Testing Continuity of a Density via g-order statistics in the Regression Discontinuity Design*

Federico A. Bugni
Department of Economics
Duke University
federico.bugni@duke.edu

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

February 13, 2020

Abstract

In the regression discontinuity design (RDD), it is common practice to assess the credibility of the design by testing the continuity of the density of the running variable at the cut-off, e.g., McCrary (2008). In this paper we propose an approximate sign test for continuity of a density at a point based on the so-called g-order statistics, and study its properties under two complementary asymptotic frameworks. In the first asymptotic framework, the number q of observations local to the cut-off is *fixed* as the sample size n diverges to infinity, while in the second framework q diverges to infinity slowly as n diverges to infinity. Under both of these frameworks, we show that the test we propose is asymptotically valid in the sense that it has limiting rejection probability under the null hypothesis not exceeding the nominal level. More importantly, the test is easy to implement, asymptotically valid under weaker conditions than those used by competing methods, and exhibits finite sample validity under stronger conditions than those needed for its asymptotic validity. In a simulation study, we find that the approximate sign test provides good control of the rejection probability under the null hypothesis while remaining competitive under the alternative hypothesis. We finally apply our test to the design in Lee (2008), a well-known application of the RDD to study incumbency advantage.

KEYWORDS: Regression discontinuity design, g-ordered statistics, sign tests, continuity, density.

JEL classification codes: C12, C14.

*The research of the first author was supported by NSF Grant SES-1729280 and the research of the second author was supported by NSF Grant SES-1530534. Special thanks go out to Tim Armstrong for pointing out a problem in Section 4.2 of an earlier manuscript. We also thank Xiaohong Chen, the Associate Editor, and three referees for helpful comments. We finally thank Jackson Bunting, Joe Long, and Deborah Kim for excellent research assistance.

1 Introduction

The regression discontinuity design (RDD) has been extensively used in recent years to retrieve causal treatment effects - see Lee and Lemieux (2010) and Imbens and Lemieux (2008) for exhaustive surveys. The design is distinguished by its unique treatment assignment rule where individuals receive treatment when an observed covariate, known as the running variable, crosses a known cut-off. Such an assignment rule allows nonparametric identification of the average treatment effect (ATE) at the cut-off, provided that potential outcomes have continuous conditional expectations at the cut-off (Hahn et al., 2001). The credibility of this identification strategy along with the abundance of such discontinuous rules have made RDD increasingly popular in empirical applications.

While the continuity assumption that is necessary for nonparametric identification of the ATE at the cut-off is fundamentally untestable, researchers routinely assess the plausibility of their RDD by exploiting two testable implications of a stronger identification assumption proposed by Lee (2008). We can describe the two implications as follows: (i) the treatment is locally randomized at the cut-off, which translates into the distribution of all observed baseline covariates being continuous at the cut-off; and (ii) individuals have imprecise control over the running variable, which translates into the density of the running variable being continuous at the cut-off. The practice of judging the reliability of RDD applications by assessing either of the two above stated implications (commonly referred to as manipulation, or falsification, or placebo tests) is ubiquitous in the empirical literature. Indeed, Table 4 surveys RDD empirical papers in four leading applied economic journals during the period 2011-2015. Out of 62 papers, 43 of them include some form of manipulation, falsification, or placebo test.

This paper proposes an approximate sign test for the null hypothesis on the second testable implication, i.e., the density of the running variable is continuous at the cut-off.[1] The approximate sign test has a number of distinctive attractive properties relative to existing methods used to test our null hypothesis of interest. First, the test does not require consistent non-parametric estimators of densities and simply exploits the fact that a certain functional of order statistics of the data is approximately binomially distributed under the null hypothesis. Second, our test controls the limiting null rejection probability under fairly mild conditions that, in particular, do not require existence of derivatives of the density of the running variable.[2] In addition, our test is valid in finite samples under stronger, yet plausible, conditions. Third, the asymptotic validity of our test holds under two alternative asymptotic frameworks; one in which the number q of observations local to the cut-off is *fixed* as the sample size n diverges to infinity, and one where q diverges to infinity slowly as

[1] It is important to emphasize that the null hypothesis we test in this paper is neither necessary nor sufficient for identification of the ATE at the cut-off; see Remark 2.2.

[2] We use the term null rejection probability as opposed to asymptotic size as a way to acknowledge that, for a given sample size, there always exists a heavily steep smooth function that is indistinguishable from a discontinuous one. Remark 4.5 discusses this further and provides important references.

n diverges to infinity. Importantly, both frameworks require similar and arguably mild conditions. Fourth, our test is simple to implement as it only involves computing order statistics, a constant critical value, and a single tuning parameter. This contrasts with existing alternatives that require local polynomial estimation of some order and either bias correction or under-smoothed bandwidth choices. Finally, we have developed a companion `Stata` package to facilitate the adoption of our test.[3]

The construction of our test is based on the simple intuition that, when the density of the running variable is continuous at the cut-off, the fraction of units under treatment and control local to the cut-off should be roughly the same. This means that the number of treated units out of the q observations closest to the cut-off, is approximately distributed as a binomial random variable with sample size q and probability $\frac{1}{2}$. To formalize this intuition, we exploit and develop properties of the so-called g-order statistics (see, e.g., Kaufmann and Reiss, 1992; Reiss, 1989) and consider the two asymptotic frameworks mentioned earlier to capture the local behavior of the density at the cut-off. In the first asymptotic framework, q is fixed as $n \to \infty$ to represent a finite sample situation where the effective number of observations used by the test is too small to credibly invoke approximations for "large" q. This may arise, for example, when the density is not so well behaved around the cut-off as illustrated in some of our simulations. This framework is similar to the one in Canay and Kamat (2018), who in turn exploit results from Canay, Romano and Shaikh (2017). It is worth noting that the hypothesis we test, the test statistic, the critical value, and most of the formal arguments are different from those in Canay and Kamat (2018) or Canay, Romano and Shaikh (2017). In the second asymptotic framework, q diverges to infinity slowly as $n \to \infty$ to represent a finite sample situation where the effective number of observations used by the test is large enough to invoke approximations for "large" q. This framework is similar to the one in McCrary (2008); Otsu et al. (2013); Cattaneo et al. (2019); Armstrong and Kolesár (2019), among others, and is in line with more traditional asymptotic arguments in non-parametric tests.

From a technical standpoint, this paper has several contributions relative to the existing literature. To start, our results exhibit two important differences relative to Canay and Kamat (2018) that go beyond the difference in the null hypotheses. First, we do not study our test as an approximate randomization test but rather as an approximate sign test. This not only requires different analytical tools, but also by-passes some of the challenges that would arise if we were to characterize our test as an approximate randomization test; see Remark 4.1 for a discussion on this. In addition, our approach in turn facilitates the analysis for the second asymptotic framework in which $q \to \infty$. Second, we develop results on g-order statistics as important intermediate steps towards our main results. Some of them may be of independent interest; e.g., Theorem 4.1. In addition, relative to the results in McCrary (2008); Otsu et al. (2013); Cattaneo et al. (2019); our test does not involve consistent estimators of density functions to either side of the cut-off and does

[3] The `Stata` package `rdcont` can be downloaded from `http://sites.northwestern.edu/iac879/software/`.

not require conditions involving existence of derivatives of the density of the running variable local to the cut-off. To the best of our knowledge, the formal asymptotic results we present are original to this paper.

It is relevant to note that similar binomial tests have been recently proposed in the RDD literature by Cattaneo et al. (2016, 2017) and Frandsen (2017). As we explain in more detail in Remark 3.3, there are important differences between these binomial tests and ours when it comes to the null hypothesis being tested, the formal arguments, and the practical implementation of the tests. Cattaneo et al. (2016, 2017) rely on finite sample arguments to justify their test construction for the hypothesis of local randomization. Frandsen (2017) also relies on finite sample arguments to test the hypothesis of manipulation of a discretely distributed running variable. In contrast, we test the hypothesis that the density of the running variable is continuous at the cut-off. Our focus on this particular null hypothesis prevents us from invoking finite sample arguments at the level of generality we consider and leads us to study the asymptotic properties of the approximate sign test. Our analysis also guides how to choose q in data-dependent way and this, in turn, leads to a distinctive implementation of the test that we propose.

The remainder of the paper is organized as follows. Section 2 introduces the notation and describes the null hypothesis of interest. Section 3 defines g-order statistics, formally describes the test we propose, and discusses all aspects related to its implementation including a data-dependent way of choosing q. Section 4 presents the main formal results of the paper, dividing those results according to the two alternative asymptotic frameworks we employ. In Section 5, we examine the relevance of our asymptotic analysis for finite samples via a simulation study. Finally, Section 6 implements our test to reevaluate the validity of the design in Lee (2008) and Section 7 concludes. The proofs of all results can be found in the Appendix.

2 Setup and notation

Let $Y \in \mathbf{R}$ denote the observed outcome of interest for an individual or unit in the population and $A \in \{0, 1\}$ denote an indicator for whether the unit is treated or not. Further denote by $Y(1)$ the potential outcome of the unit if treated and by $Y(0)$ the potential outcome if not treated. As usual, the observed outcome and potential outcomes are related to treatment assignment by the relationship

$$Y = Y(1)A + Y(0)(1 - A) \ . \tag{1}$$

The treatment assignment in the (sharp) RDD follows a discontinuous rule,

$$A = I\{Z \geq \bar{z}\} \ ,$$

where $Z \in \mathcal{Z} \equiv \text{supp}(Z)$ is an observed scalar random variable known as the running variable and $\bar{z}$ is the known threshold or cut-off value. For convenience we normalize $\bar{z} = 0$, which is without loss of generality as we can always redefine Z as $Z - \bar{z}$. This treatment assignment rule allows us to identify the average treatment effect (ATE) at the cut-off; i.e.,

$$E[Y(1) - Y(0)|Z = 0] \ .$$

In particular, Hahn et al. (2001) establish that identification of the ATE at the cut-off relies on the discontinuous treatment assignment rule and the assumption that

$$E[Y(1)|Z = z] \quad \text{and} \quad E[Y(0)|Z = z] \quad \text{are both continuous in } z \text{ at } z = 0 \ . \tag{2}$$

Reliability of the RDD thus depends on whether the mean outcome for units marginally below the cut-off identifies the true counterfactual for those marginally above the cut-off.

The continuity assumption in (2) is arguably weak, but fundamentally untestable. In practice, researchers routinely employ two specification checks in RDD that, in turn, are testable implications of a stronger sufficient condition proposed by Lee (2008, Condition 2b). The first check involves testing whether the distribution of pre-determined characteristics (conditional on the running variable) is continuous at the cut-off. See Shen and Zhang (2016) and Canay and Kamat (2018) for a recent treatment of this problem. The second check involves testing the continuity of the density of the running variable at the cut-off, an idea proposed by McCrary (2008). This second check is particularly attractive in settings where pre-determined characteristics are not available or where these characteristics are likely to be unrelated to the outcome of interest. Formally, we can state the hypothesis testing problem for the second check as

$$H_0 : f_Z^+(0) = f_Z^-(0) \quad \text{vs.} \quad H_1 : f_Z^+(0) \neq f_Z^-(0) \ , \tag{3}$$

where $f_Z^+(0)$ and $f_Z^-(0)$ are the one-sided limits of the probability density function of Z, i.e.,

$$f_Z^+(0) \equiv \lim_{z \downarrow 0} f_Z(z) \quad \text{and} \quad f_Z^-(0) \equiv \lim_{z \uparrow 0} f_Z(z) \ . \tag{4}$$

In RDD empirical studies, the aforementioned specification checks are often implemented (with different levels of formality) and referred to as falsification, manipulation, or placebo tests (see Table 4 for a survey).

In this paper we consider an approximate sign test for the null hypothesis of continuity in the density of the running variable Z at the cut-off $\bar{z} = 0$, i.e., (3). This test has three attractive features compared to existing approaches (see, e.g., McCrary, 2008; Otsu et al., 2013; Cattaneo et al., 2019). First, it does not require commonly imposed smoothness conditions on the density of Z, as it does not involve non-parametric estimation of such a density. Second, it exhibits finite sample validity under certain (stronger) easy to interpret conditions. Finally, it involves a single tuning parameter (a feature shared by the approach proposed by Cattaneo et al., 2019) as opposed to multiple ones in McCrary (2008). We discuss these features further in Section 4.

Remark 2.1. Gerard et al. (2016) study the consequences of discontinuities in the density of Z at the cut-off. In particular, the authors consider a situation in which manipulation occurs only for a subset of participants and use the magnitude of the discontinuity of $f(z)$ at $z = 0$ to identify the proportion of always-assigned units among all units close to the cut-off. Using this setup, Gerard et al. (2016) show that treatment effects in RDD are not point identified but the model still implies informative bounds. ■

Remark 2.2. It is important to emphasize that a running variable with a continuous density is neither necessary nor sufficient for the identification of the average treatment effect at the cut-off. For a discussion of this and some intuitive examples, see Lee (2008) and McCrary (2008). ■

3 Approximate sign test via g-ordered statistics

Let P be the distribution of Z and $Z^{(n)} = \{Z_i : 1 \leq i \leq n\}$ be a random sample of n i.i.d. observations from P. Let q be a small (relative to n) positive integer and $g : \mathcal{Z} \to \mathbf{R}$ be a measurable function such that $g(Z)$ has a continuous distribution function. For any $z, z' \in \mathcal{Z}$ define $\leq_g$ as

$$z \leq_g z' \quad \text{if} \quad g(z) \leq g(z') \ .$$

The ordering defined by $\leq_g$ is called a g-ordering on $\mathcal{Z}$. The g-order statistics $Z_{g,(i)}$ corresponding to $Z^{(n)}$ are defined as the values satisfying

$$Z_{g,(1)} \leq_g \cdots \leq_g Z_{g,(n)} \ ,$$

see, e.g., Reiss (1989, Section 2.1) and Kaufmann and Reiss (1992).

To construct our test statistic, we use the sign of the q values of $\{Z_i : 1 \leq i \leq n\}$ that are induced by the q smallest values of $\{g(Z_i) = |Z_i| : 1 \leq i \leq n\}$. That is, for $Z_{g,(1)}, \ldots, Z_{g,(q)}$, let

$$A_{g,(j)} \equiv I\{Z_{g,(j)} \geq 0\} \text{ for } 1 \leq j \leq q \ , \tag{5}$$

and

$$S_n \equiv \sum_{j \leq q} A_{g,(j)} \ . \tag{6}$$

The test statistic of our test only depends on the data via S_n and is defined as

$$T_q(S_n) \equiv \sqrt{q} \left| \frac{1}{q} S_n - \frac{1}{2} \right| \ . \tag{7}$$

In order to describe the critical value of our test it is convenient to recall that the cumulative distribution function (CDF) of a binomial random variable with q trials and probability of success $\frac{1}{2}$ is given by

$$\Psi_q(b) \equiv \frac{1}{2^q} \sum_{x=0}^{\lfloor b \rfloor} \binom{q}{x} I\{0 \leq b \leq q\} \ + \ I\{b > q\} \ , \tag{8}$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x. Using this notation the critical value for a significance level $\alpha \in (0,1)$ is given by

$$c_q(\alpha) \equiv \sqrt{q}\left(\frac{1}{2} - \frac{b_q(\alpha)}{q}\right) \ , \tag{9}$$

where $b_q(\alpha)$ is the unique value in $\{0, 1, \ldots, \lfloor \frac{q}{2} \rfloor\}$ satisfying

$$\Psi_q(b_q(\alpha) - 1) \leq \frac{\alpha}{2} < \Psi_q(b_q(\alpha)) \ . \tag{10}$$

The test we propose is then given by

$$\phi(S_n) = \begin{cases} 1 & \text{if } T_q(S_n) > c_q(\alpha) \\ a_q(\alpha) & \text{if } T_q(S_n) = c_q(\alpha) \\ 0 & \text{if } T_q(S_n) < c_q(\alpha) \end{cases} \ , \tag{11}$$

where

$$a_q(\alpha) \equiv 2^{q-1}\binom{q}{b_q(\alpha)}^{-1}\left[\alpha - 2\Psi_q(b_q(\alpha) - 1)\right] \ . \tag{12}$$

Intuitively, the test $\phi(S_n)$ exploits the fact that, under the null hypothesis in (3), the distribution of the treatment assignment should be locally the same to either side of the cut-off. That is, local to the cut-off, the treatment assignment behaves as purely randomized under the null hypothesis, so the fraction of units under treatment and control should be similar.

Remark 3.1. The test in (11) is possibly randomized. The non-randomized version of the test takes the form $I\{T_q(S_n) > c_q(\alpha)\}$ and, by definition, does not reject more often than $\phi(S_n)$ in (11). For our data-dependent choice of q that we describe in the next section, the randomized and non-randomized versions perform similarly in our simulations. ■

Remark 3.2. The value of $b_q(\alpha) \in \{0, 1, \ldots, \lfloor \frac{q}{2} \rfloor\}$ solving (10) is well-defined and unique for all $q \geq 1$ and $\alpha \in (0,1)$. To see this, let

$$q^*(\alpha) \equiv 1 - \frac{\log \alpha}{\log 2} \ . \tag{13}$$

When $q < q^*(\alpha)$, (10) uniquely holds for $b_q(\alpha) = 0$. In this case, $\phi(S_n)$ in (11) does not reject deterministically with positive probability. When $q \geq q^*(\alpha)$, the uniqueness of the solution is guaranteed by $\Psi_q(\cdot)$ being strictly increasing over $\{0, 1, \ldots, \lfloor \frac{q}{2} \rfloor\}$, $\Psi_q(0) = \frac{1}{2^q}$, and $\Psi_q(\frac{q}{2}) \geq \frac{1}{2}$. In this case, $\phi(S_n)$ in (11) deterministically rejects with positive probability. This shows that in order for the non-randomized version of the test to be non-trivial (see Remark 3.1), q needs to exceed $q^*(\alpha)$. To better appreciate these magnitudes, note that for $\alpha = 5\%$ this requires $q \geq 6$ while for $\alpha = 1\%$ this requires $q \geq 8$. Similarly, and given $b_q(\alpha)$, the value of $a_q(\alpha)$ in (12) is also uniquely defined and taking values in $[0, 1)$ by the same properties of $\Psi_q(\cdot)$. ■

Given q, the implementation of our test proceeds in the following five steps.

Step 1. Find the q observations closest to the cut-off, i.e., $Z_{g,(1)}, \ldots, Z_{g,(q)}$.

Step 2. Count the number of non-negative observations in $Z_{g,(1)}, \ldots, Z_{g,(q)}$, i.e., S_n as in (6).

Step 3. Compute test statistic $T_q(S_n)$ as in (7), $c_q(\alpha)$ as in (9), and $a_q(\alpha)$ as in (12).

Step 4. Compute the p-value of the non-randomized version of the test as

$$p_{\text{value}} = 2 \min \left\{ \Psi_q\left(S_n\right), \Psi_q\left(q - S_n\right) \right\} \ . \tag{14}$$

Step 5. Reject the null hypothesis in (3) using $\phi(S_n)$ in (11). If a non-randomized test is preferred, reject the null hypothesis if $p_{\text{value}} < \alpha$.

Remark 3.3. As we show in Theorems 4.1 and 4.2, the test $\phi(S_n)$ is an approximate sign or binomial test. As mentioned in the introduction, related binomial tests have been recently presented in the RDD context by Cattaneo et al. (2016), Cattaneo et al. (2017), and Frandsen (2017). The first two papers use a binomial test based on the number of observations of the running variable exceeding the cut-off in a window $[-h, h]$ for a varying bandwidth h. The authors propose to vary h until a "break-down" window size h^* is found, which is defined as the largest window such that the minimum p-value of the binomial test is larger than α for all nested (smaller) windows. The justification provided for the validity of such a test involves a finite sample argument: under the hypothesis of "local randomization/random assignment" in $[-h^*, h^*]$, a binomial test with probability π is exact. Frandsen (2017) considers an RDD model in which the running variable is discretely distributed and tests a different hypothesis from ours. Also motivated by finite sample arguments, he proposes a test that involves quantiles from binomial distributions. Contrary to these papers, our goal is not to validate a "local random assignment" hypothesis or to deal with discrete running variables in an RDD framework, but rather to test the continuity hypothesis in (3) when the running variable is continuous at the cut-off. As a result of this, we cannot exploit finite sample arguments and rather need to rely on the asymptotic analysis of our test. The formal results in Theorems 4.1, 4.2, and 4.3 are novel to this paper and, to the best of our knowledge, they provide the first formal results about approximate sign tests for the hypothesis in (3) in the RDD framework. ■

3.1 Data-dependent rule for q

In this section we discuss the practical considerations involved in the implementation of our test, highlighting how we addressed these considerations in the companion `Stata` package. The only tuning parameter of our test is the number q of observations closest to the cut-off. We propose a data-dependent way to choose q that combines a rule of thumb with a local optimization. We call this data-dependent rule the "informed rule of thumb" and its computation requires the two steps described below. For the sake of clarity, in this section we do not use the normalization $\bar{z} = 0$. Additional computational details are presented in Appendix D.

In Section 4.2 we consider the asymptotic framework where q diverges as $n \to \infty$. Under Assumption 4.1 and H_0 in (3), we show in that section that the value of q that sets the worst case asymptotic bias equal to the standard deviation is given by

$$q = n^{2/3} \left(\frac{4 f_Z^2(\bar{z})}{C_P} \right)^{2/3} , \tag{15}$$

where $f_Z(\bar{z})$ equals $f_Z^+(\bar{z}) = f_Z^-(\bar{z})$ under H_0, and C_P is the Lipschitz constant in Assumption 4.1(i'). Since the results in Theorem 4.3 also require $q^{3/2}/n \to 0$, we propose to start with an initial rule of thumb where $f_Z(\bar{z})$ and C_P are computed under the assumption $Z \sim N(\mu, \sigma^2)$ and the rate is set to $n^{1/2}$. This leads to

$$q_{\text{rot}} = n^{1/2} \left(\sigma \frac{4 \phi_{\mu,\sigma}^2(\bar{z})}{\phi_{\mu,\sigma}(\mu + \sigma)} \right)^{2/3} , \tag{16}$$

where we used that $C_P = |\phi'_{\mu,\sigma}(\mu + \sigma)| = \frac{1}{\sigma} \phi_{\mu,\sigma}(\mu + \sigma)$ when $Z \sim N(\mu, \sigma^2)$, and $\phi_{\mu,\sigma}(\cdot)$ and $\phi'_{\mu,\sigma}(\cdot)$ denote the density of $N(\mu, \sigma^2)$ and its derivative. This initial rule of thumb is location and scale invariant and, by definition, is inversely related to the asymptotic bias of the test statistic in the asymptotic framework of Section 4.2. In turn, the constant multiplying $n^{1/2}$ in (16) is fairly intuitive. First, it captures the idea that a steeper density at the cut-off should be associated with a smaller value of q. Intuitively, the steeper the density, the more it resembles a density that is discontinuous (Figure 1.(c) illustrates this in Section 5). Since the maximum slope is determined by the Lipschitz constant, the rule is inversely proportional to that. Second, it also captures the idea that q should be small if the cut-off is a point of low density. Intuitively, when $f_Z(\bar{z})$ is low, the q closest observations to $\bar{z}$ are likely to be "far" from $\bar{z}$ (Figure 1.(a) with $\mu = -2$ illustrates this in Section 5). One could alternatively replace the normality assumption with a non-parametric estimator of $f_Z(\bar{z})$ but it is unfortunately impossible to choose C_P adaptively for testing (3) (see, e.g., Low, 1997; Armstrong and Kolesár, 2018). Since any data-dependent rule for q will require a reference for C_P, we prefer to prioritize its simplicity and use normality for both $f_Z(\bar{z})$ and C_P.

The second step involves a local maximization of the asymptotic null rejection probability of the non-randomized version of the test. In particular, based on our results, we propose

$$q_{\text{irot}} = \underset{q \in \mathcal{N}(q_{\text{rot}})}{\operatorname{argmax}} \Psi_q(b_q(\alpha) - 1) , \tag{17}$$

where $\Psi_q(\cdot)$ is the CDF defined in (8), $b_q(\alpha)$ is defined in (10), and $\mathcal{N}(q_{\text{rot}})$ is a discrete neighborhood defined in (D-21) in Appendix D. This step helps the performance of the non-randomized version of the test (see Remark 3.1) as $\Psi_q(b_q(\alpha) - 1)$ is non-monotonic in q (see Figure 3) and so optimizing locally to q_{rot} over $\mathcal{N}(q_{\text{rot}})$ prevents choosing a value of q with a low value of $\Psi_q(b_q(\alpha)-1)$. In practice, we replace (μ,σ) with sample analogs to obtain the feasible informed rule of thumb $\hat{q}_{\text{irot}}$.

Remark 3.4. The recommended choice of q in (17) can be interpreted as the under-smoothed version of the rule that captures a bias-variance trade-off, where we impose normality to compute unknown constants. It also exploits the shape of the limiting null rejection probability of the non-randomized version of the test to derive a better choice of q. Even though q_{irot} is motivated by a root-mean-square error (RMSE) optimal choice of q, it is an under-smoothed rule of thumb that is not optimal in a formal sense. A formal study of an optimal choice of q in either of our asymptotic frameworks is an important topic of investigation that we leave for future research. ■

4 Asymptotic framework and formal results

In this section we derive the asymptotic properties of the test in (11) using two alternative asymptotic frameworks. The first one requires q to be fixed as $n\to\infty$, and represents a finite sample situation where the effective number of observations used by the test is too small to credibly invoke approximations for "large" q. The second framework requires $q\to\infty$ slowly as $n\to\infty$, and represents a finite sample situation where the effective number of observations used by the test is large enough to invoke approximations for "large" q.

There are three main features of our results that are worth highlighting: (i) our test exhibits similar properties under both asymptotic frameworks, (ii) the implementation of the test *does not* depend on which asymptotic framework one has in mind, and (iii) all formal results require similar, and arguably mild, conditions. We start by introducing these conditions.

Assumption 4.1. *The distribution function P is absolutely continuous on $(-\delta,\delta)$ for some $\delta>0$. On this set, the density function $f_Z(z)$ satisfies the following:*

(i) $f_Z(z)$ is bounded on $(-\delta,\delta)$ and has one-sided limits at zero given by $f_Z^+(0)$ and $f_Z^-(0)$.

(i') $\exists C_P\in(0,\infty)$ such that

$$|f_Z(z)-f_Z^+(0)|\le C_P|z| \text{ for } z\in(0,\delta) \text{ and } |f_Z(z)-f_Z^-(0)|\le C_P|z| \text{ for } z\in(-\delta,0)\ .$$

(ii) $f_Z^-(0)+f_Z^+(0)>0$.

Assumptions 4.1(i) and 4.1(i') each impose different degrees of smoothness on the density of Z local to the cut-off $\bar{z}=0$. Indeed, Assumption 4.1(i') strengthens Assumption 4.1(i) by replacing the requirement of left- and right-continuity at the cut-off with its Lipschitz version. In the formal results that follow, we use Assumption 4.1(i) in the asymptotic framework where q is fixed as $n\to\infty$ and Assumption 4.1(i') in the asymptotic framework where $q\to\infty$ as $n\to\infty$. Both assumptions allow for the distribution of Z to be discontinuous outside of a neighborhood of the cut-off.[4] More importantly, they do not require the density of Z to be differentiable anywhere. This is in contrast to McCrary (2008), who requires three continuous and bounded derivatives of the density of Z (everywhere except possibly at $\bar{z} = 0$), and Cattaneo et al. (2019) and Otsu et al. (2013), who require the density of Z to be twice continuously differentiable local to the cut-off (in the case of a local-quadratic approximation). Assumption 4.1(ii) rules out a situation where $f_Z^-(0) = f_Z^+(0) = 0$, which is implicitly assumed away in McCrary (2008) and Otsu et al. (2013), and is weaker than assuming a positive density of Z in a neighborhood of the cut-off as in Cattaneo et al. (2019). In Section 5 we explore the sensitivity of our results to violations of these conditions.

4.1 Results for fixed q

In this section we present two main results. The first result, Theorem 4.1, describes the asymptotic properties of S_n in (6) when q is fixed as $n \to \infty$. This result about g-order statistics with $g(\cdot) = |\cdot|$ represents an important milestone in proving the asymptotic validity of our test. The second result, Theorem 4.2, exploits Theorem 4.1 to show that the test in (11) controls the limiting rejection probability under the null hypothesis.

Theorem 4.1. *Let Assumptions 4.1(i) and 4.1(ii) hold and let $q \geq 1$ be fixed. Then,*

$$S_n \xrightarrow{d} S \sim \mathrm{Bi}(q, \pi_f)$$

as $n \to \infty$, where $\mathrm{Bi}(q, \pi_f)$ denotes the Binomial distribution with q trials and probability of success

$$\pi_f \equiv \frac{f_Z^+(0)}{f_Z^-(0) + f_Z^+(0)} .$$

Theorem 4.1, although fairly intuitive, does not follow from standard arguments. First, the random variables $\{A_{g,(j)} : 1 \leq j \leq q\}$ are indicators of g-order statistics so, in general, they are neither independent nor identically distributed. Second, applying results from the literature on g-order statistics (e.g., Kaufmann and Reiss, 1992, Theorem 1) requires $g(Z) = |Z|$ to have a continuous distribution function everywhere on its domain. Under Assumption 4.1(i) this is only true in $[0, \delta)$, and mass points are allowed outside of $[0, \delta)$. In the proof of Theorem 4.1 we use a smoothing transformation of Z as an intermediate step and then accommodate the results in Kaufmann and Reiss (1992, Theorem 1) to reach the desired conclusion.

The following result, which heavily relies on Theorem 4.1, is the main result of this section and characterizes the asymptotic properties of the test $\phi(S_n)$ in (11).

Theorem 4.2. *Let Assumptions 4.1(i) and 4.1(ii) hold and let $q \geq 1$ be fixed. Then, the following holds for $\alpha \in (0, 1)$:*

[4] In Appendix C we also allow for situations with a mass point at the cut-off, i.e., $P\{Z = 0\} > 0$.

(a) Under H_0 in (3),

$$\lim_{n\to\infty} E[\phi(S_n)] = 2\Psi_q(b_q(\alpha)-1) + \frac{a_q(\alpha)}{2^{q-1}}\binom{q}{b_q(\alpha)} = \alpha \ .$$

(b) Under H_1 in (3), $\lim_{n\to\infty} E[\phi(S_n)] \geq \alpha$.

Theorem 4.2 shows that $\phi(S_n)$ behaves asymptotically, as $n \to \infty$, as the two-sided sign test in an experiment where one observes $S \sim \text{Bi}(q,\pi)$ and wishes to test the hypotheses $H_0 : \pi = \frac{1}{2}$ versus $H_1 : \pi \neq \frac{1}{2}$. For this reason, we refer to $\phi(S_n)$ as an approximate sign test.

Remark 4.1. The test $\phi(S_n)$ could be alternatively characterized as an "approximate" randomization test, see Canay et al. (2017) for a general description of such tests. However, such a characterization would make the analysis of the formal properties of the test more complicated and, in particular, the results in Canay et al. (2017) would not immediately apply due to two fundamental challenges. First, Assumption 3.1(iii) in Canay et al. (2017) is immediately violated in our setting. Second, such an approach would require an asymptotic approximation to the joint distribution of $\{A_{g,(j)} : 1 \leq j \leq q\}$, which in turn would require a strengthening of Lemma B.4. Our proof approach avoids both of these technicalities by directly exploiting the binary nature of $\{A_{g,(j)} : 1 \leq j \leq q\}$ and by simply approximating the distribution of S_n, which is a scalar, as in Theorem 4.1. ■

Remark 4.2. It is possible to show that $\phi(S_n)$ in (11) is level α in finite samples whenever the distribution of Z is continuous and symmetric about the cut-off. In this case, the fundamental result in Lemma B.4 holds for S_n with $P\{Z > 0 \mid |Z| < r\} = \frac{1}{2}$ for any $r > 0$, and the proof of Theorem 4.2 can in turn be properly modified to show $E[\phi(S_n)] = \alpha$ for all $n \geq 1$. ■

4.2 Results for large q

In this section we study the properties of $\phi(S_n)$ in (11) in an asymptotic framework where q diverges to infinity as $n \to \infty$. This asymptotic framework is in line with traditional non-parametric arguments and so our results depend on the assumed smoothness of the density of Z and the rate at which q is allowed to grow. Importantly, the results in this section follow from Assumption 4.1(i')-(ii) and so, accounting for the differences between Assumptions 4.1(i) and 4.1(i'), the result below shows that the asymptotic properties of the approximate sign test under both asymptotic frameworks require similar, and arguably mild, conditions.

Theorem 4.3. *Let Assumptions 4.1(i') and 4.1(ii) hold and let q be such that $q \to \infty$ and $\frac{q^{3/2}}{n} \to 0$ as $n \to \infty$. Then,*

$$\sqrt{q}\left(\frac{1}{q}S_n - \pi_f\right) \xrightarrow{d} N\left(0, \pi_f(1-\pi_f)\right) \ , \tag{18}$$

where π_f is as in Theorem 4.1. Moreover, the following holds for $\alpha \in (0,1)$:

(a) Under H_0 in (3), $\lim_{n\to\infty} E[\phi(S_n)] = \alpha$.

(b) Under H_1 in (3), $\lim_{n\to\infty} E[\phi(S_n)] = 1$.

(c) Under a sequence of alternative distributions local to H_0 satisfying $\sqrt{q}(\pi_f - \frac{1}{2}) \to \Delta \neq 0$,

$$\lim_{n\to\infty} E[\phi(S_n)] = P\{|\zeta + 2\Delta| > z_{\alpha/2}\} > \alpha \ ,$$

where $\zeta \sim N(0,1)$ and $z_{\alpha/2}$ is the $(1-\frac{\alpha}{2})$-quantile of ζ.

Theorem 4.3, although fairly intuitive again, does not follow from standard arguments. In particular, given that the random variables $\{A_{g,(j)} : 1 \le j \le q\}$ are neither independent nor identically distributed, the result does not follow from a simple application of the central limit theorem. We instead adapt Kaufmann and Reiss (1992, Theorem 1) and prove the result using first principles and the normal approximation to the binomial distribution.

Given the result in Theorem 4.3, we can provide some insight on the properties of the data-dependent rule for choosing q that we describe in Section 3.1. Specifically, we focus on providing interpretation to q_{rot} in (16), as q_{irot} in (17) is a modification of q_{rot} to improve the performance of the non-randomized version of the test. Under H_0 in (3) and Assumption 4.1(i')-(ii), the results in Armstrong and Kolesár (2019) imply that

$$\frac{\sqrt{q}(\frac{1}{q}S_n - \pi_f)}{\sqrt{\pi_f(1-\pi_f)}} = \zeta_n + B_{n,q} + o_p(1) \ , \tag{19}$$

where $\zeta_n \xrightarrow{d} \zeta \sim N(0,1)$ and $B_{n,q}$ is a standardized bias term satisfying

$$|B_{n,q}| \le \frac{q^{3/2}}{n} \frac{C_P}{4f_Z^2(0)} \tag{20}$$

with $f_Z(0)$ equals $f_Z^+(0) = f_Z^-(0)$ under H_0. Denote by t^* the right-hand side of (20) and note that this can be interpreted as the worst (in absolute value) ratio of bias to standard deviation (sd) of the left-hand side of (19). We can then solve for q to obtain

$$q^* = n^{2/3}(t^*)^{2/3}\left(\frac{4f_Z^2(0)}{C_P}\right)^{2/3} \ . \tag{21}$$

This derivation shows that the requirement $\frac{q^{3/2}}{n} \to 0$ is analogous to under-smoothing as this is the rate condition that removes the worst-case asymptotic bias.[5] This immediately gives two alternative interpretations to the data-dependent rule q_{rot} in (16) (see Armstrong and Kolesár, 2019, Section

[5] A previous version of this paper did not include Assumption 4.1(i') and the requirement $q^{3/2}/n \to 0$, which is required to control the asymptotic bias term. We thank Tim Armstrong for pointing this out to us.

4.2, for a more detailed description). In order to describe these two interpretations, note that by (21) and (16) we obtain that $q_{\text{rot}} = q^*$ whenever

$$t^* = \left(\frac{n^{1/2}}{n^{2/3}}\right)^{3/2} \left[\frac{\phi^2_{\mu,\sigma}(0)}{f^2_Z(0)} \frac{C_P}{\frac{1}{\sigma}\phi_{\mu,\sigma}(\mu+\sigma)}\right] . \tag{22}$$

Assume for a moment that the rule-of-thumb assumption of normality is correct (which means that the term within brackets in (22) equals 1). Then, q_{rot} is equivalent to q^* for a worst ratio of bias to sd t^* given by

$$t^* = \left(\frac{n^{1/2}}{n^{2/3}}\right)^{3/2} \quad \Rightarrow \quad t^* = 0.12 \text{ for } n = 5,000 .$$

This implies the size of $\phi(S_n)$ for $\alpha = 5\%$ and $n = 5,000$ would approximately be $P\{|\zeta + 0.12| > z_{\alpha/2}\} = 5.16\%$. In this sense, q_{rot} makes the size distortion of the bias negligible when $n = 5,000$. Next, suppose that the rule-of-thumb assumption of normality over-estimates the ratio $f^2_Z(0)/C_P$. In other words, suppose that the term within brackets in (22) equals a constant $a > 1$. In this case, q_{rot} would be equivalent to q^* for a worst ratio of bias to sd t^* given by

$$t^* = a\left(\frac{n^{1/2}}{n^{2/3}}\right)^{3/2} \quad \Rightarrow \quad t^* = 0.36 \text{ for } n = 5,000 \text{ and } a = 3 .$$

This implies that the size of $\phi(S_n)$ for $\alpha = 5\%$ and $n = 5,000$ would approximately be $P\{|\zeta+0.36| > z_{\alpha/2}\} = 6.38\%$. When $f_Z(0) = \phi_{\mu,\sigma}(0)$, this means that even if the true Lipschitz constant C_P is three times larger than the one imposed by normality, $\phi(S_n)$ would still exhibit mild over-rejection under the null hypothesis. The price we pay for this robustness under the null hypothesis (in terms of performance and mild requirements) is possibly a lower power under the alternative hypothesis, a feature that we explore in the simulations of Section 5.

Remark 4.3. It may be tempting to use the first part of Theorem 4.3 to consider a variation of the test we propose; namely the test that rejects H_0 when $T_q(S_n) > \frac{1}{2}z_{\alpha/2}$ and $z_{\alpha/2}$ is the $(1-\frac{\alpha}{2})$-quantile of a standard normal random variable. However, we do not recommend this variation as it provides no theoretical advantages over $\phi(S_n)$ in the asymptotic framework where $q \to \infty$, and it is not formally justified in the asymptotic framework where q is fixed (in particular, such a variation will not inherit the finite sample properties discussed in Remark 4.2). ■

Remark 4.4. As pointed out by a referee, in the asymptotic framework where $q \to \infty$, the test statistic $T_q(S_n)$ can be shown to be proportional to a Wald-type statistic

$$W_n = |\hat{f}_Z(h_n) - \hat{f}_Z(-h_n)| ,$$

where $\hat{f}_Z(z)$ is a non-parametric kernel density estimator of $f_Z(z)$ implemented with a uniform on $[-1,1]$ kernel and bandwidth h_n. Under some conditions it will follow that W_n is asymptotically normal and a test for H_0 could be constructed by using the quantile of a normal distribution (possibly by additionally estimating the asymptotic variance). One could go a step further and use the bound on the bias term $B_{n,q}$ to construct a test that explicitly accounts for the asymptotic bias of the test following the approach proposed by Armstrong and Kolesár (2019). However, the interpretation of $\phi(S_n)$ as a test based on a Wald-type statistic with a normal critical value exclusively holds in the asymptotic framework where $q \to \infty$ and does not apply in the asymptotic framework with fixed q. For this reason, we do not emphasize this interpretation here. ■

Remark 4.5. The recent literature has obtained impossibility results in the RDD setting that apply to the hypothesis testing problem in (3); see, e.g., Low (1997), Kamat (2017), Armstrong and Kolesár (2018), and Bertanha and Moreira (2019). An implication of these impossibility results is that $\phi(S_n)$ cannot control size in a uniform sense without further restricting the set of data generating processes. These findings are reflected in the bound on the bias term presented in (20), where higher values of C_P or lower values of $f_Z(0)$ can make such a bound arbitrarily high for given values of q and n. We would therefore expect the performance of $\phi(S_n)$ to deteriorate in cases where the density at the cut-off is very low or very steep, as highlighted by the simulations we present next. ■

5 Simulations

In this section we examine the finite-sample performance of the test in (11) with a simulation study. Instead of just presenting designs where this test excels relative to competing ones, we present an array of data generating processes that hopefully illustrate its relative strengths and weaknesses. The data for the study are simulated as i.i.d. samples from the following designs.

Design 1: For $\mu \in \{-2, -1, 0\}$, $Z \sim N(\mu, 1)$.

Design 2: For $\lambda \in \{\frac{1}{3}, 1\}$,

$$Z \sim \begin{cases} V_1 & \text{with prob. } \lambda \\ V_2 & \text{with prob. } (1-\lambda) \end{cases},$$

where $V_1 \sim 2\text{Beta}(2,4) - 1$ and $V_2 \sim 1 - 2\text{Beta}(2,8)$.

Design 3: For $(\lambda_1, \lambda_2, \lambda_3) = (0.4, 0.1, 0.5)$,

$$Z \sim \begin{cases} V_1 & \text{with prob. } \lambda_1 \\ V_2 & \text{with prob. } \lambda_2 \\ V_3 & \text{with prob. } \lambda_3 \end{cases},$$

where $V_1 \sim N(-1, 1)$, $V_2 \sim N(-0.2, 0.2)$, and $V_3 \sim N(3, 2.5)$.

Design 4: For $\kappa \in \{0.05, 0.10, 0.25\}$, the density of Z is given by

$$f_Z(z) = \begin{cases} 0.75 & \text{if } z \in [-1, -\kappa] \\ 0.75 - \frac{1}{4\kappa}(z + \kappa) & \text{if } z \in [-\kappa, \kappa] \\ 0.25 & \text{if } z \in [\kappa, 1] \end{cases} .$$

Design 5: For $\kappa \in \{0.05, 0.10, 0.25\}$, the density of Z is given by

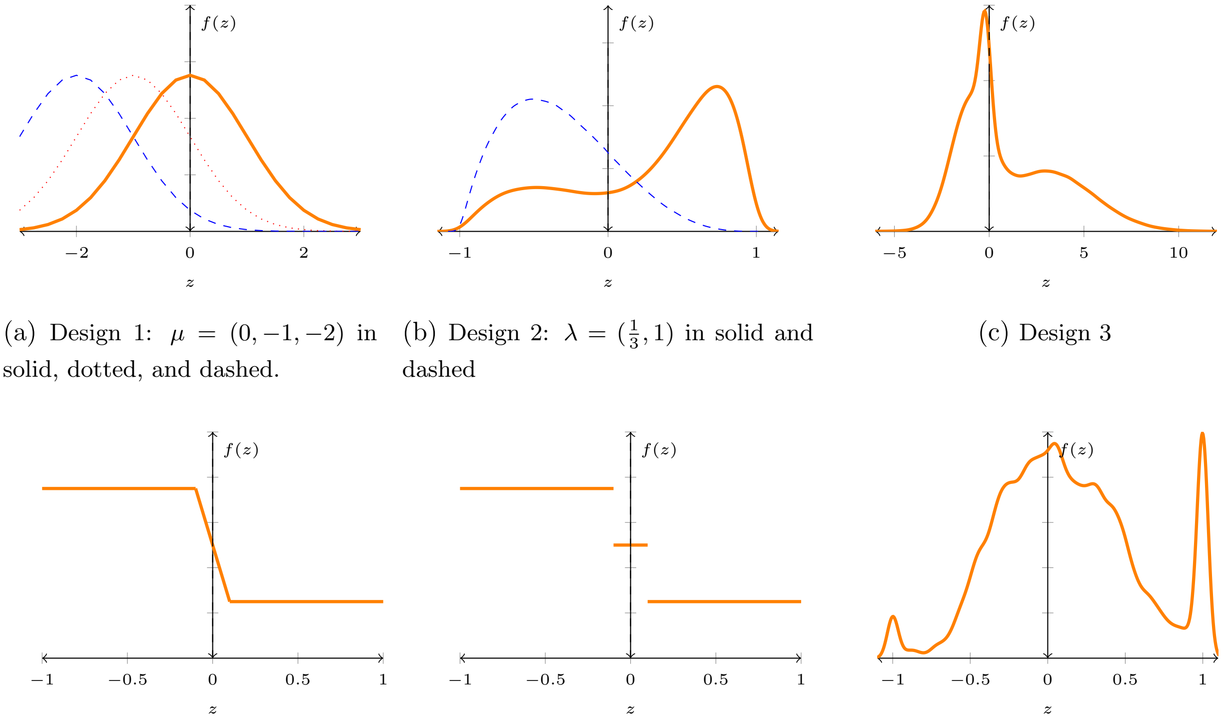
$$f_Z(z) = \begin{cases} 0.25 & \text{if } z \in [-1, -\kappa] \\ 0.50 & \text{if } z \in [-\kappa, \kappa] \\ 0.75 & \text{if } z \in [\kappa, 1] \end{cases} .$$

Design 6: We first non-parametrically estimate the density of the running variable in Lee (2008, see Section 6 for details) and then take i.i.d. draws from such a density.

Design 1 in Figure 1(a) is the canonical normal case and, by Remark 4.2, our test is expected to control size in finite samples when $\mu = 0$ but not when $\mu \in \{-2, -1\}$. Indeed, $\mu = -2$ is a challenging case due to the low probability of getting observations to the right of the cut-off. Design 2 in Figure 1(b) is taken from Canay and Kamat (2018). Design 3 in Figure 1(c) is a parametrization of the taxable income density in Saez (2010, Figure 8). This design exhibits a spike (almost a kink) to the left of the cut-off which is essentially a violation of the smoothness assumptions required by McCrary (2008) and Cattaneo et al. (2019). It also exhibits a steep density at the cut-off, which also makes it a difficult case in general. Similar to Design 3, Design 4 in Figure 1(d) also illustrates the difficulty in distinguishing a discontinuity from a very steep slope; see Low (1997), Kamat (2017), Armstrong and Kolesár (2018), and Bertanha and Moreira (2019) for a formal discussion. Here we can study the sensitivity to the slope by changing the value of κ. Design 5 in Figure 1(e) requires δ in Assumption 4.1(a) to be such that $\delta < \kappa$ in order for our approximations to be accurate, but as opposed to Design 4, it is locally symmetric around the cut-off. As κ gets smaller, we expect our test to perform worse if q is not chosen carefully. Finally, Design 6 in Figure 1(f) draws data i.i.d. from the non-parametric density estimate of the running variable in Lee (2008), i.e., Z is the difference in vote shares between Democrats and Republicans.

We consider sample sizes $n \in \{1,000; 5,000\}$, a nominal level of $\alpha = 10\%$, and perform $10,000$ Monte Carlo repetitions. Designs 1 to 6 satisfy the null hypothesis in (3). We additionally consider the same models under the alternative hypothesis by randomly changing the sign of observations in the interval $z \in [0, 0.1]$ with probability $\Pr = 0.2 - 2z$. We report results for the following tests.

AS-NR and **AS-R**: the approximate sign test we propose in this paper in its two versions. The randomized version (AS-R) in (11) and the non-randomized version (AS-NR) that rejects



(a) Design 1: $\mu = (0, -1, -2)$ in solid, dotted, and dashed.

(b) Design 2: $\lambda = (\frac{1}{3}, 1)$ in solid and dashed

(c) Design 3

(d) Design 4: $\kappa = 0.1$

(e) Design 5: $\kappa = 0.1$

(f) Design 6

Figure 1: Density functions $f(z)$ for Designs 1 to 6 used in the Monte Carlo simulations

when p_{value} in (14) is below α, see Remark 3.1. The tuning parameter q is set to

$$q \in \{20, 50, 75, \hat{q}_{\text{irot}}\} ,$$

where $\hat{q}_{\text{irot}}$ is the feasible informed rule of thumb described in Section 3 and Appendix D.

McC: the test proposed by McCrary (2008). We implement this test using the function `DCdensity` from the `R` package `rdd` (`Ver` 0.57), with the default choices for the bandwidth parameter and kernel type.

CJM: the test proposed by Cattaneo et al. (2019). We implement this test using the `rddensity` function from the `R` package `rddensity` (`Ver` 1.0). We use jackknifed standard errors and bias correction, as these are the default choices in the paper.

Tables 1 and 2 report rejection probabilities under the null and alternative hypotheses for the six designs we consider and for sample sizes of $n = 1,000$ and $n = 5,000$, respectively. We start by discussing the results under the null hypothesis. AS-NR delivers rejection probabilities under the null hypothesis closer to the nominal level than those delivered by McC and CJM in most of the designs. The two empirically motivated designs (Designs 3 and 6) illustrate this feature clearly. Designs 4 and 5 also show big differences in performance, both in cases where AS-NR delivers

	Rejection Rate under H_0							Rejection Rate under H_1						
	AS-NR				AS-R	McC	CJM	AS-NR				AS-R	McC	CJM
	q				q			q				q		
Design	20	50	75	$\hat{q}_{\text{irot}}$	$\hat{q}_{\text{irot}}$			20	50	75	$\hat{q}_{\text{irot}}$	$\hat{q}_{\text{irot}}$		
D1: $\mu = 0$	4.4	6.8	6.6	10.0	10.1	9.2	8.2	11.1	19.6	18.1	25.2	25.4	17.0	11.4
D1: $\mu = -1$	4.3	8.1	12.4	10.5	10.6	11.9	9.4	10.7	21.5	26.3	24.8	24.9	21.2	10.4
D1: $\mu = -2$	12.4	84.5	99.8	8.3	11.3	11.4	7.6	17.2	87.8	99.9	12.0	15.4	11.8	7.8
D2: $\lambda = 1$	4.0	7.0	7.9	10.4	10.6	11.2	9.6	9.4	13.4	9.7	19.5	19.7	26.4	16.8
D2: $\lambda = \frac{1}{3}$	4.2	7.0	10.3	10.6	10.7	10.7	7.6	11.9	32.0	42.6	32.1	32.3	34.6	18.2
D3	5.1	17.5	39.2	24.6	24.9	99.9	86.1	12.9	38.6	60.9	48.0	48.3	100.0	89.2
D4: $\kappa = 0.25$	4.0	8.1	12.2	10.9	11.0	11.9	9.3	12.5	32.4	42.6	34.8	35.0	33.2	18.6
D4: $\kappa = 0.10$	4.6	16.9	42.0	16.3	16.5	48.1	24.4	14.8	50.5	77.3	46.4	46.6	80.2	42.1
D4: $\kappa = 0.05$	6.9	48.0	86.2	35.9	36.1	84.9	59.9	20.2	79.4	97.4	66.8	67.0	96.4	77.4
D5: $\kappa = 0.25$	4.3	7.4	7.1	10.4	10.5	21.0	13.5	11.2	21.6	22.3	26.8	27.0	17.2	31.0
D5: $\kappa = 0.10$	4.1	6.4	6.4	9.9	10.1	17.9	33.1	11.0	21.5	21.8	26.1	26.3	23.2	13.5
D5: $\kappa = 0.05$	4.0	7.4	35.1	9.7	9.8	40.3	25.7	10.7	27.4	72.6	27.4	27.6	75.7	42.0
D6	4.2	6.2	6.6	9.4	9.6	9.5	10.6	11.5	25.1	27.8	32.8	33.0	37.0	23.8

Table 1: Rejection probabilities (in %) under H_0 and H_1 across Designs 1-6 and for $n = 1,000$.

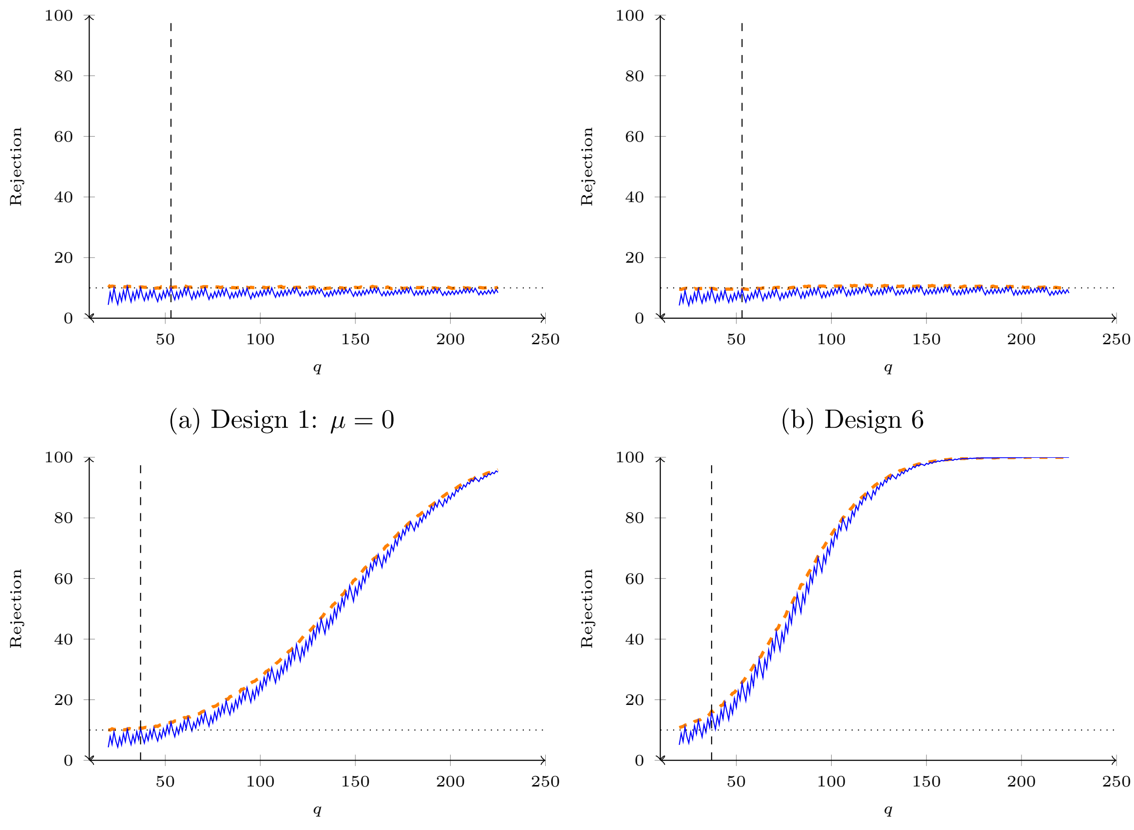
	Rejection Rate under H_0							Rejection Rate under H_1						
	AS-NR				AS-R	McC	CJM	AS-NR				AS-R	McC	CJM
	q				q			q				q		
Design	20	50	75	$\hat{q}_{\text{irot}}$	$\hat{q}_{\text{irot}}$			20	50	75	$\hat{q}_{\text{irot}}$	$\hat{q}_{\text{irot}}$		
D1: $\mu = 0$	4.1	6.4	6.8	9.8	10.0	9.2	8.8	12.5	30.5	38.4	63.7	63.9	50.0	33.2
D1: $\mu = -1$	4.0	6.3	6.5	9.5	9.7	11.8	9.0	12.0	30.2	38.4	39.1	39.4	47.2	16.6
D1: $\mu = -2$	4.5	12.4	26.0	10.2	10.6	12.2	9.6	11.8	31.5	47.9	21.2	21.7	15.7	12.5
D2: $\lambda = 1$	4.0	6.5	6.0	9.7	9.8	9.7	9.8	12.1	29.0	35.8	50.9	51.2	68.6	34.4
D2: $\lambda = \frac{1}{3}$	4.4	6.8	6.7	10.0	10.2	10.3	8.1	12.6	32.5	44.0	46.2	46.5	87.0	59.2
D3	4.2	7.2	7.9	17.2	17.4	100.0	93.9	12.4	33.6	46.6	73.7	73.9	100.0	97.8
D4: $\kappa = 0.25$	4.7	6.3	6.4	11.2	11.4	11.6	10.2	13.5	33.8	45.5	69.9	70.2	78.8	61.5
D4: $\kappa = 0.10$	4.2	7.1	8.1	16.9	17.0	68.6	22.8	13.3	36.3	51.5	80.0	80.2	99.5	72.4
D4: $\kappa = 0.05$	4.2	8.3	12.0	36.7	36.9	99.7	93.5	14.5	42.0	62.1	91.9	92.0	100.0	99.7
D5: $\kappa = 0.25$	4.0	6.2	6.0	9.7	9.8	33.4	17.0	12.1	30.8	40.8	60.1	60.4	55.2	78.5
D5: $\kappa = 0.10$	4.0	6.3	6.4	10.0	10.2	50.9	43.9	12.6	31.3	40.9	60.8	61.1	33.2	13.3
D5: $\kappa = 0.05$	4.3	6.6	6.9	10.5	10.7	45.7	30.9	12.8	31.7	40.3	60.8	61.1	98.2	63.7
D6	4.1	6.6	6.5	9.4	9.5	12.9	13.4	12.0	31.7	41.9	70.3	70.5	90.2	65.2

Table 2: Rejection probabilities (in %) under H_0 and H_1 across Designs 1-6 and for $n = 5,000$.

rejection rates equal to the nominal level (Design 5) and McC and CJM severely over-reject; as well as in cases where all tests over-reject (Design 4, $\kappa = 0.05$) but AS-NR is relatively closer to the nominal level. A particularly difficult case for AS-NR is Design 1 with $\mu = -2$, where the probability of getting observations to the right of the cut-off is below 2%. This showcases the satisfactory performance of our data-dependent rule $\hat{q}_{\text{irot}}$, which takes the lowest value in that

	D1 $\mu = 0$	D1 $\mu = -1$	D1 $\mu = -2$	D2 $\lambda = \frac{1}{3}$	D2 $\lambda = 1$	D3	D4 $\kappa = 0.25$	D4 $\kappa = 0.01$	D4 $\kappa = 0.05$	D5 $\kappa = 0.25$	D5 $\kappa = 0.01$	D5 $\kappa = 0.05$	D6
							$n = 1,000$						
AS-NR	53.0	37.0	8.5	37.0	37.0	51.7	40.5	39.3	39.2	44.2	39.7	39.2	53.0
McC	716.2	440.6	113.7	209.7	415.3	606.1	361.3	358.1	377.9	365.6	337.4	350.8	564.4
CJM	574.7	434.1	85.1	142.8	456.6	493.7	351.4	412.8	437.9	359.2	338.5	416.9	498.8
							$n = 5,000$						
AS-NR	147.0	54.1	18.0	119.0	62.0	119.0	119.0	119.0	119.0	119.0	119.0	119.0	146.9
McC	2964.3	1775.6	506.4	774.8	1782.3	2944.8	1526.3	1424.2	1612.0	1623.9	1318.4	1365.4	2258.3
CJM	2312.5	1883.5	348.0	489.2	2007.7	1879.5	1263.9	1660.5	2076.4	1363.6	1194.7	1703.4	2274.9

Table 3: Average effective sample size across simulations for each design. Effective sample size is defined as follows: $\hat{q}_{\text{irot}}$ for AS-NR, number of obs. in $[-h_n, h_n]$ for McC, with h_n being the bandwidth used to estimate the density to the left and to the right of the cutoff, and number of obs. in $[-h_{n,L}, h_{n,R}]$ for CJM, with $h_{n,L}$ and $h_{n,R}$ being the bandwidths used to estimate the density the left and to the right of the cutoff, respectively.



(a) Design 1: $\mu = 0$

(b) Design 6

(c) Design 1: $\mu = -1$

(d) Design 3

Figure 2: Rejection probabilities of AS-NR (solid blue line) and AS-R (dashed orange line) as a function of q for $n = 1,000$. The vertical dashed line denotes the value of $\hat{q}_{\text{irot}}$ and the horizontal dotted line the value of α.

particular design. Tables 1 and 2 also show negligible differences between the randomized (AS-R) and non-randomized (AS-NR) versions of our test, consistent with our discussion in Remark 3.1.

To describe the performance of the different tests under the alternative hypothesis, we focus on designs where the rejection probability under the null hypothesis is close to the nominal level for all tests: Design 1, Design 2, Design 4 with $\kappa = 0.25$, and Design 6. In those cases, we see that AS-NR has competitive power, and can sometimes even be the test with the highest rejection probability under the alternative hypothesis. For $n = 1,000$, AS-NR delivers the highest rejection probability under the alternative hypothesis in Design 1 for all values of μ and Design 4 with $\kappa = 0.25$. In the rest of the cases under consideration, McC exhibits the highest power and is followed by AS-NR. The results for $n = 5,000$ are qualitatively similar, with a few exceptions. McC has the highest rejection probability in Design 1 with $\mu = -1$, and CJM are have the second highest rejection probability in Design 2 with $\lambda = \frac{1}{3}$.

Table 3 shows the mean values of $\hat{q}_{\text{irot}}$ across simulations for all designs and sample sizes. As described in Section 3, $\hat{q}_{\text{irot}}$ takes into account both the slope and the magnitude of the density at the cut-off. As a result, $\hat{q}_{\text{irot}}$ is relatively high in designs with flat density at the cut-off and high $f_Z(0)$ (e.g., Design 1 with $\mu = 0$) and relatively low in designs with steep slopes or low $f_Z(0)$ (e.g., Design 1 with $\mu = -2$ or Design 2 with $\lambda = 1$). Table 3 also reports the average number of observations in $[-h_{n,L}, h_{n,R}]$ for McC and CJM, where $h_{n,L}$ and $h_{n,R}$ are the left and right bandwidths used to estimate $f_Z^-(0)$ and $f_Z^+(0)$, respectively (in the case of McC, $h_{n,L} = h_{n,R}$). In comparison, AS-NR uses significantly fewer observations than either McC or CJM, a feature that may support the the asymptotic framework in Section 4.1. Finally, and to gain further insight on the sensitivity of our test to the choice of q, Figure 2 displays the rejection probabilities of AS-NR and AS-R as a function of q in two types of designs. In the top row we illustrate two designs where the rejection probability is mostly insensitive to the choice of q (Design 1 with $\mu = 0$ and Design 6). These are designs where the density is rather flat around the cut-off so increasing q does not deteriorate the performance of our test. In the bottom row we illustrate two designs where the rejection probability is highly sensitive to the choice of q (Design 1 with $\mu = -1$ and Design 3). These are designs that feature a steep density at the cut-off (also low in Design 1) so increasing q very quickly deteriorates the performance of the test under the null hypothesis. The data-dependent rule $\hat{q}_{\text{irot}}$ is displayed in each case with a vertical dashed line and seems to be doing a good job at choosing relatively smaller values in the sensitive cases.

We conclude this section with two final remarks. First, one could compare the results in Tables 1 and 2 for a fix value of q to appreciate the results in Section 4.1. For example, taking $q = 75$, the rejection probability in Design 1 with $\mu = -2$ and Design 3 are 99.8 and 40.1, respectively, when $n = 1,000$. The same numbers when $n = 5,000$ are 26 and 8.4, respectively, which are closer to the nominal level as predicted by our results. Second, at the request of a referee, the results for $n = 1,000,000$ and $\alpha = 1\%$ are available upon request. Notably, AS-NR with the data-dependent rule $\hat{q}_{\text{irot}}$ delivers rejection probabilities under H_0 equal to α across all designs when $n = 1,000,000$ whereas McC and/or CJM still significantly over-reject for Designs 3, 5, and 6.

6 Empirical illustration

In this section we briefly reevaluate the validity of the design in Lee (2008). Lee studies the benefits of incumbency on electoral outcomes using a discontinuity constructed with the insight that the party with the majority wins. Specifically, the running variable Z is the difference in vote shares between Democrats and Republicans in a house election; see Figure 1(f) for a graphical illustration of the density of Z. The assignment rule then takes a cut-off value of zero that determines the treatment of incumbency to the Democratic candidate, which is used to study their outcome in the next election. The total number of observations is 6,559 with 2,740 below the cut-off. The dataset is publicly available at `http://economics.mit.edu/faculty/angrist/data1/mhe`.

Lee assesses the credibility of the design in this application by inspecting discontinuities in means of the baseline covariates, but mentions in footnote 19 the possibility of using the test proposed by McCrary (2008). Here, we frame the validity of the design in terms of the hypothesis in (3) and use the approximate sign test we describe in Section 3, using $\hat{q}_{\text{irot}}$ as our default choice for the number of observations q. This test delivers a p-value of 0.55 for $S_n = 73$ out of $\hat{q}_{\text{irot}} = 138$ observations. The null hypothesis of continuity of the density is therefore not rejected.

7 Concluding remarks

This paper presents an approximate sign test based on g-order statistics for testing the continuity of a density at a point in RDD. We study the properties of this test under two asymptotic frameworks; one in which the number q of observations local to the cut-off is fixed as the sample size n diverges to infinity, and one in which q diverges to infinity slowly as n diverges to infinity. We show that the test has limiting rejection probability under the null hypothesis not exceeding the nominal level in both asymptotic frameworks under similar and arguably mild conditions. More importantly, our test is easy to implement, asymptotically valid under weaker conditions than those used by competing methods, exhibits finite sample validity under stronger conditions than those needed for its asymptotic validity, and delivers competitive power in simulations.

A final aspect we would like to highlight of our test is its simplicity. The test only requires to count the number of non-negative observations out of the q observations closest to the cut-off (this is all that is required to compute the p-value in (14)), and does not involve kernels, local polynomials, bias correction, or bandwidth choices. Importantly, we have developed the `rdcont` `Stata` package that allows for effortless implementation of the test we propose in this paper.

A Proof of the main results

A.1 Proof of Theorem 4.1

Throughout the proof we use $\{Z_i^* : 1 \le i \le n\}$ as defined in Lemma B.1, which in turn allow us to apply Kaufmann and Reiss (1992, Theorem 1) later in the proof, when invoking Lemma B.4.

Let $Z^*_{g,(1)}, \ldots, Z^*_{g,(q)}$ denote the q values of $\{Z_i^* : 1 \le i \le n\}$ that are induced by the q smallest values of $\{g(Z_i^*) = |Z_i^*| : 1 \le i \le n\}$ and let

$$A^*_{g,(j)} \equiv I\{Z^*_{g,(j)} \ge 0\} \text{ for } 1 \le j \le q$$

and

$$S_n^* \equiv \sum_{j=1}^{q} A^*_{g,(j)} \ . \tag{A-1}$$

It is convenient to introduce the following notation. Let $p_q(s|\pi)$ denote the probability mass function (pmf) of $\mathrm{Bi}(q,\pi)$ with $\pi \in [0,1]$, i.e.,

$$p_q(s|\pi) = \binom{q}{s} \pi^s (1-\pi)^{q-s} \ . \tag{A-2}$$

Note that $p_q(s|\pi)$ is continuous in π.

Next, consider S_n in (6). Note that S_n has support $\mathbf{N}_q \equiv \{0, 1, \ldots, q\}$ for all $n \in \mathbf{N}$, and so its CDF at any $x \in \mathbf{R}$ is $\sum_{s \in \mathbf{N}_q} P\{S_n = s\} I\{s \le x\}$. From this, we conclude that $S_n \xrightarrow{d} S \sim \mathrm{Bi}(q, \pi_f)$ follows from showing that $P\{S_n = s\} \to p_q(s|\pi_f)$ for all $s \in \mathbf{N}_q$. To this end, consider the following derivation for arbitrary $s \in N_q$ and $\epsilon < \delta/2$ with δ as in Assumption 4.1.

$$\begin{aligned}
|P\{S_n = s\} - p_q(s|\pi_f)| &\le R_{n,1} + |P\{S_n^* = s\} - p_q(s|\pi_f)| \\
&= R_{n,1} + |E[P\{S_n^* = s \mid |Z^*_{g,(q+1)}|\} - p_q(s|\pi_f)]| \\
&\le R_{n,1} + E|P\{S_n^* = s \mid |Z^*_{g,(q+1)}|\} - p_q(s|\pi_f)| \\
&= R_{n,1} + E[|P\{S_n^* = s \mid |Z^*_{g,(q+1)}|\} - p_q(s|\pi_f)| I\{|Z^*_{g,(q+1)}| \le \epsilon\}] \\
&\quad + E[|P\{S_n^* = s \mid |Z^*_{g,(q+1)}|\} - p_q(s|\pi_f)| I\{|Z^*_{g,(q+1)}| > \epsilon\}] \\
&\le R_{n,1} + R_2(\varepsilon) + R_{n,3}(\varepsilon) \ ,
\end{aligned} \tag{A-3}$$

where the first inequality follows from the triangle inequality and by setting $R_{n,1} \equiv |P\{S_n = s\} - P\{S_n^* = s\}|$, the first equality follows from the law of iterated expectations, the second inequality follows from Jensen's inequality, and the last inequality follows from Lemma B.4 and by setting $R_2(\varepsilon) \equiv \sup_{r \le \epsilon} |p_q(s|\pi(r)) - p_q(s|\pi_f)|$, and $R_{n,3}(\varepsilon) \equiv P\{|Z^*_{g,(q+1)}| > \epsilon\}$. By computing sequential limits $n \to \infty$ and $\epsilon \downarrow 0$, we now show that right hand side of (A-3) converges to zero. As $n \to \infty$, $R_{n,1} = o(1)$ by Lemma B.3(b) and $R_{n,3}(\varepsilon) = o(1)$ by Lemma B.3(a). By then taking $\epsilon \downarrow 0$, Lemma B.2 implies that $\sup_{r \le \epsilon} |\pi(r) - \pi_f| = o(1)$, and this and the continuity of $p_q(s|\pi)$ in π implies that $R_2(\varepsilon) = o(1)$. ■

A.2 Proof of Theorem 4.2

By the definition of $\phi(S_n)$ in (11) and the expressions of $T(S_n)$ in (7) and $c_q(\alpha)$ in (9),

$$E[\phi(S_n)] = P\{S_n < b_q(\alpha)\} + P\{S_n > q - b_q(\alpha)\} + a_q(\alpha)\left(P\{S_n = b_q(\alpha)\} + P\{S_n = q - b_q(\alpha)\}\right) \ .$$

Theorem 4.1 shows that $P\{S_n = s\} = P\{S = s\} + o(1)$ for all $s \in \mathbf{N}_q \equiv \{0, 1, \dots, q\}$, where $S \sim \text{Bi}(q, \pi_f)$ and π_f is as in Theorem 4.1. It follows from this result and the above display that $E[\phi(S_n)] \to E[\phi(S)]$ as $n \to \infty$, where

$$E[\phi(S)] = P\{S < b_q(\alpha)\} + P\{S > q - b_q(\alpha)\} + a_q(\alpha)\left(P\{S = b_q(\alpha)\} + P\{S = q - b_q(\alpha)\}\right) \ . \tag{A-4}$$

We complete the proof by analyzing (A-4) under H_0 and H_1 in (3).

Under H_0 in (3), $S \sim \text{Bi}(q, \frac{1}{2})$. In this case,

$$P\{S < b_q(\alpha)\} + P\{S > q - b_q(\alpha)\} = 2\Psi_q(b_q(\alpha) - 1) \ ,$$

where we used that $b_q(\alpha) \in \{0, 1, \dots, \lfloor \frac{q}{2} \rfloor\}$ and $P\{S < b\} = P\{S > q - b\}$ for any $b \in \{0, \dots, \lfloor \frac{q}{2} \rfloor\}$ when $\pi_f = \frac{1}{2}$. In addition,

$$a_q(\alpha)\left(P\{S = b_q(\alpha)\} + P\{S = q - b_q(\alpha)\}\right) = 2a_q(\alpha)\frac{1}{2^q}\binom{q}{b_q(\alpha)} \ ,$$

where we used that $\binom{q}{C} = \binom{q}{q-C}$ for any $C \in \{0, \dots, q\}$. Therefore,

$$E[\phi(S)] = 2\Psi_q(b_q(\alpha) - 1) + \frac{a_q(\alpha)}{2^{q-1}}\binom{q}{b_q(\alpha)} = \alpha \ , \tag{A-5}$$

where the last equality follows by definition of $a_q(\alpha)$.

By Lehmann and Romano (2005, Example 4.2.1) (with $p_0 = 1/2$), $\phi(\cdot)$ in (11) is an unbiased test for (3). From this and (A-5), it follows that $E[\phi(S)] \geq \alpha$ under H_1 in (3), as desired. ■

A.3 Proof of Theorem 4.3

For S_n as in (6) and S_n^* as in (A-1), let

$$\xi_q(\pi) \equiv \sqrt{q}\left(\frac{1}{q}S_n - \pi\right) \quad \text{and} \quad \xi_q^*(\pi) \equiv \sqrt{q}\left(\frac{1}{q}S_n^* - \pi\right) \ .$$

For any $\pi \in (0, 1)$ and for $S \sim \text{Bi}(q, \pi)$, let $\Psi_q(x|\pi)$ denote the CDF of S and let

$$J_q(x|\pi) \equiv P\left\{\sqrt{q}\left(\frac{1}{q}S - \pi\right) \leq x\right\} \ .$$

It suffices to show that for any $\eta > 0$, there exists N such that $\forall n \geq N$,

$$\left| P\{\xi_q(\pi_f) \leq x\} - \Phi\left(\frac{x}{\sqrt{\pi_f(1 - \pi_f)}}\right) \right| \leq \eta \ .$$

Let $\{\epsilon_q : q \geq 1\}$ be a sequence in $(0, \delta/2)$ that satisfies $\sqrt{q}\epsilon_q \to 0$ and $\epsilon_q \frac{n}{q} \to \infty$. Since $\frac{q^{3/2}}{n} \to 0$ these conditions occur for all q sufficiently large if we set $\epsilon_q = \frac{1}{q^{1/2}}\left(\log\left(\frac{n}{q^{3/2}}\right)\right)^{-1}$. Consider the following decomposition for $x \in \mathbf{R}$,

$$P\{\xi_q(\pi_f) \leq x\} = \bar{R}_{n,1} + \bar{R}_{n,2} + \bar{R}_{n,3} \ , \tag{A-6}$$

with

$$
\begin{aligned}
\bar{R}_{n,1} &\equiv P\{\xi_q(\pi_f) \le x\} - P\{\xi_q^*(\pi_f) \le x\} \\
\bar{R}_{n,2} &\equiv E[P\{\xi_q^*(\pi_f) \le x \mid |Z^*_{g,(q+1)}|\} I\{|Z^*_{g,(q+1)}| > \epsilon_q\}] \\
\bar{R}_{n,3} &\equiv E[P\{\xi_q^*(\pi_f) \le x \mid |Z^*_{g,(q+1)}|\} I\{|Z^*_{g,(q+1)}| \le \epsilon_q\}] \ .
\end{aligned}
$$

First, Lemma B.3(b) implies that $\bar{R}_{n,1} = o(1)$. Second, $\bar{R}_{n,2} = o(1)$ follows from

$$
0 \le \bar{R}_{n,2} \le P\{|Z^*_{g,(q+1)}| > \epsilon_q\} = P\Big\{\frac{n}{q}|Z^*_{g,(q+1)}| > \epsilon_q \frac{n}{q}\Big\} = o(1) \ ,
$$

where the last equality follows from Lemma B.5 and $\epsilon_q \frac{n}{q} \to \infty$. Finally, let $\pi_q^+ \equiv \pi_f + \frac{1}{2}\frac{\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)}$ and consider the following derivation,

$$
\begin{aligned}
\bar{R}_{n,3} &\ge P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} \inf_{r \le \epsilon_q} P\left\{S_n^* \le \sqrt{q}x + q\pi_f \;\Big|\; |Z^*_{g,(q+1)}| = r\right\} \\
&= P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} \inf_{r \le \epsilon_q} \Psi_q\left(\sqrt{q}x + q\pi_f \Big| \pi(r)\right) \\
&\ge P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} \Psi_q\left(\sqrt{q}x + q\pi_f \Big| \pi_q^+\right) \\
&= P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} J_q\left(x - \sqrt{q}(\pi_q^+ - \pi_f) \Big| \pi_q^+\right) \\
&= P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} J_q\left(x - \frac{1}{2}\frac{\sqrt{q}\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)} \Bigg| \pi_q^+\right) \ ,
\end{aligned} \tag{A-7}
$$

where the first equality uses $\pi(r) = P\{Z > 0 \mid |Z| < r\}$ and follows from Lemma B.4, the second inequality follows from $\Psi_q(x|\pi)$ being decreasing in π and $\pi(r) \le \pi_q^+$ for $r \le \epsilon_q$ from Lemma B.6, and the last equality follows from the definition of π_q^+. By an analogous argument,

$$
\begin{aligned}
\bar{R}_{n,3} &\le P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} \sup_{r \le \epsilon_q} P\left\{S_n^* \le \sqrt{q}x + q\pi_f \;\Big|\; |Z^*_{g,(q+1)}| = r\right\} \\
&\le P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} \Psi_q\left(\sqrt{q}x + q\pi_f \Big| \pi_q^-\right) \\
&\le P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} J_q\left(x + \frac{1}{2}\frac{\sqrt{q}\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)} \Bigg| \pi_q^-\right) \ ,
\end{aligned} \tag{A-8}
$$

where in this case we define $\pi_q^- \equiv \pi_f - \frac{1}{2}\frac{\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)}$. To complete the proof, it suffices to show that that the right-hand side expressions of (A-7) and (A-8) converge to $\Phi\left(\frac{x}{\sqrt{\pi_f(1-\pi_f)}}\right)$. We only show the result for (A-7), as the result for (A-8) is analogous.

It follows by the Berry-Esseen theorem that

$$
\left| J_q\left(x - \frac{1}{2}\frac{\sqrt{q}\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)} \Bigg| \pi_q^+\right) - \Phi\left(\frac{x - \frac{1}{2}\frac{\sqrt{q}\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)}}{\sqrt{\pi_q^+(1-\pi_q^+)}}\right)\right| \le \frac{1}{2\sqrt{q}}\frac{((\pi_q^+)^2 + (1-\pi_q^+)^2)}{\sqrt{\pi_q^+(1-\pi_q^+)}} \to 0 \ , \tag{A-9}
$$

where the convergence follows from $q \to \infty$ and $\pi_q^+ \to \pi_f \in (0,1)$. Since $\sqrt{q}\epsilon_q \to 0$, the continuity of the standard normal CDF implies that

$$
\left|\Phi\left(\frac{x - \frac{1}{2}\frac{\sqrt{q}\epsilon_q C_P}{f_Z^+(0)+f_Z^-(0)}}{\sqrt{\pi_q^+(1-\pi_q^+)}}\right) - \Phi\left(\frac{x}{\sqrt{\pi_f(1-\pi_f)}}\right)\right| \to 0 \ . \tag{A-10}
$$

Finally, Lemma B.5 and $\epsilon_q \frac{n}{q} \to \infty$ imply that

$$P\{|Z^*_{g,(q+1)}| \le \epsilon_q\} = P\Big\{\frac{n}{q}|Z^*_{g,(q+1)}| \le \frac{n}{q}\epsilon_q\Big\} \to 1 \ . \qquad \text{(A-11)}$$

By combining (A-9), (A-10), and (A-11), (18) follows.

We now conclude the proof by showing parts (a)-(c) of the theorem. By the definition of $c_q(\alpha)$ in (9) and the central limit theorem, it follows that for any $\alpha \in (0,1)$,

$$c_q(\alpha) \to \frac{1}{2} z_{\alpha/2} \qquad \text{(A-12)}$$

as $q \to \infty$, where $z_{\alpha/2}$ is the $(1-\frac{\alpha}{2})$-quantile of $N(0,1)$. Since $q \to \infty$ as $n \to \infty$, this implies that $c_q(\alpha)$ converges to the $(1-\alpha/2)$-quantile of $N(0,1/4)$ as $n \to \infty$. Next, notice that

$$\sqrt{q}\left(\frac{1}{q}S_n - \frac{1}{2}\right) \ = \ \sqrt{q}\left(\frac{1}{q}S_n - \pi_f\right) + \sqrt{q}\left(\pi_f - \frac{1}{2}\right) \ . \qquad \text{(A-13)}$$

Under H_0 in (3), $\pi_f = \frac{1}{2}$, and so (18) implies that the expression in (A-13) converges in distribution to $N(0,1/4)$ as $n \to \infty$. From here, part (a) follows. Parts (b) and (c) follow from analogous arguments based on (18), (A-13), and the convergence of $c_q(\alpha)$ to the $(1-\alpha/2)$-quantile of $N(0,1/4)$ as $n \to \infty$. ■

B Auxiliary Results

Lemma B.1. *Let $\delta > 0$ be as in Assumption 4.1 and let $\{\upsilon_i : 1 \le i \le n\}$ be an i.i.d. sample such that $\upsilon_i \sim U(-\frac{\delta}{2}, \frac{\delta}{2})$ independent of $Z^{(n)}$. Define the sequence of i.i.d. random variables $\{Z^*_i : 1 \le i \le n\}$ as*

$$Z^*_i \equiv Z_i + \upsilon_i I\{|Z_i| \ge \delta\} \ .$$

Then,

(a) The distribution function of $|Z^*|$ is continuous on $\mathbf{R}$.

(b) For any $r \in (0, \frac{\delta}{2})$,

$$P\{Z^* \ge 0 \mid |Z^*| < r\} = P\{Z \ge 0 \mid |Z| < r\} \ . \qquad \text{(B-14)}$$

(c) For any $r > 0$, $P\{|Z^*| < r\} > 0$.

Proof. To prove part (a), let $E \subset \mathbf{R}$ be a set of zero Lebesgue measure and note that

$$\begin{aligned} P\{|Z^*| \in E\} &= P\{|Z + \upsilon I\{|Z| \ge \delta\}| \in E\} \\ &= P\{|Z + \upsilon I\{|Z| \ge \delta\}| \in E \cap |Z| \ge \delta\} + P\{|Z + \upsilon I\{|Z| \ge \delta\}| \in E \cap |Z| < \delta\} \\ &= P\{|Z + \upsilon| \in E \cap |Z| \ge \delta\} + P\{|Z| \in E \cap |Z| < \delta\} \\ &\le P\{|Z + \upsilon| \in E\} + P\{|Z| \in E \cap (0,\delta)\} = 0 \ , \end{aligned}$$

where the last equality holds because the distribution function of $|Z + \upsilon|$ is continuous and $E \cap (0,\delta)$ is a subset of zero Lebesgue measure in the set where the distribution function of $|Z|$ is assumed to be continuous.

For part (b), note that for any $r \in (0, \frac{\delta}{2})$, $|Z^*| < r$ implies that $Z = Z^*$ and (B-14) follows.

For part (c), use again that $P\{|Z^*| < r\} = P\{|Z| < r\}$ whenever $r \in (0, \frac{\delta}{2})$. By Assumption 4.1, for any $0 < \epsilon < \delta$,

$$\frac{1}{\epsilon} P\{|Z| < \epsilon\} = \frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} f_Z(z) dz \ .$$

Taking limits as $\epsilon \downarrow 0$, using L'Hôpital's rule, and invoking Assumption 4.1(ii) shows that $\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} P\{|Z| < \epsilon\} = f_Z^+(0) + f_Z^-(0) > 0$. Thus, there exists $\bar{\epsilon} < \delta$ such that $P\{|Z| < \epsilon\} > 0$ for all $\epsilon \in (0, \bar{\epsilon})$ and so $P\{|Z| < r\} > 0$ for all $r \in \mathbf{R}$. This completes the proof. ■

Lemma B.2. *Let Assumptions 4.1(i) and 4.1(ii) hold and let π_f be defined as in Theorem 4.1. Then, for all $\mu > 0$, there exists $\epsilon > 0$ s.t.*

$$\sup_{r \le \epsilon} |P\{Z \ge 0 \mid |Z| < r\} - \pi_f| \le \mu \ .$$

Proof. First note that, under Assumption 4.1(ii), the proof of Lemma B.1 shows that $P\{|Z| < r\} > 0$ for all $r \in \mathbf{R}$. It follows that

$$P\{Z \ge 0 \mid |Z| < \epsilon\} \ = \ \frac{\frac{1}{\epsilon}\int_0^{\epsilon} f_Z(z)dz}{\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} f_Z(z)dz} \ = \ \frac{f_Z^+(0)}{f_Z^+(0) + f_Z^-(0)} + \Delta_\epsilon \ ,$$

where the last equality holds for $\Delta_\epsilon \to 0$ as $\epsilon \to 0$ by using L'Hôpital's rule and Assumption 4.1(ii). The result then follows by definition of π_f. ■

Lemma B.3. *Let Assumptions 4.1(i) and 4.1(ii) hold and $\frac{q}{n} \to 0$ as $n \to \infty$. Then,*

(a) For any $\epsilon \in (0, \frac{\delta}{2})$, $P\{\liminf_{n\to\infty}\{|Z^*_{g,(q+1)}| \le \epsilon\}\} = P\{\liminf_{n\to\infty}\{|Z_{g,(q+1)}| \le \epsilon\}\} = 1$.

(b) $P\{\liminf_{n\to\infty}\{S_n = S_n^*\}\} = 1$, where S_n is as in (6) and S_n^* is as in (A-1).

Proof. Fix $\epsilon \in (0, \frac{\delta}{2})$ arbitrarily and set $N_n \equiv \sum_{i=1}^n I\{|Z_i| \le \epsilon\}$. Note that $N_n \ge q+1$ implies that $Z_i^* = Z_i$ and $Z^*_{g,(j)} = Z_{g,(j)}$ for at least $q+1$ observations that are within an ϵ-neighborhood of zero. It follows that for all these observations, $A_{g,(j)} = A^*_{g,(j)}$, $Z^*_{g,(j)} \le \epsilon$, and $Z_{g,(j)} \le \epsilon$. We conclude that $N_n \ge q+1$ implies

$$S_n = S_n^*, \quad Z^*_{g,(q+1)} \le \epsilon, \quad \text{and} \quad Z_{g,(q+1)} \le \epsilon \ .$$

Parts (a)-(b) thus follow from proving that $P\{\liminf_{n\to\infty}\{N_n \ge q+1\}\} = 1$. To show this, note that $N_n \sim \mathrm{Bi}(n, P\{|Z| \le \epsilon\})$. Now set $\mu \equiv \frac{1}{2}P\{|Z| \le \epsilon\}$, which is positive by the proof of Lemma B.1. It follows that

$$P\{\liminf_{n\to\infty}\{N_n \ge q+1\}\} = P\left\{\liminf_{n\to\infty}\left\{\frac{N_n}{n} \ge \frac{q+1}{n}\right\}\right\} \ge P\left\{\liminf_{n\to\infty}\left\{\frac{N_n}{n} \ge \mu\right\}\right\} = 1 \ ,$$

where the inequality holds for all $n > (q+1)/\mu$, and the last equality follows by the strong law of large numbers, i.e., $N_n/n \overset{a.s.}{\to} 2\mu > 0$. ■

Lemma B.4. *Let Assumptions 4.1(i) and 4.1(ii) hold. Fix $r \in (0, \frac{\delta}{2})$ and $q \in \{1, \ldots, n-1\}$ arbitrarily. Then, for all $s \in \mathbf{N}_q \equiv \{0, 1, \ldots, q\}$,*

$$P\{S_n^* = s \mid |Z^*_{g,(q+1)}| = r\} = p_q(s|\pi(r))$$

where $p_q(s|\pi(r))$ is the pmf defined in (A-2) with $\pi(r) \equiv P\{Z \ge 0 \mid |Z| < r\}$.

Proof. Let $X \equiv (|Z^*|, A^*)$ with $A^* = I\{Z^* \geq 0\}$ and note that the g-order statistics we defined in Section 3 using $g = |\cdot|$, could be alternatively obtained using X and $\tilde{g}$-order statistics where $\tilde{g}$ is now the projection into the first component of X, i.e.

$$\tilde{g}(X) = |Z^*| \ .$$

In this way, and for this particular choice of $\tilde{g}$, $\tilde{g}$-order statistics on X deliver

$$X_{\tilde{g},(1)} \equiv (|Z^*|_{(1)}, A^*_{[1]}) \leq_{\tilde{g}} (|Z^*|_{(2)}, A^*_{[2]}) \leq_{\tilde{g}} \cdots \leq_{\tilde{g}} (|Z^*|_{(n)}, A^*_{[n]}) \equiv X_{\tilde{g},(n)} \ ,$$

where the random variables $(A^*_{[1]}, \ldots, A^*_{[n]})$ are called *induced order statistics* or *concomitants* of order statistics, see David and Galambos (1974); Bhattacharya (1974).

Let $\tilde{X}_1, \ldots, \tilde{X}_q$ be i.i.d. bivariate random variables such that $\tilde{X} \stackrel{d}{=} \{X \mid \tilde{g}(X) < r\}$. Theorem 1 in Kaufmann and Reiss (1992) states that

$$\{(X_{\tilde{g},(1)}, \ldots, X_{\tilde{g},(q)}) \mid \tilde{g}(X_{\tilde{g},(q+1)}) = r\} \stackrel{d}{=} \{\tilde{X}_{\tilde{g},(1)}, \ldots, \tilde{X}_{\tilde{g},(q)}\} \ , \tag{B-15}$$

with $\tilde{X}_{\tilde{g},(1)}, \ldots, \tilde{X}_{\tilde{g},(q)}$ being the $\tilde{g}$-order statistics of $\tilde{X}_1, \ldots, \tilde{X}_q$, provided that (i) $\tilde{g}(X)$ has a continuous distribution and (ii) $P\{\tilde{g}(X) < r\} > 0$. Since $\tilde{g}(X) = |Z^*|$ has a continuous distribution by Lemma B.1(a) and $P\{\tilde{g}(X) < r\} = P\{|Z^*| < r\} > 0$ by Lemma B.1(c), we use (B-15) to prove our result.

Next, note that we can re-write S^*_n in (A-1) as a function of $(X_{\tilde{g},(1)}, \ldots, X_{\tilde{g},(q)})$ by using the function h that projects into the second component of X, i.e.

$$S^*_n = \sum_{j=1}^{q} A^*_{g,(j)} = \sum_{j=1}^{q} A^*_{[j]} = \sum_{j=1}^{q} h(X_{\tilde{g},(j)}) \ ,$$

where in the second equality we used that $A^*_{g,(j)} = A^*_{[j]}$ by definition. Using this characterization, it follows that

$$\begin{aligned}
P\{S^*_n = s \mid |Z^*_{g,(q+1)}| = r\} &= P\Big\{\sum_{j=1}^{q} h(X_{\tilde{g},(j)}) = s \mid |\tilde{g}(X_{\tilde{g},(q+1)})| = r\Big\} \\
&= P\Big\{\sum_{j=1}^{q} h(\tilde{X}_{\tilde{g},(j)}) = s\Big\} \\
&= P\Big\{\sum_{j=1}^{q} h(\tilde{X}_j) = s\Big\} \\
&= p_q(s|\pi(r)) \ ,
\end{aligned}$$

where the second equality follows from (B-15), the third equality follows from $\sum_{j=1}^{q} h(\tilde{X}_{\tilde{g},(j)}) = \sum_{j=1}^{q} h(\tilde{X}_j)$, and the last equality follows from $h(\tilde{X}_1), \ldots, h(\tilde{X}_q)$ being i.i.d. bivariate random variables such that $h(\tilde{X}) \stackrel{d}{=} \{h(X) \mid \tilde{g}(X) < r\}$ and $\{h(X) \mid \tilde{g}(X) < r\} = \{I\{Z^* \geq 0\} \mid |Z^*| < r\}$ being distributed Bernoulli with parameter $\pi(r) = P\{Z^* \geq 0 \mid |Z^*| < r\}$. Since $P\{Z^* \geq 0 \mid |Z^*| < r\} = P\{Z \geq 0 \mid |Z| < r\}$ for $r \in (0, \frac{\delta}{2})$ by Lemma B.1(b), this completes the proof. ∎

Lemma B.5. *Let Assumptions 4.1(i') and 4.1(ii) hold and suppose $q \to \infty$ and $\frac{q}{n} \to 0$ as $n \to \infty$. Then*

$$\frac{n}{q}|Z^*_{g,(q)}| \ \stackrel{P}{\to} \ \frac{1}{f^+_Z(0) + f^-_Z(0)} \ .$$

Proof. For any $\epsilon > 0$, it suffices to show that

$$P\Big\{\frac{n}{q}|Z^*_{g,(q)}| > \frac{1}{f_Z^+(0)+f_Z^-(0)} + \epsilon\Big\} \to 0 \quad \text{and} \quad P\Big\{\frac{n}{q}|Z^*_{g,(q)}| < \frac{1}{f_Z^+(0)+f_Z^-(0)} - \epsilon\Big\} \to 0 \ . \tag{B-16}$$

We only show the first result in (B-16), as the second one follows from symmetric arguments. By definition, $|Z^*_{g,(q)}| = |Z^*|_{(q)}$ where $|Z^*|_{(q)}$ denotes the q-th order statistic of the absolute value of Z^*. Denote by Q the CDF of $|Z^*|$ and by $U_{(q)}$ the q-th order statistic of a $U[0,1]$ distributed random variable. Lemma B.1(a) implies that $Q(\cdot)$ is a continuous CDF. Then, for $\bar{M} = \frac{1}{f_Z^+(0)+f_Z^-(0)} + \epsilon$, note that

$$\begin{aligned} P\Big\{\frac{n}{q}|Z^*_{g,(q)}| > \bar{M}\Big\} &= P\Big\{|Z^*|_{(q)} > \frac{q}{n}\bar{M}\Big\} \\ &= P\Big\{U_{(q)} > Q\left(\frac{q}{n}\bar{M}\right)\Big\} \\ &= P\Big\{\frac{n^{1/2}}{e_n}(U_{(q)} - \mu_n) > \frac{n^{1/2}}{e_n}\left(Q\left(\frac{q}{n}\bar{M}\right) - \mu_n\right)\Big\} \ , \end{aligned} \tag{B-17}$$

where $e_n^2 = \mu_n(1-\mu_n)$ and $\mu_n = q/(n+1)$. Letting

$$\gamma_n = e_n n^{1/2}\frac{1}{e_n^2}\left(Q\left(\frac{q}{n}\bar{M}\right) - \mu_n\right) \ ,$$

it follows from (B-17) and Reiss (1989, Eq. (3.1.2) in Lemma 3.1.1) that

$$P\Big\{\frac{n}{q}|Z^*_{g,(q)}| > \bar{M}\Big\} \le \exp\left(-\frac{\gamma_n^2}{3(1+\gamma_n/(e_n n^{1/2}))}\right) \ .$$

To complete the proof it suffices to show that the right-hand side expression in the display above converges to zero. To this end, it suffices to show that $\gamma_n \to \infty$ and $\gamma_n/(e_n n^{1/2})$ converges to a positive constant. In turn, this follows from showing that

$$e_n n^{1/2} \to \infty \quad \text{and} \quad \frac{1}{e_n^2}\left(Q\left(\frac{q}{n}\bar{M}\right) - \mu_n\right) \to \epsilon\left(f_Z^+(0)+f_Z^-(0)\right) > 0 \ , \tag{B-18}$$

where the limit of the second expression is positive by Assumption 4.1(ii). To show the first result in (B-18), note that $\frac{q}{n} \to 0$ implies $\frac{(e_n n^{1/2})^2}{q} = \frac{\mu_n(1-\mu_n)}{q/n} = \frac{n}{n+1}(1-\frac{q}{n+1}) \to 1$. Combined with $q \to \infty$, this then implies that $e_n n^{1/2} \to \infty$. To show the second result in (B-18), note that $\frac{q}{n} \to 0$ implies that for n sufficiently large we obtain $\frac{q}{n}\bar{M} < \delta$ and so $Q\left(\frac{q}{n}\bar{M}\right) = \int_{-\frac{q}{n}\bar{M}}^{\frac{q}{n}\bar{M}} f_Z(z)dz$ since $Z^* = Z$ on $(-\delta, \delta)$. Then,

$$\begin{aligned} \frac{1}{e_n^2}\left(Q\left(\frac{q}{n}\bar{M}\right) - \mu_n\right) &= \frac{q/n}{\mu_n(1-\mu_n)}\left(\frac{n}{q}\int_{-\frac{q}{n}\bar{M}}^{\frac{q}{n}\bar{M}} f_Z(z)dz - \frac{n}{q}\mu_n\right) \\ &= \frac{q/n}{\mu_n(1-\mu_n)}\left(\frac{n}{q}\int_0^{\frac{q}{n}\bar{M}}(f_Z(z) - f_Z^+(0))dz + \frac{n}{q}\int_{-\frac{q}{n}\bar{M}}^0 (f_Z(z) - f_Z^-(0))dz\right) \\ &\quad + \frac{q/n}{\mu_n(1-\mu_n)}\left(\frac{n}{q}\int_0^{\frac{q}{n}\bar{M}} f_Z^+(0)dz + \frac{n}{q}\int_{-\frac{q}{n}\bar{M}}^0 f_Z^-(0)dz - \frac{n}{q}\mu_n\right) \\ &\to \bar{M}\left(f_Z^+(0)+f_Z^-(0)\right) - 1 = \epsilon\left(f_Z^+(0)+f_Z^-(0)\right) \ , \end{aligned}$$

where the convergence follows from Assumption 4.1(i') and $\frac{q}{n} \to 0$, which imply that

$$\Big|\frac{n}{q}\int_0^{\frac{q}{n}\bar{M}}(f_Z(z) - f_Z^+(0))dz\Big| \le C_P\bar{M}^2\frac{q}{n} \to 0 \quad \text{and} \quad \Big|\frac{n}{q}\int_{-\frac{q}{n}\bar{M}}^0 (f_Z(z) - f_Z^-(0))dz\Big| \le C_P\bar{M}^2\frac{q}{n} \to 0 \ ,$$

and $\frac{q/n}{\mu_n(1-\mu_n)} \to 1$. ■

Lemma B.6. *Let Assumption 4.1(i') hold, π_f be as in Theorem 4.1, and $\pi(r) = P\{Z > 0 \mid |Z| < r\}$. Then, for any $r \in (0, \delta)$,*

$$|\pi(r) - \pi_f| \leq \frac{r}{2}\frac{C_P}{f_Z^+(0) + f_Z^-(0)} \ .$$

Proof. Fix $r \in (0, \delta)$ arbitrarily. Start by re-writing $\pi(r)$ as follows,

$$\begin{aligned}\pi(r) &= \frac{\frac{1}{r}\int_0^r f_Z(z)\,dz}{\frac{1}{r}\int_{-r}^0 f_Z(z)dz + \frac{1}{r}\int_0^r f_Z(z)dz} = \left(\frac{\frac{1}{r}\int_{-r}^0 f_Z(z)dz}{\frac{1}{r}\int_0^r f_Z(z)dz} + 1\right)^{-1} \\ &= \left(\frac{f_Z^-(0) + \frac{1}{r}\int_{-r}^0 (f_Z(z) - f_Z^-(0))dz}{f_Z^+(0) + \frac{1}{r}\int_0^r (f_Z(z) - f_Z^+(0))dz} + 1\right)^{-1} \ . \qquad \text{(B-19)}\end{aligned}$$

Next, note that Assumption 4.1(i') implies that

$$\left|\frac{1}{r}\int_0^r (f_Z(z) - f_Z^+(0))dz\right| \leq \frac{r}{2}C_P \quad \text{and} \quad \left|\frac{1}{r}\int_-^0 (f_Z(z) - f_Z^-(0))dz\right| \leq \frac{r}{2}C_P \ .$$

Combining these two derivations we conclude that

$$\begin{aligned}\pi(r) &\leq \left(\frac{f_Z^-(0) - \frac{r}{2}C_P}{f_Z^+(0) + \frac{r}{2}C_P} + 1\right)^{-1} = \pi_f + \frac{\frac{r}{2}C_P}{f_Z^+(0) + f_Z^-(0)} \ , \\ \pi(r) &\geq \left(\frac{f_Z^-(0) + \frac{r}{2}C_P}{f_Z^+(0)\frac{r}{2}rC_P} + 1\right)^{-1} = \pi_f - \frac{\frac{r}{2}C_P}{f_Z^+(0) + f_Z^-(0)} \ .\end{aligned}$$

This implies the desired result. ■

C Results under a mass point at the cut-off

In this section, we consider the asymptotic behavior of the proposed test when there is a mass point at the cut-off $\bar{z} = 0$. As mentioned in Section 2, this mass point implies a violation of Assumption 4.1, and so our formal results do not apply. On the other hand, a mass point at the cutoff is usually considered an extreme form of violation of the continuity of the density at the cut-off and, thus, should be regarded as part of H_1 in (3). The following result shows that the proposed test rejects with probability approaching one whenever there is a mass point zero.

Theorem C.1. *Assume that $P\{Z = 0\} > 0$ and let $\alpha \in (0, 1)$. If $q \geq q^*(\alpha)$ as in (13) and $\frac{q}{n} \to 0$,*

(a) $S_n = q$ with probability approaching one.

(b) $\lim_{n\to\infty} E[\phi(S_n)] = 1$.

Proof. Let $N_n \equiv \sum_{i=1}^n I\{Z_i = 0\}$. Note that $N_n \geq q + 1$ implies that $S_n = q$ so $T_q(S_n) = \sqrt{q}/2$. By this and $q \geq q^*(\alpha)$, $c_q(\alpha) < \sqrt{q}/2 = T_q(S_n)$ and so $\phi(S_n) = 1$. Therefore, the desired results follow from showing that $P\{\liminf_{n\to\infty}\{N_n \geq q + 1\}\} = 1$. To this end, note that $N_n \sim \text{Bi}(n, P\{Z = 0\})$ so

$$P\left\{\liminf_{n\to\infty}\{N_n \geq q+1\}\right\} = P\left\{\liminf_{n\to\infty}\left\{\frac{N_n}{n} \geq \frac{q+1}{n}\right\}\right\} \geq P\left\{\liminf_{n\to\infty}\left\{\frac{N_n}{n} \geq \frac{P\{Z=0\}}{2}\right\}\right\} = 1 \ ,$$

where the inequality holds for all n large enough such that $P\{Z = 0\}/2 > (q + 1)/n \to 0$, and the last equality follows by the strong law of large numbers, i.e., $N_n/n \overset{a.s.}{\to} P\{Z = 0\} > 0$. ■

It is relevant to note that Theorem C.1 applies to both asymptotic frameworks considered in the paper, i.e., it applies to fixed q case as long as $q \geq q^*(\alpha)$, and it applies to large q case provided that $q/n \to 0$.

D Computational details on the data-dependent rule for q

In the simulations of Section 5 and in the companion `Stata` package, the feasible informed rule of thumb is computed as follows. First, we compute

$$\hat{q}_{\text{rot}} = \left\lceil \max\left\{ q^*(\alpha), n^{1/2}\left(\hat{\sigma}\frac{4\phi^2_{\hat{\mu},\hat{\sigma}}(\bar{z})}{\phi_{\hat{\mu},\hat{\sigma}}(\hat{\mu}+\hat{\sigma})}\right)^{2/3}\right\}\right\rceil ,$$

where $q^*(\alpha) = 1 - \frac{\log \alpha}{\log 2}$, $\hat{\mu}$ is the sample mean of $\{Z_1, \ldots, Z_n\}$, $\hat{\sigma}^2$ is the sample variance of $\{Z_1, \ldots, Z_n\}$, $\bar{z}$ is the cut-off point, and n is the sample size. In principle, the value $\hat{q}_{\text{rot}}$ could be used to implement our test. However, this would ignore the non-monotonicity of the limiting null rejection probability of the non-randomized version of our test, which according to Theorem 4.2, equals $2\Psi_q(b_q(\alpha)-1)$ with $b_q(\alpha)$ defined in (10). Figure 3 displays $2\Psi_q(b_q(\alpha)-1)$ for $\alpha = 5\%$ as a function of q. The figure shows that $2\Psi_q(b_q(\alpha)-1)$ takes values very close to α for q as low as 17 (i.e., 4.9%), but could be far from α for $q = 19$ (i.e., 1.9%). We therefore propose an additional layer in the data-dependent way of choosing q that guarantees that such a value delivers a local "peak" of $2\Psi_q(b_q(\alpha)-1)$ in Figure 3.

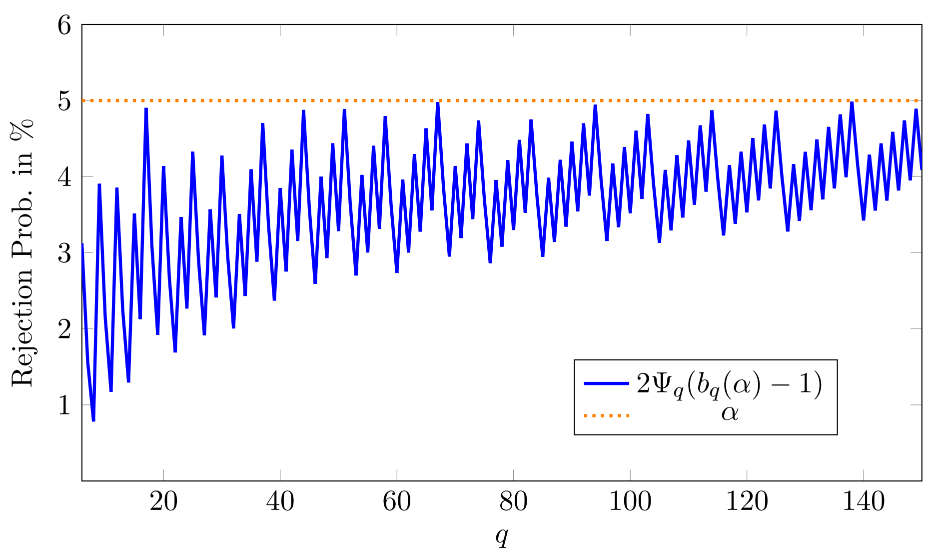


Figure 3: The solid line is the limiting null rejection probability (in %) of the non-randomized version of the test, $2\Psi_q(b_q(\alpha)-1)$, as a function of q. The dotted line is the nominal level of the test.

To be concrete, we define $\hat{q}_{\text{irot}}$ as

$$\hat{q}_{\text{irot}} = \underset{q\in\mathcal{N}(\hat{q}_{\text{rot}})}{\operatorname{argmax}}\ \Psi_q(b_q(\alpha)-1) , \tag{D-20}$$

where

$$\mathcal{N}(\hat{q}_{\text{rot}}) \equiv \{q \in \mathbf{N} : \max\{q^*(\alpha), \hat{q}_{\text{rot}} - \lceil 4\log(\hat{q}_{\text{rot}})\rceil\} \leq q \leq \hat{q}_{\text{rot}} + \lceil 4\log(\hat{q}_{\text{rot}})\rceil\} . \tag{D-21}$$

The value of window size $\lceil 4\log(\hat{q}_{\text{rot}})\rceil$ is the minimum number of points that are required to reach a local peak of $2\Psi_q(b_q(\alpha)-1)$ for values of $\alpha \in \{1\%, 5\%, 10\%\}$ and is such that, for large values of $\hat{q}_{\text{rot}}$, the window gets larger to improve the chances of getting one of the peaks closer to α as $\hat{q}_{\text{rot}}$ increases. A smaller window size may not guarantee one actually reaches a local peak. The value $\hat{q}_{\text{irot}}$ defined in (D-20) is the one we use in the simulations of Section 5 and the default value in the companion `Rdcont` `Stata` package.

E Surveyed papers on RDD

Table 4 displays the list of papers we surveyed in leading journals that use regression discontinuity designs. For a description on the criteria used to compile the list of papers in Table 4, see Canay and Kamat (2018, Appendix E).

Authors (Year)	Journal	(i): Mean Test	(ii): Density Test	Authors (Year)	Journal	(i): Mean Test	(ii): Density Test
Schmieder et al. (2016)	AER	✓	✓	Miller et al. (2013)	AEJ:AppEcon	✓	✓
Feldman et al. (2016)	AER	✓	✓	Litschig and Morrison (2013)	AEJ:AppEcon	✓	✓
Jayaraman et al. (2016)	AER	×	×	Dobbie and Skiba (2013)	AEJ:AppEcon	✓	✓
Dell (2015)	AER	✓	✓	Kazianga et al. (2013)	AEJ:AppEcon	✓	✓
Hansen (2015)	AER	✓	✓	Magruder (2012)	AEJ:AppEcon	×	×
Anderson (2014)	AER	×	×	Dustmann and Schnberg (2012)	AEJ:AppEcon	×	×
Martin et al. (2014)	AER	×	×	Clots-Figueras (2012)	AEJ:AppEcon	✓	✓
Dahl et al. (2014)	AER	✓	✓	Manacorda et al. (2011)	AEJ:AppEcon	✓	✓
Shigeoka (2014)	AER	✓	×	Chetty et al. (2014)	QJE	✓	✓
Crost et al. (2014)	AER	✓	×	Michalopoulos and Papaioannou (2014)	QJE	✓	×
Kostol and Mogstad. (2014)	AER	✓	✓	Fredriksson et al. (2013)	QJE	✓	✓
Clark and Royer (2013)	AER	✓	×	Schmieder et al. (2012)	QJE	✓	✓
Brollo et al. (2013)	AER	✓	✓	Lee and Mas (2012)	QJE	×	×
Bharadwaj et al. (2013)	AER	✓	✓	Saez et al. (2012)	QJE	×	×
Pop-Eleches and Urquiola (2013)	AER	✓	✓	Barreca et al. (2011)	QJE	×	×
Lacetera et al. (2012)	AER	×	✓	Almond et al. (2011)	QJE	✓	✓
Duflo et al. (2012)	AER	×	×	Malamud and Pop-Eleches (2011)	QJE	✓	✓
Gopinath et al. (2011)	AER	✓	✓	Fulford (2015)	ReStat	✓	×
Auffhammer and Kellogg (2011)	AER	×	×	Snider and Williams (2015)	ReStat	×	×
Duflo et al. (2011)	AER	×	×	Doleac and Sanders (2015)	ReStat	×	×
Ferraz and Finan (2011)	AER	×	×	Coşar et al. (2015)	ReStat	×	×
McCrary and Royer (2011)	AER	✓	×	Avery and Brevoort (2015)	ReStat	×	×
Beland (2015)	AEJ:AppEcon	✓	✓	Carpenter and Dobkin (2015)	ReStat	✓	×
Buser (2015)	AEJ:AppEcon	✓	✓	Black et al. (2014)	ReStat	✓	✓
Fack and Grenet (2015)	AEJ:AppEcon	✓	✓	Anderson et al. (2014)	ReStat	×	×
Cohodes and Goodman (2014)	AEJ:AppEcon	✓	✓	Alix-Garcia et al. (2013)	ReStat	✓	×
Haggag and Paci (2014)	AEJ:AppEcon	✓	✓	Albouy (2013)	ReStat	×	×
Dobbie and Fryer (2014)	AEJ:AppEcon	✓	✓	Garibaldi et al. (2012)	ReStat	✓	✓
Sekhri (2014)	AEJ:AppEcon	✓	✓	Manacorda (2012)	ReStat	✓	✓
Schumann (2014)	AEJ:AppEcon	✓	✓	Martorell and McFarlin (2011)	ReStat	✓	✓
Lucas and Mbiti (2014)	AEJ:AppEcon	✓	✓	Grosjean and Senik (2011)	ReStat	×	×

Table 4: Survey of RDD empirical papers from 2011–2015 in the following journals: American Economic Review (AER), American Economic Journal: Applied Economics (AEJ:AppEcon), Quarterly Journal of Economics (QJE), and Review of Economics and Statistics (ReStat). Implications (i) and (ii) denote the testable implications proposed by Lee (2008) described in page 1. A checkmark indicates that the corresponding implication has been tested and a cross indicates otherwise.

References

Armstrong, T. B. and Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, **86** 655–683.

Armstrong, T. B. and Kolesár, M. (2019). Simple and Honest Confidence Intervals in Nonparametric Regression. *Quantitative Economics,* forthcoming.

Bertanha, M. and Moreira, M. J. (2019). Impossible inference in econometrics: theory and applications. *Journal of Econometrics,* forthcoming.

Bhattacharya, P. (1974). Convergence of sample paths of normalized sums of induced order statistics. *The Annals of Statistics* 1034–1039.

Canay, I. A. and Kamat, V. (2018). Approximate permutation tests and induced order statistics in the regression discontinuity design. *The Review of Economic Studies*, **85** 1577–1608.

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica*, **85** 1013–1030.

Cattaneo, M. D., Jansson, M. and Ma, X. (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, **0** 1–7. Doi:10.1080/01621459.2019.1635480.

Cattaneo, M. D., Titiunik, R. and Vazquez-Bare, G. (2016). Inference in regression discontinuity designs under local randomization. *Stata Journal*, **16** 331–367.

Cattaneo, M. D., Titiunik, R. and Vazquez-Bare, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of Head Start on child mortality. *Journal of Policy Analysis and Management*, **36** 643–681.

David, H. and Galambos, J. (1974). The asymptotic theory of concomitants of order statistics. *Journal of Applied Probability* 762–770.

Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. *Advances in Econometrics*, **38** 281–315.

Gerard, F., Rokkanen, M. and Rothe, C. (2016). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable, with an application to unemployment insurance in Brazil. Working paper.

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

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142** 615–635.

Kamat, V. (2017). On nonparametric inference in the regression discontinuity design. *Econometric Theory*, **34** 694–703.

Kaufmann, E. and Reiss, R.-D. (1992). On conditional distributions of nearest neighbors. *Journal of Multivariate Analysis*, **42** 67–76.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics*, **142** 675 – 697. The regression discontinuity design: Theory and applications.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, **48** 281–355.

Lehmann, E. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.

Low, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics*, **25** 2547–2554.

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

Otsu, T., Xu, K.-L. and Matsushita, Y. (2013). Estimation and inference of discontinuity in density. *Journal of Business & Economic Statistics*, **31** 507–524.

Reiss, R.-D. (1989). *Approximate distributions of order statistics: with applications to nonparametric statistics*. Springer-Verlag, New York.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, **2** 180–212.

Shen, S. and Zhang, X. (2016). Distributional tests for regression discontinuity: Theory and empirical examples. *Review of Economics and Statistics*, **98** 685–700.