \tilde{f}_X(x)/f_X(x)$  with  $\tilde{f}_X$  a density function supported by  $\mathcal{X}$ . In such a case  $\theta_0$  represents the average of  $Y$  under the assumption that the distribution of unobserved  $\varepsilon$  is unchanged whereas the distribution of  $X$  is changed into  $\tilde{f}_X$ .

Our framework is also a generalization of the weighted average derivative studied among others by Powell et al. (1989) and Newey & Stoker (1993). Such cases correspond to  $X = Z$ ,  $(j) = (0, \dots, 1, \dots, 0)$ . Powell et al. (1989) consider the case where  $\psi(x) = f_X(x)$ , whereas Newey & Stoker (1993) consider any differentiable  $\psi$  and include the case of average marginal effect  $\theta_0 = \mathbb{E}(\partial_{x_k} m(X))$ .

Whereas Severini & Tripathi (2012b) emphasize the difficulty resulting from the endogeneity of  $X$ , they do not allow  $\theta_0$  to be a linear functional of the derivative of  $m$ . On the other hand, Powell et al. (1989) and Newey & Stoker (1993) focus on the average of derivative of  $m$  but under an exogeneity assumption of  $X$ .

We now turn to the identification of  $\theta_0$  under Assumption 1.

### 3 Identification

Before we present our result on identification, let us introduce some notation that will be used in the rest of the paper. For any nonnegative number and for any random variable  $U$ ,  $L^p(U)$  denotes the set of measurable functions  $f$  defined on  $\text{Supp}(U)$  such that  $\mathbb{E}(|f^p(U)|) < \infty$ . If  $p = 0$ ,  $L^0(U)$  is simply the set of measurable functions defined on the support of  $U$ .  $L^\infty(U)$  is the set of bounded and measurable functions of  $U$ . For  $p \geq 1$ ,  $L_0^p(U)$  denotes the space  $\{f \in L^p(U) \text{ s.t. } \mathbb{E}(f(U)) = 0\}$  and  $L_0^p(U|V)$  the space  $\{f \in L^p(U, V) \text{ s.t. } \mathbb{E}(f(U, V)|V) = 0\}$ .

Let  $A \subset L^0(X)$ ,  $A^\perp$  denotes:

$$\{h \in L^0(X) \mid \forall g \in A, \quad hg \in L_0^1(X)\}$$

Because the model relies on the nullity of the conditional expectation of  $\varepsilon$ , the condition of identification depends on the following operator:

$$\begin{aligned} T : L^1(X) &\longrightarrow L^1(Z) \\ x \mapsto g(x) &\longmapsto z \mapsto \mathbb{E}(g(X)|Z = z) \end{aligned}$$

$\text{Ker}(T)$  denotes the set of functions  $l \in L^1(X)$  such that  $T(l) = 0$ .

#### Theorem 3.1 (Identification)

(i) *The identification region of  $\theta_0$  is  $\{\theta_0 + \mathbb{E}(\psi l^{(j)}) : l \in (\mathcal{M} - \{m\}) \cap \text{Ker}(T)\}$ . So if  $\mathcal{M}$  is*

a linear space, the identification region of  $\theta_0$  is either  $\{\theta_0\}$  (in such case  $\theta_0$  is identified), either  $\mathbb{R}$  (in such case, data are uninformative about  $\theta_0$ ).

(ii)  $\theta_0$  is identified if and only if  $\psi \in \{l^{(j)} \text{ s.t. } l \in (\mathcal{M} - \{m\}) \cap \text{Ker}(T)\}^\perp$ .

When the completeness condition holds (i.e. when  $\text{Ker}(T) = \{0\}$ ) then  $m$  is identified and and so is the case for  $\theta_0$  (Newey & Powell (2003)). Moreover, Canay et al. (2013) have shown that completeness is not testable without supplementary restrictions. So achieving identification of the parameter of interest under weaker condition is attractive. The previous result generalizes those of Severini & Tripathi (2012b): if  $j = 0$  and  $\psi \in L^0(X)$ ,  $\theta_0$  is identified as soon as for any  $l \in \text{Ker}(T)$  such that  $\mathbb{E}(|\psi(X)l(X)|) < \infty$ , one has  $\mathbb{E}(\psi(X)l(X)) = 0$ . The following examples illustrate the fact that without completeness some parameters of interest could be identified.

**Example: A simple scale model on the first stage.** Let  $X = Z \cdot \eta$ , with  $Z \perp\!\!\!\perp \eta$ ,  $\eta \sim \mathcal{U}_{[-1;1]}$ ,  $\mathcal{Z}$  an interval of  $\mathbb{R}^+$  and  $\mathbb{E}(|Z|) < \infty$ . In that case,  $l \in \text{Ker}(T)$  if and only if for any  $z \in \mathcal{Z}$ ,  $\int_{-1}^1 l(zu)du = \int_{-z}^z l(u)du/z = 0$ . Derivation with respect to  $z$  gives  $(l(z) + l(-z))/z - \int_{-z}^z l(u)du/z^2 = 0$  and then implies that  $l$  is an odd function. Reciprocally, if  $l$  is an odd function, then it belongs to  $\text{Ker}(T)$ . From Theorem 3.1.(ii), it follows that the average effect under the counterfactual distribution  $\tilde{f}_X$  (i.e.  $\theta_0 = \mathbb{E}\left(\frac{\tilde{f}_X(X)}{f_X(X)}m(X)\right)$ ) is identifiable if and only if  $\tilde{f}_X$  is even. On the other hand, the average marginal effect is not identifiable because  $\psi(x) = 1$  is even whereas  $\{l' \text{ s.t. } l \in \text{Ker}(T)\}^\perp$  is the set of odd functions on  $[-1; 1]$ .

**Example: Average marginal effect under an additive first stage.** We will see in the last section of this paper, that an additive first stage implies regularity of the DGP that ensures the finiteness of the semi-parametric efficiency bound for the average marginal effect under mild conditions. Here, we discuss the identification of such parameter.

**Proposition 3.2 (Identification of Average marginal effect)**

If  $Y = m(X) + \varepsilon$  with  $Y, \varepsilon$  and  $X$  real random variables such that Assumption 1 holds, and assume that  $X = p(Z) + \eta$  with  $Z \perp\!\!\!\perp \eta$ . Then  $\theta_0 = \mathbb{E}(m'(X))$  is identified if:

$$\mathcal{M} \subset \{l \in L^1(X), l \in D^1(X), \exists P \text{ s.t. } \mathbb{P}(p(z) \in P) = 1 \text{ and } \forall p \in P, \partial_p \mathbb{E}(l(p+\eta)) = \mathbb{E}(l'(p+\eta))\}$$

The argument of identification is relies on the fact that derivative with respect to  $p \in \text{Supp}(p(Z))$  and integration with respect to  $F_\eta$  commute. It holds if for instance,  $\mathcal{M} = \{l \in L^1(X), l \in D^1(X) \text{ and } |l'(x)| \leq \sum_{s=0}^S a_s x^s\}$ , and  $\mathbb{E}(|\eta|^S) < \infty$ . The derivative of  $m$  is assumed to be bounded by a polynomial of fixed degree  $S$  and for this given  $S$ , the corresponding moment of  $\eta$  are well defined. If the derivative of  $m$  is assumed to be

bounded by a constant, this condition always holds. It is not difficult to find some DGP where such condition holds but where completeness condition is not fulfilled. Consider for instance that  $\mathbb{P}(p(z) \in \text{int Supp}(p(Z))) = 1$ ,  $\eta = \nu + u$  with  $\nu$  and  $u$  two independent variables with  $u$  uniform on an interval  $I$  not reduced to a single point. Then  $\mathbb{E}(m'(X))$  is identified whereas neither the completeness nor the bounded completeness condition hold (d'Haultfoeuille, 2011). The result holds if  $\nu = 0$  almost-surely and whatever the distribution of  $Z$ . Proof of the Proposition 3.2 is given in the Appendix, but we will present here an informal sketch of it. Under Condition (i) or (ii) of the Proposition 3.2, for any  $p \in \text{int}(\text{Supp}(p(Z)))$  we have the following equality:

$$\frac{\partial}{\partial p} \mathbb{E}(l(X)|p(Z) = p) = \mathbb{E}(l'(X)|p(Z) = p).$$

For  $l \in \text{Ker}(T)$  the left hand side is null and so is the right hand side. Theorem 3.1.(ii) ensures the identification of  $\theta_0$ . Moreover, the strategy of identification suggests  $\frac{1}{n} \sum_{i=1}^n \hat{g}(Z_i)$  with  $\hat{g}(z)$  an estimator of  $\frac{\partial}{\partial p} \mathbb{E}(Y|p(Z) = p)$  because  $\frac{\partial}{\partial p} \mathbb{E}(Y|p(Z) = p) = \mathbb{E}(m'(X)|p(Z) = p)$ . When  $p(\cdot)$  admits non vanishing derivative with respect to the component  $z_l$ , the identification argument could be restated as a kind of Wald-ratio:  $\mathbb{E}(m'(X)) = \mathbb{E}(\mathbb{E}(m'(X)|Z = z)) = \mathbb{E} \left[ \mathbb{E} \left( \frac{\partial_{z_l} E(Y|Z=z)}{\partial_{z_l} p(z)} | Z = z \right) \right]$ . Moreover when more than one instrument  $z_l$  satisfies  $\partial_{z_k} p(z) \neq 0$ , the average marginal effect is overidentified (cf. Carrasco et al. (2007) for a definition of overidentification in non-parametric or semi-parametric linear inverse problem).

In both the two previous examples, we see that completeness condition is a sufficient but not a necessary condition for identification of  $\theta_0$ . Under the conditions of identification given in Theorem 3.1, we will now examine the issue of the existence of a root-N consistent estimator of  $\theta_0$ .

## 4 Existence of root-N regular estimator

In our context of non-parametric regression and under our identification condition of  $\theta_0$ , it is natural to ask whether the parameter of interest admits a finite semi-parametric efficiency bound. Heuristically, we know that the optimal rate of convergence of estimators of  $m^{(j)}$  is always lower than  $\sqrt{N}$  and in some cases far lower. But because  $\theta$  is an "average" of  $m^{(j)}$ , the optimal rate of convergence of estimators of  $\theta$  could be higher and sometimes reaches the  $\sqrt{N}$ -rate. In that case, there is a room to build an efficient estimator with respect to any subconvex loss function (van der Vaart (2000), Theorem 25.21), uniformly most powerful tests and asymptotically uniformly most accurate confidence sets (see Choi et al. (1996)



for a discussion about asymptotically uniformly most powerful tests in semiparametric models).

The case  $j = 0$ ,  $\mathcal{M} = L^2(X)$  and  $\psi \in L^2(X)$  has been studied by Severini & Tripathi (2012b). If there exists  $q \in L^1(Z)$  such that  $\psi(X) = \mathbb{E}(q(Z)|X)$ , then for any  $l \in L^2(X) \cap \text{Ker}(T)$ ,  $\mathbb{E}(\psi(X)l(X)) = \mathbb{E}(q(Z)l(X)) = 0$  and the condition for identification of  $\theta_0$  holds. Moreover in this case, the parameter can be rewritten as  $\theta_0 = \mathbb{E}(Yq(Z))$  and estimated by the average of the  $Y_i\widehat{q}(Z_i)$ , with  $\widehat{q}$  a consistent estimator of  $q$ . Intuitively, if  $\widehat{q}$  is sufficiently "regular" the estimator will behave as a sample mean, i.e. the corresponding estimator of  $\theta_0$  will be regular and root-N consistent. Severini & Tripathi (2012b) have shown that in such a case, the existence of such a  $q$  is a necessary condition for the existence of a regular root-N estimator of  $\theta_0$ .

In this section we derive the same kind of necessary condition in a more general framework to ensure existence of regular  $\sqrt{N}$  consistent estimators of  $\theta_0$  and we derive an expression of the semi-parametric efficiency bound of  $\theta_0$ . For the sake of simplicity, we will now assume that  $Y$ ,  $m(X)$ ,  $m^{(j)}(X)$  and  $\psi(X)$  all admit finite variances. And consequently we now work in Hilbert spaces  $L^2(Y, X, Z)$ . Then  $T$  is now considered as an operator from  $L^2(X)$  into  $L^2(Z)$  and  $T^*$  is his adjoint operator.

$$\begin{aligned} T^* : L^2(Z) &\longrightarrow L^2(X) \\ z \mapsto g(z) &\longmapsto x \mapsto \mathbb{E}(g(z)|X = x) \end{aligned}$$

#### 4.1 Regularity conditions

We will assume some supplementary regularity conditions. Let  $X^+$  (respectively  $X^0$ ) denote the subset of the variables  $X_k$  such that  $j_k > 0$  (respectively  $j_k = 0$ ) and  $K^+$  (respectively  $K^0$ ) the number of such variables. We assume without loss of generality that the variables  $X$  are ordered such that  $X = (X^+, X^0)$ .

Remember that  $\psi(x) = h(f_X(x), x)$ . Because  $f_X$  is unknown but identifiable in the data, the existence of root-N consistent estimator is related to regularity conditions of  $u \mapsto h(u, x)$ .

Intuitively, the largest  $\mathcal{M}$  is, the strongest are the conditions to ensure existence of a root-N consistent estimator. If  $\mathcal{M}$  is a finite dimensional space (for instance, if  $m$  is assumed to be linear) and if the identification condition for  $\theta_0$  holds, GMM estimators are root-N consistent (as soon as  $Y$  and  $m(X)$  admit finite variances). To avoid such degenerated cases, we have to impose that  $\mathcal{M}$  is sufficiently large and a priori the largest is  $\mathcal{M}$  the strongest are the necessary condition for finiteness of the semiparametric efficiency bound. However, we show that the necessary and sufficient condition holds as soon as  $\mathcal{M}$  contains

$m + t\varphi$  for any  $\varphi \in \Phi$  and  $t$  sufficiently close to zero, where  $\Phi$  is set of test functions of Schwartz (Schwartz, 1997):

$$\Phi = \left\{ \varphi(x) = \prod_{k=1}^K \varphi_k(x_k) \text{ with } \varphi_k \text{ indefinitely differentiable with compact support} \right\}.$$

The statistical model is parameterized by  $(m, f_{\varepsilon, X, Z})$ . Consider a subset of parameter  $(m_t, f_{\varepsilon, X, Z, t})$  indexed by  $t \in T$  a neighborhood of 0 such that  $(m, f_{\varepsilon, X, Z}) = (m_0, f_{\varepsilon, X, Z, 0})$ . This subset of parameter generates a submodel indexed by  $t$ :  $f_{Y, X, Z, t}(y, x, z) = f_{\varepsilon, X, Z, t}(y - m_t(x), x, z)$ . This submodel is differentiable in quadratic mean if there exists a score  $s \in L^2(Y, X, Z)$  such that:

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} \left( f_{Y, X, Z, t}^{1/2} - f_{Y, X, Z}^{1/2} \right) - \frac{1}{2} s f_{Y, X, Z}^{1/2} \right)^2 d\nu(y, x, z) = 0.$$

The differentiability in quadratic mean is a regularity condition that ensures the efficiency of the maximum likelihood in parametric framework (van der Vaart, 2000, Chapters 7 and 8), this regularity condition is allows to extend the notion of optimality in semi-parametric framework (see for instance Bickel et al., 1993, van der Vaart, 2000 Chapter 25 or Severini & Tripathi, 2012a Section 2.3). A set of score functions  $s$  is a tangent set. The semi-parametric efficiency bound of  $\theta_0$  is always defined relative to a given tangent set. And this bound is only well defined for a parameter and a tangent set such that  $t \mapsto \theta(t) = t^{-1} \left( \int \psi(x) m_t^{(j)}(x) f_{\varepsilon, X, Z, t}(e, x, z) d\nu - \theta_0 \right)$  tends to a continuous linear form on the given tangent set when  $t \rightarrow 0$ . Without any regularity condition, the largest tangent set with respect to which  $\theta(t)$  is differentiable is not easily characterizable. But under the following regularity conditions,  $\theta(t)$  is differentiable with respect to the maximal tangent set, i.e. the set of all square integrable function with respect to  $f_{Y, X, Z}$ .

**Assumption 2 (Regularity Assumption)**

- (i)  $Y \in L^2(Y)$ ,  $\psi(X) \in L^2(X)$  and  $\mathbb{E}(\varepsilon|Z) = 0$ ,
- (ii)  $m \in \mathcal{M} \subset \{l \in L^2(X) \cap D^{(j)}(X), l^{(j)} \in L^2(X) \text{ and } \psi l^{(j)} \in L^2(X)\}$ ,
- (iii) The statistical model is dominated by  $\nu(y, x, z) = \lambda(y) \otimes \mu(x, z)$  with  $\lambda$  the Lebesgue measure
- (iv)  $h$  admits a derivative with respect to the first component  $X$ -almost-surely and  $\partial_{f_X} \psi(X) := \frac{\partial h(u, x)}{\partial u} \Big|_{u=f_X(X)}$  is in  $L^2(X)$ ,
- (v) There exists non negative  $\underline{v}$  and  $\bar{v}$  bounding the skedastik function:

$$\mathbb{E} \left( (Y - m(X))^2 | Z \right) \in [\underline{v}; \bar{v}],$$

(vi) When  $X^+ \neq \emptyset$ ,  $X^+|X^0$  admits a density  $f_{X^+|X^0}$  with respect to the Lebesgue measure on  $\mathbb{R}^{K^+}$ ,  $X^0$ -almost-surely,

(vii) For  $t$  sufficiently close to 0,  $m-t \in \mathcal{M}$  and  $e \mapsto f_{\varepsilon,Z,X}^{1/2}(e, Z, X)$  is absolutely continuous ( $Z, X$ )-almost-surely,  $\frac{\partial_\varepsilon f_{\varepsilon,X,Z}(\varepsilon, X, Z)}{f_{\varepsilon,X,Z}(\varepsilon, X, Z)} \in L^2(\varepsilon, X, Z)$  (i.e. the Fisher information for location of  $\varepsilon$  is finite) and  $\varepsilon \frac{\partial_\varepsilon f_{\varepsilon,X,Z}(\varepsilon, X, Z)}{f_{\varepsilon,X,Z}(\varepsilon, X, Z)} \in L^2(\varepsilon, X, Z)$ ,

(viii)  $\forall \varphi \in \Phi$ ,  $m + t\varphi \in \mathcal{M}$  for  $t$  sufficiently close to 0,

(ix) For  $t$  sufficiently close to 0,  $(1+t)m \in \mathcal{M}$  and  $m(X) \frac{\partial_\varepsilon f_{\varepsilon,X,Z}(\varepsilon, X, Z)}{f_{\varepsilon,X,Z}(\varepsilon, X, Z)} \in L^2(\varepsilon, X, Z)$ .

The first and the second conditions (i) and (ii) in Assumption 2 reinforce the conditions of Assumption 1, assuming that all considered variables admit finite variances. In particular,  $\psi(X)m^{(j)}(X) \in L^2(X)$  ensures that  $\theta(t)$  admits a derivative at  $t = 0$  which is a continuous linear form on the maximal tangent set of scores. Assumption 2.(iii) is very common and natural in such context. Assumption 2.(iv) is made to get a tractable formula for the semi-parametric efficiency bound of  $\theta_0$  when  $\psi(X)$  depends on  $f_X(X)$ . We do not have in mind some interesting parameters of the form  $\mathbb{E}(h(f_X(X), X)m^{(j)}(X))$  with  $h$  not differentiable with respect to the first component. For instance,  $h(u, x) = u$  for the density-weighted average or  $h(u, x) = \tilde{f}(x)/u$  for the average under the counterfactual distribution  $\tilde{f}$  are both differentiable with respect to  $u$ . Assumption 2.(v) ensures the equivalence between  $q(Z) \in L^2(Z)$  and  $q(Z)\varepsilon \in L^2(Z, \varepsilon)$ , which considerably simplifies the proof. Such an assumption also ensures that every score (corresponding to a parametric sub-model) can be approximated by a bounded score (of an other parametric sub-model). Indeed, in a step of our proof we establish some results for parametric sub-models with bounded score, and in another step we extend this result to parametric sub-models with unbounded score. Even if such condition may not be minimal to establish our result, we follow Severini & Tripathi (2012b) who also assume such quite intuitive and simple condition. Concerning Assumption 2.(vi), remember that we have previously assumed that each component of  $X^+$  is dominated by the Lebesgue measure on  $\mathbb{R}$ . Assumption 2.(vi) ensures moreover that no component of  $X^+$  is functionally determined by other components of  $X$ . This is a very usual assumption in econometrics to ensure that data allows to identify variation of one component of  $X$  keeping the others components constant. In Assumption 2.(vii), absolute continuity of  $e \mapsto f_{\varepsilon,Z,X}^{1/2}(e, Z, X)$  ensures that  $\partial_\varepsilon f_{\varepsilon,X,Z}$  is well defined (De La Vallée Poussin (1915), Section 7.40.2) and next  $\mathbb{E} \left( \left[ \frac{\partial_\varepsilon f_{\varepsilon,X,Z}(\varepsilon, X, Z)}{f_{\varepsilon,X,Z}(\varepsilon, X, Z)} \right]^2 \right)$  is also well defined. Assumptions 2.(vii) and 2.(iii) ensure that the parametric sub-model of location  $\{(m(\cdot) - t, f_{\varepsilon,X,Z}), t \in T\}$  is differentiable in quadratic mean. These two assumptions also ensure that any score (corresponding to any sub-model differentiable in quadratic mean) can be decomposed in two components: one related to the nuisance parameter  $m$  and the other one related to

$f_{\varepsilon, X, Z}$ . Assumption 2.(viii) enables us to consider a tangent set sufficiently rich to provide non-trivial restrictions on the DGP to get a finite semi-parametric efficiency bound. In particular, this Assumption rules out the case where the support of  $X$  is not finite and where  $m$  is known up to a finite parameter corresponding to the usual two-step GMM. Last, if Assumption 2.(ix) also holds then the parameter  $\theta_0$  needs to be a moment of  $(Y, Z)$  to ensure that the semi-parametric efficiency bound is finite. Assumptions 2.(vii) to 2.(ix) impose that  $\mathcal{M}$  is not degenerated. But they are compatible with a lot of restrictions of regularity to define  $\mathcal{M}$ , like those considered by Chen (2007) (Hölder, Sobolev or Besov spaces). Concerning shape restrictions such as monotonicity and/or concavity, Assumption 2 imposes that  $m$  is strictly monotonic and/or strictly concave, meaning that  $m$  belongs to the interior of the constraints that define  $\mathcal{M}$ . When  $m$  is at the boundary of  $\mathcal{M}$ , the model becomes irregular and this could help to get a higher rate of convergence<sup>3</sup>, such as in Chetverikov & Wilhelm (2015).

## 4.2 Main result

The key necessary condition for existence of a root-N regular estimator of  $\theta_0$  is given by the following Assumption.

### Assumption 3

(i) With probability 1, the weak-derivative of  $x^+ \mapsto \psi(x^+, X^0)f_X(x^+, X^0)$  is a function on  $\mathbb{R}^{K^+}$ . This function is denoted  $(\psi f_X)^{(j)}$ .

(ii)  $\mathcal{Q} := \{q \in L^2(Z) \text{ s.t. } T^*(q)f_X = (-1)^{|j|}(\psi f_X)^{(j)}\} \neq \emptyset$

The Assumption 3.(i) is related to the distribution theory developed by Schwartz (1997). For  $k = 1, \dots, K^+$ , let  $\varphi_k$  an indefinitely differentiable function with compact support from  $\mathbb{R}$  to  $\mathbb{R}$ . As  $x^+ \mapsto \psi(x^+, X^0)f_X(x^+, X^0)$  is locally integrable then the quantity  $(-1)^{|j|} \int \psi(x^+, X^0)f_X(x^+, X^0) \prod_{k=1}^{K^+} \varphi_k^{(j_k)}(x_k^+) dx_1^+ \dots dx_{K^+}^+$  is well defined for any  $(\varphi_1, \dots, \varphi_{K^+})$ . This mapping defines the weak derivative of order  $(j)$  of  $\psi(x^+, X^0)f_X(x^+, X^0)$ . This weak derivative is a function (up to an isomorphism) denoted  $(\psi f_X)^{(j)}$  if for any  $(\varphi_1, \dots, \varphi_{K^+})$ :

$$\int (\psi f_X)^{(j)}(x^+, X^0) \prod_{k=1}^{K^+} \varphi_k(x_k^+) dx_1^+ \dots dx_{K^+}^+ = (-1)^{|j|} \int \psi(x^+, X^0)f_X(x^+, X^0) \prod_{k=1}^{K^+} \varphi_k^{(j_k)}(x_k^+) dx_1^+ \dots dx_{K^+}^+.$$

When  $x^+ \mapsto \psi(x^+, X^0)f_X(x^+, X^0)$  is continuously differentiable up to order  $(j)$ , successive integrations by parts ensure that  $(\psi f_X)^{(j)}$  coincides with the usual derivative of order  $(j)$  of

<sup>3</sup>but the rate of convergence becomes specific to loss function and risk considered

$\psi f_X$ . Such a restriction will be discussed briefly in the next section. The second condition in the previous assumption is the strongest one and the most abstract one. To get the intuition about the role of this condition, we can consider the particular case where  $(j) = (0)$ . In that case Assumption 3.(i) automatically holds and the Assumption 3.(ii) is equivalent to the existence of  $q \in L^2(Z)$  such that  $\mathbb{E}(q(Z)|X) = \psi(X)$ . Under the completeness condition (i.e.  $\text{Ker}(T) = \{0\}$ ),  $m$  is identified and so is  $\mathbb{E}(\psi m)$  for every  $\psi$ . The completeness condition is equivalent to  $cl(\mathcal{R}(T^*)) = L^2(X)$ . In the favorable case  $\psi$  is in  $\mathcal{R}(T^*)$  and so there exists  $q$  such that  $\theta_0 = \mathbb{E}(\psi(X)m(X)) = \mathbb{E}(q(Z)m(X)) = \mathbb{E}(q(Z)Y)$ . But in other cases  $\psi$  is only in  $cl(\mathcal{R}(T^*)) \setminus \mathcal{R}(T^*)$ . In this latter case,  $\theta_0$  is only *poorly identified* and cannot be expressed as a known moment of the form  $\mathbb{E}(q(Z)Y)$  but only approximated by such moments. Such a situation is not specific to the non parametric instrumental variable and has been formalized in a wider framework by van der Vaart (van der Vaart, 1991 for a detailed discussion or van der Vaart, 2000, Theorem 25.32 for a more synthetic one). In the general case (i.e.  $(j) \neq (0)$ ), if Assumption 2 holds then Assumption 3 is necessary to ensure existence of root-N regular estimator, and in that case there exists  $q \in L^2(Z)$  such that  $\theta_0 = \mathbb{E}(q(Z)Y)$ . This is the main theoretical result of this paper established in the following Theorem.

**Theorem 4.1 (Semi-parametric efficiency bounds)**

*If Assumptions 1 and Assumptions 2.(i) to 2.(viii) hold and if  $\theta_0$  is identified, a regular root-N estimator of  $\theta_0$  exists only if Assumption 3 holds, and in that case there exists  $q^* \in \mathcal{Q}$  such that the semiparametric efficiency bound  $V^*$  satisfies:*

$$\begin{aligned} V^* &= \mathbb{V}[\varepsilon q^*(Z) + R(X)] \\ &= \min_{q \in \mathcal{Q}} \mathbb{V}[\varepsilon q(Z) + R(X)]. \end{aligned}$$

*with  $R(X) = (\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X) - \mathbb{E}((\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X))$ . If moreover Assumption 2.(ix) holds, a regular root-N estimator of  $\theta_0$  exists only if  $\theta_0 = \mathbb{E}(q^*(Z)Y)$ .*

Theorem 4.1 states that Assumption 3 is necessary to ensure the existence of a finite semi-parametric efficiency bound. It also states that the bound is the minimum on  $\mathcal{Q}$  of a function  $V$  and is achieved for a given  $q^* \in \mathcal{Q}$ .

The existence of regular root-N consistent estimator depends on two conditions. Assumption 3.(i) concerns the regularity of  $\psi f_X$  and will be briefly discussed in the next section. Assumption 3.(ii) is more abstract and restrictive and is crucial to ensure finiteness of the semiparametric efficiency bound. Moreover in that case,  $\theta_0$  is a moment of  $(Y, Z)$  only (unless  $(1+t)m \notin \mathcal{M}$ ). So, Theorem 4.1 states that a root-N regular consistent estimator exists only if we are able to "rule out" the endogenous variable  $X$  in the definition of  $\theta_0$ .

And the intuition suggests that such situation is very specific. So, we will discuss more extensively Assumption 3.(ii). In particular, we will focus on two special cases widely considered by theoretical and applied econometricians: the exponential first stage (i.e.  $X|Z$  belongs to an exponential family of distributions) and the additive first stage (i.e.  $X$  is the sum of a function of  $Z$  and a term  $\eta$  independent of  $Z$  but potentially correlated with  $\varepsilon$ ).

Another interesting remark about Theorem 4.1 concerns the existence of supplementary instruments in the exogenous case<sup>4</sup>. In the exogenous case, we have  $Z = X$  and next the semi-parametric efficiency bound of  $\theta_0$  is not finite if and only if  $\mathbb{E} \left[ \left( \frac{[\psi f_X]^{(j)}}{f_X} \right)^2 \right] = +\infty$  (cf. Assumption 3). In that case supplementary instruments  $Z^*$  (ie the situation where  $Z = (X, Z^*)$ ) will not help achieve a parametric rate of convergence (because if Assumption 3.(ii) holds then  $\mathbb{E} \left[ \left( \frac{[\psi f_X]^{(j)}}{f_X} \right)^2 \right] < +\infty$ ).

On the other hand, in the endogenous case (ie when  $X \not\subset Z$ ), supplementary instruments could ensure the validity of Assumption 3.(ii). Last, in all cases (endogenous or exogenous) if Assumption 3 holds for a given set of instruments it will still hold if supplementary instruments are used.

### 4.3 Alternative formulation of the semi-parametric efficiency bound

Assumption 3.(ii) and the expression of the semiparametric efficiency bound have alternative formulations, that makes more transparent the link with the two stage least square estimators used in a parametric framework. Let define the two adjoint operators:

$$\begin{aligned} K : L^2(X) &\rightarrow L^2(Z) & \text{and} & \quad K^* : L^2(Z) \rightarrow L^2(X) \\ f &\mapsto \frac{1}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} T(f) & g(z) &\mapsto T^* \left( \frac{g}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \end{aligned}$$

If Assumption 2.2 holds, the skedastic function  $z \mapsto \mathbb{E}(\varepsilon^2|Z = z)$  is assumed to be bounded and bounded away from 0 and so  $\mathcal{R}(T^*) = \mathcal{R}(K^*)$ ,  $\mathcal{R}(T) = \mathcal{R}(K)$ ,  $\mathcal{R}(TT^*) = \mathcal{R}(KK^*)$  and  $\mathcal{R}(T^*T) = \mathcal{R}(K^*K)$ .

For any linear and bounded operator  $A$  from an Hilbert  $H_1$  to an Hilbert  $H_2$ , we can define the generalized inverse  $A^\dagger$  of  $A$  on  $\mathcal{R}(A)$  by:

$$A^\dagger(y) = u \text{ if and only if } u = \text{Arg} \min_{v \in H_1: A(v)=y} \|v\|_{H_1}^2.$$

**Proposition 4.2** *If Assumptions 1 to 3 hold and if  $\theta_0$  is identified, then the semi-parametric efficiency bound of  $\theta_0$  is equal to:*

---

<sup>4</sup>We are grateful to Arthur Lewbel for suggesting to look at this question.

(i)

$$V^* = \mathbb{E} \left[ \left( K^{*\dagger} \left( (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right) \right)^2 \right], \\ + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right]$$

(ii) For any  $q \in \mathcal{Q}$ ,

$$V^* = \mathbb{E} \left[ \left( K^{*\dagger} K^* \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right)^2 \right], \\ + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right]$$

If and only if  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \in \mathcal{R}(K^* K)$ , these expressions reduce to:

(iii)

$$V^* = \mathbb{E} \left( \left( (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right) (K^* K)^\dagger \left( (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right) \right) \\ + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right],$$

(iv) For any  $q \in \mathcal{Q}$ ,

$$V^* = \mathbb{E} \left( \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) K (K^* K)^\dagger K^* \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right) \\ + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right],$$

The formula (iii) has already been established by Ai & Chen (2012), but it is valid only if  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \in \mathcal{R}(K^* K)$ , a stronger condition than Assumption 3.(ii).

Under Assumption 3.(ii) the condition  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \in \mathcal{R}(K^* K)$  is equivalent to the fact that  $K^{*\dagger} (-1)^{|j|} \left( \frac{[\psi f_X]^{(j)}}{f_X} \right) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \in \mathcal{R}(K) + \text{Ker}(K^*)$  or equivalently  $\mathcal{P}_{\overline{\mathcal{R}(K)}} \left[ K^{*\dagger} (-1)^{|j|} \left( \frac{[\psi f_X]^{(j)}}{f_X} \right) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right] \in \mathcal{R}(K)$ , where  $\mathcal{P}_{\mathcal{R}(K)}$  is the orthogonal projection on the closure of  $\mathcal{R}(K)$ . In a non-parametric context, the assumption  $\mathcal{R}(K) = \overline{\mathcal{R}(K)}$  (or equivalently  $\mathcal{R}(T) = \overline{\mathcal{R}(T)}$ ) is a restrictive condition that will be discussed below in Proposition 5.2. Then in general the orthogonal projection of an element on  $\overline{\mathcal{R}(K)}$  does not belong to  $\mathcal{R}(K)$ .

## 5 When Assumption 3 holds ?

We will discuss separately Assumption 3.(i) on the one hand and Assumption 3.(ii) on the other hand. Indeed, the first assumption is only a regularity condition on the functions  $\psi$  and  $f_X$ . For a given theoretical DGP and function  $\psi$ , it is not difficult to determine if it holds or not. We will only emphasize that the condition holds only if some regularity conditions at the boundary of the support of  $X$  hold and discuss the role of this condition on the fact that  $\theta_0$  is a moment of  $(Z, Y)$  under Assumption 2.(ix). On the contrary, Assumption 3.(ii) involves the adjoint operator  $T^*$  and is quite abstract at this stage, so it is not easy to know if it holds for a given DGP and a given function  $\psi$ . So, the largest part of the following will concern Assumption 3.(iii).

### 5.1 Discussion of Assumption 3.(i)

#### **Proposition 5.1 (Regularity condition on $\psi f_X$ )**

*If Assumption 3.(i) holds then, the function  $x \mapsto \psi(x)f_X(x)$  admits is a derivative of order  $(j') \leq (j)$  on set  $D_{(j')}$  such that  $\mathbb{P}(X \in D_{(j')}) = 1$ . For any  $k = 1, \dots, K$  such that  $0 \leq j'_k < j_k$ , with probability one,  $x_k \mapsto (\psi f_X)^{(j')}(X_1, \dots, X_{k-1}, x_k, X_{k+1}, \dots, X_K)$  (is almost-everywhere equal to a function that) vanishes at the boundary of the support of  $X_k | X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_K$ .*

*If  $x \mapsto \psi(x)f_X(x)$  is  $(j)$  times-continuously differentiable on  $\mathbb{R}^K$  then Assumption 3.(i) holds.*

The previous proposition ensure that weak derivative defined in Assumption 2.(i) is equal to a usual derivative (up to a negligible set). Moreover, this derivative needs to vanish at the boundary of the support of  $\psi f_X$ . When  $(j) > (0)$ , this rules out the function  $\psi = \mathbb{1}_C$  where  $C$  is a set on which  $f_X$  is bounded below 0.

To understand why such regularity at the boundary of the support is necessary for the existence of root-N and regular estimator, we can consider the simple case where there is only one variable  $X$  with a convex compact support  $[a; b]$  dominated by the Lebesgue measure. If  $j > 0$  then  $(\psi f_X)^{(k)}$  for  $k < j$  vanishes at  $a$  and  $b$ . If  $f_X$  (or one of its  $j - 1$  first derivatives) does not vanish at the boundary, then  $\psi$  needs to vanish sufficiently quickly at the boundary of the support of  $X$ . Then successive integrations by part of  $\varphi^{(j)}\psi f_X$  hold for any  $\varphi \in \Phi$ . Under Assumption 2.(viii), the existence of a root-N consistent and regular estimator implies that "the parts between brackets" in the successive integrations by parts



of  $\varphi^{(j)}\psi f_X$  vanish i.e.:

$$\lim_{(\underline{x}, \bar{x}) \rightarrow (a, b)} \sum_{l=0}^{j-1} (-1)^l [(\psi(x)f_X(x))^{(l)} \varphi^{(j-1-l)}(x)]_{x=\underline{x}}^{x=\bar{x}} = 0.$$

Because this result needs to hold for any  $\varphi \in \Phi$ , then the successive derivatives of  $\psi f_x$  needs to vanish at the boundaries  $a$  and  $b$ .

Such a regularity condition at the boundary of the support is commonly used as a sufficient condition to get an efficient estimator for average partial derivative of weighted average derivatives of index models with exogenous regressors (Powell et al., 1989, Newey & Stoker, 1993, Cattaneo et al., 2014). In a wider class of models, relaxing the index structure and the exogeneity of regressors, we show that such an assumption is in fact necessary for the existence of regular and root-N consistent estimators.

When  $X$  is multidimensional, the previous reasoning can be generalized, however, it involves a generalization of integration by part to a multidimensional setting that complicates the presentation of the result without adding any essential benefit.

When  $\psi f_X$  is  $(j)$  continuously differentiable then successive integration by parts could be performed and in this the weak derivatives are the usual derivatives. Interestingly, such regularity condition is often assumed to show root-N consistency of specific estimators (Cattaneo et al., 2014 for density weighted average derivative with exogenous regressors or Ai & Chen, 2007 for the endogenous regressors), or to derive the semiparametric efficiency bound of a parameter (Ai & Chen, 2012). In these papers, the assumption of continuously differentiability is *sufficient*, in this paper we show that a slightly weaker regularity condition (namely the weak differentiability) is in fact *necessary*.

**Example: The average under counterfactual distribution** corresponds to the case where  $(j) = (0, \dots, 0)$  and  $\psi = \frac{\tilde{f}_X}{f_X}$ , in that case Assumption 3.(i) does not restrict the choice of  $\tilde{f}_X$ . However, we will show in the sequel that averages under counterfactual distributions are not always root-N estimable because Assumptions 3.(ii) restricts  $\tilde{f}_X$ .

**Example: The weighted average partial derivative** corresponds to  $X^+ = X_1$  and  $(j) = (1, 0, \dots, 0)$  and the average marginal effect is a special case corresponding to  $\psi = 1$ .  $\theta_0 = \mathbb{E}(\psi(X)\partial_{x_1}m(X))$  is root-N estimable as soon as  $x_1 \mapsto \psi(x_1, X^0)f_{X_1|X^0}(x_1)$  is (equal  $X$  almost-surely to) an absolutely continuous function on  $\mathbb{R}$ . For instance, if  $X \sim \mathcal{N}(0, 1)$ , then  $\theta_0$  is root-N estimable only if  $\psi$  admits a derivative whose integral is equal (up to an additive constant) to  $\psi$ . This rules out discontinuous  $\psi$  like  $\mathbb{1}\{x \in [a; b]\}$  or  $\mathbb{1}\{x > 0\}$  but allows for kinked function such as  $x\mathbb{1}\{x > 0\}$ . Similarly, if  $X$  is drawn in a  $B(\alpha, \beta)$  distribution, then  $\theta_0 = \mathbb{E}(\psi m')$  is root-N estimable only if  $\psi$  is twice differentiable on  $]0; 1[$  and  $\psi(x) = o(x^{1-\alpha})$  and  $\psi(1-x) = o(x^{1-\beta})$  for  $x \sim 0$ . Notice that if  $\alpha \leq 1$  (respectively  $\beta \leq 1$ ) then  $\psi$  needs to vanish at 0 (respectively at 1) sufficiently quickly to ensure existence

of a root-N consistent estimator. In particular, root-N consistent estimators for the average marginal effect ( $\psi = 1$ ) or for the density weighted average derivative ( $\psi = f_X$ ) does not exist if  $\alpha \leq 1$  or  $\beta \leq 1$ . For the Beta distribution, note that Assumption 3.(ii) reinforces these conditions at the boundary of the support: the square integrability of  $q$  ensures that root-N consistent estimators for the average marginal effect ( $\psi = 1$ ) do not exist if  $\alpha \leq 2$  or  $\beta \leq 2$ .

We now turn to Assumption 3.(ii), which is particularly abstract in the endogenous case (ie  $X \neq Z$ ). There is at least two frameworks in which the condition  $\mathcal{Q} \neq \emptyset$  have more comprehensive interpretations. The first one corresponds to the case where the distribution of  $X|Z$  belongs to an exponential family (we will refer to that case as "the exponential first stage"). The second one corresponds to the case where the first stage is "additive" i.e.  $X = p(Z) + \eta$  with  $\eta \perp\!\!\!\perp Z$ .

## 5.2 When is $\mathcal{Q} \neq \emptyset$ ?

### 5.2.1 On the closedness of of range of linear operators

The Assumption 3.(ii) implies that  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} \in L^2(X)$ . Reciprocally, if  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} \in L^2(X)$ , then  $\mathcal{Q}$  is never empty if  $\mathcal{R}(T^*) = L^2(X)$ . The condition  $\mathcal{R}(T^*) = L^2(X)$  holds if and only if  $\text{Ker}(T) = \{0\}$  (completeness condition) and  $\mathcal{R}(T^*)$  is closed. The completeness condition is known to be hard to test in the data (Canay et al., 2013 and Freyberger, 2015) and impose abstract restrictions on the joint distribution of  $(X, Z)$  (see d'Haultfœuille (2011), Hu & Schennach (2008) or ? for discussion of criteria ensuring the completeness of non-completeness). The closedness of  $\mathcal{R}(T^*)$  is even more restrictive as illustrated by the following proposition.

### **Proposition 5.2 (Closedness of $\mathcal{R}(T)$ and $\mathcal{R}(T^*)$ )**

- (i) *The range of  $T$  is closed if and only if the range of  $T^*$  is closed.*
- (ii) *If  $T$  (or  $T^*$ ) is a compact operator, then  $\mathcal{R}(T)$  is closed if and only if  $\mathcal{R}(T)$  and  $\mathcal{R}(T^*)$  are finite dimensional.*
- (iii) *If the completeness condition holds (ie  $\text{Ker}(T) = \{0\}$ ) and if  $X|Z$  admits a density with respect to the Lebesgue measure  $Z$ -almost surely, then neither  $\mathcal{R}(T)$  neither  $\mathcal{R}(T^*)$  are closed.*

In our instrumental variable framework, if the product measure of  $(X, Z)$  is dominated by the product of marginal measures of  $X$  and  $Z$ , then  $T$  is a compact operator as soon

as  $\mathbb{E}(f_{X,Z}(X, Z)) < \infty$  (cf. discussion of Assumption A1 in Darolles et al., 2011). The proposition 5.2.(ii) shows that in that case,  $\mathcal{R}(T) = L^2(Z)$  and  $\mathcal{R}(T^*) = L^2(X)$  only if  $X$  and  $Z$  have finite supports (and then  $T$  and  $T^*$  are characterized by full rank matrixes). When the support of  $X$  or  $Z$  is not finite, then even in the favorable case of completeness, Proposition 5.2.(iii) shows that  $\mathcal{R}(T)$  (respectively  $\mathcal{R}(T^*)$ ) is not closed and then not equal to  $L^2(Z)$  (respectively  $L^2(X)$ ).

So because  $\mathcal{R}(T)$  is in general not closed when  $X \neq Z$ , criteria on the DGP and the parameter  $\theta_0$  ensuring that  $\mathcal{Q} \neq \emptyset$  are hard to formulate in the most general case. In the following, we investigate the condition  $\mathcal{Q} \neq \emptyset$  in two specific cases: when the first stage belongs to an exponential distribution and when the first stage is additive.

### 5.2.2 Specialization to the exponential first stage

Exponential families of distributions have many nice features: they embed many usual parametric models and provide sufficient statistics. So, they have been widely studied by econometricians such as in Gourieroux et al. (1984). Concerning the non-parametric instrumental variables, Newey & Powell (2003) have established the completeness condition for some (but not all) models in which the true conditional distribution  $X|Z$  belongs to an exponential family. When the first stage is exponential, the condition  $\mathcal{Q} \neq \emptyset$  is equivalent to the condition that an identifiable function of  $X$  (depending on  $\psi$  and on the distribution of  $(Z, X)$ ) is the Laplace transform of a measure. The range of Laplace transforms of measures is not easily characterizable, but some properties can be derived. On the other hand the range of Laplace transform of linear forms has been precisely characterized by L. Schwartz and this implies interesting necessary conditions for the non-emptiness of  $\mathcal{Q}$ . Note that the following discussion differs from the one of Ai & Chen (2007) (pp 25-26), which show that root-N consistent estimator exists for  $\theta_0 = \mathbb{E}(m^{(1)}(X))$  when  $T$  is Hilbert-Schmidt and the marginal density distribution of  $X$  is proportional to  $\exp(t(x))$  with  $t$  a finite order polynomial.

Let  $W$  denote the set of exogenous regressors i.e. the variables that are simultaneously in  $X$  and  $Z$ . Let  $\tilde{X}$  denote the set of endogenous variable and  $\tilde{Z}$  the set of excluded instruments.

#### Assumption 4 (Exponential First Stage)

- (i)  $\tilde{X}, \tilde{Z}|W$  is dominated by a product measure  $\pi_{\tilde{X}|W} \otimes \pi_{\tilde{Z}|W}$
- (ii)  $f_{\tilde{X}|\tilde{Z}=\tilde{z}, W=w}(\tilde{x})$  takes the form  $h_w(\tilde{x})g_w(\tau_w(z)) \exp(-\mu_w(\tilde{x})'\tau_w(\tilde{z}))$ , with where  $\mu_w(x)$  and  $\tau_w(z)$  are elements of  $\mathbb{R}^{p_w}$  and  $0 \in \text{Supp}(\mu_w(\tilde{X})|W = w)$  and  $\text{span}(\text{Supp}(\mu_w(X)|W = w)) = \mathbb{R}^{p_w}$ .

Assumption 4.(i) is mainly needed for the sake of simplicity, and means that the joint distribution of  $X, Z|W$  is not degenerated as in the exogenous case that has been previously studied. Assumption 4.(ii) is equivalent to  $f_{\tilde{X}|\tilde{Z},W}(\tilde{x}) = h_W(\tilde{x})\tilde{g}_W(\tilde{Z}) \exp\left(-\mu_W(\tilde{x})'\tau_W(\tilde{Z})\right)$ , because an integration with respect to  $\tilde{x}$  ensures that  $\tilde{g}_w(z)$  is a function of  $(\tau(z), w)$ . The condition  $0 \in \text{Supp}(\mu(X)|W)$  can be assumed without loss of generality and is just a normalization<sup>5</sup>. Similarly, the condition  $\text{span}(\text{Supp}(\mu_w(X)|W = w)) = \mathbb{R}^{p_w}$  can be assumed without loss of generality<sup>6</sup>, as soon as we adopt the convention  $\mathbb{R}^0 = \{0\}$ .

Newey & Powell (2003) have proved that the completeness condition holds for a subclass of models with an exponential first stage (for instance, for  $X|Z \sim \mathcal{N}(b(Z), \sigma^2)$  with  $\mathcal{R}(b)$  containing an open set of  $\mathbb{R}^K$ ). But some models with an exponential first stage do not verify the condition of completeness for at least two reasons. The first reason is the lack of variation in  $Z$ , for instance if  $Z$  is binary and  $X = Z + \varepsilon$  with  $\varepsilon|Z \sim \mathcal{N}(0, 1)$ . A second reason of failure of the completeness condition is due to the fact that for some models the dimension of the linear span of the range of  $\mu$  and  $\tau$  is "too large". For instance, if  $X|Z \sim \mathcal{N}(b(Z), \sigma^2(Z))$ , then the linear span of the range of  $\mu(X) = (X^2/2, X)$  and  $\tau(Z) = \left(-\frac{1}{\sigma^2(Z)}, \frac{b(Z)}{\sigma^2(Z)}\right)$  is  $\mathbb{R}^2$  whereas  $X$  is only real and it follows that the completeness condition does not hold in general<sup>7</sup>. In both cases, Assumption 4 holds and despite the failure of the completeness condition,  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$  will be root-N estimable for some functions  $\psi$ . And on the other hand, we will also show that even if the completeness condition holds, some simple parameter will not be root-N estimable.

For any  $\gamma \in \mathbb{R}^{p_w}$ , let  $\nu_w(\gamma) = \int g_w(t) \exp(-\gamma't) dF_{\tau_w(\tilde{Z})|W=w}(t)$ . The random set  $\{\gamma \in \mathbb{R}^{p_w} : \nu_w(\gamma) < \infty\}$  is a convex set<sup>8</sup> that contains almost-surely  $\mu_w(\tilde{X})$  (because  $h_W(X)\nu_w(\mu_W(\tilde{X})) = f_{\tilde{X}|W}$  is finite almost-surely and  $\mathbb{P}(h_W(X) = 0|W) = 0$  almost-surely). Then it contains with probability one  $\Gamma_W = \text{int co Supp}(\mu_w(\tilde{X})|W)$ , the interior of the convex hull of  $\text{Supp}(\mu_w(\tilde{X})|W)$ . The function  $\nu_w$  can be naturally extended to  $\Gamma_W + i\mathbb{R}^{p_w}$  (where  $i^2 = -1$ ), by  $\nu_w(u + iv) = \int g_w(t) \exp(-(u + iv)'t) dF_{\tau_w(\tilde{Z})|W}(t)$ . This is the Laplace transform of the signed measure  $M_W$  defined by  $dM_W(t) = g_w(t)dF_{\tau_w(\tilde{Z})|W}(t)$ . Such a Laplace transform is a function that verifies strong conditions of smoothness. Interestingly another function (denoted  $\rho_w$  in the following theorem) needs to verify similar conditions

<sup>5</sup>Indeed  $g_w(\tau(z)) \exp(-\mu(x)'\tau(z)) = \tilde{g}_w(\tau(z)) \exp(-\tilde{\mu}(x)'\tau(z))$  with  $\tilde{g}_w(\tau(z)) = g_w(\tau(z)) \exp(-\mu(x_0)'\tau(z))$  and  $\tilde{\mu}(x) = \mu(x) - \mu(x_0)$  for a given  $x_0 \in \text{Supp}(X|W)$ .

<sup>6</sup>If  $\text{span}(\text{Supp}(\mu_w(X)|W = w))$  is a proper subspace of dimension  $r_w < d_w$ , we can always reparameterize  $\mu_w$  and  $\tau_w$ . Indeed, in such case, up to a reordering of the component of  $\mu_w$  and  $\tau_w$ , we can assume that  $\mu_j = \sum_{i=1}^r \lambda_{ji}\mu_i$  for  $j > r$ , and next we have  $\sum_{i=1}^d \mu_i\tau_i = \sum_{i=1}^r \mu_i\tilde{\tau}_i$  with  $\tilde{\tau}_i = \tau_i + \sum_{j=r+1}^d \lambda_{ji}\tau_j$ .

<sup>7</sup>For instance, if  $b$  is bounded on  $\text{Supp}(Z)$ , then it exists  $\alpha$  and  $\beta$  such that  $b^2(Z) + \alpha b(Z) + \beta < 0$  on  $\text{Supp}(Z)$  and in that case,  $\mathbb{E}(X^2 + \alpha X + \beta|Z) = 0$  for  $\sigma^2(Z) = -b^2(Z) - \alpha b(Z) - \beta$ . To our best knowledge, in such case of heteroscedastic first stage, necessary conditions on  $b(z)$  and  $\sigma^2(z)$  to ensure the completeness are not well understood. See d'Haultfœuille (2011) for a discussion of the completeness assumption.

<sup>8</sup> $\exp(-(\gamma_1 + \gamma_2)'t/2) \leq \exp(-\gamma_1't) + \exp(-\gamma_2't)$  because  $ab \leq a^2 + b^2$ , and then the convexity follows.

of smoothness. Let us introduce some notions of regularity of functions from  $\mathbb{C}^p$  to  $\mathbb{C}$ .

A function  $f$  from an open set  $D \subset \mathbb{C}^p$  to  $\mathbb{C}$  is holomorphic on  $D$  if for any  $z_0 \in D$ ,  $\lim_{z \rightarrow z_0, z \neq z_0} \frac{f(z) - f(z_0)}{z - z_0}$  exists. Roughly speaking, an holomorphic function is "very" regular and admits an analytic development  $f(z) = \sum_{(n) \in \mathbb{N}^p} a_{(n)} z^{(n)}$  where  $z^{(n)} = \prod_{j=1}^p z_j^{n_j}$ .

L. Schwartz has defined the Laplace transform for a wider class of object than functions or measures by duality arguments (Schwartz, 1997, and see Appendix 7.6 for details). The Laplace transforms are always defined on a set  $C + i\mathbb{R}^p$ , where  $C$  is a convex set. When  $C$  has a non-empty interior, the Laplace transform is an holomorphic function on  $\text{int}(C) + i\mathbb{R}^p$  and verifies some conditions of domination by polynomials. All the Laplace transforms considered in this paper are Laplace transforms of signed measures. In this case, the condition of domination of the Laplace transform can be refined, and last when the measure is dominated by the Lebesgue measure additional conditions apply. Under Assumption 4,  $\mathcal{Q}$  is not empty if and only if a function  $\varrho_W(x)$  depending on  $\psi$ ,  $f_X$ ,  $\mu$  and  $\nu_W$  is the Laplace transform of a signed measure. The next Theorem uses this property to derive necessary conditions on  $\psi$  and the joint distribution of  $(X, Z)$  for the existence of a regular root-N estimator. These conditions are sufficient to ensure that  $\varrho_W$  is the Laplace transform of a linear form, so informally speaking, necessary conditions given below are almost-sufficient to ensure the non-emptiness of  $\mathcal{Q}$  and can be used to reject existence of root-N estimator in many settings.

### Theorem 5.3

If Assumptions 1, 2 and 4 hold then Assumption 3.(ii) holds only if:

- (i)  $(-1)^{|j|} \frac{[\psi(X)f_X(X)]^{(j)}}{f_X(X)} \in L^1(\mu_W(\tilde{X}), W)$ , i.e. there exists  $r_W$  such that:  
 $r_W(\mu_W(\tilde{X})) = (-1)^{|j|} \frac{[\psi(X)f_X(X)]^{(j)}}{f_X(X)}$  and  $r_W$  is integrable with respect to  $F_{\mu_W(\tilde{X})|W}$ ,  
 $W$ -almost surely,
- (ii)  $W$ -almost-surely, there exists a convex  $C_W$  such that  $\mathbb{P}(\mu_W(X) \in C_W|W) = 1$  such that  $\varrho_W = r_W \nu_W$  is the Laplace transform of a signed measure with finite totale variation dominated by  $dF_{\tau(Z)|W}$  on  $C_W + i\mathbb{R}^{p_W}$ .
- (iii)  $W$ -almost-surely,  $\varrho_W$  is an holomorphic function on  $\Gamma_W + i\mathbb{R}^{p_W}$ ,
- (iv)  $W$ -almost-surely,  $\varrho_W$  is bounded on  $C + i\mathbb{R}^{p_W}$  for any compact subset  $C$  of  $\Gamma_W$ ,
- (v) when  $\tau_W(\tilde{Z})|W$  is dominated by the Lebesgue measure on  $\mathbb{R}^{p_W}$ ,  $\varrho_W(u + iv)$  tends to 0 when  $v$  tends to infinity for any  $u \in \Gamma_W$ .

Moreover, when a root- $N$  regular consistent estimator of  $\theta_0$  exists, the semiparametric efficiency bound of  $\theta_0$  is:

$$V^* = \mathbb{V}(\varepsilon q^*(Z) + R(X)),$$

with

$$q^*(Z) = \frac{1}{g_w(\tau(\tilde{Z}))\mathbb{E}(\varepsilon^2|Z)\mathbb{E}\left[1/\mathbb{E}(\varepsilon^2|Z)|\tau_W(\tilde{Z}), W\right]} \frac{d\mathcal{L}^{-1}[\varrho_W]}{dF_{\tau_W(\tilde{Z})|W}}(\tau_W(\tilde{Z})) \\ + \frac{\mathbb{E}\left[\mathbb{E}(\varepsilon R(X)|Z)/\mathbb{E}(\varepsilon^2|Z)|\tau_W(\tilde{Z}), W\right]}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}\left[1/\mathbb{E}(\varepsilon^2|Z)|\tau_W(\tilde{Z}), W\right]} - \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)}$$

and

$$R(X) = (\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X) - \mathbb{E}\left((\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X)\right).$$

Condition (i) of the Theorem means that the quantity  $\frac{[\psi(X)f_X(X)]^{(j)}}{f_X(X)}$  is independent of  $X$  conditional on  $\mu(X)$  and  $W$ . Intuitively, this means that  $(\mu(X), W)$  is a sufficient statistic for  $\frac{[\psi(X)f_X(X)]^{(j)}}{f_X(X)}$ , i.e. there is enough variations in  $(\mu(X), W)$  to recover variations of  $[\psi f_X]^{(j)}/f_X$ .

Conditions (i) and (ii) are necessary and sufficient to ensure the existence of  $q \in L^1(Z)$  such that  $T^*(q)f_X = [\psi f_X]^{(j)}$ . And if such a function  $q$  is also in  $L^2(Z)$ , it will be in  $\mathcal{Q}$ .

Moreover, other conditions of the Theorem imply strong regularity conditions on  $\psi$  and on  $f_X$  to ensure existence of regular root- $N$  consistent estimator of  $\theta_0$ . In condition (iii), holomorphy is a necessary condition to ensure that  $\varrho_W$  is a Laplace transform (in the sense of L. Schwartz, see Proposition 6 in Chapter VIII of Schwartz (1997)). We prove in the appendix that Condition (iv) is necessary for the Laplace transform to be the Laplace transform of a signed measure. Both (iii) and (iv) are sufficient to ensure that  $\varrho$  is a Laplace transform in the sense of L. Schwartz (see Appendix 7.7.2 and Proposition 6 in Chapter VIII of Schwartz (1997) for more details). Last, Condition (v) is a refinement of (iv) when  $\tau_W(\tilde{Z})|W$  is dominated by the Lebesgue measure.

In the expression of the semiparametric efficiency bound,  $\frac{d\mathcal{L}^{-1}[\varrho_W]}{dF_{\tau_W(\tilde{Z})|W}}$  denotes the density of the inverse Laplace transform of  $\varrho_W$  with respect to the Lebesgue-Stieljes measure of  $\tau_W(\tilde{Z})|W$ . The condition (ii) of the Theorem, the injectivity of the Laplace transform and the Radon-Nikodym Theorem, ensure that the density is well defined. Under exponential first stage, Theorem 5.3 gives the solution of the minimization problem given in Theorem 4.1. Not surprisingly, the solution  $q^*$  depends on the skedastic function  $\mathbb{E}(\varepsilon^2|Z)$  and depends also on the dependence between  $\varepsilon$  and  $X$  through the quantity

$\mathbb{E}(\varepsilon R(X)|Z)$ . When  $\tau_W$  is injective, the sigma-algebra generated by  $Z$  is equal to the sigma-algebra generated by  $(\tau_W(\tilde{Z}), W)$ , and in that case  $q^*(Z)$  has a simpler expression:  $q^*(Z) = \frac{1}{g_w(\tau(\tilde{Z}))} \frac{d\mathcal{L}^{-1}[g_W]}{dF_{\tau_W(\tilde{Z})|W}}(\tau_W(\tilde{Z}))$ . More generally, this simpler expression holds as soon as  $\mathbb{E}(\varepsilon^2|Z) = \mathbb{E}(\varepsilon^2|\tau_W(\tilde{Z}), W)$  and  $\mathbb{E}(\varepsilon R(X)|Z) = \mathbb{E}(\varepsilon R(X)|\tau_W(\tilde{Z}), W)$ .

Even when  $\psi$  and  $f_X$  are strongly regular,  $\theta_0$  is not always root-N estimable as illustrated by simple examples detailed below. Indeed, the behavior of the holomorphic function  $\nu_W(u)$  when the imaginary part of  $u$  tends to infinity is related to the size of the set of parameters  $\theta_0$  that are root-N estimable: the faster  $\nu_W(x + iy)$  tends to 0 when  $|y| \rightarrow \infty$ , the stronger is the link between  $X$  and  $Z$  and the larger is the set of functions  $r_W$  such that  $r_W \nu_W$  is a Laplace transform and then the larger is the range of  $T^*$  and the set of  $\psi$  such that  $\theta_0$  is root-N estimable. In particular, some functions  $\psi$  correspond to parameters which are root-N estimable if the strength of instruments is sufficient and which is not otherwise.

**Example: Bivariate normal distribution**

Consider the bivariate normal case:

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

The first stage is  $X = \rho Z + \sqrt{1 - \rho^2} \eta$  with  $\eta|Z \sim \mathcal{N}(0, 1)$ .

Consider a shift  $\Delta m$  in the mean of  $X$  keeping constant the unobserved variable  $\varepsilon$ . In that case, the average of the outcome under such counterfactual distribution corresponds to  $\theta_0 = \mathbb{E} \left( \exp \left( -\frac{1}{2} [(X - \Delta m)^2 - X^2] \right) m(X) \right)$  and this parameter is root-N estimable.

On the other hand, the average of the outcome under a shift  $\sigma^2 - 1$  in the variance of  $X$  corresponds to  $\theta_0 = 1/\sqrt{\sigma^2} \mathbb{E} \left( \exp \left( -\frac{1}{2} [X^2(1/\sigma^2 - 1)] \right) m(X) \right)$ . This parameter is root-N estimable if and only if  $\sigma^2 > 1 - \rho^2$ . Intuition behind this result is that when  $\sigma$  tends to 0,  $\theta_0$  tends to  $m(0)$  which is not a parameter that can be consistently estimated at the root-N rate even in the exogenous case. On the other hand if the counterfactual distribution has a higher variance than the original one ( $\sigma > 1$ ) the existence of a root-N consistent estimator is always ensured as soon as  $\rho \neq 0$ . We can also consider a choice of  $\sigma$  and remark, that the semi-parametric efficiency bound is finite if and only if  $\rho^2 > 1 - \sigma^2$ . When the counterfactual distribution has a lower variance than the original ones, we are in an intermediary case in which the existence of root-N consistent estimator depends on the strength of the instruments.

Similarly, the average marginal effect  $\mathbb{E}(m'(X))$  is always root-N estimable, whereas the density weighted average derivative  $\mathbb{E}(f_X(X)m'(X))$  is root-N estimable if and only if  $\rho^2 \geq 1/2$ . The corresponding semi-parametric efficiency bounds are derived in the Appendix and are reported in Table 1.

**Example: Average Marginal Effect with normal homoscedastic first stage** Consider the case where  $\theta_0 = \mathbb{E}(m'(X))$ , with  $X|Z \sim \mathcal{N}(t(Z), \sigma^2)$  and where Assumption 1 and 2 holds. In that case Assumption 4 is fulfilled with  $\mu(X) = X$  and  $\tau(Z) = -t(Z)/\sigma^2$ ,  $h(X) = \exp(-X^2/(2\sigma^2))$ ,  $g(\tau(Z)) = (2\pi\sigma^2)^{-1/2} \exp(-\tau^2(Z)\sigma^2/2)$ . In that case, there is no variable  $W$  and  $\Gamma = \Gamma_W = \mathbb{R}$ ,  $\nu_W(x) = \nu(x) = (2\pi\sigma^2)^{-1/2} \int \exp(-t^2\sigma^2/2) \exp(-xt) dF_{\tau(Z)}(t)$  and  $r_W(x) = r(x) = -\frac{f'_X(x)}{f_X(x)} = \frac{x}{\sigma^2} + \frac{(2\pi\sigma^2)^{-1/2} \int t \exp(-t^2\sigma^2/2) \exp(-xt) dF_{\tau(Z)}(t)}{\nu(x)}$ . So, we conclude that

$$\rho_W(x) = \rho(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \int \exp(-t^2\sigma^2/2) \exp(-xt) \left( \frac{x}{\sigma^2} + t \right) dF_{\tau(Z)}(t).$$

Now, assume for the sake of simplicity that  $\mathbb{E}(|\tau(Z)| \exp(-\tau(Z)^2\sigma^2/2 - x\tau(Z))) < \infty$  for any  $x \in \mathbb{C}$ . The Condition (ii) of Theorem 5.3 holds only if  $\tau(Z)$  admits a locally absolutely continuous density with respect to the Lebesgue measure (see Lemma 7.5 in appendix). So the density of  $\tau(Z)$  is continuous and almost everywhere differentiable on  $\mathbb{R}$  and its derivative  $f'_{\tau(Z)}$  is locally integrable. If  $\tau(Z)$  admits a mass point or if  $\tau(Z)$  has a density bounded below on its compact support (such as the uniform distribution), regular root-N estimators do not exist. A bit of algebra gives:

$$\frac{1}{g(\tau(Z))} \frac{d\mathcal{L}^{-1}(\varrho)}{dF_{\tau(Z)}}(\tau(Z)) = \frac{1}{\sigma^2} \frac{f'_{\tau(Z)}(\tau(Z))}{f_{\tau(Z)}(\tau(Z))} = -\frac{f'_{t(Z)}(t(Z))}{f_{t(Z)}(t(Z))},$$

Then  $\mathcal{Q}$  is not empty only if  $\mathbb{E} \left( \frac{f'^2_{t(Z)}(t(Z))}{f^2_{t(Z)}(t(Z))} \right) < \infty$  which means that the Fisher information matrix of location of  $t(Z)$  is finite. Last, if moreover  $t$  is one to one, the semiparametric efficiency bound of the average marginal effect is:

$$V^* = \mathbb{V} \left( m'(X) - (Y - m(X)) \frac{f'_{t(Z)}(t(Z))}{f_{t(Z)}(t(Z))} \right).$$

**Example: Average Marginal Effect with normal heteroscedastic first stage** Consider the case where  $\theta_0 = \mathbb{E}(m'(X))$ , with  $X|Z \sim \mathcal{N}(t(Z), \sigma^2(Z))$ . For the sake of simplicity, we assume moreover that  $\mathcal{Z}$  is compact and that the range of  $t$  and  $\sigma^2$  are respectively included in compacts of  $\mathbb{R}$  and  $]0; \infty[$ . To our best knowledge, conditions on  $(t(Z), \sigma^2(Z))$  that ensure the completeness condition are not known. Here, we discuss existence of regular and root-N consistent estimators. In that case,  $h(X) = 1$ ,  $\mu(X) = (\mu_1, \mu_2) = (X, X^2/2)$ ,  $\tau(Z) = (\tau_1, \tau_2) = \left(-\frac{t(Z)}{\sigma^2(Z)}, \frac{1}{\sigma^2(Z)}\right)$  and  $g(\tau) = \frac{\tau_2^{1/2}}{(2\pi)^{1/2}} \exp(-\frac{\tau_1^2}{2\tau_2})$ . Note that the restrictions assumed on  $t(Z)$  and  $\sigma^2(Z)$  ensure that the range of  $\tau_1$  and  $\tau_2$  are respectively included in compacts of  $\mathbb{R}$  and  $]0; \infty[$ . It follows that any continuous function on  $\mathbb{R} \times ]0; \infty[$  of  $(\tau_1(Z), \tau_2(Z))$  admits a finite moment. The support of  $\mu(X)|Z$  does not depend on  $Z$  and is a parabola in the plan and the relative interior of its convex hull is  $\Gamma = \{(a, b) \in$



$\mathbb{R}^2$  s.t.  $b > \frac{a^2}{2}$ . For  $(\mu_1, \mu_2) \in \mathbb{C}^2$ ,  $\nu(\mu_1, \mu_2)$  is well defined and equal to:

$$\nu(\mu_1, \mu_2) = \int \frac{t_2^{1/2}}{\sqrt{2\pi}} e^{-\frac{t_1^2}{2t_2} - \mu_1 t_1 - \mu_2 t_2} dF_{\tau(Z)}(t_1, t_2),$$

and  $f_X(x) = \nu(x, x^2/2)$ . It follows that:

$$\varrho(\mu_1, \mu_2) = \int (t_1 + \mu_1 t_2) \frac{t_2^{1/2}}{\sqrt{2\pi}} e^{-\frac{t_1^2}{2t_2} - \mu_1 t_1 - \mu_2 t_2} dF_{\tau(Z)}(t_1, t_2).$$

Because  $(\mu_1, \mu_2) \in \mathbb{C}^2 \mapsto \int t_1 \frac{|t_2|^{1/2}}{\sqrt{2\pi}} e^{-\frac{t_1^2}{2t_2} - \mu_1 t_1 - \mu_2 t_2} dF_{\tau(Z)}(t_1, t_2)$  is the Laplace transform of a signed measure dominated by the push-forward measure of  $\tau(Z)$ , it follows that  $\varrho$  is a Laplace transform if and only if  $(\mu_1, \mu_2) \mapsto \mu_1 \int \frac{t_2^{3/2}}{\sqrt{2\pi}} e^{-\frac{t_1^2}{2t_2} - \mu_1 t_1 - \mu_2 t_2} dF_{\tau(Z)}(t_1, t_2)$  is itself a Laplace transform on  $C + i\mathbb{R}^2$  where  $C$  is a convex of  $\mathbb{R}^2$  such that  $\mathbb{P}((X, X^2/2) \in C) = 1$ . This means that there exists a function  $g \in L^1(\tau_1(Z), \tau_2(Z))$  such that for any  $(\mu_1, \mu_2) \in C + i\mathbb{R}^2$ :

$$\int \mu_1 \frac{t_2^{3/2}}{\sqrt{2\pi}} e^{-\frac{t_1^2}{2t_2} - \mu_1 t_1 - \mu_2 t_2} dF_{\tau(Z)}(t_1, t_2) = \int g(t_1, t_2) e^{-\mu_1 t_1 - \mu_2 t_2} dF_{\tau_1(Z), \tau_2(Z)}(t_1, t_2).$$

Fubini Theorem ensures this previous equality is equivalent to:

$$\int \mu_1 \mathcal{L}[M_{t_2}](\mu_1) \exp(-\mu_2 t_2) dF_{\tau_2(Z)}(t_2) = \int \mathcal{L}[G_{t_2}](\mu_1) \exp(-\mu_2 t_2) dF_{\tau_2(Z)}(t_2),$$

with  $M_{t_2}$  and  $G_{t_2}$  measures defined by  $dM_{t_2}(u) = \frac{t_2^{3/2}}{\sqrt{2\pi}} \exp(-\frac{u^2}{2t_2}) dF_{\tau_1(Z)|\tau_2(Z)=t_2}(u)$  and  $dG_{t_2}(u) = g(u, t_2) dF_{\tau_1(Z)|\tau_2(Z)=t_2}(u)$ .

Injectivity of the Laplace transform ensures that  $\mu_1 \mathcal{L}[M_{\tau_2(Z)}](\mu_1) = \mathcal{L}[G_{\tau_2(Z)}](\mu_1)$ . It follows from Lemma 7.5 in Appendix that  $\tau_1(Z)|\tau_2(Z)$  needs to admit a locally absolutely continuous density with respect to the Lebesgue measure, which is equivalent to the fact that  $t(Z)|\sigma^2(Z)$  admits a locally absolutely continuous density ( $\sigma^2(Z)$ -almost surely). Informally this requirement is very strong when the econometrician only have one instrument  $Z$ , but quite reasonable if the econometrician observe two continuous and different instruments  $Z_1$  and  $Z_2$ . We will just discuss informally<sup>9</sup> this point.

Imagine that  $Z$  admits a density with respect to the Lebesgue measure in  $\mathbb{R}$ . In that case,

---

<sup>9</sup>We do not discuss more precisely the precise meaning of having only one or two distincts instruments, because some difficulties come from the fact that for  $Z_1, Z_2$  drawn in a uniform distribution on  $[0; 1]^2$ , it exists an invertible function  $g$  such that  $Z = g(Z_1, Z_2)$  with  $Z$  is drawn in a uniform distribution on  $[0; 1]$ .

a finite semi-parametric efficiency bound exists only if  $Z|\sigma^2(Z)$  is not countable:

$$|\{z \in \mathcal{Z} : \sigma^2(Z) = \sigma^2(z)\}| > |\mathbb{N}| \quad \sigma^2(Z) \text{ almost-surely.}$$

Informally,  $\sigma^2(Z)$  needs to be not too informative on  $t(Z)$ . It follows that when  $z \mapsto \sigma^2(z)$  is invertible or more generally admits a finite or countable fibers, regular root-N consistent estimators do not exist. And even when  $\sigma(Z)$  takes only a few values, the existence of a locally absolutely continuous density of  $t(Z)|\sigma^2(Z)$  imposes non trivial and weird restrictions on the distribution of  $t(Z)$ . To understand why, consider the case  $X|Z \sim \mathcal{N}(Z, \sigma_0^2 \mathbf{1}_{\{Z \leq a\}} + \sigma_1^2 \mathbf{1}_{\{Z > a\}})$ : a root-N consistent estimator of the average marginal effect exists only if  $Z|Z \leq a$  and  $Z|Z > a$  admit locally absolutely continuous densities. Then  $f_Z(z)/F_Z(a) \mathbf{1}_{\{z \leq a\}}$  and  $f_Z(z)/(1 - F_Z(a)) \mathbf{1}_{\{z > a\}}$  needs to vanish at the boundary of  $\mathcal{Z}$  but also at  $z = a$ . Then  $Z$  needs to admit a locally absolutely continuous density that vanishes at  $a$  and at the boundary of its support.

On the other hand,  $Z$  admits a density with respect to the Lebesgue measure on  $\mathbb{R}^2$  (or  $\mathbb{R}^p$  for  $p > 2$ ). Then  $t(Z)|\sigma^2(Z) = u$  admits a locally absolutely continuous densities under reasonable conditions because the existence of two distinct instruments enables the econometrician to observe continuous variation of  $t(Z)$  keeping constant  $\sigma^2(Z)$ . If such reasonable conditions are fulfilled, then :

$$V^* = \mathbb{V}(\varepsilon q^*(Z) + m'(X)),$$

with:

$$q^*(Z) = \frac{1}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}(1/\mathbb{E}(\varepsilon^2|Z)|t(Z), \sigma^2(Z))} \times \left( \text{Cov}(\mathbb{E}(\varepsilon m'(X)|Z), 1/\mathbb{E}(\varepsilon^2|Z)|t(Z), \sigma^2(Z)) - \frac{f'_{t(Z)|\sigma^2(Z)}(t(Z))}{f_{t(Z)|\sigma^2(Z)}(t(Z))} \right).$$

**Example: Average marginal effect for a simple parametric first stage** Consider the case where  $X|Z$  is dominated by the Lebesgue measure on  $\mathbb{R}$ , the marginal distribution of  $Z$  is such that  $\mathbb{P}(Z > 0) = 1$  and where  $f_{X|Z}(x) = g(Z) \exp(-|x|^k Z)$  with  $k > 0$ . For  $k = 1$  we get the two sided-exponential distribution for  $X|Z$  and the case  $k = 2$  corresponds to  $X|Z \sim \mathcal{N}(0, (2Z)^{-1})$ . It follows that  $g(z) = [\int \exp(-|x|^k z) dx]^{-1} = \frac{kz^{1-1/k}}{2\Gamma(1/k)}$ . In such a case  $\mu(X)$  is a scalar. If  $Z$  is such that  $\mathbb{E}(Z^{2-1/k}) < \infty$  (this is the case if  $Z$  has a compact support included in  $]0; +\infty[$ ), then the dominated convergence Theorem ensures that  $f_X$  is differentiable and  $f'_X(x) = k|x|^{k-1} \text{sgn}(x) \int z g(z) \exp(-|x|^k z) dF_Z(z)$ , where  $\text{sgn}(x) = \mathbf{1}_{\{x > 0\}} - \mathbf{1}_{\{x < 0\}}$ . To satisfy Condition (i) of Theorem 5.3,  $f'_X(x)/f_X(x)$  needs to be a function of  $\mu(x) = |x|^k$ . Because  $f'_X(x)/f_X(x) = k\mu(x)^{1-1/k} \text{sgn}(x) \frac{\int z g(z) \exp(-\mu(x)z) dF_Z(z)}{\int g(z) \exp(-\mu(x)z) dF_Z(z)}$ , this condition holds if and only if  $\text{sgn}(x)$  is a function of  $|x| = \mu(x)^{1/k}$ , which never holds.

### 5.2.3 Specialization to the additive first stage

We now consider the case where the first stage is additive and where both instruments and residuals of the first stage admit densities with respect to the Lebesgue measure. In the following, for any  $p \geq 0$ ,  $L^p(\mathbb{R}^K)$  denotes the set of functions  $f$  which are measurable with respect to the Lebesgue measure on  $\mathbb{R}^K$  and such that  $f^p$  is integrable.

#### Assumption 5 (Additive 1st stage)

$$X = p(Z) + \eta$$

where  $\mathcal{R}(p) \subset \mathbb{R}^K$ ,  $p(Z)$  and  $\eta$  are random variables in  $\mathbb{R}^K$  that admit densities  $f_{p(Z)}$  and  $f_\eta$  with respect to the Lebesgue measure on  $\mathbb{R}^K$  and  $Z \perp\!\!\!\perp \eta$ .

The previous Assumption rules out the case where there exists non continuous variables in  $X^0$ . This restriction is made for the sake of simplicity but can be relaxed by a reasoning conditional to  $X^0$ . For every  $f$  integrable with respect to the Lebesgue measure on  $\mathbb{R}^d$ ,  $\mathcal{F}_q$  denotes the Fourier transform of  $q$ :  $\int q(u) \exp(-2\pi i x' u) du$ , and  $\overline{\mathcal{F}}_q(x) = \mathcal{F}_q(-x)$ . Note that under the previous Assumption,  $X \perp\!\!\!\perp Z | p(Z)$  and then  $\mathbb{E}(q(Z) | X) = \mathbb{E}(\mathbb{E}(q(Z) | p(Z)) | X) = \mathbb{E}(\tilde{q}(p(Z)) | X)$ . And the additivity of the first stage enables us to rewrite  $T^*(q)f_X$  as an convolution product because  $T^*(q)(x)f_X(x) = \int \tilde{q}(v)f_{p(Z)}(v)f_\eta(x-v)dv = [(\tilde{q}f_{p(Z)}) \star f_\eta](x)$ . The convolution Theorem ensures that  $\mathcal{Q} \neq \emptyset$  if and only if  $\mathcal{F}_{[\psi f_X]^{(v)}/\mathcal{F}_{f_\eta}}$  is the Fourier transform of an integrable function (more precisely the equality holds for any  $v$  such that  $\mathcal{F}_{f_\eta}(v) \neq 0$ ). Let  $\mathcal{S}$ , the Schwartz space (i.e. the space of infinitely differentiable functions  $f$  such that  $x^\alpha f^{(\beta)}(x)$  decreases to zero when  $|x| \rightarrow +\infty$  for any multi-index  $(\alpha)$  and  $(\beta)$ ). Because  $\mathcal{S} \subset L^1(\mathbb{R}^K)$ ,  $\mathcal{F}$  is well defined on  $\mathcal{S}$ . By duality arguments,  $\mathcal{F}$  can be extended to  $\mathcal{S}'$  the topological dual space of  $\mathcal{S}$  (see Schwartz (1997) for the topology considered) that contains  $L^1(\mathbb{R}^K)$ ,  $\mathcal{S}$  and  $L^2(\mathbb{R}^K)$  (up to a canonical isomorphism). On  $\mathcal{S}'$ , we have  $\overline{\mathcal{F}}\mathcal{F} = \mathcal{F}\overline{\mathcal{F}} = id$ .  $\mathcal{F}$  is a bijection from  $\mathcal{S}'$  into itself. This is also a bijection from  $\mathcal{S}$  (respectively  $L^2(\mathbb{R}^K)$ ) to itself. Then despite that  $\mathcal{F}(\mathcal{S})$  (respectively  $\mathcal{F}(L^2(\mathbb{R}^K))$  and  $\mathcal{F}(\mathcal{S}')$ ) is easily characterizable, there isn't also simple result for  $\mathcal{F}(L^1(\mathbb{R}^K))$ . But we know that  $\mathcal{F}(L^1(\mathbb{R}^K))$  is included in  $\mathcal{C}_0^0(\mathbb{R}^K)$ , the space of continuous functions from  $\mathbb{R}^K$  to  $\mathbb{C}$  vanishing at infinity. And because  $\overline{\mathcal{F}}\mathcal{F} = id$  on  $\mathcal{S}'$  and then on  $L^1(\mathbb{R}^K)$ , we also have  $\{f \in L^1(\mathbb{R}^K) \text{ s.t. } \mathcal{F}_f \in L^1(\mathbb{R}^K)\} = \{f \in L^1(\mathbb{R}^K) \text{ s.t. } \overline{\mathcal{F}}_f \in L^1(\mathbb{R}^K)\} \subset \mathcal{F}(L^1(\mathbb{R}^K)) \cap \mathcal{C}_0^0(\mathbb{R}^K)$ . Moreover, the Laplace transform defined by  $\mathcal{L}_f(z) = \int f(u) \exp(-z'u) du$  extends the Fourier transform<sup>10</sup> to  $\Gamma_f + i\mathbb{R}^d$ , where  $\Gamma_f = \{x \in \mathbb{R}^K : \int f(u) \exp(-u'x) \text{ integrable}\}$ . When the support of  $p(Z)$  is compact, then we know that the Laplace transform of  $\tilde{q}f_{p(Z)}$  is defined on  $\mathbb{C}^p$  and the Paley-Wiener Theorem gives us a sharp characterization of the

<sup>10</sup>More exactly, with the definition of  $\mathcal{F}$  and  $\mathcal{L}$  used here,  $\mathcal{L}_f(x + iy)$  extends  $iy \mapsto \mathcal{F}_f(-iy/2\pi)$  from  $i\mathbb{R}^K$  to  $\Gamma_f + i\mathbb{R}^K$ .

range of  $\mathcal{L}_{\tilde{q}f_{p(Z)}}$ . So in the next Theorem, we consider a general case and a special case where  $\text{Supp}(p(Z))$  is compact.

**Theorem 5.4**

If Assumptions 1, 2 and 5 hold and if  $\mathcal{F}_{f_\eta}$  has isolated zeros then:

- (i)  $\mathcal{Q} \neq \emptyset$  if and only if it exists a function  $q \in L^2(p(Z))$  such that  $\mathcal{F}_{[\psi_{f_X}]^{(j)}}/\mathcal{F}_{f_\eta} = (-1)^{|j|} \mathcal{F}_{qf_{p(Z)}}$ .
- (ii)  $\mathcal{Q} \neq \emptyset$  only if  $\mathcal{F}_{[\psi_{f_X}]^{(j)}}/\mathcal{F}_{f_\eta} \in \mathcal{C}_0^0(\mathbb{R}^K)$ .
- (iii) If  $\mathcal{F}_{[\psi_{f_X}]^{(j)}}/\mathcal{F}_{f_\eta} \in \mathcal{C}_0^0(\mathbb{R}^K) \cap L^1(\mathbb{R}^K)$  then  $\mathcal{Q} \neq \emptyset$  if and only if  $\mathcal{F}(\mathcal{F}_{[\psi_{f_X}]^{(j)}}/\mathcal{F}_{f_\eta}) \in L^1(\mathbb{R}^K)$ . And in that case,

$$\mathcal{Q} = \left\{ q \in L^1(Z) : \mathbb{E}(q(Z)|p(Z) = v)f_{p(Z)}(v) = (-1)^{|j|} \overline{\mathcal{F}}(\mathcal{F}_{[\psi_{f_X}]^{(j)}}/\mathcal{F}_{f_\eta}) \right\}.$$

Moreover, if the support of  $p(Z)$  is compact and  $\int f_\eta(x)e^{-t'x}dx < \infty$  for every  $t \in \mathbb{R}^K$  then:

- (iv)  $\mathcal{Q} \neq \emptyset$  only if:
  - (a)  $\mathcal{L}_{[\psi_{f_X}]^{(j)}}/\mathcal{L}_{f_\eta}$  is an holomorphic function on  $\mathbb{C}^K$  bounded on  $\Gamma + i\mathbb{R}^K$  for every compact  $\Gamma$  included in  $\mathbb{R}^K$ ,
  - (b)  $\left[ \mathcal{L}_{[\psi_{f_X}]^{(j)}}/\mathcal{L}_{f_\eta} \right](x + iy)$  tends to zero when  $|y|$  tends to infinity for any  $x \in \mathbb{R}^K$ ,
  - (c) there exists  $C, N$  such that:

$$\left\| \frac{\mathcal{L}_{[\psi_{f_X}]^{(j)}}(x + iy)}{\mathcal{L}_{f_\eta}(x + iy)} \right\| \leq C(1 + \|x + iy\|)^N \exp \left( \sup_{z \in \text{Supp}(Z)} \|p(z)\| \cdot \|y\| \right).$$

When a root- $N$  regular consistent estimator of  $\theta_0$  exists, the semiparametric efficiency bound of  $\theta_0$  is:

$$V^* = \mathbb{V}(\varepsilon q^*(Z) + R(X)),$$

with

$$q^*(Z) = \frac{(-1)^{|j|}}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)]} \frac{d\mathcal{F}^{-1}[\mathcal{F}_{(\psi_{f_X})^{(j)}}/\mathcal{F}_{f_\eta}]}{dF_{p(Z)}}(p(Z)) + \frac{\mathbb{E}[\mathbb{E}(\varepsilon R(X)|Z)/\mathbb{E}(\varepsilon^2|Z)|p(Z)]}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)]} - \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)}$$

and

$$R(X) = (\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X) - \mathbb{E}((\psi(X) + \partial_{f_X}\psi(X)f_X(X))m^{(j)}(X)).$$

The first part of the previous Theorem enables us to distinguish three cases. First, if  $\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}} \notin \mathcal{C}_0^0(\mathbb{R}^K)$ , it ensures that  $\mathcal{Q} = \emptyset$ . Second if  $\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}} \in \mathcal{C}_0^0(\mathbb{R}^K) \cap L^1(\mathbb{R}^K)$  and  $\mathcal{F}\left(\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}}\right) \in L^1(\mathbb{R}^K)$ , then  $\mathcal{Q} \neq \emptyset$ . Otherwise, if  $\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}} \in \mathcal{C}_0^0(\mathbb{R}^K) \setminus L^1(\mathbb{R}^K)$  or  $\mathcal{F}\left(\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}}\right) \notin L^1(\mathbb{R}^K)$ , we remain agnostic about the emptiness of  $\mathcal{Q}$ , we only know that there is a linear form  $S \in \mathcal{S}'$  with a Fourier transform<sup>11</sup> equals to  $\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}}$  (up to a canonical isomorphism) but this linear form does not necessarily correspond to a  $q \in L^1(Z)$  (in the sense that  $S \in \{U \in \mathcal{S}' \text{ s.t. } \exists q \in L^1(Z) \text{ s.t. } U(\varphi) = (-1)^{|j|} \int \mathbb{E}(q(Z)|p(Z) = v) f_{p(Z)}(v) \varphi(v) dv \text{ for } \varphi \in \mathcal{S}\}$ ). In this last case, it is always possible to consider  $\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}}$  as an element of  $\mathcal{S}'$ , and to verify<sup>12</sup> that  $S = \overline{\mathcal{F}}\left(\mathcal{F}_{[\psi_{f_X}]^{(j)}/\mathcal{F}_{f_\eta}}\right)$  is in  $L^1(\mathbb{R}^K)$  (up to the usual isomorphism).

The second part of the previous Theorem considers the case where the support of  $p(Z)$  is bounded and where  $\eta$  have thin tails (such as the normal distribution or distribution with bounded support). In particular, when  $X$  has a compact support then  $p(Z)$  and  $\eta$  also have compact supports and then Proposition (iii) of the previous theorem gives a criterion on the distribution of  $\eta$  to ensure that  $\mathcal{Q} \neq \emptyset$ .

**Example: Average under counterfactual distribution.** If  $Z, \eta$  are two independent normal centered distributions of variance  $\sigma_Z^2$  and  $\sigma_\eta^2$ . In that case if,  $X = Z + \eta$  then  $X \sim \mathcal{N}(0, \sigma_Z^2 + \sigma_\eta^2)$ . If we are interested to average  $m(X)$  under the counterfactual distribution  $\mathcal{N}(\mu, \sigma^2)$ , then  $\psi(x) = \frac{\sqrt{\sigma_Z^2 + \sigma_\eta^2}}{\sigma} \exp\left(-\left(\frac{(x - \mu)^2}{\sigma^2} - \frac{x^2}{(\sigma_Z^2 + \sigma_\eta^2)}\right)/2\right)$ . In such a case,  $\mathcal{L}_{\psi_{f_X}}(z) = \exp(\mu z + \sigma^2 z^2/2)$  and  $\mathcal{L}_{f_\eta}(z) = \exp(\sigma_\eta^2 z^2/2)$  with  $\Gamma_{\psi_{f_X}} = \Gamma_{f_\eta} = \mathbb{R}$ , then the ratio  $\mathcal{L}_{\psi_{f_X}}(z)/\mathcal{L}_{f_\eta}(z)$  is equal to  $\exp(\mu z + \sigma^2 z^2/2 - \sigma_\eta^2 z^2/2)$  and verifies condition (i) of the Theorem 5.4 if and only if  $\sigma^2 > \sigma_\eta^2$ . In such a case,  $\mathcal{Q}$  contains the single element  $q(z) = \frac{\sigma z}{\sqrt{\sigma^2 - \sigma_\eta^2}} \exp\left(\frac{z^2}{2\sigma^2} - \frac{(z - m)^2}{2\sqrt{\sigma^2 - \sigma_\eta^2}}\right)$ . The condition  $\sigma^2 > \sigma_\eta^2$  means that the counterfactual distribution cannot be concentrated more than the noise in the first stage.

We recover a result already established with Theorem 5.3. Interesting results can also be obtained for additive first stage which does not verify Assumption 4. For instance, if  $Z, \eta$  and  $X$  are Gamma distributions of parameter  $(k_Z, \tau)$ ,  $(k_\eta, \tau)$  and  $(k_Z + k_\eta, \tau)$ . If the counterfactual distribution is a Gamma distribution of parameter  $(k', \tau')$ , then the ratio  $\mathcal{L}_{\psi_{f_X}}(z)/\mathcal{L}_{f_\eta}(z)$  is equal to  $(1 + \tau_\eta z)^{k_\eta} (1 + \tau' z)^{-k'}$  and then Condition (ii) of Theorem 5.4

<sup>11</sup>See Schwartz (1997) for a definition of Fourier transform of linear form.

<sup>12</sup>To do so, verify that for every compact  $K$ , if  $\varphi_n$  is a sequence of functions infinitely differentiable with a support included in  $K$  that tends uniformly to 0 then  $S(\varphi_n)$  tends to 0. This first step ensures that  $S$  is identifiable to a measure (cf. Theorem III, Chapter I of Schwartz (1997)). To be identifiable to an integrable function, such measure have also to be dominated by the Lebesgue measure and of finite total variation by the Radon-Nikodim Theorem. This leads to check also that  $\sup_{\varphi \in \mathcal{S}, \|\varphi\|_\infty < 1} S(\varphi) < \infty$  and  $\sup_{A \text{ negligible Borel set}} \inf_{\varphi \in \mathcal{S}, K \text{ compact s.t. } \varphi \geq \mathbf{1}_{A \cap K}} S(\varphi) = 0$ .

holds if and only if  $k' > k_\eta$ . Because the value of  $\theta$  does not matter, it appears that in this last case the mean and the variance of the counterfactual distribution can be chosen arbitrarily small (because the both depends on  $\theta$ ) but its skewness and its kurtosis (the both depend only on  $k$ ) have to be larger than those of  $\eta$ .

The case of average derivatives leads to elegant results, as illustrated in the following proposition.

**Proposition 5.5 (Average derivatives under additive first stage)**

If Assumption 1, 2 and 5 hold, if  $\mathcal{F}_{f_\eta}$  has isolated zero and if  $\theta_0 = \mathbb{E}(m^{(j)}(X))$  then a regular and root- $N$  consistent estimator of  $\theta_0$  exists only if the weak derivative of  $f_{p(Z)}$  is a locally integrable function (denoted  $f_{p(Z)}^{(j)}$  in the sequel) and if  $\mathbb{E}\left(\left(\frac{f_{p(Z)}^{(j)}(p(Z))}{f_{p(Z)}(p(Z))}\right)^2\right) < \infty$ . In that case, the semi-parametric efficiency bound is:

$$\mathbb{V}(\varepsilon q^*(Z) + m^{(j)}(X)),$$

with

$$q^*(Z) = \frac{(-1)^{|j|} f_{p(Z)}^{(j)}(p(Z))}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)] f_{p(Z)}} + \frac{\mathbb{E}[\mathbb{E}(\varepsilon m^{(j)}(X)|Z)/\mathbb{E}(\varepsilon^2|Z)|p(Z)]}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)]} - \frac{\mathbb{E}(\varepsilon m^{(j)}(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)}.$$

and when  $p$  is one to one this expression reduced to:

$$q^*(Z) = (-1)^{|j|} \frac{f_{p(Z)}^{(j)}(p(Z))}{f_{p(Z)}(p(Z))}.$$

The proof rely on the Condition (i) of Theorem 5.4 and is given in Appendix. We briefly illustrate the implication of the previous proposition for the average marginal effect.

**Example: Average marginal effect.** Let  $X = Z + \eta$  with  $\eta$  drawn in an uniform distribution on  $[0; 1]$ ,  $Z$  drawn in an uniform distribution  $[0; \alpha]$  and  $\eta \perp\!\!\!\perp Z$ . The correlation between  $X$  and  $Z$  is  $\left(\frac{\alpha^2}{\alpha^2+1}\right)^{1/2}$ , then  $\alpha$  is a measure of the strength of the instrument. In such cases, we have shown that  $\mathbb{E}(m'(X))$  is always identified (cf. Proposition 3.2). But the previous proposition ensures that  $\mathbb{E}(m'(X))$  is not root- $N$  estimable whatever the value of  $\alpha$ . Indeed  $f_Z$  is not absolutely continuous on  $\mathbb{R}$  and its weak derivative is the signed measure  $\delta_0 - \delta_1$  (where  $\delta_a$  is the Dirac at  $a$ ) and not a function. Such an example shows that if the densities of  $Z$  and  $\eta$  are not "sufficiently" regular, root- $N$  estimation of the average marginal effect is lost even in case of identification and in presence of strong instruments. A subsequent question concerns the respective minimal regularity of

$f_Z$  and  $f_\eta$ . As soon as  $\mathcal{F}_{f'_X}(t) = (2i\pi t)\mathcal{F}_{f_X}(t)$ , then  $\mathbb{E}(m'(X))$  is root-N estimable only if  $Z$  admits an absolutely-continuous density  $f_Z$  on  $\mathbb{R}$  such that  $\int \frac{f'_Z(z)}{f_Z(z)} dz < \infty$ . Note that  $f_Z$  must at least vanish at the boundary of its support as it is often assumed and this vanishing needs to be sufficiently smooth to ensure the integrability condition. If  $Z \sim B(\alpha, \beta)$ ,  $\alpha > 2$  and  $\beta > 2$  is a necessary condition for the existence of a consistent root-N estimator. And similarly, if  $Z \sim \Gamma(k, \tau)$  or if  $Z \sim Weibull(k, \lambda)$  then  $k > 2$  is a necessary condition for the existence of a consistent root-N estimator. With Pareto distribution root-N consistent estimators never exist. But normal or Student distributions for  $Z$  ensure that  $\mathcal{Q}$  is not empty as soon as  $f_X$  is absolutely continuous. In that case, the semiparametric efficiency bound is  $\mathbb{V} \left[ m'(x) - (Y - m(x)) \frac{f'_Z(Z)}{f_Z(Z)} \right]$  and the parameter  $\theta_0$  is also equal to  $-\mathbb{E} \left( Y \frac{f'_Z(Z)}{f_Z(Z)} \right)$ . More generally for multivariate  $X$ , if Assumptions 2 and 5 hold and if the Fourier transform of  $f_\eta$  has isolated zeros, then a regular and root-N consistent estimator of  $\mathbb{E}(\partial_{x_k} m(X))$  exists only if  $p_k \mapsto f_{p(Z)}(p_1, \dots, p_K)$  is absolutely continuous (for almost-every  $p_1, \dots, p_{k-1}, p_{k+1}, \dots, p_K$ ) and if  $\mathbb{E} \left( \frac{(\partial_{p_k} f_{p(Z)})^2}{f_{p(Z)}^2}(p(Z)) \right) < \infty$ . In that case, the semiparametric efficiency bound is  $\mathbb{V}(\varepsilon q^*(Z) + \partial_{x_k} m(X))$  with

$$q^*(Z) = \frac{-1}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)]} \frac{\partial_{p_k} f_{p(Z)}}{f_{p(Z)}}(p(Z)) + \frac{\mathbb{E}[\mathbb{E}(\varepsilon \partial_{x_k} m(X)|Z)/\mathbb{E}(\varepsilon^2|Z)|p(Z)]}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E}[1/\mathbb{E}(\varepsilon^2|Z)|p(Z)]} - \frac{\mathbb{E}(\varepsilon \partial_{x_k} m(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)}.$$

## 6 Conclusion

In this paper we show that the parameter  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$  with  $m$  unknown and  $\mathbb{E}(Y - m(X)|Z) = 0$  is often not root-N estimable. We show that root-N estimability of  $\theta_0$  is closely related to the fact that  $[\psi f_X]^{(j)}/f_X$  belongs the range of  $T^*$ , with  $T^*(q)(X) = \mathbb{E}(q(Z)|X)$ . This condition is not nested with the completeness condition that ensures the identification of  $m$ . If the first stage is exponential or additive, our condition is quite restrictive when the parameter of interest is an average under a counterfactual distribution of  $X$ . It appears that for some DGPs and some counterfactual distributions,  $\theta_0$  is never root-N estimable, and sometimes for others,  $\theta_0$  is root-N estimable if and only if the strength of the instruments is above a critical value. When the parameter of interest is an average marginal effect, root-N estimation is achievable if the first stage is additive, even if the completeness condition does not hold, as soon as the density of the instrument admits a derivative that vanishes at the boundary of its support.

$\theta_0$	$\rho$	$\sqrt{N}$ -estimable	efficiency bound
Shift in Mean			
$\mathbb{E} \left( \exp \left( -\frac{1}{2} [(x - m)^2 - x^2] \right) m(X) \right)$	$\rho \neq 0$	yes	$\mathbb{V} \left( (Y - m(X)) \exp \left( \frac{m}{\rho} Z - \frac{m^2}{2\rho^2} \right) \right)$
Shift in Variance			
$1/\sqrt{\sigma^2} \mathbb{E} \left( \exp \left( -\frac{1}{2} [x^2(1/\sigma^2 - 1)] \right) m(X) \right)$	$\rho^2 \leq 1 - \sigma^2$	no	
	$\rho^2 > 1 - \sigma^2$	yes	$\frac{\rho^2}{\sigma^2 + \rho^2 - 1} \mathbb{V} \left[ (Y - m(X)) \exp \left( \frac{\rho^2 z^2}{2(1 - \rho^2)} \left[ 1 - \frac{\sigma^2 \rho^2}{\sigma^2 + \rho^2 - 1} \right] \right) \right]$
Average marginal effect			
$\mathbb{E}(m'(x))$	$\rho^2 = 0$	no	
	$\rho^2 > 0$	yes	$\mathbb{V} \left( \varepsilon \frac{Z}{\rho} + m'(X) \right)$
Density-weighted average derivative, ...			
$\mathbb{E} \left( (2\pi)^{-k/2} \exp(-kX^2/2) m'(X) \right), \quad k \in \mathbb{N}$	$\rho^2 < \frac{k}{k+1}$ or $\rho^2 = 0$	no	
	$\rho^2 = \frac{k}{k+1}$ with $k \geq 1$	yes	$(k+1)^2 \mathbb{V} \left( \varepsilon \frac{Z}{\rho} + \frac{1}{\sqrt{2\pi}^k} \exp(-kX^2/2) m'(X) \right)$
	$\rho^2 > \frac{k}{k+1}$	yes	$(k+1)^2 \mathbb{V} \left( \varepsilon Z \exp \left( Z^2 \frac{1 - \rho^2}{(k+1)\rho^2 - k} \right) + \frac{1}{\sqrt{2\pi}^k} \exp(-kX^2/2) m'(X) \right)$

Table 1: Some examples when  $(X, Z) \sim \mathcal{N}(0, \Sigma)$  with  $\Sigma_{11} = \Sigma_{22} = 1$  and  $\Sigma_{12} = \rho$



## 7 Appendix A: Proofs of theorems

### 7.1 Proof of Theorem 3.1

Without completeness assumption,  $T$  is not invertible,  $m$  is not point identified and then the set identification is

$$\{m + l, l \in \text{Ker}(T) \cap (\mathcal{M} - \{m\})\} = \{m\} + \text{Ker}(T) \cap (\mathcal{M} - \{m\}).$$

It follows that the set identification of  $m^{(j)}$  is

$$\{m^{(j)} + l^{(j)}, l \in \text{Ker}(T) \cap (\mathcal{M} - \{m\})\}.$$

Because  $\psi$  is known (up to  $f_X$  which is identified), the set identification of  $\theta_0$  is

$$\{\theta_0 + \mathbb{E}(\psi l^{(j)}), l \in \text{Ker}(T) \cap (\mathcal{M} - \{m\})\}.$$

When  $\mathcal{M}$  is a vectorial space,  $\mathcal{M} - \{m\} = \mathcal{M}$  and because  $l \mapsto \mathbb{E}(\psi l^{(j)})$  is a linear form on vectorial space  $\text{Ker}(T) \cap \mathcal{M}$  and its range is  $\{0\}$  or  $\mathbb{R}$  and this proves (i).

To prove (ii), remark that  $\theta_0$  is identified if and only if  $\mathbb{E}(\psi l^{(j)}) = 0$  for every  $l \in \text{Ker}(T) \cap (\mathcal{M} - \{m\})$ , i.e.

$$\psi \in (l^{(j)} \text{ s.t. } l \in \text{Ker}(T) \cap (\mathcal{M} - \{m\}))^\perp.$$

### 7.2 Proof of Theorem 3.2

We first prove the result under the first condition. The region of identification of  $\theta_0$  is  $\{\mathbb{E}(m'(X) + l'(X)), l \in \mathcal{M} \cap \text{Ker}(T)\}$ . For all  $l \in \mathcal{M} \cap \text{Ker}(T)$ , we have both equalities:

$$\mathbb{E}(l(X)|p(Z) = u) = 0$$

$$\partial_u \mathbb{E}(l(X)|p(Z) = u) = \mathbb{E}(l'(X)|p(Z) = u)$$

The second equality is a simple consequence of the dominated convergence Theorem using  $\sum_{s=1}^S a_s x^s$  for the domination. All together, these equalities imply that  $\mathbb{E}(l'(X)) = 0$ , and next the region of identification is reduced to a single point.

Under the second condition, we have:

$$\mathbb{E}(l(X)|p(Z) = p) = \int_{\inf(I)}^{\sup(I)} \int_{\text{Supp}(\nu)} l(p+u+v) dF_\nu(v) du = \int_{\inf(I)+p}^{\sup(I)+p} \int_{\text{Supp}(\nu)} l(u+v) dF_\nu(v) du,$$

derivation with respect to  $p$  gives:  $\int_{\text{Supp}(\nu)} l(\sup(I) + p + v) dF_\nu(v) - \int_{\text{Supp}(\nu)} l(\inf(I) + p + v) dF_\nu(v) = \int_{\text{Supp}(\nu)} \int_{\inf(I)}^{\sup(I)} l'(p + u + v) du dF_\nu = \mathbb{E}(l'(X)|p(Z) = p)$ , And next,  $\theta_0$  is identified.

### 7.3 Proof of Theorem 4.1

#### 7.3.1 Some Lemmas

#### Lemma 7.1 (Derivation in quadratic mean "with respect to $m(X)$ ")

If Assumption 1 and 2 holds then  $r(e, x, z) = \partial_e f_{\varepsilon, X, Z}(e, x, z) / f_{\varepsilon, X, Z}(e, x, z) \mathbb{1}\{f_{\varepsilon, X, Z}(e, x, z) > 0\}$  is well defined and square integrable, and for any  $l \in L^2(X)$  such that  $rl \in L^2(\varepsilon, X, Z)$ , we have:

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} \left( f_{\varepsilon, X, Z}^{1/2}(e - tl(x), x, z) - f_{\varepsilon, X, Z}^{1/2}(e, x, z) \right) - \frac{l(x)}{2} r(e, x, z) f_{\varepsilon, X, Z}^{1/2}(e, x, z) \right)^2 d\nu(e, x, z) = 0$$

Moreover,  $\mathbb{E}(r|X, Z) = 0$ .

Proof: Let  $s_u(e, x, z) = f_{\varepsilon, X, Z}^{1/2}(e - u, x, z)$ , because  $x \mapsto x^2$  is locally Lipschitz continuous on  $\mathbb{R}^+$ ,  $u \mapsto s_u^2(y, x, z)$  is absolutely continuous and the chain rule applies: i.e.  $\partial_u s_u^2(e, x, z) = 2s_u(e, x, z) \partial_u s_u(e, x, z)$ , and next  $r(e - u, x, z) = \partial_u s_u^2(e, x, z) / s_u^2(e, x, z) \mathbb{1}_{s_u(e, x, z) > 0} = 2\partial_u s_u(e, x, z) / s_u(e, x, z)$  is well defined. The square integrability of  $r$  is ensured by Assumption 2.(vii). Next, we have:

$$\begin{aligned} s_{tl(x)}(e, x, z) - s_0(e, x, z) &= \int_0^{tl(x)} \partial_u s_u(e, x, z) d\lambda(u) \\ &= \frac{1}{2} \int_0^{tl(x)} r(e - u, x, z) f_{\varepsilon, X, Z}^{1/2}(e - u, x, z) d\lambda(u) \\ &= \frac{tl(x)}{2} \int_0^1 r(e - utl(x), x, z) f_{\varepsilon, X, Z}^{1/2}(e - utl(x), x, z) d\lambda(u) \end{aligned}$$

Next, the Cauchy-Schwartz inequality and the Fubini Theorem ensure that:

$$\begin{aligned} \frac{1}{t^2} \mathbb{E} \left[ (s_{tl(X)} - s_0)^2 \right] &= \frac{1}{4} \mathbb{E} \left[ l^2(X) \left( \int_0^1 r(e - utl(X), X, Z) f_{\varepsilon, X, Z}^{1/2}(e - utl(X), X, Z) d\lambda(u) \right)^2 \right] \\ &\leq \frac{1}{4} \mathbb{E} \left[ l^2(X) \left( \int_0^1 r^2(\varepsilon - utl(X), X, Z) f_{\varepsilon, X, Z}(\varepsilon - utl(X), X, Z) d\lambda(u) \right) \right] \\ &= \frac{1}{4} \int_0^1 \int_{\text{Supp}(X, Z)} \left( \int_{\mathbb{R}} r^2(e - utl(x), x, z) f_{\varepsilon, X, Z}(e - utl(x), x, z) d\lambda(e) \right) l^2(X) d\mu(x, z) d\lambda(u) \\ &= \frac{1}{4} \int_0^1 \int_{\text{Supp}(X, Z)} \left( \int_{\mathbb{R}} r^2(e, x, z) f_{\varepsilon, X, Z}(e, x, z) d\lambda(e) \right) l^2(X) d\mu(x, z) d\lambda(u) \\ &= \frac{1}{4} \mathbb{E} (r^2(\varepsilon, X, Z) l^2(X)) < \infty \end{aligned}$$

$X, Z$ -almost-surely,  $t \mapsto f_{\varepsilon, X, Z}^{1/2}(e - t, X, Z)$  is absolutely continuous and then admits a

derivative at  $t = 0$  for almost-every  $e \in \mathbb{R}$ , then:

$$\left( \frac{1}{t} (s_{U(X)}(\varepsilon, X, Z) - s_0(\varepsilon, X, Z)) - \frac{l(X)}{2} r(\varepsilon, X, Z) s_0(\varepsilon, X, Z) \right)^2 \rightarrow 0 \quad a.s.$$

The Proposition 2.29 of van der Vaart (2000) ensures the nullity of the limit given in the Lemma. Next, consider the case where  $l(X) = 1$  and use the fact that  $\nu(e, x, z) = \lambda(e) \otimes \nu(x, z)$  to deduce that  $X, Z$ -almost surely:

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} (f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e)) - \frac{1}{2} r(e, X, Z) f_{\varepsilon|X,Z}^{1/2}(e) \right)^2 d\lambda(e) = 0.$$

Next, bound  $\mathbb{E}(r|X, Z)$ :

$$\begin{aligned} |\mathbb{E}(r|X, Z)| &= \left| \int \frac{1}{t} (f_{\varepsilon|X,Z}(e-t) - f_{\varepsilon|X,Z}(e)) - r(e, X, Z) f_{\varepsilon|X,Z}(e) d\lambda(e) \right| \\ &\leq \left| \int \left[ \frac{1}{t} (f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e)) - \frac{r(e, X, Z)}{2} f_{\varepsilon|X,Z}^{1/2}(e) \right] \left[ f_{\varepsilon|X,Z}^{1/2}(e-t) + f_{\varepsilon|X,Z}^{1/2}(e) \right] d\lambda(e) \right| \\ &\quad + \left| \frac{1}{2} \int r(e, X, Z) f_{\varepsilon|X,Z}^{1/2}(e) \left[ f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e) \right] d\lambda(e) \right| \\ &\leq 2 \left( \int \left[ \frac{1}{t} (f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e)) - \frac{1}{2} r(e, X, Z) f_{\varepsilon|X,Z}^{1/2}(e) \right]^2 d\lambda(e) \right)^{1/2} \\ &\quad + \frac{1}{2} \mathbb{E}(r^2(\varepsilon, X, Z)|X, Z) \left[ \int (f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e))^2 d\lambda(e) \right]^{1/2} \end{aligned}$$

Theorem 9.5 of Rudin (1987) ensures that  $\lim_{t \rightarrow 0} \int (f_{\varepsilon|X,Z}^{1/2}(e-t) - f_{\varepsilon|X,Z}^{1/2}(e))^2 d\lambda(e) = 0$ .

**Lemma 7.2 (DQM "with respect to the nuisance parameter  $f_{\varepsilon, X, Z}$ ")** *If Assumption 2 holds, and if  $f_t(e, x, z)$  is a density with respect to  $\nu$  and  $s \in L^2(\varepsilon, X, Z)$  such that:*

$$\lim_{t \rightarrow 0} \frac{1}{t^2} \int f_t(e, x, z) \mathbb{1}_{\{f_{\varepsilon, X, Z}(e, x, z) = 0\}} d\nu(e, x, z) = 0,$$

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} (f_t^{1/2}(e, x, z) - f_{\varepsilon, X, Z}^{1/2}(e, x, z)) - \frac{1}{2} s(e, x, z) f_{\varepsilon, X, Z}^{1/2}(e, x, z) \right)^2 d\nu(e, x, z) = 0$$

then for any  $\varphi \in \Phi$ :

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} (f_t^{1/2}(e - t\varphi(x), x, z) - f_{\varepsilon, X, Z}^{1/2}(e, x, z)) - \frac{s - r\varphi}{2} f_{\varepsilon, X, Z}^{1/2}(e, x, z) \right)^2 d\nu(e, x, z) = 0$$

Let  $W = (X, Z)$  and  $w = (x, z)$ .

Assumptions imply that  $\mathbb{E}(s^2(\varepsilon, W)) = \int s^2(e, w) f_{\varepsilon, W}(e, w) d\nu(e, w) < \infty$ .

Moreover, we have:

$$\begin{aligned}
& \int \left( \frac{1}{t} \left( f_t^{1/2}(e - t\varphi(x), w) - f_{\varepsilon, W}^{1/2}(e, w) \right) - \frac{1}{2}(s(e, w) + r(e, w)\varphi(x))f_{\varepsilon, W}^{1/2}(e, w) \right)^2 d\nu(e, w) \\
& \leq 2 \int \left( \frac{1}{t} \left( f_t^{1/2}(e - t\varphi(x), w) - f_{\varepsilon, W}^{1/2}(e - t\varphi(x), w) \right) - \frac{1}{2}s(e - t\varphi(x), w)f_{\varepsilon, W}^{1/2}(e - t\varphi(x), w) \right)^2 d\nu(e, w) \\
& + 2 \int \left( \frac{1}{t} \left( f_{\varepsilon, W}^{1/2}(e - t\varphi(x), w) - f_{\varepsilon, W}^{1/2}(e, w) \right) - \frac{1}{2}r(e, w)\varphi(x)f_{\varepsilon, W}^{1/2}(e, w) \right)^2 d\nu(e, w) \\
& + 2 \int \left( s(e - t\varphi(x), w)f_{\varepsilon, W}^{1/2}(e - t\varphi(x), w) - s(e, w)f_{\varepsilon, W}^{1/2}(e, w) \right)^2 d\nu(e, x, z)
\end{aligned}$$

The first and the third terms tend to zero because  $\nu(e, x, z) = \lambda(e) \otimes (x, z)$  and the Lebesgue measure  $\lambda$  is translation invariant. The second term tends to zero by assumption.

**Lemma 7.3 (Density of bounded scores)** *Let*

$$\mathcal{S} = \{s \in L^2(\varepsilon, X, Z) : \mathbb{E}(s) = 0, \mathbb{E}(\varepsilon s | Z) = 0\},$$

*Under Assumption 2,  $\mathcal{S}^b = \mathcal{S} \cap L^\infty(\varepsilon, X, X)$  is dense in  $\mathcal{S}$  for the norm of  $L^2(\varepsilon, X, Z)$ .*

Consider  $s \in \mathcal{S}$  and  $s_n = s \mathbf{1}_{\{|s| < n\}} - \mathbb{E}(s \mathbf{1}_{\{|s| < n\}})$ . We have  $s_n \in L^{20}(\varepsilon, X, Z) \cap L^\infty(\varepsilon, X, Z)$  and by dominated convergence Theorem,

$$\lim_n \|s_n - s\|_2^2 = \lim_n \mathbb{E}((s_n - s)^2) = 0.$$

Let introduce  $p_n$  the orthogonal projection of  $s_n$  on closure of  $\{q(Z)\varepsilon + c : q \in L^2(Z), c \in \mathbb{R}\}$ , defined by:

$$\mathbb{E}[(s_n - p_n)(q(Z)\varepsilon + c)] = 0 \quad \forall c, q \in \mathbb{R} \times L^2(Z).$$

We can check that

$$p_n = \frac{\mathbb{E}[s_n \varepsilon | Z]}{\mathbb{E}[\varepsilon^2 | Z]} \varepsilon.$$

Note that by definition  $s_n - p_n$  is the projection of  $s_n$  on  $\bar{\mathcal{S}} = \mathcal{S}$ , so  $s_n - p_n - s \in \mathcal{S}$  and  $p_n \in \mathcal{S}^\perp$ , we have by Pythagoras Theorem:

$$\|s_n - s\|_2^2 = \|s_n - p_n - s\|_2^2 + \|p_n\|_2^2 \geq \|s_n - p_n - s\|_2^2.$$

Let  $g_k(Z) = \mathbb{E}(\varepsilon^2 \mathbf{1}_{\{|\varepsilon| \geq k\}} | Z)$ , by dominated convergence Theorem:

$$\lim_k g_k(Z) = 0, \quad Z - a.s.,$$

then  $g_k(Z)$  converges in probability to 0. It follows that it exists  $k_l$  a strictly increasing sequence in  $\mathbb{N}$ , such that  $\mathbb{P}(g_{k_l}(Z) \geq \frac{1}{l}) \leq 1/l$ .

Let  $r_l(Y, X, Z) = \varepsilon \mathbb{1}_{\{|\varepsilon| < k_l\}} \mathbb{1}_{\{g_{k_l}(Z) < 1/l\}}$ .  $r_l$  converges in probability to  $\varepsilon$  because

$$\begin{aligned} \mathbb{P}(|r_l - \varepsilon| > a) &\leq \mathbb{P}(|r_l - \varepsilon| \neq 0) \\ &\leq \mathbb{P}(\{|\varepsilon| \geq k_l\} \cap \{g_{k_l}(Z) < 1/l\}) + \mathbb{P}(g_{k_l}(Z) \geq 1/l) \\ &\leq \mathbb{P}(|\varepsilon| \geq k_l) + \mathbb{P}(g_{k_l}(Z) \geq \frac{1}{l}) \\ &\leq \mathbb{P}(|\varepsilon| \geq k_l) + 1/l \end{aligned}$$

Moreover  $\mathbb{E}(r_l^2|Z)$  converges almost-surely (and then in probability) to  $\mathbb{E}(\varepsilon^2|Z)$  because:

$$\begin{aligned} |E(\varepsilon^2 - r_l^2|Z)| &= \mathbb{E}\left(\varepsilon^2 \left[ \mathbb{1}_{\{|\varepsilon| \geq k_l\}} \mathbb{1}_{\{g_{k_l}(Z) < 1/l\}} + \mathbb{1}_{\{g_{k_l}(Z) \geq 1/l\}} \right]\right) \\ &\leq \mathbb{E}\left(g_{k_l}(Z) \mathbb{1}_{\{g_{k_l}(Z) < 1/l\}}\right) + \mathbb{E}\left(\mathbb{E}[\varepsilon^2|Z] \mathbb{1}_{\{g_{k_l}(Z) \geq 1/l\}}\right) \\ &\leq \frac{1}{l} + \bar{v} \mathbb{P}(g_{k_l}(Z) \geq \frac{1}{l}) \\ &\leq \frac{1+\bar{v}}{l}. \end{aligned}$$

It follows that  $\mathbb{E}(r_l^2|Z) \geq \underline{v} - \frac{1+\bar{v}}{l} > 0$  for  $l \geq \frac{1+\bar{v}}{\underline{v}}$ .

Then for a given  $n$ , by continuous mapping Theorem,  $\mathbb{E}(s_n \varepsilon|Z)r_l/\mathbb{E}(r_l^2|Z)$  converge in probability to  $p_n$  when  $l$  tends to infinity.

Moreover,

$$\begin{aligned} \frac{\mathbb{E}(s_n \varepsilon|Z)^2 r_l^2}{\mathbb{E}(r_l^2|Z)^2} &\leq \mathbb{E}(s_n^2|Z) \mathbb{E}(\varepsilon^2|Z) \frac{\varepsilon^2}{(\underline{v} + \frac{1+\bar{v}}{n})^2} \\ &\leq 4n^2 \bar{v}^2 \frac{\varepsilon^2}{(\underline{v} + \frac{1+\bar{v}}{n})^2} \end{aligned}$$

By Theorem 17.4 in Jacod & Protter (2000), for every  $n$ ,

$$\lim_l \left\| \frac{\mathbb{E}(s_n \varepsilon|Z)r_l}{\mathbb{E}(r_l^2|Z)} - p_n \right\|_2 = 0.$$

Then for every  $n$ , let

$$h_{n,l} = \frac{\mathbb{E}(s_n \varepsilon|Z)r_l}{\mathbb{E}(r_l^2|Z)} - \mathbb{E}\left(\frac{\mathbb{E}(s_n \varepsilon|Z)r_l}{\mathbb{E}(r_l^2|Z)}\right).$$

We have:

$$\lim_l \|h_{n,l} - p_n\|_2 = 0,$$

and because  $r_l^2 = r_l \varepsilon$ :

$$\begin{aligned} \mathbb{E}[(s_n - h_{n,l})\varepsilon|Z] &= 0 \\ \mathbb{E}(h_{n,l}) &= \mathbb{E}(s_n) = 0. \end{aligned}$$

It follows that

$$s_n - h_{n,l} \in \mathcal{S} \cap L^\infty(\varepsilon, X, Z)$$

and

$$\begin{aligned}\|s_n - h_{n,l} - s\|_2 &\leq \|s_n - p_n - s\|_2 + \|p_n - h_{n,l}\|_2 \\ &\leq \|s_n - s\|_2 + \|p_n - h_{n,l}\|_2.\end{aligned}$$

And then we deduce that:

$$\lim_n \lim_l \|s_n - h_{n,l} - s\|_2 = 0.$$

Then  $s \in \overline{\mathcal{S} \cap L^\infty(\varepsilon, X, Z)}$ .

### 7.3.2 The Proof

#### The tangent set.

For a parametric sub-model indexed by  $t$  in a neighborhood of 0 in  $\mathbb{R}$ , let  $\theta(t)$  the value of the parameter for the DGP corresponding to  $t$ .

$$\begin{aligned}\theta(t) &= \int \psi_t(x) m_t^{(j)}(x) f_{Y,X,Z,t}(y, x, z) d\lambda(y, x, z) \\ &= \int h(f_{X,t}(x), x) m_t^{(j)}(x) f_{Y,X,Z,t}(y, x, z) d\lambda(y, x, z)\end{aligned}$$

Consider a family of parametric submodel with score  $l \in \mathcal{S} \subset L^2(Y, X, Z)$  such that  $\theta(t)$  is a sufficiently smooth function of  $t$  to ensure that it exists  $f \in L^2(Y, X, Z)$  such that  $\partial_t \theta(0) = \mathbb{E}(fl)$ . The set of score  $l$  is called a tangent set,  $f$  is called an influence function, and such case we say that  $\theta$  is differentiable with respect to the tangent set  $\mathcal{S}$  (see van der Vaart (2000), Chapter 25). A necessary condition for existence of a regular root-N convergent estimator is that it exists such an influence function for the parameter  $\theta = \mathbb{E}(\psi(X)m^{(j)}(X))$ . The semi-parametric efficiency bound (with respect to the tangent set  $\mathcal{S}$ ) of  $\theta_0$  is simply the smallest variance among variance of the influence functions.

In our case, we consider the tangent set of scores  $\mathcal{S} = \cup_{\varphi \in \Phi} \mathcal{S}_\varphi = \cup_{\varphi \in \Phi} \cup_{c \geq 0} \mathcal{S}_{\varphi,c}$ , where  $\mathcal{S}_{\varphi,c}$  denotes the set of scores of parametric models indexed by  $t$  and defined by:

$$Y = m(X) + t\varphi(X) + \varepsilon, \quad \mathbb{E}(\varepsilon|Z) = 0, \quad (\varepsilon, X, Z) \text{ has density } f(t) \text{ with respect to } \nu$$

such that the "true" density  $f_{\varepsilon,X,Z}$  corresponds to  $f_0$  and

$$\lim_{t \rightarrow 0} \frac{1}{t^2} \int f_t(e, x, z) \mathbf{1}_{\{f_0(e,x,z)=0\}} d\nu(e, x, z) = 0,$$

$$\lim_{t \rightarrow 0} \int \left( \frac{1}{t} \left( f_t^{1/2}(e, x, z) - f_0^{1/2}(e, x, z) \right) - \frac{1}{2} s(e, x, z) f_0^{1/2}(e, x, z) \right)^2 d\nu(e, x, z) = 0,$$

and

$$\max \left( \int e^2 f_t(e, x, z) d\nu(e, x, z), \mathbb{E}(\varepsilon^2 s^2) \right) < c.$$

Lemma 7.2 implies that the sub-models considered are differentiable in quadratic means with a score of the form  $s - r\varphi$ . It follows that the tangent set considered here is  $\mathcal{S}_0 - r\Phi$  (with  $\mathcal{S}_0$  corresponds to  $\mathcal{S}_\varphi$  with  $\varphi = 0$  and  $r$  defined in Lemma 7.1). The restriction  $\mathbb{E}(\varepsilon|Z) = 0$  imposes that:

$$\int qe f_t d\nu = 0 \text{ for any } q \in L^\infty(Z) \text{ and } t \text{ close to } 0,$$

On the other side,

$$\begin{aligned} -\mathbb{E}(q(Z)\varepsilon s(\varepsilon, X, Z)) &= \int qe \left( \frac{1}{t} (f_t - f_0) - s f_0 \right) d\nu \\ &= \int qe \left( \frac{1}{t} (f_t^{1/2} - f_0^{1/2}) - \frac{1}{2} s f_0^{1/2} \right) (f_t^{1/2} + f_0^{1/2}) d\nu(e, x, z) \\ &\quad + \frac{1}{2} \int qe s f_0^{1/2} (f_t^{1/2} - f_0^{1/2}) d\nu \end{aligned}$$

The Cauchy-Schwartz inequality ensures that:

$$\begin{aligned} |\mathbb{E}(q\varepsilon s)| &\leq \left( \int q^2 e^2 (f_t^{1/2} + f_0^{1/2})^2 d\nu \right)^{1/2} \left( \int \left( \frac{1}{t} (f_t^{1/2} - f_0^{1/2}) - \frac{1}{2} s f_0^{1/2} \right)^2 d\nu \right)^{1/2} \\ &\quad + \left( \int q^2 e^2 s^2 f_0 d\nu \right)^{1/2} \left( \int (f_t^{1/2} - f_0^{1/2})^2 d\nu \right)^{1/2} \end{aligned}$$

The triangle inequality ensures that:  $\left( \int q^2 e^2 (f_t^{1/2} + f_0^{1/2})^2 d\nu \right)^{1/2} \leq 2\sqrt{\max(c, \bar{\nu})} \|q\|_\infty$  and moreover  $\left( \int q^2 e^2 s^2 f_0 d\nu(e, x, z) \right)^{1/2} \leq \|q\|_\infty \sqrt{c}$ . So, the differentiability in quadratic mean ensures the right hand side of the previous inequality tends to 0 and next

$$\mathbb{E}(q\varepsilon s) = 0.$$

Then for any  $s \in \mathcal{S}_0$  we have  $\mathbb{E}(\varepsilon s|Z) = 0$ .

Lemma 7.3 ensures that we can restrict without loss of generality the tangent set to  $\mathcal{S}_0^b - r\Phi$  (because the efficient influence function is the projection of any influence function on  $\mathcal{S}_0 - r\Phi$  and then on  $\mathcal{S}_0^b - r\Phi$  by density of  $\mathcal{S}_0^b$  in  $\mathcal{S}_0$ ).

We also have  $\int q(z)(e - t\varphi(x))f_0(e - t\varphi(x), x, z)d\nu(e, x, z) = 0$  for any  $q \in L^\infty(Z)$ ,  $\varphi \in \Phi$ , and  $t$  close to 0. Then,

$$\int q(z)e \left( \frac{1}{t} (f_0(e - t\varphi(x), x, z) - f_0(e, x, z)) \right) d\nu(e, x, z) = \int q(z)\varphi(x)f_0(e - t\varphi(x), x, z)d\nu(e, x, z).$$

And translation invariance of the Lebesgue measure ensures that

$$\int q(z)\varphi(x)f_0(e - t\varphi(x), x, z)d\nu(e, x, z) = \int q(z)\varphi(x)f_0(e, x, z)d\nu(e, x, z).$$

It follows that:

$$\begin{aligned}
& |\mathbb{E}(q\varphi) - \mathbb{E}(q\varepsilon r\varphi)| = \left| \int qe \left( \frac{1}{t} (f(e - t\varphi(x), x, z) - f(e, x, z)) - \varphi(x)r(e, x, z)f(e, x, z) \right) d\nu(e, x, z) \right| \\
& \leq \left( \int q(z)^2 e^2 (f^{1/2}(e - t\varphi(x), x, z) + f^{1/2}(e, x, z))^2 d\nu(e, x, z) \right)^{1/2} \\
& \left( \int \left( \frac{1}{t} (f^{1/2}(e - t\varphi(x), x, z) - f^{1/2}(e, x, z)) - \frac{\varphi(x)}{2} r(e, x, z) f^{1/2}(e, x, z) \right)^2 d\nu(e, x, z) \right)^{1/2} \\
& + \left( \int q(z)^2 e^2 s^2(e, x, z) f(e, x, z) d\nu(e, x, z) \right)^{1/2} \left( \int (f^{1/2}(e - t\varphi(x), x, z) - f^{1/2}(e, x, z))^2 d\nu(e, x, z) \right)^{1/2}
\end{aligned}$$

Lemma 7.1 ensures the differentiability in quadratic mean of  $t \mapsto f_0^{1/2}(e - t\varphi(x), x, z)$ , and we conclude that for any  $\varphi \in \Phi$ ,  $\mathbb{E}(\varepsilon r\varphi|Z) = \mathbb{E}(\varphi|Z)$ .

### Differentiability of $\theta$ with respect to the tangent set

For any  $s \in S_0^b$  and  $\varphi \in \Phi$ , the score  $s - r\varphi$  is generated for instance by the parametric sub-model  $f_t(e, x, z) = f_0(e - t\varphi(x), x, z) [1 + ts(e, x, z)]$  for  $|t| \leq \|s\|_\infty$ . This sub-model generates a parametric model for the marginal density of  $X$ :  $f_{X,t} = \int f_t(e, x, z) d\lambda(e) d\rho(z)$ , where  $\rho$  is the pushforward measure of  $\mu$  under  $\pi_Z(x, z) = z$ . The projected model  $f_{X,t}$  is also differentiable in quadratic mean (cf. Proposition 5 of Appendix A.5 of Bickel et al. (1993)). A direct calculation ensures that  $f_{X,t}(x) = f_X(x)(1 + tb(x))$  and  $b$  is the score of  $t \mapsto f_{X,t}$  defined by  $b(X) = \mathbb{E}(s - r\varphi|X) = \mathbb{E}(s|X)$  (remember that  $\mathbb{E}(r|X) = 0$ ). Let  $\psi_t(x) = h(f_{X,t}(x), x)$  with  $f_{X,t}(x) = f_X(x)(1 + tb(x))$ , its derivative at  $t = 0$  is  $\partial_t \psi_t(x)|_{t=0} = \partial_{f_X} \psi(x) b(x) f_X(x)$ . Then for any  $b \in L^\infty(X)$ , we have:

$$\begin{aligned}
\frac{\theta(t) - \theta}{t} &= \frac{1}{t} \int (\psi_t(x) - \psi(x)) m^{(j)}(x) f_{X,t}(x) d\rho(x) \\
&+ \frac{1}{t} \int \psi(x) m^{(j)}(x) (f_{X,t}(x) - f_X(x)) d\rho(x) \\
&+ \int \psi_t(x) \varphi^{(j)}(x) f_{X,t}(x) d\rho(x)
\end{aligned}$$

The term  $\int \psi_t(x) \varphi^{(j)}(x) f_{X,t}(x) d\rho(x)$  is  $\int \psi_t(x) \varphi^{(j)}(x) f_X(x) d\rho(x) + t \int \psi_t(x) b(x) \varphi^{(j)}(x) f_X(x) d\rho(x)$ , and then dominated convergence Theorem ensures that

$$\lim_{t \rightarrow 0} \int \psi_t(x) \varphi^{(j)}(x) f_X(x) d\rho(x) = \int \psi(x) \varphi^{(j)}(x) f_X(x) d\rho(x) = \mathbb{E}(\psi \varphi^{(j)}).$$

The fact that  $\psi m^{(j)} \in L^2(X)$  ensures that

$$\lim_{t \rightarrow 0} \frac{1}{t} \int \psi(x) m^{(j)}(x) (f_{X,t}(x) - f_X(x)) d\rho(x) = \mathbb{E}(\psi m^{(j)} b) = \mathbb{E}(\psi m^{(j)} (s - r\varphi)).$$

And last, Assumption 2.(iv) ensures that

$$\lim_{t \rightarrow 0} \frac{1}{t} \int (\psi_t(x) - \psi(x)) m^{(j)}(x) f_{X,t}(x) d\rho(x) = \mathbb{E}(\partial_{f_X} \psi f_X b m^{(j)}) = \mathbb{E}(\partial_{f_X} \psi f_X (s - r\varphi) m^{(j)}).$$



So:

$$\lim_{t \rightarrow 0} \frac{\theta(t) - \theta}{t} = \mathbb{E}([\partial_{f_X} \psi f_X + \psi] m^{(j)}(s - r\varphi) + \psi \varphi^{(j)})$$

If it exists a regular  $\sqrt{n}$ -consistent estimator  $\hat{\theta}$  of  $\theta$ , then Theorem 2.1 of van der Vaart (1991) ensures that:

1.  $\theta(t)$  is differentiable with respect to the tangent set, i.e. it exists  $f \in L^2(\varepsilon, X, Z)$  such that  $\lim_t \frac{\theta(t) - \theta}{t} = \mathbb{E}(f(s - r\varphi))$  for any  $s \in \mathcal{S}_0$  and  $\varphi \in \Phi$ ,
2. and the asymptotic distribution of  $\hat{\theta}$  is the convolution of a distribution with a normal  $\mathcal{N}(0, V^*)$ , with  $V^* = \mathbb{V}(\Pi_{\mathcal{S}_0 - r\Phi}(f))$ , where  $\Pi_{\mathcal{S}_0 - r\Phi}$  is the orthogonal projection on  $\mathcal{S}_0 - r\Phi$ .

Note that in the previous claim,  $f$  is not necessarily unique, but  $\Pi_{\mathcal{S}_0 - r\Phi}(f)$  does not depend on the choice of  $f$ .

**Deriving the conditions and the form of semi-parametric efficiency bound** At this stage, we have show that a necessary condition for existence of regular  $\sqrt{N}$ -consistent estimator of  $\theta$  is the existence of an influence function  $f$ . If we consider the parametric sub-models with scores  $\mathcal{S}_0^b$ , the previous derivation of  $\theta(t)$  and the projection on then

$$f \in \{[\partial_{f_X} \psi f_X + \psi] m^{(j)}\} + \mathcal{S}_0^\perp.$$

Usual algebra about projections on convex in Hilbert space ensures that  $\mathcal{S}_0^\perp = \{q(Z)\varepsilon + c : q \in L^2(Z), c \in \mathbb{R}\}$ .

Now if we consider parametric sub model with scores in  $\mathcal{S}_\varphi^b = \mathcal{S}_0^b - r\varphi$ , existence of  $\sqrt{n}$  consistent and regular estimator of  $\theta$  is given by implies that it exists  $q \in L^2(Z)$  and  $c \in \mathbb{R}$  such that:

$$\forall \varphi \in \Phi : \mathbb{E}(\psi \varphi^{(j)}) = \mathbb{E}((q\varepsilon + c)r\varphi)$$

But we know that  $\mathbb{E}(\varepsilon r\varphi|Z) = \mathbb{E}(\varphi|Z)$  and  $\mathbb{E}(r|X, Z) = 0$ , so we get:

$$\forall \varphi \in \Phi : \mathbb{E}(\psi \varphi^{(j)}) = \mathbb{E}(T^*(q)\varphi).$$

Then, up to  $(-1)^{|j|}$ ,  $T^*(q)f_X$  is the weak derivative of order  $(j)$  of  $\psi f_X$  (in the sense of the linear forms on  $\Phi$ , cf. Schwartz (1997), Chapter II and III). It follows that Assumption 3 is necessary for existence of a regular and root-N consistent estimator.

So, the expression of the semi-parametric efficiency bound is:

$$V^* = \min_{q \in \mathcal{Q}} \mathbb{V} [\varepsilon q(Z) + (\psi(X) + \partial_{f_X} \psi(X) f_X(X)) m^{(j)}(X)],$$

Where  $\mathcal{Q}$  is a closed convex subset in  $L^2(Z)$  and the objective function is a quadratic form

of  $q$  that tends to  $+\infty$  when  $\|q\|_{L^2(Z)} \rightarrow +\infty$ , this minimisation admits a unique solution  $q^*$ .

**Representation of  $\theta$  under Assumption 2.(ix)** If we have  $rm \in L^2(\varepsilon, X, Z)$ , then the parametric sub-model defined by  $Y = (1+t)m(X) + \varepsilon$ , with  $\mathbb{E}(\varepsilon|Z) = 0$  and  $(\varepsilon, X, Z) \sim f_t(e, x, z)$  and  $f_t^{1/2}$  differentiable in quadratic means is regular. Following the same reasoning as previously, we know that existence of regular estimator of  $\theta$  implies that  $\theta = \mathbb{E}(\psi(X)m(X)^{(j)}) = \mathbb{E}(q(Z)m(X))$  for any  $q \in \mathcal{Q}$ . So, the parameter can be expressed as the expectation of  $q^*(Z)Y$ .

#### 7.4 Proof of Proposition 4.2

In Theorem 4.1, the bound is the variance of the sum of a function  $R(X)$  and an element of  $\varepsilon L^2(X)$ , then a projection of  $R(X)$  on  $\varepsilon L^2(Z)$  and Pythagorean Theorem ensures that:

$$\begin{aligned} V^* &= \min_{q \in \mathcal{Q}} \mathbb{E} \left[ (\varepsilon q(Z) + R(X))^2 \right] \\ &= \min_{q \in \mathcal{Q}} \mathbb{E} \left[ \varepsilon^2 \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right] + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right] \\ &= \min_{q \in \mathcal{Q}} \mathbb{E} \left[ \mathbb{E}(\varepsilon^2|Z) \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right] + \mathbb{E} \left[ \left( R(X) - \varepsilon \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right] \end{aligned}$$

The second term does not depend on  $q$ , so the minimization on  $\mathcal{Q}$  only concerns the first term. This first term could be expressed as a function of a generalized inverse of an operator  $K^*$ .

When  $q$  varies in  $\mathcal{Q}$  then  $s = \mathbb{E}(\varepsilon^2|Z) \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2$  varies in  $\mathcal{T}$  defined by:

$$\mathcal{T} = \left\{ s \in L^2(Z) : K^*(s) = (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right\}$$

And next:

$$\min_{q \in \mathcal{Q}} \mathbb{E} \left[ \mathbb{E}(\varepsilon^2|Z) \left( q(Z) + \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} \right)^2 \right] = \min_{s \in \mathcal{T}} \mathbb{E} [s^2(Z)].$$

So we are lead back to find the element  $s^*$  of minimum norm in  $\mathcal{T}$  which is the preimage of  $\left\{ (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right\}$  under  $K^*$ . By definition,  $s^*(Z) = \text{Arg min}_{s \in \mathcal{T}} \mathbb{E}(s^2(Z))$ , is the generalized inverse  $K^{*\dagger} \left[ (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right) \right]$  which is well defined if

and only if  $(-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} \in \mathcal{R}(K^*) = \mathcal{R}(T^*)$ . And then we get the formula (i) and (ii) of the Proposition 4.2.

Moreover, Corollary 3.2.2 in Groetsch (1977) ensures that:

$$K^{*\dagger} = \lim_{t \downarrow 0} (tI + KK^*)^{-1} K.$$

Let  $h = (-1)^{|j|} \frac{[\psi f_X]^{(j)}}{f_X} + K^* \left( \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\sqrt{\mathbb{E}(\varepsilon^2|Z)}} \right)$ , under Assumption 3 it exists  $u$  such that  $K^*u = h$  and  $K^\dagger u = \lim_{t \downarrow 0} (tI + K^*K)^{-1} K^*u$  is well-defined if and only if  $u \in \mathcal{R}(K) + \ker K^*$  (see Groetsch (1977), Chapter 3), then

$$\begin{aligned} K^{*\dagger} h &= \lim_{t \downarrow 0} (tI + KK^*)^{-1} K (tI + KK^*)^{-1} K (tI + K^*K) (tI + K^*K)^{-1} K^*u \\ &= \lim_{t \downarrow 0} K (tI + K^*K)^{-1} K^*u \\ &= KK^\dagger u \text{ if and only if } u \in \mathcal{R}(K) + \ker K^* \text{ or equivalently } h \in \mathcal{R}(K^*K) \end{aligned}$$

Last because  $K^\dagger u$  is the mean square solution of the equation  $Kx = u$ , this is also the mean square solution of  $K^*Kx = K^*u$  and then  $K^\dagger u = (K^*K)^\dagger K^*u$ . The formula (iii) and (iv) of Proposition 4.2 follows.

## 7.5 Proof of Proposition 5.2

The first part of the proposition is given by the Theorem 1.2.4 of Groetsch (1977). The second part of the proposition is given by the Theorem 3.1.3 of Groetsch (1977) and the fact that  $T$  is normal if and only if  $T^*$  is normal. For the third part of the proposition, assume that  $\text{Ker}((T)) = \{0\}$  and that  $\mathbb{E}(g(x)|Z) = \int f_{X|Z}(x)g(x)d\lambda(x)$  where  $\lambda$  is the Lebesgue measure on  $\mathbb{R}^K$ . For  $g_n(x) = \cos(n(x_1 + x_2 + \dots + x_K))$ , we have  $\int f_{X|Z}(x)g_n(x)d\lambda(x) \rightarrow 0$  when  $n$  tends to  $+\infty$  because the Fourier transform of  $\sum_{i=1}^K X_i|Z$  tends to zero at the infinity. Moreover the Cauchy-Schwartz inequality ensures that  $[\int f_{X|Z}(x)g_n(x)d\lambda(x)]^2 \leq [\int f_{X|Z}(x)d\lambda(x)] [\int f_{X|Z}(x)g_n^2(x)d\lambda(x)] \leq 1$ . Then the dominated convergence Theorem ensures that  $\|T(g_n)\|_{L^2(Z)}$  tends to zero when  $n$  tends to  $\infty$ . On the other hand  $\|g_n\|_{L^2(Z)}^2 = \int \cos^2(n(x_1 + \dots + x_K))f_{X|Z}(x)d\lambda(x) = \frac{1}{2} + \frac{1}{2} \int \sin(2n(x_1 + \dots + x_K))f_{X|Z}(x)d\lambda(x)$  and tends to  $1/2$  when  $n$  tends to infinity. Then, Theorem 1.2.3 in Groetsch (1977) ensures that  $\mathcal{R}(T)$  is not closed and next by the first part of the proposition  $\mathcal{R}(T^*)$  is not closed.

## 7.6 Definition and properties of Fourier and Laplace transforms of linear forms.

As explained in Section 5.2.2, the condition given in the Theorem 5.3 are sufficient to ensure that  $\varrho$  is a Laplace transform of a linear form. We will briefly present the definition of the Laplace transform of a linear form.

Let define  $\mathcal{S}$ , the Schwartz space of  $\mathbb{R}^p$ : this is the set of complex-valued functions that are indefinitely differentiable such that themselves and all their derivatives tend to 0 faster than any inverse power of the norm (see Schwartz (1997), Chapter VII) :  $\lim_{|x| \rightarrow \infty} \prod_{i=1}^p x_i^{k_i} \varphi^{(j)}(x) = 0$ . The usual Fourier transform that maps  $\varphi$  to  $\mathcal{F}_\varphi(t) = \int \exp(-2\pi i v' t) \varphi(v) dv$  is a bijection on  $\mathcal{S}$ . The topology of  $\mathcal{S}$  is the topology of the following convergence

$$\varphi_j \rightarrow 0 \text{ if and only if } \forall ((k), (l)) \in (\mathbb{N}^p)^2, \sup_{x \in \mathbb{R}^p} \left| \prod_{i=1}^p x_i^{k_i} \varphi_j^{(l)}(x) \right| \rightarrow 0.$$

For any continuous linear form  $T \in \mathcal{S}'$ ,  $\varphi \in \mathcal{S} \mapsto T(\mathcal{F}_\varphi)$  is well-defined because  $\mathcal{F}_\varphi \in \mathcal{S}$ . This is a continuous linear form on  $\mathcal{S}$ , denoted  $\mathcal{F}[T]$  and called Fourier transform of  $T$ . The Fourier transform of  $T \in \mathcal{S}'$  is related to the usual Fourier transform on  $\mathcal{S}$  through the formula:

$$\mathcal{F}[T](\varphi) = T(\mathcal{F}_\varphi).$$

For the definition of Laplace transform of  $T$ , let  $\check{\mathcal{F}}_\varphi$  defined by  $\check{\mathcal{F}}_\varphi(t) = \mathcal{F}_\varphi\left(\frac{t}{2\pi}\right)$ , this also defined a bijection on  $\mathcal{S}$ .

For any continuous linear form  $T \in \mathcal{S}'$ ,  $\varphi \in \mathcal{S} \mapsto T(\check{\mathcal{F}}_\varphi)$  is well-defined linear form on  $\mathcal{S}$ , denoted  $\check{\mathcal{F}}[T]$ . Moreover, this linear form is continuous  $\check{\mathcal{F}}[T] \in \mathcal{S}'$ .

For any  $T \in \mathcal{S}'$ , let  $\Gamma_T$  the convex set containing 0, defined by:

$$u \in \Gamma_T \Leftrightarrow [\varphi \mapsto T(\exp(-u' \cdot) \varphi(\cdot))] \text{ is a well-defined element of } \mathcal{S}'.$$

For instance, consider the case where  $T(\varphi) = \int f(x) \varphi(x) dx$ . If  $f$  is integrable with a compact support on  $\mathbb{R}^p$  then  $\Gamma_T = \mathbb{R}^p$ , if  $f(x) = \prod_{i=1}^p \mathbf{1}_{\{x_i > 0\}}$  then  $\Gamma_T = \mathbb{R}_+^p$  and if  $f(x) = 1$  then  $\Gamma_T = \{0\}$ .

For  $u \in \Gamma_T$ ,  $\exp(-u' \cdot) T$  denotes the element of  $\mathcal{S}'$  defined by  $[\exp(-u' \cdot) T](\varphi) = T(\tilde{\varphi})$  with  $\tilde{\varphi}(x) = \exp(-u' x) \varphi(x)$ . And  $u \in \Gamma_T$  if and only if  $\exp(-u' \cdot) T \in \mathcal{S}'$ .

The Laplace transform of  $T$  is defined from  $\Gamma_T$  to  $\mathcal{S}'$  by:

$$\mathcal{L}[T] : u \in \Gamma_T \mapsto [\varphi \mapsto \check{\mathcal{F}}[\exp(-u' \cdot) T](\varphi)].$$

For instance, consider the case where  $T(\varphi) = \int f(x)\varphi(x)dx$  for  $f$  a bounded function, then  $\mathcal{L}[T](u)(\varphi) = \int f(x)e^{-u'x} \left( \int e^{-iv'x}\varphi(v)dv \right) dx$ . Then  $\mathcal{L}[T]$  is related to the usual Laplace transform  $\mathcal{L}_f : (u + iv) \mapsto \int f(x)e^{-(u+iv)x}dx$  via the following canonical mapping:

$$\mathcal{L}[T](u)(\varphi) = \int \mathcal{L}_f(u + iv)\varphi(v)dv.$$

Similarly, if  $T(\varphi) = \int \varphi(x)d\mu(x)$  for  $\mu$  a signed measure, then  $\mathcal{L}[T](u)(\varphi) = \int \mathcal{L}_\mu(u + iv)\varphi(v)dv$  with  $\mathcal{L}_\mu(u + iv) = \int e^{-(u+iv)'x}d\mu(x)$ . More generally, for any  $T \in \mathcal{S}'$  it exists a function  $\mathcal{L}_T$  defined on  $\Gamma_T + i\mathbb{R}^p$  such that  $\mathcal{L}(T)(u)(\varphi) = \int \mathcal{L}_T(u + iv)\varphi(v)dv$  (Schwartz, 1997, Chapter VIII), and such function is strongly regular : for instance if  $\Gamma_T$  is an open set, then  $\mathcal{L}_T$  is holomorphic on  $\Gamma_T + i\mathbb{R}^p$ . Because it exists a canonical isomorphism between  $\mathcal{L}[T] \in (\Gamma_T)^{\mathcal{S}'}$  and  $\mathcal{L}_T \in (\Gamma_T + i\mathbb{R}^p)^{\mathbb{C}}$ , it is not worthwhile to make difference between these two objects.

## 7.7 Proof of Theorem 5.3

Before to prove the Theorem, we will prove two useful Lemma to establish (iii).

### 7.7.1 Lemma

#### **Lemma 7.4 (Approximation of compact sets by polytopes)**

*Let  $S$  a convex set in  $\mathbb{R}^p$  such that  $\text{int}(S) \neq \emptyset$  and  $K$  a compact included in  $\text{int}(S)$ , then it exists a polytope  $P$  such that  $K \subset P \subset \text{int}(S)$ .*

The case  $p = 1$  is obvious because in that case  $\text{int}(S)$  is an open interval of  $\mathbb{R}$ , and we can choose  $P = [\inf(K); \sup(K)]$ . Because  $K$  is bounded it exists a closed ball of radius  $r$  and center  $x_0$  containing  $K$ , we can assume without loss of generality that  $S$  is bounded (if  $S$  is unbounded replace  $S$  by  $S \cap \overline{B}(x_0, r)$  in the following). Let  $K' = \text{co}(K)$  the convex hull of  $K$ ,  $K'$  is compact because this is the range of the continuous mapping  $(\lambda, x) \in \Lambda \times K \mapsto \lambda \cdot x$ , where  $\Lambda$  is the compact  $\{\lambda \in [0; 1]^p : \sum_i^p \lambda_i = 1\}$  (the Thykhonov Theorem ensures that  $\Lambda \times K$  is compact in  $\mathbb{R}^{2p}$ ). For  $(x, y) \in \mathbb{R}^p$ ,  $d_2(x, y) = \|x - y\|_2$  denotes the Euclidean distance between  $x$  and  $y$  and for non-empty  $A \subset \mathbb{R}^p$ ,  $d_2(A, y) = \inf_{x \in A} d_2(x, y)$ . For any  $A \subset \mathbb{R}^p$ ,  $A^c$  denotes the complementary of  $A$  in  $\mathbb{R}^p$ . Let  $d = \inf_{y \in S^c} d_2(y, K') > 0$ , compactness of  $K'$  ensures that  $d > 0$  (if  $d = 0$  it exists a sequence  $x_n \in K'$  such that  $d_2(x_n, S^c) < 1/n$  and a subsequence that converges to  $x_\infty \in K' \cap \text{cl}(S^c) = K' \cap \text{int}(S)$ ). For any non-empty subsets  $A$  and  $B$  in  $\mathbb{R}^p$ , let  $d_H(A, B) = \sup\{\sup_{x \in A} d_2(B, x), \sup_{y \in B} d_2(A, y)\}$  the Hausdorff metric between  $A$  and  $B$ . It exists a polytope  $P$  such that  $d_H(P, K') < d/2$  and  $K' \subset P$  (Bronstein (2008)). For any non-empty convex  $C \subset \mathbb{R}^p$  and any  $x \in \mathbb{R}^p$ , it exists a unique

$\pi_C(x) \in \text{cl}(C)$  such that  $d_2(x, \pi_C(x)) = d_2(x, C)$  (Hiriart-Urruty & Lemaréchal (2001), Section A.3). For any  $y \in S^c$  :

$$\begin{aligned}
d \leq d_2(y, \pi_{K'}(y)) &\leq d_2(y, \pi_{K'}(\pi_P(y))) \\
&\leq d_2(y, \pi_P(y)) + d_2(\pi_P(y), \pi_{K'}(\pi_P(y))) \\
&\leq d_2(y, \pi_P(y)) + \inf_{u \in K'} d_2(\pi_P(y), u) \\
&\leq d_2(y, \pi_P(y)) + \sup_{v \in P} \inf_{u \in K'} d_2(v, u) \\
&\leq d_2(y, \pi_P(y)) + d_H(P, K') \\
&\leq d_2(y, \pi_P(y)) + d/2
\end{aligned}$$

It follows that  $d_2(y, \pi_P(y)) \geq d/2$  for any  $y \in S^c$  and then  $\inf_{y \in S^c} \inf_{x \in P} d_2(x, y) > 0$ . This ensures that  $P \cap \text{cl}(S^c) = \emptyset$  and then  $P \subset \text{int}(S)$ .

### 7.7.2 Theorem 5.3

For any  $q \in \mathcal{Q}$ , let  $\tilde{q}(\tau_W(\tilde{Z}), W) = \mathbb{E}(q(\tilde{Z}, W) | \tau_W(\tilde{Z}), W)$ . Under Assumption 4,  $\tilde{X} \perp\!\!\!\perp \tilde{Z} | \tau_W(\tilde{Z}), W$  then for any  $q \in \mathcal{Q}$ ,  $T^*(q) = T^*(\tilde{q})$ .

Moreover, under Assumption 4 the Bayes formula implies:

$$\mathbb{E}(\tilde{q}(\tau(\tilde{Z}, W))) | \tilde{X} = \tilde{x}, W = w = \frac{\int \tilde{q}(t, w) g_w(t) \exp(-\mu_w(\tilde{x})'t) dF_{\tau_W(Z)|W=w}(t)}{\nu_w(t)}$$

Assumption 3 ensures that  $\forall \varphi \in \Phi$ :

$$\begin{aligned}
\int \psi(x) \varphi^{(j)}(x) dF_X(x) &= \int \mathbb{E}(q(Z) | X = x) \varphi(x) dF_X(x) \\
&= \int \frac{\int \tilde{q}(t, w) g_w(t) \exp(-\mu_w(x)'t) dF_{\tau(Z)|W=w}(t)}{\nu_w(t)} \varphi(\tilde{x}, w) dF_X(\tilde{x}, w) \\
&= \int r_w(\mu_w(\tilde{x})) \varphi(\tilde{x}, w) dF_X(\tilde{x}, w).
\end{aligned}$$

Then if  $\mathcal{Q} \neq \emptyset$  :

$$(-1)^{|j|} \frac{[\psi(X) f_X(X)]^{(j)}}{f_X(X)} = r_W(\mu_W(\tilde{X})) = \frac{\int \mathbb{E}(q(Z) | \tau(Z) = t, W) g_W(t) \exp(-\mu(X)'t) dF_{\tau(Z)|W}(t)}{\nu_W(\mu_W(\tilde{X}))}.$$

And (i) holds because  $r_W(\mu_W(\tilde{X})) = \mathbb{E}(q(Z) | X)$  which is integrable with respect to  $F_{\tilde{X}|W}$ . Remark that for any  $v \in \mathbb{R}^{p_W}$ ,  $\varrho_W(\mu_W(\tilde{X}) + iv)$  is (almost-surely) well defined and finite because  $\int |\tilde{q}(t, W) g_W(t) \exp(-(\mu_W(\tilde{X}) + iv)'t)| dF_{\tau(Z)|W}(t) \leq \nu_W(\mu_W(\tilde{X})) \mathbb{E}(|q(Z)| | \mu_W(\tilde{X}), W)$ . By definition,  $\rho_W$  is the Laplace transform of the measure  $M_W$  defined by  $dM_W(t) = \tilde{q}(t, W) g_W(t) dF_{\tau_W(\tilde{Z})|W}(t)$ . Domain of a Laplace transforms are always of the form  $\Gamma + i\mathbb{R}^p$  where  $\Gamma$  is convex (see for instance Schwartz (1997), Chapter VIII). It follows that  $\rho_W$  is defined on  $C_W + i\mathbb{R}^{p_W}$  where  $C_W$  is a convex set such that  $\mathbb{P}(\mu_W(\tilde{X}) \in C_W | W) = 1$ , then  $\text{Supp}(\mu_W(\tilde{X}) | W) \subset \text{cl}(C_W)$  and  $\text{int co Supp}(\mu_W(\tilde{X}) | W) \subset \text{int cl}(C_W)$ . It follows

$\Gamma_W \subset \text{int}(C_W)$  (a convex set in  $\mathbb{R}^d$  has the same interior than its closure, see for instance Hiriart-Urruty & Lemaréchal, 2001, pages 36 and 37). Moreover, Proposition 6 in Chapter VIII of Schwartz (1997) ensures that  $\rho_W$  is holomorphic on  $\text{int}C_W + i\mathbb{R}^{p_W}$  and then on  $\Gamma_W + i\mathbb{R}^{p_W}$  and Condition (ii) holds.

For any  $K_W$  compact included in  $\Gamma_W$ , Lemma 7.4 ensures that it exists  $(u_1, \dots, u_{p_W})$  (depending on  $W$ ) such that  $K_W \subset \text{co}(u_1, \dots, u_{p_W}) \subset \Gamma_W$ , then for every  $u + iv \in K_W + i\mathbb{R}^p$ :

$$\begin{aligned} |\rho_W(u + iv)| &= \left| \int \tilde{q}(t, W) g_W(t) \exp(-(u + iv)'t) dF_{\tau(Z)|W}(t) \right| \\ &\leq \int |\tilde{q}(t, W)| g_W(t) \exp(-u't) dF_{\tau(Z)|W}(t) \\ &\leq \sum_{i=1}^l \lambda_i \int |\tilde{q}(t, W)| g_W(t) \exp(-u_i't) dF_{\tau(Z)|W}(t), \text{ with } \sum_{i=1}^l \lambda_i u_i = u \\ &\leq \sum_{i=1}^l \int |\tilde{q}(t, W)| g_W(t) \exp(-u_i't) dF_{\tau(Z)|W}(t) \\ &\leq \sum_{i=1}^l \mathbb{E}(|q(Z)| | \mu(X) = u_i, W) \nu_W(u_i) < \infty \end{aligned}$$

Then (iii) holds.

Last if  $\tau(Z)|W$  is dominated by the Lebesgue measure, then it exists  $f(t, w) \geq 0$  such that  $\mathbb{E}(h(\tau_W(\tilde{Z}))) = \int h(t) f(t, w) dt$ . Let  $\xi_{u,w}(t) = \tilde{q}(t, w) g_w(t) f(t, w) \exp(-u't)$ . For any  $v \in \mathbb{R}^{p_W}$ ,  $\tilde{v}$  denotes the vector of  $\mathbb{R}^p$  with components  $\tilde{v}_k = 0$  if  $|v_k| < \max_{k'=1, \dots, p_W} |v_{k'}|$  and  $\tilde{v}_k = \pi/v_k$  if  $|v_k| = \max_{k'=1, \dots, p_W} |v_{k'}|$ . Then  $\rho_W(u + iv) = \int_{\mathbb{R}^p} \xi_{u,w}(t) \exp(-iv't) dt = - \int_{\mathbb{R}^p} \xi_{u,w}(t) \exp(-iv'(t - \tilde{v})) dt = - \int_{\mathbb{R}^p} \xi_{u,w}(t + \tilde{v}) \exp(-iv't) dt$ .

It follows that:

$$2|\rho_W(u + iv)| \leq \int |\xi_{u,w}(t) - \xi_{u,w}(t + \tilde{v})| dt$$

Continuity of translations in  $L^1(\mathbb{R}^p)$  (cf. Rudin (1987), Theorem 9.5) ensures that  $\rho_W(u + iv)$  tends to 0 when  $\max_{k'=1, \dots, p_W} |v_{k'}|$  tends to  $\infty$  or equivalently when  $v$  tends to  $\infty$ . And the Condition (iv) holds.

To derive the form of the semi-parametric efficiency bound, note that the Laplace transform is injective. It follows that  $\tilde{q}(t, W) = \frac{1}{g_W(t)} \frac{d\mathcal{L}^{-1}[q_W]}{dF_{\tau_W(\tilde{Z})|W}}(t)$  does not depend on the choice of  $q \in \mathcal{Q}$ . Then if  $\mathcal{Q}$  is not empty, we have:

$$\mathcal{Q} = \tilde{q}(\tau_W(\tilde{Z}), W) + \{\eta \in L^2(Z) \text{ such that } \mathbb{E}(\eta(Z) | \tau_W(\tilde{Z}), W) = 0\}$$

Let  $q^*(Z) = \tilde{q}(\tau_W(\tilde{Z}), W) + \eta^*(Z)$  the element of  $\mathcal{Q}$  that achieves the bound  $V^*$ . For any  $\eta \in L^2(Z)$  such that  $\mathbb{E}(\eta(Z) | \tau_W(\tilde{Z}), W) = 0$  and any  $\delta \in \mathbb{R}$ :

$$\mathbb{V}(\varepsilon q^*(Z) + R(X)) \leq \mathbb{V}(\varepsilon(q^*(Z) + \delta\eta(Z)) + R(X)),$$

or equivalently, after an expansion:

$$\delta^2 \mathbb{V}(\varepsilon\eta(Z)) + 2\delta \text{Cov}(\varepsilon\eta(Z), \varepsilon q^*(Z) + R(X)) \geq 0.$$

It follows that for any  $\eta \in L^2(Z)$  such that  $\mathbb{E}(\eta(Z)|\tau_W(\tilde{Z}), W) = 0$ :

$$\text{Cov}(\varepsilon\eta(Z), \varepsilon q^*(Z) + R(X)) = 0,$$

Because  $\mathbb{E}(\varepsilon\eta(Z)) = \mathbb{E}(\mathbb{E}(\varepsilon|Z)\eta(Z)) = 0$ , then for any  $\eta \in L^2(Z)$  such that  $\mathbb{E}(\eta(Z)|\tau_W(\tilde{Z}), W) = 0$  we have:

$$\mathbb{E} \left[ (\mathbb{E}(\varepsilon^2|Z)q^*(Z) + \mathbb{E}(\varepsilon R(X)|Z)) \eta(Z) \right] = 0.$$

Consider  $\eta(Z) = \mathbb{E}(\varepsilon^2|Z)q^*(Z) + \mathbb{E}(\varepsilon R(X)|Z) - \mathbb{E}(\mathbb{E}(\varepsilon^2|Z)q^*(Z)|\tau_W(\tilde{Z}), W) + \mathbb{E}(\varepsilon R(X)|\tau_W(\tilde{Z}), W)$ , to obtain that:

$$\mathbb{E} \left( \mathbb{V} \left[ \mathbb{E}(\varepsilon^2|Z)q^*(Z) + \mathbb{E}(\varepsilon R(X)|Z) | \tau_W(\tilde{Z}), W \right] \right) = 0.$$

It follows that:

$$\mathbb{E}(\varepsilon^2|Z)q^*(Z) + \mathbb{E}(\varepsilon R(X)|Z) = \mathbb{E} \left( \varepsilon^2 q^*(Z) | \tau_W(\tilde{Z}), W \right) + \mathbb{E}(\varepsilon R(X) | \tau_W(\tilde{Z}), W).$$

Because  $q^*(Z) = \tilde{q}(\tau_W(\tilde{Z})) + \eta^*(Z)$ , it exists a function  $h$  such that:

$$\eta^*(\tilde{Z}) = -\tilde{q}(\tau_W(\tilde{Z}), W) - \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)} + \frac{h(\tau_W(\tilde{Z}), W)}{\mathbb{E}(\varepsilon^2|Z)},$$

and then, the condition  $\mathbb{E}(\eta^*(Z)|\tau_W(\tilde{Z}), W) = 0$  ensures:

$$h(\tau_W(\tilde{Z}), W) = \frac{\tilde{q}(\tau_W(\tilde{Z}), W) + \mathbb{E} \left[ \mathbb{E}(\varepsilon R(X)|Z) / \mathbb{E}(\varepsilon^2|Z) | \tau_W(\tilde{Z}), W \right]}{\mathbb{E} \left[ 1 / \mathbb{E}(\varepsilon^2|Z) | \tau_W(\tilde{Z}), W \right]},$$

and next:

$$q^*(Z) = \frac{\tilde{q}(\tau_W(\tilde{Z}), W)}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E} \left[ 1 / \mathbb{E}(\varepsilon^2|Z) | \tau_W(\tilde{Z}), W \right]} + \frac{\mathbb{E} \left[ \mathbb{E}(\varepsilon R(X)|Z) / \mathbb{E}(\varepsilon^2|Z) | \tau_W(\tilde{Z}), W \right]}{\mathbb{E}(\varepsilon^2|Z)\mathbb{E} \left[ 1 / \mathbb{E}(\varepsilon^2|Z) | \tau_W(\tilde{Z}), W \right]} - \frac{\mathbb{E}(\varepsilon R(X)|Z)}{\mathbb{E}(\varepsilon^2|Z)}$$

and the form of the semiparametric efficiency bound follows.

## 7.8 Proof of Theorem 5.4

If it exists a regular and root-N consistent estimator, then  $\psi f_X$  admits a weak derivative of order  $(j)$  and there exists  $q \in L^2(Z)$  such that  $(-1)^{(j)}\mathbb{E}(q(Z)|X = x)f_X(x) = [\psi f_X]^{(j)}(x)$ .



Assumption 5 ensures that  $Z \perp\!\!\!\perp X|p(Z)$ , then

$$T^*(q)(x) = \mathbb{E}(q(Z)|X = x) = \mathbb{E}(\mathbb{E}(q(Z)|p(Z))|X = x).$$

Let  $\tilde{q}(v) = \mathbb{E}(q(Z)|p(Z) = v)$ , we have  $\tilde{q} \in L^2(p(Z))$  and:

$$\begin{aligned} \int \psi(x)\varphi^{(j)}(x)dF_X(x) &= \int \varphi(x)T^*(q)(x)f_X(x)dx \\ &= \int \varphi(x)\mathbb{E}(\tilde{q}(p(Z))|X = x)f_X(x)dx \\ &= \int \varphi(x) [(\tilde{q}f_{p(Z)}) \star f_\eta](x)dx \end{aligned}$$

Then if there exists a regular and root-N consistent estimator, we have:

$$[\psi f_X]^{(j)} = (-1)^{|j|} (\tilde{q}f_{p(Z)}) \star f_\eta.$$

The Fourier transform are well defined and next:

$$\mathcal{F}_{[\psi f_X]^{(j)}} = (-1)^{|j|} \mathcal{F}_{\tilde{q}f_{p(Z)}} \mathcal{F}_{f_\eta}.$$

Because  $\mathcal{F}_{f_\eta}$  has isolated zeros, we get the "only if part" of the Proposition (i) of Theorem 5.4. The "if part" is achieved by the fact that if such  $\tilde{q}$  exists then  $\tilde{q} \in \mathcal{Q}$ .

Because  $q \in L^1(Z)$  then  $\tilde{q}f_{p(Z)}$  is integrable with respect to the Lebesgue measure and next  $\mathcal{F}_{\tilde{q}f_{p(Z)}} \in \mathcal{C}_0^0(\mathbb{R}^K)$  by the Riemann-Lebesgue lemma and (ii) holds.

Moreover, when  $\mathcal{F}_{\tilde{q}f_{p(Z)}} \in L^1(\mathbb{R}^K) \cap \mathcal{C}_0^0(\mathbb{R}^K)$ , then usual inversion formula applies i.e.  $\mathcal{F}^{-1} = \overline{\mathcal{F}}$ . Then  $\mathcal{Q}$  is not empty if and only if it exists  $(-1)^{|j|} \overline{\mathcal{F}} \left( \mathcal{F}_{[\psi f_X]^{(j)}} / \mathcal{F}_{f_\eta} \right) \in L^1(\mathbb{R}^K)$ , which is equivalent to  $\mathcal{F} \left( \mathcal{F}_{[\psi f_X]^{(j)}} / \mathcal{F}_{f_\eta} \right) \in L^1(\mathbb{R}^K)$ . And next, the full characterization of  $\mathcal{Q}$  follows. Then (iii) holds.

If  $p(Z)$  has a compact support then  $\varphi \mapsto \int \tilde{q}(v)f_{p(Z)}(v)\varphi(v)dv$  is a linear form compactly supported and then its Laplace transform verifies conditions (iv).(a), (iv).(b) and (iv).(c) (cf. Schwartz (1997)). Moreover, the identity

$$\mathcal{L}_{[\psi f_X]^{(j)}} = (-1)^{|j|} \mathcal{L}_{\tilde{q}f_{p(Z)}} \mathcal{L}_{f_\eta}$$

holds on  $\Gamma_{f_\eta} + i\mathbb{R}^K = \mathbb{C}^K$  and next

$$\frac{\mathcal{L}_{[\psi f_X]^{(j)}}}{\mathcal{L}_{f_\eta}} = (-1)^{|j|} \mathcal{L}_{\tilde{q}f_{p(Z)}}$$

on  $\mathbb{C}^K$ .

Last, the expression of  $q^*$  is achieved by the same reasoning than those used in the proof of Theorem 5.3.

## 7.9 Proof of Propostion 5.5

It follows from Condition (i) of Theorem 5.4 that a regular and root-N estimator exists only if the weak derivative of  $f_X$  is a function (locally integrable)  $f_X^{(j)}$  and we have (cf. Schwartz (1997), Equation VII.7.3, page 253):  $\mathcal{F}_{f_X^{(j)}}(t) = (2\pi i)^{|j|} t^{(j)} \mathcal{F}_{f_X}(t)$  with  $t^{(j)} = \prod_{k=1}^K t_k^{j_k}$ . The additive first stage ensures that  $\mathcal{F}_{f_X^{(j)}}(t) = (2\pi i)^{|j|} t^{(j)} \mathcal{F}_{f_{p(Z)}}(t) \mathcal{F}_{f_\eta}(t)$ . The first condition of Theorem 5.4 ensures that there exists  $q \in L^2(p(Z))$  such that:

$$(2\pi i)^{|j|} t^{(j)} \mathcal{F}_{f_{p(Z)}}(t) = (-1)^{|j|} \mathcal{F}_{qf_{p(Z)}}(t).$$

Because  $(2\pi i)^{|j|} t^{(j)}$  is the weak derivative of order  $(j)$  of the Dirac at 0, it follows (cf. Schwartz, 1997, Chapter VII, Theorem XV) that  $(-1)^{|j|} qf_{p(Z)}$  is the weak derivative of order  $(j)$  of  $f_{p(Z)}$ . Then the weak derivative of  $f_{p(Z)}$  is a function of the form  $(-1)^{|j|} qf_{p(Z)}$  with  $q \in L^2(p(Z))$  and next  $\mathbb{E} \left( \left( \frac{f_{p(Z)}^{(j)}(p(Z))}{f_{p(Z)}(p(Z))} \right)^2 \right) < \infty$ . The expression of  $q^*$  follows from Theorem 5.4.

## 7.10 A Lemma on Fourier transform

### Lemma 7.5 (Fourier transforms and derivatives)

Let  $P$  a probability measure on  $\mathbb{R}$ , and  $H$  and  $G$  two signed measures dominated by  $P$  such that:

$P$  is dominated by  $H$ ,

$\frac{dH}{dP}$  is locally absolutely continuous, and

$$\text{for any } x \in \mathbb{R} : \quad 2\pi i x \mathcal{F}[H](x) = \mathcal{F}[G](x),$$

then  $P$  admits a locally absolutely continuous density with respect to the Lebesgue measure.

Proof:

As a signed measured  $H \in \mathcal{S}'$  and because the weak derivative of the Dirac at 0 admits a compact support,  $2\pi i x \mathcal{F}[H] = \mathcal{F}[D\delta \star H] = \mathcal{F}[DH]$  where  $DH$  is the weak derivative of  $H$  and  $D\delta$  is the weak derivative of the Dirac at 0. The injectivity of the Fourier transform on  $\mathcal{S}'$  ensures that the weak derivative of the measure  $H$  needs to be the measure  $G$ . Theorem II, Chapter II in Schwartz (1997) ensures that it exists a function  $\tilde{h}$  (with bounded variation on every finit interval) such that  $dH(t) = \tilde{h}(t)d\lambda(t)$  where  $\lambda$  is the Lebesgue measure on  $\mathbb{R}$ . Then  $dP(t) = \tilde{h}(t) \frac{dP}{dH}(t) d\lambda(t)$  and next,  $dG(t) = \tilde{g}(t) d\lambda(t)$  with

$\tilde{g} = \tilde{h}(t) \frac{dP}{dH}(t) \frac{dG}{dP}(t)$  and  $\tilde{g} \in L^1(\lambda)$ . So  $\tilde{g}$  is the weak derivative of  $\tilde{h}$ . The Theorem III in Chapter II of Schwartz ensures that  $\tilde{h}$  is absolutely continuous. The product of locally absolutely continuous functions is a locally absolutely continuous function. And next  $\tilde{h}(t) \frac{dP}{dH}(t)$  is locally absolutely continuous, this is the density of  $P$  with respect to the Lebesgue measure.

## 8 Appendix B: Details on Table 1

### 8.1 Proof of the results about the normal bivariate case

We consider the case where:

$$(X, Z) \sim \mathcal{N} \left( \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right) \right),$$

Following the notations of Assumption 4 and Theorem 4 (and dropping  $W$  that does not appear here), we have:

$$\mu(x) = x, \tau(z) = -\frac{\rho z}{1 - \rho^2},$$

$$g(t) = \frac{1}{\sqrt{1 - \rho^2}} \exp\left(-\frac{t^2(1 - \rho^2)}{2}\right), h(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1 - \rho^2)}\right).$$

Remember that  $\nu(x) = \mathbb{E}(g(\tau(Z)) \exp(-x\tau(Z))) = \frac{1}{\sqrt{2\pi(1 - \rho^2)}} \int \exp\left(\frac{x\rho z}{1 - \rho^2} - \frac{z^2}{2(1 - \rho^2)}\right) dz$ , and next a bit of algebra gives:

$$\nu(x) = \exp\left(\frac{x^2 \rho^2}{2(1 - \rho^2)}\right).$$

Then for a parameter  $\theta_0 = \mathbb{E}(\psi(X)m^{(j)}(X))$ , we have to check that the following function  $\varrho(x)$  is a Laplace transform:

$$\varrho(x) = (-1)^j \left[ \psi(x) \exp\left(-\frac{x^2}{2}\right) \right]^{(j)} \exp\left(\frac{x^2}{2(1 - \rho^2)}\right).$$

1. Shift in mean: If  $\theta_0 = \mathbb{E}(\psi(X)m(X))$  with  $\psi(x) = h\left(\frac{\exp(-(x^2/2))}{\sqrt{2\pi}}, x\right)$  with  $h(y, x) = \frac{1}{\sqrt{2\pi}} \frac{e^{-(x-m)^2/2}}{y}$ . Note that choosing  $h$  like that means that we are interesting by the average effect under a counterfactual distribution  $\mathcal{N}(m, 1)$  in a case where  $f_X$  is unknown and needs to be estimated. In that case,  $R(X) = (\psi(X) + \partial_{f_X} \psi(X) \cdot f_X(X))m(X) - \mathbb{E}((\psi(X) + \partial_{f_X} \psi(X) \cdot f_X(X))m(X)) = 0$ . A less realistic problem

could also be considered, assuming that  $f_X$  is known by econometrician, in that case  $h$  would take the form  $h(y, x) = \exp\left(-\frac{(x-m)^2 - x^2}{2}\right)$  and even if the parameter is the same, the expression of the semi parametric efficiency bound would be different because in that case  $R(X) = \psi(X)m'(X) - \mathbb{E}(\psi(X)m'(X))$ .

In that case,

$$\varrho(x) = \exp\left(-\frac{(x-m)^2}{2} + \frac{x^2}{2(1-\rho^2)}\right) = \exp\left(-\frac{m^2}{2}\right) \exp\left(\frac{x^2\rho^2}{2(1-\rho^2)} - xm\right).$$

For  $\rho \neq 0$ , usual formula of Laplace transform ensures that:

$$\varrho(x) = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}|\rho|} \exp\left(-\frac{m^2}{2}\right) \mathcal{L}\left[\exp\left(-\frac{(z-m)^2(1-\rho^2)}{2\rho^2}\right)\right],$$

and next, if  $\mathcal{L}^{-1}[\varrho]$  denotes the measure that is the inverse Laplace of  $\varrho(x)$ :

$$d\mathcal{L}^{-1}[\varrho] = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}|\rho|} \exp\left(-\frac{m^2}{2}\right) \exp\left(-\frac{(t-m)^2(1-\rho^2)}{2\rho^2}\right) dt.$$

On the other side:

$$g(t)dF_{\tau(Z)}(t) = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}|\rho|} \exp\left(-\frac{t^2(1-\rho^2)}{2\rho^2}\right) dt,$$

so, we obtain:

$$\frac{1}{g(t)} \frac{d\mathcal{L}^{-1}[\varrho(x)]}{dF_{\tau(Z)}}(t) = \frac{\sqrt{2\pi}(1-\rho^2)^{3/2}}{|\rho|} \exp\left(-\frac{m^2}{2\rho^2}\right) \exp\left(\frac{tm(1-\rho^2)}{\rho^2}\right).$$

Because  $\tau(Z)$  is one to one,  $\mathbb{E}(\varepsilon^2|Z) = \mathbb{E}(\varepsilon^2|\tau(Z))$  and  $\mathbb{E}(\varepsilon R(X)|Z) = \mathbb{E}(\varepsilon R(X)|\tau(Z))$ .

It follows that the semi-parametric efficiency bound of  $\theta_0$  is:

$$\mathbb{V}(\varepsilon q^*(Z)).$$

With  $q^*(Z) = \exp\left(\frac{mZ}{\rho} - \frac{m^2}{2\rho^2}\right)$ .

2. Shift in variance: If  $\theta_0 = \mathbb{E}(\psi(X)m(X))$  with  $\psi(x) = h\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, x\right)$  with  $h(y, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{e^{-x^2/(2\sigma^2)}}{y}$ . Note that choosing  $h$  like that means that we are interesting by the average effect under a counterfactual distribution  $\mathcal{N}(0, \sigma^2)$  in a case where  $f_X$  is unknown and needs to be estimated and next  $R(X) = 0$ . We have  $\psi(x) =$

$(\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - 1\right)\right)$ , then

$$\varrho(x) = (\sigma^2)^{-1/2} \exp\left(-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - \frac{1}{1-\rho^2}\right)\right).$$

Because  $\Gamma = \mathbb{R}$ , it follows that conditions (iii) and (iv) hold only if  $\frac{1}{\sigma^2} - \frac{1}{1-\rho^2} < 0$ , ie only if  $\sigma^2 > 1 - \rho^2$ . Reciprocally, if  $\sigma^2 > 1 - \rho^2$ , we have:

$$\varrho(x) = \frac{\sqrt{(1-\rho^2)}}{\sqrt{2\pi(\sigma^2 + \rho^2 - 1)}} \mathcal{L}\left[\exp\left(-\frac{z^2\sigma^2(1-\rho^2)}{2(\sigma^2 + \rho^2 - 1)}\right)\right],$$

and in this case if  $\mathcal{L}^{-1}[\varrho]$  denotes the measure that is the inverse Laplace of  $\varrho(x)$ :

$$d\mathcal{L}^{-1}[\varrho](t) = \frac{\sqrt{(1-\rho^2)}}{\sqrt{2\pi(\sigma^2 + \rho^2 - 1)}} \exp\left(-\frac{t^2\sigma^2(1-\rho^2)}{2(\sigma^2 + \rho^2 - 1)}\right) dt.$$

On the other side:

$$g(t) dF_{\tau(Z)}(t) = \frac{\sqrt{1-\rho^2}}{\sqrt{2\pi}|\rho|} \exp\left(-\frac{t^2(1-\rho^2)}{2\rho^2}\right) dt,$$

then:

$$\frac{1}{g(t)} \frac{d\mathcal{L}^{-1}[\varrho(x)]}{dF_{\tau(Z)}}(t) = \frac{|\rho|}{\sqrt{\sigma^2 + \rho^2 - 1}} \exp\left(\frac{t^2(1-\rho^2)^2(\sigma^2 - 1)}{2\rho^2(\sigma^2 + \rho^2 - 1)}\right).$$

It follows that the semi-parametric efficiency bound of  $\theta_0$  is:

$$\frac{\rho^2}{\sigma^2 + \rho^2 - 1} \mathbb{V}\left(\varepsilon \exp\left(\frac{Z^2(\sigma^2 - 1)}{2(\sigma^2 + \rho^2 - 1)}\right)\right).$$

3. Average marginal effect: If  $\theta_0 = \mathbb{E}(m'(X))$ , then if  $q \in \mathcal{Q}$ , it follows that  $\mathbb{E}[q(Z)|X] = X$  and then  $q(Z) = \frac{Z}{\rho}$  and the semi parametric efficiency bound is:

$$\mathbb{V}\left((Y - m(X))\frac{Z}{\rho} + m'(X)\right)$$

4. average derivative weighted by power of density: Let  $\theta_0 = \mathbb{E}(\psi(X)m'(X))$  with  $\psi(X) = h((2\pi)^{-1/2} \exp(-x^2/2), x)$ , with  $h(y, x) = y^k$ , for  $k = 0$  we get the average marginal effect and for  $k = 1$  we get the density-weighted average derivative.  $R(X) = \frac{k+1}{(2\pi)^{k/2}} \exp\left(-\frac{kx^2}{2}\right) m'(X) - \mathbb{E}\left(\frac{k+1}{(2\pi)^{k/2}} \exp\left(-\frac{kx^2}{2}\right) m'(X)\right)$ . The function  $\varrho$  is:

$$\varrho(x) = (k+1)x \exp\left(\frac{x^2}{2} \left(\frac{1}{1-\rho^2} - (k+1)\right)\right),$$

which is not the Laplace transform of a function if  $\left(\frac{1}{1-\rho^2} - (k+1)\right) < 0$ . Then the parameter  $\theta_0$  is root-N estimable only if  $\rho^2 \geq \frac{k}{k+1}$ . For  $\rho^2 > \frac{k}{k+1}$ , a bit of algebra ensure that the semi-parametric efficiency bound is:

$$(k+1)^2 \mathbb{V} \left( \varepsilon Z \exp \left( Z^2 \frac{1-\rho^2}{(k+1)\rho^2 - k} \right) + \frac{1}{\sqrt{2\pi}^k} \exp(-kX^2/2) m'(X) \right).$$

And for  $\rho^2 = \frac{k}{k+1}$  and  $k \geq 1$ , we get:

$$(k+1)^2 \mathbb{V} \left( \varepsilon \frac{Z}{\rho} + \frac{1}{\sqrt{2\pi}^k} \exp(-kX^2/2) m'(X) \right).$$

Last, if  $k = 0$  and  $\rho^2 = 0$   $\mathcal{Q}$  is empty.

## References

- Ai, C. & Chen, X. (2003), ‘Efficient estimation of models with conditional moment restrictions containing unknown functions’, *Econometrica* **71**, 1795–1843.
- Ai, C. & Chen, X. (2007), ‘Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables’, *Journal of Econometrics* **141**, 5–43.
- Ai, C. & Chen, X. (2012), ‘The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions’, *Journal of Econometrics* **170**, 442–457.
- Bickel, P. J., Klaassen, C. A., Ritov, Y. & Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer-Verlag New York. ISBN 0-387-98473-9.
- Blundell, R., Chen, X. & Kristensen, D. (2007), ‘Semi-nonparametric iv estimation of shape-invariant engel curves’, *Econometrica* **75**, 1613–1669.
- Breunig, C. & Johannes, J. (2016), ‘Adaptive estimation of functionals in nonparametric instrumental regression’, *Econometric Theory* **32**.
- Bronstein, E. M. (2008), ‘Approximation of convex sets by polytopes’, *Journal of Mathematical Sciences* **153**(6), 727–762.
- Canay, I. A., Santos, A. & Shaikh, A. M. (2013), ‘On the testability of identification in some nonparametric models with endogeneity’, *Econometrica* **81**(6), 2535–2559.
- Carrasco, M., Florens, J. & Renault, E. (2007), Linear inverse problems and structural econometrics: Estimation based on spectral decomposition and regularization, in J. Heckman & E. Leamer, eds, ‘Handbook of Econometrics’, Vol. 6B, North-Holland, pp. 5633–5751.
- Carrasco, M., Florens, J. & Renault, E. (2014), Asymptotic normal inference in linear inverse problems, in J. Racine, L. Su & A. Ullah, eds, ‘The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics’, Oxford University Press, pp. 65–96.
- Cattaneo, M. D., Crump, R. K. & Jansson, M. (2014), ‘Small bandwidth asymptotics for density-weighted average derivatives’, *Econometric Theory* **30**, 176–200.
- Chamberlain, G. (1987), ‘Asymptotic efficiency in estimation with conditional moment restrictions’, *Journal of Econometrics* **34**, 305–334.

- Chen, X. (2007), Large sample sieve estimation of semi-nonparametric models, *in* J. Heckman & E. Leamer, eds, ‘Handbook of Econometrics’, Vol. 6B, North-Holland, pp. 5549–5632.
- Chen, X. & Pouzo, D. (2012), ‘Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals’, *Econometrica* **80**(1), 277–321.
- Chen, X. & Pouzo, D. (2015), ‘Sieve wald and qlr inferences on semi/nonparametric conditional moment models’, *Econometrica* **83**(3), 1013–1079.
- Chetverikov, D. & Wilhelm, D. (2015), ‘Nonparametric instrumental variable estimation under monotonicity’, *CEMMAP working paper* (39).
- Choi, S., Hall, W. J. & Schick, A. (1996), ‘Asymptotically uniformly most powerful tests in parametric and semiparametric models’, *The Annals of Statistics* **24**(2), 841–861.
- Darolles, S., Fan, Y., Florens, J. & Renault, E. (2011), ‘Nonparametric instrumental regression’, *Econometrica* **79**(5), 1541–1565.
- De La Vallée Poussin, C.-J. (1915), ‘Sur l’intégrale de lebesgue’, *Transactions of the American Mathematical Society* **16**(4), 435–501.
- d’Haultfoeuille, X. (2011), ‘On the completeness condition in nonparametric instrumental problems’, *Econometric Theory* **27**(03), 460–471.
- Florens, J. (2003), Inverse problems and structural econometrics: The example of instrumental variables, *in* R. Dewatripont, L. Hansen & S. Turnovsky, eds, ‘Advances in Economics and Econometrics: Theory and Applications’, Cambridge University Press, pp. 284–311.
- Florens, J., Johannes, J. & Bellegem, S. V. (2011), ‘Identification and estimation by penalization in nonparametric instrumental regression’, *Econometric Theory* **27**, 472–496.
- Florens, J., Johannes, J. & Bellegem, S. V. (2012), ‘Instrumental regression in partially linear models’, *Econometric Journal* **15**, 304–324.
- Florens, J. & Simoni, A. (2012), ‘Nonparametric estimation of an instrumental variables regression: A quasi bayesian approach based on a regularized posterior’, *Journal of Econometrics* **170**, 458–475.
- Freyberger, J. (2015), On completeness and consistency in nonparametric instrumental variable models.  
webpage: [http://www.ssc.wisc.edu/~jfreyberger/Completeness\\_Freyberger](http://www.ssc.wisc.edu/~jfreyberger/Completeness_Freyberger).



- Gourieroux, C., Montfort, A. & Trognon, A. (1984), ‘Pseudo maximum likelihood methods: Theory’, *Econometrica* **52**(03), 681–700.
- Groetsch, C. W. (1977), *Generalized Inverse of Linear Operators, Representation and Approximation*, first edn, MARCEL DEKKER. ISBN 0-8247-6615-6.
- Hall, P. & Horowitz, J. (2005), ‘Nonparametric methods for inference in the presence of instrumental variables’, *The Annals of Statistics* **33**, 2904–2929.
- Hiriart-Urruty, J.-B. & Lemaréchal, C. (2001), *Fundamentals of Convex Analysis*, first edn, Springer. ISBN 3-540-42205-6.
- Hu, Y. & Schennach, S. M. (2008), ‘Instrumental variable treatment of nonclassical measurement error models’, *Econometrica* **76**(1), pp. 195–216.
- Jacod, J. & Protter, P. (2000), *Probability Essentials*, Springer Berlin Heidelberg. ISBN 3-540-66419-X.
- Newey, W. K. (1988), ‘Adaptative estimation of regression models via moment restrictions’, *Journal of Econometrics* **38**, 301–339.
- Newey, W. K. (1990a), ‘Efficient instrumental variables estimation of nonlinear models’, *Econometrica* **58**(4), 809–837.
- Newey, W. K. (1990b), ‘Semiparametric efficiency bound’, *Journal of Applied Econometrics* **5**(2), 99–135.
- Newey, W. K. & Powell, J. L. (2003), ‘Instrumental variable estimation of nonparametric models’, *Econometrica* **71**(5), 1565–1578.
- Newey, W. K. & Stoker, T. M. (1993), ‘Efficiency of weighted average derivative estimators and index models’, *Econometrica* **61**(5), 1199–1223.
- Powell, J. L., Stock, J. H. & Stoker, T. M. (1989), ‘Semiparametric estimation of index coefficients’, *Econometrica* **57**(6), 1403–1430.
- Rudin, W. (1987), *Real and Complex Analysis*, third edn, McGraw-Hill. ISBN 0-07-054234-1.
- Santos, A. (2011), ‘Instrumental variables methods for recovering continuous linear functionals’, *Journal of Econometrics* **161**, 129–146.
- Schwartz, L. (1997), *Théorie des distributions*, third edn, Hermann. ISBN 2-7056-5551-4.

Severini, T. A. & Tripathi, G. (2012*a*), ‘Semiparametric efficiency bounds for microeconomic models: A survey’, *Foundations and Trends<sup>®</sup> in Econometrics* **6**(3-4), 163–397.

Severini, T. & Tripathi, G. (2012*b*), ‘Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors’, *Journal of Econometrics* **170**, 491–498.

van der Vaart, A. W. (1991), ‘On differentiable functionals’, *Annals of Statistics* **19**(1), 178–204.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press. ISBN 0-521-78450-6.