A Discontinuity Test of Endogeneity*

Maria Carolina Caetano

UC Berkeley, Department of Economics

December 2009 (Last update: March 12, 2010)

PRELIMINARY AND INCOMPLETE, COMMENTS WELCOME

Abstract

This paper develops a nonparametric test of endogeneity without the need of instrumental variables. The test ensues from the novel observation that the potentially endogenous variable x is often of a nature such that the distribution of the unobservable q conditional on x and covariates z is discontinuous in x at a known value in its range. This relationship arises, for example, when x is subject to corner solutions, default contracts, social norms or law imposed restrictions, and may be argued using both economic theory and empirical evidence. If also x has a continuous effect on the dependent variable y, any discontinuity of y that is not accounted by the discontinuities in the covariates z is evidence that q and y are dependent conditional on z, i.e. it is evidence of the endogeneity of x. The analysis develops the test statistics and derives the asymptotic distribution for three versions of the test: linear, partially linear (nonparametric only on x but not on covariates) and non-parametric. Finally, the partially linear version of the test is applied to the estimation of the effect of maternal smoking on birth weight and on the probability of low birth weight (LBW). For the most detailed specification in the literature (Almond, Chay, and Lee (2005)), the test finds strong evidence of endogeneity in the case of birth weight, and very weak evidence in the case of the probability of LBW.

*I am indebted to Jim Powell for his advice throughout the entire duration of this project, and also to David Brillinger, Gregorio Caetano, Bin Chen, David Freedman and Nese Yildiz for their time and valuable input. This paper also benefited from discussions and comments from David Card, Matias Cattaneo, Ken Chay, Andres Donangelo, Shari Ely, Jeff Greenbaum, Bryan Graham, Michael Jansson, Pat Kline, Demian Pouzo, Ram Rajagopal, and Paul Ruud. Martha Berbert helped immensely in the selection of articles from the medical literature. Ken Chay provided support during a year of this project. Douglas Almond, Ken Chay and David Lee graciously provided the data used in this paper. I am thankful to all of them. All errors are my own.

1 Introduction

A test of endogeneity is in general a test of a condition satisfied by the data generating process under the assumption of exogeneity. A rejection in such test can only be interpreted within the assumptions made for the data generating process. A typical concern is when the data is modeled in a parametric framework and the test statistic derived in that context. In that case, the problem of endogeneity is indistinguishable from the problem generated by the wrong choice of model. Though the test perceives both issues in the same way, the solutions are entirely different. Endogeneity requires that a specific effort be made either to account for such unobservables with proxy variables, fixed effects in panel data sets, etc., or to eliminate their influence by the use of instrumental variables (IV). Wrong choice of model is solved by searching and testing different specifications. Nonparametric tests of endogeneity allow the interpretation of the rejection to mean exclusively the problem of endogeneity, and from that follows their importance.

Nonparametric tests of endogeneity are not abundant in the literature. This is in part due to the recency of the research on nonparametric IV estimators. Blundell and Powell (2003) and Hall and Horowitz (2005) discuss the difficulties involved in such undertaking, due to the fact that the identifying condition is an "ill-posed inverse problem." Nonparametric IV estimators of the structural function have been proposed in Darolles, Florens, and Renault (2003), Blundell, Chen, and Kristensen (2007), Newey and Powell (2003), and Hall and Horowitz (2005). The available tests of endogeneity suppose either that the potentially omitted variables are observed (see Fan and Li (1996), Chen and Fan (1999) and Li and Racine (2007)), or that an instrumental variable exists and is observed (see Blundell and Horowitz (2007) and Horowitz (2009)). In both cases, the effects are identified and can be consistently estimated, and the test is useful in the decision of which estimation strategy to pursue. This is no small concern in nonparametric estimation, because the rates of convergence of the estimators decrease considerably if irrelevant covariates are included, and even more if an instrumental variable approach is used where unrequited. The potential efficiency losses are therefore much more substantial than in the parametric cases. The test presented in this paper does not require that the omitted variables be observable, nor that an instrumental variable exist. Since most omitted variables are so because of being unobserved and since good instrumental variables are often not readily available, a test of endogeneity with no such requirements is of considerable interest. Its usefulness is twofold: first, the researcher can use the test for guidance in choosing an appropriate model, even with a selection on observables assumption. Second, in case the research finds evidence of endogeneity for any model with the selection on observables assumption, the researcher is alerted that another measure needs to be taken (search for IVs, search for another data set with more observable variables, etc.). To the author's knowledge, the discontinuity test is the first nonparametric test of endogeneity in the structural function without instrumental variables and where the omitted variables may be truly unobservable.

The discontinuity test ensues from the new observation that the presence of endogenous variables often generates discontinuities in the data generating process. Such is the case when the distribution of the unobservable variable conditional on the potentially endogenous variable (referred hereafter as the "running variable") is discontinuous at a certain value of the latter. This relationship between the running variable and other variables arises naturally, for example, when a selected part of the population concentrates at a point of the running variable. Examples of such phenomena include when the running variable is censored, in the sense that it cannot be chosen at a value above or below a certain threshold. The group that chose exactly the threshold point may be discontinuously different from the groups that chose immediately above the threshold. This is commonly observed when the running variable is a consumption good, which cannot be chosen in negative amounts. The argument in this case is that the observations at zero are discontinuously different from the observations at positive amounts. The discontinuity may exist because among everyone who chose zero there are not only those who would have optimally chosen zero in an unconstrained problem (who could indeed be similar to those who chose immediately positive amounts), but also those whose would have chosen negative amounts if they could (which can presumably be very different from those who chose immediately positive amounts). Other examples can be found in law imposed restrictions, such as minimum age required to drop out of school when the running variable is years of education, or minimum salary when the running variable is hourly wage.

Censored running variables are just one example where selected concentration happens. Another example is when the running variable is a choice variable for which default values are specified. For example, if the running variable is a level of insurance coverage and there exists a standard contract which can be tailored, the observations at the standard level may be discontinuously different from the observations at the tailored levels near the standard level. A very different example of the same nature is when the default value is not a result of the existence of a standard contract, but rather of a social norm. If the running variable is a continuous measure of inequality in the distribution of inheritance among the progeny, the observations where the division was perfectly equal may be discontinuously different from the observations at small levels of inequality.

There are other running variables for which selected concentration at a given point is present which are not of the types described above. An example is the running variable "weekly hours worked." The group that reports exactly forty hours may be discontinuously more likely to contain individuals in professions or positions with a fixed workload, such as the typical "9 to 5" worker. This phenomenon may generate a discontinuity in the distribution of many variables related to the choice of profession, industry or position conditional on the number of weekly hours worked at the exact level of forty hours.[1]

The discontinuity of the distribution of the unobservable variable at a given value

[1] Evidence of this can be found using Current Population Survey (CPS) data and is available from the author by request.

of the running variable cannot be proved, but it can be argued by showing that the observable covariates are discontinuous at the same value.[2] In the application section, the running variable "average cigarettes smoked per day" among pregnant women is studied, and it can be shown that the levels of education, marriage status, race distribution, prenatal visits, age, etc. are all discontinuous at zero cigarettes. Such discontinuities may be assumed to also hold for at least one unobserved variable. Examples of unobservable variables that may be discontinuous at zero cigarettes are whether the pregnancy was planned (or desired), how responsible or talented is the mother, etc.

For the applicability of the test, two conditions are necessary (and, given other regularity conditions, sufficient): that the distribution of the unobservable variables conditional on the running variable be discontinuous at a given point, and that the true structural function be continuous on the running variable. Take the example of maternal smoking, and suppose that the child's birth weight is explained by the average daily cigarettes and other observed and unobserved variables. The requirement is that cigarettes alone cannot explain a discontinuous change in birth weight. If this is accepted, then if the expected birth weight conditional on smoking and covariates is discontinuous in cigarettes at zero, it cannot be due to the effect of cigarettes on birth weight. The discontinuity is then attributed to the discontinuous change in the unobservable variables when comparing positive and zero cigarettes.

In principle, such a test could be performed by a nonparametric regression of the dependent variable on the running variables and covariates, and testing whether the resulting relationship is discontinuous at a given point for a given value of the covariates. The rate of convergence of such regressions is typically very slow, and hence such test would have little power. A much higher rate of convergence can be achieved through aggregation, i.e. by estimating, for example, the average discontinuity, or the correlation between the discontinuities and a function of the covariates, etc. For a wide variety of such tests, this paper shows that the rate of convergence is the same as that of a univariate nonparametric regression. Therefore, the aggregation allows for controlling the influence of the observable covariates without loss of power due to slower rates of convergence. This is a property observed in the literature of partial means (see Newey (1994)), or marginal integration (see Linton and Nielsen (1995)).

The discontinuities will be estimated in a similar fashion to what is done in the regression discontinuity literature, by estimating the one sided limits of the conditional expectation at a point. This entails nonparametric estimation at the boundary, which must be considered carefully due to the high boundary bias of the most commonly used estimators, such as Nadaraya-Watson, or series estimators using orthogonal bases like B-splines. Specifically, the bias at the boundary of the Nadaraya-Watson estimator is of the order of the bandwidth h, which is very large compared to the order h^2 of the bias at interior points (see Fan and Gijbels (1996) for a discussion). The results in this paper use local polynomial regression instead, which has bias of order at most h^{p+1}, where p is the degree of the polynomial, irrespective of the position of the point

[2] An analogous argument is made in the Regression Discontinuity Design literature. See Lee (2008).

in the support. Moreover, the local polynomial estimator adapts automatically to the estimation at the boundary, and therefore requires no more discretion from the applied researcher than for the estimation at an interior point. Local polynomial estimators are the preferred nonparametric approach in the regression discontinuity literature, as can be seen in Imbens and Lemieux (2008) and Porter (2003).

Following up on the parallel with regression discontinuity, it can be argued that the discontinuity test arises from an inversion of the identification assumptions in the regression discontinuity design. In the latter, the distribution of the unobservables conditional on the running variable has to be continuous at the threshold point, but the treatment has to be discontinuous at that point. In the discontinuity test setup, it is the treatment that has to be continuous at the threshold while the distribution of the unobservables is discontinuous at that point.

This paper is divided in the following way. It begins by detailing the test strategy and essential requirements somewhat informally in the context of a regression with no covariates in section 2.1. The intention is to provide a restrictive framework where the test is intuitively understood, so that no reader is lost on the details of the general case. To better illustrate the point, this section is written explicitly within the example of the effects of maternal smoking in birth weight. The following section (section 2.2) formally defines endogeneity and develops conditions for the identification of a parameter which equals zero if the running variable is exogenous. This parameter aggregates over some measure of the covariates the potential discontinuities of the conditional expectation of the dependent variable for each value of the covariates.

Section 2.3 focuses on the estimation of the parameter which identifies the endogeneity. If the researcher is willing to make assumptions about the functional form of the expectation of the dependent variable conditional on the running variable and covariates, more accurate test statistics can be developed. In the interest of applied research, section 2.3.1 provides the test statistic and asymptotic distribution when the conditional expectation is linear in both the running variable and the covariates, and section 2.3.2 does the same when the conditional expectation is partially linear, i.e. separably linear in the covariates and nonparametric in the running variable. These tests are naturally sensitive to the wrong choice of model, but they are easy to implement. The partially linear case is particularly flexible, because it allows for the inclusion of a very large number of covariates, which in practice is not always possible in fully nonparametric specifications. Section 2.3.3 presents the fully nonparametric test, which has the same rates of convergence as the partially linear case.

The assumptions are all expressed in terms of conditional expectations and probability distributions. However, inside of a specific model it is possible to propose primitive conditions that may be more interesting from the applied researcher's point of view. Throughout all the theory sections, an example is carried out of a model where the discontinuity in the distribution of the unobservable variable is created by censoring in the running variable. For example, the identification section has a subsection (2.2.1) where the identification strategy is explicitly shown in a model with censoring. Examples 1, 2 and 3 propose specific shapes of the functions in the censored model that imply that the conditional expectation is linear, partially linear or nonparametric respectively. Finally, the application section 3 presents a practical example of a problem that can be modeled within the censoring framework. It is an implementation of the test in the partially linear case to the effects of maternal smoking in both birth weight and in the probability of low birth weight (LBW), defined as birth weight lower than 2500 grams.

The test is not hard to implement. The linear case is trivial, and for the partially linear and fully nonparametric cases, all that is required for the estimation of the test statistic and its variance is the computation of some local polynomial regressions at the threshold point and some sample averages. The discretion requirements are only the choice of bandwidth, kernel type and the degree of the polynomial.

Section 3 discusses the difficulties involved in experimentation in the study of the effects of maternal smoking in birth weight, which justifies the need of careful studies using non-experimental data, even with the assumption of selection on observables. In this context, an IV-free test of endogeneity is of particular interest. Almond et al. (2005) is to the author's knowledge the most exhaustive of these studies, and in section 3 the partially linear version of the discontinuity test is applied to the most complete specification of that paper.[3] For all bandwidths, the discontinuity test shows strong evidence of endogeneity in the birth weight equation at the 95% confidence level. In the equation of the probability of LBW, Almond et al. (2005)'s specification is weakly rejected at the 95% confidence level when the optimal bandwidth according to the cross-validation method is used, and not rejected at the 99% confidence level. For all the other bandwidths, their specification is not rejected at the 95% confidence level. If this is taken as evidence of none or low endogeneity in the probability of LBW equation, Almond et al. (2005)'s specification can be used to estimate how the probability of low birth weight is affected by smoking each additional daily cigarette, that is, the effect of smoking one cigarette versus none, two versus one, and so on.[4]

2 The modeling framework and the test statistic

Notation 1. *Throughout the paper, $\mathbb{P}$ refers to the probability function defined for events in a probability space. For example, $\mathbb{P}(u \leqslant c)$ is the probability of $u \leqslant c$. $F(c) = \mathbb{P}(u \leqslant c)$ refers to the cumulative distribution function, and $dF = F(x + dx) - F(x)$, which is the probability density function when F is differentiable. If $u = (x, z)$ is multivariate, define $\mathcal{Z}_x$ as the closure of the set $\{z\,;\, dF(x, z) > 0\}$ (i.e. $\mathcal{Z}_x$ is the support of $dF(z \mid x)$), $F(x) = \int_{-\infty}^{x} \int dF(x, z)$ the marginal distribution of x, and*

[3] It should be noted that the specification used in this paper is the same as in Almond et al. (2005) with respect to the covariates. A crucial difference in the approaches is that Almond et al. (2005)'s main explanatory variable is a binary variable for whether the mother smoked during pregnancy, while in the application in this paper the main explanatory variable is the daily number of cigarettes smoked, which is allowed to enter non parametrically in the structural equation.

[4] See Cattaneo (2009) for a study of the effects on birth weight of smoking 1-5 cigarettes versus none, 6-10 cigarettes versus 1-5 and so on, using the same specification as Almond et al. (2005) under the selection on observables assumption.

$\mathcal{X}$ is the closure of the set $\{x\,;\,dF(x) > 0\}$ (i.e. it is the support of $dF(x)$). If $u = (u_1, \ldots, u_l)$, $||u|| = \sqrt{u_1^2 + \cdots + u_l^2}$, and u^T is the transpose of u. $\mathbb{E}$ denotes the expectation operator, $\mathbb{C}ov(u, u') = \mathbb{E}(uu'^T) - \mathbb{E}(u)\mathbb{E}(u')^T$, and $\mathbb{V}ar(u) = \mathbb{C}ov(u, u)$ when these expectations exist. For derivatives, $g^{(r)}(u)$ means the r^{th}-derivative of the function g with respect to u, which can also be expressed as $\frac{d^r}{du^r}g(u)$. If g is a vector, the derivative of the vector is the vector of the derivatives. The notation $u \uparrow c$ implies that u converges to c and $u > c$, and $u \uparrow c$ means the same when $u > c$. The notation $u \sim \mathcal{N}(\mu, v)$ means that the random variable u is distributed as a gaussian with mean μ and variance v. If $v = 0$, then $u = \mu$ with probability one. Where omitted, assume all written moments exist.

2.1 A simple model

This section presents an (informal) account of the main identification idea in a model with no covariates. The formal exposition in the next section will of course account for covariates and a more complex model structure. To help make the exposition clearer, an example where birthweight is modeled as a function of maternal smoking is carried over in this section as the motivation for the model.

Let birth weight be represented by the random variable y, and maternal smoking by the continuous random variable x. Smoking and birth weight have a structural relationship expressed in the model

$$y = f(x, q) + \varepsilon,$$

where q and ε are unobservable, ε is independent of both x and q, and f is continuous in x. The interest is to determine whether birth weight is also affected by a variable q that is dependent of x. To keep this preliminary section on an intuitive level, q will be referred to as the mother's "type." The mother type can be related to the level of smoking and the birth weight of her offspring, but assume that the relation between smoking and the type is of a discontinuous nature, more specifically that the distribution of the type conditional on the level of smoking be discontinuous at zero and continuous for positive levels of smoking. In more mundane words, assume that the mothers that didn't smoke are of discontinuously different types from the mothers that smoke positive amounts, even if small. This condition can be understood within the context of censored variables: suppose there are types of mothers that would choose negative amounts of smoking, but the restriction to non negative values would force this group to smoke only zero. This way of understanding why the unobservables would be discontinuously different at zero cigarettes may be helpful, but it is not necessary. Many justifications can be offered for this phenomenon, and in the application section (3) it is shown that the discontinuity at zero cigarettes in fact exists for several observable covariates. What remains of this section will model smoking as a real-valued censored variable, and its relation to the mother type, q, as continuous and invertible (and therefore monotonic), which will make all the equations explicit.

Suppose that mothers could choose how much to smoke in positive and negative levels. This latent variable is expressed by x^*, so that $x = \max\{0, x^*\}$, and it is related to the mother type, q, through the function g, which is continuous and invertible (and therefore monotonic), hence

$$x^* = g(q)$$

which is to say that there is a one-to-one relation between the mother type and how much she would smoke in case she could choose any amount in the real line. Assume that a mother who would choose to smoke zero or a positive amount will actually do so, and a mother who would chose to smoke negative amounts will smoke zero. Suppose g is decreasing, so higher types smoke less. Then, mothers who smoked a positive amount x are in fact of the type $g^{-1}(x)$, and the mothers that did not smoke are of the types $Q(0) = \{q; q \geqslant g^{-1}(0)\}$. If the set of mothers who would choose to smoke strictly negative amounts has positive probability, then the expected type of the mothers that did not smoke is higher than $g^{-1}(0)$, and therefore, there is a discontinuity in the distribution of the types conditional on the smoking level at zero cigarettes.

This discontinuity can be used to gauge whether birth weight is affected by the mother type. The expected birth weight conditional on the smoking level is given by

$$\mathbb{E}(y \,|\, x) = \begin{cases} f(x, g^{-1}(x)), & \text{if } x > 0 \\ \mathbb{E}\left(f(0, g^{-1}(x^*)) \mid x^* \leqslant 0\right), & \text{if } x = 0 \end{cases}$$

If the mother's type does not affect birth weight (i.e. q is independent of y), then f is constant in q with probability one, and therefore $\mathbb{E}(y \mid x) = f(x, g^{-1}(0))$. Hence, since f is continuous in x, $\mathbb{E}(y \mid x)$ will be continuous in x at $x = 0$. In other words, if birth weight is a continuous function of smoking, a discontinuity at zero is reflecting the presence of something else that affects birth weight and is discontinuous at zero, an unobservable variable correlated with smoking which here is called the mother's type. In that case, x is endogenous. Moreover, the test derives its power from the fact that

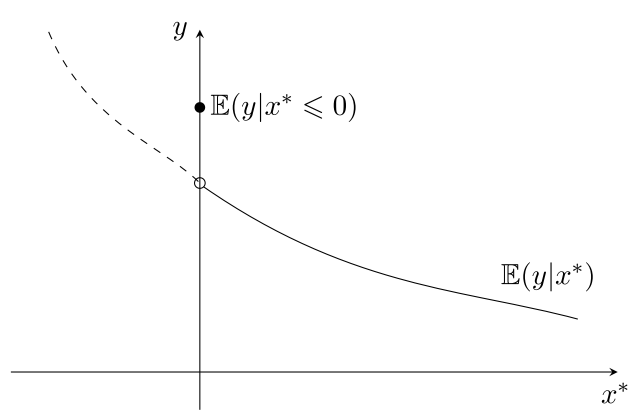


Figure 1: Here, f varies in q, and hence $\mathbb{E}(y \mid x) = \mathbb{E}(y \mid x^*) = f(x^*, g^{-1}(x^*))$ when $x > 0$, but $\mathbb{E}(y \mid x = 0) = \mathbb{E}(f(0, x^* \mid x^* \leqslant 0) < f(0, g^{-1}(0))$, and hence $\mathbb{E}(y \mid x)$ is discontinuous at $x = 0$.

the more birth weight is affected by the mother's type, the larger the discontinuity of $\mathbb{E}(y \mid x)$ at $x = 0$ (see figure 1).

The discontinuity test of endogeneity consists of estimating the discontinuity of $\mathbb{E}(y \mid x)$ at zero and testing whether it is significant. If covariates are added to the model, the discontinuity in figure 1 would exist for the expected birth weight conditional on cigarette number for each value of the covariates. The discontinuity test in this more complex context requires that these discontinuities be aggregated somehow, which is done in the following section.

The fundamental assumption for identification of the test is that birthweight is a continuous function of smoking (f is continuous in x). For the test to have power, the two fundamental assumptions are that there are mothers of the highest types ($\mathbb{P}(q > g^{-1}(0)) > 0$, or $\mathbb{P}(x^* < 0) > 0$) and that the mothers that did not smoke are of discontinuously different types than the mothers that smoked ($\mathbb{E}(q \mid x)$ is discontinuous at $x = 0$). The driving assumptions in the general case with covariates are similar to these.

2.2 Identification

Assumption 2.1. *Let x be a continuous observable random variable $\mathcal{X} \subset \mathbb{R}$ such that $\bar{x} \in \mathcal{X}$. Let z be a vector of observable random variables, and q be a scalar unobservable variable. Then*

1. $\mathbb{E}(y \,|\, x, z, q)$ *is continuous in x at $x = \bar{x}$ for all the values of z and q.*
2. $\lim_{x \downarrow \bar{x}} dF(q \,|\, x, z)$ *and* $\lim_{x \uparrow \bar{x}} dF(q \,|\, x, z)$ *exist for all the values of z.*
3. *There exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that $\mathcal{N} \subset \mathcal{X}$. Also, the sets $\mathcal{Z}_x$ are identical, $\forall x \in \mathcal{N}$.*

Define the quantity

$$\Delta(z) = \mathbb{E}(y \,|\, x = \bar{x}, z) - \left(\alpha \lim_{x \downarrow \bar{x}} \mathbb{E}(y \,|\, x, z) + (1 - \alpha) \lim_{x \uparrow \bar{x}} \mathbb{E}(y \,|\, x, z)\right),$$

where if $\bar{x}$ is a lower boundary point in $\mathcal{X}$ (as in the maternal smoking example), $\alpha = 1$, and if $\bar{x}$ is at the upper boundary, $\alpha = 0$. $\Delta(z)$ is the weighted right and left discontinuity of $\mathbb{E}(y \,|\, x, z)$ at $x = \bar{x}$. Let θ be the aggregation of the $\Delta(z)$ defined as

$$\theta = \int G(\Delta(z), z) d\nu(z) \tag{1}$$

for a known function G and a measure ν on the range of the z.

Definition 1. *Let y be the dependent variable, and x and z be observable explanatory variables. Then x is said to be exogenous if*

$$\mathbb{P}(\mathbb{E}(y \,|\, x, z, q) = \mathbb{E}(y \,|\, x, z)) = 1.$$

for any variable q such that q and x are not independent conditional on z. Otherwise, x is endogenous.

This paper presents a test of the null hypothesis H_0, that x is exogenous, against the alternative hypothesis, H_1, that x is endogenous. The test will depend on the estimation of the parameter θ. The following result states that x exogenous implies $\theta = 0$. This is a fundamental result in establishing that a test based on θ is well defined, in the sense that it has the correct asymptotic size under H_0.

Theorem 1. *If ν is identified, $G(0, z) = 0$, $\forall z$ and assumptions 2.1 and A.1 (see remark 2.2.2) are satisfied, then θ is identified and is equal to zero if x is exogenous.*

The proof is in appendix A.1.1. Section 2.3 shows that θ is in fact estimable (pending more conditions), and its estimator is the discontinuity test statistic. An example of an identified ν is the distribution of z, which yields $\theta = \mathbb{E}(\Delta(z), z))$. Another possibility would be to measure the average square of the discontinuities, $\theta = \mathbb{E}(\Delta(z)^2)$, using $G(a, b) = a^2$.

Of particular interest is the case when $G(a, b) = a\, g(b)$, for a given real valued function g in the domain of z, and $\nu(z) = F(z \,|\, x = \bar{x})$. In this case,

$$\begin{aligned}\theta =& \mathbb{E}(\Delta(z) g(z) \,|\, x = \bar{x}) \\ =& \mathbb{E}(y\, g(z) \,|\, x = \bar{x}) - \\ & - \left[\alpha E \left(\lim_{x \downarrow \bar{x}} \mathbb{E}(y \,|\, x, z)\, g(z) \,\middle|\, x = \bar{x} \right) + (1 - \alpha) E \left(\lim_{x \uparrow \bar{x}} \mathbb{E}(y \,|\, x, z)\, g(z) \,\middle|\, x = \bar{x} \right) \right],\end{aligned} \tag{2}$$

because of the law of iterated expectations. This parameter is useful because its estimation does not require the estimation of $\mathbb{E}(y \,|\, x = \bar{x}, z)$, and because $\mathbb{E}(y\, g(z) \,|\, x = \bar{x})$ can be estimated at the rate $\sqrt{n}$ if $P(x = \bar{x}) > 0$.

The following assumption determines in which cases the discontinuity has power.

Assumption 2.2. *Let $\lim_{x \downarrow \bar{x}} dF(q \,|\, x, z)$ exist if $\bar{x}$ is an interior point or the left boundary point in $\mathcal{X}$, and $\lim_{x \uparrow \bar{x}} dF(q \,|\, x, z)$ exist if $\bar{x}$ is an interior point or the right boundary point in $\mathcal{X}$. If $\bar{x} \leqslant x$, $\forall x \in \mathcal{X}$, let $\alpha = 1$, and if $\bar{x} \geqslant x$, $\forall x \in \mathcal{X}$, then $\alpha = 0$. Define*

$$\zeta(q, z) := dF(q \,|\, x = \bar{x}, z) - \left(\alpha \lim_{x \downarrow \bar{x}} dF(q \,|\, x, z) + (1 - \alpha) \lim_{x \uparrow \bar{x}} dF(q \,|\, x, z) \right), \quad \text{for } \alpha \in [0, 1].$$

Then, $\mathbb{P}(\zeta(q, z) \neq 0 \,|\, x) > 0$, for all the values of x in a neighborhood of $\bar{x}$.

Assumption 2.2 implies that $dF(q \,|\, x, z)$ discontinuous at $\bar{x}$. The discontinuity may be different from the right or left hand side, or even exist in only one of the sides. The assumption also stipulates that the discontinuity is one sided if $\bar{x}$ is at the boundary of $\mathcal{X}$. If $\bar{x}$ is an interior point, previous knowledge of the process can be used to choose α more effectively. If the right and left limits of $dF(q \,|\, x, z)$ are the same, the choice of α is irrelevant, and $\zeta(q, z) := dF(q \,|\, x = \bar{x}, z) - \lim_{x \to \bar{x}} dF(q \,|\, x, z)$.

Assumption 2.2 defines $\bar{x}$ implicitly. In other words, the situations where the discontinuity test has power are those where a value $\bar{x}$ in $\mathcal{X}$ can be found such that assumption 2.2 is believed to hold. The following result will be useful in proving that the asymptotic power of the test converges to one under H_1 as the sample size increases.

Result 1. *If assumptions 2.1, 2.2 and A.1 hold, then in general, if x is endogenous, $\theta \neq 0$.*

It will be said that the result is true whenever x endogenous implies $\theta \neq 0$. To understand this result, observe that by assumption A.1,

$$\begin{aligned}
\theta = \int G\Big(\Big[\int \mathbb{E}(y \,|\, x = \bar{x}, z, q) dF(q \,|\, x = \bar{x}, z) - \\
&- \alpha \int \mathbb{E}(y \,|\, x = \bar{x}, z, q) \lim_{x \downarrow \bar{x}} dF(q \,|\, x, z) \\
&- (1-\alpha) \int \mathbb{E}(y \,|\, x = \bar{x}, z, q) \lim_{x \uparrow \bar{x}} dF(q \,|\, x, z)\Big], z\Big) d\nu(z) \\
&= \int G\Big(\int \mathbb{E}(y \,|\, x = \bar{x}, z, q) \zeta(q, z), z\Big) d\nu(z)
\end{aligned} \tag{3}$$

Assumption 2.2 guarantees that $\zeta(q, z) \neq 0$ with positive probability. However, it cannot be guaranteed that $\theta \neq 0$ unless stronger requirements are made in all G, ν, $\mathbb{E}(y \mid x, z)$ and even the shape of $\xi(q, z)$. Such requirements can easily be made, and section 2.2.1 presents an example where primitive conditions are derived such that result 1 holds always instead of only in general. This paper refrains from presenting such conditions in the general case because they would be unnecessarily restrictive. In other words, the cases where $\theta = 0$ even though x is endogenous have zero measure in the functional spaces where the estimators will be defined, and are of no concern. If such a possibility is feared, different choices of G and ν should be attempted. Hence, the results concerning the power of the test will hold if result 1 holds.

Remark 2.2.1. *The discontinuity test could simply consist on the estimation of $\Delta(z)$ for some value of z and then testing whether it is equal to zero. However, such a test may have little power because $\Delta(z)$ can in general be estimated only at very low rates of convergence. In the interest of the accuracy of the estimation, and to avoid the problem that a wrong choice of z could occasion, it is preferable to aggregate the discontinuities somehow, and from this derives the interest of the parameter θ.*

Remark 2.2.2. *Condition A.1 in appendix A.1.1 requires the interchangeability of the integral and the limits in the following non-trivial specification. Let a sequence $x_n \downarrow \bar{x}$, define $f_n(q) = \mathbb{E}(y \,|\, x_n, z, q)$, $f(q) = \mathbb{E}(y \,|\, \bar{x}, z, q)$, $\mu_n(q) = F(q \,|\, x_n, z)$ and $\mu(q) = d \lim_{n\to\infty} F(q \,|\, x_n, z)$. By assumption 2.1 (1), $f_n \to f$ pointwise on q. Observe that by the definition of the Riemann-Stieltjes integral,*

$$\lim_{n\to\infty} \int f_n \, d\mu_n = \lim_{n\to\infty} \lim_{\Delta q \to 0} \sum f_n(q^c) \mu_n(\Delta q),$$

where q^c is any point in the intervals of size Δq. Then, assumption A.1 can be expressed as $\lim_{n\to\infty} \int f_n \, d\mu_n = \int f \, d\mu$. Though primitive conditions for this are not *specified here, they can be established with measure theory convergence theorems, and by changing the order of the limits and requiring that the support of $dF(q)$ be compact.*

Remark 2.2.3. *The parameter used in the case with no covariates z cannot be used in the case where the z are present. In that case, $\mathbb{E}(y \mid x = \bar{x})$ is compared with the limit $\lim_{x \to \bar{x}} \mathbb{E}(y \mid x)$. Observe that the parameter θ controls the distribution of z, because it uses the fixed measure ν to weight the different z. In the simple comparison of $\lim_{x \to \bar{x}} \mathbb{E}(y \mid x)$ and $\mathbb{E}(y \mid x = \bar{x})$, the distribution of z, which is often discontinuous at $x = \bar{x}$ can be responsible for a difference even when x is exogenous. To see this, notice that if x is exogenous, $\mathbb{E}(y \mid x, z, q) = \mathbb{E}(y \mid x, z)$, and provided the limit can exchange places with the integral sign,*

$$\begin{aligned}
\lim_{x \to \bar{x}} \mathbb{E}(y \mid x) &= \lim_{x \to \bar{x}} \int \mathbb{E}(y \mid x, z) dF(z \mid x) \\
&= \int \mathbb{E}(y \mid \bar{x}, z) \lim_{x \to \bar{x}} dF(z \mid x)
\end{aligned}$$

and

$$\mathbb{E}(y \mid x = \bar{x}) = \int \mathbb{E}(y \mid \bar{x}, z)\, dF(z \mid x = \bar{x})$$

Since $\lim_{x \to \bar{x}} dF(z \mid x)$ and $dF(z \mid x = \bar{x})$ can be and often are different, $\mathbb{E}(y \mid x)$ can be discontinuous at $\bar{x}$ even when x is exogenous, and therefore this comparison is useless for the detection of endogeneity.

Remark 2.2.4. *Generalization for multivariate q is straightforward. It is enough to understand $dF(q \mid x, z)$ as a multivariate probability distribution, and the limits $\lim_{x \downarrow \bar{x}} dF(q \mid x, z)$ and $\lim_{x \uparrow \bar{x}} dF(q \mid x, z)$ as multivariate limits. All conditions and theorems remain the same.*

Remark 2.2.5. *Theorem 1 allows for random variables to enter the model in very flexible ways. Suppose*

$$y = f(x, z, q, \varepsilon)$$

where ε is independent of x, z and q. Provided f is continuous in x at $\bar{x}$, it is easy to show that $\mathbb{E}(y \mid x, z, q)$ will also be continuous at $\bar{x}$. See proof of this in appendix A.1.2.

2.2.1 An example in censoring

Primitive conditions for assumptions 2.1 and 2.2 can be established in a more intuitive setup inside a model where x is censored. Such situations could naturally develop, for example, when x is the result of a cornered optimization problem. This is the case in the smoking example, where smoking is a choice variable that cannot be chosen in negative values. This model and the suggested assumptions are not the weakest for identification of θ; they just illustrate the point in an intuitive way.

Suppose an unobservable variable x^* is only observed in its censored form $x = \max\{x^*, 0\}$, ε is independent of the variables x, z and q, and the variables y, x, x^*, z,

q and ε are related in the structural equations

$$y = f_1(x, z, q) + \varepsilon \quad \text{and} \quad x^* = f_2(z, q).$$

Then

$$\implies \mathbb{E}(y \mid x, z) = \begin{cases} f_1(x, z, f_2^{-1}(x; z)), & \text{if } x > 0, \\ \mathbb{E}(f_1(0, z, f_2^{-1}(x^*; z)) \mid x^* \leqslant 0, z), & \text{if } x = 0. \end{cases} \tag{4}$$

In this model, instead of assumptions 2.1 and 2.2, consider the following assumption:

Assumption 2.3.

1. *f_1 is continuous in x at 0, and if f_1 varies on q, it is continuous and increasing in q.*
2. *f_2 is strictly decreasing in x^*, and $f_2(\cdot; z)^{-1}$ is continuous in x^*, $\forall z$.*
3. *$F(x^* \mid z) > 0$, for some value $x^* < 0$, $\forall z$.*

Given the model and assumption 2.3,

$$\Delta(z) = \mathbb{E}(f_1(0, z, f_2^{-1}(x^*; z)) \mid x^* \leqslant 0, z) - f_1(0, z_i, f_2^{-1}(0, z_i)) > 0$$

if and only if f_1 varies in q. Hence, if $G(\Delta(z), z) = \Delta(z)g(z)$, for a strictly positive function g, and suppose ν is not zero everywhere, then $\theta = \int \Delta(z)g(z)\, d\nu(z) > 0$ if and only if x is endogenous.

In the smoking example, if birthweight is y, smoking is x, x^* is "intended" smoking, and the z are a set of covariates, assumption 2.3 implies that even if the covariates are held constant, the average birthweight of babies born to nonsmoker mothers will be discontinuously higher than the birthweight of babies born to mothers that smoked positive amounts if and only if smoking is endogenous.

2.3 The discontinuity test of endogeneity

The discontinuity test consists of the estimation of θ and testing whether it is equal to zero. A natural approach would be to adopt

$$\hat{\theta} = \int G(\hat{\Delta}(z), z) d\hat{\nu}(z),$$

where $\hat{\Delta}(z) = \hat{\mathbb{E}}(y \mid x = \bar{x}, z) - \left(\alpha \hat{\mathbb{E}}(y \mid \bar{x}, z)^{\downarrow} + (1 - \alpha)\hat{\mathbb{E}}(y \mid \bar{x}, z)^{\uparrow}\right)$, $\hat{\mathbb{E}}(y \mid \bar{x}, z)^{\downarrow}$ is an estimator of $\lim_{x \downarrow \bar{x}} \mathbb{E}(y \mid x, z)$ and $\hat{\mathbb{E}}(y \mid \bar{x}, z)^{\uparrow}$ is an estimator of $\lim_{x \uparrow \bar{x}} \mathbb{E}(y \mid x, z)$.

As explained in section 2.2, the tests where $G(\Delta(z), z) = \Delta(z)\, g(z)$ and $\mu(z) = F(z \mid x = \bar{x})$ are of particular interest, because they eliminate one step in the estimation of θ. The rest of this section will develop the discontinuity test for such choices of G and μ. Let the data satisfy

Assumption 2.4.

1. *The observations* (y_i, x_i, z_i), $i = 1, ..., n$ *are i.i.d.*, $z_i = (z_i^1, \ldots, z_i^d)^T$. *Define* $\epsilon_i = y_i - \mathbb{E}(y_i \mid x_i, z_i)$.
2. $0 < \mathbb{P}(x_i = \bar{x}) < 1$.
3. $\mathbb{E}(|\Delta(z_i)\, g(z_i)|^{2+\xi_1} \mathbf{1}(x_i = \bar{x})) < \infty$ *for some* $\xi_1 > 0$. $V_A := \mathbb{V}ar(\Delta(z_i)\, g(z_i) \mid x_i = \bar{x})$.

Define $\hat{p}_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(x_i = \bar{x})$, the estimator of $P(x_i = \bar{x})$. The suggested estimator of θ is, from equation (2),

$$\hat{\theta} = \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - \alpha \hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\downarrow} - (1-\alpha)\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\uparrow} \right] g(z_i)\, \mathbf{1}(x_i = \bar{x}). \tag{5}$$

Define $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\downarrow} := \lim_{x \downarrow \bar{x}} \mathbb{E}(y_i \mid x_i = x, z_i)$, $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\uparrow} := \lim_{x \uparrow \bar{x}} \mathbb{E}(y_i \mid x_i = x, z_i)$, and also define $\hat{\Gamma}(z)^{+} := \hat{\mathbb{E}}(y \,|\, \bar{x}, z)^{\downarrow} - \mathbb{E}(y \,|\, \bar{x}, z)^{\downarrow}$, and $\hat{\Gamma}(z)^{-} := \hat{\mathbb{E}}(y \,|\, \bar{x}, z)^{\uparrow} - \mathbb{E}(y \,|\, \bar{x}, z)^{\uparrow}$. then it is possible to write

$$\hat{\theta} - \theta = A_n - B_n$$

where

$$A_n = \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^{n} \Delta(z_i) g(z_i) \mathbf{1}(x_i = \bar{x}) - \mathbb{E}(\Delta(z_i) g(z_i) \,|\, x_i = \bar{x}) \tag{6}$$

$$B_n = \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^{n} [\alpha \hat{\Gamma}(z_i)^{+} + (1-\alpha)\hat{\Gamma}(z_i)^{-}] g(z_i) \mathbf{1}(x_i = \bar{x}). \tag{7}$$

Under the null hypothesis that x_i is exogenous, $\Delta(z_i) = 0$, and therefore $A_n = 0$, though results are shown when $A_n \neq 0$ for power consideriations. The asymptotic distribution of A_n does not depend on the choice of estimators. Assumption 2.4 item (1) and the LLN imply that $\hat{p}_{\bar{x}} \xrightarrow{p} \mathbb{P}(x_i = \bar{x})$, and items (1) and (3) and the CLT imply that $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \Delta(z_i) g(z_i) \mathbf{1}(x_i = \bar{x}) - \mathbb{E}(\Delta(z_i) g(z_i) \mathbf{1}(x_i = \bar{x})\right)$ is asymptotically normally distributed. Finally, item (2), the Continuous Mapping theorem and Slutsky's theorem imply that

$$\sqrt{n} A_n \xrightarrow{d} \mathcal{N}(0, V_A) \tag{8}$$

The asymptotic behavior of B_n, and hence of $\hat{\theta}$, depends on the assumptions one is willing to make on the nature of $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\downarrow}$ and $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\uparrow}$, and the related choice of estimators. The following sections propose increasingly complex models of $\mathbb{E}(y_i | x_i, z_i)$ and corresponding estimators, and derive the asymptotic distributions of the test under appropriate assumptions. Section 2.3.1 presents the linear case, the partially linear case is developed in section 2.3.2, and finally section 2.3.3 presents the fully non-parametric case. The three sections are written so they stand alone. No assumption or result is shared across the sections, and therefore they can be read and consulted independently.

2.3.1 The linear case

Suppose that for $x > \bar{x}$, the conditional expectation satisfies

$$\mathbb{E}(y \,|\, x, z) = \beta^+ x + z^T \gamma^+, \tag{9}$$

and for $x < \bar{x}$, the conditional expectation satisfies

$$\mathbb{E}(y \,|\, x, z) = \beta^- x + z^T \gamma^-.$$

If $\bar{x}$ is the left boundary of $\mathcal{X}$, then $\beta^- = 0$ and $\gamma^- = 0$. If $\bar{x}$ is the right boundary of $\mathcal{X}$, then $\beta^+ = 0$ and $\gamma^+ = 0$.

Example 1. *(Censoring) Equation (9) can be derived inside the censoring model presented in section 2.2.1. Suppose*

$$\begin{aligned} f(x, z, q) &= \alpha_x x + z^T \alpha_z + \alpha_q q \\ g(z, q) &= z^T \pi_z + q \end{aligned}$$

then, substituting into equation (4) for $x > 0$,

$$\mathbb{E}(y \,|\, x, z) = (\alpha_x + \alpha_q)\, x + z^T (\alpha_z - \alpha_q \pi_z),$$

which translates into equation (9) if $\beta^+ := \alpha_x + \alpha_q$ and $\gamma^+ := \alpha_z - \alpha_q \pi_z$.

In the linear case, $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = \beta^+ \bar{x} + z_i^T \gamma^+$, and $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\uparrow} = \beta^- \bar{x} + z_i^T \gamma^-$. β^+ and γ^+ can be estimated by simply regressing y_i on x_i and z_i using only the observations for which $x_i > \bar{x}$, so that $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = \hat{\beta}^+ \bar{x} + z_i^T \hat{\gamma}^+$, and analogously for $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\uparrow}$.

Let $X_i = (x_i, z_i^T)^T$, $\delta^+ = (\beta^+, \gamma^{+T})^T$ and $\delta^- = (\beta^-, \gamma^{-T})^T$, then if $\bar{x}$ is an interior point,

$$\begin{aligned} \hat{\delta}^+ &= \left(\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i > \bar{x}) \right)^{-1} \sum_{i=1}^n X_i \, y_i \, \mathbf{1}(x_i > \bar{x}), \\ \hat{\delta}^- &= \left(\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i < \bar{x}) \right)^{-1} \sum_{i=1}^n X_i \, y_i \, \mathbf{1}(x_i < \bar{x}). \end{aligned}$$

If $\bar{x}$ is the left boundary of the $\mathcal{X}$, then $\hat{\delta}^- = 0$. If $\bar{x}$ is the right boundary of the $\mathcal{X}$, then $\hat{\delta}^+ = 0$. Let $\hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x}) := \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i = \bar{x}) g(z_i)$ and $\hat{\mathbb{E}}(g(z_i) z_i \mid x_i = \bar{x}) := \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i = \bar{x}) g(z_i) z_i$, then from equation (7),

$$B_n = \begin{bmatrix} \hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})\, \bar{x} \\ \hat{\mathbb{E}}(g(z_i) z_i \mid x_i = \bar{x}) \end{bmatrix}^T \Big(\alpha (\hat{\delta}^+ - \delta^+) + (1 - \alpha)(\hat{\delta}^- - \delta^-) \Big)$$

Assumption 2.5.

1. $\mathbb{E}(|g(z_i)| \mid x_i = \bar{x}) < \infty$ *and* $\mathbb{E}(||g(z_i)z_i|| \mid x_i = \bar{x}) < \infty$.
2. $\mathbb{V}ar(\epsilon_i \mid x_i, z_i) = \sigma^2 < \infty$. *(See below remark 2.3.1 about relaxing this condition.)*
3. $\mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x})) < \infty$ *is positive definite, and* $\mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x})) < \infty$ *is positive definite.*
4. *If* $\bar{x}$ *is the left boundary of* $\mathcal{X}$*, then then* $\alpha = 1$*, and if* $\bar{x}$ *is the right boundary of* $\mathcal{X}$*, then* $\alpha = 0$.

Theorem 2. *If assumptions 2.1, 2.4 and 2.5 hold, then*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V_A + V_B) \tag{10}$$

where

$$V_B = \sigma^2 \begin{bmatrix} \mathbb{E}(g(z_i) \mid x_i = \bar{x})\, \bar{x} \\ \mathbb{E}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix}^T \big[\alpha^2 \mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))^{-1} + \\ + (1-\alpha)^2 \mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))^{-1}\big] \begin{bmatrix} \mathbb{E}(g(z_i) \mid x_i = \bar{x})\, \bar{x} \\ \mathbb{E}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix}.$$

The proof is similar to the classical proofs of the asymptotic properties of the OLS estimator. The absence of a term to account for the correlation of A_n and B_n follows because $\big(\alpha(\hat{\delta}^+ - \delta^+) + (1-\alpha)(\hat{\delta}^- - \delta^-)\big)$ is independent of A_n and of $\hat{\mathbb{E}}(g(z_i)z_i \mid x_i = \bar{x})$, since the two latter only use observations for which $x_i = \bar{x}$, while the former has zero mean and only uses observations for which $x_i \neq \bar{x}$. The absence of a cross term in V_B happens because $\hat{\delta}^+$ and $\hat{\delta}^+$ are built using different parts of the sample, and are therefore independent. See the proof in detail in the appendix A.2.1.

Theorem 3. *Under* H_0*:* x_i *is exogenous,* $\theta = 0$ *and* $\sqrt{n}\hat{\theta} \xrightarrow{d} \mathcal{N}(0, V_B)$*. If assumptions 2.1, 2.4 and 2.5 hold,* V_B *can be consistently estimated by*

$$\hat{V}_B = \hat{\sigma}^2 \begin{bmatrix} \hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})\, \bar{x} \\ \hat{\mathbb{E}}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix}^T \big[\alpha^2 \hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))^{-1} + \\ + (1-\alpha)^2 \hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))^{-1}\big] \begin{bmatrix} \hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})\, \bar{x} \\ \hat{\mathbb{E}}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix},$$

where

$$\hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i > \bar{x})) = \frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i > \bar{x}),$$

$$\hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i < \bar{x})) = \frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i < \bar{x}),$$

$$\hat{\sigma}^2 = \frac{1}{1 - \hat{p}_{\bar{x}}}\left[\alpha \frac{1}{n}\sum_{i=1}^n (y_i - X_i^T \hat{\gamma}^+)^2 \mathbf{1}(x_i > \bar{x}) + (1-\alpha)\frac{1}{n}\sum_{i=1}^n (y_i - X_i^T \hat{\gamma}^-)^2 \mathbf{1}(x_i < \bar{x})\right].$$

The convergence in probability of $\hat{\sigma}^2$ to σ^2 is established by noticing that $\hat{\sigma}^2$ is simply a weighted average of two standard estimators of the variance of ϵ_i using weighted least squares. The convergence of $\hat{V}_B$ follows from the LLN applied to $\hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})$, $\hat{\mathbb{E}}(g(z_i)z_i \mid x_i = \bar{x})$, $\hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))$ and $\hat{\mathbb{E}}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))$ (given assumptions 2.4 (1) and 2.5 (1) and (3)) and Slutsky's theorem. The following theorem gives the discontinuity test properties in the linear case.

Theorem 4. *Let $0 \leqslant \lambda \leqslant 1$, Φ be the standard normal cumulative distribution function, and $c_\lambda = \Phi^{-1}(\lambda)$. Then if theorems 1, 2 and 3 hold, under H_0: x is exogenous,*

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\sqrt{\hat{V}_B}} \leqslant c_\lambda\right) \to \lambda \quad \textit{as} \quad n \to \infty.$$

moreover, if result 1 is true, then under H_1: x is endogenous,

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\sqrt{\hat{V}_B}} > c_\lambda\right) \to 1 \quad \textit{as} \quad n \to \infty,$$

and under the local alternatives $\frac{\theta}{\sqrt{n}}$,

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\sqrt{\hat{V}_B}} \leqslant c_\lambda\right) \to \Phi\left(\frac{c_\lambda\sqrt{V_B} - \theta}{\sqrt{V_A + V_B}}\right) \quad \textit{as} \quad n \to \infty.$$

See proof in appendix A.2.2.

Remark 2.3.1. *Homoskedasticity can easily be relaxed. Let X be the matrix whose rows are the X_i^T, X^+ be the matrix whose rows are the $\mathbf{1}(x_i > \bar{x})X_i^T$, and X^- be the matrix whose rows are the $\mathbf{1}(x_i < \bar{x})X_i^T$. Suppose $\mathbb{V}ar(\epsilon \mid X) = \Sigma$, then*

$$V_B = \begin{bmatrix} \mathbb{E}(g(z_i) \mid x_i = \bar{x})\,\bar{x} \\ \mathbb{E}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix}^T \left[\alpha^2 V_1 + (1-\alpha)^2 V_2\right] \begin{bmatrix} \mathbb{E}(g(z_i) \mid x_i = \bar{x})\,\bar{x} \\ \mathbb{E}(g(z_i)z_i \mid x_i = \bar{x}) \end{bmatrix},$$

where

$$V_1 = \mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))^{-1} \mathbb{E}\left(\underset{n\to\infty}{plim} \frac{X^{+T}\Sigma X^+}{n}\right) \mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))^{-1},$$
$$V_2 = \mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))^{-1} \mathbb{E}\left(\underset{n\to\infty}{plim} \frac{X^{-T}\Sigma X^-}{n}\right) \mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))^{-1},$$

and $\underset{n\to\infty}{plim}$ denotes the limit in probability, supposing the limits exist. V_1 can be estimated using the Eicker-White covariance matrix of an OLS regression of the y_i onto x_i and z_i using only observations such that $x_i > \bar{x}$, and V_2 can be estimated analogously, using only observations such that $x_i < \bar{x}$.

2.3.2 The partially linear case

Suppose that for $x > \bar{x}$, the conditional expectation satisfies

$$\mathbb{E}(y_i \,|\, x_i, z_i) = \tau^+(x_i) + z_i^T \gamma^+, \tag{11}$$

and for $x < \bar{x}$, the conditional expectation satisfies

$$\mathbb{E}(y_i \,|\, x_i, z_i) = \tau^-(x_i) + z_i^T \gamma^-,$$

where $\tau^+(\bar{x})^{\downarrow} := \lim_{x \downarrow \bar{x}} \tau^+(x)$ and $\tau^-(\bar{x})^{\uparrow} := \lim_{x \uparrow \bar{x}} \tau^-(x)$ exist. If $\bar{x}$ is the left boundary of $\mathcal{X}$, then $\tau^-(x_i) = 0$ for all x_i and $\gamma^- = 0$. If $\bar{x}$ is the right boundary of $\mathcal{X}$, then $\tau^+(x_i) = 0$ for all x_i and $\gamma^+ = 0$.

Example 2. *(Censoring) Equation (11) can be derived inside the censoring model presented in section 2.2.1. Suppose*

$$\begin{aligned} f(x, z, q) &= \psi_1(x) + z'\alpha_z + \alpha_q q, \\ g(z, q) &= \psi_2(z'\pi_z + q), \end{aligned}$$

where ψ_2 is invertible. Then, substituting into equation (4) for $x > 0$,

$$\mathbb{E}(y \,|\, x, z) = (\psi_1(x) + \alpha_q \psi_2^{-1}(x))\, x + z^T(\alpha_z - \alpha_q \pi_z).$$

which translates into equation (11) if $\tau^+(x) := \psi_1(x) + \alpha_q \psi_2^{-1}(x)\ \forall\, x$, and $\gamma^+ := \alpha_z - \alpha_q \pi_z$.

In the partially linear case, $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = \tau^+(\bar{x})^{\downarrow} + z_i^T \gamma^+$, and $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\uparrow} = \tau^-(\bar{x})^{\uparrow} + z_i^T \gamma^-$. Hence, $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = \hat{\tau}^+(\bar{x})^{\downarrow} + z_i^T \hat{\gamma}^+$, and $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\uparrow} = \hat{\tau}^-(\bar{x})^{\uparrow} + z_i^T \hat{\gamma}^-$. Define $\hat{\mathbb{E}}(g(z_i) \,|\, x_i = \bar{x}) = \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^n g(z_i) \mathbf{1}(x_i = \bar{x})$ and $\hat{\mathbb{E}}(g(z_i) z_i \,|\, x_i = \bar{x}) = \frac{1}{\hat{p}_{\bar{x}}} \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i = \bar{x}) g(z_i) z_i$, then

$$\begin{aligned} B_n = {} & \hat{\mathbb{E}}(g(z_i) \,|\, x_i = \bar{x}) \left[\alpha(\hat{\tau}^+(\bar{x})^{\downarrow} - \tau^+(\bar{x})^{\downarrow}) + (1-\alpha)(\hat{\tau}^-(\bar{x})^{\uparrow} - \tau^-(\bar{x})^{\uparrow})\right] \\ & + \hat{\mathbb{E}}(g(z_i) z_i \,|\, x_i = \bar{x})^T \left[\alpha(\hat{\gamma}^+ - \gamma^+) + (1-\alpha)(\hat{\gamma}^- - \gamma^-)\right] \end{aligned} \tag{12}$$

The following discussion refers to the estimation of $\tau^+(\bar{x})^{\downarrow}$ and γ^+. $\tau^-(\bar{x})^{\uparrow}$ and γ^- are estimated analogously. The estimation of the parametric component in the partially linear regression has been widely discussed in the literature. In the later papers, the generally adopted technique is that of subtracting the conditional expectation of y_i given x_i so as to eliminate the nonparametric part. The resulting equation is

$$y_i - \mathbb{E}(y_i \,|\, x_i) = (z_i - \mathbb{E}(z_i \,|\, x_i))^T \gamma^+ + \epsilon_i, \quad \text{for} \quad x_i > \bar{x}. \tag{13}$$

The coefficient of the constant term among the covariates is not identified and is eliminated in the subtraction, so z_i in this equation does not include a constant term.

Robinson (1988) first suggested this approach. He estimated the conditional expectations using kernel regression, and peformed an OLS regression of $y_i - \hat{\mathbb{E}}(y_i \,|\, x_i)$ on $z_i - \hat{\mathbb{E}}(z_i \,|\, x_i)$, to obtain $\hat{\gamma}^+$. Robinson showed that the estimated $\hat{\gamma}^+$ converges to γ^+ at the rate $\sqrt{n}$, even though the regression includes nonparametric plugins. The following literature established the same $\sqrt{n}$ rate of convergence and the asymptotic distribution of γ^+ for an array of different nonparametric plugins. See for example Linton (1995) when the nonparametric component is estimated using local polynomial regression, and Li (2000) when the nonparametric component is estimated using series or spline orthogonal bases.

The basic technique for the estimation of the nonparametric component is rather intuitive. It consists of a nonparametric regression of $y_i - z_i^T \hat{\gamma}^+$ on x_i, and the variations depend on the nature of $\hat{\gamma}^+$ and the regression technique chosen. Since the rates of convergence of this component are slower than $\sqrt{n}$, the asymptotic behavior of the estimated nonparametric component is a simple extension of the results for regular nonparametric regression, because the estimated parametric component is estimated at the faster rate $\sqrt{n}$. The case of interest in this paper is more delicate, because the value of interest is $\tau^+(\bar{x})^\downarrow$, which is the limit of the nonparametric component at a boundary point. There are two difficulties, the first is that nonparametric estimation at boundary points requires especial attention in the choice of the estimator and in the asymptotic treatment. For this reason $\tau^+(\bar{x})^\downarrow$ is estimated using local polynomial regression, since this technique has been shown to possess excellent boundary properties. Though other techniques could also be used, such as for example a simple kernel regression using boundary kernels, the local polynomial regression is also desirable in that it requires no especial tailoring for the boundaries. Hence, the researcher needs to apply no extra discretion than for a regular nonparametric regression. Porter (2003) developed the asymptotic theory for the local polynomial estimator of the discontinuity in the regression discontinuity design. His method is to estimate the right and left limits of the discontinuous function at the point of discontinuity using local polynomial regression, and he derives results for arbitrary choice of the polynomial degree. This paper provides the extension of his results to the partially linear case, in which the dependent variable in the local polynomial regression, $y_i - z_i^T \hat{\gamma}^+$, contains a plugin estimator of the parametric component. Though from an asymptotic point of view the extension is very simple, this paper explicits the variance terms up to the $O(h)$ magnitude, which requires the careful consideration of the covariances between the parametric and nonparametric parts of the estimation. Moreover, the results are presented for a generic nonparametric plugin for $\hat{\mathbb{E}}(y_i \mid x_i)$ and $\hat{\mathbb{E}}(y_i \mid x_i)$, so that the plugins can be estimated with other, sometimes more practical techniques, such as series estimators.

The second difficulty is that $\bar{x}$ is a point with positive probability in $\mathcal{X}$. The available theory on local polynomial estimators relies on the existence of a density function in a neighborhood of $\bar{x}$. However, when using local polynomial estimators to estimate the limit of a function at a point, the observations at the point itself are not used. In fact, although Porter (2003) requires the existence of a density function, the proofs do not use the entire support of $dF(x)$ at once, but rather separate the observations to the right and to the left of $\bar{x}$. This paper adapts Porter's result using distribution functions conditional on $x_i \neq \bar{x}$, which have a density function by assumption, though with possibly different right and left limits at $\bar{x}$. As a consequence the same results as in Porter (2003) can be derived in terms of limits, therefore generalizing Porter's results to allow both for the positive probability of $x_i = \bar{x}$, and also for the density function of $F(x_i \mid x_i = x)$ to have different right and left limits at $\bar{x}$. It is important to notice that because the limits may be different, the variance estimator suggested by Porter in theorem 4 cannot be used in this case. Theorem 6 below proposes a different estimator which allows for the different right and left limits of the density at $\bar{x}$.

If $\bar{x}$ is an interior point, or is at the left boundary of the $\mathcal{X}$, the estimator $\tau^+(\bar{x})^{\downarrow}$ is defined in the following way. Given the kernel function k, the smoothing parameter h, the polynomial degree p, and let $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p$ be the solution the problem

$$\min_{a_0,\ldots,a_p} \frac{1}{n} \sum_{j=1}^{n} k\left(\frac{x_j - \bar{x}}{h}\right) \mathbf{1}(x_j > \bar{x}) \left[y_j - z_j^T \hat{\gamma}^+ - a_0 - a_1(x_j - \bar{x}) - \ldots - a_p(x_j - \bar{x})^p\right]^2,$$

the local polynomial estimator of $\tau^+(\bar{x})^{\downarrow}$ is given by

$$\hat{\tau}^+(\bar{x})^{\downarrow} = \hat{a}_0 = e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1} \tilde{X}^T W^+ (Y - Z\hat{\gamma}^+), \tag{14}$$

where $e_1 = (1, 0, \ldots, 0)^T$ has dimension $1 \times (p+1)$, $\tilde{X}$ has rows equal to $(1, (x_j - \bar{x}), \ldots, (x_j - \bar{x})^p)$, for $j = 1, ..., n$, W^+ is a $n \times n$ diagonal matrix with diagonal $\left\{\mathbf{1}(x_1 > \bar{x})\, k\left(\frac{x_1 - \bar{x}}{h}\right), \ldots \ \ldots, \mathbf{1}(x_n > \bar{x})\, k\left(\frac{x_n - \bar{x}}{h}\right)\right\}$, $Y = (y_1, \ldots, y_n)^T$, and $Z = [z_1 \ \ldots \ z_n]^T$. If $\bar{x}$ is at the right boundary of $\mathcal{X}$, the estimator $\hat{\tau}^+(\bar{x})^{\downarrow} = 0$.

The next conditions make it possible to obtain the asymptotic distribution of B_n given in equation (12). The essence of the proof can be understood by observing that when $\hat{\tau}^+(\bar{x})^{\downarrow}$ is defined as in (14),

$$\begin{aligned} \hat{\tau}^+(\bar{x})^{\downarrow} &:= e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1} \tilde{X}^T W^+ (Y - Z\hat{\gamma}^+) \qquad (15) \\ &= e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1} \tilde{X}^T W^+ (Y - Z\gamma^+) - e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1} \tilde{X}^T W^+ Z(\hat{\gamma}^+ - \gamma^+), \end{aligned}$$

The first term is a simple local polynomial estimator of a boundary point seen, as discussed, in Porter (2003), but also examined in Fan and Gijbels (1996). Deriving its asymptotic distribution in this case needs only a small modification to account for the fact that x does not have a density function, since $\mathbb{P}(x_i = \bar{x}) > 0$. It converges to a normally distributed random variable at the rate $\sqrt{nh}$. The second term can be considered jointly with the second term in equation (12), which converges at the rate $\sqrt{n}$. For testing in smaller samples, the results consider the effect of the estimation of γ^+ and γ^-. However both the bias and variance of $\hat{\tau}^+(\bar{x})^{\downarrow}$ and $\hat{\tau}^-(\bar{x})^{\uparrow}$ dominate the asymptotic behavior of $\hat{\theta}$.

Assumption 2.6.

1. *$\mathbb{E}(|g(z_i)|^{2+\xi_2} \mid x_i = \bar{x}) < \infty$ and $\mathbb{E}(||g(z_i)z_i||^{2+\xi_2} \mid x_i = \bar{x}) < \infty$, for some $\xi_2 > 0$.*

2. *If $\bar{x}$ is an interior point in $\mathcal{X}$, then the estimators $\hat{\gamma}^+$ and $\hat{\gamma}^-$ are defined as*

$$
\begin{aligned}
&\hat{\gamma}^+ = (\tilde{Z}_+^T \tilde{Z}_+)^{-1} \tilde{Z}_+^T \tilde{Y}_+, && \hat{\gamma}^- = (\tilde{Z}_-^T \tilde{Z}_-)^{-1} \tilde{Z}_-^T \tilde{Y}_-,\\
&\tilde{y}_{i+} = (y_i - \hat{\mathbb{E}}(y_i \mid x_i)^+)\mathbf{1}(x_i > \bar{x}), && \tilde{y}_{i-} = (y_i - \hat{\mathbb{E}}(y_i \mid x_i)^-)\mathbf{1}(x_i < \bar{x})\\
&\tilde{z}_{i+} = (z_i - \hat{\mathbb{E}}(z_i \mid x_i)^+)\mathbf{1}(x_i > \bar{x}), && \tilde{z}_{i-} = (z_i - \hat{\mathbb{E}}(z_i \mid x_i)^-)\mathbf{1}(x_i < \bar{x})\\
&\hat{\mathbb{E}}(y_i \mid x_i)^+ = \textstyle\sum_{j=1}^n \mathbf{1}(x_j > \bar{x})\, T_{i,j}^+ y_j, && \hat{\mathbb{E}}(y_i \mid x_i)^- = \textstyle\sum_{j=1}^n \mathbf{1}(x_j < \bar{x})\, T_{i,j}^- y_j,\\
&\hat{\mathbb{E}}(z_i \mid x_i)^+ = \textstyle\sum_{j=1}^n \mathbf{1}(x_j > \bar{x})\, T_{i,j}^+ z_j, && \hat{\mathbb{E}}(z_i \mid x_i)^- = \textstyle\sum_{j=1}^n \mathbf{1}(x_j < \bar{x})\, T_{i,j}^- z_j,
\end{aligned}
$$

for some $T_{i,j}^+$ and $T_{i,j}^-$ which are a function exclusively of the observations such that $x_i > \bar{x}$ and $x_i < \bar{x}$ respectively. Also, $\sup_i \left\| \sum_{j=1}^n \mathbf{1}(x_j > \bar{x})\, T_{i,j}^+ u_j - \mathbb{E}(u_i \mid x_i) \right\| = o_p(1)$ for $u_i = z_i, \epsilon_i^2, \mathbb{E}(z_i \mid x_i)\epsilon_i^2, z_i \mathbb{E}(\epsilon_i^2 \mid x_i)$ and $\mathbb{E}(z_i \mid x_i)\mathbb{E}(\epsilon_i^2 \mid x_i)$. $\hat{\gamma}^+$ and $\hat{\gamma}^-$ satisfy

$$
\sqrt{n} \begin{bmatrix} \hat{\gamma}^+ - \gamma^+ \\ \hat{\gamma}^- - \gamma^- \\ A_n \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{V}_\gamma^+ & 0 & 0 \\ 0 & \mathcal{V}_\gamma^- & 0 \\ 0 & 0 & V_A \end{bmatrix} \right),
$$

and there exist $\hat{\mathcal{V}}_{\gamma n}^+$ and $\hat{\mathcal{V}}_{\gamma n}^-$, functions exclusively of data for which $x_i > \bar{x}$ and $x_i < \bar{x}$ respectively, and such that $\hat{\mathcal{V}}_{\gamma n}^+ \xrightarrow{p} \mathcal{V}_{\gamma n}^+$ and $\hat{\mathcal{V}}_{\gamma n}^- \xrightarrow{p} \mathcal{V}_{\gamma n}^-$. Moreover, $\mathbb{E}(\|\sqrt{n}(\hat{\gamma}^+ - \gamma^+)\|^{2+\xi_3})$ and $\mathbb{E}(\|\sqrt{n}(\hat{\gamma}^- - \gamma^-)\|^{2+\xi_3})$ are uniformly bounded for all n and some $\xi_3 > 0$. If $\bar{x}$ is the left boundary of $\mathcal{X}$, all is true except that $\hat{\gamma}^- = 0$, $\mathcal{V}_\gamma^- = 0$, and $\hat{\mathcal{V}}_\gamma^- = 0$. If $\bar{x}$ is the right boundary of $\mathcal{X}$, all is true except that $\hat{\gamma}^+ = 0$, $\mathcal{V}_\gamma^+ = 0$, and $\hat{\mathcal{V}}_\gamma^+ = 0$.

3. *There exist $x^-, x^+ \in \mathbb{R}$, with $x^- < \bar{x} < x^+$ such that $F(x)$ is twice continuously differentiable in $[x_-, \bar{x}) \cup (\bar{x}, x^+]$ with first derivative bounded away from zero and second derivative uniformly bounded in $[x_-, \bar{x}) \cup (\bar{x}, x^+]$. Define*

$$
\begin{aligned}
&\phi(\bar{x})^\downarrow := \lim_{x \downarrow \bar{x}} \tfrac{d}{dx} F(x), && \phi(\bar{x})^\uparrow := \lim_{x \uparrow \bar{x}} \tfrac{d}{dx} F(x),\\
&\phi'(\bar{x})^\downarrow := \lim_{x \downarrow \bar{x}} \tfrac{d^2}{dx^2} F(x), && \phi'(\bar{x})^\uparrow := \lim_{x \uparrow \bar{x}} \tfrac{d^2}{dx^2} F(x),
\end{aligned}
$$

then all of these quantities exist. Moreover, there exist $\hat{\phi}(\bar{x})^\downarrow$ and $\hat{\phi}(\bar{x})^\downarrow$, consistent estimators of $\phi(\bar{x})^\downarrow$ and $\phi(\bar{x})^\downarrow$ respectively (see remark 2.3.2).

4. *The function $\tau^+(x)$ is at least $p+2$ times continuously differentiable in $(\bar{x}, x^+]$, and the function $\tau^-(x)$ is at least $p+2$ times continuously differentiable in $[x^-, \bar{x})$. Define*

$$
\tau^{+(m)}(\bar{x})^\downarrow := \lim_{x \downarrow \bar{x}} \tfrac{d^m}{dx^m} \tau^+(x), \qquad \tau^{-(m)}(\bar{x})^\uparrow := \lim_{x \downarrow \bar{x}} \tfrac{d^m}{dx^m} \tau^-(x),
$$

then these quantities exist for $m = 1, \ldots, p+2$.

5. *The variances $\sigma^2(x) := \mathbb{E}(\epsilon_i^2 \mid x_i = x)$ are at least $p+2$ continuously differentiable in $[x^-, \bar{x}) \cup (\bar{x}, x^+]$. The errors $\epsilon_i^{\epsilon^2} = \epsilon_i^2 - \sigma^2(x_i)$ have moments $\mathbb{E}(|\epsilon_i^{\epsilon^2}|^{2+\xi_4} \mid x_i)$ uniformly bounded for some $\xi_5 > 0$. Define*

$$
\sigma^2(\bar{x})^\downarrow := \lim_{x \downarrow \bar{x}} \sigma^2(x), \qquad \sigma^2(\bar{x})^\uparrow := \lim_{x \uparrow \bar{x}} \sigma^2(x),
$$

then these quantities exist.

6. *The kernel k is symmetric and has bounded support. For all j odd integers, $\int k^x(u)u^j du = 0$. Define $\upsilon_j = \int_0^\infty k(u)u^j du$ and $\omega_j = \int_0^\infty k^2(u)u^j du$, then*

$$\Upsilon_j = \begin{bmatrix} \upsilon_j \\ \vdots \\ \upsilon_{j+p} \end{bmatrix} \Lambda_j = \begin{bmatrix} \upsilon_j & \dots & \upsilon_{j+p} \\ \vdots & & \vdots \\ \upsilon_{j+p} & \dots & \upsilon_{j+2p+1} \end{bmatrix} \Omega_j = \begin{bmatrix} \omega_j \\ \vdots \\ \omega_{j+p} \end{bmatrix} \Omega = \begin{bmatrix} \omega_j & \dots & \omega_{j+p} \\ \vdots & & \vdots \\ \omega_{j+p} & \dots & \omega_{j+2p+1} \end{bmatrix}.$$

7. $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh = \infty$, *and* $\lim_{n\to\infty} h^{p+1}\sqrt{n} < \infty$.

8. *The functions $\mathbb{E}(z_i^d \mid x_i = x)$ and $\mathbb{E}(\epsilon_i^2 z_i^d \mid x_i = x)$ are at least $p+2$ times continuously differentiable in $[x^-, \bar{x}) \cup (\bar{x}, x^+]$. The $\epsilon_i^z := z_i - \mathbb{E}(z_i \mid x_i)$ have moments $\mathbb{E}(||\epsilon_i^z||^{2+\xi_5} \mid x_i)$ uniformly bounded for some $\xi_5 > 0$. Define*

$$\mathbb{E}(z_i \mid x_i = \bar{x})^\downarrow := \lim_{x\downarrow\bar{x}} \mathbb{E}(z_i \mid x_i = x), \qquad \mathbb{E}(z_i \mid x_i = \bar{x})^\uparrow := \lim_{x\uparrow\bar{x}} \mathbb{E}(z_i \mid x_i = x),$$
$$\mathbb{E}(z_i z_i^T \mid x_i = \bar{x})^\downarrow := \lim_{x\downarrow\bar{x}} \mathbb{E}(z_i z_i^T \mid x_i = x), \quad \mathbb{E}(z_i z_i^T \mid x_i = \bar{x})^\uparrow := \lim_{x\uparrow\bar{x}} \mathbb{E}(z_i z_i^T \mid x_i = x),$$
$$\mathbb{E}(z_i \epsilon_i^2 \mid x_i = \bar{x})^\downarrow := \lim_{x\downarrow\bar{x}} \mathbb{E}(z_i \sigma^2(x_i, z_i) \mid x_i = x),$$
$$\mathbb{E}(z_i \epsilon_i^2 \mid x_i = \bar{x})^\uparrow := \lim_{x\uparrow\bar{x}} \mathbb{E}(z_i \sigma^2(x_i, z_i) \mid x_i = x),$$

then all of these quantities exist. Finally, define the notation

$$\Sigma_z(\bar{x})^\downarrow := \mathbb{E}(z_i z_i^T \mid x_i = \bar{x})^\downarrow - \mathbb{E}(z_i \mid x_i = \bar{x})^\downarrow \mathbb{E}(z_i \mid x_i = \bar{x})^{\downarrow T}$$
$$\Sigma_z(\bar{x})^\uparrow := \mathbb{E}(z_i z_i^T \mid x_i = \bar{x})^\uparrow - \mathbb{E}(z_i \mid x_i = \bar{x})^\uparrow \mathbb{E}(z_i \mid x_i = \bar{x})^{\uparrow T}$$
$$c_{z\epsilon^2}(\bar{x})^\downarrow := \mathbb{E}(z_i \epsilon_i^2 \mid x_i = \bar{x})^\downarrow - \mathbb{E}(z_i \mid x_i = \bar{x})^\downarrow \sigma^2(\bar{x})^\downarrow$$
$$c_{z\epsilon^2}(\bar{x})^\uparrow := \mathbb{E}(z_i \epsilon_i^2 \mid x_i = \bar{x})^\uparrow - \mathbb{E}(z_i \mid x_i = \bar{x})^\uparrow \sigma^2(\bar{x})^\uparrow$$

9. *If $\bar{x}$ is the left boundary of $\mathcal{X}$, then $\alpha = 1$, and if $\bar{x}$ is the right boundary of $\mathcal{X}$, then $\alpha = 0$.*

Theorem 5. *If assumptions 2.1, 2.4 and 2.5 hold, then*

$$\sqrt{nh}\mathcal{V}_n^{-1/2}(\hat{\theta} - \theta - \mathcal{B}_n) \xrightarrow{d} \mathcal{N}(0, 1),$$

where

$$\mathcal{B}_n = \mathbb{E}(g(z_i) \,|\, x_i = \bar{x}) \left[\alpha \mathcal{B}_n^+ + (1-\alpha)\mathcal{B}_n^-\right],$$

$$\mathcal{B}_n^+ = \begin{cases} h^{p+1} \frac{\tau^{+(p+1)}(\bar{x})^{\lim}}{(p+1)!} e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+1}), & \textit{if } p \textit{ is odd,} \\ \\ h^{p+2} \left[\frac{\tau^{+(p+1)}(\bar{x})^{\lim}}{(p+1)!} \frac{\phi'(\bar{x})^\downarrow}{\phi(\bar{x})^\downarrow}\right] e_1^T \Lambda_0^{-1}(\Upsilon_{p+2} - \Lambda_1 \Lambda_0 \Upsilon_{p+1}) & \\ \quad + \left[\frac{\tau^{+(p+2)}(\bar{x})^{\lim}}{(p+2)!}\right] e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+2}) & \textit{if } p \textit{ is even,} \end{cases}$$

and analogously for $\mathcal{B}_n^-$, substituting the "+" by "−" in the notation.

$$\mathcal{V}_n = \alpha^2\Big[\mathcal{V}_\tau^+ + 2\sqrt{h}C_+^T\mathcal{C}_{\tau\gamma}^+ + hC_+^T\mathcal{V}_\gamma^+C_+\Big] + (1-\alpha)^2\Big[\mathcal{V}_\tau^- + 2\sqrt{h}C_-^T\mathcal{C}_{\tau\gamma}^- + hC_-^T\mathcal{V}_\gamma^-C_-\Big] + hV_A + o(h),$$

where if $\bar{x}$ is an interior point or is at the left boundary of $\mathcal{X}$,

$$\begin{aligned}
\mathcal{V}_\tau^+ &= \mathbb{E}(g(z_i)\,|\,x_i=\bar{x})^2\mathcal{V}^+,\\
\mathcal{V}^+ &= \frac{\sigma^2(\bar{x})^\downarrow}{\phi(\bar{x})^\downarrow}e_1^T\Lambda_0^{-1}\Omega\,\Lambda_0^{-1}e_1,\\
C_+ &= \mathbb{E}(g(z_i)z_i\,|\,x_i=\bar{x}) - \mathbb{E}(g(z_i)\,|\,x_i=\bar{x})\,\mathbb{E}(z_i \mid x_i=\bar{x})^\downarrow,\\
\mathcal{C}_{\tau\gamma}^+ &= \left(\Sigma_z(\bar{x})^\downarrow\right)^{-1}c_{z\epsilon^2}(\bar{x})^\downarrow,
\end{aligned}$$

and if $\bar{x}$ is an interior point or is at the right boundary of the support of the x_i, $\mathcal{V}_\tau^-$, $\mathcal{V}_-$, C_- and $\mathcal{C}_{\tau\gamma}^-$ are defined analogously, substituting the "+" by "−" and $\downarrow$ by $\uparrow$ in the notation.

The proof is in section A.3.1 in the appendix. The following definitions concern the estimation of the variance $\mathcal{V}_n$. Define the operator

$$P_t^+ = e_t^T(\tilde{X}^TW^+\tilde{X})^{-1}\tilde{X}^TW^+.$$

Then, observe that $\hat{\tau}(\bar{x})^\downarrow = P_1^+(Y - Z\hat{\gamma}^+)$. Whenever $\bar{x}$ is an interior point or is the left boundary of $\mathcal{X}$, the quantities $\hat{C}_+$, $\hat{\Sigma}_z(\bar{x})^\downarrow$, $c_{z\epsilon^2}(\bar{x})^\downarrow$ and $\hat{\sigma}^2(\bar{x})^\downarrow$ are defined in equations (16)-(18) below:

$$\begin{aligned}
&\hat{\mathbb{E}}(z_i \mid x_i=\bar{x})^\downarrow = (P_1^+Z)^T,\\
&\hat{C}_+ := \alpha\,\hat{\mathbb{E}}(g(z_i)z_i\,|\,x_i=\bar{x}) - \hat{\mathbb{E}}(g(z_i)\,|\,x_i=\bar{x})\,\hat{\mathbb{E}}(z_i \mid x_i=\bar{x})^\downarrow.
\end{aligned} \tag{16}$$

Let $U_{ls} = \begin{bmatrix} z_1^lz_1^s \\ \vdots \\ z_n^lz_n^s \end{bmatrix}$, $\quad \hat{E}(z_iz_i^T \mid x_i=\bar{x})^\downarrow = \begin{bmatrix} P_1^+U_{11} & \dots & P_1^+U_{1d} \\ \vdots & & \vdots \\ P_1^+U_{d1} & \dots & P_1^+U_{dd} \end{bmatrix}$, then

$$\hat{\Sigma}_z(\bar{x})^\downarrow = \hat{E}(z_iz_i^T \mid x_i=\bar{x})^\downarrow - \hat{\mathbb{E}}(z_i \mid x_i=\bar{x})^\downarrow\,\hat{\mathbb{E}}(z_i^T \mid x_i=\bar{x})^\downarrow. \tag{17}$$

Let $R_z^+ = \begin{bmatrix} (y_1 - z_1^T\hat{\gamma}^+)^2z_1^T \\ \vdots \\ (y_n - z_n^T\hat{\gamma}^+)^2z_n^T \end{bmatrix}$, then

$$\begin{aligned}
&\hat{\mathbb{E}}(z_i(y_i - z_i\hat{\gamma}^+)^2 \mid x_i=\bar{x})^\downarrow = P_1^+R_z^{1+},\\
&\hat{c}_{z\epsilon^2}(\bar{x})^\downarrow = \hat{\mathbb{E}}(z_i(y_i - z_i\hat{\gamma}^+)^2 \mid x_i=\bar{x})^\downarrow - \hat{\mathbb{E}}(z_i \mid x_i=\bar{x})^\downarrow\hat{\mathbb{E}}((y_i - z_i\hat{\gamma}^+)^2 \mid x_i=\bar{x})^\downarrow.
\end{aligned} \tag{18}$$

Let $R^+ = \begin{bmatrix} (y_1 - z_1^T \hat{\gamma}^+)^2 \\ \vdots \\ (y_n - z_n^T \hat{\gamma}^+)^2 \end{bmatrix}$, then

$$
\begin{aligned}
&\hat{\mathbb{E}}((y_i - z_i\hat{\gamma}^+)^2 \mid x_i = \bar{x})^{\downarrow} = P_1^+ R^+ \\
&\hat{\sigma}^2(\bar{x})^{\downarrow} = \mathbb{E}((y_i - z_i^T\hat{\gamma}^+)^2 \mid x_i = \bar{x})^{\downarrow} - (\hat{\tau}(\bar{x})^{\lim +})^2.
\end{aligned} \tag{19}
$$

Finally, if $\bar{x}$ is an interior point or is the right boundary of $\mathcal{X}$, then $\hat{C}_-$, $\hat{\Sigma}_z(\bar{x})^{\uparrow}$, $c_{z\epsilon^2}(\bar{x})^{\uparrow}$ and $\hat{\sigma}^2(\bar{x})^{\uparrow}$ are defined analogously, substituting "+" by "-" and "↓" by "↑" in the notation.

Theorem 6. *Under H_0: x_i is exogenous, $\theta = 0$ and $V_A = 0$. If assumptions 2.1, 2.4 and 2.6 hold, then if*

$$
\hat{\mathcal{V}}_n = \alpha^2\left[\hat{\mathcal{V}}_\tau^+ + 2\sqrt{h}\hat{C}_+^T\hat{\mathcal{C}}_{\tau\gamma}^+ + h\hat{C}_+^T\hat{\mathcal{V}}_\gamma^+\hat{C}_+\right] + (1-\alpha)^2\left[\hat{\mathcal{V}}_\tau^- + 2\sqrt{h}\hat{C}_-^T\hat{\mathcal{C}}_{\tau\gamma}^- + h\hat{C}_-^T\hat{\mathcal{V}}_\gamma^-\hat{C}_-\right],
$$

where

$$
\begin{aligned}
\hat{\mathcal{V}}_\tau^+ &= \hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})^2\hat{\mathcal{V}}^+, & \hat{\mathcal{V}}_\tau^- &= \hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x})^2\hat{\mathcal{V}}^-, \\
\hat{\mathcal{V}}^+ &= \tfrac{\hat{\sigma}^2(\bar{x})^{\downarrow}}{\hat{\phi}(\bar{x})^{\downarrow}} e_1^T\Lambda_0^{-1}\Omega\,\Lambda_0^{-1}e_1, & \hat{\mathcal{V}}^- &= \tfrac{\hat{\sigma}^2(\bar{x})^{\uparrow}}{\hat{\phi}(\bar{x})^{\uparrow}} e_1^T\Lambda_0^{-1}\Omega\,\Lambda_0^{-1}e_1, \\
\hat{\mathcal{C}}_{\tau\gamma}^+ &= \left(\hat{\Sigma}_z(\bar{x})^{\downarrow}\right)^{-1}\hat{c}_{z\epsilon^2}(\bar{x})^{\downarrow} & \hat{\mathcal{C}}_{\tau\gamma}^- &= \left(\hat{\Sigma}_z(\bar{x})^{\uparrow}\right)^{-1}\hat{c}_{z\epsilon^2}(\bar{x})^{\uparrow},
\end{aligned}
$$

then $\hat{\mathcal{V}}_n - \mathcal{V}_n = o_p(1)$.

The proof is in section A.3.2. The following theorem gives the discontinuity test properties in the partially linear case.

Theorem 7. *Let $0 \leqslant \lambda \leqslant 1$, Φ be the standard normal cumulative distribution function, and $c_\lambda = \Phi^{-1}(\lambda)$. If theorems 1, 5 and 6 hold and $\sqrt{nh}h^{p+1} \to 0$ as $n \to \infty$, then under H_0: x is exogenous,*

$$
\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} \leqslant c_\lambda\right) \to \lambda \quad as \quad n \to \infty.
$$

moreover, if result 1 is true, under H_1: x is endogenous,

$$
\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) \to 1 \quad as \quad n \to \infty,
$$

and under the local alternatives $\frac{\theta}{\sqrt{nh}}$,

$$
\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} \leqslant c_\lambda\right) \to \Phi\left(c_\lambda - \frac{\theta}{\sqrt{\alpha^2\mathcal{V}_\tau^+ + (1-\alpha)^2\mathcal{V}_\tau^-}}\right) \quad as \quad n \to \infty.
$$

See proof in section A.3.3. Observe that the variance of the estimation of the nonparametric terms τ^+ and τ^+ is the only variance that affects the local power of the test in large samples. This occurs because the other components in $\mathcal{V}_n$ are $o(nh)$, more specifically, $O(nh^{3/2})$.

Remark 2.3.2. *The estimation of $\phi(\bar{x})^{\downarrow}$ and $\phi(\bar{x})^{\uparrow}$ is not a trivial application of the literature of density estimation. When estimating limits of densities at boundary points, the same concerns as with the estimation of conditional expectations at boundary points arise, so $\hat{\phi}(\bar{x})^{\downarrow}$ and $\phi(\bar{x})^{\uparrow}$ must be chosen mindful of their boundary properties. Although local polynomial estimators have excellent boundary properties, they cannot be naturally transformed for density estimation, as it can be done with kernels. One solution is to estimate $\phi(\bar{x})^{\downarrow}$ with boundary kernels, as in Jones (1993). The application section uses a different approach, based on the estimator proposed in Lejeune and Sarda (1992), which consists on the local polynomial regression of the empirical distribution function $\hat{F}(x_i)$ on x_i using only observations such that $x_i > 0$. The coefficient of the constant term is an estimator of $\lim_{x\downarrow\bar{x}} F(x)$, but the coefficient of the linear term is actually an estimator of $\lim_{x\downarrow\bar{x}} \frac{d}{dx}F(x)$, which is exactly $\phi(\bar{x})^{\downarrow}$. Hence, in this case*

$$\hat{\phi}(\bar{x})^{\downarrow} = e_2^T(\tilde{X}^T W^+ \tilde{X})^{-1}\tilde{X}^T W^+ \hat{F} = P_2^+ \hat{F},$$

where $\hat{F} = (\hat{F}_1, \ldots, \hat{F}_n)^T$, $\hat{F}_j = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(x_i \leqslant x_j)$. Analogously for $\hat{\phi}(\bar{x})^{\uparrow}$.

2.3.3 The nonparametric case

Let the conditional expectation be represented by the function f, so that

$$f(x_i, z_i) := \mathbb{E}(y_i \,|\, x_i, z_i), \tag{20}$$

and define $f(\bar{x}, z_i)^{\downarrow} := \lim_{x\downarrow\bar{x}} f(x, z_i)$, $f(\bar{x}, z_i)^{\uparrow} := \lim_{x\uparrow\bar{x}} f(x, z_i)$, and suppose that these limits exist for all z_i.

Example 3. *(Censoring) Equation (20) can be parameterized inside the censoring model presented in section 2.2.1. From equation (4), observe that for $x > 0$,*

$$f(x, z) = f_1(x, z, f_2^{-1}(x; z)).$$

Assumption 2.3 can be modified to serve as a primitive of assumption 2.7 (3).

In this case, $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = f(\bar{x}, z_i)^{\downarrow}$, and $\mathbb{E}(y_i \,|\, \bar{x}, z_i)^{\uparrow} = f(\bar{x}, z_i)^{\uparrow}$. Hence, $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\downarrow} = \hat{f}(\bar{x}, z_i)^{\downarrow}$, and $\hat{\mathbb{E}}(y_i \,|\, \bar{x}, z_i)^{\uparrow} = \hat{f}(\bar{x}, z_i)^{\uparrow}$. Define $\hat{\mathbb{E}}(g(z_i) \,|\, x_i = \bar{x}) = \frac{1}{\hat{p}_{\bar{x}}}\frac{1}{n}\sum_{i=1}^n g(z_i)\mathbf{1}(x_i = \bar{x})$ and $\hat{\mathbb{E}}(g(z_i)z_i \,|\, x_i = \bar{x}) = \frac{1}{\hat{p}_{\bar{x}}}\frac{1}{n}\sum_{i=1}^n \mathbf{1}(x_i = \bar{x})g(z_i)z_i$, then equation (7) cannot be simplified as in the previous cases. The present case will assume that the z_i are random variables which can take a finite number of values. Similar results could be derived when the z_i can take a countable number of values, and also when the z_i are continuous or mixed random variables. The decision to present results in the finite case has the advantage of the simplicity, but is also done for practical reasons, as is explained in remark 2.3.4 below. The following exposition refers to the estimation of $f(\bar{x}, z_i)^{\downarrow}$, and $f(\bar{x}, z_i)^{\uparrow}$ is estimated analogously.

Let the $z_i \in \{z^1, \ldots, z^M\}$, and define the estimator of $f(\bar{x}, z^m)^{\downarrow}$ in the following way. Given the kernel function k, the smoothing parameter h, the polynomial degree p, and let $\hat{a}_0, \ldots, \hat{a}_p$ be the solution to the problem

$$\min_{a_0,\ldots,a_p} \frac{1}{n} \sum_{j=1}^{n} k\left(\frac{x_j - \bar{x}}{h}\right) \mathbf{1}(x_j > \bar{x}) \left[y_j - a_0 - a_1(x_j - \bar{x}) - \ldots - a_p(x_j - \bar{x})^p\right]^2.$$

If $\bar{x}$ is an interior point or is the left boundary of $\mathcal{X}$, the local polynomial estimator of $f(\bar{x}, z^m)^{\downarrow}$ is given by

$$\hat{f}(\bar{x}, z^m)^{\downarrow} = \hat{a}_0 = e_1^T (\tilde{X}^T W_m^+ \tilde{X})^{-1} \tilde{X}^T W_m^+ Y, \tag{21}$$

where $e_1 = (1, 0, \ldots, 0)^T$ has dimension $1 \times (p+1)$, $\tilde{X}$ has rows equal to $(1, (x_j - \bar{x}), \ldots, (x_j - \bar{x})^p)$, $j = 1, \ldots, n$, W_m^+ is a $n \times n$ diagonal matrix with diagonal elements $\left\{\mathbf{1}(z_1 = z^m)\mathbf{1}(x_1 > \bar{x})\, k\left(\frac{x_1 - \bar{x}}{h}\right), \ldots,\ \mathbf{1}(z_n = z^m)\mathbf{1}(x_n > \bar{x})\, k\left(\frac{x_n - \bar{x}}{h}\right)\right\}$, and $Y = (y_1, \ldots, y_n)^T$. If $\bar{x}$ is the right boundary of $\mathcal{X}$, then $\hat{f}(\bar{x}, z^m)^{\downarrow} = 0$.

Let $\hat{p}_{\bar{x}}^m := \left(\sum_{i=1}^n \mathbf{1}(x_i = \bar{x})\right)^{-1} \sum_{i=1}^n \mathbf{1}(z_i = z^m)\mathbf{1}(x_i = \bar{x})$ be an estimator of $p_{\bar{x}}^m := \mathbb{P}(z_i = z^m \mid x_i = \bar{x})$, hence

$$B_n = \alpha \sum_{m=1}^{M} \hat{p}_{\bar{x}}^m\, \hat{\Gamma}(z^m)^+ g(z^m) + (1 - \alpha) \sum_{m=1}^{M} \hat{p}_{\bar{x}}^m\, \hat{\Gamma}(z^m)^- g(z^m).$$

The next assumption provides conditions that allow the derivation of the asymptotic distribution of B_n.

Assumption 2.7.

1. *$dF(x, z^m) > 0$, for all m and $x \in (x^-, x^+) \cap \mathcal{X}$ (see remark 2.3.4 below for when this condition fails).*
2. *There exist $x^-, x^+ \in \mathbb{R}$, with $x^- < \bar{x} < x^+$ such that $\mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)$ is twice continuously differentiable in x with first derivative bounded away from zero and second derivative uniformly bounded for x in $(x_-\,,\, \bar{x}) \cup (\bar{x}\,,\, x^+)$ and all m. Define $\phi(\bar{x}, z^m)^{\downarrow} := \lim_{x \downarrow \bar{x}} \frac{d}{dx} \mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)$, $\phi(\bar{x}, z^m)^{\uparrow} := \lim_{x \uparrow \bar{x}} \frac{d}{dx} \mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)$, $\phi'(\bar{x}, z^m)^{\downarrow} := \lim_{x \downarrow \bar{x}} \frac{d^2}{dx^2} \mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)$, and $\phi'(\bar{x}, z^m)^{\uparrow} := \lim_{x \uparrow \bar{x}} \frac{d^2}{dx^2} \mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)$, then all of these quantities exist. Moreover, there exist $\hat{\phi}(\bar{x}, z^m)^{\downarrow}$ and $\hat{\phi}(\bar{x}, z^m)^{\uparrow}$, consistent estimators of $\phi(\bar{x}, z^m)^{\downarrow}$ and $\phi(\bar{x}, z^m)^{\uparrow}$ respectively (see remark 2.3.3 below).*
3. *The function $f(x, z^m)$ is at least $p+2$ times continuously differentiable in x in $(x^-, \bar{x}) \cap (\bar{x}, x^+)$ for all m. Define $f^{(l)}(\bar{x})^{\downarrow} := \lim_{x \downarrow \bar{x}} \frac{d^l}{dx^l} f(x, z^m)$, and $f^{(l)}(\bar{x}, z^m)^{\uparrow} := \lim_{x \downarrow \bar{x}} \frac{d^l}{dx^l} f(x, z^m)$, then these quantities exist for $l = 1, \ldots, p+2$ and all m.*
4. *The variances $\sigma^2(x, z^m) := \mathbb{E}(\epsilon_i^2 \mid x_i = x, z_i = z^m)$ are continuous in $(x_-\,,\, \bar{x}) \cup (\bar{x}\,,\, x^+)$, and the limits $\sigma^2(\bar{x}, z^m)^{\downarrow} := \lim_{x \downarrow \bar{x}} \sigma^2(x, z^m)$ and $\sigma^2(\bar{x}, z^m)^{\uparrow} := \lim_{x \uparrow \bar{x}} \sigma^2(x, z^m)$ exist for all m. Moreover, the moments $\mathbb{E}(|\epsilon_i^2|^{2+\xi_6} \mid x_i = x, z_i = z^m)$ are uniformly bounded for some $\xi_6 > 0$.*

5. *The kernel k is continuous, symmetric and has bounded support. For all j odd integers, $\int k(u)u^j du = 0$.*
6. $\lim_{n\to\infty} h = 0$*,* $\lim_{n\to\infty} nh = \infty$*, and* $\lim_{n\to\infty} h^{p+1}\sqrt{nh} < \infty$*.*
7. *If $\bar{x}$ is the left boundary of $\mathcal{X}$, then $\alpha = 1$, and if $\bar{x}$ is the right boundary of $\mathcal{X}$, then $\alpha = 0$.*

Theorem 8. *If assumptions 2.1, 2.4 and 2.7 hold, then*

$$\sqrt{nh}\mathcal{V}_n^{-1/2}(\hat{\theta} - \theta - \mathcal{B}_n) \xrightarrow{d} \mathcal{N}(0,1)$$

where

$$\mathcal{B}_n = \sum_{m=1}^{M} \hat{p}_{\bar{x}}^m \, g(z^m) \left[\alpha\mathcal{B}_{m,n}^+ + (1-\alpha)\mathcal{B}_{m,n}^-\right]$$

$$\mathcal{B}_{m,n}^+ = \begin{cases} h^{p+1} \frac{f^{+(p+1)}(\bar{x},z^m)^{\lim}}{(p+1)!} e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+1}), & \text{if } p \text{ is odd,} \\ h^{p+2} \left[\frac{f^{+(p+1)}(\bar{x},z^m)^{\lim}}{(p+1)!} \frac{\phi'(\bar{x},z^m)^{\downarrow}}{\phi(\bar{x},z^m)^{\downarrow}}\right] e_1^T \Lambda_0^{-1}(\Upsilon_{p+2} - \Lambda_1\Lambda_0\Upsilon_{p+1}) \\ \qquad + \left[\frac{f^{+(p+2)}(\bar{x},z^m)^{\lim}}{(p+2)!}\right] e_1^T \Lambda_0^{-1}\Upsilon_{p+1} + o(h^{p+2}), & \text{if } p \text{ is even,} \end{cases}$$

and analogously for $\mathcal{B}_{m,n}^-$, just substitute the "+" by "−" in the notation. Finally, Λ_0, Λ_1, Υ_{p+1} and Υ_{p+2} are defined in assumption 2.6 (6).

$$\mathcal{V}_n = \mathcal{V} + hV_A + o(h),$$

$$\mathcal{V} = \sum_{m=1}^{M} \left(p_{\bar{x}}^m\right)^2 \, g(z^m)^2 \left[\alpha^2\mathcal{V}_m^+ + (1-\alpha)^2\mathcal{V}_m^-\right],$$

$$\mathcal{V}_m^+ = \frac{\sigma^2(\bar{x},z^m)^{\downarrow}}{\phi(\bar{x},z^m)^{\downarrow}} e_1^T \Lambda_0^{-1}\Omega\,\Lambda_0^{-1} e_1,$$

and analogously for $\mathcal{V}_m^-$, just substitute the "+" by "−" in the notation. Finally, Ω is defined in assumption 2.6 (6).

The proof is in section A.4.1 in the appendix. Its essence can be understood by observing that, since M is finite, the asymptotic distribution of B_n as defined in equation (7) can be trivially derived if the convergence of the $\hat{\Gamma}(z^m)^+$ and $\hat{\Gamma}(z^m)^-$ is ascertained. Since each z^m has positive probability in a neighborhood of $\bar{x}$, the results in Porter (2003) can be applied with the same modifications as in the partially linear case to account for the fact that $\bar{x}$ is a mass point.

The estimation of the variance depends on the estimation of $\sigma^2(\bar{x},z^m)^{\downarrow}$ and $\sigma^2(\bar{x},z^m)^{\downarrow}$. This step requires the estimation of the residuals. Define the operator

$$P_{t,m,x}^+ = e_t^T(\tilde{X}_x^T W_{x,m}^+ \tilde{X}_x)^{-1}\tilde{X}_x^T W_{x,m}^+.$$

where $\tilde{X}_x$ has rows equal to $(1, (x_i - x), \dots, (x_i - x)^p)$, and $W^+_{x,m}$ is a diagonal matrix with diagonal elements equal to $\left\{\mathbf{1}(x_1 > \bar{x}\,,\, z_1 = z^m)\, k\left(\frac{x_1 - \bar{x}}{h}\right), \dots, \mathbf{1}(x_n > \bar{x}\,,\, z_n = z^m)\, k\left(\frac{x_n - \bar{x}}{h}\right)\right\}$. Whenever $\bar{x}$ is an interior point or is the left boundary of $\mathcal{X}$, $\hat{f}(\bar{x}, z^m)^{\downarrow} = P^s_{1,m,\bar{x}} Y$, and $\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow}$ is defined in equation (22) below. Define

$$
\begin{aligned}
&\hat{f}^+(x_i, z^m) = P^+_{1,m,x} Y \\
&\hat{\epsilon}^+_i = y_i - \hat{f}^+(x_i, z_i) \\
&R = ((\hat{\epsilon}^+_1)^2, \dots, (\hat{\epsilon}^+_n)^2)^T, \\
&\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow} = P^+_{1,m,\bar{x}} R
\end{aligned} \tag{22}
$$

and if $\bar{x}$ is the right boundary of $\mathcal{X}$, $\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow} = 0$. Analogously for $\hat{\sigma}^2(\bar{x}, z^m)^{\uparrow}$, substituting "+" by "-" and "↓" by "↑" in the notation.

Assumption 2.8.

1. *The variances $\sigma^2(x, z^m)$ are at least $p + 2$ times continuously differentiable in $(x_-\,,\, \bar{x}) \cup (\bar{x}\,,\, x^+)$ for all m. Moreover, $\lim_{x \downarrow \bar{x}} d^{\,(l)} \sigma^2(x, z^m)$ and $\lim_{x \downarrow \bar{x}} d^{\,(l)} \sigma^2(x, z^m)$ exist for $l = 1, \dots, p+2$ and all m.*
2. *The moments $\mathbb{E}\left(\left(\epsilon_i^2 - \sigma_\epsilon^2(x_i, z_i)\right)^2 \mid x_i = x, z_i = z^m\right)$ are continuous and uniformly bounded in $(x_-\,,\, \bar{x}) \cup (\bar{x}\,,\, x^+)$, and the right and left limits when $x \to \bar{x}$ exist for all m.*
3. $hn^{1/3} (\log n)^{-1/3} \to \infty$

Theorem 9. *Suppose assumptions 2.1, 2.4, 2.7 and 2.8 hold. Under H_0: x_i is exogenous, $\theta = 0$ and $V_A = 0$. Then*

$$
\sqrt{nh}\, \hat{\mathcal{V}}_n^{-1/2} (\hat{\theta} - \mathcal{B}_n) \xrightarrow{d} \mathcal{N}(0, 1),
$$

where

$$
\hat{\mathcal{V}}_n = \hat{\mathcal{V}} := \sum_{m=1}^{M} (\hat{p}^m_{\bar{x}})^2 \; g(z^m)^2 \left[\alpha^2 \hat{\mathcal{V}}^+_m + (1 - \alpha)^2 \hat{\mathcal{V}}^-_m\right]
$$

with

$$
\hat{\mathcal{V}}^+_m = \frac{\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow}}{\hat{\phi}(\bar{x}, z^m)^{\downarrow}} e_1^T \Lambda_0^{-1} \Omega\, \Lambda_0^{-1} e_1, \quad \textit{and} \quad \hat{\mathcal{V}}^-_m = \frac{\hat{\sigma}^2(\bar{x}, z^m)^{\uparrow}}{\hat{\phi}(\bar{x}, z^m)^{\uparrow}} e_1^T \Lambda_0^{-1} \Omega\, \Lambda_0^{-1} e_1
$$

The proof is in section A.4.2. It relies on Masry (1996)'s result about the uniform convergence of the local polynomial estimator applied to the estimated $\hat{\epsilon}_i^2$. The following theorem gives the discontinuity test properties in the partially linear case.

Theorem 10. *Let $0 \leqslant \lambda \leqslant 1$, Φ be the standard normal cumulative distribution function, and $c_\lambda = \Phi^{-1}(\lambda)$. If theorems 1, 8 and 9 hold and $\sqrt{nh} h^{p+1} \to 0$ as $n \to \infty$, then under H_0: x is exogenous,*

$$
\mathbb{P}\left(\sqrt{nh} \frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} \leqslant c_\lambda\right) \to \lambda \quad \textit{as} \quad n \to \infty.
$$

moreover, if result 1 is true, under H_1: x is endogenous,

$$\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) \to 1 \quad \text{as} \quad n \to \infty,$$

and under the local alternatives $\frac{\theta}{\sqrt{nh}}$,

$$\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} \leqslant c_\lambda\right) \to \Phi\left(c_\lambda - \frac{\theta}{\sqrt{\mathcal{V}}}\right) \quad \text{as} \quad n \to \infty.$$

See proof in section A.4.3. If $\sqrt{nh}h^{p+1} \to 0$, $\sqrt{nh}\mathcal{B}^+_{m,n} \to 0$ and $\sqrt{nh}\mathcal{B}^-_{m,n} \to 0$. Hence $\sqrt{nh}\mathcal{B}_n \to 0$. The rest of the theorem is an immediate consequence of the theorems invoked and Slutsky's theorem, so no further proof is necessary.

Remark 2.3.3. *The estimation of $\phi(\bar{x}, z^m)^\downarrow$ can be done as in the partially linear case (see remark 2.3.2), following the approach proposed in Lejeune and Sarda (1992). The values $\hat{F}_m(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(x_i \leqslant x)\mathbf{1}(z_i = z^m)$ are consistent estimators of $\mathbb{P}(x_i \leqslant x, z_i = z^m)$. The approach consists on the local polynomial regression of the function $\hat{F}_m(x_i)$ on x_i at $\bar{x}$, using only observations such that $z_i = z^m$ and $x_i > \bar{x}$. The coefficient of the constant term is an estimator of $\lim_{x\downarrow\bar{x}} \mathbb{P}(x_i \leqslant x, z_i = z^m)$, but the coefficient of the linear term is actually an estimator of $\lim_{x\downarrow\bar{x}} \frac{d}{dx}\mathbb{P}(x_i \leqslant x, z_i = z^m)$, which is exactly $\phi(\bar{x}, z^m)^\downarrow$. Hence, in this case*

$$\hat{\phi}(\bar{x}, z^m)^\downarrow = P^+_{2,m,\bar{x}}\hat{F}_m,$$

where $\hat{F}_m = (\hat{F}_m(x_1), \ldots, \hat{F}_m(x_n))^T$. Analogously for $\hat{\phi}(\bar{x})^\uparrow$.

Remark 2.3.4. *The measure ν in $\theta = \int G(\Delta(z), z)d\nu(z)$ is chosen by the researcher. The measure chosen for the derivation of the estimators is $F(z \mid x_i = \bar{x})$, and from that derives the requirement that if $\mathbb{P}(z_i = z^m, x_i = \bar{x}) > 0$, then for estimation purposes it is necessary that $dF(x, z^m) > 0$, for all x in a neighborhood of $\bar{x}$. All the results can be derived in exactly the same way if the measure chosen is $F(z \mid x_i = \bar{x}, z \in \bar{\mathcal{A}})$, where $\bar{\mathcal{A}}$ is a finite subset of $\mathcal{A} := \{z\,;\, dF(x, z) > 0, \forall x \in (x^-, x^+) \cap \mathcal{X}\}$, as long as $\bar{\mathcal{A}}$ is not empty. Hence, A_n and B_n in equations (6) and (7) are substituted by*

$$A_n = \frac{1}{\hat{p}_{\bar{x},\bar{\mathcal{A}}}}\frac{1}{n}\sum_{i=1}^n \Delta(z_i)g(z_i)\mathbf{1}(x_i = \bar{x}, z_i \in \bar{\mathcal{A}}) - \mathbb{E}(\Delta(z_i)g(z_i)\,|\, x_i = \bar{x}, z_i \in \bar{\mathcal{A}})$$
$$B_n = \frac{1}{\hat{p}_{\bar{x},\bar{\mathcal{A}}}}\frac{1}{n}\sum_{i=1}^n [\alpha\hat{\Gamma}(z_i)^+ + (1-\alpha)\hat{\Gamma}(z_i)^-]g(z_i)\mathbf{1}(x_i = \bar{x}, z_i \in \bar{\mathcal{A}}).$$

where $\hat{p}_{\bar{x},\bar{\mathcal{A}}} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(x_i = \bar{x}, z_i \in \bar{\mathcal{A}})$. Assumption 2.1 remains the same, assumption 2.4 remains the same, except for the new definition of $V_A := \mathbb{V}ar(\Delta(z_i)g(z_i \mid x_i = \bar{x}, z_i \in \bar{\mathcal{A}})$, and assumptions 2.7 and 2.8 remain the same as long as $\{z^1, \ldots, z^M\} = \bar{\mathcal{A}}$. The results in theorems 8, 9 and 10 remain unchanged.

Even when the $dF(z_i)$ does not have a finite support, the discontinuity test using a measure ν that integrates over a finite subset of the support of the $dF(z_i)$ may be valid. For example, the same approach as above can be used when the support of $dF(z_i)$ is countable, on when it is continuous but there exists a subset of values z such that $\mathbb{P}(z_i = z) > 0$. Define $\mathcal{A} = \{z\,;\, dF(x,z) > 0, \forall x \in (x^-, x^+) \cap \mathcal{X}\,,\, \mathbb{P}(x_i = \bar{x}\,,\, z_i = z) > 0\,,$ and $\mathbb{P}(x_i \in (x^-, x^+) \backslash \{\bar{x}\}\,,\, z_i = z) > 0\}$. Define $\bar{\mathcal{A}} = \{z^1, \ldots, z^m\} \subset \mathcal{A}$. As long as $\bar{\mathcal{A}} \neq \emptyset$, the procedure and the results hold exactly as in the case above.

3 An application to the effects of maternal smoking in birth weight

The effects of smoking during the pregnancy, known as "maternal smoking," on the birthweight of the child is an important topic of research in the medical literature, both because birthweight is seen as the primary measure of the newborn's health and as an excellent predictor of infant's survival and development (see Almond et al. (2005) p. 1032), but also because early studies in the effects of smoking in birthweight claimed impressive effects in the ballpark of 500 grams (see Sexton and Hebel (1984)).

Let the variable CIG represent the average cigarettes smoked per day by the mother during pregnancy, BW be the weight of the child at birth, and z represent a set of d covariates which include detailed information about the mother, the father and the pregnancy. The interest is to uncover the causal relation between CIG and BW , which is expressed in the model

$$BW_i = m(CIG_i, z_i, q_i) + \varepsilon_i. \tag{23}$$

This relation is identified if $m(CIG_i, z_i, q_i) = m(CIG_i, z_i, 0)$ with probability one. Otherwise, further measures must be taken to account for the presence of q_i, such as searching for more complete datasets where hopefully q_i can be observed, searching for instrumental variables, proxy variables etc.

The effect of maternal smoking in birth weight is an example where experiments that randomly and directly change the quantities smoked by the mothers cannot be generated for ethical reasons. Randomized trials in the field try to influence the amounts smoked indirectly through some kind of propaganda[5] directed to a randomly selected part of the sample. A case can be made in favor of the reduced form effect on birth weight, which consists in that the true parameter of interest is not the effect of smoking on birth weight, but rather the effect of the smoke-related intervention on birth weight. However, propaganda can be of many different kinds, and may have radically different effects in different parts of the population depending on its content, way of transmission and scope. In this case, the effect of one kind of propaganda on birth weight is not necessarily a good predictor of the effects of other kinds of propaganda,

[5]The word propaganda will be used here to denote the set of smoke-related interventions that were randomly provided in such studies, such as informational phone calls, house visits etc.

and therefore the effects of smoking may constitute a better source of information to extrapolate between different options of public policy. Additionally, the smoking rates can be affected not only through public policy, but also through medical recommendations, which reinforces even further the relevance of knowing the effect of smoking versus the effect of smoke-related interventions. Moreover, studies that use propaganda to influence smoking behavior may also affect birthweight through other means, by providing information or raising health concerns that can make *all* pregnant women (including those who did not quit smoking) to change other behaviors. If the actual direct effect of smoking in birth weight is small although the estimated reduced form effect is large, then policy and medical attention directed at smoking may have comparatively less effect than the same resources directed at changing other habits, such as promoting propaganda for pregnant women to stop drinking, to eat better or to have more frequent prenatal visits.

According to the Cochrane Review (see Lumley, Chamberlain, Dowswell, Oliver, Oakley, and Watson (2009)), in randomized trials the smoking cessation interventions had on average a significant but imprecisely estimated effect on birth weight. On average 6 out of 100 mothers quit smoking because of the intervention, and the average reduced-form effect of the intervention is 55 grams, with a 95% confidence interval between 10 grams and 90 grams, which implies that the effect of smoking cessation on birth weight is around 915 grams, with a 95% confidence interval between 167 grams and 1500 grams. Sexton and Hebel (1984), one of the most well-known among such studies, shows a great effect in smoking cessation (20% people quit smoking because of the intervention) and a reduced-form effect of 93 grams, with a 95% confidence interval between 15 and 170 grams, implying an effect of smoking cessation between 77 and 845 grams. These imprecise estimates present an even more ambiguous picture with regard to the relative importance of smoking cessation and smoke-related interventions.[6]

Due to the difficulties associated with experimental studies mentioned above, the literature in the field has focused in non-experimental data sources where large samples and a wide array of control variables is observed. All these studies rely on an assumption of selection on observables. Therefore, a test of endogeneity without instruments is more than a convenience, it is a necessity not only because it can help in the detection of endogeneity, but it can also contribute to validate a certain choice of covariates over another.

Almond et al. (2005) is, to the author's knowledge, the most exhaustive analysis in this question using non-experimental data.[7] More specifically, they estimate the

[6] These imprecise estimates seem to be mostly due to small samples. The Cochrane Review (Lumley et al. (2009)), a systematic review of the field concerning only experimental studies, analyzes 72 trials, which amount to a total sample size of just over 25,000 observations, with on average around 350 observations per study.

[7] Almond et al. (2005) provide a detailed analysis of the costs of LBW using two independent empirical approaches, each employing a different source of variation on birthweight. The first approach uses variation of birthweight across twins in order to control for determinants of birthweight that are constant within a family, such as maternal smoking and gestation period. The second approach, of interest to this application, uses only singletons and explores variation on birth weight due to maternal smoking, which vary across

difference in the birthweight and probability of LBW between women who smoked and women who did not smoke during pregnancy, using the population of births of singletons from Pennsylvania from 1989 to 1991 and controlling for a rich set of covariates. Almond et al. (2005) directly compare nonsmoking with smoking mothers, disregarding the actual quantities smoked. They find that the children born of smoking mothers weigh 200 grams less than those of nonsmoking mothers, with a 95% confidence interval between 199 and 207 grams. For the case of LBW, they found that children of smoking mothers are 3.5% more likely to be of LBW than those of nonsmoking mothers, with a 95% confidence interval between 3.3% and 3.7%.

The remaining of this section will apply the discontinuity test to the full specification in Almond et al. (2005), using the same data set as in that paper, which is the annual, linked birth and infant death micro data produced by the National Center for Health Statistics (NCHS). This rich data set contains information for every newborn in Pennsylvania between 1989 and 1991 (488,144 observations, 94,205 smokers) such as mother's and father's demographic characteristics, mother's behaviors during pregnancy, mother's health history and risk factors, sex of the newborn, birth order of the newborn and whether the newborn was part of a multiple birth (i.e., whether the newborn is a singleton). The data also contains relevant information such as mother's and father's age, level of education and race, mother's marital status, foreign born status, number of previous live births and number of previous births where the newborn died. Other information includes maternal risk factors that are believed not to be affected by pregnancy smoking such as chronic hypertension, cardiac disease, lung disease and diabetes. Finally, the data has information related to maternal behavior such as number and timing of prenatal visits, whether the mother drinks and with which frequency, and number of cigarettes smoked per day.[8] In the context of the discontinuity test, $CIG = x$, $BW = y$, and the covariates chosen for each specification are denoted z.

The argument for the applicability of the discontinuity test in the case of the effects of maternal smoking in birthweight depends on two crucial assumptions. The first is that the effect of smoking in birthweight is continuous. In equation (23), it means that m is continuous in CIG .

The second crucial assumption is that any unobservable variable correlated with CIG conditional on z has a distribution conditional on CIG and z that is discontinuous in CIG at a certain value of CIG. Since CIG cannot be negative, a candidate to be such a threshold is $CIG = 0$. In more empirical terms, the requirement is that the mothers that did not smoke during pregnancy have to be discontinuously different with regard to the unobservable variable from the mothers that smoked, even conditional on the covariates z.

Though this cannot be confirmed for the unobservable variables q, this phenomenon can be tested for the observable covariates z.[9] The test would be essentially the same as

families. The authors use both regression adjusted methods and subclassification on the propensity score to control for potential endogeneity due to family unobservables.

[8] For a full list of the variables used, see note 36 of Almond et al. (2005) in p.1064.

[9] This heuristic evidence is analogous to the evidence provided in the applied Regression Discontinuity

the discontinuity test, except that it would be performed in a variable z^s, $s = 1, \ldots, d$, instead of in the dependent variable y, and using the rest of the covariates as controls. No matter which continuous function of z^s is tested in the discontinuity test, if the distribution of z^s conditional on CIG and the rest of the covariates is continuous, so should be the function. If some of the z^s were found to be discontinuous at $CIG = 0$, this would be understood as evidence that an unobservable correlated with cigarettes is also discontinuous at $CIG = 0$.[10]

The following figures provide heuristic evidence that the expectation of the observable covariates conditional on CIG is discontinuous at $CIG = 0$.[11] The figures were cropped at $CIG = 40$, although the observed CIG goes up to 98. However, $CIG = 41$ to 98 account for only 0.05% of the full sample, and 0.2% of the mothers that smoked positive amounts. Table 3 in appendix C shows the number of observations for each level of CIG . The dots correspond to the averages per CIG level and the lines show the 95% confidence interval of the mean per CIG level for low levels of CIG only.[12] Figures 2, 3, 4 and 5 are examples of covariates referring to the mother's demographic variables where there is a clear difference in the averages per level of CIG for zero versus just above zero cigarettes. Figure 2 shows the mother's education in years, with fairly constant averages of a little below 12 years for $0 < CIG < 8$, and increasing one full year of education for $CIG = 0$. Figure 3 shows the mother's age, which averages around 25 years old for low-level smoking mothers, and increases to 27 among the nonsmoking mothers. The marital status shifts from 50% of unmarried low-level smoking mothers to only 24% of unmarried nonsmoking mothers. The proportion of black women among the women surveyed has higher variation, but is constantly above 20%, and often closer to 30% for low-level smokers, and is only 14% for the nonsmokers.

The father's demographic variables present even higher differences for low-level smoking mothers relative to nonsmoking mothers. The education level, shown in figure 6, changes from below 11 years among the fathers of

children of low-level smoking mothers to 12.6 years for fathers of children of nonsmoking mothers, increasing more than 1.5 years of education. Figure 7 shows that the average age of the father is at most 25 years old for $0 < CIG < 10$, but increases to an average of 28 years of age among the fathers of children of nonsmoking mothers.

literature that covariates are continuous at the threshold, suggesting that unobservables are also continuous at the threshold. See Lee (2008) for an example.

[10]This is true unless z^s is a proxy for such a variable, in which case there is no identification issue in the first place.

[11]It is not possible to guarantee from these figures that any of the variables below would still be discontinuous conditional on the rest of the covariates. However, if the figures below are taken as evidence of discontinuities in the expectation of some of the z^s conditional on CIG, then at least one of the z^s has to be discontinuous in CIG conditional on the rest of the covariates.

[12]The confidence intervals are shown only for small levels of cigarettes for exposition reasons. The confidence intervals become larger for $CIG > 20$, which correspond to only 1% of the full sample (6% of the sample of smokers). As suggested by table 3 of appendix C, the confidence intervals increase as a reflection of the smaller sample sizes per CIG value. For values of CIG between 26 and 29, 31 and 34 and 36 and 39, sample size are extremely small per level of CIG , rarely above 10 and never above 15 observations.

Mother's demographic characteristics.

Figure 2: Education (years)

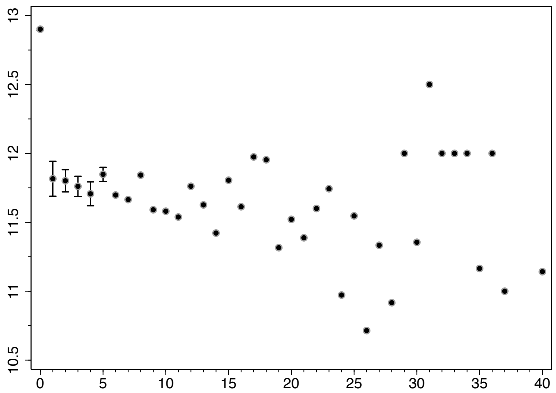


Figure 3: Age

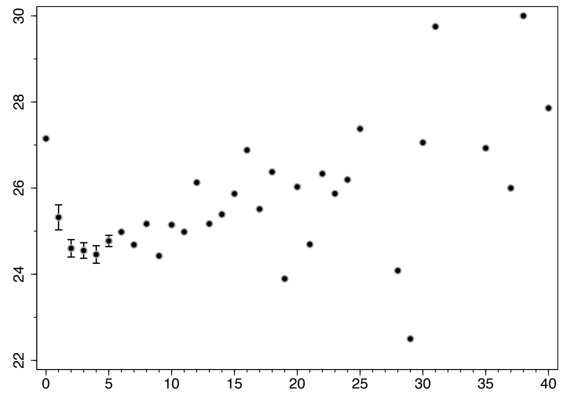


Figure 4: Unmarried

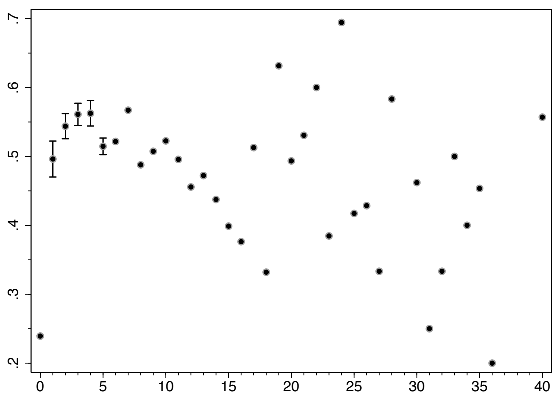
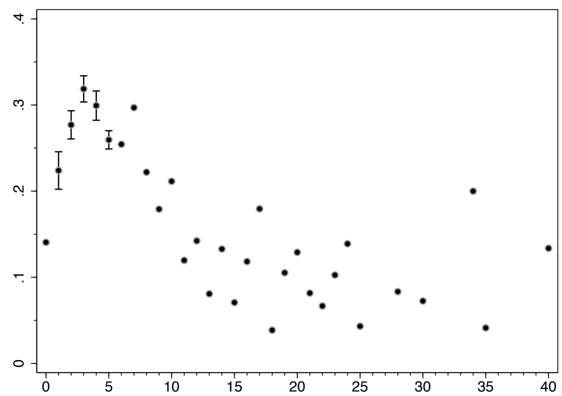


Figure 5: Race: black



Father's demographic characteristics.

Figure 6: Education (years)

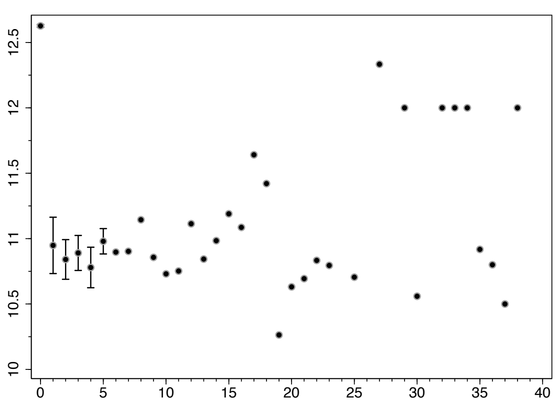
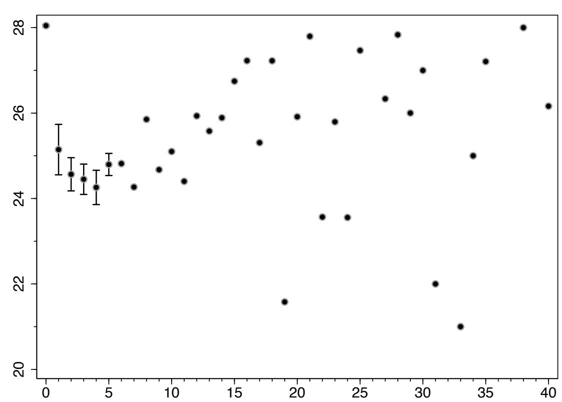


Figure 7: Age



Figures 2 to 7: Dots represent average values referring to the pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean.

The behavioral characteristics of the mother also seem to change significantly when comparing low-level smoking mothers to nonsmoking mothers. Around 10% of the

Mother's behavior variables.

Figure 8: Consumed alcohol

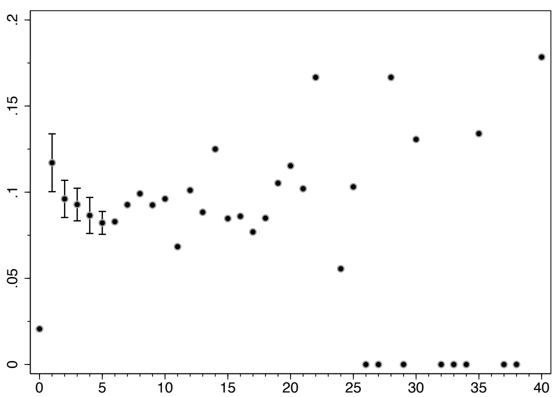
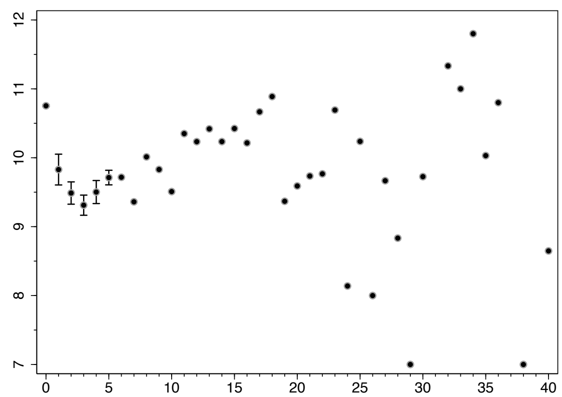


Figure 9: Number of prenatal visits



Figures 8 and 9: Dots represent average values among pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean.

mothers consumed alcohol during pregnancy for all smoking levels until CIG=20, while only 2% of the nonsmoking mothers did the same. Low-level smoking mothers on average visited doctors for prenatal visits around 10 times, which is one less time than in the case of nonsmoking mothers.

Although the behavior of mothers seem to be discontinuously different at zero

Mother's behavior covariates.

Figure 10: Gender of Newborn

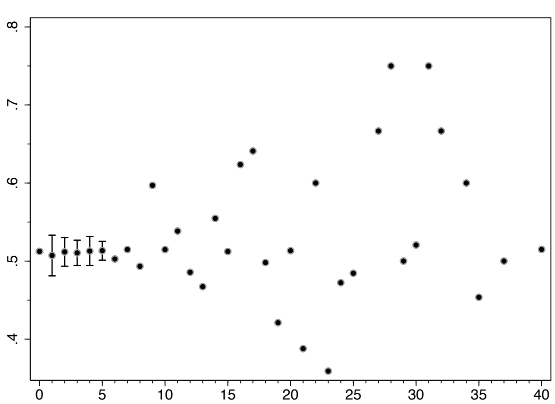
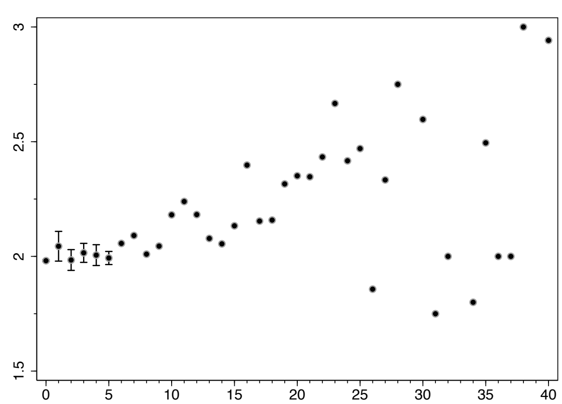


Figure 11: Order of Newborn



Figures 10 and 11: Dots represent average values among pregnant mothers for each level of daily cigarette consumption. The vertical lines represent the 95% confidence interval of the mean. Order of Newborn in figure 11 represents the order among live births.

cigarettes, some contingencies that may have an influence on mother's behavior during pregnancy were not found to be discontinuous at zero cigarettes, such as the gender of the newborn (constant around 50% of males) and the birth order of the newborn (constant around the average of second birth).

The two outcome variables tested were birthweight and the probability of low birthweight (LBW), defined as weight below 2500 grams. However, for simplicity the notation of the outcome variable in the rest of this section will remain BW. The test was performed for the dependent variables without the covariates (specification I), and for the most complete specification provided in Almond et al. (2005).

For the implementation of the test, it is assumed that for $CIG_i > 0$,

$$\mathbb{E}(BW_i \mid CIG_i, z_i) = \tau(CIG_i) + z_i^T \gamma.$$

This is equivalent to equation (11) with the superscripts "+" omitted, since in this application $\bar{x} = 0$ is the left boundary point. Though this specification is not as flexible as a fully nonparametric approach, it allows the use of a high number of control variables, and therefore the immediate comparison with the most complete specification of Almond et al. (2005). The test statistic is calculated as in section 2.3.2. The only step which is not specified in the description is the estimator of $\mathbb{E}(BW_i \mid CIG_i)$ and $\mathbb{E}(z_i \mid CIG_i)$ used in the estimation of γ, though some requirements about this estimator are made in assumption 2.6 (2). Most kernel based estimators such as the Nadaraya-Watson or the local polynomial, as well as series estimators satisfy these requirements under roughly the same conditions, but the kernel-based techniques require one regression per different value of CIG_i in the sample, while the series estimators requires only one regression for the estimation of all the values required in the estimation of γ. A series estimator was therefore preferred over the kernel-based for practical reasons. The basis chosen is that of cubic B-splines, which have better local properties than classic global bases such as Fourier or power series. The knots of the spline basis were chosen to be $0.5, 3.5, 6.5, 9.5, 12.5, 17.5, 27.5, 37.5, 47.5, \ldots, 107.5$. Many other combinations of knots were attempted with virtually identical results. Let the $\rho_j(CIG_i)$ represent the j-th element in the basis evaluated at CIG_i , and let ρ be the matrix whose rows are $(\rho_1(CIG_i)\mathbf{1}(CIG_i > 0), \ldots, \rho_N(CIG_i)\mathbf{1}(CIG_i > 0))$, where N is the number of elements of the basis used in the regression. Let $P_\rho^+ = \rho(\rho^T\rho)^{-1}\rho^T$ and I^+ be the $n \times n$ diagonal matrix with $\{\mathbf{1}(CIG_1 > 0), \ldots, \mathbf{1}(CIG_n > 0)\}$ in the diagonal. The estimator of the variance matrix of $\hat{\gamma}$ is given by

$$\hat{\mathcal{V}}_\gamma = n^+ \, (Z^T(I^+ - P_\rho^+)Z)^{-1} Z^T(I^+ - P_\rho^+)\hat{\Sigma}(I^+ - P_\rho^+)Z(Z^T(I^+ - P_\rho^+)Z)^{-1},$$

where $n^+ = \sum_{i=1}^n \mathbf{1}(CIG_i > 0)$. $\hat{\mathcal{V}}_\gamma$ is the Eicker-White covariance matrix of an OLS regression of $(I^+ - P_\rho^+)Y$ on $(I^+ - P_\rho^+)Z$. This can be useful if the researcher intends to estimate the standard errors using theorem 6 (see Li (2000) for the asymptotic behavior of the estimator of the parametric term in the partially linear model using series plugins). This paper reports standard errors acquired instead by a bootstrap approach, which will be described later. An important restriction on the covariate list is that Z does not contain a constant term, because as can be seen in equation (13), in the partially linear models the constant term cannot be identified separately from $\tau(CIG)$.

For the local polynomial step, as described in equation (15), the choice parameters are the kernel, the degree of the polynomial, and the bandwidth size. The kernel used is epanechnikov (rectangular and triangular kernels were also tested) and the polynomial degree is 3, although degrees 2 and 1 were also tested with very similar results. The bandwidth was chosen by a cross-validation technique, which consisted in the estimation of $\tau(CIG)$ for $CIG = 1, \ldots, 20$ by a local polynomial regression of the $BW_i - z_i^T \hat{\gamma}$ using only observations for which $CIG_i > 0$, and $CIG_i \neq CIG$, for each bandwidth $h = 2, 3, ..., 20$, which yielded the values $\hat{\tau}_h(CIG)$, $h = 1, \ldots, 20$, $CIG = 1, \ldots, 20$. The chosen h^* is the one that satisfies

$$h^* = arg \min_{h=1,\ldots,20} \sum_{i=1}^{n} \left(BW_i - z_i^T \hat{\gamma} - \hat{\tau}_h(CIG_i)\right)^2 \mathbf{1}(0 < CIG_i \leqslant 20).$$

The bandwidth that performed the best was $h = 2$, corresponding to roughly 1.5% of the observations such that $CIC > 0$, followed by $h = 3$, $h = 10$, $h = 11$ and $h = 6$, corresponding to 5%, 26%, 60% and 19% of the observations such that $CIG > 0$ respectively.

Tables 1 and 2 show the discontinuity test results for the birth weight and the probability of LBW equations. As stated previously, the standard errors were estimated by a bootstrap approach, which consisted in drawing 500 bootstrap samples of the data, and calculating $\hat{\theta}$ for each of those independently, exactly in the same way described above. The resulting standard deviations of the 500 values of $\hat{\theta}$ are the standard errors reported in tables 1 and 2.

The results in table 1 present strong evidence of endogeneity for all specifications of birth weight and for all bandwidths. Table 2 indicates only weak evidence of endogeneity for the main specification of the probability of LBW and the preferred bandwidths ($h = 2$ and $h = 3$), and no evidence of endogeneity for larger bandwidths. For $h = 2$ and $h = 3$, specification II is rejected with 95% confidence but not rejected with 99% of confidence, and for all other bandwidths specification II is not rejected even with 90% confidence.

Table 1:					**Table 2:**				
Birthweight					**$\mathbb{P}$(Birthweight<2500g)**				
	C.V.		**I**	**II**		**C.V.**		**I**	**II**
h=2	1	$\hat{\boldsymbol{\theta}}$	196**	121**	**h=2**	1	$\hat{\boldsymbol{\theta}}$	-0.043**	-0.016*
(1.5%)		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(14)	(14)	**(1.5%)**		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(0.007)	(0.007)
h=3	2	$\hat{\boldsymbol{\theta}}$	194**	121**	**h=3**	2	$\hat{\boldsymbol{\theta}}$	-0.043**	-0.016*
(5%)		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(17)	(17)	**(5%)**		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(0.008)	(0.008)
h=6	5	$\hat{\boldsymbol{\theta}}$	199**	145*	**h=6**	5	$\hat{\boldsymbol{\theta}}$	-0.041*	-0.019
(19%)		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(52)	(62)	**(19%)**		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(0.017)	(0.027)
h=10	3	$\hat{\boldsymbol{\theta}}$	178**	140**	**h=10**	3	$\hat{\boldsymbol{\theta}}$	-0.037**	-0.023
(26%)		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(30)	(32)	**(26%)**		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(0.014)	(0.015)
h=11	4	$\hat{\boldsymbol{\theta}}$	176**	122**	**h=11**	4	$\hat{\boldsymbol{\theta}}$	-0.040**	-0.022
(60%)		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(25)	(25)	**(60%)**		$(\mathbb{SE}(\hat{\boldsymbol{\theta}}))$	(0.012)	(0.013)

Tables 1 and 2: In the first column, h is the bandwidth, and the percentage in parenthesis is the proportion of the sample of smokers used in the local polynomial regression for each value of the bandwidth. C.V. shows the position of the bandwidth in the cross-validation results. $\hat{\theta}$ is the discontinuity test statistic. The standard errors are the result of a bootstrap of the original sample with 200 repetitions. Specification I has no covariates and II is the same specification used in Almond et al. (2005) (see footnote 8). "**" means that the discontinuity test rejects at the 99% confidence level, "*" means that the test rejects at the 95% confidence level, but not at the 99% confidence level.

Figures 12 and 13 depict the main results from tables 1 and 2, respectively. Figure 12 shows the average birth weight for each level of CIG (black dots) and two other marks at zero cigarettes. The hollow dot and the "×" point represent the predicted birth weight at zero cigarettes using specification I and II respectively. It can be seen in figure 12 that the covariates of specification II help reduce the discontinuity of actual birthweight and predicted birthweight, but not enough for it to vanish.

Figure 13, which depicts the results for the probability of LBW analogously to 12, shows that the covariates of specification II help reduce the discontinuity of actual LBW and predicted LBW to a third of its original value. The results for the probability of LBW show that the discontinuity is small, so that if there is endogeneity in specification II, it is of low importance for LBW.

Figure 12: Birth Weight, Specification II

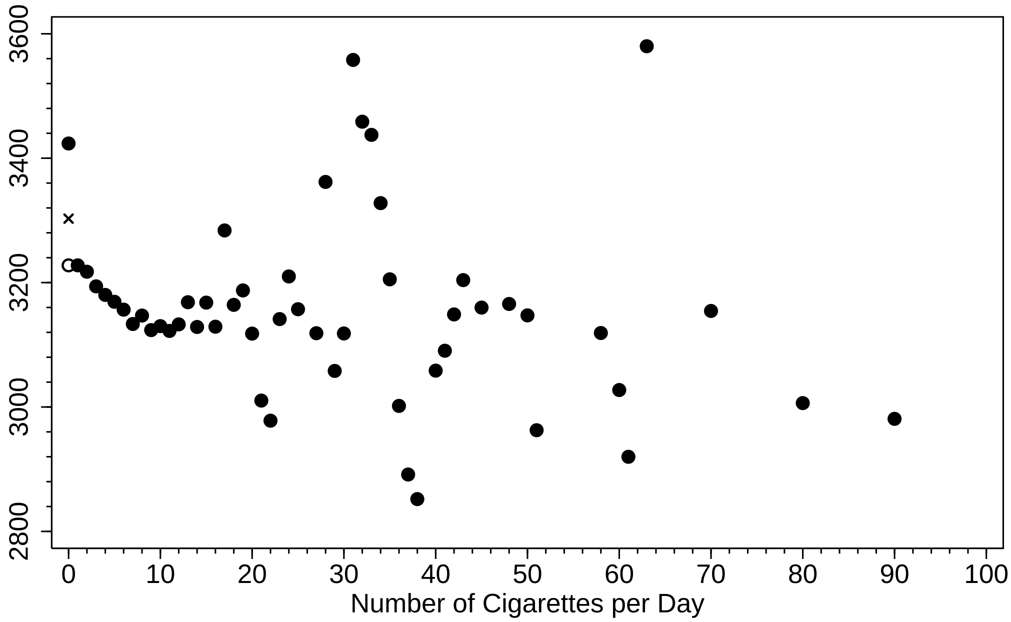
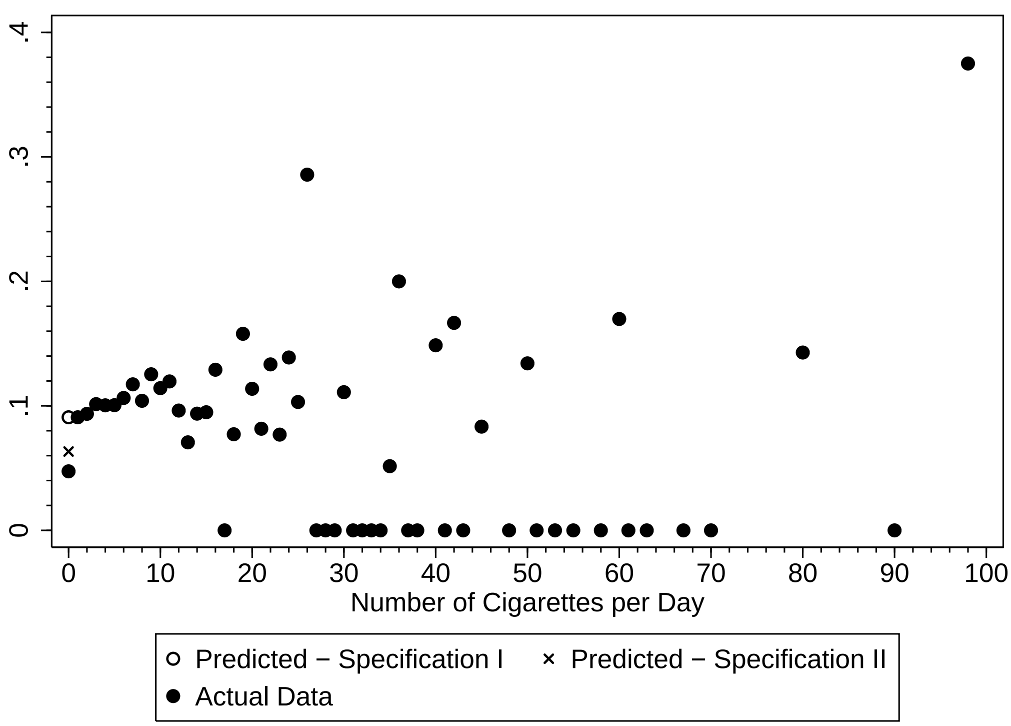


Figure 13: Probability of Low BW, Specification II



The solid dots in figure 12 represent average birth weights among the pregnant mothers for each level of daily cigarette consumption. The hollow dot is the local polynomial predictor of the birth weight at zero cigarettes. The point "×" is the predicted birth weight after the effect of the covariates is removed. The solid and hollow dots, as well as the "×" point in figure 13 represent the same as in figure 12, but for the incidence of LBW (birth weight <2500 g) instead of the birth weight itself. In both cases, $\hat{\theta}$ is the difference between the $\times$ and the solid points at $CIG = 0$.

4 Conclusion

This paper develops a nonparametric test of endogeneity which does not require instrumental variables. The two crucial assumptions for the applicability of the test are that the distribution of the unobservable conditional on the observable variables be discontinuous in the running variable at a given threshold, and that the structural equation relating the dependent variable and the observables be continuous in the running variable. The test consists in estimating such discontinuities, averaging them over a given distribution of the covariates, and then testing for whether this average is equal to zero.

The paper provides test statistics and asymptotic distributions for the average of the discontinuities interacted with arbitrary functions of the covariates, averaged over the distribution of the covariates at the threshold. This type of test eliminates one step in the estimation process. The estimation of the discontinuities is done for three different specifications of the conditional expectation of the dependent variable given the covariates when the running variable is different than the threshold. The first assumes that it is linear, the second that it is partially linear (nonparametric in the running variable and additively linear in the covariates), and the third that it is fully nonparametric, although with certain smoothness conditions. The test statistic is shown to converge at $\sqrt{n}$ rates in the linear case, and at the rate $\sqrt{nh}$ in the partially linear and nonparametric cases. This rate is the same as that of a nonparametric regression with a single right-hand side variable, and this is achieved in spite of the presence of the covariates due to the aggregation over the measure of the covariates.

The estimation has to be sensible to the boundary nature of the threshold, even when it is not in fact a boundary point in the domain of the running variable. This is the case because this paper allows for the functional forms, as well as conditional distribution functions, variances etc., to be different at the right and left sides of the threshold. Hence, the threshold is treated as a boundary point in all cases, and estimation has to be mindful of boundary biases. The nonparametric estimators use the local polynomial method, known for its automatic boundary carpentry and low bias. In that regard this paper is in accordance with the regression discontinuity literature, which uses the same kind of estimator. However that literature also assumes that the probability densities are continuous across the threshold, while this paper allows for differences at the different sides of the threshold, which requires different estimators for the variance of the estimator.

The test was applied to the estimation of the effects of maternal smoking in both birth weight and the probability of low birthweight, an example where many covariates are shown to be discontinuous at the threshold of zero cigarettes. The unobservable, if existent, is assumed to be discontinuous at zero cigarettes conditional on the covariates, and the effect of smoking on birthweight and on the probability of LBW is assumed to be continuous. Using the same data set as in Almond et al. (2005), the discontinuity test was performed on and their most complete specification. For the case of birth weight, the test shows strong evidence of endogeneity for all bandwidths. For the probability of LBW the evidence of endogeneity is weak, only existent at the 95% confidence level for the optimal bandwidth according to the cross-validation technique, and not existent for all the other bandwidths or when the confidence level is 99%.

One of the two crucial assumptions of the discontinuity test of endogeneity may not be valid in the case of maternal smoking: that the effect of smoking on birth weight or in the probability of LBW is continuous at zero cigarettes. If that is the case, then one cannot disentangle the part of the discontinuity found in the test that is due to the discontinuous treatment effect and the part that is due to the endogeneity. In the results shown in this paper, the discontinuities become smaller when more covariates are added, which may be an indication that at least part of the discontinuities are due to endogeneity, hence suggesting the necessity of even better data sets or of the search for quasi-experimental variations of smoking.

References

D. Almond, K.Y. Chay, and D.S. Lee. The Costs of Low Birth Weight*. The Quarterly Journal of Economics, 120(3):1031–1083, 2005. Cited on pages 1, 6, 30, 31, 32, 36, 38, and 40.

Patrick Billingsley. Probability & Measure. Wiley, New York, 3rd. edition, 1995. Cited on page 57.

R. Blundell and J.L. Powell. Endogeneity in nonparametric and semiparametric regression models. In Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, volume 2, pages 655–679, 2003. Cited on page 2.

Richard Blundell, Xiaohong Chen, and Dennis Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. Econometrica, 75(6):1613–1669, 2007. Cited on page 2.

R.W. Blundell and J.L. Horowitz. A Non-parametric test of exogeneity. Review of Economic Studies, 74(4):1035–1058, 2007. Cited on page 2.

M.D. Cattaneo. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. Journal of Econometrics, 2009. Cited on page 6.

X. Chen and Y. Fan. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics, 91(2):373–401, 1999. Cited on page 2.

Y.S. Chow and H. Teicher. Probability theory: independence, interchangeability, martingales. Springer Verlag, New York, 3rd edition, 1997. Cited on page 57.

Serge Darolles, Jean-Pierre Florens, and Eric Renault. Non parametric instrumental regression. IDEI Working Papers 228, Institut d'Économie Industrielle (IDEI), Toulouse, 2003. Cited on page 2.

J. Fan and I. Gijbels. Local polynomial modelling and its applications. Chapman & Hall, 1996. Cited on pages 4 and 20.

Yanqin Fan and Qi Li. Consistent model specification tests: Omitted variables and semiparametric functional forms. Econometrica, 64(4):865–890, 1996. Cited on page 2.

Peter Hall and Joel L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. The Annals of Statistics, 33(6):2904–2929, 2005. Cited on page 2.

Joel L. Horowitz. Applied nonparametric instrumental variables estimation. Department of Economics, Northwestern University, June 2009. Cited on page 2.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142(2):615–635, February 2008. Cited on page 5.

MC Jones. Simple boundary correction for kernel density estimation. Statistics and Computing, 3(3):135–146, 1993. Cited on page 25.

D.S. Lee. Randomized experiments from non-random selection in US House elections. Journal of Econometrics, 142(2):675–697, 2008. Cited on pages 4 and 33.

M. Lejeune and P. Sarda. Smooth estimators of distribution and density functions. Computational Statistics & Data Analysis, 14(4):457–471, 1992. Cited on pages 25 and 29.

Q. Li and J.S. Racine. Nonparametric econometrics: theory and practice. Princeton Univ Press, 2007. Cited on page 2.

Qi Li. Efficient estimation of additive partially linear models. International Economic Review, 41(4):1073–1092, 2000. Cited on pages 19 and 36.

O. Linton and J.P. Nielsen. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika, 82(1):93–100, 1995. Cited on page 4.

Oliver Linton. Second order approximation in the partially linear regression model. Econometrica, 63(5):1079–1112, 1995. Cited on page 19.

J. Lumley, C. Chamberlain, T. Dowswell, S. Oliver, L. Oakley, and L. Watson. Interventions for promoting smoking cessation during pregnancy (Cochrane Review). The Cochrane Library, 8(3), July 2009. Cited on page 31.

E. Masry. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis, 17(6):571–600, 1996. Cited on page 55.

Whitney K. Newey and James L. Powell. Instrumental variable estimation of nonparametric models. Econometrica, 71(5):1565–1578, 09 2003. Cited on page 2.

W.K. Newey. Kernel estimation of partial means and a general variance estimator. Econometric Theory, pages 233–253, 1994. Cited on page 4.

J. Porter. Estimation in the regression discontinuity model. Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison, 2003. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.133.540. Cited on pages 5, 19, 20, 27, 47, 48, and 54.

PM Robinson. Root-N-consistent semiparametric regression. Econometrica, 56(4): 931–954, 1988. Cited on page 18.

D. Ruppert and MP Wand. Multivariate locally weighted least squares regression. The annals of statistics, pages 1346–1370, 1994. Cited on page 52.

M. Sexton and R. Hebel. A clinical trial of change in maternal smoking and its effect on birth weight. JAMA, 251(7), Feb 1984. Cited on pages 30 and 31.

A Proof of theorems

A.1 Identification theorems

A.1.1 Theorem 1:

The proof requires the following assumption:

Assumption A.1.

$$\lim_{x\downarrow\bar{x}} \int \mathbb{E}(y\,|\,x,z,q)\,dF(q\,|\,x,z) = \int \mathbb{E}(y\,|\,\bar{x},z,q)\lim_{x\downarrow\bar{x}} dF(q\,|\,x,z),\ \textit{and}$$

$$\lim_{x\uparrow\bar{x}} \int \mathbb{E}(y\,|\,x,z,q)\,dF(q\,|\,x,z) = \int \mathbb{E}(y\,|\,\bar{x},z,q)\lim_{x\uparrow\bar{x}} dF(q\,|\,x,z).$$

Proof. First, observe that assumption 2.1 (3) assures that $\mathbb{E}(y \mid x = \bar{x}, z)$, $\lim_{x\downarrow\bar{x}} \mathbb{E}(y\,|\,x,z)$ and $\lim_{x\uparrow\bar{x}} \mathbb{E}(y\,|\,x,z)$ are identified for all $z \in \mathcal{Z}_{\bar{x}}$, unless $\bar{x}$ is a boundary point, in which case either the right or left limit will not be identified. However, item (2) assures $\Delta(z)$ will be identified, because when one of its parts is not identified, α is such that the part is null. Identification of θ follows because G is known and ν is identified.

From equation (1),

$$\theta = \int G\Big(\Big[\int \mathbb{E}(y\,|\,x=\bar{x},z,q)dF(q\,|\,x=\bar{x},z) - \\ - \alpha\lim_{x\downarrow\bar{x}}\int \mathbb{E}(y\,|\,x,z,q)dF(q\,|\,x,z) - (1-\alpha)\lim_{x\uparrow\bar{x}}\int \mathbb{E}(y\,|\,x,z,q)dF(q\,|\,x,z)\Big],z\Big)d\nu(z)$$

If x is exogenous, $\int \mathbb{E}(y \,|\, x, z, q) dF(q \,|\, x, z) = \mathbb{E}(y \,|\, x, z) \int dF(q \,|\, x, z) = \mathbb{E}(y \,|\, x, z)$, hence $\theta = \int G\left([\mathbb{E}(y \,|\, x = \bar{x}, z) - \alpha \lim_{x \downarrow \bar{x}} \mathbb{E}(y \,|\, x, z) - (1-\alpha) \lim_{x \uparrow \bar{x}} \mathbb{E}(y \,|\, x, z)], z\right) d\nu(z)$. From assumption 2.1 (1), $\mathbb{E}(y \,|\, x, z)$ is continuous, and therefore $\theta = \int G(0, z)\, d\nu(z) = 0$. $\square$

A.1.2 Proof of Remark 2.2.5:

f continuous in x at $\bar{x}$ implies that $\forall \epsilon > 0$, $\exists\, \delta > 0$ such that $|x - \bar{x}| < \delta \implies |f(x, z, q, \varepsilon) - f(\bar{x}, z, q, \varepsilon)| < \epsilon$. Hence,

$$\begin{aligned}
|\mathbb{E}(y \mid x, z, q) - \mathbb{E}(y \mid x = \bar{x}, z, q)| =& \Big| \int f(x, z, q, \varepsilon)\, dF(\varepsilon \mid x, z, q) \\
& - \int f(\bar{x}, z, q, \varepsilon)\, dF(\varepsilon \mid x = \bar{x}, z, q) \Big| \\
=& \left| \int (f(x, z, q, \varepsilon) - f(\bar{x}, z, q, \varepsilon))\, dF(\varepsilon) \right| \\
\leqslant& \int |f(x, z, q, \varepsilon) - f(\bar{x}, z, q, \varepsilon)|\, dF(\varepsilon) < \epsilon.
\end{aligned}$$

A.2 Estimation in the linear case

A.2.1 Theorem 2:

First, observe that

$$\sqrt{n} \begin{bmatrix} \hat{\delta}^+ - \delta^+ \\ \hat{\delta}^- - \delta^- \end{bmatrix} = \begin{bmatrix} \left[\frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i > \bar{x})\right]^{-1} & 0 \\ 0 & \left[\frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i < \bar{x})\right]^{-1} \end{bmatrix} \cdot \\
\cdot \quad \sqrt{n}\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} X_i\, \epsilon_i\, \mathbf{1}(x_i > \bar{x}) \\ X_i\, \epsilon_i\, \mathbf{1}(x_i < \bar{x}) \end{bmatrix}.$$

Assumptions 2.4 (1), assumption 2.5 (3), the LLN and the continuous mapping theorem guarantee that $\left[\frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i > \bar{x})\right]^{-1} \xrightarrow{p} \mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x}))^{-1}$ and $\left[\frac{1}{n}\sum_{i=1}^n X_i X_i^T \mathbf{1}(x_i < \bar{x})\right]^{-1} \xrightarrow{p} \mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x}))^{-1}$. Since the ϵ_i are functions of y_i, x_i and z_i, they are i.i.d. Moreover,

$$\begin{aligned}
\mathbb{C}ov(X_i\, \epsilon_i\, \mathbf{1}(x_i > \bar{x}), X_i\, \epsilon_i\, \mathbf{1}(x_i < \bar{x})) &= \mathbb{E}(X_i\, \epsilon_i\, \mathbf{1}(x_i > \bar{x}))\mathbb{E}(X_i\, \epsilon_i\, \mathbf{1}(x_i < \bar{x})) \\
&= \mathbb{E}(X_i\, \mathbb{E}(\epsilon_i \mid X_i)\, \mathbf{1}(x_i > \bar{x}))\mathbb{E}(X_i\, \mathbb{E}(\epsilon_i \mid X_i)\, \mathbf{1}(x_i < \bar{x})) = 0.
\end{aligned}$$

Therefore, assumptions 2.4 (1) and 2.5 (2) and the vector CLT imply that

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^n \begin{bmatrix} X_i\, \epsilon_i\, \mathbf{1}(x_i > \bar{x}) \\ X_i\, \epsilon_i\, \mathbf{1}(x_i < \bar{x}) \end{bmatrix} \xrightarrow{d} \\
\xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbb{E}(X_i X_i^T \mathbf{1}(x_i > \bar{x})) & 0 \\ 0 & \mathbb{E}(X_i X_i^T \mathbf{1}(x_i < \bar{x})) \end{bmatrix} \right).$$

Finally, Slutsky's theorem implies that

$$\sqrt{n}\begin{bmatrix}\hat{\delta}^{+}-\delta^{+}\\ \hat{\delta}^{-}-\delta^{-}\end{bmatrix}\xrightarrow{d}\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\sigma^2\begin{bmatrix}\mathbb{E}(X_iX_i^T\mathbf{1}(x_i>\bar{x})) & 0\\ 0 & \mathbb{E}(X_iX_i^T\mathbf{1}(x_i<\bar{x}))\end{bmatrix}^{-1}\right).$$

By the continuous mapping theorem,

$$\sqrt{n}a_n := \left(\alpha\sqrt{n}(\hat{\delta}^{+}-\delta^{+})+(1-\alpha)\sqrt{n}(\hat{\delta}^{-}-\delta^{-})\right)\xrightarrow{d}\mathcal{N}(0,v),$$

where

$$v=\sigma^2\left(\alpha^2\mathbb{E}(X_iX_i^T\mathbf{1}(x_i>\bar{x}))^{-1}+(1-\alpha)^2\sigma^2\mathbb{E}(X_iX_i^T\mathbf{1}(x_i<\bar{x}))^{-1}\right).$$

By assumption 2.4 item (1), assumption 2.5 item (1) and the strong LLN, $\hat{\mathbb{E}}(g(z_i)\mid x_i=\bar{x})\xrightarrow{p}\mathbb{E}(g(z_i)\mid x_i=\bar{x})$ and $\hat{\mathbb{E}}(g(z_i)z_i\mid x_i=\bar{x})\xrightarrow{p}\mathbb{E}(g(z_i)z_i\mid x_i=\bar{x})$, and since the limits are scalar, the convergence holds for the vector. By Slutsky's theorem,

$$\sqrt{n}B_n\xrightarrow{d}\mathcal{N}(0,V_B).$$

It is easy to establish the joint convergence of A_n and B_n with the same arguments as above. Observe that since $(\hat{\delta}^{+}-\delta^{+})$ and $(\hat{\delta}^{-}-\delta^{-})$ use only data for which $x_i\neq\bar{x}$,

$$\begin{aligned}\mathbb{C}ov(A_n,B_n)=\mathbb{E}&\left(\frac{1}{\hat{p}_{\bar{x}}^2}(\Delta(z_i)g(z_i)\mathbf{1}(x_i=\bar{x})-\theta)[\bar{x}\;\; g(z_i)z_i\mathbf{1}(x_i=\bar{x})]\right)\cdot\\ &\cdot\mathbb{E}\left(\alpha(\hat{\delta}^{+}-\delta^{+})+(1-\alpha)(\hat{\delta}^{-}-\delta^{-})\right)=0\end{aligned}\tag{24}$$

because the weighted least squares estimators δ^{+} and δ^{-} are unbiased. Equation (24) and the continuous mapping theorem imply that

$$\sqrt{n}(\hat{\theta}-\theta)=\sqrt{n}A_n-\sqrt{n}B_n\xrightarrow{d}\mathcal{N}(0,V_A+V_B).$$

A.2.2 Theorem 4

The convergence of $\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\hat{V}_B}\leqslant c_\lambda\right)$ to λ as $n\to\infty$ under H_0 is a trivial consequence of theorem 3. Under H_1,

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\sqrt{\hat{V}_B}}>c_\lambda\right)=\mathbb{P}\left(\sqrt{n}\left(\frac{\hat{\theta}-\theta}{\sqrt{V_A+V_B}}\right)-\frac{c_\lambda\left(\sqrt{\hat{V}_B}-\sqrt{V_B}\right)}{\sqrt{V_A+V_B}}>\frac{c_\lambda\sqrt{V_B}}{\sqrt{V_A+V_B}}-\sqrt{n}\frac{\theta}{\sqrt{V_A+V_B}}\right)$$

From theorem 3 and the continuous mapping theorem, $\sqrt{\hat{V}_B}-\sqrt{V_B}\xrightarrow{p}0$, and therefore, by the same theorem and Slutsky's theorem, $\sqrt{n}\left(\frac{\hat{\theta}-\theta}{\sqrt{V_A+V_B}}\right)-\frac{c_\lambda\left(\sqrt{\hat{V}_B}-\sqrt{V_B}\right)}{\sqrt{V_A+V_B}}\xrightarrow{d}\mathcal{N}(0,1)$. Since $-\sqrt{n}\frac{\theta}{\sqrt{V_A+V_B}}\to-\infty$ as $n\to\infty$, it is easy to prove that $\mathbb{P}\Big(\sqrt{n}\left(\frac{\hat{\theta}-\theta}{\sqrt{V_A+V_B}}\right)-$

$\frac{c_\lambda\left(\sqrt{\hat{V}_B}-\sqrt{V_B}\right)}{\sqrt{V_A+V_B}} > \frac{c_\lambda\sqrt{V_B}}{\sqrt{V_A+V_B}} - \sqrt{n}\frac{\theta}{\sqrt{V_A+V_B}}\Big) \to 1$ as $n \to \infty$.

Under the alternatives $\theta/\sqrt{n}$, the same manipulations as above yield

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{\theta}}{\sqrt{\hat{V}_B}} \leqslant c_\lambda\right) = \mathbb{P}\left(\sqrt{n}\left(\frac{\hat{\theta}-\theta/\sqrt{n}}{\sqrt{V_A+V_B}}\right) - \frac{c_\lambda\left(\sqrt{\hat{V}_B}-\sqrt{V_B}\right)}{\sqrt{V_A+V_B}} \leqslant \frac{c_\lambda\sqrt{V_B}-\theta}{\sqrt{V_A+V_B}}\right)$$

and since $\sqrt{n}\left(\frac{\hat{\theta}-\theta/\sqrt{n}}{\sqrt{V_A+V_B}}\right) - \frac{c_\lambda\left(\sqrt{\hat{V}_B}-\sqrt{V_B}\right)}{\sqrt{V_A+V_B}} \xrightarrow{d} \mathcal{N}(0,1)$, the result of the theorem follows immediately.

A.3 Estimation in the partially linear case

A.3.1 Theorem 5:

From equation (15), equation (12) can be rewritten as

$$\begin{aligned}
&B_n = B_n^1 + B_n^2\\
&B_n^1 := \hat{\mathbb{E}}(g(z_i)\,|\,x_i=\bar{x})\left(\alpha[\tilde{\tau}^+(\bar{x})^{\lim} - \tau^+(\bar{x})^{\lim}] + (1-\alpha)[\tilde{\tau}^-(\bar{x})^{\lim} - \tau^-(\bar{x})^{\lim}]\right)\\
&B_n^2 := \alpha\left[\left(\hat{\mathbb{E}}(g(z_i)z_i\,|\,x_i=\bar{x})^T - \hat{\mathbb{E}}(g(z_i)\,|\,x_i=\bar{x})\,e_1^T(\tilde{X}^TW^+\tilde{X})^{-1}\tilde{X}^TW^+Z\right)(\hat{\gamma}^+-\gamma^+)\right] +\\
&\quad +(1-\alpha)\left[\left(\hat{\mathbb{E}}(g(z_i)z_i\,|\,x_i=\bar{x})^T - \hat{\mathbb{E}}(g(z_i)\,|\,x_i=\bar{x})\,e_1^T(\tilde{X}^TW^-\tilde{X})^{-1}\tilde{X}^TW^-Z\right)(\hat{\gamma}^- - \gamma^-)\right].
\end{aligned}$$

It will be shown that B_n^1 converges at the rate $\sqrt{nh}$ and $(A_n + B_n^2)$ converges at the rate $\sqrt{n}$, and therefore $\mathbb{V}ar(\sqrt{nh}B_n^2) = O(h)$. The consequence of the disparity between the rates is that only the variance of B_n^1 will affect the asymptotic variance of $\hat{\theta}$. However, in order to study the influence of B_n^2 in small samples, the results will consider variance terms that are at least $O(h)$, which is the same to say that any term which is $o(h)$ will be considered irrelevant in the variance calculations and not taken into account.

For the distribution of B_n^1, observe that $\tilde{\tau}^+(\bar{x})^{\lim}$ and $\tilde{\tau}^-(\bar{x})^{\lim}$ are local polynomial regressions of $y_i - z_i^T\gamma^+$ and $y_i - z_i^T\gamma^-$ on x_i using only observations for which $x_i > \bar{x}$ and $x_i < \bar{x}$ respectively. These are standard local polynomial regressions of the kind used in Porter (2003) for the estimation of the sides of the discontinuity in the regression discontinuity design. There is one crucial difference: Porter assumes that the running variable x_i has a density function in a neighborhood of $\bar{x}$. Since here $\mathbb{P}(x_i = \bar{x}) > 0$, this is no longer possible. However, by assumption 2.6 (3) the distribution function $F(x \mid x \neq \bar{x})$ has a density function in $[x^-, \bar{x}) \cap (\bar{x}, x^+]$, and it is equal to

$$\varphi(x) := \frac{\frac{d}{dx}F(x)}{\mathbb{P}(x_i \neq \bar{x})}.$$

Let the random variables $\tilde{x}_i^+$ be defined in $[\bar{x}, \infty) \cap \mathcal{X}$ with density function $\tilde{\varphi}(x)^+ = \varphi(x)$ in $(\bar{x}, x^+]$ and $\tilde{\varphi}(\bar{x})^+ = \lim_{x\downarrow\bar{x}}\varphi(x)$. Then $\mathbb{P}(\tilde{x}_i^+ = x_i\mathbf{1}(x_i > \bar{x})) = 1$. Define $\tilde{x}_i^-$

analogously. Though in theorem 3 Porter assumes that the x_i have a density function in an open set $\mathcal{N} \ni \bar{x}$, all the equations use either $x_i \mathbf{1}(x_i > \bar{x})$ or $x_i \mathbf{1}(x_i < \bar{x})$, and the results only require that $x_i \mathbf{1}(x_i \geqslant \bar{x})$ has a density in $[\bar{x}, x^+)$ and that $x_i \mathbf{1}(x_i \leqslant \bar{x})$ has a density in $(x^-, \bar{x}]$. Hence, theorem 3 in Porter (2003) can be applied to $\tilde{x}_i^+$ and $\tilde{x}_i^-$, and the results will be valid to $x_i \mathbf{1}(x_i > \bar{x})$ and $x_i \mathbf{1}(x_i < \bar{x})$ respectively with probability one. Assumption 2.6 (4)-(7) complete the requirements of the theorem. Let $\tilde{n} := \sum_{i=1}^n \mathbf{1}(x_i \neq \bar{x})$, Porter shows that

$$\sqrt{h\tilde{n}} \begin{pmatrix} \tilde{\tau}^+(\bar{x})^{\lim} - \tau^+(\bar{x})^{\downarrow} - \tilde{\mathcal{B}}_n^+ \\ \tilde{\tau}^-(\bar{x})^{\lim} - \tau^-(\bar{x})^{\uparrow} - \tilde{\mathcal{B}}_n^- \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tilde{\mathcal{V}}^+ & 0 \\ 0 & \tilde{\mathcal{V}}^- \end{bmatrix} \right) \tag{25}$$

where if p is odd,

$$\tilde{\mathcal{B}}_n^+ = h^{p+1} \frac{\tau^{+(p+1)}(\bar{x})^{\lim}}{(p+1)!} e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+1}) = \mathcal{B}_n^+$$

and if p is even,

$$\begin{aligned} \tilde{\mathcal{B}}_n^+ &= h^{p+2} \left[\frac{\tau^{+(p+1)}(\bar{x})^{\lim}}{(p+1)!} \frac{\tilde{\varphi}'(\bar{x})^+}{\tilde{\varphi}(\bar{x})^+} \right] e_1^T \Lambda_0^{-1} (\Upsilon_{p+2} - \Lambda_1 \Lambda_0 \Upsilon_{p+1}) \\ &\quad + \left[\frac{\tau^{+(p+2)}(\bar{x})^{\lim}}{(p+2)!} \right] e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+2}) \\ &= h^{p+2} \left[\frac{\tau^{+(p+1)}(\bar{x})^{\lim}}{(p+1)!} \frac{\phi'(\bar{x})^{\downarrow}}{\phi(\bar{x})^{\downarrow}} \right] e_1^T \Lambda_0^{-1} (\Upsilon_{p+2} - \Lambda_1 \Lambda_0 \Upsilon_{p+1}) \\ &\quad + \left[\frac{\tau^{+(p+2)}(\bar{x})^{\lim}}{(p+2)!} \right] e_1^T \Lambda_0^{-1} \Upsilon_{p+1} + o(h^{p+2}) = \mathcal{B}_n^+ \end{aligned}$$

and analogously for $\mathcal{B}_n^-$. Observe that $\mathbb{E}(\sigma_\epsilon^2(x_i, z_i) \mid x_i = x \,,\, x_i \neq \bar{x}) = \sigma^2(x)$ for all x in $(\bar{x}, x^+)$. Hence, if p is even or odd,

$$\tilde{\mathcal{V}}^+ = \frac{\mathbb{V}ar(\epsilon_i \mid \tilde{x}_i^+ = \bar{x})}{\tilde{\varphi}(\bar{x})^+} e_1^T \Lambda_0^{-1} \Omega\, \Lambda_0^{-1} e_1 = \mathbb{P}(x_i \neq \bar{x}) \frac{\sigma^2(\bar{x})^{\downarrow}}{\phi(\bar{x})^{\downarrow}} e_1^T \Lambda_0^{-1} \Omega\, \Lambda_0^{-1} e_1 = \mathbb{P}(x_i \neq \bar{x})\, \mathcal{V}^+,$$

and analogously for $\mathcal{V}^-$. By assumption 2.4 (1) and the LLN, $\tilde{n}/n \xrightarrow{p} \mathbb{P}(x_i \neq \bar{x}) > 0$, and by the Continuous Mapping theorem and Slutsky's theorem,

$$\sqrt{hn} \begin{pmatrix} \tilde{\tau}^+(\bar{x})^{\lim} - \tau^+(\bar{x})^{\downarrow} - \mathcal{B}_n^+ \\ \tilde{\tau}^-(\bar{x})^{\lim} - \tau^-(\bar{x})^{\uparrow} - \mathcal{B}_n^- \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{V}^+ & 0 \\ 0 & \mathcal{V}^- \end{bmatrix} \right). \tag{26}$$

Also, by assumption 2.4 item (1), assumption 2.5 item (1) and the LLN, $\hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x}) \xrightarrow{p} \mathbb{E}(g(z_i) \mid x_i = \bar{x})$. Hence, Slutsky's theorem and the continuous mapping theorem imply

$$\sqrt{nh}\left(B_n^1 - \mathcal{B}_n\right) \xrightarrow{d} \mathcal{N}\Big(0, \mathbb{E}(g(z_i) \mid x_i = \bar{x})^2 \left[\alpha^2 \mathcal{V}^+ + (1-\alpha)^2 \mathcal{V}^-\right]\Big).$$

For determining the asymptotic distribution of $B_n^2 + A_n$, denote

$$\begin{aligned}
a_n^+ &:= e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1}\tilde{X}^T W^+ Z,\\
a_n^- &:= e_1^T(\tilde{X}^T W^+ \tilde{X})^{-1}\tilde{X}^T W^- Z,\\
b_n^+ &:= \alpha\left[\hat{\mathbb{E}}(g(z_i)z_i \,|\, x_i = \bar{x})^T - \hat{\mathbb{E}}(g(z_i) \,|\, x_i = \bar{x})\, a_n^+\right],\\
b_n^- &:= (1-\alpha)\left[\hat{\mathbb{E}}(g(z_i)z_i \,|\, x_i = \bar{x})^T - \hat{\mathbb{E}}(g(z_i) \,|\, x_i = \bar{x})\, a_n^-\right],
\end{aligned}$$

then

$$B_n^2 + A_n = b_n^+(\hat{\gamma}^+ - \gamma^+) + b_n^-(\hat{\gamma}^- - \gamma^-) + A_n = \begin{bmatrix} b_n^+ & b_n^+ & 1\end{bmatrix}\begin{bmatrix} \hat{\gamma}^+ - \gamma^+ \\ \hat{\gamma}^- - \gamma^- \\ A_n \end{bmatrix}.$$

First, observe that assumptions 2.6 (3) and (6)-(8) and Theorem 3 in Porter (2003) guarantee that $a_n^+ \xrightarrow{p} \mathbb{E}(z_i \mid x_i = \bar{x})^{\downarrow}$ and $a_n^- \xrightarrow{p} \mathbb{E}(z_i \mid x_i = \bar{x})^{\uparrow}$. By assumption 2.4 item (1), assumption 2.6 item (1) and the LLN, $\hat{\mathbb{E}}(g(z_i) \mid x_i = \bar{x}) \xrightarrow{p} \mathbb{E}(g(z_i) \mid x_i = \bar{x})$, and $\hat{\mathbb{E}}(g(z_i)z_i \mid x_i = \bar{x}) \xrightarrow{p} \mathbb{E}(g(z_i)z_i \mid x_i = \bar{x})$. Hence, by Slutsky's theorem, $[b_n^{+T} \quad b_n^{-T} \quad 1] \xrightarrow{p} [\alpha C_+^T \quad (1-\alpha)C_-^T \quad 1]$. From assumption 2.6 (2) and Slutsky's theorem,

$$\begin{aligned}
\sqrt{n}(B_n^2 + A_n) &\xrightarrow{d} \begin{bmatrix} \alpha C_+^T & (1-\alpha)C_-^T & 1 \end{bmatrix} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{V}_\gamma^+ & 0 & 0 \\ 0 & \mathcal{V}_\gamma^- & 0 \\ 0 & 0 & V_A \end{bmatrix}\right)\\
&\sim \mathcal{N}\left(0, \alpha^2 C_+^T \mathcal{V}_\gamma^+ C_+ + (1-\alpha)^2 C_-^T \mathcal{V}_\gamma^- C_- + V_A\right). \qquad (27)
\end{aligned}$$

Since $\sqrt{nh}(B_n^2 + A_n) \xrightarrow{p} 0$, Slutsky's theorem guarantees the joint convergence of $\sqrt{nh}(B_n^1 - \mathcal{B}_n)$ and $\sqrt{nh}(B_n^2 + A_n)$. The only remaining task is to calculate the covariance $nh\,\mathbb{C}ov(B_n^1 - \mathcal{B}_n, B_n^2 + A_n)$ up to the $O(h)$ level. This result requires that one return to the proof of theorem 3 in Porter (2003), p. 44., and refer to equations (17) to (22) in that paper. B_n^1 can be rewritten as

$$\sqrt{nh}(B_n^1 - \mathcal{B}_n) = e_1^T\left[\alpha D_{n+}\sum_{s=17}^{19} E_s + (1-\alpha)D_{n-}\sum_{s=20}^{22} E_s\right]$$

where E_s is the numerator in equation (s) in p.44 in Porter (2003), the notation translates here as

$$\begin{aligned}
&d_i = \mathbf{1}(x_i > \bar{x}), \qquad Z_i = \frac{1}{h}\tilde{X}_i, \qquad D_{n+} = (n^{-1}\tilde{X}^T W^+ \tilde{X})^{-1}, \qquad B_{n+} = \mathcal{B}_n^+,\\
&y_i^+ = \tau^+(x_i) - \tau^+(\bar{x})^{\downarrow} - \tau^{+(1)}(\bar{x})^{\lim}(x_i - \bar{x}) - \cdots - \frac{1}{p!}\tau^{+(p)}(\bar{x})^{\lim}(x_i - \bar{x})^p + \varepsilon_i,\\
&\mu_j^+(x) = \tau^+(x) - \left[\tau^+(\bar{x})^{\downarrow} + \tau^{+(1)}(\bar{x})^{\lim}(x_i - \bar{x}) + \cdots + \frac{1}{j!}\tau^{+(j)}(\bar{x})^{\lim}(x_i - \bar{x})^j\right],
\end{aligned}$$

and the terms with the "-" sign are defined analogously. Hence,

$$\frac{1}{h}nh\,\mathbb{C}ov(B_n^1-\mathcal{B}_n,B_n^2+A_n)=\Bigg[\alpha\sum_{s=17}^{19}\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,\sqrt{n}(B_n^2+A_n))$$
$$+(1-\alpha)\sum_{s=20}^{22}\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n-}E_s,\sqrt{n}(B_n^2+A_n))\Bigg]$$

We will deal with the first term, for $s=17,18$ and 19. The second term, for $s=20,21$ and 22 is analogous.

$$\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,\sqrt{n}(B_n^2+A_n))=\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,\sqrt{n}B_n^2)+\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,\sqrt{n}A_n)$$

A_n is composed exclusively by observations for which $x_i=\bar{x}$, while $D_{n+}E_s$ is composed exclusively by observations for which $x_i>\bar{x}$. They are therefore independent, and since $\mathbb{E}(A_n)=0$, $\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,\sqrt{n}A_n)=0$. $B_n^2=b_n^+(\hat{\gamma}^+-\gamma^+)+b_n^-(\hat{\gamma}^--\gamma^-)$. The term $b_n^-(\hat{\gamma}^--\gamma^-)$ is composed exclusively of observations for which $x_i=\bar{x}$ and $x_i>\bar{x}$. It is therefore independent of $D_{n+}E_s$, and $\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_s,b_n^-\sqrt{n}(\hat{\gamma}^--\gamma^-))=0$. The term E_{19} is not random, and therefore, $\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_{19},b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))=0$. For the term E_{18}, we will use Hölder's inequality:

$$\begin{aligned}\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_{18},b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))&=\mathbb{E}(e_1^TD_{n+}E_{18}b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))\\&-\mathbb{E}(e_1^TD_{n+}E_{18})\mathbb{E}(b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))\\&\leqslant\mathbb{E}((e_1^TD_{n+}E_{18})^2)^{1/2}\mathbb{E}(n(b_n^+(\hat{\gamma}^+-\gamma^+))^2)^{1/2}\\&+\mathbb{E}(|e_1^TD_{n+}E_{18}|)\mathbb{E}(|b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+)|)\end{aligned}$$

Porter shows that $\mathbb{V}ar(e_1^TD_{n+}E_{18})=o(h^{-(p+1)})$, and from assumption 2.6 (2), $\mathbb{E}(n(b_n^+(\hat{\gamma}^+-\gamma^+))^2)^{1/2}$ is uniformly bounded. Hence, $\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_{18},b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))=o(1)$.

The only remaining term is $\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^TD_{n+}E_{17},b_n^+\sqrt{n}(\hat{\gamma}^+-\gamma^+))$. It is necessary to have a better understanding of $\hat{\gamma}^+$. $b_n^+(\hat{\gamma}^+-\gamma^+)=b_n^+(\tilde{Z}_+^T\tilde{Z}_+)^{-1}\tilde{Z}_+^T(\tilde{Y}_+-\tilde{Z}_+\gamma^+)$, and for observations such that $x_i>\bar{x}$,

$$\begin{aligned}\tilde{y}_{i+}-\tilde{z}_{i+}\gamma^+&=y_i-z_i^T\gamma^+-\sum_{j=1}^n\mathbf{1}(x_j>\bar{x})\,T_{i,j}^+(y_j-z_j^T\gamma^+)\\&=\tau^+(x_i)+\epsilon_i-\sum_{j=1}^n\mathbf{1}(x_j>\bar{x})\,T_{i,j}^+(\tau^+(x_j)+\epsilon_j)\end{aligned}$$

Let $T^+=[T_{i,j}^+\mathbf{1}(x_j>\bar{x})]$ the $n\times n$ matrix with entry $T_{i,j}\mathbf{1}(x_i>\bar{x},x_j>\bar{x})$ in line i, column j, $\tau^+=(\tau^+(x_1),\ldots,\tau^+(x_n))^T$, and $P_\gamma^+:=b_n^+(\tilde{Z}_+^T\tilde{Z}_+)^{-1}\tilde{Z}_+^T$. Then $b_n^+(\hat{\gamma}^+-\gamma^+)=P_\gamma^+(I^+-T^+)(\tau^++\epsilon)$, where $I^+=Diag\{\mathbf{1}(x_1>\bar{x}),\ldots,\mathbf{1}(X_n>\bar{x})\}$. Also, $e_1D_{n+}E_{17}=\sqrt{nh}P_1^+\epsilon$. Hence, since it can be easily shown that $E(\frac{1}{\sqrt{h}}e_1^TD_{n+}E_{17})=$

$o(1)$ and since $b_n^+\sqrt{n}(\hat{\gamma}^+ - \gamma^+))$ is uniformly bounded,

$$\frac{1}{\sqrt{h}}\mathbb{C}ov(e_1^T D_{n+}E_{17}, b_n^+\sqrt{n}(\hat{\gamma}^+ - \gamma^+)) = n\mathbb{E}(P_1^+\epsilon(\hat{\gamma}^+ - \gamma^+)^T b_n^{+T})$$

$$\begin{aligned}
&\left(P_1^+\epsilon\right) b_n^+(\hat{\gamma}^+ - \gamma^+) = \left(P_1^+\epsilon\right)(\hat{\gamma}^+ - \gamma^+)^T b_n^{+T} = P_1^+\epsilon(\tau^{+T} + \epsilon^T)(I^+ - T^+)^T P_\gamma^{+T} \\
&\implies \mathbb{E}(\left(P_1^+\epsilon\right) b_n^+(\hat{\gamma}^+ - \gamma^+)) = \mathbb{E}(P_1^+\epsilon^2(I^+ - T^+)^T P_\gamma^{+T})
\end{aligned}$$

where $\epsilon^2 = Diag\{\epsilon_1^2, \ldots, \epsilon_n^2\}$. Define $\mathbb{E}(\epsilon^2 \mid X) = Diag\{\mathbb{E}(\epsilon_i^2 \mid x_i = x_1), \ldots, \mathbb{E}(\epsilon_i^2 \mid x_i = x_n)\}$ and $\hat{\mathbb{E}}(\epsilon^2 \mid X) = Diag\{\hat{\mathbb{E}}(\epsilon_i^2 \mid x_i = x_1\}, \ldots, \hat{\mathbb{E}}(\epsilon_i^2 \mid x_i = x_n))^T$. We can then rewrite $\epsilon^2(I^+ - T^+)^T P_\gamma^{+T}$ as

$$\begin{aligned}
P_\gamma^T(I^+ - T^+)\epsilon^2 =& b_n^+(\tilde{Z}_+^T\tilde{Z}_+)^{-1}Z^T(I^+ - T^{+T})(I^+ - T^+)\epsilon^2 \\
=& b_n^+(\tilde{Z}_+^T\tilde{Z}_+)^{-1}\Big[Z^T\epsilon^2 - \mathbb{E}(Z \mid X)^T\epsilon^2 - Z^T\mathbb{E}(\epsilon^2 \mid X) + \mathbb{E}(Z \mid X)^T\mathbb{E}(\epsilon^2 \mid X) \\
&- (\hat{\mathbb{E}}(Z \mid X) - \mathbb{E}(Z \mid X))^T\epsilon^2 - Z^T(\hat{\mathbb{E}}(\epsilon^2 \mid X) - \mathbb{E}(Z \mid X)) \\
&+ (\hat{\mathbb{E}}(Z \mid X) - \mathbb{E}(Z \mid X))^T\hat{\mathbb{E}}(\epsilon^2 \mid X) + \mathbb{E}(Z \mid X)(\hat{\mathbb{E}}(\epsilon^2 \mid X) - \mathbb{E}(\epsilon^2 \mid X))\Big].
\end{aligned}$$

Hence

$$\begin{aligned}
nP_1^+\epsilon^2(I^+ - T^+)^T P_\gamma^{+T} = \Big[&P_1^+\epsilon^2 Z - P_1^+\epsilon^2\mathbb{E}(Z \mid X) - P_1^+\mathbb{E}(\epsilon^2 \mid X)Z \\
&+ P_1^+\mathbb{E}(\epsilon^2 \mid X)\mathbb{E}(Z \mid X)\Big]\left(\frac{\tilde{Z}_+^T\tilde{Z}_+}{n}\right)^{-1} b_n^{+T} + U_n,
\end{aligned}$$

$$\begin{aligned}
|U_n| \leqslant& \left|\frac{1}{n}\sum_{i=1}^n \frac{1}{h}k\left(\frac{x_i - \bar{x}}{h}\right)\mathbf{1}(x_i > 0)\tilde{X}_i\epsilon_i^2\right|\sum_{s=1}^d \sup_i\left|\hat{\mathbb{E}}(z_i^s \mid x_i) - \mathbb{E}(z_i^s \mid x_i)\right||b_n^s| \\
&+ \left\|\frac{1}{n}\sum_{i=1}^n \frac{1}{h}k\left(\frac{x_i - \bar{x}}{h}\right)\mathbf{1}(x_i > 0)\tilde{X}_i z_i\right\|\sum_{s=1}^d \sup_i\left|\hat{\mathbb{E}}(\epsilon_i^2 \mid x_i) - \mathbb{E}(\epsilon_i^2 \mid x_i)\right||b_n^s| \\
&+ \left|\frac{1}{n}\sum_{i=1}^n \frac{1}{h}k\left(\frac{x_i - \bar{x}}{h}\right)\mathbf{1}(x_i > 0)\tilde{X}_i\right|\sup_i\left|\hat{\mathbb{E}}(\epsilon_i^2 \mid x_i)\right|\sum_{s=1}^d \sup_i\left|\hat{\mathbb{E}}(z_i \mid x_i) - \mathbb{E}(z_i \mid x_i)\right||b_n^s| \\
&+ \left\|\frac{1}{n}\sum_{i=1}^n \frac{1}{h}k\left(\frac{x_i - \bar{x}}{h}\right)\mathbf{1}(x_i > 0)\tilde{X}_i\mathbb{E}(z_i \mid x_i)\right\|\sum_{s=1}^d \sup_i\left|\hat{\mathbb{E}}(\epsilon_i^2 \mid x_i) - \mathbb{E}(\epsilon_i^2 \mid x_i)\right||b_n^s|.
\end{aligned}$$

Because the terms with the kernels can easily be shown to be bounded, $\sup_i\left\|\hat{\mathbb{E}}(\epsilon_i^2 \mid x_i)\right\|$ is asymptotically bounded, because $\sup_i\left\|\sum_{j=1}^n \mathbf{1}(x_j > \bar{x})\,T_{i,j}^+\epsilon_j^2 - \mathbb{E}(\epsilon_i^2 \mid x_i)\right\| = o_p(1)$, and the other terms are $o_p(1)$ by assumption 2.6 (2), $U_n = o_p(1)$.

Finally, again by assumption 2.6 (2),

$$\left(\frac{\tilde{Z}_+^T\tilde{Z}_+}{n}\right) \xrightarrow{p} \mathbb{V}ar(z_i \mid x_i)$$
$$\implies nP_1^+\epsilon^2(I^+ - T^+)^T P_\gamma^{+T} \xrightarrow{p} \alpha(\mathbb{E}(\epsilon_i^2 z_i \mid x_i = \bar{x})^\downarrow - \mathbb{E}(\epsilon_i^2 \mid x_i = \bar{x})^\downarrow \mathbb{E}(z_i z_i \mid x_i = \bar{x})^\downarrow)\, C^+$$
$$= \alpha\, C^+(\Sigma_z(\bar{x})^\downarrow)^{-1} c_{z\epsilon^2}(\bar{x})^\downarrow.$$

Therefore,

$$\alpha \sum_{s=17}^{19} \frac{1}{\sqrt{h}} \mathbb{C}ov(e_1^T D_{n+} E_s, \sqrt{n}(B_n^2 + A_n)) \to \alpha^2\, C^+(\Sigma_z(\bar{x})^\downarrow)^{-1} c_{z\epsilon^2}(\bar{x})^\downarrow$$

and the result for $s = 20, 21$ and 22 is analogous.

A.3.2 Theorem 6:

The convergence of $\hat{\mathcal{V}}$ follows from observing that $\alpha\hat{C}_+ = b_n^+$ and $(1-\alpha)\hat{C}_- = b_n^-$ and its convergence is established in the previous section. Theorem 4 in Porter (2003) guarantees the convergence of $\hat{\mathcal{V}}^+$ and $\hat{\mathcal{V}}^-$, as long as $\hat{\sigma}^2(\bar{x})^{\lim s} \to \sigma^2(\bar{x})^{\lim s}$ and $\hat{\phi}(\bar{x})^{\lim s} \to \phi(\bar{x})^{\lim s}$ as $n \to \infty$. The latter is guaranteed by item (3) in assumtpion 2.6. We show the former for $s =$ "$+$", the result for $s =$ "$-$" is analogous.

$$\mathbb{E}((y_i - z_i^T\hat{\gamma}^+)^2 \mid x_i = \bar{x})^\downarrow = e_1^T(\tilde{X}^T W^{s+}\tilde{X})^{-1}\tilde{X}^T W^+ R^+ \tag{28}$$
$$= e_1^T(\tilde{X}^T W^+\tilde{X})^{-1}\tilde{X}^T W^+ \tilde{R}^+$$
$$- e_1^T(\tilde{X}^T W^+\tilde{X})^{-1}\tilde{X}^T W^+ (R^+ - \tilde{R}^+) \tag{29}$$

where $\tilde{R}^+ = (\tilde{R}_1^+, \ldots, \tilde{R}_n^+)^T$, $\tilde{R}_i^+ = (y_i - z_i^T\gamma^+)^2$. We begin by showing that the second term is $o_p(1)$. Notice that

$$R_i^+ - \tilde{R}_i^+ = (y_i - z_i^T\hat{\gamma}^+)^2 - (y_i - z_i^T\gamma^+)^2$$
$$= \left[(y_i - z_i^T\hat{\gamma}^+) + (y_i - z_i^T\gamma^+)\right] z_i^T(\hat{\gamma}^+ - \gamma^+)$$

Let

$$\hat{U}^+ = e_1^T(\tilde{X}^T W^+\tilde{X})^{-1} \sum_{i=1}^n \mathbf{1}(x_i > \bar{x})\, k\left(\frac{x_i - \bar{x}}{h}\right)(y_i - z_i^T\hat{\gamma}^+) z_i^T$$
$$\tilde{U}^+ = e_1^T(\tilde{X}^T W^+\tilde{X})^{-1} \sum_{i=1}^n \mathbf{1}(x_i > \bar{x})\, k\left(\frac{x_i - \bar{x}}{h}\right)(y_i - z_i^T\gamma^+) z_i^T$$

Then the second term in (28) is

$$e_1^T(\tilde{X}^T W^+\tilde{X})^{-1}\tilde{X}^T W^+(R^+ - \tilde{R}^+) = (\hat{U}^+ + \tilde{U}^+)(\hat{\gamma}^+ - \gamma^+)$$

From equation (15) and the proof of theorem 5, it's easy to see that

$$\hat{U}^{+} + \tilde{U}^{+} \xrightarrow{p} 2\,(\tau^{+}(\bar{x})^{\downarrow})\,\mathbb{E}(z_i \mid x_i = \bar{x})^{\downarrow}$$

and since $\hat{\gamma}^{+} - \gamma^{+} \xrightarrow{p} 0$ by assumption 2.6 item (2), the second term in (28) is $o_p(1)$.

The first term in (28) is a local polynomial regression of $(y_i - z_i^T\gamma^{+})^2$ on x_i at $\bar{x}$. Hence, from assumption 2.6 items (2)-(7) and theorem 4.1 in Ruppert and Wand (1994),

$$e_1^T(\tilde{X}^T W^{+} \tilde{X})^{-1}\tilde{X}^T W^{+} \tilde{R}^{+} \xrightarrow{p} \lim_{x\downarrow\bar{x}} \mathbb{E}(y_i - z_i^T\gamma^{+})^2 \mid x_i = x) = \lim_{x\downarrow\bar{x}} \mathbb{E}(\epsilon_i^2 \mid x_i = x) = \sigma^2(\bar{x})^{\downarrow}.$$

The proof for the convergence of $\hat{c}_{z\epsilon^2}(\bar{x})^{\lim}$ is analogous. It is only necessary to observe that

$$\mathbb{E}(z_i(y_i - z_i\gamma^{+})^2 \mid x_i) - \mathbb{E}(z_i \mid x_i)\mathbb{E}((y_i - z_i\gamma^{+})^2 \mid x_i) = \mathbb{E}(z_i\epsilon_i^2 \mid x_i) - \mathbb{E}(z_i \mid x_i)\mathbb{E}(\epsilon_i^2 \mid x_i).$$

A.3.3 Theorem 7:

Observe that

$$\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) = \mathbb{P}\left(\sqrt{nh}\left(\frac{\hat{\theta} - \theta - \mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}} > c_\lambda - \frac{\sqrt{nh}\theta}{\sqrt{\mathcal{V}_n}}\right).$$

From theorem 6 and the continuous mapping theorem, $\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n} \xrightarrow{p} 0$. Moreover, $\sqrt{nh}\mathcal{B}_n \to 0$, since $\sqrt{nh}h^{p+1} \to 0$. Hence, by theorems 5 and Slutsky's, $\sqrt{nh}\left(\frac{\hat{\theta}-\theta-\mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n}-\sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}} \xrightarrow{d} \mathcal{N}(0,1)$. Under H_0, $\theta = 0$, and the first result follows immediately. Under H_1, since $h \to 0$, $\mathcal{V}_n \to \alpha^2\mathcal{V}_\tau^{+} + (1-\alpha)^2\mathcal{V}_\tau^{-}$, and therefore $-\frac{\sqrt{nh}\theta}{\sqrt{\mathcal{V}_n}} \to -\infty$, from which the second result follows.

Under the alternatives $\theta/\sqrt{nh}$, observe that

$$\begin{aligned}\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) = \mathbb{P}\Bigg(&\sqrt{nh}\left(\frac{\hat{\theta} - \theta/\sqrt{nh} - \mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}} \\ &+ \theta\left(\frac{1}{\sqrt{\mathcal{V}_n}} - \frac{1}{\sqrt{\alpha^2\mathcal{V}_\tau^{+} + (1-\alpha)^2\mathcal{V}_\tau^{-}}}\right) > c_\lambda - \frac{\theta}{\sqrt{\alpha^2\mathcal{V}_\tau^{+} + (1-\alpha)^2\mathcal{V}_\tau^{-}}}\Bigg).\end{aligned}$$

and since $\theta\left(\frac{1}{\sqrt{\mathcal{V}_n}} - \frac{1}{\sqrt{\alpha^2\mathcal{V}_\tau^{+}+(1-\alpha)^2\mathcal{V}_\tau^{-}}}\right) \xrightarrow{p} 0$, by Slutsky's theorem the third result of the theorem follows.

A.4 Estimation in the nonparametric case

A.4.1 Theorem 8:

The proof is similar to the proof of the convergence of the nonparametric term in the partially linear case. The essence of the argument is that since the support if $dF(z_i)$ is finite, all arguments can be done separately for each possible value of z_i. We begin by deriving the asymptotic distribution of $\hat{\Gamma}(z^m)^+$. This is a standard local polynomial regression of the kind used in Porter (2003) for the estimation of one side of the discontinuity in the regression discontinuity design. There are two differences. First, $\hat{\Gamma}(z^m)^+$ uses only data for which $z_i = z^m$. Second, the results in Porter assume that the variable x_i has a density function in a neighborhood of $\bar{x}$. Assumption 2.7 item (1) implies that $\mathbb{P}(x_i = \bar{x} \mid z_i = z^m) > 0$, so this is no longer possible. However, from item (2), the conditional distribution function

$$\mathbb{P}(x_i \leqslant x \mid x_i > \bar{x}\,,\, z_i = z^m) = \frac{\mathbb{P}(x_i \leqslant x\,,\, z_i = z^m) - \mathbb{P}(x_i \leqslant \bar{x}\,,\, z_i = z^m)}{\mathbb{P}(x_i > \bar{x}\,,\, z_i = z^m)}$$

has a density function in $(\bar{x}, x^+)$, and it is equal to

$$\varphi_m(x) := \frac{\frac{d}{dx}\mathbb{P}(x_i \leqslant x\,,\, z_i = z^m)}{\mathbb{P}(x_i > \bar{x}\,,\, z_i = z^m)}.$$

Though theorem 3 in Porter dependens on the existence of a density function in $(\bar{x}, x^+)$, it is not dependent on the existence of a density function at the discontinuity point $\bar{x}$, as long as the right limit of $\varphi_m(x)$ at $\bar{x}$ exists. From assumption 2.7 item (2), this is true and

$$\varphi_m(\bar{x})^{\downarrow} := \lim_{x \downarrow \bar{x}} \varphi_m(x) := \frac{\phi(\bar{x}, z^m)^{\downarrow}}{\mathbb{P}(x_i > \bar{x}\,,\, z_i = z^m)}.$$

Assumption 2.7 (3)-(6) complete the requirements of Theorem 3 in Porter (2003). Let $n_m^+ := \sum_{i=1}^n \mathbf{1}(x_i > \bar{x})\mathbf{1}(z_i = z^m)$,

$$\sqrt{hn_m^+}\left(\hat{\Gamma}(z^m)^+ - \tilde{\mathcal{B}}_{m,n}^+\right) \xrightarrow{d} \mathcal{N}\left(0\,,\, \tilde{\mathcal{V}}_m^+\right) \tag{30}$$

where if p is odd,

$$\tilde{\mathcal{B}}_n^+ = h^{p+1}\frac{f^{+(p+1)}(\bar{x}, z^m)^{\lim}}{(p+1)!}e_1^T\Lambda_0^{-1}\Upsilon_{p+1} + o(h^{p+1}) = \mathcal{B}_{m,n}^+$$

and if p is even,

$$\begin{aligned}\tilde{\mathcal{B}}_n^+ &= h^{p+2}\left[\frac{f^{+(p+1)}(\bar{x},z^m)^{\lim}}{(p+1)!}\frac{\varphi'(\bar{x},z^m)^{\downarrow}}{\varphi(\bar{x},z^m)^{\downarrow}}\right] e_1^T\Lambda_0^{-1}(\Upsilon_{p+2}-\Lambda_1\Lambda_0\Upsilon_{p+1}) \\ &\quad+\left[\frac{f^{+(p+2)}(\bar{x},z^m)^{\lim}}{(p+2)!}\right] e_1^T\Lambda_0^{-1}\Upsilon_{p+1}+o(h^{p+2}) \\ &= h^{p+2}\left[\frac{f^{+(p+1)}(\bar{x},z^m)^{\lim}}{(p+1)!}\frac{\phi'(\bar{x},z^m)^{\downarrow}}{\phi(\bar{x},z^m)^{\downarrow}}\right] e_1^T\Lambda_0^{-1}(\Upsilon_{p+2}-\Lambda_1\Lambda_0\Upsilon_{p+1}) \\ &\quad+\left[\frac{f^{+(p+2)}(\bar{x},z^m)^{\lim}}{(p+2)!}\right] e_1^T\Lambda_0^{-1}\Upsilon_{p+1}+o(h^{p+2}) = \mathcal{B}_{m,n}^+\end{aligned}$$

Also, observe that $\mathbb{E}(\sigma_\epsilon^2(x_i,z_i)\mid x_i=x\,,\,x_i>\bar{x},z_i=z^m)=\sigma_m^2(x)$ for all x in $(\bar{x},x^+)$. Hence, if p is even or odd,

$$\begin{aligned}\tilde{\mathcal{V}}_m^+ &= \frac{\sigma^2(\bar{x})^{\downarrow}}{\varphi(\bar{x},z^m)^{\downarrow}}e_1^T\Lambda_0^{-1}\Omega\,\Lambda_0^{-1}e_1=\mathbb{P}(x_i>\bar{x},z_i=z^m)\frac{\sigma^2(\bar{x})^{\downarrow}}{\phi(\bar{x},z^m)^{\downarrow}}e_1^T\Lambda_0^{-1}\Omega\,\Lambda_0^{-1}e_1 \\ &= \mathbb{P}(x_i>\bar{x}\,,\,z_i=z^m)\,\mathcal{V}_m^+\end{aligned}$$

By assumption 2.4 (1) and the LLN, $n_m^+/n\xrightarrow{p}\mathbb{P}(x_i>\bar{x}\,,\,z_i=z^m)$, and by the continuous mapping theorem and Slutsky's theorem,

$$\sqrt{nh}\left(\hat{\Gamma}(z^m)^+-\mathcal{B}_{m,n}^+\right)\xrightarrow{d}\mathcal{N}\left(0\,,\,\mathcal{V}_m^+\right).$$

The exact same reasoning applied to $\hat{\Gamma}(z^m)^-$ will yield the equivalent result for the left limit. Moreover, the result in Porter (2003) states the joint convergence of $\sqrt{nh}\left(\hat{\Gamma}(z^m)^+-\mathcal{B}_{m,n}^+\right)$ and $\sqrt{nh}\left(\hat{\Gamma}(z^m)^--\mathcal{B}_{m,n}^-\right)$ by the Cràmer Wold device. The $\hat{\Gamma}(z^m)^+$ are independent for all m, and also independent from the $\hat{\Gamma}(z^m)^-$, because they are built using different parts of the sample, hence by continuous mapping theorem,

$$\alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^m)^+-\mathcal{B}_{m,n}^+\right)+(1-\alpha)\alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^m)^--\mathcal{B}_{m,n}^-\right)\xrightarrow{d}\mathcal{N}\left(0\,,\,\alpha^2\mathcal{V}_m^++(1-\alpha)^2\mathcal{V}_m^-\right).$$

Assumption 2.4 (1) and (2), the LLN and Slutsky's theorem imply that $\hat{p}_{\bar{x}}^m\xrightarrow{p}p_{\bar{x}}^m$ jointly for all m. By Slutsky's theorem again, $\sqrt{nh}\,(B_n-\mathcal{B}_n)=$

$$=\begin{bmatrix}\hat{p}_{\bar{x}}^1 & \dots & \hat{p}_{\bar{x}}^M\end{bmatrix}\begin{bmatrix}\alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^1)^+-\mathcal{B}_{1,n}^+\right)+(1-\alpha)\alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^1)^--\mathcal{B}_{1,n}^-\right)\\ \vdots \\ \alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^M)^+-\mathcal{B}_{M,n}^+\right)+(1-\alpha)\alpha\sqrt{nh}\,\left(\hat{\Gamma}(z^M)^--\mathcal{B}_{M,n}^-\right)\end{bmatrix}\xrightarrow{d}$$

$$\xrightarrow{d}\mathcal{N}\left(0\,,\,\begin{bmatrix}p_{\bar{x}}^1 & \dots & p_{\bar{x}}^M\end{bmatrix}\,Diag\{\alpha^2\mathcal{V}_m^++(1-\alpha)^2\mathcal{V}_m^-\}\begin{bmatrix}p_{\bar{x}}^1 & \dots & p_{\bar{x}}^M\end{bmatrix}^T\right)=\mathcal{N}\left(0\,,\,\mathcal{V}\right)$$

The joint convergence of $\sqrt{nh}B_n$ and $\sqrt{nh}A_n$ is guaranteed by Slutsky's theorem, because $\sqrt{nh}A_n\xrightarrow{p}0$. In order to derive the small sample covariance, the same considerations as in the correlation between the $\hat{\Gamma}(z^m)^+$ and the $\hat{\Gamma}(z^m)^-$ for all m apply here, namely that they are independent from A_n because they are built using different observations. A_n may be correlated with $\mathcal{B}_n$. Equation (8) and lemma B.2 imply that $nh\,\mathbb{E}(A_n\mathcal{B}_n) = \sqrt{h}\,O\left(h^{p+1}\right) = h\,O(h^{p+1/2})$, which is of order smaller than h, and therefore the correlation os negligible. Hence, the small sample variance is

$$\mathcal{V} + hV_A + o(h) = \mathcal{V}_n$$

which concludes the demonstration.

A.4.2 Theorem 9:

We showed that $\hat{p}^m_{\bar{x}} \xrightarrow{p} p^m_{\bar{x}}$. It only remains to prove that $\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow} \xrightarrow{p} \sigma^2(\bar{x}, z^m)^{\downarrow}$ and $\hat{\sigma}^2(\bar{x}, z^m)^{\uparrow} \xrightarrow{p} \sigma^2(\bar{x}, z^m)^{\uparrow}$ for all m. In the beginning of section 8 in the appendix, we showed that the restriction to the observations such that $x_i \in (\bar{x}, x^+)$ and $z_i = z^m$ has a density function in $(\bar{x}, x^+)$ equal to $\varphi_m(x)$. We will use Masry (1996)'s result on the uniform convergence of the multivariate local polynomial. From assumption 2.7 and Theorem 6 in that article, if $z_i = z^m$,

$$\sup_{x\in(\bar{x},x^+)} \left|\hat{f}^+(x_i, z_i) - f(x_i, z_i)\right| = \sup_{x\in(\bar{x},x^+)} \left|\hat{f}^+(x_i, z^m) - f(x_i, z^m)\right| = O\left(\left(\frac{\log n}{nh}\right)^{1/2} + h^{p+1}\right)$$

almost surely. Define $D^+_m = \{i\,;\, x_i > \bar{x} \text{ and } z_i = z^m\}$, then by the continuous mapping theorem,

$$\sup_{i\in D^+_m} \left|(\hat{\epsilon}^s_i)^2 - \epsilon^2_i\right| = O\left(\left(\frac{\log n}{nh}\right)^{1/2} + h^{p+1}\right) \quad \text{a.s.}$$

Let $\tilde{R} = (\epsilon^2_1, \ldots, \epsilon^2_n)^T$,

$$\hat{\sigma}^2(\bar{x}, z^m)^{\downarrow} = P^+_{1,m,\bar{x}}\tilde{R} + P^+_{1,m,\bar{x}}(R - \tilde{R})$$

The first term is a simple local polynomial regression of the ϵ^2_i onto x_i at $\bar{x}$, and by theorem 3 in Porter (2003), it is a consistent estimator of $\lim_{x\downarrow\bar{x}} \mathbb{E}(\epsilon^2_i \mid x_i = x, z_i = z^m) = \sigma^2(\bar{x}, z^m)^{\downarrow}$. For the second term, let $(v)_i$ denote the i-th element of vector v, and since $\left(P^+_{1,m,\bar{x}}\right)_i$ is different from zero only if $i \in D^+_m$,

$$\begin{aligned}
\left|P^+_{1,m,\bar{x}}(R - \tilde{R})\right| &= \left|\left(\frac{\tilde{X}^T_x W^s_{x,m}\tilde{X}_x}{nh}\right)^{-1} \frac{\tilde{X}^T_x W^s_{x,m}(R - \tilde{R})}{nh}\right| \\
&\leqslant \left\|\left(\frac{\tilde{X}^T_x W^s_{x,m}\tilde{X}_x}{nh}\right)^{-1}\right\| \left\|\frac{\tilde{X}^T_x W^s_{x,m}(R - \tilde{R})}{h}\right\| \\
&\leqslant \left\|\left(\frac{\tilde{X}^T_x W^s_{x,m}\tilde{X}_x}{nh}\right)^{-1}\right\| \sup_{i\in D^+_m} \left|\left(\frac{\tilde{X}^T_x W^s_{x,m}}{nh}\right)_i \left((\hat{\epsilon}^s_i)^2 - \epsilon^2_i\right)\right| \\
&\leqslant \left\|\left(\frac{\tilde{X}^T_x W^s_{x,m}\tilde{X}_x}{nh}\right)^{-1}\right\| \sup_{i\in D^+_m} \left|\left(\frac{\tilde{X}^T_x W^s_{x,m}}{h}\right)_i\right| \sup_{i\in D^+_m} \left|(\hat{\epsilon}^s_i)^2 - \epsilon^2_i\right|
\end{aligned}$$

Observe that $\left(\frac{\tilde{X}_x^T W_{x,m}^s}{h}\right)_i = \mathbf{1}(i \in D_m^+)\frac{1}{h}k\left(\frac{x_i-\bar{x}}{h}\right)(a_0 + a_1(x_i - \bar{x}) + \cdots + a_p(x_i - \bar{x})^p)$. From assumption 2.7 (5), the kernel has bounded support, and since k is continuous, there exists $\bar{k}$ such that $|k(u)| \leqslant \bar{k}$ for all u. Let $u^{\max} := \sup_u \{u\,;\, k(u) \neq 0\}$, define $x_h^{\max} := \bar{x} + u^{\max}h$. Hence,

$$
\begin{aligned}
\left|\left(\frac{\tilde{X}_x^T W_{x,m}^s}{h}\right)_i\right| &\leqslant \frac{1}{h}\bar{k}\,\left[|a_0| + |a_1||x_h^{\max} - \bar{x}| + \cdots + |a_p||x_h^{\max} - \bar{x}|^p\right] \\
&\leqslant \frac{1}{h}\bar{k}\,\left[|a_0| + |a_1 u^{\max}|\,h + \cdots + |a_p(u^{\max})^p|\,h^p\right] \\
&\leqslant \frac{C}{h}, \quad \text{for } n \text{ large enough.}
\end{aligned}
$$

$$
\implies \left|P_{1,m,\bar{x}}^{+}(R - \tilde{R})\right| \leqslant \frac{C}{h}\left\|\left(\frac{\tilde{X}_x^T W_{x,m}^s \tilde{X}_x}{nh}\right)^{-1}\right\| \sup_{i \in D_m^+}\left|(\hat{\epsilon}_i^s)^2 - \epsilon_i^2\right|
$$

By the convergence of $P_{1,m,\bar{x}}^{+}\tilde{R}$ and the continuous mapping theorem, there exists a $(p+1)\times(p+1)$ positive definite matrix M such that for all $\delta > 0$,

$$
\mathbb{P}\left(\left\|\left(\frac{\tilde{X}_x^T W_{x,m}^s \tilde{X}_x}{nh}\right)^{-1} - M^{-1}\right\| > \delta\right) \to 0
$$

Hence,

$$
\begin{aligned}
\implies \left|P_{1,m,\bar{x}}^{+}(R - \tilde{R})\right| &\leqslant \frac{C}{h}\left(\left\|M^{-1}\right\| + o_p(1)\right)\sup_{i \in D_m^+}\left|(\hat{\epsilon}_i^s)^2 - \epsilon_i^2\right| \\
&= \left[\left(\frac{(\log n)^{1/3}}{n^{1/3}h}\right)^{3/2} + h^p\right] O_p\left(1\right) \quad \text{a.s.}
\end{aligned}
$$

From assumption 2.8 (3), $hn^{1/3}(\log n)^{-1/3} \to \infty$, and from assumption 2.7 (7), $h \to 0$. Hence, $\left|P_{1,m,\bar{x}}^{+}(R - \tilde{R})\right| \xrightarrow{p} 0$. The proof of the convergence of $\hat{\sigma}^2(\bar{x}, z^m)^{\uparrow}$ is analogous.

A.4.3 Theorem 10:

Analogously to the proof of theorem 7,

$$
\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) = \mathbb{P}\left(\sqrt{nh}\left(\frac{\hat{\theta} - \theta - \mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}} > c_\lambda - \frac{\sqrt{nh}\theta}{\sqrt{\mathcal{V}_n}}\right).
$$

From theorem 9 and the continuous mapping theorem, $\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n} \xrightarrow{p} 0$. Moreover, if $\sqrt{nh}h^{p+1} \to 0$, $\sqrt{nh}\mathcal{B}_{m,n}^{+} \to 0$ and $\sqrt{nh}\mathcal{B}_{m,n}^{-} \to 0$. Hence $\sqrt{nh}\mathcal{B}_n \to 0$. Hence, by theorems 8 and Slutsky's, $\sqrt{nh}\left(\frac{\hat{\theta}-\theta-\mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n}-\sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}} \xrightarrow{d} \mathcal{N}(0,1)$. Under H_0, $\theta = 0$, and the first result follows immediately. Under H_1, since $h \to 0$, $\mathcal{V}_n \to \mathcal{V}$, and therefore $-\frac{\sqrt{nh}\theta}{\sqrt{\mathcal{V}_n}} \to -\infty$, from which the second result follows.

Under the alternatives $\theta/\sqrt{nh}$, observe that

$$\mathbb{P}\left(\sqrt{nh}\frac{\hat{\theta}}{\sqrt{\hat{\mathcal{V}}_n}} > c_\lambda\right) = \mathbb{P}\left(\sqrt{nh}\left(\frac{\hat{\theta} - \theta/\sqrt{nh} - \mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right) - \frac{c_\lambda\left(\sqrt{\hat{\mathcal{V}}_n} - \sqrt{\mathcal{V}_n}\right)}{\sqrt{\mathcal{V}_n}} + \frac{\sqrt{nh}\mathcal{B}_n}{\sqrt{\mathcal{V}_n}}\right.$$
$$\left. + \theta\left(\frac{1}{\sqrt{\mathcal{V}_n}} - \frac{1}{\sqrt{\mathcal{V}}}\right) > c_\lambda - \frac{\theta}{\sqrt{\mathcal{V}}}\right).$$

and since $\theta\left(\frac{1}{\sqrt{\mathcal{V}_n}} - \frac{1}{\sqrt{\mathcal{V}}}\right) \xrightarrow{p} 0$, by Slutsky's theorem the third result of the theorem follows.

B Lemmas and definitions

Given how much the term "uniform integrability" is used here (particularly in the proof of theorem 5), explicitating the definition may be useful.

Definition 2. *A sequence of random variables $\{\kappa_n,\, n \geqslant 1\}$ is called uniformly integrable if for every $\varepsilon > 0$, there corresponds a $\delta > 0$ such that $\sup_{n\geqslant 1} \mathbb{E}(|\kappa_n| \mid \kappa_n \in A) < \varepsilon$ whenever $\mathbb{P}(A) < \delta$ and, in addition, $\sup_{n\geqslant 1} \mathbb{E}(|\kappa_n|) < \infty$.*

Lemma B.1. *(Chow and Teicher (1997)) If $\{\kappa_n,\, n \geqslant 1\}$ are random variables with $\sup_{n\geqslant 1} \mathbb{E}(|\kappa_n|^s) < \infty$ for some $s > 0$, then the sequence $\{|\kappa_n|^r,\, n \geqslant 1\}$ is uniformly integrable for all $0 < r < s$.*

Lemma B.2. *(Billingsley (1995)) If the random variables $\{|\kappa_n|^r,\, n \geqslant 1\}$ are uniformly integrable for some $r > 0$ and $\kappa_n \xrightarrow{d} \kappa$, then $\mathbb{E}(|\kappa|^p) < \infty$ and $\mathbb{E}(|\kappa_n|^r) \to \mathbb{E}(|\kappa|^p)$.*

C Empirical Appendix

Data Frequency

CIG	**Frequency**	**Percent**	**Cumulative**
0	393,939	80.70	80.70
1	1,469	0.30	81.00
2	2,986	0.61	81.61
3	3,759	0.77	82.38
4	2,890	0.59	82.98
5	6,838	1.40	84.38
6	2,618	0.54	84.91
7	1,758	0.36	85.27
8	1,644	0.34	85.61
9	335	0.07	85.68
10	32,720	6.70	92.38
11	117	0.02	92.41
12	801	0.16	92.57
13	396	0.08	92.65
14	128	0.03	92.68
15	4,568	0.94	93.61
16	93	0.02	93.63
17	39	0.01	93.64
18	259	0.05	93.69
19	19	0.00	93.70
20	25,333	5.19	98.89
21	49	0.01	98.90
22	30	0.01	98.90
23	39	0.01	98.91
24	36	0.01	98.92
25	417	0.09	99.00
26	7	0.00	99.01
27	3	0.00	99.01
28	12	0.00	99.01
29	2	0.00	99.01
30	2,993	0.61	99.62
31	4	0.00	99.62
32	3	0.00	99.62
33	2	0.00	99.62
34	5	0.00	99.62
35	97	0.02	99.64
36	5	0.00	99.65
37	2	0.00	99.65
38	1	0.00	99.65
39	0	0.00	99.65
40	1,474	0.30	99.95
> 40	254	0.05	100.00
Total	488,144		