Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2014-27

Regression Discontinuity Design with Continuous Measurement Error in the Running Variable

L. DAVEZIES[1]
T. LE BARBANCHON[2]

1 CREST. laurent.davezies@ensae.fr

2 CREST. (*Address correspondence* : 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France) thomas.le-barbanchon@ensae.fr Tel. : +33 623786708 +33 (0)141176035 -

Regression Discontinuity Design with Continuous Measurement Error in the Running Variable *

Laurent Davezies[†] Thomas Le Barbanchon[‡]

August 4, 2014

Abstract

Since the late 90s, regression discontinuity designs have been widely used to estimate local treatment effects. When the running variable is observed with continuous errors, identification fails even if the dispersion of measurement errors is small. Assuming non-differential measurement errors, we propose a consistent nonparametric estimator of the LATE when the true running variable is observed in an auxiliary sample of treated individuals. Such auxiliary information is usually collected by agencies in charge of delivering the treatment.

JEL codes: C21, C14, C51

Keywords: regression discontinuity design, measurement error

*We thank seminar participants at Crest for their helpful comments. We are especially grateful to Elise Coudin, Yingyao Dong and Xavier d'Haultfoeuille for comments on our research. We thank Benjamin Walter for excellent research assistance.

[†]CREST, laurent.davezies@ensae.fr

[‡]CREST. Address correspondence to thomas.le-barbanchon@ensae.fr, 15 Boulevard Gabriel Péri 92245 Malakoff Cedex France - Tel.: +33623786708 - Fax.:+33141176035

1 Introduction

The regression discontinuity (RD) design has been widely emphasized for its internal validity to estimate local treatment effects and has been widely used since the late 1990s (Imbens & Lemieux, 2008, Lee & Lemieux, 2010). Such a method relies on the fact that assignment to the treatment T is determined (at least partially) by a continuous running variable Z being on either side of a fixed threshold. If the joint distribution of the realized outcome Y, the treatment T and the running variable Z is observed, then a local treatment effect is identified (Hahn et al., 2001). However, for many practical reasons, variables are often subject to measurement errors, especially with survey data (Bound et al., 2001). In RD designs, smoothly distributed measurement errors in Z have drastic consequences : even with a small dispersion of measurement errors, the discontinuity in the assignment probability vanishes (see for example Hullegie & Klein, 2010 or Cahuc et al., 2014). Hence, usual RD estimators using the running variable observed with measurement error are inconsistent. The problem is similar to that of IVs with weak instruments.

In this paper, we show that the identification of the treatment effect can be recovered, when auxiliary information about the distribution of the true running variable is observed. More precisely, we consider the typical case when the econometrician observes the discrepancy between the running variable and its noisy measure in a sample of treated individuals. This happens when the agency in charge of delivering the treatment collects data on treated individuals, specifically their eligibility (the true running variable). We can restore identification by assuming that conditional on the true running variable, its noisy measure is independent of treatment and potential outcomes, i.e. the measurement error is non-differential (see Bound et al., 2001). This allows us to use the structure of the error identified on treated individuals to infer the true running variable for non-treated individuals. Fuzzy RD designs with treated individuals on both sides of the threshold are particularly adapted to our approach. In that case, the econometrician observes the error structure on the whole support of the true running variable. When there are no treated individuals below the threshold, supplementary assumptions on the error structure are needed. Under classical measurement error, we can ensure nonparametric identification using a deconvolution strategy (Schennach, 2004, Hu & Ridder, 2012).

To give an intuition of our approach in the fuzzy RD design, let us denote Z^* the true running variable, T the treatment and Z a noisy measure of Z^* such that $Z \perp\!\!\!\perp T|Z^*$. If the distribution of (T, Z, TZ^*) is identified by the data, following D'Haultfœuille (2010), the assignment probability $p(Z^*) = \mathbb{E}(T|Z^*)$ is identified by the conditional moment $\mathbb{E}\left([p(Z^*)]^{-1}|T = 1, Z\right) = [\mathbb{E}(T|Z)]^{-1}$. This conditional moment condition can be written in a form similar to usual nonparametric IV conditions: $\mathbb{E}(T/p(Z^*) - 1|Z) = 0$. We thus adapt sieve estimation strategies from the nonparametric IV literature (Ai & Chen, 2003, Newey & Powel, 2003, Chen, 2007, Chen & Pouzo, 2012, Peter & Joel, 2005, Darolles et al., 2011). More precisely, as estimating the treatment effect in the RD design requires pointwise convergence of the estimator $\hat{p}(.)$ of $p(.)$ at the threshold value, we follow strategies that ensure uniform convergence (Ai & Chen, 2003, Chen, 2007 and Chen & Pouzo, 2012). We adapt the regularity conditions of usual nonparametric IV estimation to our framework. First, we take into account the discontinuity of the take-up $p(.)$ at the threshold. Second, we relax usual conditions on the density of the variable Z to cope with the fact that measurement error usually vanishes at the boundary of its support.[1] We show that under those relaxed conditions, our sieve estimator is consistent. Monte-Carlo simulations show that our estimator outperforms naive RD estimator that ignores the measurement error.

Our paper relates to the burgeoning literature on measurement issues in the running variable of RD designs. Most of the theoretical RD literature focuses on discreteness or rounding error. Lee & Card (2008) show that, in RD designs with a discrete running variable, researchers need to account for specification errors in the model. This affects the precision of the estimated treatment effect. However Lee & Card (2008) do not really consider that

[1]In the usual nonparametric IV framework, the density of the instrument Z is assumed to be bounded away from zero on its support (see for example Ai & Chen, 2003 or Chen & Pouzo, 2012).

discreteness of the observed variable is the result of measurement error of a true continuous underlying variable.

More directly related to our paper is thus the contribution of Dong (2014) which explicitly addresses issues raised by rounding errors in a parametric setting. Such errors typically prevent researchers from observing individuals just above and just below the threshold, and the source of identification in the RD design is lost. To restore identification, Dong (2014) uses auxiliary information on the distribution of the true running variable, as we do. In addition, her approach makes use of parametric assumptions and of the deterministic relation between rounding errors and the true running variable. In our paper, the precise form of measurement error is a priori unknown and our identification strategy is nonparametric.

Another close contribution is Pei (2011), which assumes classical measurement error in sharp or one-sided fuzzy RD designs. He further assumes that both the observed and the true running variables are discrete and have bounded support. Then, when there is no treatment on one side of the threshold, the true distribution of the running variable is identified (using a deconvolution argument). In his context, the assignment status (which is observed without error) informs about the position of the true running variable with respect to the threshold. In our paper, auxiliary information on the measurement error helps us to deal with continuous running variables and two-sided fuzzy RD designs.

When confronted with continuous measurement errors, applied researchers tend to adopt a fully parametric approach. For example, Hullegie & Klein (2010) estimate the impact of private health insurance on expenditures and health. They clearly illustrate that the discontinuity in the assignment probability disappears when their running variable (which is income) is measured in a survey. To recover identification, they make parametric assumptions about the relation between the potential outcomes and the true running variable and assume that income is measured with Berkson type error (see Wansbeek & Meijer, 2000, Section 2.5 for a simple exposition of Berkson's model or the original paper of Berkson, 1950). Our contribution is to show how ad-hoc parametric specifications can be abandoned provided that auxiliary information is available.

Our paper also complements other applied RD contributions that focus on *contaminated* or *corrupted* data sampling, in which the observed running variable Z is a mixture of the true running variable and of a noisy proxy (i.e. $\mathbb{P}(Z = Z^*) > 0$, see Horowitz & Manski, 1995). In a paper that investigates the causal effect of retirement on consumption with *contaminated* data, Battistin et al. (2009) assume that the measurement error is non-differential (i.e. $Z \perp\!\!\!\perp (Y(0), Y(1), T)|Z^*$) and that $z \mapsto \mathbb{P}(Z = Z^*|Z = z)$ is continuous at the threshold, and thus recover the parameter of interest. However, this last assumption of continuity can be violated when rounding errors generate heaping patterns in the data. In the case of heaped data, Barreca, Guldi, Lindo & Waddell (2011) and Barreca, Lindo & Waddell (2011) propose to remove observations at heaping values and estimate the treatment effect in the decontaminated sample (this approach is called Donut-RD). Our paper addresses a different problem where there is no mass of observations with correct values (i.e. $\mathbb{P}(Z = Z^*) = 0$). Then there is no useable information on the true running variable in the main sample. This difficulty can be circumvented with the observation of auxiliary information on the treated individuals.

The paper is organized as follows: in the second section, we show in detail how a measurement error in the running variable Z smoothes any discontinuity in assignment and leads to the loss of identification. In the third section, we present and discuss our identification results. We distinguish two cases depending on the support of $T|Z^*$. In the case (a), when there are treated individuals below and above the true threshold, we only assume that the measurement error is non-differential. In the case (b), when there are no treated individuals below the threshold, we further assume that the measurement error is classical. The fourth section is devoted to estimation issues in the more general case (a), we then provide Monte-Carlo simulations to investigate the finite sample behavior of our estimator. The last section concludes.

2 Framework

2.1 Regression Discontinuity design

Let T be a binary variable of treatment, Z^* a continuous random variable with support $\mathcal{Z} \subset \mathbb{R}$ and 0 an interior point of $\mathcal{Z}$, 0 is the cutoff of the Regression Discontinuity (RD) design. Following Rubin's framework, we define $(Y(0), Y(1))$ the potential outcomes with respect to T and $Y = Y(0)(1-T)+Y(1)T$ is the observed outcome. The usual assumptions of RD design are as follows:

Assumption 1 (RD Design)

1. $\lim_{z\to 0^+} \mathbb{E}\left(T|Z^* = z\right) > \lim_{z\to 0^-} \mathbb{E}\left(T|Z^* = z\right)$
2. *It exists $\mathcal{Z}_0$ a neighborhood of 0, such that almost surely it exists an increasing random function τ from $\mathcal{Z}_0$ to $\{0;1\}$ such that $\tau(\omega)(Z^*(\omega)) = T(\omega)$ for all $\omega \in Z^{*-1}(\mathcal{Z}_0)$.*
3. *$z \mapsto \mathbb{E}(Y(t)|Z^* = z, \tau(0^+), \tau(0^-))$ is continuous (almost-surely) for $t = 0, 1$.*

Assumption 1.1 states that the assignment probability (or take-up) is discontinuous at the cutoff 0, which is known by the econometrician. Assumption 1.2 is a form of monotonicity condition. It rules out the existence of defiers, individuals who would abandon the treatment had they crossed the cutoff. Assumption 1.3 states that the conditional expectations of the potential outcomes are continuous at the cutoff. Under Assumption 1, it is well known that the Local Average Treatment Effect (LATE)

$$\theta = \mathbb{E}\left[Y(1) - Y(0)|Z^* = 0, \tau(0^+) > \tau(0^-)\right] \tag{2.1}$$

is equal to a Wald's ratio $\frac{\mathbb{E}(Y|Z^*=0^+)-\mathbb{E}(Y|Z^*=0^-)}{\mathbb{E}(T|Z^*=0^+)-\mathbb{E}(T|Z^*=0^-)}$ (see Hahn et al., 2001).

It follows that if the joint distributions of (Y, Z^*) and (T, Z^*) are identified by the observation of a large number of independent realizations, θ is identified. Estimation in such contexts relies on consistent estimators of the four quantities $\mathbb{E}(Y|Z^* = 0^+)$, $\mathbb{E}(Y|Z^* = 0^-)$, $\mathbb{E}(T|Z^* = 0^+)$, and $\mathbb{E}(T|Z^* = 0^-)$. Our general framework is less favorable: we observe a proxy variable of Z^*.

2.2 Measurement error

We denote Z the noisy measure of the true running variable Z^*. We adopt the following assumptions on the measurement error generating process:

Assumption 2 (Measurement error)

1. *(Z, Z^*) admits a density with respect to the Lebesgue measure in $\mathbb{R}^2$.*

2. *The measurement error is non-differential: $Z \perp\!\!\!\perp (T, Y(0), Y(1))|Z^*$.*

The first assumption describes "continuous" measurement error. Specifically it rules out rounding errors and *contaminated* data. The second assumption characterizes non-differential measurement error (Bound et al., 2001). It states that the noisy measure does not yield any supplementary information on the variables of interest, once we condition on the true running variable. Classical measurement error verifies those assumptions. More generally, any *transformation* of a classical measurement error model verifies those assumptions. (Z, Z^*) is a *transformation* of a classical measurement error model if there exist μ and ν increasing C^1 diffeomorphisms with derivatives bounded away from zero on $\text{Supp}(Z)$ and $\text{Supp}(Z^*)$ such that $\mu(Z)$ is a classical measurement error of $\nu(Z^*)$. A direct consequence of the *transformation* definition is that the difference $\mu(Z) - \nu(Z^*)$ is independent of Z^*. A multiplicative error such that $Z = Z^* \times \varepsilon$ is a *transformation* model.

2.3 Loss of identification

The following proposition shows that if instead of Z^* we observe Z a noisy measure of Z^*, the probability to be treated is a continuous function of Z. As a consequence, the usual framework of the RD design fails to identify the LATE θ.

Proposition 2.1 (Continuity of the take-up)
Under Assumptions 1 and 2, $z \mapsto \mathbb{E}(T|Z = z)$ is a continuous function on the interior of the support of Z if one of the following condition holds:

1. $z \mapsto f_{Z|Z^*=z^*}(z)$ *is continuous* z^**-almost-everywhere and* $\mathbb{E}\left(\sup_z f_{Z|Z^*}(z)\right) < \infty$.

2. (Z, Z^*) *is a transformation of a classical measurement error model and the density of* Z^* *is bounded.*

The first part of Proposition 2.1 states that under mild conditions on the measurement error, the take-up $z \mapsto \mathbb{E}(T|Z = z)$ is continuous. In particular, there is no discontinuity at the RD design cutoff. The denominator of the usual Wald's ratio $\frac{\mathbb{E}(Y|Z=0^+)-\mathbb{E}(Y|Z=0^-)}{\mathbb{E}(T|Z=0^+)-\mathbb{E}(T|Z=0^-)}$, defined on the observed noisy running variable, is null. The second part of Proposition 2.1 considers the specific case of a *transformation* of a classical measurement error. In such a case, the boundedness of f_{Z^*} ensures the continuity of the take-up. No extra assumption on the density of the error is needed.[2]

Proposition 2.1 shows the continuity of the take-up. Similar arguments can be used to show the continuity of the outcome with respect to the noisy running variable. As a consequence, the numerator of the usual Wald's ratio is also null. We discuss below the implications of those continuity results for the naive estimation of the Wald ratio.

Let K be a symmetric kernel function, and h_n a decreasing sequence tending to 0. A popular estimation of the Wald's ratio in the RD framework is based on local linear regression:

$$\frac{a_Y^+ - a_Y^-}{a_T^+ - a_T^-}, \text{ with } a_U^\pm = \arg\min_\alpha \min_\beta \sum_{i=1}^n (U_i - \alpha - \beta Z_i^*)^2 K\left(\frac{Z_i^*}{h_n}\right) \mathbb{1}\{Z_i^* \in \mathbb{R}^\pm\} \qquad (2.2)$$

For a given value of h_n, this estimator boils down to the weighted two stage least square of Y on T with covariates $Z^* \times \mathbb{1}\{Z^* \in \mathbb{R}^+\}$ and $Z^* \times \mathbb{1}\{Z^* \in \mathbb{R}^-\}$, and excluded instrument $\mathbb{1}\{Z^* \in \mathbb{R}^+\}$. This estimator is widely popular, because the usual inference is valid in such a semi-parametric model (Hahn et al., 2001). However, when only a noisy measure of Z is available, the naive adaption of such an estimator, replacing Z^* by Z in Equation 2.2, leads to dramatic results.

[2]Note that in the case of a transformation model, the result holds even if f_{Z^*} is not bounded provided that the error density is bounded. Then the first condition of the proposition is verified. To illustrate this point, let us consider the case of classical measurement error. Then $f_{Z|Z^*}(z) = f_\varepsilon(z - z^*)$ is bounded if f_ε is bounded.

Proposition 2.2 *Let (Z, Z^*) be such that one of the two following conditions holds:*

1. *$z \mapsto f_{Z|Z^*=z^*}(z)$ is twice continuously differentiable z^*-almost-everywhere and $\mathbb{E}\left(\sup_z f^{(j)}_{Z|Z^*}(z)\right) < \infty$ for $j = 1, 2$.*

2. *(Z, Z^*) is a transformation of a classical measurement error model and the density of Z^* is twice continuously differentiable.*

Moreover let us assume that it exists $\delta > 2$ such that $\mathbb{E}\left(|Y|^\delta\right) < \infty$, that $z^* \mapsto \mathbb{E}(Y^2|Z^* = z^*)$ is bounded and that Assumptions 1 and 2 hold. Then for $h_n \sim n^{-1/5}$ and for any K bounded, symmetric and nonnegative-valued kernel with compact support,

$$\widehat{\theta}^{naive}_{LLR} = \frac{a_Y^+ - a_Y^-}{a_T^+ - a_T^-}, \text{ with } a_U^\pm = \arg\min_\alpha \min_\beta \sum_{i=1}^n (U_i - \alpha - \beta Z_i)^2 K\left(\frac{Z_i}{h_n}\right) \mathbb{1}\{Z_i \in \mathbb{R}^\pm\}$$

tends in distribution to a Cauchy of location $\frac{\mathbb{C}ov(Y,T|Z=0)}{\mathbb{V}(T|Z=0)}$ and scale $\left(\frac{\mathbb{V}(Y|Z=0)}{\mathbb{V}(T|Z=0)} - \frac{\mathbb{C}ov^2(Y,T|Z=0)}{\mathbb{V}^2(T|Z=0)}\right)^{1/2}$.

The two first conditions in Proposition 2.2 reinforce the conditions of Proposition 2.1. This ensures that $\mathbb{E}(T|Z)$ are twice continuously differentiable. Associated with the boundedness of $\mathbb{E}(Y^2|Z^*)$, this also ensures also that $\mathbb{E}(Y|Z)$, $\mathbb{E}(Y^2|Z)$ and $\mathbb{V}(Y|Z)$ are twice differentiable. The condition $\mathbb{E}(|Y|^\delta) < \infty$ is mild but allows us to apply the Lyapounov's Central Limit Theorem to derive the asymptotic properties of the estimator. If K is not symmetric, the limit distribution is no more a Cauchy but a ratio of normal with non null expectation. The assumption on the support of K is made for simplicity but can also be relaxed with simple conditions on the tails of K.

The main message of Proposition 2.2 is that the naive estimator does not converge to θ_0 (and nor to any value!). The situation is similar to what happens in the two-stage least squares with completely uninformative instruments. In that case, the IV estimator is also inconsistent.

Finally, under the assumptions of Proposition 2.1, the marginal density of the noisy running variable Z is continuous (see proof in the Appendix). This means that the McCrary test of non-manipulation of the running variable is never rejected, when there is measurement error in the running variable. Similarly any tests of continuity of the covariates are not rejected.[3]

3 Identification with auxiliary information

To recover the identification of the LATE in the presence of measurement errors in the running variable, we rely on an auxiliary sample of treated individuals, for whom we observe the true running variable Z^*. This naturally occurs when individuals apply to an independent agency in order to be treated and then, declare their running variable on their application form. In such a context, it is very likely that the agency in charge of the treatment checks the eligibility conditions and keeps a record of the correct running variable for the treated. Many programs, which could be evaluated in RD designs, feature this institutional process: means-tested treatment as in Hullegie & Klein (2010), conditional subsidies to firms as in Cahuc et al. (2014), etc.

In the following, we distinguish two types of RD designs, depending on the support of the score $\mathbb{P}(T = 1|Z^*)$. In the first subsection, we consider two-sided fuzzy designs, i.e $\mathbb{P}(T = 1|Z^*) > 0$ almost-surely. In the second subsection, we consider the case when individuals below the cutoff cannot apply for treatment ($\mathbb{P}(T = 1|Z^*) = 0$ with a positive probability).

3.1 RD Design with support condition on the score

Assumption 3 (Observation from the data)

We observe an iid sample S of (Y, T, Z) with n observations and an iid auxiliary sample S_a of $(Z, Z^*)|T = 1$ with n_a observations.

Assumption 3 holds when we observe an iid sample of (Y, Z, T, TZ^*), i.e. when S_a is a

[3] This is true as long as the measurement error is non-differential in the most general terms: for any variable X, $Z \perp\!\!\!\perp X|Z^*$

subset of S and when we can match the two samples. However Assumption 3 is more general: samples may not be nested or cannot be matched.

Assumption 4 (Completeness Condition)
$\forall g$ such that $\mathbb{E}(|g(Z^*)|) < +\infty$, $\mathbb{E}\left(g(Z^*)|Z\right) = 0 \Rightarrow g = 0$.

Assumption 4 is equivalent to the rank condition when Z^* and Z have finite support: $rk[\mathbb{P}(Z^* = i|T = 1, Z = j)]_{i=1,\dots,I,j=1,\dots,J} = I$ where I is the dimension of the support of Z^*. Intuitively, this means that there is enough variation in Z to identify $g \in L^1(Z^*)$ when we observe $\mathbb{E}(g(Z^*)|Z)$. Assumption 4 is usual in the nonparametric IV framework. Examples of data generating processes such that the completeness condition holds can be found in Newey & Powel (2003) or D'Haultfœuille (2011). Interestingly, D'Haultfœuille (2011) shows that in the context of a transformation model, the completeness condition holds for all the usual classes of parametric distributions of the measurement error (as soon as the error characteristic function has isolated zeros). For estimation issues, we will further assume that $\mathbb{E}(Y|Z^*)$ is bounded and that there exists $\underline{c} > 0$ such that $\underline{c} < P(T = 1|Z^*)$, then the completeness condition may be replaced by the weaker bounded completeness condition.[4]
We now state the main result of identification when the support condition on the score is verified.

Theorem 3.1 (Identification of $F_{Z^*,Z,T,Y}$ and θ)
Under Assumptions 1, 2, 3 and 4, when the support condition on the score is verified, i.e. $\mathbb{P}(T = 1|Z^*) > 0$ Z^*-almost-surely, the joint distribution of Z^*, Z, T, Y and the LATE (θ) are identified.

The complete proof of the identification of the joint distribution (Z^*, Z, T, Y) and of the LATE is reported in the Appendix. We now briefly give some intuition about the identification of θ.

[4]The bounded completeness condition holds if the implication is true for any bounded function g(.).

To identify θ, we need to identify $\mathbb{P}(T=1|Z^*=z^*)=p(z^*)$ and $\mathbb{E}(Y|Z^*=z^*)=m(z^*)$ in the neighborhood of $z^*=0$. Under Assumption 2 and the support condition on the score, m and p are solutions of the following moment conditions:

$$\mathbb{E}\left(\frac{1}{p(Z^*)}|T=1,Z\right)=\frac{1}{\mathbb{E}(T|Z)} \tag{3.1}$$

$$\mathbb{E}\left(\frac{m(Z^*)}{p(Z^*)}|T=1,Z\right)=\frac{\mathbb{E}(Y|Z)}{\mathbb{E}(T|Z)} \tag{3.2}$$

Under Assumption 3, the right-hand sides of these equations are identified because the distribution of (Y,T,Z) is identified from the main sample S. Moreover, $\forall$ known function f, $\mathbb{E}(f(Z^*)|T=1,Z)$ is identified, because the distribution of $(Z^*,Z)|T=1$ is identified from the auxiliary sample S_a. Hence, the region of identification of $1/p(z^*)$ (respectively $m(z^*)/p(z^*)$) is the set of functions f such that $\mathbb{E}(f(Z^*)|T=1,Z)=1/\mathbb{E}(T|Z)$ (respectively $\mathbb{E}(Y|Z)/\mathbb{E}(T|Z)$). Assumption 4 ensures that these regions reduce to a single element because $\mathbb{E}(f(Z^*)|T=1,Z)=0$ implies that $f(Z^*)p(Z^*)=0$, and that $f(Z^*)=0$, because of the support condition. So $1/p(z^*)$ and $m(z^*)/p(z^*)$ are identified and then, $m(z^*)$, $p(z^*)$, and finally θ are identified.

A similar reasoning can be performed to prove the identification of $\mathbb{E}(g(Y,T)|Z^*=z^*)$ for any function $g(.)$. As a consequence, the conditional distribution $(Y,T)|Z^*$ is identified. The identification of the full joint distribution (Z^*,Z,T,Y) naturally follows (see the proof in the Appendix).

Under mild conditions on the measurement error, the support condition is likely to be necessary to obtain identification. We cannot directly adapt the previous proof to the case without support condition. To see this, let us consider that the econometrician knows the set $\mathcal{S}=\{z^*:p(z^*)>0\}$. If $\mathbb{P}(Z^*\in\mathcal{S})<1$, then Equation 3.1 has to be adapted into:

$$\mathbb{E}\left(\frac{\mathbb{1}\{Z^*\in\mathcal{S}\}}{p(Z^*)}|T=1,Z\right)=\frac{\mathbb{P}(Z^*\in\mathcal{S}|Z)}{\mathbb{E}(T|Z)}.$$

In such a case, identification fails because $\mathbb{P}(Z^*\in\mathcal{S}|Z)$ is not identified without any supplementary assumption. The auxiliary sample of treated individuals only identifies $(Z,Z^*)|Z^*\in\mathcal{S},T=1$, which is not informative on the distribution of $(Z,Z^*)|Z^*\notin\mathcal{S}$. To recover identification, we need supplementary assumptions that enable us to extrapolate outside of $\mathcal{S}$.

In many practical cases, institutional rules ensure that $\mathbb{P}(T = 1|Z^*) = 0$ if Z^* is below (or respectively over) a fixed threshold. For instance, many public benefits are means-tested. In such cases, the support condition fails to hold. So, we next reinforce our assumptions to ensure the LATE identification.

3.2 RD Design without support condition on the score

To extrapolate the distribution of Z^* outside of the support of the score, we use a supplementary Assumption on the structure of error. We assume a classical measurement error. Such a structure is sufficiently informative so that we can relax the assumption on what we observe from the data. Measurement error is identified even when the true running variable and its noisy proxy are observed in different samples.

Assumption 5 (Observation from the data)
We observe an iid sample of (Y, T, Z) and an iid sample of $Z^*|T = 1$.

Assumption 6 (Classical measurement error)

1. $Z = Z^* + \varepsilon$ *with* $\varepsilon \perp\!\!\!\perp Y(0), Y(1), T, Z^*$.

2. (Z^*, ε) *is dominated by the Lebesgue measure on* $\mathbb{R}^2$.

3. *The sets of zeros of the Fourier transforms of* f_ε *and of* $f_{Z^*|T=1}$ *have an empty interior.*

As already mentioned Assumption 6 implies Assumptions 2 and 4. More precisely, the classical measurement error is non-differential (Assumptions 6.1 and 6.2 imply Assumption 2). Furthermore, Assumption 6.3 is a form of completeness assumption.

Theorem 3.2 (Identification of $F_{Z^*,Z,T,Y}$ and θ)
Under Assumptions 1, 5 and 6, the joint distribution of Z^*, Z, T, Y and the LATE (θ) are identified.

Identification relies on the properties of Fourier transforms. Let $\mathcal{F}_f(t) = \int f(x)e^{itx}dx$ be the Fourier transform of an integrable function f. $\mathcal{F}_f$ is continuous and tends to 0 at infinity. Such a transform is injective and $\mathcal{F}_{f_\varepsilon}$ then characterizes the distribution of ε. Moreover, if we denote $\star$ the convolution product of two integrable functions, then $\mathcal{F}_{f\star g} = \mathcal{F}_f\mathcal{F}_g$. Under Assumption 6, we have $f_Z = f_\varepsilon \star f_{Z^*}$ and $f_{Z|T=1} = f_\varepsilon \star f_{Z^*|T=1}$. Then,

$$\mathcal{F}_{f_\varepsilon} = \mathcal{F}_{f_{Z|T=1}} / \mathcal{F}_{f_{Z^*|T=1}}$$

Assumption 5 ensures that the denominator is identified by the sample of $Z^*|T = 1$ and the numerator is identified by the sample of (Y, Z, T), then $\mathcal{F}_{f_\varepsilon}(t)$ is identified $\forall t$ such that $\mathcal{F}_{Z^*|T=1}(t) \neq 0$. Given Assumption 6.3, the continuity of $\mathcal{F}_{f_\varepsilon}$ ensures that $\mathcal{F}_{f_\varepsilon}$ is identified. For any integrable function g of (Y, T), we also have $\mathbb{E}(g|Z)f_Z = f_\varepsilon \star (\mathbb{E}(g|Z^*)f_{Z^*})$. It follows that $\mathbb{E}(g|Z^*)f_{Z^*}$ is identified for any g, then the distribution of $(Y, T)|Z^*$ is identified, which is sufficient to identify θ. The identification of the joint distribution (Y, Z, Z^*, T) naturally follows.

4 Nonparametric estimation

We now propose an estimation strategy of the LATE in the case with support condition (two-sided fuzzy design). When the support condition holds, our identification strategy relies on the nullity of conditional moments that are similar to those involved in the estimation of nonparametric IV models. So, following Newey & Powel (2003), Ai & Chen (2003), Chen (2007), Chen & Pouzo (2012), Blundell et al. (2007), we adopt a sieve estimator. We prove consistency. Further, we perform Monte-Carlo simulations to illustrate its finite sample performance and how it can be used in practice.

4.1 Consistency

The LATE depends on the values in the neighborhood of 0 of three functions: $p(z^*) = \mathbb{E}(T|Z^*)$, $m_0(z^*) = \mathbb{E}(Y|T=0, Z^*=z^*)$ and $m_1(z^*) = \mathbb{E}(Y|T=1, Z^*=z^*)$. It writes:

$$\theta_0 = \frac{m_0(0^+)(1-p(0^+)) + m_1(0^+)p(0^+) - m_0(0^-)(1-p(0^-)) - m_1(0^-)p(0^-)}{p(0^+) - p(0^-)}$$

In this section, we write the LATE as a function of both m_0 and m_1, whereas it could have been written as a function of $m(z^*) = \mathbb{E}(Y|Z^*=z^*)$ only. As it will become clearer below, this distinction is useful when the main and auxiliary samples can be matched. Accordingly, we adapt the moment conditions 3.1 and 3.2 to identify p, m_0 and m_1. Denoting $W = (T, Z, Z^*, Y)$, the conditional moment conditions write:

$$\mathbb{E}(\rho_p(W;p)|Z) := \mathbb{E}(T/p(Z^*) - 1|Z) = 0 \tag{4.1}$$

$$\mathbb{E}(\rho_0(W;p,m_0)|Z) := \mathbb{E}(m_0(Z^*)(1/p(Z^*) - 1)T - Y(1-T)|Z) = 0 \tag{4.2}$$

$$\mathbb{E}(\rho_1(W;m_1)|Z) := \mathbb{E}((m_1(Z^*) - Y)T|Z) = 0 \tag{4.3}$$

The previous identification conditions ensure that

$$Q(\widetilde{p}, \widetilde{m_0}, \widetilde{m_1}) := \mathbb{E}\left(\mathbb{E}\left(\rho_p(W;\widetilde{p})|Z\right)^2\right) + \mathbb{E}\left(\mathbb{E}\left(\rho_0(W;\widetilde{p},\widetilde{m_0})|Z\right)^2\right) + \mathbb{E}\left(\mathbb{E}\left(\rho_1(W;\widetilde{m_1})|Z\right)^2\right)$$

is null only for $(\widetilde{p}, \widetilde{m_0}, \widetilde{m_1}) = (p, m_0, m_1)$. Our estimation strategy is based on the minimization of an empirical counterpart of Q. We consider a sieve GMM estimator of (p, m_0, m_1) (or equivalently a sieve minimum distance (MD) estimator, see Chen (2007), Section 2.2.4 for a discussion of the relation between sieve-MD and sieve-GMM).

First, we define series estimators of the conditional moments. Recall that n and n_a are the sizes of the main and auxiliary samples, respectively S and S_a. Let $\mathcal{I}^p_{n,n_a}$ (respectively $\mathcal{I}^0_{n,n_a}$ and $\mathcal{I}^1_{n,n_a}$) be a sequence of finite dimensional subspaces of $L^\infty(Z)$, such that $\bigcup_{n,n_a} \mathcal{I}^p_{n,n_a}$ is dense in $L^\infty(Z)$ for the supremum norm. Let $B^p(z) = (b^p_1(z), ..., b^p_{l^p(n,n_a)}(z))$ (respectively $B^0(z)$, $B^1(z)$) be a row vector of elements of $L^\infty(Z)$ such that $\text{span}(B^p) = \mathcal{I}^p_{n,n_a}$. $l^p(n,n_a)$ is the dimension of $\mathcal{I}^p_{n,n_a}$. The series estimator of $\mathbb{E}(\rho_j(W)|Z=z)$ based on B^j (for $j = p, 0, 1$)

is:

$$\widehat{\mathbb{E}}(\rho_j(W)|Z=z) = B^j(z)\widehat{\mathbb{E}}(B^{j'}(Z)B^j(Z))^{-1}\widehat{\mathbb{E}}(B^{j'}(Z)\rho_j(W)).$$

It is natural to define $\widehat{\mathbb{E}}(B^{j'}(Z)B^j(Z))$ as the mean in the main sample. ρ_j can be written as $\rho_j(W) = q_j(Z^*)T + r_j(Y,T)$, then a consistent estimator for $\widehat{\mathbb{E}}(B^{j'}(Z)\rho_j(W))$ is:

$$\widehat{\mathbb{E}}(B^{j'}(Z)\rho_j(W)) = \left(\frac{1}{n}\sum_{i\in S} T_i\right)\left(\frac{1}{n_a}\sum_{i\in S_a} B^{j\prime}(Z_i)q_j(Z_i^*)\right) + \left(\frac{1}{n}\sum_{i\in S} B^{j\prime}(Z_i)r_j(Y_i,T_i)\right)$$

Given the above definition of series estimators, the sieve-GMM estimator $\widehat{p}, \widehat{m_0}, \widehat{m_1}$ is the solution to the following minimization program :

$$\min_{(p,m_0,m_1)\in\mathcal{H}_{n,n_a}} Q_{(n,n_a)}(p,m_0,m_1) := \min_{(p,m_0,m_1)\in\mathcal{H}_{n,n_a}} \sum_{j=p,0,1} \frac{1}{n}\sum_{i\in S}\widehat{\mathbb{E}}(\rho_j(W)|Z=Z_i)^2,$$

where $\mathcal{H}_{n,n_a} = \mathcal{H}^p_{n,n_a} \times \mathcal{H}^0_{n,n_a} \times \mathcal{H}^1_{n,n_a}$ is a sequence of finite dimensional functional spaces such that $\bigcup_{n,n_a}\mathcal{H}_{n,n_a}$ is dense for a given norm in $\mathcal{H} = \mathcal{H}^p \times \mathcal{H}^0 \times \mathcal{H}^1$ a functional space containing (p, m_0, m_1). To avoid that the minimization of Q_{n,n_a} gives an infinity of solutions, we naturally impose that $dim(\mathcal{I}^p_{n,n_a}) \geq dim(\mathcal{H}^p_{n,n_a})$, $dim(\mathcal{I}^0_{n,n_a}) \geq dim(\mathcal{H}^0_{n,n_a})$ and $dim(\mathcal{I}^1_{n,n_a}) \geq dim(\mathcal{H}^1_{n,n_a})$.

Note that, in the previous program, we have, for $j = p, 0, 1$:

$$\tfrac{1}{n}\textstyle\sum_{i\in S}\widehat{\mathbb{E}}(\rho_j(W)|Z=Z_i)^2 \;=\; \widehat{\mathbb{E}}(B^{j'}(Z)\rho_j(W))'\left[\tfrac{1}{n}\textstyle\sum_{i\in S} B^{j\prime}(Z_i)B^j(Z_i)\right]^{-1}\widehat{\mathbb{E}}(B^{j\prime}(Z)\rho_j(W)).$$

Note also that, when the two samples can be matched, $S_a = \{i \in S : T_i = 1\}$ and then $\widehat{\mathbb{E}}(B^{p'}(Z)\rho_p(W))$ reduces to $\frac{1}{n}\sum_{i\in S} B^p(Z_i)\left(\frac{T_i}{p(Z_i^*)} - 1\right)$. Similar simplifications occur for the two other moments. Then, our framework is very close to Ai & Chen (2003), except that, in their case, the model is such that the finite dimension parameter θ_0 is $\sqrt{n}$ estimable.

The convergence of $(\widehat{p}, \widehat{m_0}, \widehat{m_1})$ depends on the degree of ill-posedness of the problem, and on the rate of uniform convergence in probability of Q_{n,n_a} towards Q, which we control assuming the following regularity conditions on the data generating process.

Assumption 7 (Regularity Conditions)

1. *The support of* Z^* *is* $[-1;1]$,

2. *The conditional probability* $p(z^*) = P(T = 1|Z^* = z^*)$ *is bounded below by a known constant* $\underline{c} > 0$*, and* m_0 *and* m_1 *are bounded by a known constant* c*,*

3. p*,* m_0 *and* m_1 *are* $\mathcal{C}^1$ *on* $[-1; 0[$ *and* $]0; 1]$ *and their first derivatives are bounded by a known constant* C*,*

4. f_Z *has a compact support* $[l, u]$*, is bounded below on any compact included in* $]l, u[$*, is differentiable with derivative f' and it exists* $C_u, C_l, \alpha_u, \alpha_l > 0$ *such that* $f_Z(z) \sim_{z\sim u} C_u(u - z)^{\alpha_u}$*,* $f_Z(z) \sim_{z\sim l} C_l(z - l)^{\alpha_l}$*,* $f'_Z(z) \sim_{z\sim u} -\alpha_u C_u(u - z)^{\alpha_u - 1}$ *and* $f'_Z(z) \sim_{z\sim l} \alpha_l C_l(z - l)^{\alpha_l - 1}$*,*

5. $z \mapsto f_{Z|Z^*=z^*}(z)$ *is continuously differentiable* z^**-almost-everywhere and for any* $z_0 \in]l, u[$*, it exists a neighborhood* $V(z_0)$ *such that* $\mathbb{E}(\sup_{z\in V(z_0)} |f'_{Z|Z^*}|) < \infty$*.*

Assumption 7.1 essentially means that Z^* has a compact support with the discontinuity threshold inside this compact. The choice of $[-1; 1]$ is simply a normalization that can be assumed without loss of generality. Assumption 7.2 is a small reinforcement of the support condition that is necessary for identification. Assumption 7.3 is a convenient way to ensure that the model is well-separated. Indeed in this case, (p, m_0, m_1) belongs to a compact and the continuity of the theoretical objective function ensures that the model is well-separated. This kind of Assumption is usual in parametric framework (see for instance Chapter 5, page 46 and Problem 5.27 of van der Vaart (2000)). Assumption 7.3 also imposes a regularity condition that ensures that the problem is not ill-posed, which is usual in nonparametric frameworks (see for instance Chen (2007) or Newey & Powel (2003) for the case of sieve estimators).

Assumption 7.4 is necessary to control the rate of uniform convergence of Q_{n,n_a} to Q. Newey (1997), Burman & Chen (1989), Huang (1998), Blundell et al. (2007) or Chen & Pouzo (2012) use a stronger assumption, assuming that f_Z is bounded below on its support. In our framework, this is not a credible assumption, because measurement errors entail smoothness and continuity at the boundary of the support of f_Z. For example, in the case of classical measurement error, $Z = Z^* + \varepsilon$ with Z^* and ε independent, with convex compact support and f_{Z^*} and f_ε bounded below by positive constants on their supports, then f_Z tends towards 0 at the bounds of its support. However it verifies Assumption 7.4 with $\alpha_u = \alpha_l = 1$. More generally, in the case of a *transformation* of a classical measurement error model (such that $Z = \mu^{-1}(\nu(Z^*) + \varepsilon)$ with ν and μ increasing C-1 diffeomorphisms), if $f_{Z^*}(x) \sim_1 C^+_{Z^*}(1-x)^{\alpha^+_{Z^*}}$, $\nu'(x) \sim_1 C^+_{\nu'}(1-x)^{\alpha^+_{\nu'}}$, $(\mu^{-1})'(x) \sim_u C^+_{(\mu^{-1})'}(u-x)^{\alpha^+_{(\mu^{-1})'}}$ and $f_\varepsilon(x) \sim_{\overline{\varepsilon}} C^+_\varepsilon(\overline{\varepsilon}-x)^{\alpha^+_\varepsilon}$ for $\overline{\varepsilon} = \sup \mathrm{Supp}(\varepsilon)$, then Assumption 7.4 is verified with $\alpha_u = \alpha^+_{Z^*} + \alpha^+_\varepsilon + 1 - \alpha^+_{\nu'} + \alpha^+_{(\mu^{-1})'}$. As a consequence, the decreasing to 0 at the boundary of the support of f_Z prevents us from using usual results from the approximation theory (see Huang (1998)) and we find lower rate of convergence. Assumptions 7.4 and 7.5 ensure that if $\mathcal{G}$ is a class of functions uniformly bounded, then the functions $z \mapsto \mathbb{E}(g(Z^*)|Z=z)$ are uniformly bounded and satisfy a Lipschitz condition (see Lemma A.3 in the Appendix).

In practice, we consider for $\mathcal{H}^0_{n,n_a}$ the piecewise linear functions bounded by the known constant c and with Lipschitz constant C. More precisely, it exists $\underline{\delta}^0, \overline{\delta}^0$ (independent of (n,n_a)), integers k^{0+}_{n,n_a}, k^{0-}_{n,n_a} and knots $(1 = z^{0+}_{k^{0+}_{n,n_a}+1} > z^{0+}_{k^{0+}_{n,n_a}} > \ldots > z^{0+}_1 > z^{0+}_0 = 0 = z^{0-}_0 > \ldots > z^{0-}_{k^{0-}_{n,n_a}} > z^{0-}_{k^{0-}_{n,n_a}+1} = -1)$ verifying $\frac{\underline{\delta}}{k^{0\pm}_{n,n_a}} \leq |z^{0\pm}_j - z^{0\pm}_{j-1}| \leq \frac{\overline{\delta}}{k^{0\pm}_{n,n_a}}$, such that:

$$\mathcal{H}^0_{n,n_a} = \left\{ \begin{array}{ll} f: & \exists (a^+_j)_{j=1,\ldots,k^{0+}_{n,n_a}}, (a^-_j)_{j=1,\ldots,k^{0-}_{n,n_a}} \text{ such that} \\ & \quad f(z^*) = \ f(0^+)\mathbb{1}_{\{z^*>0\}} + \sum_{j=0}^{k^{0+}_{n,n_a}} a^+_j (z^* - z^{0+}_j)\mathbb{1}_{\{z^*-z^{0+}_j>0\}} \\ & \qquad\qquad\quad +f(0^-)\mathbb{1}_{\{z^*<0\}} + \sum_{j=0}^{k^{0-}_{n,n_a}} a^-_j (z^* - z^{0-}_j)\mathbb{1}_{\{z-z^{0-}_j<0\}} \\ & \quad \sup_{z^*} |f(z^*)| < c \text{ and } \sup_{z^*} |f'(z^*)| < C \end{array} \right\}$$

We obtain a sequence of such functional spaces $\mathcal{H}^0_{n,n_a}$ by increasing the number of knots with n and n_a. The union of the resulting sequence enables us to approach any function m_0 verifying Assumptions 7.1, 7.2 and 7.3. Similar spaces are considered for $\mathcal{H}^1_{n,n_a}$, associated with constants $\overline{\delta}^1$, $\underline{\delta}^1$ and $k^{1\pm}_{n,n_a}$ knots $z^{1\pm}_j$. $\mathcal{H}^p_{n,n_a}$ is defined similarly except that the condition $\sup_{z^*} |f(z^*)| < c$ is replaced by $\underline{c} \leq f \leq 1$.

Theorem 4.1 (Consistency)

Under Assumptions 1, 2, 3, 4 and 7, if $\frac{n}{n_a} \to \lambda \in]0; +\infty[$, *then:*

$$\widehat{\theta} = \frac{\widehat{m_0}(0^+)(1-\widehat{p}(0^+)) + \widehat{m_1}(0^+)\widehat{p}(0^+) - \widehat{m_0}(0^-)(1-\widehat{p}(0^-)) - \widehat{m_1}(0^-)\widehat{p}(0^-)}{\widehat{p}(0^+) - \widehat{p}(0^-)}$$

converges in probability to θ_0 *if* $\min_{j=p,0,1}(k_{n,n_a}^{j+}, k_{n,n_a}^{j-}) \to \infty$ *with* $\max_{j=p,0,1}(\dim(\mathcal{I}_{n,n_a}^j)) = o(n^{1/(2+\max(\alpha_u,\alpha_l))})$.

The proof of Theorem 4.1 is reported in the Appendix. In Theorem 4.1, we show that, when n and n_a tend to infinity, the estimator $(\widehat{p}, \widehat{m_0}, \widehat{m_1})$ consistently estimates (p, m_0, m_1), if the dimension of $\mathcal{H}_{n,n_a}$ tends to infinity and the dimension of $\mathcal{I}_{n,n_a}^j$ (for all $j = p, 0, 1$) tends to infinity sufficiently slowly.

When the samples can be matched, an alternative estimator may be of interest. It consists in estimating m_1 by local linear regression on the one hand, and p and m_0 using Equations 4.1 and 4.2 on the other hand. A straightforward adaptation of the proof of Theorem 4.1 ensures the consistency of this alternative estimator.

4.2 Monte-Carlo Simulations

In this section, we investigate the finite sample properties of our main sieve estimator by Monte-Carlo simulations. We assume that Z^* is uniformly on $[-1;1]$ and that $\mathbb{P}(T = 1|Z^* = z^*) = 1/8 + 1/4 \cdot \Phi(5 \cdot z^*) + 1/2 \cdot \mathbb{1}\{z^* \geq 0\}$, in which Φ is the cdf of the standard normal distribution. The conditional probability to be treated increases with Z^* from 1/8 in -1 to 7/8 in 1 and jumps from 1/4 to 3/4 when Z^* crosses the threshold 0. Consequently, the proportion of compliers is 1/2 whereas always-takers (respectively never-takers) represent 1/4 of the population. The DGPs of the potential outcomes are:

$$\begin{aligned} Y(0) &= 4 + 3Z^* + v_0, \\ Y(1) &= \mathbb{1}\{C\} + 2\mathbb{1}\{AT, NT\} + 3Z^* + v_1, \end{aligned}$$

where $\mathbb{1}\{C\}$ and $\mathbb{1}\{AT, NT\}$ are dummies for the types of individuals (compliers versus always or never-takers), and $(v_0, v_1)|Z^*, C, AT, NT \sim \mathcal{N}(0, \Sigma)$ with $\Sigma = \frac{1}{16}\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$.
The true LATE $\theta_0 = \mathbb{E}(Y(1) - Y(0)|Z^* = 0, C)$ is then equal to $1 - 4 = -3$.
The noisy running variable is drawn from the following multiplicative process:

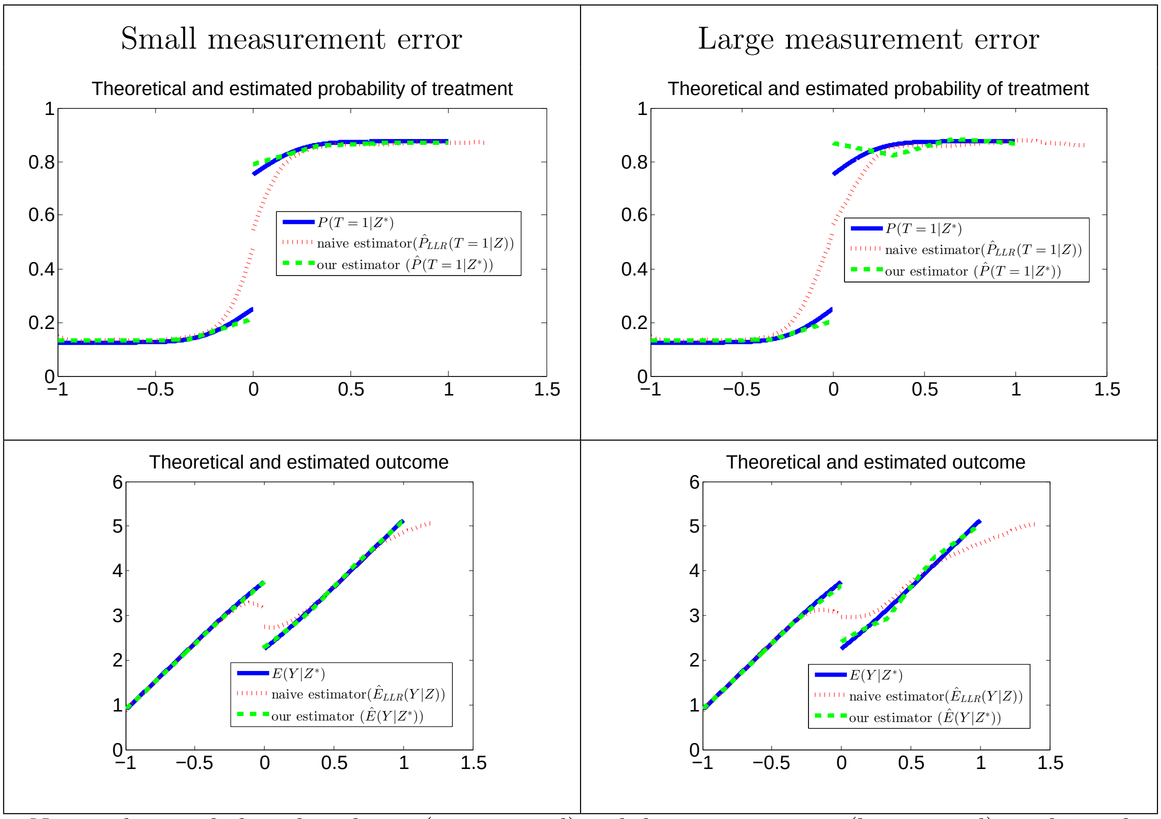
$$Z + 1 = (Z^* + 1)(\varepsilon + 1),$$

with ε uniformly distributed and independent of $(Z^*, T, Y(0), Y(1))$. To investigate the impact of the size of the measurement error, we let the dispersion of ε vary. Below we consider two cases: (i) small measurement error with $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$, and (ii) large measurement error with $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$. In the Appendix, we also present simulations in the case of additive classical measurement error.

We compare our sieve estimator ($\widehat{p}(Z^*) = \widehat{\mathbb{E}}(T|Z^*)$ or $\widehat{m}(Z^*) = \widehat{\mathbb{E}}(Y|Z^*)$) to the naive estimator obtained by ignoring measurement error ($\widehat{E}_{LLR}(T|Z)$ or $\widehat{E}_{LLR}(Y|Z)$). The naive estimation relies on a standard local linear regression (LLR) where Z^* is directly replaced by Z. We make the following assumptions on the parameters of our sieve estimation. We choose the same space for $\mathcal{I}^p_{n,n_a}$,$\mathcal{I}^0_{n,n_a}$ and $\mathcal{I}^1_{n,n_a}$, namely linear splines with equidistant knots on $\text{Supp}(Z)$. Similarly, $\mathcal{H}^p_{n,n_a}$, $\mathcal{H}^0_{n,n_a}$ and $\mathcal{H}^1_{n,n_a}$ have the same equidistant knots. The numbers of knots are chosen such that $\dim(\mathcal{I}^p_{n,n_a}) = \dim(\mathcal{H}^p_{n,n_a})$. Consequently, all the functional spaces manipulated have the same dimension. Last, concerning the bounds $\underline{c}$, c, and C of functions in $\mathcal{H}_{n,n_a}$, we have chosen 0.05, 15 and 10. Note that, given the underlying DGP, every value lower than 1/8 is admissible for $\underline{c}$, every value larger than 7 is admissible for c and every constant larger than $\max(\frac{5}{4\sqrt{2\pi}}, 3) \simeq 3.14$ is admissible for C.

Figure 1 plots the take-up (upper panel) and the mean outcome (lower panel) conditional on the true or noisy running variables. These are usual RD graphs used in applied research. The left panel plots one simulation obtained with a small measurement error (case (i)), while the right panel corresponds to the case (ii) with large measurement error. We simulate 25,000 observations. On each graph, we plot the true conditional expectation ($\mathbb{E}(T|Z^*)$ or $\mathbb{E}(Y|Z^*)$), the naive estimate of the conditional expectation obtained by ignoring measurement error ($\widehat{E}_{LLR}(T|Z)$ or $\widehat{E}_{LLR}(Y|Z)$) and our sieve estimate ($\widehat{p}(Z^*) = \widehat{\mathbb{E}}(T|Z^*)$ or $\widehat{m}(Z^*) = \widehat{\mathbb{E}}(Y|Z^*)$). Figure 1 clearly illustrates that the discontinuity vanishes when measurement error is ignored: dashed lines do not reproduce the discontinuity of full lines. The loss of discontinuity is clear in all graphs, except maybe in the left lower panel. When the measurement error is small, the conditional expectation is steep at the cutoff value. Consequently, it may appear as discontinuous if the bandwidth of the local linear regression is too large. Figure 1 also illustrates that our proposed estimator is able to recover the discontinuity of the true conditional expectation.

Figure 1: Monte-Carlo simulations



Notes: the panel plots the take-up (upper panel) and the mean outcome (lower panel) conditional on the running variables. The left panel plots the simulations obtained with a small measurement error, while the right panel corresponds to the case with large measurement error. On each graph, we plot the true conditional expectation ($\mathbb{E}(T|Z^*)$ or $\mathbb{E}(Y|Z^*)$), the naive estimation of the conditional expectation obtained by ignoring measurement error ($\widehat{\mathbb{E}}_{LLR}(T|Z)$ or $\widehat{\mathbb{E}}_{LLR}(Y|Z)$) and our sieve estimator ($\widehat{p}(Z^*) = \widehat{\mathbb{E}}(T|Z^*)$ or $\widehat{m}(Z^*) = \widehat{\mathbb{E}}(Y|Z^*)$). The naive estimation relies on a a standard local linear regression with bandwidth around 0.1, where Z^* is directly replaced by Z. Our sieve estimator is obtained with three positive and three negative knots. We select for each column one simulation of the DGP described in section 4.2 with 25,000 observations.

Table 1 reports the bias, variance and MSE of various estimators, averaged over 1,000

Monte-Carlo simulations. In Columns 3 to 5, we report results concerning our sieve estimator; in Columns 6-8, those concerning the naive estimator; and, in Columns 9-11, those concerning the unfeasible estimator obtained by LLR assuming that Z^* is observed. In the upper panel, we report results when the measurement error is small; in the lower panel when it is large. We report results for three different sample sizes 1,000, 5,000 and 25,000 observations. We also let the number of knots of our sieve estimation vary. When the number of knots is null, the approximating functions are linear on both sides of the threshold. When the number of knots is one (resp. two), we allow for one (resp. two) change in slope on each side.

Overall, the absolute bias of our estimator is below one, whatever the size of the measurement error, the sample size or the number of knots. It outperforms the naive estimator which is almost always over one in Table 1. As expected, the absolute bias of the unfeasible estimator is below 0.05. The variance of our estimator is always lower than the one of the naive estimator. Consistently with Proposition 2.2, the bias of the naive estimator and its variance are quite erratic across the Table, and its variance does not decrease with the sample size.

Table 1 also informs us about the influence of the number of knots on our sieve estimator. For a given sample size, the bias tends to decrease with the number of knots and the variance tends to increase. This reflects the usual influence of smoothing parameters on the trade-off between bias and efficiency.

Our estimator performs better, when the measurement error is small. When the difference between the bias with small and large errors is significant, it is lower in the case of small measurement errors. Moreover, the variance is smaller with small measurement error, except for the estimation with only one segment (or 0 knot).

In the Appendix, we report supplementary Monte-Carlo simulations. First, we explicitly consider the case when the main and auxiliary samples are matched. As explained above, we can estimate the conditional mean of the outcome on the treated by LLR. The results are very close to those of our main sieve estimator (see Columns 6-8 of Table 2 in the

Table 1: Estimation in finite samples, Multiplicative Error

Nb. of	Sample	Our estimator			Naive estimator			Unfeasible estimator		
knots	size	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
					A. Small Measurement Error					
	1000	-0.969	10.63	11.57	-1.840	6689	6693	0.021	0.168	0.168
0	5000	0.344	1.009	1.128	-1.526	3.993	6.323	0.011	0.043	0.043
	25000	0.678	0.024	0.484	-2.013	577.3	581.4	0.005	0.011	0.011
	1000	-0.53	24.98	25.26	-0.881	53.66	54.44	0.040	0.154	0.156
1	5000	0.351	0.328	0.451	-1.463	4.929	7.070	0.005	0.043	0.043
	25000	0.512	0.078	0.340	-2.597	6.486	13.23	0.011	0.010	0.011
	1000	-0.738	325.0	325.5	-1.715	734.1	737.1	0.014	0.169	0.170
2	5000	0.074	0.578	0.584	-1.484	15.41	17.61	0.011	0.045	0.045
	25000	0.172	0.097	0.127	-2.355	52.38	57.93	0.004	0.012	0.012
					B. Large Measurement Error					
	1000	-0.009	2.818	2.818	9.125	40095	40178	0.021	0.168	0.168
0	5000	0.681	0.127	0.591	136.3	2×10^7	2×10^7	0.011	0.043	0.043
	25000	0.806	0.018	0.667	-60.54	3×10^6	3×10^6	0.005	0.011	0.011
	1000	-0.188	31.85	31.89	-0.794	34973	34973	0.040	0.154	0.156
1	5000	0.349	0.423	0.545	-2.919	50204	50212	0.005	0.043	0.043
	25000	0.512	0.110	0.373	2.605	1076	1083	0.011	0.010	0.011
	1000	0.249	943.9	943.9	-91.40	7×10^6	8×10^6	0.014	0.169	0.170
2	5000	0.142	2.639	2.660	0.897	1140	1141	0.011	0.045	0.045
	25000	0.174	0.573	0.603	-4.893	25264	25288	0.004	0.012	0.012

Note : Computation obtained with 1000 simulations. The same set of simulations is used for all the estimators on the same line. The set of simulations changes across lines.

$Z+1=(Z^*+1)\cdot(1+\varepsilon)$ with $\varepsilon\sim\mathcal{U}_{[-0.1;0.1]}$ for the DGP with *small* measurement error and $\varepsilon\sim\mathcal{U}_{[-0.2;0.2]}$ for the DGP with *large* measurement error. Number of knots equal to 0 means that p, m_0 and m_1 are approximated by linear functions on $[-1;0]$ and $[0;1]$. When the number of knots is 1 (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$).

Appendix). Second, we compare our main estimator to the Donut estimator (see Dong, 2014 or Barreca, Guldi, Lindo & Waddell, 2011). The Donut estimator corresponds to the naive estimator in a truncated sample. Observations around the threshold (according to the noisy measure) are removed from the estimation sample. The bias of the Donut estimator is large (around 2) and greater than the bias associated with the naive estimator (see Columns 9-11 of Table 2 in the Appendix). Third, we repeat all the previous Monte-Carlo exercises, replacing the multiplicative error by an additive classical measurement error. Our estimator clearly outperforms the naive estimator when the sample size is larger than 5,000. The influence of the number of knots on the bias and on the variance is qualitatively similar to the case with multiplicative error (see Tables 3 and 4 in the Appendix).

5 Conclusion

Non-differential measurement error in the running variable has dramatic consequences for the identification of treatment effects in Regression Discontinuity designs. As soon as there is no mass of individuals with correct values of their running variable, all discontinuities are smoothed out in the noisy data. The usual estimator of the Local Average Treatment Effect (LATE) is then inconsistent. In this paper, we proposed to take advantage of naturally-occurring auxiliary data to recover identification. Agencies in charge of delivering the treatment usually keep record of the correct running variable for the treated individuals. Under the assumption of non-differential measurement error, the auxiliary information can be used to extrapolate the true running variable distribution on the non-treated, and to identify the joint distribution of the true running variable, the treatment and the potential outcomes. We then proposed a sieve estimator for the LATE, showed its consistency and investigated its performance in finite samples. Our simulations suggest that our estimator outperforms the naive estimator that ignores measurement error.

The estimator we proposed can be extensively applied. In many surveys, running variables, such as income, firms' size, even test scores, are measured with errors. Hullegie & Klein

(2010) and Cahuc et al. (2014) are perfect examples of the loss of discontinuity in the take-up when the running variable is noisy. Provided that administrative data about eligibility are matched with those samples, the LATE of those different programs could be easily estimated using our approach. Our approach could also be applied to the estimation of the return of schooling using discontinuity in tuition costs with respect to the parental income. Schools only check and record the parental income of students that effectively decide to follow a supplementary year of schooling, whereas, for individuals who stop their schooling, the econometrician only observes a proxy of income (for instance, parental income in the previous academic year).

A Appendix: Proofs

A.1 Proof of Proposition 2.1

We begin by proving the first part of the proposition. This is a direct application of the dominated convergence theorem. Our function of interest verifies:

$$\begin{aligned} \mathbb{E}[T|Z=z]f_Z(z) &= \mathbb{E}[\mathbb{E}[T|Z^*=z^*, Z=z]|Z=z]f_Z(z) \\ &= \mathbb{E}[\mathbb{E}[T|Z^*=z^*]|Z=z]f_Z(z) \\ &= \int \mathbb{E}[T|Z^*=z^*]f_{Z|Z^*=z^*}(z)f_{Z^*}(z^*)dz^*, \end{aligned}$$

where we use that the error is non differential ($Z \perp\!\!\!\perp T|Z^*$). We also have:

$$f_Z(z) = \int f_{Z|Z^*=z^*}(z)f_{Z^*}(z^*)dz^* .$$

The dominated convergence theorem then ensures that $z \mapsto \mathbb{E}[T|Z=z]f_Z(z)$ and $z \mapsto f_Z(z)$ are both continuous for all z if the first condition holds. As a consequence, $z \mapsto \mathbb{E}[T|Z=z]$ is continuous on any interior point of $\text{Supp}(Z)$.

We now prove that the second condition is also sufficient to ensure the continuity of $z \mapsto \mathbb{E}[T|Z=z]$. Note that because μ^{-1} is continuous, the continuity of $z \mapsto \mathbb{E}(T|Z=z)$ is equivalent to the continuity of $\tilde{z} \mapsto \mathbb{E}(T|\mu(Z)=\tilde{z})$. Moreover, the boundedness of the density of Z^* is equivalent to the boundedness of the density of $\nu(Z^*)$ because $\inf_{z^* \in \text{Supp}(Z^*)} \nu'(z^*) > 0$. So without loss of generality, we can restrict the proof to the case where $\mu = \nu = id$, i.e. the case of classical measurement error (where $Z = Z^* + \varepsilon$ with $\varepsilon \perp\!\!\!\perp Z^*$). Note also that boundedness of f_ε ensures that the first condition of the proposition holds, because $f_{Z|Z^*=z^*}(z) = f_\varepsilon(z - z^*)$. Then we obtain directly the continuity of $z \mapsto \mathbb{E}(T|Z=z)$. The proof is different in the case of classical measurement error with a bounded density on f_{Z^*}.

Let us begin with some usual notation. For any function f, let $\tau_h f$ denote the function $z \mapsto f(h-z)$. For any measurable function f, let $||f||_1$ and $||f||_\infty$ denote respectively the L^1 and the supremum norm of f: $||f||_1 = \int |f(z)|dz$ and $||f||_\infty = \sup_z |f(z)|$ (both could be equal to $+\infty$).

Let us consider our quantity of interest. In the case of classical measurement error, it writes as follows:

$$\begin{aligned} f_Z(z) &= \int f_{Z|Z^*=z^*}(z) f_{Z^*}(z^*)dz^* \\ f_Z(z) &= \int f_\varepsilon(z-z^*).f_{Z^*}(z^*)dz^* \quad . \\ f_Z(z) &= \int \tau_z f_\varepsilon(z^*).f_{Z^*}(z^*)dz^* \end{aligned}$$

We are now able to prove the continuity of $z \mapsto f_Z(z)$, using Lemma A.1, because

$$\begin{aligned} |f_Z(z+h) - f_Z(z)| &\leq ||f_{Z^*}||_\infty ||\tau_{z+h} f_\varepsilon - \tau_z f_\varepsilon||_1 \\ &\leq ||f_Z^*||_\infty ||\tau_h f_\varepsilon - f_\varepsilon||_1. \end{aligned}$$

The proof is similar for the continuity of $z \mapsto \mathbb{E}(T|Z=z)f_Z(z)$. It follows that $z \mapsto \mathbb{E}(T|Z=z)$ is continuous on the interior of the support of Z.

Lemma A.1 $\lim_{h\to 0} ||\tau_h f - f||_1 = 0$ *for any integrable* f*.*

Proof of Lemma A.1: For any continuous function g with a compact support, the dominated convergence Theorem ensures that $\lim_{h\to 0} ||\tau_h g - g||_1 = 0$. Because the space of continuous functions with a compact support is dense (for the norm $||.||_1$) in space of integrable functions, we know that for any integrable function f and any $\delta > 0$, it exists g continuous with compact support such that $||f - g||_1 < \delta$ and then:

$$\begin{aligned} ||\tau_h f - f||_1 &\leq ||\tau_h f - \tau_h g||_1 + ||\tau_h g - g||_1 + ||g - f||_1 \\ &\leq 2\delta + ||\tau_h g - g||_1. \end{aligned}$$

A.2 Proof of Proposition 2.2

Let us introduce some notations:

$$
\begin{aligned}
B^+ &= \frac{\left(\int_0^{+\infty} u^2K(u)du\right)^2 - \left(\int_0^{+\infty} u^3K(u)du\right)\left(\int_0^{+\infty} uK(u)du\right)}{2\left(\int_0^{+\infty} u^2K(u)du\right)\left(\int_0^{+\infty} K(u)du\right) - \left(\int_0^{+\infty} uK(u)du\right)^2} \\
B^- &= \frac{\left(\int_{-\infty}^{0} u^2K(u)du\right)^2 - \left(\int_{-\infty}^{0} u^3K(u)du\right)\left(\int_{-\infty}^{0} uK(u)du\right)}{2\left(\int_{-\infty}^{0} u^2K(u)du\right)\left(\int_{-\infty}^{0} K(u)du\right) - \left(\int_{-\infty}^{0} uK(u)du\right)^2} \\
V^+ &= \frac{\int_0^{+\infty}\left[\left(\int_0^{+\infty} s^2K(s)ds\right) - \left(\int_0^{+\infty} sK(s)ds\right)u\right]^2 K(u)^2du}{f_Z(0)\left[\left(\int_0^{+\infty} u^2K(u)du\right)\left(\int_0^{+\infty} K(u)du\right) - \left(\int_0^{+\infty} uK(u)du\right)^2\right]^2} \\
V^- &= \frac{\int_{-\infty}^{0}\left[\left(\int_{-\infty}^{0} s^2K(s)ds\right) - \left(\int_{-\infty}^{0} sK(s)ds\right)u\right]^2 K(u)^2du}{f_Z(0)\left[\left(\int_{-\infty}^{0} u^2K(u)du\right)\left(\int_{-\infty}^{0} K(u)du\right) - \left(\int_{-\infty}^{0} uK(u)du\right)^2\right]^2}
\end{aligned}
$$

Under Condition 1 of Proposition 2.2, the dominated convergence Theorem ensures that f_Z and $\mathbb{E}(T|Z)$ are twice differentiable on the interior of the support of Z, similar reasoning holds for $\mathbb{E}(Y^2|Z)$ and $\mathbb{E}(Y|Z)$, because $\mathbb{E}(Y^2|Z^*)$ and then $|\mathbb{E}(Y|Z^*)|$ are bounded. Then Assumptions 1, 2, 3 and 5 of Hahn et al. (1999) hold. Under Condition 2 of Proposition 2.2, the dominated convergence Theorem ensures that the convolution product of $f_{\nu(Z^*)}$ and f_ε is twice continuously differentiable, then Assumptions 1, 2, 3 and 5 of Hahn et al. (1999) also holds in such case.

The conditions on the kernel K and the bandwidth ensure that Assumptions 4 and 7 of Hahn et al. (1999) hold.

Last, the condition $\mathbb{E}(|Y|^\delta|Z) < \infty$ of Proposition 2.2 ensure that $\mathbb{E}(|Y - \mathbb{E}(Y|Z)|^\delta|Z) < \infty$, which is a sufficient condition to Assumption 6 of Hahn et al. (1999), when $\delta \geq 3$.

Hence, we can directly apply Hahn et al. (1999) when $\delta \geq 3$. Moreover, their reasoning, which is based on Lyapounov's central limit Theorem, also holds for $\delta \in]2;3[$. We obtain:

$$
n^{2/5}\begin{pmatrix} \widehat{a}_Y^+ - \mathbb{E}(Y|Z=0) \\ \widehat{a}_T^+ - \mathbb{P}(T=1|Z=0) \end{pmatrix} \rightarrow \mathcal{N}\left(B^+\begin{pmatrix} \partial_z^2\mathbb{E}(Y|Z=0^+) \\ \partial_z^2\mathbb{P}(T=1|Z=0^+) \end{pmatrix}, V^+\begin{pmatrix} \mathbb{V}(Y|Z=0^+) & \mathbb{C}ov(Y,T|Z=0^+) \\ \mathbb{C}ov(Y,T|Z=0^+) & \mathbb{V}(T|Z=0^+) \end{pmatrix}\right)
$$

$$
n^{2/5}\begin{pmatrix} \widehat{a}_Y^- - \mathbb{E}(Y|Z=0) \\ \widehat{a}_T^- - \mathbb{P}(T=1|Z=0) \end{pmatrix} \rightarrow \mathcal{N}\left(B^-\begin{pmatrix} \partial_z^2\mathbb{E}(Y|Z=0^-) \\ \partial_z^2\mathbb{P}(T=1|Z=0^-) \end{pmatrix}, V^-\begin{pmatrix} \mathbb{V}(Y|Z=0^-) & \mathbb{C}ov(Y,T|Z=0^-) \\ \mathbb{C}ov(Y,T|Z=0^-) & \mathbb{V}(T|Z=0^-) \end{pmatrix}\right)
$$

The symmetry of K ensures that $B^+ = B^-$ and $V^+ = V^-$. Moreover, the continuity of $z \mapsto \partial_z^2\mathbb{E}(T|Z=z)$ on the interior of the support of Z ensures that $\partial_z^2\mathbb{E}(T|Z=0^+) =$

$\partial_z^2\mathbb{E}(T|Z=0^-) = \partial_z^2\mathbb{E}(T|Z=0)$. Similar argument holds for $\partial_z^2\mathbb{E}(Y|Z=0)$, $\mathbb{V}(Y|Z=0)$, $\mathbb{C}ov(Y,T|Z=0)$ and $\mathbb{V}(T|Z=0) = \mathbb{E}(T|Z=0)(1-\mathbb{E}(T|Z=0))$. The continuous mapping Theorem ensures that:

$$n^{2/5}\begin{pmatrix} \widehat{a}_Y^+ - \widehat{a}_Y^- \\ \widehat{a}_T^+ - \widehat{a}_T^- \end{pmatrix} \to \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2V^+\begin{pmatrix} \mathbb{V}(Y|Z=0) & \mathbb{C}ov(Y,T|Z=0) \\ \mathbb{C}ov(Y,T|Z=0) & \mathbb{V}(T|Z=0) \end{pmatrix}\right)$$

It follows that $\widehat{\theta}_{LLR}$ tends in distribution to a Cauchy of location $\frac{\mathbb{C}ov(Y,T|Z=0)}{\mathbb{V}(T|Z=0)}$ and scale $\left(\frac{\mathbb{V}(Y|Z=0)}{\mathbb{V}(T|Z=0)} - \frac{\mathbb{C}ov^2(Y,T|Z=0)}{\mathbb{V}^2(T|Z=0)}\right)^{1/2}$.

A.3 Proof of Theorem 3.1

We follow the steps of D'Haultfœuille (2010). We first prove that $\mathbb{P}(T=1|Z^*) = p(Z^*)$ is identified. Under Assumption 2, the support condition and the law of iterated expectation, we have:

$$\mathbb{E}\left(\frac{1}{\mathbb{P}(T=1|Z^*)}|T=1,Z\right) = \frac{1}{\mathbb{P}(T=1|Z)} \tag{A.1}$$

Under Assumption 3, the right hand side of this equation is identified because the distribution of (Y,T,Z) is identified from the main sample. Moreover, for any known function f, $\mathbb{E}(f(Z^*)|T=1,Z)$ is identified because the distribution of $(Z^*,Z)|T=1$ is identified from the auxiliary data. It follows that the region of identification of $1/p(z^*)$ is the set of functions f such that $\mathbb{E}(f(Z^*)|T=1,Z) = 1/\mathbb{E}(T|Z)$. Suppose that there exist two functions f and g verifying equation A.1. Then their difference verifies: $\mathbb{E}(f(Z^*) - g(Z^*)|T=1,Z) = 0$. Using Assumption 2 and the law of iterated expectation, we have: $\mathbb{E}((f(Z^*)-g(Z^*)).p(Z^*)|Z) = \mathbb{E}(f(Z^*)-g(Z^*)|T=1,Z).\mathbb{P}(T=1|Z)$. This ensures that $\mathbb{E}((f(Z^*)-g(Z^*)).p(Z^*)|Z) = 0$. Following the completeness condition (Assumption 4), this implies that $(f(Z^*)-g(Z^*)).p(Z^*) = 0$. Because of the support condition, we obtain that $f = g$. The region of identification reduces to one single element and $p(Z^*)$ is identified.

For any function g, we now prove identification of $\mathbb{E}(g(Y,T)|Z^*=z^*) = h(z^*)$. We follow the same reasoning as above. The identifying moment condition writes:

$$\mathbb{E}\left(\frac{h(z^*)}{\mathbb{P}(T=1|Z^*)}|T=1,Z\right)=\frac{\mathbb{E}(g(Y,T)|Z)}{\mathbb{P}(T=1|Z)} \tag{A.2}$$

We obtain that the ratio $h(z^*)/p(z^*)$ is identified. As $p(z^*)$ is identified above, $h(z^*) = \mathbb{E}(g(Y,T)|Z^* = z^*)$ is identified. As this is true for any function g, the joint distribution of $(Y,T)|Z^*$ is identified. This is sufficient to identify the Wald ratio θ.

The distribution of $Z|Z^*, T = 1$ is identified from the sample of treated and Assumption 2 ensures that the distribution of $Z|Z^*$ is identified and then the distribution of $(Y,T,Z)|Z^*$. The distribution of Z^* is identified because $\mathbb{P}(T = 1|Z^*)$ is identified by Equation 3.1 and the distribution of T and $Z^*|T = 1$ are directly identified from the data. It follows that the distribution of (Z^*, Z, T, Y) is identified.

A.4 Proof of Theorem 4.1

Let $\xi_0 = (p, m_0, m_1)$ and $\widehat{\xi} = (\widehat{p}, \widehat{m_0}, \widehat{m_1})$. Note that $Q(\xi) \geq 0$ for any $\xi \in \mathcal{H}$ and the condition of identification ensures that $Q(\xi) = 0 \Leftrightarrow \xi = \xi_0$. Let $||\widehat{\xi} - \xi_0||_\infty = \sup(||\widehat{p} - p||_\infty, ||\widehat{m_0} - m_0||_\infty, ||\widehat{m_1} - m_1||_\infty)$ We will prove that for any $\delta > 0$, $\mathbb{P}\left(||\widehat{\xi} - \xi_0||_\infty \geq \delta\right)$ tends to zero.

For any sequence $\xi_{n,n_a} \in \mathcal{H}_{n,n_a}$ the following inequalities hold:

$$\begin{aligned}
Q(\widehat{\xi}) &\leq Q(\widehat{\xi}) - Q_{n,n_a}(\widehat{\xi}) + Q_{n,n_a}(\widehat{\xi}) - Q_{n,n_a}(\xi_{n,n_a}) \\
&\quad + Q_{n,n_a}(\xi_{n,n_a}) - Q(\xi_{n,n_a}) + Q(\xi_{n,n_a}) \\
&\leq Q_{n,n_a}(\widehat{\xi}) - Q_{n,n_a}(\xi_{n,n_a}) + 2\sup\nolimits_{\xi\in\mathcal{H}_{n,n_a}} |Q_{n,n_a}(\xi) - Q(\xi)| + Q(\xi_{n,n_a})
\end{aligned}$$

Let $U_{n,n_a} = \inf_{\xi\in\mathcal{H}_{n,n_a}, ||\xi-\xi_0||_\infty\geq\delta} Q(\xi)$. Assume that $Q(\xi_{n,n_a}) = o\left(U_{n,n_a}\right)$ and $\sup_{\xi\in\mathcal{H}_{n,n_a}} |Q_{n,n_a}(\xi) - Q(\xi)| = o_p\left(U_{n,n_a}\right)$. In that case:

$$\limsup_{n,n_a} \mathbb{P}\left(||\widehat{\xi} - \xi_0||_\infty \geq \delta\right) \leq \limsup_{n,n_a} \mathbb{P}\left(U_{n,n_a} \leq Q_{n,n_a}(\widehat{\xi}) - Q_{n,n_a}(\xi_{n,n_a}) + o_p(U_{n,n_a})\right)$$

The right hand side tends to zero because $Q_{n,n_a}(\widehat{\xi}) - Q_{n,n_a}(\xi_{n,n_a}) \leq 0$ and $U_{n,n_a} > 0$, and in that case the consistency of our estimator is ensured.

So the proof is decomposed is three steps:

1. Control of $\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)$

2. Existence of a sequence ξ_{n,n_a} such that $Q(\xi_{n,n_a}) = o\left(\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)\right)$

3. Uniform control on $\mathcal{H}_{n,n_a}$: $\sup_{\xi\in\mathcal{H}_{n,n_a}} |Q_{n,n_a}(\xi) - Q(\xi)| = o_p\left(\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)\right)$

In the first step of the proof, we will show that it exists $c(\delta)$ an increasing function of δ that does not depend on n, n_a such that $\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi) \geq c(\delta) > 0$. Then, in the second and third step, we only need to show that $Q(\xi_{n,n_a}) = o_p(1)$ and $\sup_{\xi\in\mathcal{H}_{n,n_a}} |Q_n(\xi) - Q(\xi)| = o_p(1)$.

In the following for any integer $d > 0$ and any vector in $v \in \mathbb{R}^d$, $||v||_2$ denotes the Euclidian norm of v.

1. First step: Control of $\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)$.

   Let $C_p = 1 + C, C_0 = C_1 = c + C$, and B_R closed balls of radius $C.$ of the Hölder space $\mathcal{C}^{1,1}(]-1;0[\cup]0;1[)$, i.e. $||f||_H = ||f||_\infty + \sup_{z\neq z'} \frac{|f(z)-f(z')|}{|z-z'|} < R$. We have $\mathcal{H}_{n,n_a} \subset \mathcal{H} \subset B_{C_p} \times B_{C_0} \times B_{C_1}$. The Arzelà-Ascoli Theorem ensures that B_{C_p} and $B_{C_{0,1}}$ are compact for the supremum norm. Then $B_{C_p} \times B^2_{C_{0,1}}$ is a compact space (for the norm $||\xi||_\infty = \sup(||p||_\infty, ||m_0||_\infty, ||m_1||_\infty)$). As a close subset of a compact, $\mathcal{H}$ and then $\mathcal{H} \cap \{\xi : ||\xi - \xi_0||_\infty \geq \delta\}$ are compact.

   Moreover $Q(\xi)$ is continuous for the supremum norm on $\mathcal{H}$, then $Q(\mathcal{H} \cap \{\xi : ||\xi - \xi_0||_\infty \geq \delta\})$ is compact. And the condition of identification ensures that Q is minimum (and null) only for $\xi = \xi_0$.

   So, it exists $\xi^* \in \mathcal{H} \cap \{\xi : ||\xi - \xi_0||_\infty \geq \delta\}$ that does not depend on n, n_a such that $\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi) \geq \inf_{\xi\in\mathcal{H},||\xi-\xi_0||_\infty\geq\delta} Q(\xi) \geq Q(\xi^*) > 0$. Then it follows (with a slight abuse of notation) that "$o_p\left(\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)\right) = o_p(1)$" and "$o\left(\inf_{\xi\in\mathcal{H}_{n,n_a},||\xi-\xi_0||_\infty\geq\delta} Q(\xi)\right) = o(1)$".

2. Second step: Existence of a sequence ξ_{n,n_a} such that $Q(\xi_{n,n_a}) = o(1)$.

The spline properties ensures that for every ξ_0 in $\mathcal{H}$, it exists $\xi_{n,n_a} = (p_{n,n_a}, m_{0,n,n_a}, m_{1,n,n_a})$ such that $||\xi_{n,n_a} - \xi_0||_\infty = O((min(k^{p+}_{n,n_a}, k^{p-}_{n,n_a}, k^{0+}_{n,n_a}, k^{0-}_{n,n_a}, k^{1+}_{n,n_a}, k^{1-}_{n,n_a}))^{-1})$. The condition of the Theorem 4.1 ensures that $||\xi_{n,n_a} - (p, m_0, m_1)||_\infty = o(1)$ and then by continuity of Q on $\mathcal{H}$, $Q(\xi_{n,n_a})$ tends to $Q(\xi_0) = 0$.

3. Third step: uniform control of $Q_{n,n_a}(\xi) - Q(\xi)$ on $\mathcal{H}_{n,n_a}$.

We have:

$$\sup_{\xi\in\mathcal{H}_{n,n_a}} |Q_{n,n_a}(\xi) - Q(\xi)| \leq \sum_{j=p,0,1} \sup_{\xi^j\in\mathcal{H}_{n,n_a}} \left| \tfrac{1}{n} \sum_{i\in S} \widehat{E}\left(\rho_j(W,\xi)|Z=Z_i\right)^2 - E\left(\rho_p(W,\xi)|Z=Z_i\right)^2 \right|$$

In the following, we prove that:

$$\sup_{\xi\in\mathcal{H}_{n,n_a}} \left| \frac{1}{n} \sum_{i\in S} \widehat{E}\left(\rho_p(W,\xi)|Z=Z_i\right)^2 - E\left(\rho_j(W,\xi)|Z=Z_i\right)^2 \right| = O_p(l_{n,n_a}/n)$$

The same reasoning and the same results hold for the two others terms $(j = 0,1)$ of the previous sum.

First, we restrict the proof to the case where we observe an iid sample of $W = (Y, Z, TZ^*, T)$, in this case $\sum_{i\in S_a} T_i = n_a$ and $S_a = \{i \in S : T_i = 1\}$.

For any $\xi \in \mathcal{H}^p_{n,n_a}$, let $\widehat{g}_p(z,\xi) = \widehat{\mathbb{E}}(\rho_p(W,\xi)|Z=z) = B^p(z)\widehat{\mathbb{E}}(B^{p'}(Z)B^p(Z))^{-1}\widehat{\mathbb{E}}(B^{p'}(Z)\rho_p(W,\xi))$ and $g_p(z,\xi) = \mathbb{E}(\rho_p(W_1,\xi)|Z_1=z)$, with $\rho_p(W,\xi) = \frac{T}{\xi(Z^*)} - 1$. For any $\xi \in \mathcal{H}^p_{n,n_a}$ and any $i \in S$, we have:

$$\begin{aligned}
\widehat{g}^2_p(Z_i,\xi) - g^2_p(Z_i,\xi) &= \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 + 2g_p(Z_i,\xi)\left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right) \\
&\leq \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 + 2\left|g_p(Z_i,\xi)\right|\left|\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right|
\end{aligned}$$

Then by Cauchy-Schwartz inequality,

$$\begin{aligned}
\tfrac{1}{n}\textstyle\sum_{i=1}^n \widehat{g}^2_p(Z_i,\xi) - g^2_p(Z_i,\xi) &\leq \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 \\
&\quad +2\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n g_p(Z_i,\xi)^2\right)^{1/2}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2\right)^{1/2} \\
&\leq \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 \\
&\quad +2\sup\left(1, \tfrac{1}{\underline{c}} - 1\right)\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2\right)^{1/2}
\end{aligned}$$

Then to control $\frac{1}{n}\sum_{i=1}^n \widehat{g}_p^2(Z_i,\xi) - g_p^2(Z_i,\xi)$, uniformly on $\mathcal{H}_{n,n_a}^p$, we have to control uniformly $\frac{1}{n}\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2$.

Adapting the proof of Theorem 1 of Newey (1997), we can show that for any $\xi \in \mathcal{H}_{n,n_a}^p$ that $\frac{1}{n}\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 = O_p\left(\frac{l_{n,n_a}}{n} + l_{n,n_a}^{-\gamma}\right)$. However this result is not sufficient because it is not uniform. To show that this holds uniformly on $\mathcal{H}_{n,n_a}^p$, we will use various theorems related to the behavior of empirical process, as explained in van der Vaart (2000), Chapter 19 or in van der Vaart & Wellner (1996).

Up to an affine change from $[0;1]$ to $[l;u]$, the base B considered in Lemma A.2 verifies Assumption 2 of Newey (1997), i.e. $\mathbb{E}\left(B'(Z)B(Z)\right)$ has a smallest eigenvalue bounded away from 0 uniformly in $l_{n,n_a} := \dim\left(\mathcal{I}_{n,n_a}^p\right)$ by $\underline{\lambda}$ and $\sup_z ||B(z)||_2 \leq \zeta_0(l_{n,n_a}) = l_{n,n_a}^{[\max(\alpha_u,\alpha_l)+1]/2}$. The condition of the Theorem 4.1 ensures that $\zeta_0(l_{n,n_a})^2 l_{n,n_a}/n \to 0$. Let $\overline{B}(z) = B(z)\mathbb{E}\left(B'(Z)B(Z)\right)^{-1/2}$, $\overline{B}$ is such that $\sup_z ||\overline{B}(z)||_2 \leq \overline{\zeta}_0(l_{n,n_a}) = \frac{1}{\underline{\lambda}}\zeta_0(l_{n,n_a})$, with $\overline{\zeta}_0(l_{n,n_a})^2 l_{n,n_a}/n \to 0$ and $\mathbb{E}\left(\overline{B}'(Z)\overline{B}(Z)\right) = I_{l_{n,n_a}+1}$. Because means square prediction is invariant by linear transformation of regressors, we can assume without loss of generality that $\overline{B}(Z)$ is used as the base of $\mathcal{I}_{n,n_a}^p$.

Let $A_n = \mathbb{1}\{\inf_{u\in\mathbb{R}^k} u'\widehat{\mathbb{E}}\left(\overline{B}'(Z)\overline{B}(Z)\right)u \geq ||u||_2^2/2\}$, the dummy variable that the smallest eigenvalue of the empirical estimator $\widehat{\mathbb{E}}\left(\overline{B}'(Z)\overline{B}(Z)\right) = \frac{1}{n}\sum_{i=1}^n \overline{B}'(Z_i)\overline{B}'(Z_i)$ is greater than $1/2$ (or equivalently the dummy variable that the highest eigenvalue of $\left[\frac{1}{n}\sum_{i=1}^n \overline{B}'(Z_i)\overline{B}'(Z_i)\right]^{-1}$ is lower than 2). Under the conditions of Theorem 4.1, namely $l_{n,n_a}^p = o(n^{1/(2+\max(\alpha_u,\alpha_l))})$, A_n tends to 1 in probability.

Let $\overline{B}$ the matrix of size $n\times k$ of elements $\overline{B}_j(Z_i)$, and let $G_p(\xi)$ the column vector of component $\mathbb{E}(\rho_p(W,\xi)|Z = Z_i)$. We define $\widetilde{g}_p(Z_i,\xi) = \overline{B}(Z_i)(\overline{B}'\overline{B})^{-1}\overline{B}'G_p(\xi)$. Following the usual strategy (see for instance Newey (1997) or Chen & Pouzo (2012)), we use the triangle inequality to split $\frac{1}{n}\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2$ in three terms:

$$\begin{aligned}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2\right]^{1/2} \leq\ & \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - \widetilde{g}_p(Z_i,\xi)\right)^2\right]^{1/2} \\ & + \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widetilde{g}_p(Z_i,\xi) - \overline{B}(Z_i)\pi_\xi\right)^2\right]^{1/2} \\ & + \left[\tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\overline{B}(Z_i)\pi_\xi - g_p(Z_i,\xi)\right)^2\right]^{1/2},\end{aligned}$$

The second term can be bounded by the third one, because of the projection properties of $\overline{B}$:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\overline{B}(Z_i)\pi_\xi-\widetilde{g}_p(Z_i,\xi)\right)^2 &= \frac{1}{n}\sum_{1=1}^{n}\left(\overline{B}(Z_i)(\overline{B}'\overline{B})^{-1}\overline{B}'\left(B\pi_\xi-G_p(\xi)\right)\right)^2\\
&= \frac{1}{n}\left(B\pi_\xi-G_p(\xi)\right)'\overline{B}(\overline{B}'\overline{B})^{-1}\overline{B}'\left(B\pi_\xi-G_p(\xi)\right)\\
&\leq \frac{1}{n}\left(B\pi_\xi-G_p(\xi)\right)'\left(B\pi_\xi-G_p(\xi)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\overline{B}(Z_i)\pi_\xi-g_p(Z_i,\xi)\right)^2
\end{aligned}$$

Applying Lemma A.3 to the function $z^*\mapsto\mathbb{E}(\rho_p(W)|Z^*=z^*)$, we know that there exists π_ξ such $\frac{1}{n}\sum_{i=1}^{n}\left(\overline{B}(Z_i)\pi_\xi-g_p(Z_i,\xi)\right)^2=O_p(l_{n,n_a}^{-\gamma})$ uniformly on $\mathcal{H}_{n,n_a}$.

The rest of the proof is dedicated to bound the first term of inequality 3. This is sufficient to bound this term under the condition of event A_n (because A_n tends to 1 in probability). Let $\varepsilon(\xi)$ the vector of component $\varepsilon_i(\xi)=\rho_p(W_i,\xi)-g_p(Z_i,\xi)$. When the smallest eigenvalue of $\widehat{\mathbb{E}}(\overline{B}'(Z)\overline{B}(Z))=\frac{1}{n}\overline{B}'\overline{B}$ is greater than $1/2$ ($A_n=1$), we have:

$$\begin{aligned}
&A_n\sup_{\xi\in\mathcal{H}_{n,n_a}}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{g}_p(Z_i,\xi)-\widetilde{g}_p(Z_i,\xi)\right)^2\\
&\leq A_n\sup_{\xi\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{g}_p(Z_i,\xi)-\widetilde{g}_p(Z_i,\xi)\right)^2\\
&= A_n\sup_{\xi\in\mathcal{H}}\frac{1}{n}\varepsilon(\xi)'\overline{B}\left(\overline{B}'\overline{B}\right)^{-1}\overline{B}'\varepsilon(\xi)\\
&\leq 2A_n\sup_{\xi\in\mathcal{H}}\frac{1}{n^2}\varepsilon(\xi)'\overline{B}\overline{B}'\varepsilon(\xi)\\
&= 2A_n\sup_{\xi\in\mathcal{H}}\sum_{j=1}^{l_{n,n_a}}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2\\
&\leq 2A_n\sum_{j=1}^{l_{n,n_a}}\sup_{\xi\in\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2
\end{aligned}$$

Moreover, the Markov inequality ensures that there exists a constant M (uniform in l_{n,n_a}) such that :

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{l_{n,n_a}}\sup_{\xi\in\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2>M\frac{l_{n,n_a}}{n}\right)\\
&\leq \frac{n}{Ml_{n,n_a}}\mathbb{E}\left(\sum_{j=1}^{l_{n,n_a}}\sup_{\xi\in\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2\right)\\
&\leq \frac{1}{M}\max_{1\leq j\leq l_{n,n_a}}\mathbb{E}\left(\sup_{\xi\in\mathcal{H}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2\right)\\
&= \frac{1}{M}\max_{1\leq j\leq l_{n,n_a}}\mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\overline{B}_j(Z_i)\varepsilon_i(\xi)\right|\right)^2\right)
\end{aligned}$$

Conditions of regularity imply that the class $\mathcal{E}_j$ of functions $f(W_i) = \overline{B}_j(Z_i)\varepsilon_i(\xi)$ indexed by $\xi \in \mathcal{H}^p$ has for enveloppe function

$$F^{\mathcal{E}_j}(W_i) = |\overline{B}_j(Z_i)| \times \max\left(\left|\frac{T_i}{\underline{c}} - \mathbb{E}(T|Z = Z_i)\right|, \left|T_i - \mathbb{E}(\frac{T}{\underline{c}}|Z = Z_i)\right|\right),$$

which is always square integrable and such that:

$$\overline{B}_j(Z)^2(1+\underline{c}^{-1})^2 \geq \mathbb{E}\left(F^{\mathcal{E}_j}(W)^2|Z\right) \geq \overline{B}_j(Z)^2(1-\underline{c})^2.$$

Then for any $j = 1, ..., l_{n,n_a}$, because $\mathbb{E}\left(\overline{B}_j(Z_i)\varepsilon_i(\xi)\right) = 0$, Theorem 2.14.5 of van der Vaart & Wellner (1996) ensures that it exists an universal constant M_0 such that:

$$\begin{array}{l} \mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{B}_j(Z_i)\varepsilon_i(\xi)\right|\right)^2\right) \\ \leq M_0\mathbb{E}\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{B}_j(Z_i)\varepsilon_i(\xi)\right|\right) \\ \quad +M_0(1+\underline{c}^{-1}) \end{array}$$

Theorem 2.14.2 of van der Vaart & Wellner (1996) ensures that it exists another universal constant M_1 such that:

$$\mathbb{E}\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \overline{B}_j(Z_i)\varepsilon_i(\xi)\right|\right) \leq M_1(1+\underline{c}^{-1})\int_0^1 \left(1 + LogN_{[]}\left(u(1-\underline{c}), \mathcal{E}_j, ||.||_{L^2(W)}\right)\right)^{1/2} du$$

where, for a class of function $\mathcal{F} \subset L^r(W)$, the bracketing number $N_{[]}(u, \mathcal{F}, L^r(W))$ denotes the minimum number of u-bracket necessary to cover $\mathcal{F}$. A u-bracket in $L^r(W)$ is a set of the form $\{f \in \mathcal{F} : \underline{f} \leq f \leq \overline{f}\}$ with $\overline{f}, \underline{f} \in L^r(W)$ and $||\overline{f} - \underline{f}||_{L^r(W)} \leq u$.

Let $\mathcal{O}_j$ the class of functions $f(W_i) = \overline{B}_j(Z_i)\rho_p(W_i, \xi)$ indexed by $\xi \in \mathcal{H}$. For any $f_1, f_2 \in \mathcal{E}_j$ it exists $\xi_1, \xi_2 \in \mathcal{H}^p$ such that $f_q(W) = \overline{B}_j(Z)\rho_p(W, \xi_q) - \overline{B}_j(Z)\mathbb{E}(\rho_p(W, \xi_q)|Z)$. The triangle inequality ensures:

$$\begin{array}{rcl} ||f_1 - f_2||_{L^2(W)} & \leq & ||\overline{B}_j(Z)\rho_p(W,\xi_1) - \overline{B}_j(Z)\rho_p(W,\xi_2)||_{L^2(W)} \\ & + & ||\overline{B}_j(Z)\mathbb{E}(\rho_p(W,\xi_1)|Z) - \overline{B}_j(Z)\mathbb{E}(\rho_p(W,\xi_2)|Z)||_{L^2(W)} \\ & \leq & 2||\overline{B}_j(Z)\rho_p(W,\xi_1) - \overline{B}_j(Z)\rho_p(W,\xi_2)||_{L^2(W)} \end{array}$$

It follows that $N_{[]}\left(u, \mathcal{E}_j, ||.||_{L^2(W)}\right) \leq N_{[]}\left(\frac{u}{2}, \mathcal{O}_j, ||.||_{L^2(W)}\right)$.

Moreover, for any $f_1, f_2 \in \mathcal{O}_j$, it exists $\xi_1, \xi_2 \in \mathcal{H}^p$ such that $|f_1(w) - f_2(w)| \leq \frac{|\overline{B}_j(z)|}{\underline{c}^2} ||\xi_1 - \xi_2||_\infty$ with $\left(\mathbb{E}\left(\frac{\overline{B}_j(Z)^2}{\underline{c}^4}\right)\right)^{1/2} = \frac{1}{\underline{c}^2}$.

Theorem 2.7.11 of van der Vaart & Wellner (1996) ensures that:

$$N_{[]}\left(\frac{2u}{\underline{c}^2}, \mathcal{O}_j, ||.||_{L^2(W)}\right) \leq N\left(u, \mathcal{H}^p, ||.||_\infty\right)$$

where the covering number $N(u, \mathcal{F}, L^r(W))$ denotes the minimal number of $L^r(W)$ balls of radius u needed to cover the functional set $\mathcal{F}$.

Under assumptions 7.1, 7.2, 7.3 defining $\mathcal{H}^p$, Theorem 2.7.1 of van der Vaart & Wellner (1996) ensures that it exists an universal constant M_2 such that:

$$logN\left(u, \mathcal{H}^p, ||.||_\infty\right) \leq M_2 u^{-1}.$$

It follows that:

$$\begin{array}{l}\mathbb{P}\left(\sum_{j=1}^{l_{n,n_a}} \sup_{\xi\in\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^n \overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2 > M\frac{n}{l_{n,n_a}}\right) \\ \leq \frac{M_0}{M}(1+\underline{c}^{-1})\left[1 + M_1\int_0^1\left(1+\frac{4M_2}{(1-\underline{c})\underline{c}^2}u^{-1}\right)^{1/2}du\right]\end{array}$$

Then $\sum_{j=1}^{l_{n,n_a}} \sup_{\xi\in\mathcal{H}}\left(\frac{1}{n}\sum_{i=1}^n \overline{B}_j(Z_i)\varepsilon_i(\xi)\right)^2 = O_p(l_{n,n_a}/n)$.

We now extend this result to the general case of Assumption 3, when the two samples cannot be matched. Let $\breve{g}_p(z,\xi)$ the previous unfeasible estimator of $\mathbb{E}(\rho_p(W)|Z=z)$ computed under the assumption that (Y, T, Z, TZ^*) is observed in the main sample.

By triangle inequality,

$$\begin{aligned}\left[\frac{1}{n}\sum_{i=1}^n\left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2\right]^{1/2} \quad &\leq \quad \left[\frac{1}{n}\sum_{i=1}^n\left(\widehat{g}_p(Z_i,\xi) - \breve{g}_p(Z_i,\xi)\right)^2\right]^{1/2} \\ &\quad + \left[\frac{1}{n}\sum_{i=1}^n\left(\breve{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2\right]^{1/2},\end{aligned}$$

We already have shown that the second term is such that $\sup_\xi \frac{1}{n}\sum_{i=1}^n\left(\breve{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 = O_p(l_{n,n_a}/n)$.

The first term is such that:

$$
\begin{aligned}
& A_n \sup_\xi \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\widehat{g}_p(Z_i,\xi) - \check{g}_p(Z_i,\xi)\right)^2 \\
\leq\ & 2A_n \textstyle\sum_{j=1}^{l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n}\left(\sum_{i\in S} T_i \tfrac{1}{n_a}\right)\sum_{i\in S_a} \overline{B}_j(Z_i)/\xi(Z_i^*) - \sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*)\right)^2 \\
\leq\ & 6A_n \textstyle\sum_{j=1}^{l_{n,n_a}} \sup_\xi \left[\tfrac{1}{n}\sum_{i\in S} T_i\right]^2 \left[\tfrac{1}{n_a}\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right]^2 \\
& +6A_n \textstyle\sum_{j=1}^{l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n}\left(\sum_{i\in S} T_i\right)\mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z_i^*)\right)\right)^2 \\
& +6A_n \textstyle\sum_{j=1}^{l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n}\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right)^2 \\
\leq\ & 6A_n l_{n,n_a} \max_{1\leq j\leq l_{n,n_a}} \sup_\xi \left[\tfrac{1}{n_a}\textstyle\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right]^2 \\
& +\tfrac{6A_n l_{n,n_a}}{n}\left(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i\in S} T_i - \mathbb{P}(T=1)\right)^2 \max_{1\leq j\leq l_{n,n_a}} \sup_\xi \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)^2 \\
& +6A_n l_{n,n_a} \max_{1\leq j\leq l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n}\textstyle\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right)^2
\end{aligned}
$$

The first inequality holds because $A_n u'\left[\overline{B}'\overline{B}\right]^{-1} u \leq 2A_n u'u$ and $\sup_x \sum_k f_k(x) \leq \sum_k \sup_x f_k(x)$. The second because $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. The third inequality holds because $T\leq 1$ and $\mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)\right) = \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\mathbb{P}(T=1)$ and $\sum_{k=1}^K \sup_x f_k(x) \leq K\max_k \sup_x f_k(x)$.

We have

$$
\begin{aligned}
\mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)^2 &\leq \mathbb{E}\left(\overline{B}_j(Z)^2/\xi(Z^*)^2|T=1\right) \\
&\leq \tfrac{1}{\underline{c}^2\mathbb{P}(T=1)}\mathbb{E}\left(T\overline{B}_j(Z)^2\right) \\
&\leq \tfrac{1}{\underline{c}^3},
\end{aligned}
$$

and $\left(\frac{1}{\sqrt{n}}\sum_{i\in S} T_i - \mathbb{P}(T=1)\right)^2 = O_p(1)$. Then,

$$
\frac{6A_n l_{n,n_a}}{n}\left(\frac{1}{\sqrt{n}}\sum_{i\in S} T_i - \mathbb{P}(T=1)\right)^2 \max_{1\leq j\leq l_{n,n_a}} \sup_\xi \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)^2 = O_p(l_{n,n_a}/n).
$$

Moreover (by Markov inequality),

$$
\begin{aligned}
& \mathbb{P}\left(\max_{1\leq j\leq l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n}\textstyle\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right)^2 > M\tfrac{l_{n,n_a}}{n}\right) \\
\leq\ & \tfrac{1}{M}\max_{1\leq j\leq l_{n,n_a}} \mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\tfrac{1}{\sqrt{n}}\textstyle\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right|\right)^2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbb{P}\left(\max_{1\leq j\leq l_{n,n_a}} \sup_\xi \left(\tfrac{1}{n_a}\textstyle\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right)^2 > M\tfrac{l_{n,n_a}}{n}\right) \\
\leq\ & \tfrac{1}{M}\max_{1\leq j\leq l_{n,n_a}} \mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\tfrac{1}{\sqrt{n_a}}\textstyle\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right|\right)^2\right)
\end{aligned}
$$

The classes $\mathcal{F}_j = \{f : f(z,z^*,t) = \overline{B}_j(z)t/\xi(z^*), \xi \in \mathcal{H}^p\}$ (respectively $\mathcal{F}_j^1 = \{f : f(z,z^*) = \overline{B}_j(z)/\xi(z^*), \xi \in \mathcal{H}^p\}$) has for envelope function $F^{\mathcal{F}_j}(z,z^*,t) = \overline{B}_j(z)/\underline{c}$ (respectively $F^{\mathcal{F}_j^1}(z,z^*) = \overline{B}_j(z)/\underline{c}$). We have $\mathbb{E}\left(F^{\mathcal{F}_j}(Z,Z^*,T)^2\right) = \underline{c}^{-2}$ and $\mathbb{E}\left(F^{\mathcal{F}_j^1}(Z,Z^*,T)^2|T=1\right) = \mathbb{E}\left(\overline{B}_j(Z)^2|T=1\right)\underline{c}^{-2} \in [\underline{c}^{-1};\underline{c}^{-3}]$.

Theorems 2.14.5 and 2.14.2 of van der Vaart & Wellner (1996) ensure that it exists positive numbers $M_3,...M_8$ (depending only on $\underline{c}$) such that:

$$\mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n}}\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right|\right)^2\right)$$
$$\leq M_3 + M_4\int_0^1\left(1+LogN_{[]}(M_5u,\mathcal{F}_j,||.||_{L^2(W)})\right)^{1/2}du$$

and

$$\mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\frac{1}{\sqrt{n_a}}\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right|\right)^2\right)$$
$$\leq M_6 + M_7\int_0^1\left(1+LogN_{[]}(M_8u,\mathcal{F}_j^1,||.||_{L^2(W|T=1)})\right)^{1/2}du$$

Moreover, $|\frac{\overline{B}_j(z)t}{\xi_1(z^*)} - \frac{\overline{B}_j(z)t}{\xi_2(z^*)}| \leq \frac{|\overline{B}_j(z)|t}{\underline{c}^2}||\xi_1-\xi_2||_\infty$, with $\mathbb{E}(\overline{B}_j(Z)^2T) \leq 1$ and $|\frac{\overline{B}_j(z)}{\xi_1(z^*)} - \frac{\overline{B}_j(z)}{\xi_2(z^*)}| \leq \frac{|\overline{B}_j(z)|}{\underline{c}^2}||\xi_1-\xi_2||_\infty$, with $\mathbb{E}(\overline{B}_j(z)^2|T=1) \leq \underline{c}^{-1}$. Then Theorems 2.7.11 and 2.7.1 of van der Vaart & Wellner (1996) imply that it exists M_9 and M_{10} depending only on $\underline{c}$ and C such that:

$$N_{[]}(u,\mathcal{F}_j,||.||_{L^2(W)}) \leq N_{[]}(u||\overline{B}_j(Z)T||_{L^2(W)},\mathcal{F}_j,||.||_{L^2(W)}) \leq N(u/2,\mathcal{H}^P,||.||_\infty) \leq \exp(M_9u^{-1}),$$

and

$$N_{[]}(u,\mathcal{F}_j^1,||.||_{L^2(W|T=1)}) \leq N_{[]}(u\underline{c}^{1/2}||\overline{B}_j(Z)||_{L^2(W|T=1)},\mathcal{F}_j^1,||.||_{L^2(W|T=1)})$$
$$\leq N(u\underline{c}^{1/2}/2,\mathcal{H}^P,||.||_\infty) \leq \exp(M_{10}u^{-1}).$$

It follows that:

$$\sum_{j=1}^{l_{n,n_a}}\sup_\xi\left[\frac{1}{n_a}\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right]^2 = O_p(l_{n,n_a}/n),$$
$$\sum_{j=1}^{l_{n,n_a}}\sup_\xi\left[\frac{1}{n}\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right]^2 = O_p(l_{n,n_a}/n).$$

And then $\frac{1}{n}\sum_{i=1}^n\left(\widehat{g}_p(Z_i,\xi) - g_p(Z_i,\xi)\right)^2 = O_p(l_{n,n_a}/n)$.

Lemma A.2 (Smallest eigenvalue)

Let f be a positive continuous integrable function from $[0;1]$, bounded away from 0 on *every compact included in* $]0;1[$ *and* $f(t) \sim_{t\sim 1} C_1(1-t)^{\alpha_1}$ *and* $f(t) \sim_{t\sim 0} C_0 t^{\alpha_0}$. *Let* $\underline{\delta} \leq 1 \leq \overline{\delta}$, $t_0 = 0 < t_1 < ... < t_k = 1$ *such that* $t_{i+1} - t_i \in [\underline{\delta}/k; \overline{\delta}/k]$ *and* $b_i(t) = \frac{t-t_{i-1}}{t_i-t_{i-1}} \mathbb{1}_{[t_{i-1};t_i]}(t) + \frac{t_{i+1}-t}{t_{i+1}-t_i} \mathbb{1}_{[t_i;t_{i+1}]}(t)$ *for* $i = 1, ..., k-1$, $b_0(t) = \frac{t_1-t}{t_1} \mathbb{1}_{[0;t_1]}(t)$ *and* $b_k(t) = \frac{t-t_{k-1}}{1-t_{k-1}} \mathbb{1}_{[t_{k-1};1]}(t)$. *Let* $B_k(t) = [b_0(t), ..., b_k(t)]$ *the row vector of size* $k+1$. *The smallest eigenvalue of* $k^{\max(\alpha_0,\alpha_1)+1} \int_{[0;1]} B_k'(t) B_k(t) f(t) dt$ *is bounded away from zero.*

Proof of Lemma A.2:

Let $u = (u_1, ..., u_{k+1}) \in \mathbb{R}^{k+1}$, we have:

$$u \left(\int_{[0;1]} B_k'(t) B_k(t) f(t) dt \right) u' = \frac{1}{t_1^2}(u_1, u_2) \int_0^{t_1} \begin{pmatrix} (t_1-t)^2 & (t_1-t)t \\ (t_1-t)t & t^2 \end{pmatrix} f(t) dt (u_1, u_2)'$$
$$+ \sum_{i=1}^{k-2} \frac{1}{(t_{i+1}-t_i)^2} (u_{i+1}, u_{i+2}) \int_{t_i}^{t_{i+1}} \begin{pmatrix} (t_{i+1}-t)^2 & (t_{i+1}-t)(t-t_{i)} \\ (t_{i+1}-t)(t-t_i) & (t-t_i)^2 \end{pmatrix} f(t) dt (u_{i+1}, u_{i+2})'$$
$$+ \frac{1}{(1-t_{k-1})^2} (u_k, u_{k+1}) \int_{t_{k-1}}^{1} \begin{pmatrix} (t-1)^2 & (t_{k-1}-t)(t-1) \\ (t_{k-1}-t)(t-1) & (t_{k-1}-t)^2 \end{pmatrix} f(t) dt (u_k, u_{k+1})'$$

For sufficiently large k then $f(t) \geq \min(f(t_1), f(1-t_{k-1})) \geq \min(C_0, C_1) k^{-\max(\alpha_0,\alpha_1)}$ for any $t \in [t_1; t_{k-1}]$, we have:

$$\sum_{i=1}^{k-2} \frac{1}{(t_{i+1}-t_i)^2} (u_{i+1}, u_{i+2}) \int_{t_i}^{t_{i+1}} \begin{bmatrix} (t_{i+1}-t)^2 & (t_{i+1}-t)(t-t_{i)} \\ (t_{i+1}-t)(t-t_i) & (t-t_i)^2 \end{bmatrix} f(t) dt (u_{i+1}, u_{i+2})'$$
$$\geq \frac{\underline{\delta}}{3k} \left(u_2^2/2 + \sum_{i=3}^{k-1} u_i^2 + u_k^2/2 \right) C k^{-\max(\alpha_0,\alpha_1)}$$

Because $\int_{t_i}^{t_{i+1}} \begin{bmatrix} (t_{i+1}-t)^2 & (t_{i+1}-t)(t-t_{i)} \\ (t_{i+1}-t)(t-t_i) & (t-t_i)^2 \end{bmatrix} dt = (t_{i+1}-t_i)^3/3 \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$

The first term is bounded below by:

$$\frac{1}{t_1^2}(u_1,u_2)\int_0^{t_1}\begin{bmatrix}(t_1-t)^2 & (t_1-t)t\\ (t_1-t)t & t^2\end{bmatrix} f(t)dt(u_1,u_2)'$$
$$\geq \frac{C_0}{2t_1^2}(u_1,u_2)\int_0^{t_1}\begin{bmatrix}(t_1-t)^2 & (t_1-t)t\\ (t_1-t)t & t^2\end{bmatrix} t^{\alpha_0}dt(u_1,u_2)'$$
$$= \frac{C_0}{2t_1^2}(u_1,u_2)\begin{bmatrix} t_1^{\alpha_0+3}(\frac{1}{\alpha_0+1}-\frac{2}{\alpha_0+2}+\frac{1}{\alpha_0+3}) & t_1^{\alpha_0+3}(\frac{1}{\alpha_0+2}-\frac{1}{\alpha_0+3})\\ t_1^{\alpha_0+3}(\frac{1}{\alpha_0+2}-\frac{1}{\alpha_0+3}) & t_1^{\alpha_0+3}(\frac{1}{\alpha_0+3})\end{bmatrix}(u_1,u_2)'$$
$$\geq \tfrac{C_0\underline{\delta}^{\alpha_0+1}\underline{\lambda}}{2k^{\alpha_0+1}}(u_1^2+u_2^2)$$

Where $\underline{\lambda}$ is smallest eigenvalue of

$$\begin{bmatrix}\frac{1}{\alpha_0+1}-\frac{2}{\alpha_0+2}+\frac{1}{\alpha_0+3} & \frac{1}{\alpha_0+2}-\frac{1}{\alpha_0+3}\\ \frac{1}{\alpha_0+2}-\frac{1}{\alpha_0+3} & \frac{1}{\alpha_0+3}\end{bmatrix}$$

Similarly, the last term is bounded below by $\frac{K_1}{k^{\alpha_1+1}}(u_k^2+u_{k+1}^2)$, where K_1 depends only on α_1, C_1 and $\underline{\delta}$.

At least, $k^{\max(\alpha_0,\alpha_1)+1}\int_{[0;1]} B_k'(t)B_k(t)f(t)dt$ is bounded away from zero.

Lemma A.3 *Let R be the set of function from $[-1;1]$ to $\mathbb{R}$ bounded by 1 and B the base of linear normalized B-splines $[l,u]$ of cardinal $k+1$. Under Assumption 7.4 and 7.5, it exists a constant M and $\gamma>0$ such that for any $\rho\in R$ it exists $\pi_\rho\in\mathbb{R}^{k+1}$ such that:*

$$\mathbb{E}\left([\mathbb{E}(\rho(Z^*)|Z)-B(Z)\pi_\rho]^2\right)\leq Mk^{-\gamma}$$

Consequently,

$$\lim_{M\to\infty}\sup_{\rho\in[-1;1]^{[-1;1]}}\sup_{n\in\mathbb{N}}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\mathbb{E}(\rho(Z^*)|Z=Z_i)-B(Z_i)\pi_\rho)^2\right|>Mk^\gamma\right)=0$$

Proof of Lemma A.3:

Let $I=\left[l+\lfloor k^\beta\rfloor/k;u-\lfloor k^\beta\rfloor/k\right]$ with $\beta<1$, for any $z\in I$:

$$|f_Z'(z)|=\left|\int_{-1}^1 f'_{Z|Z^*=z^*}f_{Z^*}(z^*)dz^*\right|\leq\int_{-1}^1\sup_{z\in I}\left|f'_{Z|Z^*=z^*}\right|f_{Z^*}(z^*)dz^*$$

This implies that it exists D_1 such $\sup_{z\in I}|f'(z)| \leq D_1\left(1+\left(\lfloor k^\beta\rfloor/k\right)^{\min(\alpha_u,\alpha_l)-1}\right)$, similarly it exists D_2 such that $\sup_{z\in I} 1/|f_Z(z)| \leq D_2\left(\lfloor k^\beta\rfloor/k\right)^{\max(\alpha_l,\alpha_u)}$.

$$\frac{\partial}{\partial z}\mathbb{E}(\rho(Z^*)|Z=z) = -\frac{f_Z'(z)}{f_Z(z)}\mathbb{E}(\rho(Z^*)|Z=z) + \frac{1}{f_Z(z)}\int \rho(z^*) f'_{Z|z=z^*}(z) f_{Z^*}(z^*)dz^*$$

So it exists D_3 such that : $\left|\frac{\partial}{\partial z}\mathbb{E}(\rho(Z^*|Z=z)\right| \leq D_3\left(1+k^{(1-\beta)(1-\min(\alpha_l,\alpha_u))}\right)k^{(1-\beta)(\max(\alpha_l,\alpha_u))}$. Let π_ρ the vector of size k with ith component equal to 0 for $i=0,...,\lfloor k^\beta\rfloor-1$ and $i=k-\lfloor k^\beta\rfloor+1,...,k$ and ith component equal to $\mathbb{E}\left(\rho(Z^*)|Z=i/k\right)$ otherwise. Let $\overline{\gamma} = \max\left(\alpha_l,\alpha_u,1-\min(\alpha_l,\alpha_u)+\max(\alpha_l,\alpha_u)\right)$. It exists D_4 such that for all ρ: $\sup_{z\in I}|\mathbb{E}(\rho(Z^*)|Z=z)-B(z)\pi_\rho| \leq D_4k^{(1-\beta)\overline{\gamma}-1}$. It follows that, $\int_I\left[\mathbb{E}(\rho(Z^*)|Z)-B(Z)\pi_\rho\right]^2 f_Z(z)dz \leq D_4^2k^{2(1-\beta)\overline{\gamma}-2}$. Moreover, it exists D_5, D_6 such that:

$$\int_l^{l+\lfloor k^\beta\rfloor/k}\left[\mathbb{E}(\rho(Z^*)|Z)-B(Z)\pi_\rho\right]^2 f_Z(z)dz \leq \int_l^{l+\lfloor k^\beta\rfloor/k} f_Z(z)dz \leq D_5k^{(\beta-1)(\alpha_l+1)}$$

$$\int_{u-\lfloor k^\beta\rfloor/k}^{u}\left[\mathbb{E}(\rho(Z^*)|Z)-B(Z)\pi_\rho\right]^2 f_Z(z)dz \leq \int_{u-\lfloor k^\beta\rfloor/k}^{u} f_Z(z)dz \leq D_6k^{(\beta-1)(\alpha_u+1)}$$

For β sufficiently close to 1, we have $\gamma := \min\left(2-2(1-\beta)\overline{\gamma},(1-\beta)(\alpha_u+1),(1-\beta)(\alpha_l+1)\right) \geq 0$ and $M=\max(D_4^2,D_5,D_6)$ such that:

$$\mathbb{E}\left(\left[\mathbb{E}(\rho(Z^*)|Z)-B(Z)\pi_\rho\right]^2\right) \leq Mk^{-\gamma}$$

The Markov inequality implies that:

$$\lim_{M\to\infty}\sup_{\rho\in[-1;1]^{[-1;1]}}\sup_{n\in\mathbb{N}}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n(\mathbb{E}(\rho(Z^*)|Z=Z_i)-\Pi_\rho(Z_i))^2\right| > Mk^\gamma\right) = 0$$

B Appendix: Tables

Table 2: Estimation in finite samples, Multiplicative Error

		$\hat{\theta}$			$\tilde{\theta}$			Donut estimator		
Nb. of knots	Sample size	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
		A. Small Measurement Error								
	1000	-0.969	10.63	11.57	-1.091	10.61	11.80	1.867	0.059	3.543
0	5000	0.344	1.009	1.128	0.212	1.033	1.078	1.846	0.012	3.421
	25000	0.678	0.024	0.484	0.541	0.025	0.318	1.854	0.002	3.438
	1000	-0.530	24.98	25.26	-0.547	20.44	20.73	1.870	0.062	3.557
1	5000	0.351	0.328	0.451	0.291	0.334	0.419	1.854	0.012	3.449
	25000	0.512	0.078	0.34	0.444	0.079	0.277	1.850	0.002	3.425
	1000	-0.738	325.0	325.5	-0.790	312.0	312.7	1.875	0.059	3.572
2	5000	0.074	0.578	0.584	0.054	0.562	0.565	1.857	0.012	3.461
	25000	0.172	0.097	0.127	0.137	0.100	0.119	1.849	0.002	3.421
		B. Large Measurement Error								
	1000	-0.009	2.818	2.818	-0.105	2.827	2.839	2.392	0.153	5.874
0	5000	0.681	0.127	0.591	0.561	0.133	0.448	2.361	0.025	5.598
	25000	0.806	0.018	0.667	0.674	0.018	0.472	2.365	0.005	5.597
	1000	-0.188	31.85	31.89	-0.223	22.60	22.65	2.389	0.149	5.856
1	5000	0.349	0.423	0.545	0.277	0.382	0.458	2.376	0.029	5.675
	25000	0.512	0.110	0.373	0.422	0.104	0.283	2.362	0.005	5.586
	1000	0.249	943.9	943.9	0.044	700.2	700.2	2.419	0.160	6.009
2	5000	0.142	2.639	2.660	0.122	1.430	1.445	2.367	0.027	5.631
	25000	0.174	0.573	0.603	0.119	0.441	0.456	2.363	0.005	5.589

Note : Computation obtained with 1000 simulations. The same set of simulations is used for all the estimators on the same line. The set of simulations changes across lines.
$Z + 1 = (Z^* + 1) \cdot (1 + \varepsilon)$ with $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$ for the DGP with *small* measurement error and $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$ for the DGP with *large* measurement error. Number of knots equal to 0 means that p, m_0 and m_1 are approximated by linear functions on $[-1; 0]$ and $[0; 1]$. When the number of knots is 1 (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$). $\hat{\theta}$ refers to the estimator we present in the paper. $\tilde{\theta}$ differs from $\hat{\theta}$ by the fact that m_1 is estimated by local linear regression on the treated. For the Donut estimator, the Wald ratio is estimated using averages of Y and T, on individuals whose Z belong to $[-0.2; -0.1]$ and $[0.1; 0.2]$.

Table 3: Estimation in finite samples, Additive Error

		Our estimator			Naive estimator			Unfeasible estimator		
Nb. of knots	Sample size	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
					A. Small Measurement Error					
	1000	-5.006	269.3	294.3	3.815	23638	23653	0.021	0.168	0.168
0	5000	-1.729	8.133	11.12	-1.050	106.5	107.6	0.011	0.043	0.043
	25000	-0.443	2.688	2.884	-2.740	27.93	35.44	0.005	0.011	0.011
	1000	-1.250	120.4	121.9	-1.045	24.15	25.24	0.040	0.154	0.156
1	5000	-0.445	3.475	3.673	-1.557	6.614	9.038	0.005	0.043	0.043
	25000	-0.612	0.882	1.257	-2.726	18.53	25.96	0.011	0.010	0.011
	1000	1.939	836.7	840.4	-2.321	655.6	661.0	0.014	0.169	0.170
2	5000	-0.124	18.80	18.81	-1.511	15.28	17.56	0.011	0.045	0.045
	25000	0.027	1.535	1.536	-2.524	4.734	11.10	0.004	0.012	0.012
					B. Large Measurement Error					
	1000	-4.040	123.9	140.2	-0.835	1526	1526	0.021	0.168	0.168
0	5000	-1.326	6.770	8.529	-2.688	2418	2425	0.011	0.043	0.043
	25000	-0.149	1.843	1.865	-7.633	38282	38340	0.005	0.011	0.011
	1000	-2.231	890.4	895.4	-1.856	3535	3538	0.040	0.154	0.156
1	5000	-0.22	3.520	3.569	0.237	2250	2250	0.005	0.043	0.043
	25000	-0.309	1.122	1.218	0.383	10663	10663	0.011	0.010	0.011
	1000	-15.91	2×10^5	2×10^5	-2.564	9040	9046	0.014	0.169	0.17
2	5000	0.661	285.2	285.6	28.798	8×10^5	8×10^5	0.011	0.045	0.045
	25000	0.211	9.276	9.320	1.890	15287	15291	0.004	0.012	0.012

Note : Computation obtained with 1000 simulations. The same set of simulations is used for all the estimators on the same line. The set of simulations changes across lines.
$Z = Z^* + \varepsilon$ with $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$ for the DGP with *small* measurement error and $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$ for the DGP with *large* measurement error. Number of knots equal to 0 means that p, m_0 and m_1 are approximated by linear functions on $[-1;0]$ and $[0;1]$. When the number of knots is 1 (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$).

Table 4: Estimation in finite samples, Additive Error

		$\hat{\theta}$			$\tilde{\theta}$			Donut estimator		
Nb. of knots	Sample size	Bias	Var.	MSE	Bias	Var.	MSE	Bias	Var.	MSE
				A. Small Measurement Error						
	1000	-5.006	269.3	294.3	-5.244	287.2	314.7	1.851	0.059	3.483
0	5000	-1.729	8.133	11.12	-1.938	8.313	12.07	1.838	0.012	3.39 0
	25000	-0.443	2.688	2.884	-0.644	2.760	3.174	1.843	0.002	3.401
	1000	-1.250	120.4	121.9	-1.344	96.30	98.10	1.859	0.065	3.522
1	5000	-0.445	3.475	3.673	-0.594	3.509	3.862	1.843	0.012	3.409
	25000	-0.612	0.882	1.257	-0.767	0.901	1.490	1.841	0.002	3.392
	1000	1.939	836.7	840.4	0.995	793.6	794.6	1.861	0.060	3.525
2	5000	-0.124	18.8	18.81	-0.237	15.61	15.67	1.846	0.012	3.421
	25000	0.027	1.535	1.536	-0.046	1.357	1.359	1.839	0.002	3.386
				B. Large Measurement Error						
	1000	-4.040	123.9	140.2	-4.243	125.3	143.3	2.335	0.136	5.586
0	5000	-1.326	6.77	8.529	-1.528	6.899	9.234	2.301	0.024	5.321
	25000	-0.149	1.843	1.865	-0.344	1.897	2.015	2.307	0.005	5.325
	1000	-2.231	890.4	895.4	-2.263	1008	1013	2.328	0.131	5.550
1	5000	-0.220	3.520	3.569	-0.37	3.474	3.612	2.315	0.027	5.385
	25000	-0.309	1.122	1.218	-0.456	1.106	1.315	2.302	0.005	5.307
	1000	-15.91	2×10^5	2×10^5	-4.12	20937	20954	2.356	0.135	5.685
2	5000	0.661	285.2	285.6	0.426	189.6	189.8	2.309	0.025	5.357
	25000	0.211	9.276	9.320	0.097	6.973	6.982	2.303	0.005	5.309

Note : Computation obtained with 1000 simulations. The same set of simulations is used for all the estimators on the same line. The set of simulations changes across lines.
$Z = Z^* + \varepsilon$ with $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$ for the DGP with *small* measurement error and $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$ for the DGP with *large* measurement error. Number of knots equal to 0 means that p, m_0 and m_1 are approximated by linear functions on $[-1;0]$ and $[0;1]$. When the number of knots is 1 (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$). $\hat{\theta}$ refers to the estimator we present in the paper. $\tilde{\theta}$ differs from $\hat{\theta}$ by the fact that m_1 is estimated by local linear regression on the treated. For the Donut estimator, the Wald ratio is estimated using averages of Y and T, on individuals whose Z belong to $[-0.2;-0.1]$ and $[0.1;0.2]$.

References

Ai, C. & Chen, X. (2003), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **71**(6), pp. 1795–1843.

Barreca, A. I., Guldi, M., Lindo, J. M. & Waddell, G. R. (2011), 'Saving babies? revisiting the effect of very low birth weight classification', *The Quarterly Journal of Economics* **126**(4), 2117–2123.

Barreca, A. I., Lindo, J. M. & Waddell, G. R. (2011), Heaping-induced bias in regression-discontinuity designs, Working Paper 17408, National Bureau of Economic Research.

Battistin, E., Brugiavini, A., Rettore, E. & Weber, G. (2009), 'The retirement consumption puzzle: Evidence from a regression discontinuity approach', *American Economic Review* **99**(5), 2209–26.

Berkson, J. (1950), 'Are there two regressions?', *Journal of the American Statistical Association* **45**(250), 164–180.

Blundell, R., Chen, X. & Kristensen, D. (2007), 'Semi-nonparametric iv estimation of shape-invariant engel curves', *Econometrica* **75**(6), pp. 1613–1669.

Bound, J., Brown, C. & Mathiowetz, N. (2001), Measurement error in survey data, *in* J. Heckman & E. Leamer, eds, 'Handbook of Econometrics', Vol. 5 of *Handbook of Econometrics*, Elsevier, chapter 59, pp. 3705–3843.

Burman, P. & Chen, K.-W. (1989), 'Nonparametric estimation of a regression function', *Annals of Statistics* **17**(4), 1567–1596.

Cahuc, P., Carcillo, S. & Le Barbanchon, T. (2014), Do Hiring Credits Work in Recessions? Evidence from France, IZA Discussion Papers 8330, Institute for the Study of Labor (IZA).

Chen, X. (2007), Large sample sieve estimation of semi-nonparametric models, *in* 'Handbook of Econometrics', Vol. 6 - Part B, Elsevier, chapter 76, pp. 5549–5632.

Chen, X. & Pouzo, D. (2012), 'Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals', *Econometrica* **80**(1), 277–321.

Darolles, S., Fan, Y., Florens, J. & Renault, E. (2011), 'Nonparametric instrumental regression', *Econometrica* **79**(5), pp. 1541–1565.

D'Haultfœuille, X. (2010), 'A new instrumental method for dealing with endogenous selection', *Journal of Econometrics* **154**, 1–15.

D'Haultfœuille, X. (2011), 'On the completeness condition in nonparametric instrumental problems', *Econometric Theory* **27**(03), 460–471.

Dong, Y. (2014), 'Regression discontinuity applications with rounding errors in the running variable', *Journal of Applied Econometrics* pp. n/a–n/a.

Hahn, J., Todd, P. & der Klaauw, W. V. (1999), Evaluating the effect of an antidiscrimination law using a regression-discontinuity design, Working Paper 7131, National Bureau of Economic Research.

Hahn, J., Todd, P. & Van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **69**(1), pp. 201–209.

Horowitz, J. L. & Manski, C. F. (1995), 'Identification and robustness with contaminated and corrupted data', *Econometrica* **63**(2), pp. 281–302.

Hu, Y. & Ridder, G. (2012), 'Estimation of nonlinear models with mismeasured regressors using marginal information', *Journal of Applied Econometrics* **27**(3), pp. 347–385.

Huang, J. Z. (1998), 'Projection estimation in multiple regression with application to functional anova models', *Annals of Statistics* **26**(1), 242–272.

Hullegie, P. & Klein, T. J. (2010), 'The effect of private health insurance on medical care utilization and self-assessed health in Germany', *Health Economics* **19**(9), 1048–1062.

Imbens, G. W. & Lemieux, T. (2008), 'Regression discontinuity designs: A guide to practice', *Journal of Econometrics* **142**(2), 615–635.

Lee, D. S. & Card, D. (2008), 'Regression discontinuity inference with specification error', *Journal of Econometrics* **142**(2), 655 – 674. The regression discontinuity design: Theory and applications.

Lee, D. S. & Lemieux, T. (2010), 'Regression Discontinuity Designs in Economics', *Journal of Economic Literature* **48**(2), 281–355.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**, pp. 147–168.

Newey, W. K. & Powel, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), pp. 1565–1578.

Pei, Z. (2011), Regression discontinuity design with measurement error in the assignment variable.

Peter, H. & Joel, H. L. (2005), 'Nonparametric methods for inference in the presence of instrumental variable', *The Annals of Statistics* **33**(6), pp. 147–168.

Schennach, S. (2004), 'Estimation of nonlinear models with measurement error', *Econometrica* **72**(1), pp. 33–75.

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

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes With Applications to Statistics*, Springer. ISBN 0-387-94640-3.

Wansbeek, T. & Meijer, E. (2000), *Measurement error and Latent Variables in Econometrics*, Vol. 37 of *Advanced Textbooks in Economics*, Elsevier. ISBN 0-444-88100-X.