I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10801

Regression Discontinuity Design with Continuous Measurement Error in the Running Variable

Laurent Davezies
Thomas Le Barbanchon

MAY 2017

DISCUSSION PAPER SERIES

IZA DP No. 10801

Regression Discontinuity Design with Continuous Measurement Error in the Running Variable

Laurent Davezies
CREST

Thomas Le Barbanchon
Bocconi University, CEPR and IZA

MAY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

Since the late 90s, Regression Discontinuity (RD) designs have been widely used to estimate Local Average Treatment Effects (LATE). When the running variable is observed with continuous measurement error, identification fails. Assuming non-differential measurement error, we propose a consistent nonparametric estimator of the LATE when the discrepancy between the true running variable and its noisy measure is observed in an auxiliary sample of treated individuals, and when there are treated individuals at any value of the true running variable - two-sided fuzzy designs. We apply our method to estimate the effect of receiving unemployment benefits.

JEL Classification: C21, C14, C51

Keywords: regression discontinuity design, measurement error

Corresponding author:
Thomas Le Barbanchon
Bocconi University
Via Roentgen 1
20136 Milano
Italy
E-mail: thomas.lebarbanchon@unibocconi.it

* We especially thank Yingyao Hu, Tom Wansbeek and two anonymous referees. We thank seminar participants at Bocconi, Crest, the 12th CSEF-IGIER symposium, and the 2016 IAAE conference, for their helpful comments. We are especially grateful to Elise Coudin, Yingyao Dong, Xavier d'Haultfoeuille, Roland Rathelot, Tom Schmitz, Diego Ubfal, Arne Uhlendorff for comments on our research. Thomas Le Barbanchon is an IGIER affiliate. We thank Benjamin Walter for excellent research assistance.

1 Introduction

The regression discontinuity (RD) design has been widely emphasized for its internal validity to estimate local average 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 is determined (at least partially) by a continuous running variable being on either side of a fixed threshold. If the joint distribution of the realized outcome, the treatment and the running variable is observed, then a local average treatment effect (LATE) is identified (Hahn et al., 2001). However, running variables may be subject to measurement error. In some instances, survey data or administrative data at hand are not specifically designed to measure the running variable of interest and they lack the information needed to compute the running variable according to the exact definition of the eligibility criteria. Then, smoothly distributed measurement errors in the running variable have important consequences: the discontinuity in the assignment probability vanishes (see e.g., Hullegie & Klein, 2010 or Cahuc et al., 2014). As a consequence, the identification of the LATE fails and the usual RD estimators using the running variable observed with measurement error are inconsistent.[1]

In this paper, we show that the identification of the LATE can be recovered when the econometrician observes auxiliary information on the discrepancy between the running variable and its noisy measure *in a sample of treated individuals*.[2] This happens when the agency/institution in charge of delivering the treatment collects data on treated individuals, specifically their eligibility - the true running variable. The auxiliary information is particularly useful in two-sided fuzzy RD designs with treated individuals at any value of the running variable, as the econometrician then observes the error structure on the treated on the whole support of the true running variable. In that case, to recover identification, we only need to assume that conditional on the true running variable, its noisy measure is independent of the treatment and of the potential outcomes, i.e. the measurement error is non-differential (see, e.g., Bound et al., 2001). In particular, we do not need to assume that the difference between the noisy and the true running variable is independent of the true running variable, i.e. classical measurement errors, and we do not need supplementary parametric assumptions. As usual in the nonparametric IV literature (Newey & Powell, 2003), we assume that the noisy measure is sufficiently related to the true running variable, which formally corresponds to a completeness condition. These conditions allow us to

[1] The only formal discussion of identification failure that we are aware of, is in the Appendix of Battistin et al. (2009). Consequently, we provide a formal proof of this result in this paper.

[2] The use of auxiliary data on the measurement error is a usual route to recover identification, as described by Carroll et al. (2006) or Chen et al. (2011)

use the structure of the error observed on treated individuals to identify the true running variable for non-treated individuals. More generally, we identify the joint distribution of the treatment, the mismeasured and the true running variables and the realized outcome. Consequently, the LATE is identified. Besides, the density of the true running variable is also identified and can be used to test for non-manipulation in the running variable in the spirit of McCrary (2008).

We then propose a nonparametric estimator of the LATE. We adopt a sieve estimation strategy, commonly used in the nonparametric IV literature (e.g. Ai & Chen, 2003; Newey & Powell, 2003; Chen, 2007; Chen & Pouzo, 2012). To take into account the specificity of our framework, we adapt the usual regularity conditions of the nonparametric IV estimation literature. First, we take into account the discontinuity of our functions of interest, the take-up rate and the expected outcome, at the threshold of the RD design. Second, we relax usual conditions on the density of the instrumental variable to cope with the fact that a mismeasured variable usually vanishes at the boundary of its support. We prove that under those relaxed conditions, our sieve estimator is consistent. Monte-Carlo simulations show that our estimator behaves well in finite samples, while naive estimators ignoring measurement error behave quite erratically.

We apply our method to the analysis of the effect of receiving Unemployment Insurance (UI) benefits. We analyze the French context, where dismissed workers are eligible for UI if they have worked more than 182 days or more than 910 hours in the previous year. We use Social Security (SS) data, covering both UI claimants and non-claimants, and we match the data with the UI register to determine claiming status. First, as we do not observe the number of hours worked in our main SS data, the eligibility rule gives rise to a two-sided fuzzy design with respect to the number of days worked.[3] Second, we only observe a noisy measure of the number of days worked in our main SS data. Consequently, the UI take-up conditional on the noisy running variable is continuous at the RD threshold. The matching of our main SS data with the UI registers also yields the number of days worked considered by the UI agency to grant eligibility. Assuming non-differential measurement error, we then follow our proposed estimation procedure. We find that claimants induced to start a UI claim (lasting at most 7 months) because they worked more than 182 days in their previous job, are more likely to find jobs in the same industry as the pre unemployment job. Ignoring measurement error in this application yields to estimators with non finite variance and then to unstable estimates of the LATE. On the contrary, our method provides a consistent estimator. Interestingly, we estimate a new

[3] Below the days-threshold, some workers are eligible because they worked enough hours. Above the threshold, the non-compliance may be explained by stigma effects.

treatment effect - receiving UI vs. not receiving at all - while most RD analysis document the effect of UI at the intensive margin - receiving UI during a long period vs. receiving UI during a short period (for example, Card et al., 2007; Lalive, 2008). The focus on the intensive margin in the previous literature could be partly explained by the difficulty to observe without error the running variables for non-claimants.

Our main contribution is to provide a new tool for practitioners who face continuous measurement error in RD designs with treated individuals on the whole support of the true running variable. Our approach enables the RD analysis of government programs (UI, health insurance, hiring credits...), when the econometrician only observes a proxy of one running variable and the take-up is then a continuous function of the proxy variable. This happens when the econometrician cannot compute from the data the running variable following the exact definition stated by the program agency. For example, in Hullegie & Klein (2010), the survey only reports monthly income, while annual income is used by the government to allocate the treatment. In Cahuc et al. (2014), administrative data do not contain the workers' labor contract types, while this information is used to compute the firm size conditioning the access to treatment. Other examples of common proxies of the running variable are past values, or predicted values based on covariates. More generally, our approach can be applied as soon as individuals (or firms) self-select into the treatment, whose cost is a discontinuous function of private information, disclosed for the treatment group only. For example, our approach is well suited to the estimation of the returns to an extra year of education - the treatment of interest here - when households with income below a given threshold are eligible to reduced tuition fees. The population of interest is then pupils at school in year $N-1$, while treated individuals are those who stay at school in year N. Schools usually keep a record of the income of the current pupils' parents to apply reduced fees. Then the econometrician only observes the year $N-1$ income for the untreated pupils and the probability to stay at school in year N would be a continuous function of year $N-1$ income. However our method could be applied as the evolution of parental income between year $N-1$ and year N is observed for the treated sample and there are pupils going to school in year N for any level of year N income.

Our paper complements previous RD work facing *non-continuous* forms of mismeasurement in the running variable. Dong (2014), Barreca, Guldi, Lindo & Waddell (2011) and Barreca, Lindo & Waddell (2011) consider *rounding errors*.[4] Battistin et al. (2009) consider *contaminated* or *corrupted* data sampling, in which the observed running variable Z

[4]Lee & Card (2008) also study the consequences of discrete running variables, however they do not explicitly consider that the discreteness of the observed variable is the result of measurement error of a true continuous underlying variable. Their approach is to allow for specification errors in the model, which affects the precision of the estimated treatment effect.

is a mixture of the true running variable (Z^*) and of a noisy proxy (i.e. $\mathbb{P}(Z = Z^*) > 0$, see Horowitz & Manski, 1995). Our paper also complements recent contributions adopting *parametric* approaches in one-sided RD designs, where individuals cannot be treated if their true running variable is below the RD threshold (Hullegie & Klein, 2010, Pei, 2011). In a working paper version (Davezies & Le Barbanchon, 2014), we adapt our approach to one-sided RD designs. As Hullegie & Klein (2010) and Pei (2011), we then need to add structure on the measurement error and use a deconvolution argument to restore identification.

The paper is organized as follows: in the second section, we discuss how a continuous measurement error in the running variable smooths any discontinuity in assignment and leads to the loss of identification and of consistency of usual RD estimators. In the third section, we present and discuss our identification results. The fourth section is devoted to estimation issues, we then provide Monte-Carlo simulations to investigate the finite sample properties of our estimator. In the fifth section, we apply our method to the analysis of the effect of receiving unemployment benefits. The last section concludes. We provide on our personal websites programs to perform our proposed estimation.[5]

2 Framework

2.1 Regression Discontinuity design

Let T be a binary variable of treatment, Z^* a continuous random variable with support $\text{Supp}(Z^*) \subset \mathbb{R}$ and 0 an interior point of this support, 0 is the cutoff/threshold 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 actual treatment status T depends on the running variable Z^*. We also define the potential treatment $T(z^*)$ for any $z^* \in \text{Supp}(Z^*)$. We consider the following assumptions of the RD design:

Assumption 1 (RD Design)

1. *Discontinuity in the take-up:* $\lim_{z\to 0^+} \mathbb{E}(T|Z^* = z) > \lim_{z\to 0^-} \mathbb{E}(T|Z^* = z)$
2. *Monotonicity: It exists a neighborhood of* 0 *included in Supp*(Z^*)*, such that with probability 1,* $z^* \mapsto T(z^*)$ *is non-decreasing for* z^* *in that neighborhood.*

[5]Matlab programs are posted. Stata programs are under production.

3. *Continuity in the potential outcomes:* $z^* \mapsto \mathbb{E}(Y(0)|Z^* = z^*, T(0^+), T(0^-) = 0)$ *and* $z^* \mapsto \mathbb{E}(Y(1)|Z^* = z^*, T(0^+) = 1, T(0^-))$ *are almost-surely continuous at* $z^* = 0$.

4. *Continuity in the proportions of individual types (compliers, never- and always-takers):* $z^* \mapsto \mathbb{P}\left(T(0^-) = t^-, T(0^+) = t^+|Z^* = z^*\right)$ *is continuous at* $z^* = 0$ *for* $(t^-, t^+) = (0, 1)$ *or* $(0, 0)$ *or* $(1, 1)$.

Assumption 1.1 states that the assignment probability (or take-up) is discontinuous at the cutoff, whose value is known by the econometrician. Assumption 1.2 is a form of monotonicity condition. It rules out the existence of defiers who would abandon the treatment had they crossed the cutoff, ie individuals such that $T(0^-) = 1$ and $T(0^+) = 0$. Compliers correspond to individuals such that $T(0^-) = 0$ and $T(0^+) = 1$. Never-takers are such that $T(0^-) = T(0^+) = 0$ and always-takers such that $T(0^-) = T(0^+) = 1$ (Imbens & Angrist, 1994). Assumption 1.3 states that the conditional expectations of the potential outcomes for compliers (respectively for always-takers and for never-takers) are continuous at the cutoff (Imbens & Lemieux, 2008). Last, Assumption 1.4 ensures that the proportions of compliers, never-takers and always-takers are the same just below and just above the threshold, so that the change in the take-up rate at the cutoff equals the proportion of compliers (at the cutoff). In their seminal paper, instead of Assumptions 1.3 and 1.4, Hahn et al. (2001) assume a local independence condition that ensures the continuity of the average treatment effects and of the proportions of compliers, never-takers and always-takers around the threshold.[6] Here we directly state our assumptions in term of continuity of the potential outcomes and of the types proportion, as in Assumption E in Rokkanen (2015) or Assumption 1 in Gerard et al. (2015). Under Assumption 1, the law of iterated expectation applied to $\mathbb{E}(Y|Z^* = 0^\pm)$ ensures that the Local Average Treatment Effect (LATE, see Imbens & Angrist, 1994 and Hahn et al., 2001),

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

is equal to a Wald's ratio: $\dfrac{\mathbb{E}(Y|Z^* = 0^+) - \mathbb{E}(Y|Z^* = 0^-)}{\mathbb{E}(T|Z^* = 0^+) - \mathbb{E}(T|Z^* = 0^-)}$.

It follows that if the joint distributions of (Y, Z^*) and (T, Z^*) are identified, θ is identified. In this paper, our general framework is less favorable: we observe a proxy variable of Z^*.

[6] We refer here to Theorem 3 of Hahn et al. (2001).

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 (Non-differential Continuous Measurement error)

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

2. $Z|Z^* = z^*$ *admits a continuous density with respect to the Lebesgue measure for all* z^*.

The first part of this assumption states that the noisy measure does not yield any supplementary information on the treatment status or outcomes of interest if information on the true running variable is available (Bound et al., 2001). Classical measurement error, where $Z = Z^* + \varepsilon$ with ε is independent of $(T, Y(0), Y(1), Z^*)$, verifies this assumption. More generally, if $Z = h(Z^*, \varepsilon)$ with h an unknown function and $\varepsilon \perp\!\!\!\perp (T, Y(0), Y(1), Z^*)$, such assumption also holds.
The second part of Assumption 2 specifies that the measurement error is continuous. In the RD context, where Z^* is a continuous variable, assuming that the measurement variable is continuous seems a natural starting point. Indeed, it is empirically relevant in many contexts: when the running variable is the score to test A and only scores to test B are observed, when the running variable is the population size of a given area at date t and only size at date $t-1$ is observed, when the running variable is total income and only taxable income is observed... The assumption of continuity makes our approach complementary to Battistin et al. (2009) and Dong (2014). Battistin et al. (2009) considers *contaminated* data, where only a fraction of the data is observed with error. Dong (2014) considers that the running variable is rounded. In practice, the econometrician may apply each methodology depending on the type of error at hand. Note that our main identification result can be extended to situations with noncontinuous measurement error. However, the completeness assumption presented below, and which is essential to our result, is more likely to hold when measurement error is continuous. That is why we prefer to assume continuity in the first place.

2.3 Lack of identification

When the measurement error in the running variable is continuous, it is known that the identification of the LATE fails (see for example the formal discussion of the consequences of classical measurement error in the Appendix of Battistin et al., 2009). Intuitively, when the econometrician only observes the noisy variable, it is no longer possible to isolate individuals whose true running variable is just above or just below the cutoff. More precisely, Proposition 1 below states that non-differential continuous measurement error in the running variable smooths out the discontinuity in the take-up rate (as a function of the noisy variable) at the cutoff. Consequently, the denominator of the previous naive Wald ratio $\mathbb{E}(T|Z=0^+) - \mathbb{E}(T|Z=0^-)$ is equal to zero.

Proposition 1 (Continuity of the take-up and of the marginal density)
Under Assumption 2, if the measurement error is such that $\mathbb{E}\left(\sup_z f_{Z|Z^*}(z)\right) < \infty$, *then both functions* $z \mapsto \mathbb{E}(T|Z=z)$ *and* $z \mapsto f_Z(z)$ *are continuous on the interior of the support of* Z.

The proof of Proposition 1 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. As a consequence, $z \mapsto \mathbb{E}[T|Z=z]$ is continuous on any interior point of $\text{Supp}(Z)$.

Similarly, the expectation of the outcome conditional on the noisy variable is continuous at the cutoff and the numerator of the Wald ratio is zero. As a consequence, the naive estimator of the LATE, where the running variable is just proxied by the noisy variable, features pathological properties (see Proposition 2 in the Appendix). This contrasts with the consequences of covariate measurement error in a treatment effect analysis relying on conditional independence assumptions, for example matching estimators (Battistin & Chesher, 2014). In such an analysis, ignoring measurement error leads usual estimators to converge at the standard rate to a biased value. The asymptotic bias can be approximated at the first order when the measurement error variance goes to zero, and the bias becomes negligible for sufficiently small variance of measurement error. In such a context, Battistin & Chesher (2014) advocate to perform sensitivity analysis based on bias correction.

Proposition 2 in the Appendix shows that continuous measurement error in the running variable has more adverse consequences: even a very small variance of the measurement error results in the inconsistency of the usual estimator. The strategy followed by Battistin & Chesher (2014) cannot be applied in our context. Our situation is similar to what happens for a two-stages least squares estimation where the instrument is uncorrelated with the endogenous variable (the denominator of the estimator tends to zero).
The result of Proposition 1 also contrasts with what happens when the data are *only* contaminated, i.e. when only a fraction of the data is observed with error. Battistin et al. (2009) show that, provided that the measurement error is non-differential, the LATE is identified in contaminated data and the usual naive Wald estimator is consistent. Intuitively, when a fraction of individuals is observed without error, there is still a discontinuous jump in the take-up rate at the cutoff. Ironically, the asymptotic theory reveals that small continuous measurement errors in the running variable for all individuals have more dramatic consequences on identification than large errors on a fraction of the data.
Lastly, when the measurement error is continuous, the marginal density of the noisy running variable Z is also continuous (see Proposition 1). This means that the widely used McCrary test of non-manipulation of the observed noisy running variable never rejects the null hypothesis of non-manipulation (McCrary, 2008). Similarly, the continuity of the conditional expectation of the covariates with respect to the mismeasured running variable can never be rejected.[7] The presence of continuous measurement error has also consequences on the econometrician's ability to evaluate the credibility of the RD design.
In the next section, we show that if auxiliary information is available on the treated (and only on the treated), we are able to identify the LATE and the density of the true running variable Z^*. Namely, the identification of the density Z^* allows to investigate manipulation behaviors, revealed by discontinuities of f_{Z^*}, as in McCrary (2008).

3 Identification with auxiliary information

To recover the identification of the LATE in the presence of continuous measurement errors in the running variable, we rely on an auxiliary sample of treated individuals, for whom we observe both the noisy and the true running variables (Z, Z^*). Such auxiliary data is particularly informative when there are treated individuals at any value of the true running variable (large support condition). In this section, we first state the assumptions necessary to derive our identification result and discuss empirical applications where they are likely

[7] This is true as long as the measurement error is non-differential in the most general terms: for any covariate X, if $Z \perp\!\!\!\perp X|Z^*$ then under the weak regularity conditions given in Proposition 1, $z \mapsto \mathbb{E}(X|Z = z)$ is continuous.

to hold. Second, we present our main identification result.

3.1 Assumptions

Assumption 3 (Observation from the data)
$F_{Y,T,Z}$ is identified by the observation of a main iid sample and $F_{Z,Z^*|T=1}$ is identified by the observation of an auxiliary iid sample of treated units.

Assumption 3 holds when we observe a sample of (Y, Z, T, TZ^*), i.e. when the auxiliary sample is a subset of the main sample and when we can match the two samples. However Assumption 3 is more general: samples may not be nested.[8] Observing the true running variable for the treated naturally occurs when individuals apply voluntarily 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 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. The agency data can then be matched with the main sample to obtain the joint distribution of Z, Z^* for the treated individuals. 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), unemployment insurance as in Section 5 etc.

Assumption 4 (Large support condition/ two-sided fuzzy RD design)
For any $z^* \in Supp(Z^*)$, $\mathbb{P}(T = 1|Z^* = z^*) > 0$

This support condition on the take-up (propensity score) states that there are treated individuals on the whole support of the true running variable. We refer hereafter to this Assumption as two-sided fuzzy Regression Discontinuity, to convey the idea that there are treated individuals on both sides of the threshold. Combined with Assumption 3, Assumption 4 implies that the econometrician observes the distribution of the measurement error for any value of the true running variable. As a result, the identification of the LATE, proposed below, will not need any parametric assumptions on the measurement error structure.
Assumption 4 rules out one-sided RD, i.e. when $\mathbb{P}(T = 1|Z^* < 0) = 0$, such as sharp RD. It also rules out designs where there exists $\eta < 0$ such that $\mathbb{P}(T = 1|Z^* < \eta) = 0$. We adapt

[8] Note that we cannot adopt the approach of Chen et al. (2005) who also use information from an auxiliary sample to recover identification in the presence of measurement errors. Their approach require that the auxiliary sample identifies $F_{Y,T,Z^*|Z}$. However, our auxiliary sample is less informative, as it is restricted to treated individuals and it does not contain the outcome variable. Note also that the approach of Chen et al. (2005) requires a parametric model for $T(z^*)$, while we do not need any parametric assumption here.

our method to these designs in our working paper (Davezies & Le Barbanchon, 2014). In such designs, the auxiliary data is less informative and we need parametric assumptions on the measurement error structure, as in Hullegie & Klein (2010) or Pei (2011). Namely, we assume classical measurement error, where $Z = Z^* + \varepsilon$ with Z^* and ε independent, and we use a deconvolution argument to restore identification.

Two-sided fuzzy designs occur when the treatment eligibility is granted to individuals satisfying at least one condition among several criteria. Let us consider an agency delivering a program if one of the two running variables (Z^* or S^*) crosses the 0-threshold: $T = \max(\mathbb{1}(Z^* > 0), \mathbb{1}(S^* > 0))$. When the econometrican observes only a proxy Z of Z^* and has no information on the second criteria S^* for the whole population, the support condition is verified provided that 0 is an interior point of $Supp(S^*|Z^* = z^*)$ for all z^*. Our application in Section 5 illustrates this situation: eligibility to unemployment insurance depends on several criteria, but we only observe one of them.

Two-sided fuzzy designs also occur more generally when the discontinuous rule is an instrument for a "second-round" treatment. Let us take the example of the estimation of the returns to schooling. The treatment is defined as attending school during an extra year (Card, 2001). It is often the case that reductions in tuition fees are granted if the parents' income are below a given threshold. This rule generates a discontinuity in the take-up rate at the parents' income cutoff, although some kids above the threshold attend school for an extra year anyway (Assumption 4 holds). Because schools usually record the income of the parents of the current pupils, Assumption 3 holds. However, the income for former pupils who stopped schooling, could only be proxied by the parents' income observed in the school records of the previous year (Assumption 2 holds). Alternatively the income for former pupils could be observed in another dataset and probably with error.

The example of the estimation of the returns to schooling highlights that our Assumptions are likely to hold in the general case when individuals (or firms) select themselves into the treatment, whose cost is a discontinuous function of a private information Z^*, only disclosed for the treated.

Lastly, we make the following technical assumption, which is untestable without supplementary restrictions (Canay et al., 2013).

Assumption 5 (Completeness Condition)
$\forall g$ *such that* $\mathbb{E}(|g(Z^*)|) < +\infty$, $\mathbb{E}(g(Z^*)|Z) = 0 \Rightarrow g = 0$.

Assumption 5 means that there is enough variation in Z to identify $g \in L^1(Z^*)$ when we observe $\mathbb{E}(g(Z^*)|Z)$. Note that, given Assumptions 2 and 4, the completeness condition also holds conditional on treatment $T = 1$, which is used in the proof of Theorem 1.[9]

[9] Consider g such that $\mathbb{E}(|g|(Z^*)|T = 1) < \infty$ and $\mathbb{E}(g(Z^*)|T = 1, Z) = 0$ and let $\widetilde{g}(Z^*) = g(Z^*)\mathbb{E}(T|Z^*)$.

Assumption 5 is common in the nonparametric IV framework (Newey & Powell, 2003) or in models with measurement error (Hu & Schennach, 2008). It means that the mapping from the unknown distribution of Z^* to the observed distribution of Z is injective. It generalizes the rank condition necessary for identification of the distribution of Z^* when random variables Z^* and Z have finite support.[10] Examples of data generating processes such that the completeness condition holds can be found in Newey & Powell (2003) and D'Haultfœuille (2011). Let us detail some examples. A first example is a Berkson type error model: $Z^* = Z + \eta$ with $\eta \perp\!\!\!\perp Z$, where η has a nonzero characteristic function and has heavy tails (see Wansbeek & Meijer, 2000, Section 2.5 for a simple exposition of Berkson's model or the original paper of Berkson, 1950). This type of error arises when the econometrician imputes the running variable, such that $Z = \mathbb{E}(Z^*|X)$ where X are observed covariates. The underlying imputation model is then $Z^* = \mathbb{E}(Z^*|X)+\eta$ with $\eta \perp\!\!\!\perp X$, which follows the Berkson type error structure. Second the standard classical measurement error model $Z = Z^* + \epsilon$ with $\epsilon \perp\!\!\!\perp Z^*$ also verifies the completeness condition under regularity condition on the distribution of ϵ. The more general error model $Z = \mu\left(\nu(Z^*) + \epsilon\right)$ with $\epsilon \perp\!\!\!\perp Z^*$, ensures that Assumption 5 holds, provided that μ and ν are bijective and the Fourier transform of ϵ has isolated zeros (see Proposition 2.4 in D'Haultfœuille, 2011).

3.2 Main identification result

Theorem 1 (Identification of $F_{Z^*,Z,T,Y}$ and of the LATE θ)
Under Assumptions 1, 2, 3, 4 and 5, 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 guidelines about the identification of the LATE θ.

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, we have $\mathbb{E}(T|Z, Z^*) = p(Z^*)$, $\mathbb{E}(Y|Z, Z^*) = m(Z^*)$ and $\mathbb{E}(Y|Z) = \mathbb{E}(m(Z^*)|Z)$. Next, under Assumptions 2 and 4, the law of iterated expectation ensures that m and p are solutions of the following moment

Assumption 4 ensures that $g = 0$ if and only if $\widetilde{g} = 0$. The law of iterated expectation and Assumptions 2 and 4 ensure that $\mathbb{E}(|\widetilde{g}|(Z^*)) = \mathbb{E}(|g|(Z^*)T) = \mathbb{E}(|g|(Z^*)|T = 1)\mathbb{E}(T) < \infty$ and $\mathbb{E}(\widetilde{g}(Z^*)|Z) = \mathbb{E}(g(Z^*)\mathbb{E}(T|Z, Z^*)|Z) = \mathbb{E}(g(Z^*)|Z, T = 1)\mathbb{E}(T|Z) = 0$. The completeness condition ensures that $\widetilde{g} = 0$ and we conclude that $g = 0$.

[10] If Z and Z^* were discrete, the rank condition writes $rk[\mathbb{P}(Z^* = i|Z = j)]_{i=1,\dots,I,j=1,\dots,J} = I$ where I is the cardinality of the support of Z^*.

conditions:

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

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

The left hand side of these equations can not be directly estimated, because the joint distribution of (Z, Z^*, T) is not *directly* identified from the data. But using again the law of iterated expectation, these equations are equivalent to:

$$\mathbb{E}\left(\frac{1}{p(Z^*)}|T=1,Z\right) = \frac{1}{\mathbb{E}(T|Z)} \tag{3.3}$$

$$\mathbb{E}\left(\frac{m(Z^*)}{p(Z^*)}|T=1,Z\right) = \frac{\mathbb{E}(Y|Z)}{\mathbb{E}(T|Z)} \tag{3.4}$$

Under Assumption 3, the right-hand sides of these equations are identified because the distribution of (Y, T, Z) is identified from the data. Moreover, for given p and m the left-hand sides are also identified because the distribution of $(Z^*, Z)|T = 1$ is identified from the auxiliary sample. Assumptions 4 and 5 then ensure that $1/p(z^*)$ and $m(z^*)/p(z^*)$ are identified and then, $m(z^*)$, $p(z^*)$, and finally θ are identified.
The first moment condition (3.1) is close to the one used by D'Haultfœuille (2010) in a different framework, namely in a sample selection model. Applying Theorem 2.3 of D'Haultfœuille (2010) would actually yield the identification of $p(Z^*)$. On the contrary, the second moment condition (3.2) differs from D'Haultfœuille (2010) and the proof of identification of θ and of the full distribution (Y, T, Z, Z^*) is more involved in this case.

Theorem 1 also allows the econometrician to adapt the usual tests of the RD assumptions, when there is continuous measurement error in the running variable. McCrary (2008) proposes to test for the presence of discontinuity in the density of the running variable at the threshold. Theorem 1 actually implies that the density of the true running variable is identified. The intuition is as follows. Recall that $\mathbb{P}(T = 1|Z^* = z^*) = p(z^*)$ is identified by the moment condition (3.1) and that $f_{T=1}(z^*)$ and $\mathbb{P}(T = 1)$ are observed. Then, the Bayes formula, together with the support condition, yields the identification of $f_{Z^*}(z^*)$:

$$f_{Z^*}(z^*) = \frac{\mathbb{P}(T=1)}{\mathbb{P}(T=1|Z^*=z^*)} f_{Z^*|T=1}(z^*). \tag{3.5}$$

The full adaptation of the McCrary test involves taking into account the uncertainty associated with the estimation of the conditional take-up rate, which is out of the scope of this paper.

4 Nonparametric estimation

We now propose an estimation strategy of the LATE, based on our identification result. Our identification strategy relies on conditional moments being equal to zero. Hence, following Ai & Chen (2003), Newey & Powell (2003), Blundell et al. (2007), Chen (2007), and Chen & Pouzo (2012), we adopt a sieve estimator. First, we prove the consistency of our estimator. Second, we perform Monte-Carlo simulations to illustrate its finite sample performance.

4.1 Consistency

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

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

In this section, we write the LATE as a function of both m_0 and m_1, whereas in the previous section it is written as a function of $m(z^*) = \mathbb{E}(Y|Z^* = z^*)$ only. Accordingly, we adapt the moment conditions (3.3) and (3.4) to identify the three functions 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 moment condition (4.1) is the same as the moment condition (3.1). The moment conditions (4.2) and (4.3) are obtained following the same lines as in the proof of Theorem 1.[11] Note that rewriting the moment conditions into three components shows that m_1 can be estimated separately from p and m_0. This is particularly useful when the auxiliary data also include the outcome variable: $F_{Y,Z,Z^*|T=1}$ is identified. Then m_1 can be estimated by local linear regression of Y on Z^* on the sample of treated individuals, while the conditions (4.1) and (4.2) are sufficient to estimate p and m_0.

[11] Indeed, for any function $g(Y, T)$ we have: $\mathbb{E}\left(\frac{T\mathbb{E}(g(Y,T)|Z^*)}{p(Z^*)}|Z\right) = \mathbb{E}(g(Y, T)|Z)$. The moment condition (4.2) corresponds to $g(Y, T) = Y(1 - T)$ and the moment condition (4.3) corresponds to $g(Y, T) = YT$.

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 estimator, see Chen (2007), Section 2.2.4).

First, we define series estimators of the conditional moments. Let us denote S and S_a the main and auxiliary samples of Assumption 3 with respective sizes n and n_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 $l^p(n,n_a)$ be the dimension of $\mathcal{I}^p_{n,n_a}$. 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 $\text{span}(B^p) = \mathcal{I}^p_{n,n_a}$. 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 choose $\widehat{\mathbb{E}}(B^{j'}(Z)B^j(Z))$ as the empirical mean in the main sample. We can define two functions $q_j(z^*)$ and $r_j(y,t)$ such that $\rho_j(W) = q_j(Z^*)T + r_j(Y,T)$. 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:

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

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 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 $dim(\mathcal{I}^1_{n,n_a}) \geq dim(\mathcal{H}^1_{n,n_a})$.

The convergence of $(\widehat{p},\widehat{m_0},\widehat{m_1})$ depends 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. In the following assumptions, $f(z) \sim_{z\downarrow 0} g(z)$ means that $f(z) = g(z) \times (1 + o(1))$ where $o(1)$ is a function tending to 0 when z decreases to 0.

Assumption 6 (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 continuously differentiable on $[-1; 0[$ and on $]0; 1]$ and their first derivatives are bounded by a known constant C,*

4. *The random variable Z admits a density f_Z with a compact support $[l; u]$, such that f_Z is bounded below on any compact included in $]l; u[$, and such that it exists $C_u, C_l > 0$ and $\alpha_u, \alpha_l \geq 0$ such that:*

$$f_Z(u - z) \sim_{z\downarrow 0} C_u z^{\alpha_u} \text{ and } f_Z(l + z) \sim_{z\downarrow 0} C_l z^{\alpha_l},$$

5. *The conditional distribution of $Z^*|Z$ admits a density $f_{Z^*|Z}$ with respect to the Lebesgue measure and it exists $K > 0$, $\kappa \geq 0$ such that for any $\delta > 0$ and any $z, z' \in]l+\delta; u-\delta[$*

$$\int |f_{Z^*|Z=z}(z^*) - f_{Z^*|Z=z'}(z^*)|dz^* \leq K|z - z'|\delta^{-\kappa}.$$

Assumption 6.1 essentially means that Z^* has a compact support, which comprises the discontinuity threshold. The choice of $[-1; 1]$ is a normalization that can be assumed without loss of generality.

Assumption 6.2, which implies a reinforcement of the support condition (Assumption 4), and Assumption 6.3 define the space $\mathcal{H}$ containing $\xi = (p, m_0, m_1)$.[12] Assumptions about bounds for (p, m_0, m_1) and their derivatives ensure that (p, m_0, m_1) belongs to a compact.[13] Such a compactness assumption has two advantages. First, (p, m_0, m_1) is a well-separated minimum of Q. This kind of property is necessary even in parametric framework (see for instance Chapter 5.2 and Problem 5.27 of van der Vaart, 2000). Second, the compactness of

[12] Under Assumptions 6.2 and 6.3, the consistency of our estimator holds under the bounded completeness condition, which is a weaker version of the completeness condition (Hu & Schennach, 2008, D'Haultfœuille, 2011). The bounded completeness condition holds if the implication in Assumption 5 holds for any bounded function g(.).

[13] Compactness is defined relative to the topology of the uniform convergence (see the Arzela-Ascoli Theorem).

$\mathcal{H}$ ensures that our estimation strategy is a well-posed inverse problem.[14] In nonparametric frameworks, such a compactness assumption has been used by Ai & Chen, 2003 and Newey & Powell, 2003 (cf. in particular Section 3 for more detailed explanations). Relaxing the compactness assumption is possible using high level conditions as in Chen (2007), and/or using penalization, or regularization, as suggested by Carrasco et al. (2007) or by Chen & Pouzo (2012). Such an extension is out of the scope of this paper.
Assumption 6.4 is necessary to control the rate of uniform convergence of Q_{n,n_a} to Q. Burman & Chen (1989), Newey (1997), 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 that the density f_Z converges to zero at the boundary of its support. 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, f_Z tends towards 0 at the bounds of its support. However, classical measurement error verifies Assumption 6.4 with $\alpha_u = \alpha_l = 1$. More generally, Lemma A.1 in the Appendix gives sufficient conditions to ensure simultaneously Assumptions 5, 6.4 and 6.5 when $Z = \mu(\nu(Z^*) + \varepsilon)$ with μ and ν two increasing functions and $\varepsilon \perp\!\!\!\perp Z^*$.
Last, for the choice of $\mathcal{I}^p_{n,n_a}$ presented below, Assumptions 6.4 and 6.5 altogether ensure that $\mathbb{E}(\rho_p(W;p)|Z = .)$ could be approximated uniformly in $p \in \mathcal{H}^p$ by an element of $\mathcal{I}^p_{n,n_a}$, and that similar approximations hold for $\mathbb{E}(\rho_0(W)|Z = .)$ and $\mathbb{E}(\rho_1(W)|Z = .)$.

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 $(1 = z^{0+}_{k^{0+}_{n,n_a}+1} > z^{0+}_{k^{0+}_{n,n_a}} > ... > z^{0+}_1 > z^{0+}_0 = 0 = z^{0-}_0 > ... > z^{0-}_{k^{0-}_{n,n_a}} > z^{0-}_{k^{0-}_{n,n_a}+1} = -1)$ verifying $\frac{\underline{\delta}^0}{k^{0\pm}_{n,n_a}} \leq |z^{0\pm}_j - z^{0\pm}_{j-1}| \leq \frac{\overline{\delta}^0}{k^{0\pm}_{n,n_a}}$, such that:

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

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 6.1, 6.2 and 6.3. 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

[14]Under compactness of $\mathcal{H}$, $(\widetilde{p}, \widetilde{m}_0, \widetilde{m}_1) \in \mathcal{H} \mapsto (\mathbb{E}(\rho_p(W;\widetilde{p})|Z), \mathbb{E}(\rho_0(W;\widetilde{p},\widetilde{m}_0)|Z), \mathbb{E}(\rho_1(W;\widetilde{p},\widetilde{m}_1)|Z))$ is an homeomorphism and then admits a continuous inverse (see for instance: Theorem 26.6 in Munkres, 2000).

condition $\sup_{z^*}|f(z^*)| < c$ is replaced by $\underline{c} \leq f \leq 1$.
Last, for $\mathcal{I}^p_{n,n_a}$ (and similarly for $\mathcal{I}^0_{n,n_a}$, $\mathcal{I}^1_{n,n_a}$), we consider linear splines with $l_{n,n_a} - 1$ interior and approximatively equidistant knots on $[l;u]$. So it exist $\overline{\delta}$ and $\underline{\delta}$ (independent of n, n_a) and knots ($l = z_0 < z_1 < ... < z_{l_{n,n_a}} = u$) verifying $\frac{\underline{\delta}}{l_{n,n_a}-1} \leq |z_j - z_{j-1}| \leq \frac{\overline{\delta}}{l_{n,n_a}-1}$, such that:

$$\mathcal{I}^p_{n,n_a} = \left\{ \begin{array}{ll} f: & \exists (a_j)_{j=0,1,\dots,l_{n,n_a}-1} \text{ such that} \\ & f(z) = a_0 + \sum_{j=1}^{l_{n,n_a}-1} a_j(z-z_j)\mathbb{1}_{\{z-z_j>0\}} \end{array} \right\}.$$

The following Theorem ensures consistency of our estimator.

Theorem 2 (Consistency)
Under Assumptions 1, 2, 3, 4, 5 and 6, 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* $\min_{j=p,0,1}(k^{j+}_{n,n_a}, k^{j-}_{n,n_a}) \to \infty$ *with* $\max_{j=p,0,1}(\dim(\mathcal{I}^j_{n,n_a})) = o(n^{1/(2+\max(\alpha_u,\alpha_l))})$.

The proof of Theorem 2 is reported in the Appendix. In Theorem 2, 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 $\mathcal{I}^j_{n,n_a}$ (for all $j = p, 0, 1$) tends to infinity sufficiently slowly. This last condition also constrains the increase in the dimension of $\mathcal{H}_{n,n_a}$, as $\dim(\mathcal{I}^j_{n,n_a}) \geq \dim(\mathcal{H}^j_{n,n_a}) = k^{j+}_{n,n_a} + k^{j-}_{n,n_a} + 2$. This means that the number of knots needs to tend to infinity only at a $o(n^{1/(2+\max(\alpha_u,\alpha_l))})$ rate to ensure consistency of $\widehat{\theta}$. This is a lower rate than $o(\sqrt{n})$, the rate obtained under the assumption that the density of Z is bounded below on its support (in that case $\alpha_l = \alpha_u = 0$). Here, the higher the rate of decreasing of f_Z at the boundaries of its support, the more we have to smooth the estimation of $p(.), m_0(.)$ and $m_1(.)$ to ensure the consistency of $\widehat{\theta}$.

Deriving the asymptotic behavior of $\widehat{\theta}$ and the corresponding inference properties is left for future research. In the application in Section 5, we use percentile bootstrap for inference. Chen & Pouzo (2015) prove that bootstrapping is valid in sieve semi-parametric IV estimation in a framework that slightly differs from ours. Adapting their result to our framework is also left for future research.

Contrary to other sieve estimator in nonparametric IV frameworks, our parameter of interest is local and we cannot expect its rate of convergence to be $\sqrt{n}$. Fundamentally, in a RD design, even without measurement errors, the rate of convergence of the estimators of the LATE depends on the local smoothness of $z^* \mapsto \mathbb{E}(Y|Z^* = z^*)$ and $z^* \mapsto \mathbb{E}(T|Z^* = z^*)$, around the threshold. We thus provide, in the next section, Monte-Carlo simulations to show the finite sample properties of our 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 distributed on $[-1;1]$ and that $\mathbb{P}(T=1|Z^*=z^*) = 1/8 + 1/4\Phi(5z^*) + 1/2\cdot\mathbb{1}\{z^*\geq 0\}$, where 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:

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

where $\mathbb{1}\{CO\}$ and $\mathbb{1}\{AT,NT\}$ are dummies for the types of individuals (compliers and always- or never-takers) with $CO, AT, NT \perp\!\!\!\perp Z^*$ and $(v_0,v_1)|Z^*,CO,AT,NT \sim \mathcal{N}(0,\Sigma)$ with $\Sigma_{11}=\Sigma_{22}=\frac{1}{16}$ and $\Sigma_{12}=\frac{1}{32}$. The true LATE $\theta_0 = \mathbb{E}(Y(1)-Y(0)|Z^*=0,CO)$ is then equal to $1-4=-3$.
The noisy running variable is drawn from the following multiplicative process:

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

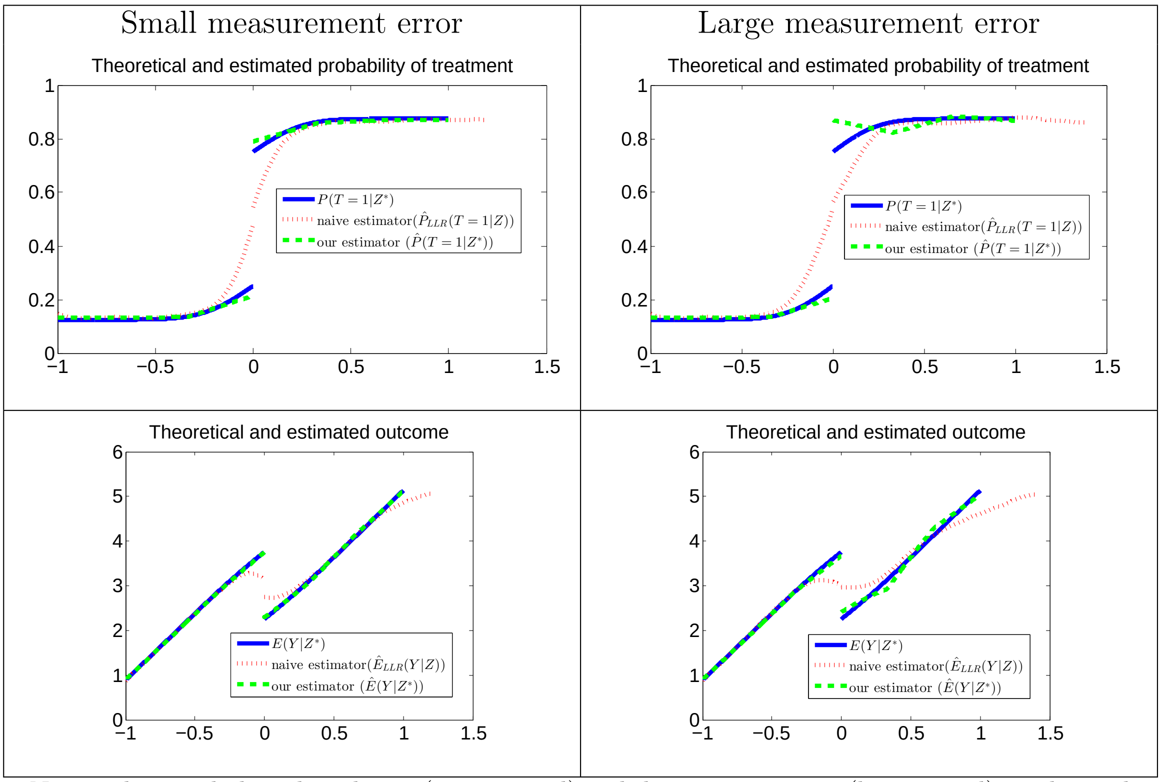
with ε uniformly distributed and independent of $(Z^*,T,Y(0),Y(1))$. To investigate the impact of the size of the measurement error, we let the dispersion of ε vary. Below we consider two cases: (i) small measurement error with $\varepsilon\sim\mathcal{U}_{[-0.1;0.1]}$, and (ii) large measurement error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

Following the usual practice in RD-design, we first give a graphical illustration plotting estimated take-up $\widehat{p}(z^*)$ and estimated mean outcome $\widehat{m}(z^*) = \widehat{m}_1(z^*)\widehat{p}(z^*)+\widehat{m}_0(z^*)(1-\widehat{p}(z^*))$. Figure 1 plots the take-up (upper panel) and the mean outcome (lower panel) conditional on the true running variable, obtained with both our estimation and the naive estimation ignoring measurement error, for two samples of 25,000 observations. The left panel corresponds to a sample with a small measurement error (case (i)), while the right panel corresponds to a sample with a large measurement error (case (ii)). Here, the naive estimates correspond to a local linear regression with a gaussian kernel on each side of the threshold. Figure 1 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.

This first graphical illustration is not enough to investigate finite sample properties on $\widehat{\theta}$, which is our main parameter of interest. For various sample sizes (1000, 5000 and 25000), we simulate 1000 samples according to the DGPs presented previously. For each sample, we estimate the LATE θ with various methods. We consider two types of naive estimators ignoring measurement error. The first naive estimator, denoted IK hereafter, is based on local linear regressions, with an MSE-optimal choice of bandwidth proposed by Imbens & Kalyanaraman (2012). The second one, denoted CCT hereafter, is a bias-corrected estimator proposed by Calonico et al. (2014*b*). Both naive estimations are performed using the Stata instruction `rdrobust` - provided by Calonico et al. (2014*a*) - with default options (in particular it uses Epanechnikov kernels). Concerning our sieve estimation derived in Section 4, 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 k 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 corresponding to $k+2$. We investigate how the finite sample properties of our estimators vary when the number of knots k vary. When the number of knots is null, the estimated functions $\widehat{p}$, $\widehat{m}_0$ and $\widehat{m}_1$ 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. Last, concerning the bounds $\underline{c}$, c, and C of functions in $\mathcal{H}_{n,n_a}$, we choose 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.

Table 1 reports the average bias, variance, median and interquartile range of the LATE estimates. Moreover for the naive estimates (IK and CCT), we report the average size of the data-driven bandwidths (denoted h_n for the bandwidth used in IK estimation and b_n for the supplementary bandwidth used to estimate asymptotic bias in CCT estimation).
Overall, the magnitude of the average bias of our estimators is below one, whatever the size of the measurement error, the sample size or the number of knots. It outperforms both naive estimators which are almost always over one in Table 1. The variance of our estimator is always lower than the one of the naive estimators. Consistently with Proposition 2, the mean bias and empirical variance of the IK estimator are quite erratic across

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: Estimation of the LATE in finite samples, Multiplicative Error

		A. Small Error, $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$			B. Large Error, $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$		
		Sample Size			Sample Size		
Estimator	Stat.	1000	5000	25000	1000	5000	25000
	Bias	-0.969	0.344	0.678	-0.009	0.681	0.806
Our ($k = 0$)	Variance	10.63	1.009	0.024	2.818	0.127	0.018
	Median	-2.555	-2.381	-2.331	-2.410	-2.270	-2.191
	IQ range	2.318	0.2874	0.149	0.794	0.246	0.165
	Bias	-0.530	0.351	0.512	-0.188	0.349	0.512
Our ($k = 1$)	Variance	24.98	0.328	0.078	31.85	0.423	0.110
	Median	-2.740	-2.493	-2.443	-2.663	-2.482	-2.443
	IQ range	1.343	0.538	0.328	1.344	0.767	0.460
	Bias	-0.738	0.074	0.172	0.249	0.142	0.174
Our ($k = 2$)	Variance	325.0	0.578	0.097	943.9	2.639	0.573
	Median	-2.948	-2.756	-2.796	-2.426	-2.766	-2.778
	IQ range	1.833	0.829	0.417	3.319	1.705	1.094
	Bias	1.396	1.018	2.058	0.226	5.433	1.495
	Variance	540.9	2083	1702	317.6	5975	284.1
Naive (IK)	Median	-2.477	-2.431	-1.811	-1.799	-1.629	-1.607
	IQ range	1.964	2.574	2.865	3.056	3.104	3.004
	Bdw. h_n	0.276	0.179	0.115	0.292	0.209	0.156
	Bias	-71.33	-588.3	-15.35	0.599	738.3	-84.96
	Variance	1×10^6	2×10^8	7×10^7	3×10^5	5×10^8	5×10^6
Naive (CCT)	Median	-2.501	-2.528	-2.174	-1.880	-1.887	-1.959
	IQ range	2.438	2.681	2.556	2.931	3.079	2.966
	Bdw. b_n	0.508	0.423	0.352	0.535	0.449	0.363

Note: Computation obtained with 1000 simulations. The same set of simulations is used in each column. For our estimators k is the common number of knots selected to defined our linear splines. When $k = 0$, functions p, m_0 and m_1 are approximated by linear functions on $[-1; 0]$ and $[0; 1]$. When $k = 1$ (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$). For the Naive estimators, (IK) is the estimator proposed by Hahn et al. (2001) using the bandwidth h_n proposed by Imbens & Kalyanaraman (2012) and (CCT) is the bias-corrected estimator proposed by Calonico et al. (2014*b*) with bandwidth b_n used to estimate asymptotic bias of (IK).

the Table. The variance and the interquartile range of the IK estimates do not tend to 0 with the sample size illustrating the lake of consistency of the IK estimator. According to Proposition 2, the IK estimator is a Cauchy for large samples. Given our choice of DGPs, we compute its theoretical median (-1.85 for the small error and -1.64 for the large one) and interquartile range (2.93 for the small error and 3.06 for the large one).[15] We verify that the empirical median and its interquartile range in Table 1 are close to these limits. Although we have no theoretical results on the limit distribution of the CCT estimator, Table 1 confirms that it also behaves quite erratically. This is not surprising as the CCT estimator corrects the IK estimates with an estimation of its asymptotic bias, which is not well-defined in our case according to Proposition 2.

Table 1 informs us about the influence of the number of knots on our sieve estimator. For sufficiently large sample size (5000 for the large error and 25000 for the small one), when the number of knots k increases, the bias decreases and the variance increases, accordingly to the usual influence of smoothing parameters on the trade-off between bias and variance.

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 m_1 by a local linear regression. The results are very close to those of our main sieve estimator (see Table 3 in the Appendix). Second, we compare our main estimator to the Donut estimator sometimes used in contexts with measurement error (see Barreca, Guldi, Lindo & Waddell, 2011 or Dong, 2014). The Donut estimator corresponds to a 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 estimators (see the last rows of Table 3 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 estimators when the sample size is larger than 5,000. The influence of sample size on the interquartile range is qualitatively similar to the case with multiplicative error, but sample sizes have to be larger to get reliable estimates (see Tables 4 and 5 in the Appendix).

[15]These limits have been estimated using formula given in Proposition 2 and simulations based on 10^7 drawn. For these simulations, we first draw Z^* in the distribution of $Z^*|Z = 0$ which is a truncated exponential distribution for the multiplicative error. Next, we draw (Y, T) in the distribution of $(Y, T)|Z^*$.

5 Application

In this section, we use our method to evaluate of the effects of more generous unemployment insurance (UI) on across-industries mobility. The Regression Discontinuity design has been used by a significant number of papers to estimate the effect of UI (for example, Card et al., 2007; Lalive, 2008). All these papers estimate the effect of an increase in potential benefit duration for UI claimants. For example, Card et al. (2007) estimate the effect of increasing the potential benefit duration from 20 to 30 weeks, using as a running variable the past tenure of claimants. Our method enables us to go further and to estimate the effect of UI at the extensive margin, i.e. the effect of having some UI benefits as opposed to having no benefits at all. Usual RD designs that rely on UI data only, cannot answer this question. Indeed, assessing the extensive margin requires not only data on unemployed people eligible for benefits, but also data on unemployment people who do not qualify for benefits. We thus need a broader data source, such as Social Security (SS) data. In this data source, the running variables of the RDD (for example, past tenure) may be observed with measurement error. For instance, it may be difficult to compute the UI-agency definition of tenure in the SS data. Our approach addresses this measurement error issue and we apply it to estimate the effect of UI in the French context. We first describe the institutional context and data, and we then discuss our estimation results.

5.1 Institutions and data

In France, between 2003 and 2006, the unemployed were eligible for UI if they had worked at least 182 days or 910 hours in the two years before becoming unemployed. When eligible, the unemployed could draw benefits during 7 months.[16] This UI eligibility rule naturally generates a Regression Discontinuity design, as in Card et al. (2007). We can define two running variables: Z^* the past tenure in the last firm before becoming unemployed and S^* the number of hours worked. As we do not observe S^* in our data, the design is a priori fuzzy with individuals treated on both sides of the 182 days cutoff. In addition, our design has another source of fuzziness, as Z^* only measures the tenure in the last job, while the UI rules enable claimants to add up tenure in different firms to obtain eligibility. We verify below that the design is indeed a two-sided fuzzy design with treated individuals at any value of the running variable Z^*.

We match the French employment registers, i.e. *Déclarations Annuelles de Données Sociales (DADS)*, with the administrative records of the Unemployment Insurance agency,

[16]The potential benefit duration increased to 23 months when the unemployed had worked at least 14 months before becoming unemployed.

i.e. *Fichier Historique (FH)*, at the individual level.[17] From the DADS, we compute for any worker separating from her firm, the past tenure in her last firm (Z) and the outcome of interest (Y) defined below. In the UI records, we observe for every claimant the past employment duration in the last job (Z^*) used by the UI agency to compute UI entitlement.[18] Matching the two registers, we define $T = 1$ if the individual appears in the UI data and starts a claim after her job separation. As a result, we observe the joint distribution of $(Y, Z, T, T \times Z^*)$.

We now describe the DADS in greater detail and highlight some technical reasons for the measurement error in employment duration. The DADS records, for every firm worker year pair, the annual wage and the first and last dates of employment in the firm within the calendar year. The DADS also records the number of hours worked, but the information is missing for a large fraction of the records. The main use of the DADS is the computation of a specific payroll tax -*Contribution Sociale Généralisée* (CSG)-, which is not affected by the actual job duration. Consequently the administration does not make specific quality controls on employment dates and employers have no monetary incentives associated with the declaration of spells. We can thus expect that the employment dates provided are often inaccurate and that the employment duration Z computed from them is measured with error. There are also several technical reasons why the variable Z differs from the true running variable, as defined by the UI agency. First, the DADS only contains the first and last dates of employment in the firm within the calendar year. The difference between these two dates is larger than the true past tenure, if individuals have several non-consecutive spells in the same firm. Second, DADS dates are coded on a scale from 1 to 360, with every month made up of 30 days. This recording type tends to underestimate the true tenure. Third, errors in firm/worker identifiers make it difficult to follow individuals over time, and lead us to underestimate past tenure.

From the DADS, we also compute our outcome (Y), which relates to the mobility across industries of the unemployed. It is equal to one, if the job-seeker finds a new job in the same sector as her job before separation, and 0 otherwise.[19] This is an important outcome, because when cash-constrained individuals lose their jobs, UI may help them to wait for high-paying jobs, making use of their sector-specific skills (Mortensen, 1977). Matching job-seekers with jobs in the same sector as their previous jobs would probably lead to less skill-mismatch, and provide a rationale for UI.

Our main sample comprises 382,037 workers who separate from their firms between 2003

[17]More precisely, we use a random subsample, comprising all individuals born in October of even years.
[18]We do not observe the number of hours worked in the data that the UI agency shared with us.
[19]Note that we observe job findings up to December 2004. Employers' industries are split into 36 groups.

and 2004. Out of these workers, 11.5% claim UI benefits. While this fraction may seem to be low at first sight, one should keep in mind that the data do not allow us to concentrate on lay offs. Many workers actually quit their job for a job-to-job transition and are therefore not eligible for UI. Moreover, even eligible workers do not necessarily claim UI benefits, for example if they fear the stigma associated with social insurance. Usual estimates of the UI take-up rate in the U.S. range from 30% to 50% (Anderson and Meyer, 1997).

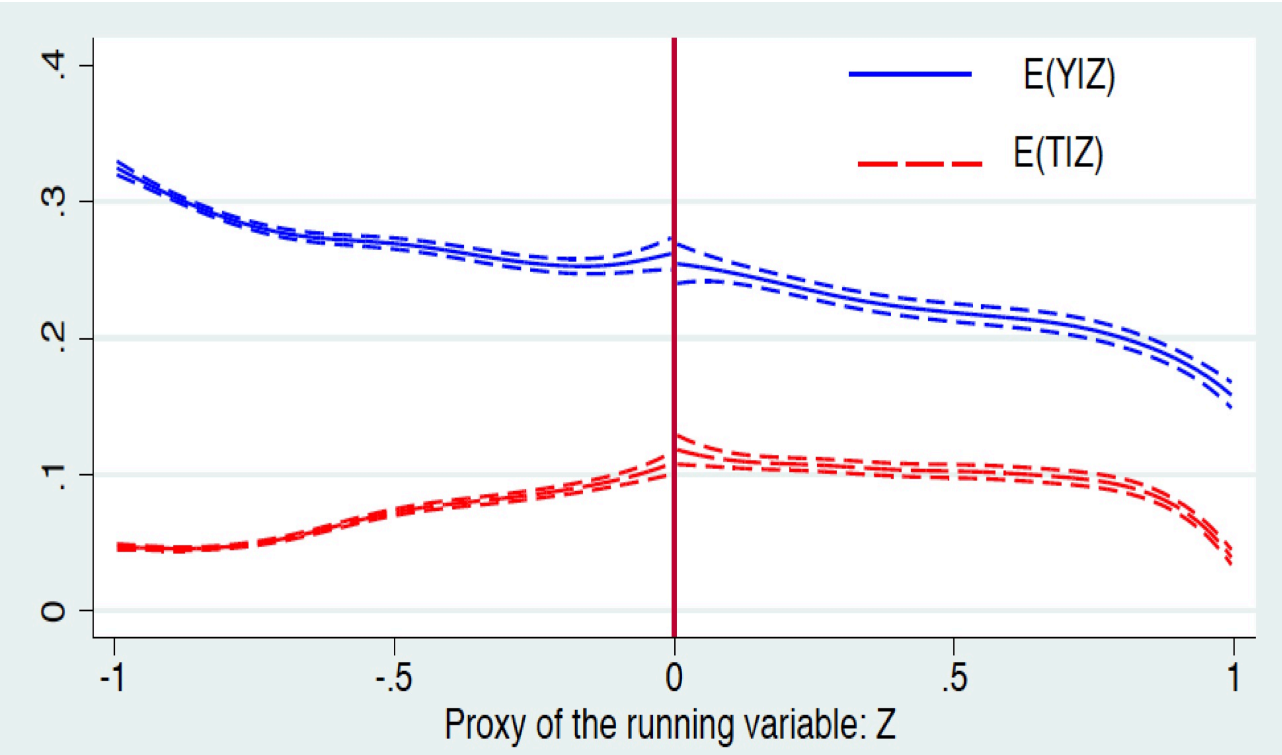
Around 20% of workers separating from their firms between 2003 and 2004 actually find a job in the same sector by the end of December 2004. This low fraction does not necessary mean that labor mobility across sectors is very high in France, as some workers do not find any jobs at all before our data end. For these workers, we set $Y = 0$. An alternative strategy would be to focus on workers who find a job before December 2004. However, we prefer to analyze unconditional across-sector mobility, as mobility conditional on employment is actually conditional on an endogenous outcome.

Figure 2 plots the take-up rate and the probability to find a job in the same sector, as functions of the past tenure measured in the DADS (Z). Past tenure is centered on 182 days and divided by 182 days, so that the cutoff is equal to 0. Workers with one day of work experience have Z close to -1 and workers who spend one year in their previous firm have Z close to 1. We adopt this normalization in the remainder of the Section. In Figure 2, we plot the estimates of $\mathbb{E}(T|Z)$ and $\mathbb{E}(Y|Z)$ using local linear regression with a Gaussian kernel of bandwidth 0.1. The take-up rate slightly increases from 5% for workers with very short past tenure to around 10% at the cutoff value. There is no significant discontinuity at the cutoff value in the take-up rate.[20] Consequently, a standard RD analysis fails to identify any treatment effects. Similarly, there is no significant discontinuity in the mean outcome at the cutoff value.

Before turning to the results obtained with our alternative method, we report in Figure 3 the density of the true running variable (Z^*) on the treated sample and the densities of the proxy (Z) on both the treated and the non-treated sample. As the density of the true running variable of claimants is strictly positive over its whole support, Assumption 4 is verified (two-sided fuzzy RD design). The density of the proxy of the running variable for UI claimants is shifted to the left of the density of the true running variable, highlighting the extent of measurement error. The characteristics of the measurement error are investigated further in the next section. The density of the mismeasured running variable on non-claimants is further shifted to the left, reflecting selection into treatment. The unemployed with lower tenure are less likely to be eligible to UI and claim benefits.

[20] The robust tests developed by Calonico et al. (2014*b*) do not reject the continuity of the take-up and zero treatment effect at the 5% level.

Figure 2: Policy effects retrieved from (Y,T,Z)



Source: DADS (Insee). Notes: This Figure plots the take-up rate ($E(T|Z)$) and the probability to find a job in the same sector ($E(Y|Z)$), as functions of the past tenure measured in the DADS (Z). Past tenure is centered on 182 days and divided by 182 days. Consequently the cutoff value of UI eligibility is at 0. Estimates are obtained using local linear regressions with a Gaussian kernel of bandwidth 0.1. Dashed lines represent 95% confidence intervals.

Figure 3: Observed densities of the true running variable on claimants and of its proxy on both claimants and non-claimants

Source: DADS (Insee). Note: kernel estimation (Epanechnikov) with bandwidth 0.0686

5.2 Results

Figure 4 plots our main estimation results: the take-up rate and the mean outcome, as functions of the true running variable. Estimation is carried out along the lines of Section 4. The dimension of the sieve spaces is 3 on each side of the cutoff and we choose exactly the same dimensions for our instruments. We use equidistant linear splines. Figure 4 also plots 90% confidence intervals, obtained by bootstrap.

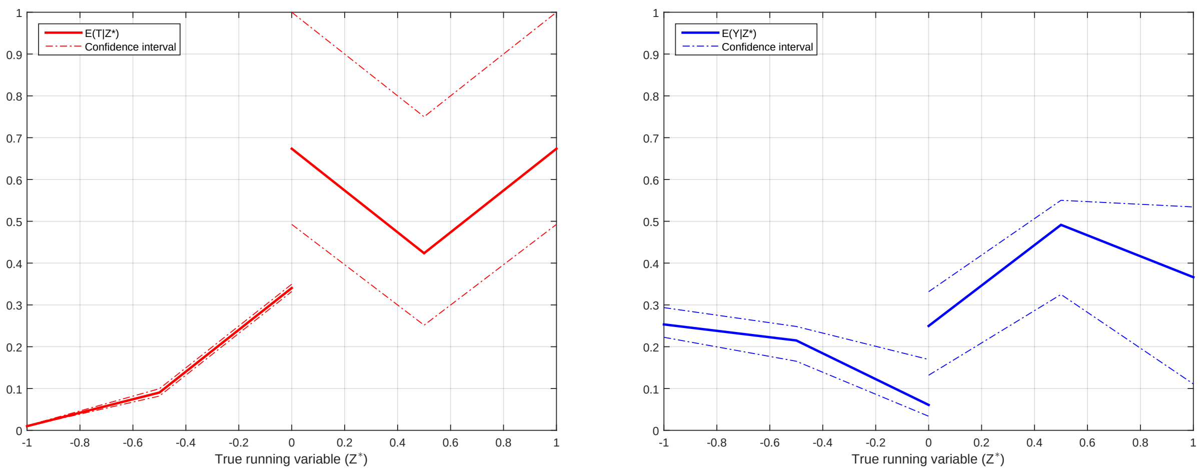
The left panel of Figure 4 clearly shows that there is a significant jump in the take-up rate at the cutoff value. The fraction of claimants increases by 33 percentage points, from 34% to 67%. The right panel shows that the fraction of workers finding jobs in the same sector increases from 6% to 25% at the cutoff value. This is a large increase, although it is not statistically significant at the 10% level. The resulting LATE estimate is 56.7 percentage points. The lower bound of the confidence interval is 8.3, so that the LATE is significantly different from 0. In sum, compliers who are induced to claim unemployment benefits because their last job crosses the 182 days threshold and are entitled to 7 months of benefits, are more likely to find a new job in their past industry. This result illustrates the strength of our new method, as it would not have been obtained with a standard RD design ignoring measurement error. Thus, unemployment insurance enables job-seekers to find jobs in industries where they have already accumulated some specific human capital. In other words, UI seems to limit downgrading effects associated with unemployment shocks. However, low mobility across sector may also be inefficient, especially when the economy is hit by permanent asymmetric shocks across industries and labor should be reallocated across industries. Analyzing the efficiency consequences of this treatment effect is left for future research.

To apply our identification strategy, we make a few assumptions. Some of these have testable implications. For example, the non-differentiality of the measurement error implies that there is no supplementary information in the noisy running variable about potential outcomes, once we condition on the true running variable, especially for the *treated sample.* Indeed, Assumption 2 implies that $Y \perp\!\!\!\perp Z | Z^*, T = 1$. We thus regress Y on Z and Z^* in the treated sample, where we observe all the variables, and test if the coefficient of Z is equal to zero. Table 2 reports the estimation results and shows that we cannot reject the non-differentiality of the measurement error at the 1% level in the treated sample.

Finally, one crucial assumption of the RD design is the absence of sorting around the cutoff. McCrary (2008) develops a test of the manipulation of the running variable, based on the absence of discontinuity in the density of the running variable. Actually, our methodology

Figure 4: Policy effects retrieved from (Y, T, Z) and $(Z, Z^*)|T = 1$

Take-up: $\mathbb{P}(T = 1|Z^*)$ Average outcome: $\mathbb{E}(Y|Z^*)$



Data: DADS (Insee). Notes: This Figure plots the estimated treatment probability in the left panel and mean outcome in the right panel, as a function of the true running variable (Z^*). The cutoff value is 0. The dimension of each sieve subspace is 3. The functional basis is compounded of linear splines with equidistant knots. 90% confidence interval are computed by bootstrap (100 replications) and plotted in dashed lines.

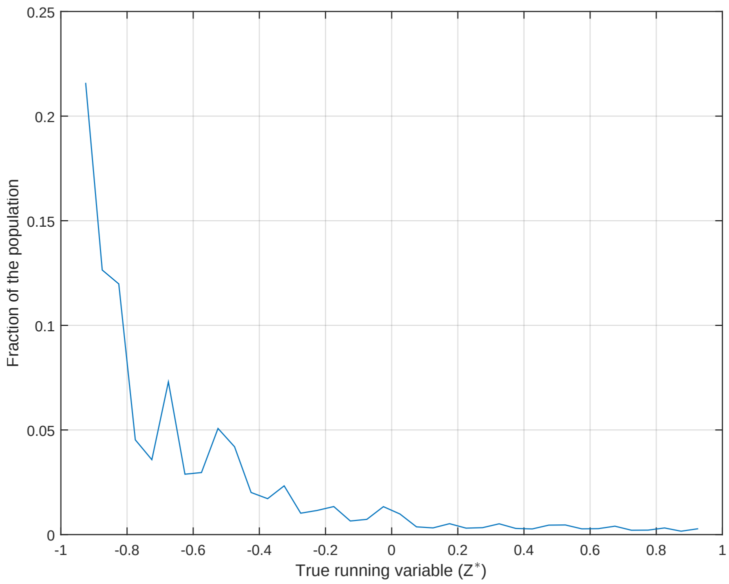
Table 2: Test of non-differential measurement error on the treated sample

	Y
Z	-0.00555
	(0.00625)
Z^*	-0.0346***
	(0.00658)
Constant	0.206***
	(0.00363)
Observations	15,931
R-squared	0.002

Note: This Table presents the estimation results of the regression of Y on Z and Z^* on the treated sample. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

identifies the density of the true running variable. Figure 5 plots the estimated density which does not feature bunching around the cutoff.

Figure 5: Estimated density of the true running variable on both claimants and non claimants



Source: DADS (Insee). Notes: This Figure plots the density of the true running variable (Z^*). The cutoff value is 0. The bin width is 0.05. The Y-axis is the estimated fraction of the population in each bin.

6 Conclusion

Continuous measurement errors in the running variable have 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/institutions in charge of delivering the treatment usually keep records of the correct running variable for the treated individuals. Under the assumption of non-differential measurement error and under a large support condition, 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 outcomes. We then proposed a sieve estimator of the LATE, showed its consistency and investigated its performance in finite samples. We illustrated the usefulness of our method by applying it to the estimation of the effect of receiving unemployment benefits.

A Appendix: Proofs

A.1 Proof of the inconsistency of the usual RD estimators

Proposition 2 *Let us assume that Assumption 2 holds and that $Z|Z^*$ admits a twice continuously differentiable density with respect to the Lebesgue measure (continuous measurement error) such that $\mathbb{E}\left(sup_z f^{(j)}_{Z|Z^*}(z)\right) < \infty$ for $j = 1, 2$. Moreover let us assume that it exists $\delta > 2$ such that $\mathbb{E}\left(|Y|^{\delta}\right) < \infty$ and that $z^* \mapsto \mathbb{E}(Y^2|Z^* = z^*)$ is bounded. For a sample of n iid observations, for a decreasing sequence $h_n = O(n^{-1/5})$ and for any K bounded, symmetric and nonnegative valued kernel with compact support, let the naive adaption of the estimator of the Wald ratio based on local linear regression defined by:*

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

Then

$$\widehat{\theta}^{naive}_{LLR} \xrightarrow[n\to\infty]{law} \mathcal{C}$$

where $\mathcal{C}$ is Cauchy of location $\frac{\mathbb{C}ov(Y,T|Z=0)}{\mathbb{V}(T|Z=0)}$ and scale $\left(\frac{\mathbb{V}(Y|Z=0)}{\mathbb{V}(T|Z=0)} - \frac{\mathbb{C}ov^2(Y,T|Z=0)}{\mathbb{V}^2(T|Z=0)}\right)^{1/2}$.

Let us introduce some notations:

$$\begin{aligned}
B^+ &= \frac{1}{2}\frac{\left(\int_0^{+\infty} u^2K(u)du\right)^2 - \left(\int_0^{+\infty} u^3K(u)du\right)\left(\int_0^{+\infty} uK(u)du\right)}{\left(\int_0^{+\infty} u^2K(u)du\right)\left(\int_0^{+\infty} K(u)du\right) - \left(\int_0^{+\infty} uK(u)du\right)^2} \\
B^- &= \frac{1}{2}\frac{\left(\int_{-\infty}^{0} u^2K(u)du\right)^2 - \left(\int_{-\infty}^{0} u^3K(u)du\right)\left(\int_{-\infty}^{0} uK(u)du\right)}{2\left(\int_{-\infty}^{0} u^2K(u)du\right)\left(\int_{-\infty}^{0} K(u)du\right) - \left(\int_{-\infty}^{0} uK(u)du\right)^2} \\
V^+ &= \frac{\int_0^{+\infty}\left[\left(\int_0^{+\infty} s^2K(s)ds\right) - \left(\int_0^{+\infty} sK(s)ds\right)u\right]^2 K(u)^2du}{f_Z(0)\left[\left(\int_0^{+\infty} u^2K(u)du\right)\left(\int_0^{+\infty} K(u)du\right) - \left(\int_0^{+\infty} uK(u)du\right)^2\right]^2} \\
V^- &= \frac{\int_{-\infty}^{0}\left[\left(\int_{-\infty}^{0} s^2K(s)ds\right) - \left(\int_{-\infty}^{0} sK(s)ds\right)u\right]^2 K(u)^2du}{f_Z(0)\left[\left(\int_{-\infty}^{0} u^2K(u)du\right)\left(\int_{-\infty}^{0} K(u)du\right) - \left(\int_{-\infty}^{0} uK(u)du\right)^2\right]^2}
\end{aligned}$$

Under assumptions of Proposition 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.
The conditions on the kernel K ensure that Assumptions 4 of Hahn et al. (1999) holds.
Last, the condition $\mathbb{E}(|Y|^{\delta}|Z) < \infty$ of Proposition 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 use Lemma 1 to Lemma 7 of Hahn et al. (1999) when $\delta \geq 3$ and $h_n \sim n^{-1/5}$. Moreover, their reasoning, which is based on Lyapounov's central limit Theorem, also holds for $\delta \in]2; 3[$ and all their asymptotic approximations are valid for

$h_n = O(n^{-1/5})$. So, we obtain:

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

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

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

$$(nh_n)^{1/2}\begin{pmatrix} \widehat{a}_Y^+ - \widehat{a}_Y^- \\ \widehat{a}_T^+ - \widehat{a}_T^- \end{pmatrix} \to \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2V^+\begin{pmatrix} \mathbb{V}(Y|Z=0) & \mathbb{C}ov(Y,T|Z=0) \\ \mathbb{C}ov(Y,T|Z=0) & \mathbb{V}(T|Z=0) \end{pmatrix}\right)$$

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 $\left(\frac{\mathbb{V}(Y|Z=0)}{\mathbb{V}(T|Z=0)} - \frac{\mathbb{C}ov^2(Y,T|Z=0)}{\mathbb{V}^2(T|Z=0)}\right)^{1/2}$.

Finally we discuss how technical assumptions could be relaxed. The domination condition $\mathbb{E}\left(sup_z f^{(j)}_{Z|Z^*}(z)\right) < \infty$ for $j = 1, 2$ 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. This domination condition is made for technical convenience and can be replaced by alternative restrictions (see Davezies & Le Barbanchon, 2014, Proposition 2.2). 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. It is well known that such ratio do not have finite 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.

A.2 Proof of Theorem 1

We first prove that $\mathbb{P}(T=1|Z^*) = p(Z^*)$ is identified, following an argument of D'Haultfœuille (2010). Let us consider the derivation of the following moment condition:

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

Under Assumption 4, we can consider the following expectation:

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

We then apply the law of iterated expectation on the left-hand side of the previous equation and simplify its expression using Assumption 2:

$$\begin{aligned}
\mathbb{E}\left(\frac{T}{\mathbb{P}(T=1|Z^*)}|Z\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{T}{\mathbb{P}(T=1|Z^*)}|Z,Z^*\right)|Z\right) && \text{(A.3)}\\
&= \mathbb{E}\left(\frac{\mathbb{E}\left(T|Z,Z^*\right)}{\mathbb{P}(T=1|Z^*)}|Z\right) && \text{(A.4)}\\
&= \mathbb{E}\left(\frac{\mathbb{E}\left(T|Z^*\right)}{\mathbb{P}(T=1|Z^*)}|Z\right) && \text{(A.5)}\\
&= 1. && \text{(A.6)}
\end{aligned}$$

Combining expressions A.2 and A.6 yields the moment condition 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 5), 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 $\mathbb{E}(g(Y,T)|Z^*=z^*)=h(z^*)$. Let us consider the derivation of the following moment condition:

$$\mathbb{E}\left(\frac{h(z^*)}{\mathbb{P}(T=1|Z^*)}|T=1,Z\right) = \frac{\mathbb{E}(g(Y,T)|Z)}{\mathbb{P}(T=1|Z)}. \tag{A.7}$$

It follows the same steps as above. Under Assumption 4, we can consider the following expectation:

$$\mathbb{E}\left(\frac{h(Z^*).T}{\mathbb{P}(T=1|Z^*)}|Z\right) = \mathbb{E}\left(\frac{h(Z^*)}{\mathbb{P}(T=1|Z^*)}|T=1,Z\right)\mathbb{P}(T=1|Z). \tag{A.8}$$

We then apply the law of iterated expectation on the left-hand side of the previous equation and simplify its expression using Assumption 2:

$$
\begin{aligned}
\mathbb{E}\left(\frac{h(Z^*).T}{\mathbb{P}(T=1|Z^*)}|Z\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{h(Z^*).T}{\mathbb{P}(T=1|Z^*)}|Z,Z^*\right)|Z\right) && \text{(A.9)}\\
&= \mathbb{E}\left(\frac{h(Z^*).\mathbb{E}\left(T|Z,Z^*\right)}{\mathbb{P}(T=1|Z^*)}|Z\right) && \text{(A.10)}\\
&= \mathbb{E}\left(\frac{h(Z^*).\mathbb{E}\left(T|Z^*\right)}{\mathbb{P}(T=1|Z^*)}|Z\right) && \text{(A.11)}\\
&= \mathbb{E}\left(\mathbb{E}(g(Y,T)|Z^*)|Z\right) && \text{(A.12)}\\
&= \mathbb{E}\left(\mathbb{E}(g(Y,T)|Z^*,Z)|Z\right) && \text{(A.13)}\\
&= \mathbb{E}\left(g(Y,T)|Z\right). && \text{(A.14)}
\end{aligned}
$$

Combining expressions A.8 and A.14 yields the moment condition A.7. 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.
Assumption 3 ensures that $F_{Z|Z^*,T=1}$ is identified. And next, Assumption 2 ensures that $F_{Z|Z^*}$ is identified. Because $F_{Y,T|Z^*}$ and $F_{Z|Z^*}$ are identified, $F_{Y,T,Z|Z^*}$ is also identified under the Assumption 2. Identification of $\mathbb{P}(T=1|Z^*)$ using Equation (A.1), identification of F_T and $F_{Z^*|T=1}$ by Assumption 3 and the Bayes formula altogether ensure that F_{Z^*} is identified. It follows that $F_{Z^*,Z,T,Y}$ is identified.

A.3 Discussion of Assumptions 5, 6.4 and 6.5

The following Lemma shows that Assumptions 5, 6.4 and 6.5 are compatible with classical measurement error model, but more generally with a large class of measurement error process.

Lemma A.1 (A Class of models of measurement error)
Let Z^* a random variable with support $[-1;1]$ that admits a density f_{Z^*} with respect to the Lebesgue measure such that f_{Z^*} is continuous on $]-1;1[$.
Let ε a random variable with support $[\underline{\varepsilon};\overline{\varepsilon}]$ independent of Z^* that admits a density f_ε with respect to the Lebesgue measure such that f_ε admits a bounded total variation.
Let

$$Z = \mu(\nu(Z^*) + \varepsilon),$$

with ν an increasing and bounded C^1 diffeomorphism from $]-1;1[$ to $]\underline{\nu};\overline{\nu}[$ and μ an increasing and bounded C^1 diffeormorphism from $]\underline{\nu}+\underline{\varepsilon};\overline{\nu}+\overline{\varepsilon}[$ to $]l;u[$.
Assume that it exists constants $C_k^{\pm} > 0$ and $\alpha_k^{\pm}$ such that for $x \downarrow 0$:

1. $f_{Z^*}(1-x) \sim C_1^+ x^{\alpha_1^+}$ *and* $f_{Z^*}(-1+x) \sim C_1^- x^{\alpha_1^-}$

2. $(\nu^{-1})'(\overline{\nu}-x) \sim C_2^+ x^{\alpha_2^+}$ *and* $(\nu^{-1})'(\underline{\nu}+x) \sim C_2^- x^{\alpha_2^-}$,

3. $(\mu^{-1})'(\overline{\nu}+\overline{\varepsilon}-x) \sim C_3^+ x^{\alpha_3^+}$ *and* $(\mu^{-1})'(\underline{\nu}+\underline{\varepsilon}+x) \sim C_3^+ x^{\alpha_3^-}$

4. $f_\varepsilon(\overline{\varepsilon}-x) \sim C_4^+ x^{\alpha_4^+}$ *and* $f_\varepsilon(\underline{\varepsilon}+x) \sim C_4^- x^{\alpha_4^-}$,

5. $\alpha_1^+ + \alpha_2^+ + \alpha_1^+\alpha_2^+ \geq 0$ *and* $\alpha_1^- + \alpha_2^- + \alpha_1^-\alpha_2^- \geq 0$.

Assumptions 5, 6.4 and 6.5 hold with
$\alpha_u = 1 + \alpha_1^+ + \alpha_2^+ + 2\alpha_3^+ + \alpha_4^+ + \alpha_1^+\alpha_2^+ + \alpha_1^+\alpha_3^+ + \alpha_2^+\alpha_3^+ + \alpha_1^+\alpha_2^+\alpha_3^+ + \alpha_3^+\alpha_4^+$,
$\alpha_l = 1 + \alpha_1^- + \alpha_2^- + 2\alpha_3^- + \alpha_4^- + \alpha_1^-\alpha_2^- + \alpha_1^-\alpha_3^- + \alpha_2^-\alpha_3^- + \alpha_1^-\alpha_2^-\alpha_3^- + \alpha_3^-\alpha_4^-$
and $\kappa = \max(\alpha_u - \alpha_3^+, \alpha_l - \alpha_3^-, 0) - \min(\alpha_3^+, \alpha_3^-, 0)$
as soon as $\alpha_u \geq 0$ *and* $\alpha_l \geq 0$.

If μ and ν admit derivatives bounded away from 0 with bounded variations, then Conditions 2 and 3 hold with $\alpha_2^+ = \alpha_2^- = \alpha_3^+ = \alpha_3^- = 0$. In that case, f_{Z^*} and f_ε fulfill the other conditions as soon as:

- they admit non negative finite limits at the boundary of the support (i.e. uniform, truncated normal, any continuous density bounded away from 0 with bounded variation close to the boundary of its support),

- or they admit limits 0 at the boundary of the support with a polynomial decay (triangular, Beta distribution with parameter greater than 1)

- or they are any convolutions of such distributions.

Proof:
First note that f_ε has bounded variation and then is bounded, it follows that $\alpha_4^+ \geq 0$ and $\alpha_4^- \geq 0$.
We first prove the completeness condition. Because ε has a bounded support, it has finite exponential moment: $\mathbb{E}(|e^{u\varepsilon}|) < \infty$ for any $u \in \mathbb{C}$. Then $u \mapsto \mathbb{E}(e^{u\varepsilon})$ is a Laplace-Fourier transform with domain $\mathbb{C}$. Any Laplace-Fourier transform on $\mathbb{C}$ is an entire function (see for instance Schwartz (1997), Chapter VIII) and then admits only isolated zeros. Then, Proposition 2.4 in D'Haultfœuille (2011) ensures that completeness condition holds.
Let $f_{\mu^{-1}(Z)}$ the convolution product of $f_{\nu(Z^*)}(v) = f_{Z^*}(\nu^{-1}(v))[\nu^{-1}]'(v)$ and f_ε, $f_{\mu^{-1}(Z)}$ is a density with respect to the Lebesgue measure of $\mu^{-1}(Z) = \nu(Z^*) + \varepsilon$.
Because f_{Z^*}, (respectively ν^{-1} and $[\nu^{-1}]'$) is continuous respectively on $]-1;1[$ (respectively on the interior of $]\underline{\nu};\overline{\nu}[$), $f_{\nu(Z^*)}(v)$ is continuous on $]\underline{\nu};\overline{\nu}[$ and next is bounded on $\mathbb{R}$ as soon

as $\lim_{x\uparrow\overline{\nu}} f_{\nu(Z^*)}$ and $\lim_{x\downarrow\underline{\nu}} f_{\nu(Z^*)}$ are well defined and finite. For $x\downarrow 0$: $1-\nu^{-1}(\overline{\nu}-x)\sim C_2^+ x^{\alpha_2^+ +1}$ and $\nu^{-1}(\overline{\nu}-x)=1-C_2^+ x^{\alpha_2^+ +1}+o(x^{\alpha_2^+ +1})$ and then:

$$\begin{aligned}
f_{\nu(Z^*)}(\overline{\nu}-x) &= f_{Z^*}(\nu^{-1}(\overline{\nu}-x))(\nu^{-1})'(\overline{\nu}-x)\\
&= f_{Z^*}(1-C_2^+ x^{\alpha_2^+ +1}+o(x^{\alpha_2^+ +1}))\times(C_2^+ x^{\alpha_2^+}+o(x^{\alpha_2^+}))\\
&\sim C_1^+(C_2^+)^{\alpha_1^+ +1}x^{\alpha_1^+ +\alpha_2^+ +\alpha_1^+\alpha_2^+},
\end{aligned}$$

and for $x\uparrow 0$: $f_{\nu(Z^*)}(\underline{\nu}+x)\sim C_1^-(C_2^-)^{\alpha_1^- +1}x^{\alpha_1^- +\alpha_2^- +\alpha_1^-\alpha_2^-}$.
Then $\left|f_{\mu^{-1}(Z)}(v_1)-f_{\mu^{-1}(Z)}(v_2)\right|\leq ||f_{\nu(Z^*)}||_\infty \int |f_\varepsilon(v_1-u)-f_\varepsilon(v_2-u)|du$. Because f_ε has a bounded total variation,

$$V(e)=\sup\{\sum_{i=1}^N |f_\varepsilon(e_i)-f_\varepsilon(e_{i-1})|; N\in\mathbb{N}^*, e_0<e_1<...<e_N\leq e\}$$

is an increasing and bounded function which is null for $e<\underline{\varepsilon}$ and constant on $]\overline{\varepsilon};+\infty[$. Let $TV(f_\varepsilon)$ the total variation of f_ε, ie $\sup_{e\in\mathbb{R}^+}V(e)$.Following the Jordan decomposition, let $f_\varepsilon^+(e)=\frac{1}{2}(V(e)+f_\varepsilon(e))$ and $f_\varepsilon^-(e)=\frac{1}{2}(V(e)-f_\varepsilon(e))$. This two functions are both non decreasing functions such that $f_\varepsilon(e)=f_\varepsilon^+(e)-f_\varepsilon^-(e)$, $f_\varepsilon^+(e)=f_\varepsilon^-(e)=0$ for $e<\underline{\varepsilon}$ and $f_\varepsilon^+(e)=f_\varepsilon^-(e)=TV(f_\varepsilon)/2$ for $e\geq\overline{\varepsilon}$. Let $\widetilde{f}_\varepsilon^+(e)=\lim_{x\downarrow e} f_\varepsilon^+(x)$ and $\widetilde{f}_\varepsilon^-(e)=\lim_{x\downarrow e} f_\varepsilon^-(x)$. We always have the relations $f_\varepsilon(e)=\widetilde{f}_\varepsilon^+(e)-\widetilde{f}_\varepsilon^-(e)$, $\widetilde{f}_\varepsilon^+(e)=\widetilde{f}_\varepsilon^-(e)=0$ for $e<\underline{\varepsilon}$ and $\widetilde{f}_\varepsilon^+(e)=\widetilde{f}_\varepsilon^-(e)=TV(f_\varepsilon)/2$ for $e\geq\overline{\varepsilon}$ but now the variation of $\widetilde{f}_\varepsilon^\pm$ can be expressed as Lebesgue-Stieljes integrals:
$\widetilde{f}_\varepsilon^\pm(e_1)-\widetilde{f}_\varepsilon^\pm(e_2)=\int \mathbb{1}_{\{e_2\leq x\leq e_1\}}d\widetilde{f}_\varepsilon^\pm(x)$.
For $v_1>v_2$ we have:

$$\begin{aligned}
\int |f_\varepsilon(v_1-u)-f_\varepsilon(v_2-u)|\,du &= \int |f_\varepsilon(v_1-v_2+u)-f_\varepsilon(u)|\,du\\
&\leq \int \widetilde{f}_\varepsilon^+(v_1-v_2+u)-\widetilde{f}_\varepsilon^+(u)du\\
&\quad +\int \widetilde{f}_\varepsilon^-(v_1-v_2+u)-\widetilde{f}_\varepsilon^-(u)du\\
&= \int\left(\int \mathbb{1}_{\{u\leq x\leq v_1-v_2+u\}}d\widetilde{f}_\varepsilon^+(x)\right)du\\
&\quad +\int\left(\int \mathbb{1}_{\{u\leq x\leq v_1-v_2+u\}}d\widetilde{f}_\varepsilon^-(x)\right)du\\
&= \int\left(\int \mathbb{1}_{\{x-(v_1-v_2)\leq u\leq x\}}du\right)d\widetilde{f}_\varepsilon^+(x)\\
&\quad +\int\left(\int \mathbb{1}_{\{x-(v_1-v_2)\leq u\leq x\}}du\right)d\widetilde{f}_\varepsilon^-(x)\\
&= (v_1-v_2)\left(\int d\widetilde{f}_\varepsilon^+(x)+\int d\widetilde{f}_\varepsilon^-(x)\right)\\
&= (v_1-v_2)\left(\widetilde{f}_\varepsilon^+(\overline{\varepsilon})+\widetilde{f}_\varepsilon^-(\overline{\varepsilon})\right)\\
&= (v_1-v_2)TV(f_\varepsilon).
\end{aligned}$$

And similar reasoning holds if $v_1 < v_2$ and thus:

$$\int |f_\varepsilon(v_1 - u) - f_\varepsilon(v_2 - u)|\, du \le TV(f_\varepsilon)|v_1 - v_2|.$$

It follows that $f_{\mu^{-1}(Z)}$ is Lipschitz continuous on $]\underline{\nu} + \underline{\varepsilon}; \overline{\nu} + \overline{\varepsilon}[$. Because μ^{-1} and $(\mu^{-1})'$ are continuous on $]l;u[$, $f_Z(z) = f_{\mu^{-1}(Z)}(\mu^{-1}(z)) \times |(\mu^{-1})'(z)|$ is a density of Z, continuous on $]l;u[$.

Because ν and μ are C^1 diffeormorphisms, we have $\infty > \mu' > 0$, $\infty > \nu' > 0$, $\infty > (\mu^{-1})' > 0$ and $\infty > (\nu^{-1})' > 0$. For $z \in]l;u[$, if $f_Z(z) = 0$ we have successively:

$$(f_{\nu(Z^*)} \star f_\varepsilon)(\mu^{-1}(z)) = 0,$$

$$\int f_{\nu(Z^*)}(\mu^{-1}(z) - e) f_\varepsilon(e) de = 0,$$

$f_{Z^*}(\nu^{-1}(\mu^{-1}(z) - e)) f_\varepsilon(e) = 0$, for almost all $e \in [\min(\mu^{-1}(z) - \overline{\nu}, \underline{\varepsilon}); \max(\mu^{-1}(z) - \underline{\nu}, \overline{\varepsilon})]$, which contradicts the fact that the support of f_{Z^*} and f_ε are conex set $[-1;1]$ and $[\underline{\varepsilon}; \overline{\varepsilon}]$. So $f_Z(z) > 0$ for almost all $z \in]l;u[$ and continuous on $]l;u[$, then f_Z is bounded below on any compact of $]l;u[$.

To ensure that Assumption 6.4 holds, we have to derive the behavior of f_Z at the boundary of its support. We consider here the limit for $z \uparrow u$, but similar reasoning holds for the lower bound l. For $x \downarrow 0$:

$$1 - \nu^{-1}(\overline{\nu} - x) \sim C_2^+ x^{\alpha_2^+ + 1},$$

$$\nu^{-1}(\overline{\nu} - x) = 1 - C_2^+ x^{\alpha_2^+ + 1} + o(x^{\alpha_2^+ + 1})$$

$$\begin{aligned}
f_{\nu(Z^*)}(\overline{\nu} - x) &= f_{Z^*}(\nu^{-1}(\overline{\nu} - x))(\nu^{-1})'(\overline{\nu} - x) \\
&= f_{Z^*}(1 - C_2^+ x^{\alpha_2^+ + 1} + o(x^{\alpha_2^+ + 1})) \times (C_2^+ x^{\alpha_2^+} + o(x^{\alpha_2^+})) \\
&\sim C_1^+ (C_2^+)^{\alpha_1^+ + 1} x^{\alpha_1^+ + \alpha_2^+ + \alpha_1^+ \alpha_2^+},
\end{aligned}$$

$$\begin{aligned}
(f_{\nu(Z^*)} \star f_\varepsilon)(\overline{\nu} + \overline{\varepsilon} - x) &= x \int_0^1 f_{\nu(Z^*)}(\overline{\nu} - x + vx) f_\varepsilon(\overline{\varepsilon} - vx) dv \\
&\sim C_1^+ (C_2^+)^{\alpha_1^+ + 1} C_4^+ x^{\alpha_4^+ + 1 + \alpha_1^+ + \alpha_2^+ + \alpha_1^+ \alpha_2^+} \int_0^1 (1 - v)^{\alpha_1^+ + \alpha_2^+ + \alpha_1^+ \alpha_2^+} v^{\alpha_4^+} dv \\
&\sim C_1^+ (C_2^+)^{\alpha_1^+ + 1} C_4^+ x^{\alpha_4^+ + 1 + \alpha_1^+ + \alpha_2^+ + \alpha_1^+ \alpha_2^+} \frac{\Gamma((1+\alpha_1^+)(1+\alpha_2^+))\Gamma(1+\alpha_4^+)}{\Gamma((1+\alpha_1^+)(1+\alpha_2^+) + 1 + \alpha_4^+)},
\end{aligned}$$

$$\mu^{-1}(u) - (\mu^{-1})(u - x) \sim C_3^+ x^{\alpha_3^+ + 1},$$

$$\begin{aligned}
\mu^{-1}(u - x) &= \mu^{-1}(u) - C_3^+ x^{\alpha_3^+ + 1} + o(x^{\alpha_3^+ + 1}) \\
&= \overline{\nu} + \overline{\varepsilon} - C_3^+ x^{\alpha_3^+ + 1} + o(x^{\alpha_3^+ + 1}),
\end{aligned}$$

$$
\begin{aligned}
f_Z(u-x) &= (f_{\nu(Z^*)} \star f_\varepsilon)(\mu^{-1}(u-x)) \times (\mu^{-1})'(u-x) \\
&= (f_{\nu(Z^*)} \star f_\varepsilon)(\overline{\nu} + \overline{\varepsilon} - C_3^+ x^{\alpha_3^+ +1} + o(x^{\alpha_3^+ +1})) \times (C_3^+ x^{\alpha_3^+} + o(x^{\alpha_3^+}) \\
&\sim C_1^+ (C_2^+)^{\alpha_1^+ +1} C_4^+ (C_3^+ x^{\alpha_3^+ +1})^{\alpha_4^+ +1+\alpha_1^+ +\alpha_2^+ +\alpha_1^+\alpha_2^+} \tfrac{\Gamma((1+\alpha_1^+)(1+\alpha_2^+))\Gamma(1+\alpha_4^+)}{\Gamma((1+\alpha_1^+)(1+\alpha_2^+)+1+\alpha_4^+)} \times C_3^+ x^{\alpha_3^+} \\
&\sim C_u x^{\alpha_u},
\end{aligned}
$$

with $C_u = C_1^+ (C_2^+)^{\alpha_2^+ +1} (C_3^+)^{2+\alpha_1^+ +\alpha_2^+ +\alpha_1^+\alpha_2^+ +\alpha_4^+} C_4^+ \frac{\Gamma((1+\alpha_1^+)(1+\alpha_2^+))\Gamma(1+\alpha_4^+)}{\Gamma((1+\alpha_1^+)(1+\alpha_2^+)+1+\alpha_4^+)}$ and
$\alpha_u = 1 + \alpha_1^+ + \alpha_2^+ + 2\alpha_3^+ + \alpha_4^+ + \alpha_1^+\alpha_2^+ + \alpha_1^+\alpha_3^+ + \alpha_2^+\alpha_3^+ + \alpha_1^+\alpha_2^+\alpha_3^+ + \alpha_3^+\alpha_4^+$.
Similar reasoning hold for the lower l. And then Assumption 6.4 holds.
We now prove that Assumption 6.5 holds.
Let $\delta > 0$ and $(z, z') \in I = [l+\delta; u-\delta]$. Let $g(z)$ and $g(z')$ denote respectively $f_{\mu^{-1}(Z)}(\mu^{-1}(z))$ and $f_{\mu^{-1}(Z)}(\mu^{-1}(z'))$. Let $g(z|z^*)$ and $g(z'|z^*)$ denote respectively $f_{\mu^{-1}(Z)|Z^*=z^*}(\mu^{-1}(z))$ and $f_{\mu^{-1}(Z)|Z^*=z^*}(\mu^{-1}(z'))$:

$$
\begin{aligned}
\int |f_{Z^*|Z=z}(z^*) - f_{Z^*|Z=z'}(z^*)|dz^* &= \int \left| \tfrac{g(z|z^*)}{g(z)} - \tfrac{g(z'|z^*)}{g(z')} \right| f_{Z^*}(z^*)dz^* \\
&\leq \tfrac{1}{2} \left| \tfrac{1}{g(z)} - \tfrac{1}{g(z')} \right| \times \int [g(z|z^*) + g(z'|z^*)]\, f_{Z^*}(z^*)dz^* \\
&\quad + \tfrac{1}{2} \left(\tfrac{1}{g(z)} + \tfrac{1}{g(z')} \right) \times \int |g(z|z^*) - g(z'|z^*)|\, f_{Z^*}(z^*)dz^* \\
&= \tfrac{1}{2} \left| \tfrac{1}{g(z)} - \tfrac{1}{g(z')} \right| \times (g(z) + g(z')) \\
&\quad + \tfrac{1}{2} \left(\tfrac{1}{g(z)} + \tfrac{1}{g(z')} \right) \times \int |g(z|z^*) - g(z'|z^*)|\, f_{Z^*}(z^*)dz^* \\
&= \tfrac{1}{2} \left(\tfrac{1}{g(z)} + \tfrac{1}{g(z')} \right) \times \\
&\quad \left\{ |g(z) - g(z')| + \int |g(z|z^*) - g(z'|z^*)|\, f_{Z^*}(z^*)dz^* \right\} \\
&\leq \left(\tfrac{1}{g(z)} + \tfrac{1}{g(z')} \right) \times \int |g(z|z^*) - g(z'|z^*)|\, f_{Z^*}(z^*)dz^*.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\int |g(z|z^*) - g(z'|z^*)| f_{Z^*}(z^*)dz^* &= \int |f_\varepsilon(\mu^{-1}(z) - \nu(z^*)) - f_\varepsilon(\mu^{-1}(z') - \nu(z^*))| f_{Z^*}(z^*)dz^* \\
&\leq \int |f_\varepsilon(\mu^{-1}(z) - u) - f_\varepsilon(\mu^{-1}(z') - u)|\, f_{\nu(Z^*)}(u)du \\
&\leq ||f_{\nu(Z^*)}||_\infty TV(f_\varepsilon)\, |\mu^{-1}(z) - \mu^{-1}(z')|
\end{aligned}
$$

with $|\mu^{-1}(z) - \mu^{-1}(z')| \leq \sup_{x\in I} |(\mu^{-1})'(x)| \times |z - z'|$.
Last, for $\delta \downarrow 0$, it exists $K_1, K_2 > 0$ such that $\inf_{z\in I} g(z) \sim K_1 \delta^{\max(\gamma^+, \gamma^-, 0)}$ with $\gamma^+ = \alpha_u - \alpha_3^+$ and $\gamma^- = \alpha_l - \alpha_3^-$ and $\sup_{x\in I} |(\mu^{-1})'(x)| \sim K_2 \delta^{\min(\alpha_3^+, \alpha_3^-, 0)}$. Then Assumption 6.5 holds with $\kappa = \max(\gamma^+, \gamma^-, 0) - \min(\alpha_3^+, \alpha_3^-, 0)$.

A.4 Proof of Theorem 2

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:

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

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 $\xi_{n,n_a} \in \mathcal{H}_{n,n_a}$ 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 $\sup_{\xi \in \mathcal{H}_{n,n_a}} |Q_n(\xi) - Q(\xi)| = o_p\left(1\right)$.
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 $\inf_{\xi \in \mathcal{H}_{n,n_a}, ||\xi-\xi_0||_\infty \geq \delta} Q(\xi) \geq \inf_{\xi \in \mathcal{H}, ||\xi-\xi_0||_\infty \geq \delta} Q(\xi) \geq Q(\xi^*) > 0$. Because $n/n_a \to \lambda \in]0; +\infty[$, it follows that $X_{n,n_a} = o_p\left(\inf_{\xi \in \mathcal{H}_{n,n_a}, ||\xi-\xi_0||_\infty \geq \delta} Q(\xi)\right)$ if and only if $X_{n,n_a} = o_p(1)$ and $x_{n,n_a} = o\left(\inf_{\xi \in \mathcal{H}_{n,n_a}, ||\xi-\xi_0||_\infty \geq \delta} Q(\xi)\right)$ if and only if $x_{n,n_a} = o(1)$.

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 $||\xi_{n,n_a} - \xi_0||_\infty = O((min(k^{p+}_{n,n_a}, k^{p-}_{n,n_a}, k^{0+}_{n,n_a}, k^{0-}_{n,n_a}, k^{1+}_{n,n_a}, k^{1-}_{n,n_a}))^{-1})$. Because $min(k^{p+}_{n,n_a}, k^{p-}_{n,n_a}, k^{0+}_{n,n_a}, k^{0-}_{n,n_a}, k^{1+}_{n,n_a}, k^{1-}_{n,n_a}) \to +\infty$, we have $||\xi_{n,n_a} - (p, m_0, m_1)||_\infty = o(1)$ 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:

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

   In the following, we prove that:

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

   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:

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

Then by Cauchy-Schwartz inequality,

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

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 $\mathcal{H}^p_{n,n_a}$, 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 $\xi\in\mathcal{H}^p_{n,n_a}$ 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 $\mathcal{H}^p_{n,n_a}$, we will use various theorems related to the behavior of empirical process, as explained in van der Vaart & Wellner (1996) or in van der Vaart (2000), Chapter 19.

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 $l_{n,n_a}:=\dim\left(\mathcal{I}^p_{n,n_a}\right)$ 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 2 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 $\mathcal{I}^p_{n,n_a}$.

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 2, namely $l^p_{n,n_a}=o(n^{1/(2+\max(\alpha_u,\alpha_l))})$, 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:

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

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

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

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:

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

Moreover, the Markov inequality ensures that there exists a constant M (uniform in

l_{n,n_a}) such that:

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

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:

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

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 6.1, 6.2, 6.3 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:

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

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 $\check{g}_p(z,\xi)$ 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,

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

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

The first term is such that:

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

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 $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ and again and $\sup_x \sum_k f_k(x) \leq \sum_k \sup_x f_k(x)$. 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

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

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),

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

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:

$$\begin{aligned}
&\mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\tfrac{1}{\sqrt{n_a}}\textstyle\sum_{i\in S_a}\overline{B}_j(Z_i)/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)/\xi(Z^*)|T=1\right)\right|\right)^2\right)\\
&\leq M_3 + M_4\textstyle\int_0^1\left(1+LogN_{[]}(M_5u,\mathcal{F}_j^1,||.||_{L^2(W|T=1)})\right)^{1/2}du\\
&\text{and}\\
&\mathbb{E}\left(\left(\sup_{\xi\in\mathcal{H}}\left|\tfrac{1}{\sqrt{n}}\textstyle\sum_{i\in S}\overline{B}_j(Z_i)T_i/\xi(Z_i^*) - \mathbb{E}\left(\overline{B}_j(Z)T/\xi(Z^*)\right)\right|\right)^2\right)\\
&\leq M_6 + M_7\textstyle\int_0^1\left(1+LogN_{[]}(M_8u,\mathcal{F}_j,||.||_{L^2(W)})\right)^{1/2}du.
\end{aligned}$$

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:

$$\begin{aligned}
&N_{[]}(u,\mathcal{F}_j^1,||.||_{L^2(W|T=1)}) \leq N_{[]}(u\underline{c}^{1/2}||\overline{B}_j(Z)||_{L^2(W|T=1)},\mathcal{F}_j^1,||.||_{L^2(W|T=1)})\\
&\leq N(u\underline{c}^{1/2}/2,\mathcal{H}^P,||.||_\infty) \leq \exp(M_9u^{-1}),\\
&\text{and}\\
&N_{[]}(u,\mathcal{F}_j,||.||_{L^2(W)}) \leq N_{[]}(u||\overline{B}_j(Z)T||_{L^2(W)},\mathcal{F}_j,||.||_{L^2(W)}) \leq N(u/2,\mathcal{H}^P,||.||_\infty) \leq \exp(M_{10}u^{-1}).
\end{aligned}$$

It follows that:

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

And last, because $\frac{n}{n_a} \to \lambda \in]0;+\infty[$,

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

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:

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

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:

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

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:

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

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 6.4 and 6.5, 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 $\beta < 1$, it exists $D_1 > 0$ such that for any $(z, z') \in I$ and for any ρ:

$$|\mathbb{E}(\rho(Z^*|Z = z) - \mathbb{E}(\rho(Z^*|Z = z')| \leq \int |\rho(z^*)||f_{Z^*|Z=z}(z^*) - f_{Z^*|Z=z'}(z^*)| \leq D_1|z-z'|k^{(1-\beta)\kappa}.$$

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 $\mathbb{E}(\rho(Z^*)|Z = i/k)$ otherwise. It exists D_2 such that for all ρ: $\sup_{z\in I} |\mathbb{E}(\rho(Z^*)|Z = z) - B(z)\pi_\rho| \leq D_2 k^{(1-\beta)\lambda - 1}$. It follows that $\int_I [\mathbb{E}(\rho(Z^*)|Z) - B(Z)\pi_\rho]^2 f_Z(z)dz \leq D_2^2 k^{2(1-\beta)\lambda-2}$.
Moreover, it exists D_3, D_4 such that:

$$\int_l^{l+\lfloor k^\beta \rfloor/k} [\mathbb{E}(\rho(Z^*)|Z) - B(Z)\pi_\rho]^2 f_Z(z)dz \leq \int_l^{l+\lfloor k^\beta \rfloor/k} f_Z(z)dz \leq D_3 k^{(\beta-1)(\alpha_l+1)}$$

$$\int_{u-\lfloor k^\beta \rfloor/k}^{u} [\mathbb{E}(\rho(Z^*)|Z) - B(Z)\pi_\rho]^2 f_Z(z)dz \leq \int_{u-\lfloor k^\beta \rfloor/k}^{u} f_Z(z)dz \leq D_4 k^{(\beta-1)(\alpha_u+1)}$$

For β sufficiently close to 1, we have $\gamma := \min\left(2 - 2(1-\beta)\lambda, (1-\beta)(\alpha_u+1), (1-\beta)(\alpha_l+1)\right) \geq 0$ and $M = \max(D_2^2, D_3, D_4)$ such that:

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

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: Supplementary Tables

Table 3: Estimation of the LATE in finite samples, Multiplicative Error

		A. Small Error, $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$			B. Large Error, $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$		
		Sample Size			Sample Size		
Estimator	Stat.	1000	5000	25000	1000	5000	25000
$\hat{\theta}$ $(k=0)$	Bias	-0.969	0.344	0.678	-0.009	0.681	0.806
	Variance	10.63	1.009	0.024	2.818	0.127	0.018
$\hat{\theta}$ $(k=1)$	Bias	-0.530	0.351	0.512	-0.188	0.349	0.512
	Variance	24.98	0.328	0.078	31.85	0.423	0.110
$\hat{\theta}$ $(k=2)$	Bias	-0.738	0.074	0.172	0.249	0.142	0.174
	Variance	325.0	0.578	0.097	943.9	2.639	0.573
$\tilde{\theta}$ $(k=0)$	Bias	-1.091	0.212	0.541	-0.105	0.561	0.674
	Variance	10.61	1.033	0.025	2.827	0.133	0.018
$\tilde{\theta}$ $(k=1)$	Bias	-0.547	0.291	0.444	-0.223	0.277	0.422
	Variance	20.44	0.334	0.079	22.60	0.382	0.104
$\tilde{\theta}$ $(k=2)$	Bias	-0.790	0.054	0.137	0.044	0.122	0.119
	Variance	312.0	0.562	0.100	700.2	1.430	0.441
Donut	Bias	1.897	1.846	1.854	2.392	2.361	2.365
	Variance	0.059	0.012	0.002	0.153	0.025	0.005

Note: Computation obtained with 1000 simulations. $Z+1 = (Z^*+1)\cdot(1+\varepsilon)$ 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 4: Estimation of the LATE in finite samples, Additive Error

Estimator	Stat.	A. Small Error, $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$ Sample Size 1000	5000	25000	B. Large Error, $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$ Sample Size 1000	5000	25000
	Bias	-5.006	-1.729	-0.443	-4.040	-1.326	-0.149
Our ($k=0$)	Variance	269.3	8.133	2.688	123.9	6.770	1.843
	Median	-3.993	-3.262	-2.569	-3.122	-2.793	-2.488
	IQ range	6.653	4.197	0.902	5.902	3.725	0.243
	Bias	-1.250	-0.445	-0.612	-2.231	-0.220	-0.309
Our ($k=1$)	Variance	120.4	3.475	0.882	890.4	3.520	1.122
	Median	-2.898	-2.776	-3.899	-2.888	-2.693	-3.290
	IQ range	3.424	1.976	1.657	3.267	2.046	1.815
	Bias	1.939	-0.124	0.027	-15.91	0.661	0.211
Our ($k=2$)	Variance	836.7	18.80	1.535	2×10^5	285.2	9.320
	Median	-2.524	-2.736	-2.851	-2.294	-2.564	-2.683
	IQ range	3.237	2.022	1.964	4.274	2.753	2.070
	Bias	-1.198	-1.792	-0.702	2.287	4.009	1.935
	Variance	2518	2766	4419	584.8	1×10^4	2×10^4
Naive (IK)	Median	-2.506	-2.431	-1.761	-1.750	-1.720	-1.583
	IQ range	1.900	2.500	2.855	2.736	2.996	2.919
	Bdw. h_n	0.278	0.182	0.114	0.295	0.211	0.155
	Bias	-43.51	343.4	-1479	483.0	2350	-2704
	Variance	3×10^8	2×10^8	1×10^9	2×10^8	5×10^9	3×10^{10}
Naive (CCT)	Median	-2.555	-2.533	-2.136	-1.942	-1.982	-1.896
	IQ range	2.436	2.543	2.669	2.560	2.776	2.757
	Bdw. b_n	0.509	0.426	0.345	0.541	0.460	0.365

Note: Computation obtained with 1000 simulations. For our estimators k is the common number of knots selected to defined our linear splines. When $k=0$, functions p, m_0 and m_1 are approximated by linear functions on $[-1;0]$ and $[0;1]$. When $k=1$ (resp. 2), change in slope is allowed at $-1/2$ and $1/2$ (resp. $-2/3$, $-1/3$, $1/3$, $2/3$). For the Naive estimators, (IK) is the estimator proposed by Hahn et al. (2001) using the bandwidth h_n proposed by Imbens & Kalyanaraman (2012) and (CCT) is the bias-corrected estimator proposed by Calonico et al. (2014*b*) with bandwidth b_n used to estimate asymptotic bias of (IK).

Table 5: Estimation of the LATE in finite samples, Additive Error

		A. Small Error, $\varepsilon \sim \mathcal{U}_{[-0.1;0.1]}$			B. Large Error, $\varepsilon \sim \mathcal{U}_{[-0.2;0.2]}$		
		Sample Size			Sample Size		
Estimator	Stat.	1000	5000	25000	1000	5000	25000
$\hat{\theta}$ $(k=0)$	Bias	-5.006	-1.729	-0.443	-4.040	-1.326	-0.149
	Variance	269.3	8.133	2.688	123.9	6.770	1.843
$\hat{\theta}$ $(k=1)$	Bias	-1.250	-0.445	-0.612	-2.231	-0.220	-0.309
	Variance	120.4	3.475	0.882	890.4	3.520	1.122
$\hat{\theta}$ $(k=2)$	Bias	1.939	-0.124	0.027	-15.91	0.661	0.211
	Variance	836.7	18.80	1.535	2×10^5	285.2	9.320
$\tilde{\theta}$ $(k=0)$	Bias	-5.244	-1.938	-0.644	-4.243	-1.528	-0.344
	Variance	287.2	8.313	2.760	125.3	6.899	1.897
$\tilde{\theta}$ $(k=1)$	Bias	-1.344	-0.594	-0.767	-2.263	-0.37	-0.456
	Variance	96.30	3.509	0.901	1008	3.474	1.106
$\tilde{\theta}$ $(k=2)$	Bias	0.995	-0.237	-0.046	-4.12	0.426	0.097
	Variance	793.6	15.61	1.357	20937	189.6	6.973
Donut	Bias	1.861	1.846	1.839	2.356	2.309	2.303
	Variance	0.060	0.012	0.002	0.135	0.025	0.005

Note: Computation obtained with 1000 simulations.
$Z = Z^* + \varepsilon$ 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., Guldi, M., Lindo, J. & Waddell, G. (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.

Battistin, E. & Chesher, A. (2014), 'Treatment effect estimation with covariate measurement error', *Journal of Econometrics* **178**, 707–715.

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).

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014*a*), 'Robust data-driven inference in the regression-discontinuity design', *Stata-Journal* **14**(4), 909–946.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014*b*), 'Robust nonparametric confidence intervals for regression-discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Canay, I. A., Santos, A. & Shaikh, A. M. (2013), 'On the testability of identification in some non-parametric models with endogeneity', *Econometrica* **81**(6), 2535–2559.

Card, D. (2001), 'Estimating the return to schooling: Progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–60.

Card, D., Chetty, R. & Weber, A. (2007), 'Cash-on-hand and competing models of intertemporal substitution: New evidence from the labor market', *Quarterly Journal of Economics* **122**(4), 1511 – 1560.

Carrasco, M., Florens, J.-P. & Renault, E. (2007), Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization, *in* 'Handbook of Econometrics', Vol. 6 - Part B, Elsevier, chapter 77, pp. 5633–5751.

Carroll, R., Ruppert, D., Stefanski, L. & Crainiceanu, C. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press.

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., Hong, H. & Nekipelov, D. (2011), 'Nonlinear models of measurement errors', *Journal of Economic Literature* **49**(4), 901–37.
URL: *http://www.aeaweb.org/articles?id=10.1257/jel.49.4.901*

Chen, X., Hong, H. & Tamer, E. (2005), 'Measurement error models with auxiliary data', *Review of Economic Studies* **72**(2), 343–366.

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

Chen, X. & Pouzo, D. (2015), 'Sieve wald and qlr inferences on semi/nonparametric conditional moment models', *Econometrica* **83**(3), 1013–1079.

Davezies, L. & Le Barbanchon, T. (2014), Regression Discontinuity Design with Continuous Measurement Error in the Running Variable, Working Papers 2014-27, Centre de Recherche en Economie et Statistique.

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.

Gerard, F., Rothe, C. & Rokkanen, M. (2015), Identification and inference in regression discontinuity designs with a manipulated running variable, Working Paper 9604, Forschungsinstitut zur Zukunft der Arbeit.

Hahn, J., Todd, P. & van der Klaauw, W. (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. & Schennach, S. M. (2008), 'Instrumental variable treatment of nonclassical measurement error models', *Econometrica* **76**(1), pp. 195–216.

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. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), pp. 467–475.

Imbens, G. W. & Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *Review of Economic Studies* **79**(3), pp. 933–959.

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

Lalive, R. (2008), 'How do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach', *Journal of Econometrics* **142**(2), 785–806.

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.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics* **142**(2), 698–714.

Mortensen, D. T. (1977), 'Unemployment insurance and job search decisions', *Industrial and Labor Relations Review* **30**(4), 505–517.

Munkres, J. R. (2000), *Topology*, 2nd edn, Prentice Hall. ISBN 0-131-81629-2.

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

Newey, W. K. & Powell, 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, Unpublished manuscript.

Rokkanen, M. (2015), Exam schools, ability, and the effects of affirmative action: Latent factor extrapolation in the regression discontinuity design, Unpublished manuscript.

Schwartz, L. (1997), *Théorie des distributions*, third edn, Hermann. ISBN 2-7056-5551-4.

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.