cemmap

centre for microdata methods and practice

Approximate permutation tests and induced order statistics in the regression discontinuity design

Ivan A. Canay
Vishal Kamat

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP27/15

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design *

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

Vishal Kamat
Departments of Economics
Northwestern University
v.kamat@u.northwestern.edu

June 18, 2015

Abstract

This paper proposes an asymptotically valid permutation test for a testable implication of the identification assumption in the regression discontinuity design (RDD). Here, by testable implication, we mean the requirement that the distribution of observed baseline covariates should not change discontinuously at the threshold of the so-called running variable. This contrasts to the common practice of testing the weaker implication of continuity of the means of the covariates at the threshold. When testing our null hypothesis using observations that are "close" to the threshold, the standard requirement for the finite sample validity of a permutation does not necessarily hold. We therefore propose an asymptotic framework where there is a fixed number of closest observations to the threshold with the sample size going to infinity, and propose a permutation test based on the so-called induced order statistics that controls the limiting rejection probability under the null hypothesis. In a simulation study, we find that the new test controls size remarkably well in most designs. Finally, we use our test to evaluate the validity of the design in Lee (2008), a well-known application of the RDD to study incumbency advantage.

KEYWORDS: Regression discontinuity design, permutation tests, randomization tests, induced ordered statistics, rank tests.

JEL classification codes: C12, C14.

*We would like to thank Azeem Shaikh for helpful comments on this paper. The research of the first author was supported by National Science Foundation Grants SES-1530534. This research was supported in part through the computational resources and staff contributions provided for the Social Sciences Computing cluster (SSCC) at Northwestern University.

1 Introduction

The regression discontinuity design (RDD) has become widespread in recent years to retrieve causal treatment effects; see Lee and Lemieux (2010) and Imbens and Lemieux (2008) for exhaustive surveys. A unique element of the design is the discontinuous treatment rule. This rule allocates a binary treatment when an observed predetermined variable, termed the running variable, crosses a known threshold. The design allows the researcher to identify causal effects at the threshold under a weak continuity assumption; shown by Hahn et al. (2001). Moreover, the RDD is commonly compared to a local randomized experiment, which further strengthens its internal validity; see Lee (2008). This credible identification strategy along with the abundance of discontinuous rules in practice is probably the main reason that has made the RDD increasingly popular in empirical applications - see, for recent examples, Zimmerman (2014) for test score discontinuities, Card et al. (2008) for time discontinuities, and Dell (2010) for geography discontinuities.

The identification assumption in the RDD states that individuals have imprecise control over the running variable, which translates into the distribution of observed baseline covariates and unobserved individual characteristics to be continuous in the running variable. However, due to the public knowledge of the threshold, agents may have the ability to sort precisely around the threshold in many applications, and this in turn invalidates the basic requirement for the RDD design to deliver valid estimates (see Urquiola and Verhoogen, 2009, for an explicit example using class size constraints as a discontinuity). It is thus important to try and verify the validity of the assumption to ensure credibility of the estimates. Yet, our ability to perform a test is fundamentally limited as the identification condition involves a statement about the distribution of observed and unobserved characteristics. A natural alternative in such context is to focus on a testable implication of the identification condition, which is simply a statement on the distribution of observed covariates. More specifically, in the absence of sorting we would expect the distribution of the covariates to vary smoothly at the threshold; see Lee (2008) and Section 2 for a formalization. It is then potentially possible to construct a test for this hypothesis, and a rejection would therefore indicate a failure of the identification assumption. As a matter of fact, a weaker implication of continuity of means at the threshold is instead commonly tested in empirical work. Another popular test that is indicative of sorting is the one proposed by McCrary (2008). This test uses the density of the running variable to examine if there is a disproportionate mass of individuals on one side of the threshold, which represents an alternative implication of the identification assumption.

In this paper we propose an approximate permutation test for a testable implication of the identification assumption in the RDD. We test the hypothesis of continuity of the distribution of the covariates at the threshold by using an approximate permutation test based on induced order statistics. This test provides certain benefits, over current tests, due to the novel asymptotic framework we propose for its validity. First, our permutation test controls the asymptotic null rejection probability under fairly mild conditions, and delivers finite sample validity under stronger, but yet plausible, conditions. Second, our test is more powerful against some alternatives than tests based on the difference in means at either side of the threshold, which appears to be common practice in applied work. Third, our test is arguably simple to implement as it only involves computing empirical cdf's with a fixed number of observations closest to the threshold, a tuning parameter in our framework. This contrasts with existing alternatives that require local linear estimation and delicate bandwidth choices.

The framework for our test is based on the simple intuition that observations close to the threshold are *approximately* randomly assigned to either side of it when the null hypothesis holds. This allows us to permute these observations to conduct an approximately valid test. By approximate test, we mean that the justification for its validity is asymptotic in nature and exploits the fact that the invariance inherently required for permutation tests may not hold in finite samples but it does hold in the limit experiment, as the sample size grows. To be specific, the asymptotic framework we use consists of a fixed number of closest observations to the threshold with the sample size going to infinity. These asymptotics are intended to capture a situation where the number of effective observations close to the threshold is small relative to the sample size. Formally, we exploit the framework, with novel additions, from Canay, Romano and Shaikh (2014), which first developed the insight of approximating randomization tests in this manner. Further, in an important intermediate stage, we use induced order statistics, c.f. Bhattacharya (1974), to frame our problem and develop some insightful results of independent interest.

Despite the many differences, this paper is fundamentally related to the numerous testing procedures present in the RDD. Conceptually, the paper is most closely related to Shen and Zhang (2014), which tests the hypothesis of equal outcome distributions at the threshold. This test, similar to common practice, uses an asymptotic framework that depends on a bandwidth parameter that vanishes with sample size. In contrast to ours, such a framework requires the number of effective observations used in estimation to become large as the sample size grows so that critical values of some asymptotic distributional approximation can be used. For further examples using this asymptotic approach see Calonico et al. (2014) for robust confidence interval, Imbens and Kalyanaraman (2012) for optimal bandwidth choice, and Frandsen et al. (2012) for quantiles. In addition, our framework is considerably different from the traditional finite sample one pursued in Cattaneo et al. (2015), which performs randomization inference in the RDD.

The structure of the paper is organized as follows. Section 2 introduces the notation and the hypotheses of interest. Section 3 describes our permutation test based on a fixed number of closest observations and highlights how it does not directly map into the framework of Canay et al. (2014). Section 4 presents the formal properties of the new test. To study the finite sample properties our test, we perform simulations in Section 5. In particular, we compare our test to the one proposed by Shen and Zhang (2014). In Section 6, we implement our test to reevaluate the validity of the design in Lee (2008), a familiar application of the RDD to study incumbency advantage.

2 Setup and Notation

Assume there are random variables $(\{Y(a)\}_{a\in\mathcal{A}}, W, Z) \sim Q \in \mathbf{Q}$ and let $\{(\{Y_i(a)\}_{a\in\mathcal{A}}, W_i, Z_i) : 1 \le i \le n\}$ be a random sample from Q. The response function $Y(\cdot)$ is a mapping from treatment assignments taking values in $\mathcal{A} = \{0, 1\}$ to outcomes in $\mathbf{R}$. We refer to the vector of observed characteristics $(W, Z) \in \mathcal{W} \times \mathcal{Z}$, as covariates or pretreatment variables. The unobserved heterogeneity is denoted by U. The observed outcome Y from the experiment is given by

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i) \ , \tag{1}$$

where the assignment of the treatment is deterministic and specifically follows a discontinuous rule,

$$A_i = I\{Z_i \ge 0\} \ .$$

The discontinuous assignment rule above allows us to identify the average treatment effect at the threshold by basically comparing marginal individuals *just* above and below it. Formally, Hahn et al. (2001) established that continuity of $E[Y(1)|Z]$ and $E[Y(0)|Z]$ at the threshold is sufficient to identify the average treatment at the threshold; a similar reasoning also holds for other parameters of interest at the threshold.

Despite the continuity assumption appearing weak, Lee (2008) states two practical limitations. First, it is difficult to determine whether the assumption is plausible as it is not a description of a treatment-assigning process. Second, the assumption is fundamentally untestable. Motivated by these limitations, Lee (2008, Proposition 2) shows that an alternative identification condition, which requires that the cdf of Z conditional on (W, U) is continuously differentiable in Z at $Z = 0$, translates into the following simple condition:

Assumption 2.1. *$P\{W \le w, U \le u | Z = z\}$ is continuous at $z = 0$ for all w and u.*

This assumption, referred to as local randomization assumption, has a clear behavioral interpretation. It allows individuals to have *imprecise* control over the running variable, yet restricts *deterministic* sorting around the threshold. In addition, a test can be constructed to empirically assess this assumption. Although it cannot be tested directly due to the presence of the unobservable U, we can instead evaluate the implication of continuity of the distribution of the observed predetermined variables W at the threshold $Z = 0$. See Lee (2008) and Lee and Lemieux (2010) for a further exposition of this assumption and an interpretation of it in terms of a locally randomized experiment.

In this paper we propose a test for this null hypothesis of continuity in the distributions of the predetermined variables W at the threshold $Z = 0$. There are certain elements of our hypothesis of interest that are worth emphasizing. First, it is testing an implication of Assumption 2.1 that integrates out the unobserved random variable U. It is however superior, in terms of power, to the weaker one of continuity in means that practitioners commonly test. Specifically, our test will be more powerful against a class of alternatives consisting of different distributions with equal means. Second, our hypothesis is also conceptually different to one of interest in the popular manipulation test proposed in McCrary (2008). This test exploits another implication of the identification assumption proposed by Lee (2008); that is, the continuity of the unconditional density of Z at the threshold.

To formalize the testing problem, let $X^{(n)} = \{(Y_i, W_i, Z_i) : 1 \leq i \leq n\} \sim P_n \in \mathbf{P}_n$ denote the observed data, where $\mathbf{P}_n$ is a set of distributions on a sample space $\mathcal{X}_n$. To better understand the test we propose in this paper, it is useful to represent the covariates W using the potential outcomes notation. To this end, let the observed covariate W taking values in $\mathcal{W}$ be

$$W_i = W_i(1)A_i + W_i(0)(1 - A_i) \ , \tag{2}$$

where $A_i = I\{Z_i \geq 0\}$ as before, so that $W_i = W_i(0)$ if $Z_i < 0$ and $W_i = W_i(1)$ if $Z_i \geq 0$. We denote by $H_0(w|z)$ the conditional on $Z = z$ cdf of $W(0)$ and by $H_1(w|z)$ the conditional on $Z = z$ cdf of $W(1)$. Using this notation, the researcher is interested in testing

$$H_0 : P_n \in \mathbf{P}_{n,0} \text{ versus } H_1 : P_n \in \mathbf{P}_n \setminus \mathbf{P}_{n,0} \ , \tag{3}$$

where

$$\mathbf{P}_{n,0} = \{P_n \in \mathbf{P}_n : H_0(w|0) = H_1(w|0) \text{ for all } w \in \mathcal{W}\} \ ,$$

is the subset of distributions such that the conditional distribution of W is continuous at the threshold $Z = 0$.

Remark 2.1. The testing problem in (3) is motivated with aim of verifying the validity of the identification assumption. However, conceptually, the test proposed in this paper can be used to perform distributional inference on the outcome at the threshold as in Shen and Zhang (2014). In our simulations in Section 5, we compare the performance our test to the one proposed in Shen and Zhang (2014) in the context of the testing problem in (3). ■

3 A permutation test based on induced ordered statistics

The test we propose is based on $2q$ values of $\{W_i : 1 \leq i \leq n\}$, such that q of those are associated with the q closets values of $\{Z_i : 1 \leq i \leq n\}$ to the right of the threshold, and the remaining q are associated with the q closet values of $\{Z_i : 1 \leq i \leq n\}$ to the left of the threshold. The number of observations, q, on either side of the threshold is a tuning parameter in our framework, and our asymptotics are intended to approximate a situation where there are few observations "close" to the threshold on either side. In other words, we consider a framework where q is fixed and $n \to \infty$.

To be precise in the construction of our test, denote by

$$Z_{n,(1)} \le Z_{n,(2)} \le \cdots \le Z_{n,(n)} \tag{4}$$

the order statistics of the sample $\{Z_i : 1 \le i \le n\}$ and by

$$W_{n,[1]}, W_{n,[2]}, \ldots, W_{n,[n]} \tag{5}$$

the corresponding values of the sample $\{W_i : 1 \le i \le n\}$, i.e., $W_{n,[j]} = W_k$ if $Z_{n,(j)} = Z_k$ for $k = 1, \ldots, n$. The random variables in (5) are called *induced order statistics* or *concomitants* of order statistics, see David and Galambos (1974); Bhattacharya (1974).

In order to construct our test statistic, we first take the q closest values in (4) to the right of zero and the q closest values to the left of zero. We denote these ordered values by

$$Z^-_{n,(q)} \le \cdots \le Z^-_{n,(1)} < 0 \text{ and } 0 \le Z^+_{n,(1)} \le \cdots \le Z^+_{n,(q)} , \tag{6}$$

respectively, and the corresponding induced values in (5) by

$$W^-_{n,[q]}, \ldots, W^-_{n,[1]} \text{ and } W^+_{n,[1]}, \ldots, W^+_{n,[q]} . \tag{7}$$

Note that while the values in (6) are ordered, those in (7) are not necessarily ordered.

Using the representation in (2), the random variables $(W^-_{n,[1]}, \ldots, W^-_{n,[q]})$ are viewed as an independent sample of $W(0)$ conditional on Z being "close" to zero from the left, while $(W^+_{n,[1]}, \ldots, W^+_{n,[q]})$ are viewed as an independent sample of $W(1)$ conditional on Z being "close" to zero from the right. We therefore use each of these two samples to compute empirical cdfs as follows,

$$\hat{H}^-_n(w) = \frac{1}{q}\sum_{j=1}^{q} I\{W^-_{n,[j]} \le w\} \text{ and } \hat{H}^+_n(w) = \frac{1}{q}\sum_{j=1}^{q} I\{W^+_{n,[j]} \le w\} . \tag{8}$$

Finally, letting

$$S_n = (S_{n,1}, \ldots, S_{n,2q}) = (W^-_{n,[1]}, \ldots, W^-_{n,[q]}, W^+_{n,[1]}, \ldots, W^+_{n,[q]}) , \tag{9}$$

denote the pooled sample of induced order statistics, we can define our test statistic as

$$T(S_n) = \int_{-\infty}^{\infty} (\hat{H}^-_n(w) - \hat{H}^+_n(w))^2 d\bar{H}_n(w) , \tag{10}$$

where $\bar{H}_n(w) = (1/2)\hat{H}^-_n(w) + (1/2)\hat{H}^+_n(w)$. The statistic $T(S_n)$ in (10) is the usual Cramér Von Mises test statistic, see Hajek et al. (1999, p. 101).

We propose to compute the critical values of our test by a permutation test as follows. Let $\mathbf{G}$ denote the set of all permutations $\pi = (\pi(1), \ldots, \pi(2q))$ of $\{1, \ldots, 2q\}$ and let

$$S_n^{\pi} = (S_{n,\pi(1)}, \ldots, S_{n,\pi(2q)}) ,$$

be the permuted vector S_n in (9) according to π. Let $M = |\mathbf{G}|$ be the cardinality of $\mathbf{G}$ and denote by

$$T^{(1)}(S_n) \le T^{(2)}(S_n) \le \cdots \le T^{(M)}(S_n)$$

the ordered values of $\{T(S_n^{\pi}) : \pi \in \mathbf{G}\}$. For $\alpha \in (0, 1)$, let $k = \lceil M(1-\alpha) \rceil$ and define

$$\begin{aligned} M^{+}(S_n) &= |\{1 \le j \le M : T^{(j)}(S_n) > T^{(k)}(S_n)\}| \\ M^{0}(S_n) &= |\{1 \le j \le M : T^{(j)}(S_n) = T^{(k)}(S_n)\}| . \end{aligned} \tag{11}$$

Using this notation, the proposed test is given by

$$\phi(S_n) = \begin{cases} 1 & T(S_n) > T^{(k)}(S_n) \\ a(S_n) & T(S_n) = T^{(k)}(S_n) \\ 0 & T(S_n) < T^{(k)}(S_n) \end{cases} , \tag{12}$$

where

$$a(S_n) = \frac{M\alpha - M^{+}(S_n)}{M^{0}(S_n)} .$$

It is important to understand that the test in (12) is not necessarily level $\alpha \in (0, 1)$ in finite samples, and the justification for using this test relies on asymptotic arguments that we describe in Section 4. To see why this is the case, note that under the null hypothesis in (3) it is not necessarily true that the distribution of S_n is invariant to permutations of $\{1, \ldots, 2q\}$ - which is required for the finite sample validity of a permutation test, see Lehmann and Romano (2005). The lack of invariance in finite samples lies behind the fact that the random variables in S_n are not draws from $H_0(w|0)$ and $H_1(w|0)$, but rather from $H_0(w|Z^{-}_{n,(j)})$ and $H_1(w|Z^{+}_{n,(j)})$, $j \in \{1, \ldots, q\}$. Under the null hypothesis in (3), these two distributions are not necessarily the same and therefore permuting the elements of S_n may not keep the joint distribution unaffected. However, if each of these conditional cdfs are continuous in Z, it follows a sample from $H_0(w|Z^{-}_{n,(j)})$ exhibits a similar behavior to a sample from $H_0(w|0)$, at least for n sufficiently large. This insight of approximating randomization tests when the randomization hypothesis does not hold in finite samples, but is satisfied in the limit, was first developed by Canay et al. (2014) in a context where the group of transformations $\mathbf{G}$ was essentially sign-changes. In this paper we show that such insight can be applied to permutations tests under a set of conditions that capture the fact that several test statistics for our problem, including the one in (10), are discontinuous and do not satisfy the no-ties condition required in Canay et al. (2014).

To see this last point more clearly, it is convenient to write the test statistic in (10) using an alternative representation. This representation may be also more convenient for implementing the test in practice, and it makes some of the arguments used in the proof of Theorem 4.2 more transparent. Let

$$R_{n,i} = \sum_{j=1}^{2q} I\{S_{n,j} \leq S_{n,i}\} \ , \tag{13}$$

be the rank of $S_{n,i}$ in the pooled vector S_n in (9). Let $R^*_{n,1} < R^*_{n,2} < \cdots < R^*_{n,q}$ denote the increasingly ordered ranks $R_{n,1}, \ldots, R_{n,q}$ corresponding to the first sample (i.e., first q values) and $R^*_{n,q+1} < \cdots < R^*_{n,2q}$ denote the increasingly ordered ranks $R_{n,q+1}, \ldots, R_{n,2q}$ corresponding to the second sample (i.e., last q values). Letting

$$T^*(S_n) = \frac{1}{q}\sum_{i=1}^{q}(R^*_{n,i} - i)^2 + \frac{1}{q}\sum_{j=1}^{q}(R^*_{n,q+j} - j)^2 \tag{14}$$

it follows that

$$T(S_n) = \frac{1}{q}T^*(S_n) - \frac{4q^2 - 1}{12q} \ ,$$

see Hajek et al. (1999, p. 102). The expression in (14) immediately shows two properties of the statistics $T(s)$. First, $T(s)$ is not a continuous function of s as the ranks make discrete changes with s. Second, $T(s) = T(s')$ whenever s and s' share the same ranks, which immediately follows from the definition of $T^*(s)$. This last property is what makes rank test statistics violate the no-ties condition in Canay et al. (2014).

Remark 3.1. When M is too large the researcher may use a stochastic approximation to $\phi(S_n)$ without affecting the properties of our test. More formally, let

$$\hat{\mathbf{G}} = \{\pi_1, \ldots, \pi_B\} \ , \tag{15}$$

where $\pi_1 = \{1, \ldots, 2q\}$ is the identity permutation and $\pi_2, \ldots, \pi_B$ are i.i.d. Uniform($\mathbf{G}$). Theorem 4.2 in the next section remains true if, in the construction of $\phi(S_n)$, $\mathbf{G}$ is replaced by $\hat{\mathbf{G}}$. ■

Remark 3.2. The test in (12) is possibly randomized. In case one prefers not to randomize, note that the non-randomized test that rejects if $T(S_n) > T^{(k)}(S_n)$ is asymptotically level α by Theorem 4.2. In our simulations, this test has rejection probability under the null hypothesis only slightly less than α when M is not too small. ■

Remark 3.3. The formal results we present in the next section are not restricted to the Cramér Von Mises test statistic in (10) and apply to other rank statistics satisfying our assumptions, e.g., the Kolmogorov Smirnov statistics. We restrict our discussion to the statistic in (10) for simplicity of exposition. ■

4 Main Results

In this section we derive the asymptotic properties of the test $\phi(S_n)$ in (12) using an asymptotic framework where q is fixed and $n \to \infty$. These asymptotics are intended to capture a situation where there are few observations that are "close" to the threshold $Z = 0$ on either side. We do this in two parts. We first derive a result on the asymptotic properties of induced order statistics in (7) that may be of independent interest and provides an important milestone in proving the asymptotic validity of our test. We then use this intermediate result and use arguments similar to those in Canay et al. (2014) to prove our main theorem.

4.1 A result on induced order statistics

Consider the order statistics in (4) and the induced order statistics in (5). As in the previous section, denote the q closest values in (4) to the right of zero and the q closest values to the left of zero by

$$Z^-_{n,(q)} \le \cdots \le Z^-_{n,(1)} < 0 \text{ and } 0 \le Z^+_{n,(1)} \le \cdots \le Z^+_{n,(q)} ,$$

respectively, and the corresponding induced values in (5) by

$$W^-_{n,[q]}, \ldots, W^-_{n,[1]} \text{ and } W^+_{n,[1]}, \ldots, W^+_{n,[q]} .$$

This representation implies that $W_i = W_i(0)$ if $Z_i < 0$ and $W_i = W_i(1)$ if $Z_i \ge 0$, and therefore it follows that $W^-_{n,[j]} = W_{n,[j]}(0)$ and $W^+_{n,[j]} = W_{n,[j]}(1)$, by construction. Recall that $H_0(w|z)$ is the conditional on $Z = z$ cdf of $W(0)$ and $H_1(w|z)$ the conditional on $Z = z$ cdf of $W(1)$. To prove the main result in this section we make the following assumptions.

Assumption 4.1. *The random variable Z has a continuous distribution with cdf F and density f such that $f(z) > 0$ for all z in a neighborhood of zero.*

Assumption 4.2. *The conditional cdfs $H_0(w|z)$ and $H_1(w|z)$ are continuous functions of z.*

Theorem 4.1. *Let Assumptions 4.1 and 4.2 hold. Then,*

$$\Pr\left\{\bigcap_{j=1}^{q}\{W^-_{n,[j]} \le w^-_j\} \bigcap_{j=1}^{q}\{W^+_{n,[j]} \le w^+_j\}\right\} = \Pi_{j=1}^{q} H_0(w^-_j|0) \cdot \Pi_{j=1}^{q} H_1(w^+_j|0) + o(1) , \qquad (16)$$

as $n \to \infty$, for any $(w^-_1, \ldots, w^-_q, w^+_1, \ldots, w^+_q) \in \mathbf{R}^{2q}$.

Theorem 4.1 states that the joint distribution of the induced order statistics are asymptotically independent, with the first q random variables having limit distribution $H_0(w|0)$ and the last q random variables having limit distribution $H_1(w|0)$. The proof relies on the fact the induced order statistics $S_n = (W^-_{n,[q]}, \ldots, W^-_{n,[1]}, W^+_{n,[1]}, \ldots, W^+_{n,[q]})$ are conditionally independent given $(Z_1, \ldots, Z_n)$, with conditional cdfs

$$H_0(w|Z^-_{n,(q)}), \ldots, H_0(w|Z^-_{n,(1)}), H_1(w|Z^+_{n,(1)}), \ldots, H_1(w|Z^+_{n,(q)}) \ .$$

The result then follows by showing that $Z^-_{n,(j)} = o_p(1)$ and $Z^+_{n,(j)} = o_p(1)$ for all $j = 1, \ldots, q$, and invoking standard properties of weak convergence.

Theorem 4.1 plays a fundamental role in the proof of Theorem 4.2 in the next section. It is the intermediate step that guarantees that, under the null hypothesis in (3), we have

$$S_n \stackrel{d}{\to} S = (S_1, \ldots, S_{2q}) \ , \tag{17}$$

where $(S_1, \ldots, S_{2q})$ are i.i.d. with cdf $H_0(w|0) = H_1(w|0)$. This implies that $S^\pi \stackrel{d}{=} S$ for all $\pi \in \mathbf{G}$, which means that the limit random variable S is indeed invariant to permutations. In addition to Assumptions 4.1 and 4.2, we also require that the random variable W is either continuous or discrete to prove the main result of the next section.

Assumption 4.3. *The scalar random variables $W(0)$ and $W(1)$ are continuously distributed conditional on $Z = 0$.*

Assumption 4.4. *The scalar random variables $W(0)$ and $W(1)$ are discretely distributed with $m \in \mathbf{N}$ points of support.*

We note that Theorem 4.1 does not require either Assumption 4.3 or 4.4. We however use each of these assumptions as a primitive condition of Assumptions 4.5 and 4.6 below, which are the high-level assumptions we use to prove the asymptotic validity of the permutation test in (12) for the scalar case. For ease of exposition, we present the extension to the case where $W(0)$ and $W(1)$ are vectors of possibly continuous and discrete random variables in Appendix C.

4.2 Asymptotic validity under approximate invariance

In this section, we present our theory of permutation tests under approximate invariance. The treatment in this section employs the asymptotic framework developed by Canay et al. (2014), but with two important differences. First, our arguments illustrate a concrete case in which the framework in Canay et al. (2014) can be used for the group of permutations as opposed to the group of sign-changes. The result in Theorem 4.1 provides a fundamental step in this direction. Second, we adjust the arguments in Canay et al. (2014) to accommodate rank test statistics, which happen to be discontinuous and do not satisfy the so-called no-ties condition in Canay et al. (2014). We do this by exploiting the specific structure of rank test statistics, together with the requirement that the limit random variable S is either continuously or discretely distributed. We formalize our requirements for the continuous case in the following assumption.

Assumption 4.5. *If $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$, then*

(i) $S_n = S_n(X^{(n)}) \xrightarrow{d} S$ *under* P_n.

(ii) $S^\pi \stackrel{d}{=} S$ *for all* $\pi \in \mathbf{G}$.

(iii) S *is an absolutely continuous random variable taking values in* $\mathcal{S} \subseteq \mathbf{R}$.

(iv) $T : \mathcal{S} \to \mathbf{R}$ *is invariant to rank, i.e., it only depends on the order of the elements in* $(S_1, \ldots, S_{2q})$.

Assumption 4.5 states the high-level conditions that we use to show the asymptotic validity of the permutation test we propose in (12) and formally state in Theorem 4.2 below. The assumption is also written in a way that facilitates the comparison with the conditions in Canay et al. (2014). In the particular testing problem we consider in this paper, Assumption 4.5 follows from Assumptions 4.1-4.3, which may be easier to interpret and impose clear restrictions on the primitives of the model. To quickly see this, note that Theorem 4.1, and the statement in (17) in particular, imply that Assumptions 4.5.(i)-(ii) follow from Assumptions 4.1-4.2. In turn, Assumption 4.5.(iii) follows directly from Assumption 4.3. Finally, Assumption 4.5.(iv), as explained in Section 3, holds for several rank test statistics and for the test statistic in (10) in particular. We formalize all these results in Theorem 4.2, which shows that the permutation test defined in (12) leads to a test that is asymptotically level α whenever Assumption 4.5 holds.

We next formalize our requirements for the discrete case in the following assumption.

Assumption 4.6. *If $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$, then*

(i) $S_n = S_n(X^{(n)}) \xrightarrow{d} S$ *under* P_n.

(ii) $S^\pi \stackrel{d}{=} S$ *for all* $\pi \in \mathbf{G}$.

(iii) S *and* S_n *are discrete random variables taking values in* $\mathcal{S} \subseteq \mathbf{R}$ *with* $|\mathcal{S}| = m$ *for all* $n \geq 1$.

Parts (i) and (ii) of Assumption 4.6 coincide with parts (i) and (ii) of Assumption 4.5 and, accordingly, follow from Assumptions 4.1-4.2. Assumption 4.6.(iii) accommodates a case not allowed by Assumption 4.5.(iii), which required S to be absolutely continuous. This is important as many covariates are discrete in applications; see Section 6. Note that here we also require the random variable S_n to be discrete, as opposed to the continuous case. However, Assumption 4.6 does not impose any requirement on the test statistic $T : \mathcal{S} \to \mathbf{R}$.

We now formalize our main result in Theorem 4.2, which shows that the permutation test defined in (12) leads to a test that is asymptotically level α whenever either Assumption 4.5 or Assumption 4.6 hold. In addition, the same theorem also shows that Assumptions 4.1-4.4 are sufficient primitive conditions for the asymptotic validity of our test.

Theorem 4.2. *Suppose $X^{(n)} \sim P_n \in \mathbf{P}_n$ and consider the problem of testing (3). Let $S_n : \mathcal{X}_n \to \mathcal{S}$, $T : \mathcal{S} \to \mathbf{R}$ and $\mathbf{G} : \mathcal{S} \to \mathcal{S}$ be such that either Assumption 4.5 or Assumption 4.6 holds. Then, for any $\alpha \in (0,1)$, $\phi(S_n)$ defined in (12) satisfies*

$$E_{P_n}[\phi(S_n)] \to \alpha \tag{18}$$

as $n \to \infty$ whenever $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$. Moreover, if $T : \mathcal{S} \to \mathbf{R}$ is the Cramér Von Mises test statistic in (10) and Assumptions 4.1-4.2 and 4.3 hold, then Assumption 4.5 also holds and (32) follows. Additionally, if instead Assumptions 4.1-4.2 and 4.4 hold, then Assumption 4.6 also holds and (32) follows.

Remark 4.1. Theorem 4.2 implies that the proposed test is asymptotically similar, i.e., has limiting rejection probability equal to α if $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$. Although this property is not necessarily shared by the non-randomized version of our test, see Remark 3.2, we find in our simulations that the randomized and non-randomized versions of our test deliver similar rejection probabilities under the null hypothesis in all designs. ■

Remark 4.2. Earlier work on the asymptotic behavior of randomization tests includes Hoeffding (1952), Romano (1989), Romano (1990), and more recently, Chung and Romano (2013). The arguments in these papers involve showing that the "randomization distribution" (see, e.g., Chapter 15 of Lehmann and Romano, 2005) settles down to a fixed distribution as $|\mathbf{G}| \to \infty$. Our asymptotic framework follows that in Canay et al. (2014), where $|\mathbf{G}|$ is fixed and the "randomization distribution" will generally not settle down at all. ■

Remark 4.3. An asymptotically valid p-value for the test $\phi(S_n)$ defined in (12) can be computed as

$$\hat{p} = \hat{p}(S_n) = \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I\{T(S_n^{\pi}) \geq T(S_n)\} \ . \tag{19}$$

We report these p-values in Section 6. The same construction is also valid when $M = |\mathbf{G}|$ is large and the researcher uses a stochastic approximation that replaces $\mathbf{G}$ with $\hat{\mathbf{G}}$, see Remark 3.1. ■

Remark 4.4. Theorem 4.2 shows the validity of the test in (12) when the scalar random variable W is either discrete or continuous. However, the test statistic in (10) and the test construction in (12) immediately apply to the case where W is a vector consisting of a combination of discrete and continuously distributed random variables. In Appendix C we show the validity of the test in (12) for the vector case, which is a result we use in the empirical application of Section 6. ■

5 Monte Carlo Simulations

In this section, we illustrate the finite sample performance of our procedure for the testing problem in (3). We use the following designs to generate the sequence of random variables.

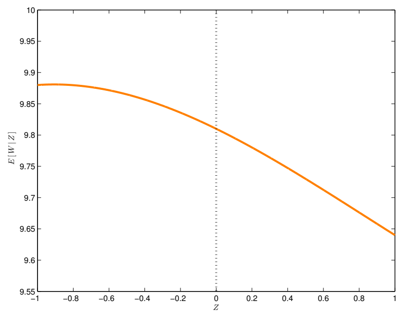

Figure 1: Model A

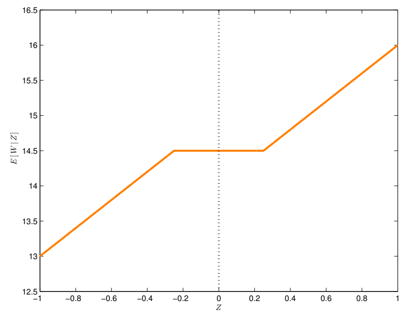

Figure 2: Model B

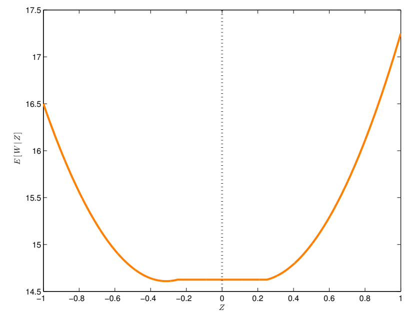

Figure 3: Model C

Model A: Following Shen and Zhang (2014), let

$$W = 9.81 - 0.14 \cdot Z - 0.05 \cdot Z^2 + 0.02 \cdot Z^3 + U \ ,$$

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.2^2)$. We use this design to highlight a model where the permutation hypothesis holds only approximately. The arguments for the validity of our test in this case are asymptotic.

Model B: Let

$$W = \begin{cases} 15 + 2.0 \cdot (Z - 0.5) + U & \text{if } Z \geq 0.25 \\ 15 + 2.0 \cdot (-0.25) + U & \text{if } -0.25 < Z < 0.25 \\ 15 + 2.0 \cdot Z + U & \text{if } Z \leq -0.25 \end{cases} ,$$

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.2^2)$. We use this design to highlight a model where the permutation hypothesis holds true in a neighborhood around the threshold. In this case, our test provides exact finite sample validity when all the observations used fall inside this neighborhood.

Model C: Let

$$W = \begin{cases} 15 + 2.5 \cdot (Z - 0.5) + 4 \cdot (Z - 0.5)^2 + U & \text{if } Z \geq 0.25 \\ 15 + 2.5 \cdot (-0.25) + 4 \cdot (-0.25)^2 + U & \text{if } -0.25 < Z < 0.25 \\ 15 + 2.5 \cdot Z + 4 \cdot Z^2 + U & \text{if } Z \leq -0.25 \end{cases} ,$$

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.2^2)$. This design is an extension of the previous one allowing for a quadratic function.

Model D: Following an adapted version of the design from Imbens and Kalyanaraman (2012), let

$$W = \begin{cases} 0.52 + 0.84 \cdot Z - 3.00 \cdot Z^2 + 7.99 \cdot Z^3 - 9.01 \cdot Z^4 + 3.56 \cdot Z^5 + U & \text{if } Z \geq 0 \\ 0.52 + 1.27 \cdot Z + 7.18 \cdot Z^2 + 20.21 \cdot Z^3 + 21.54 \cdot Z^4 + 7.33 \cdot Z^5 + U & \text{if } Z < 0 \end{cases} ,$$

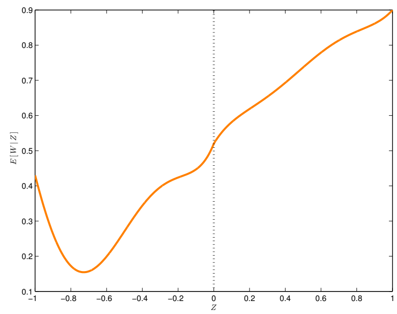

Figure 4: Model D

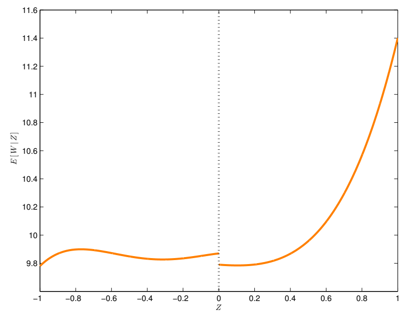

Figure 5: Model P

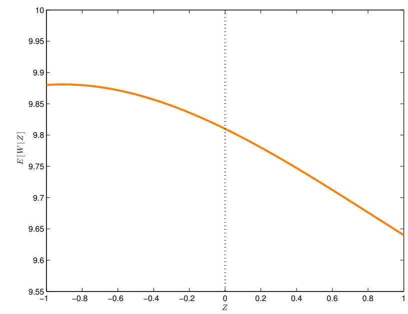

Figure 6: Model Q

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.25^2)$. We use this design to demonstrate the asymptotic performance of our test in a somewhat difficult case. It specifically illustrates the limitations of the asymptotic framework of most tests used in the RDD; see Kamat (2015) for a formal treatment of these limitations.

Model P: Following Shen and Zhang (2014), let

$$W = \begin{cases} 9.79 - 0.90 \cdot Z + 0.28 \cdot Z^2 + 0.86 \cdot Z^3 + 0.56 \cdot Z^4 + U & \text{if } Z \geq 0 \\ 9.87 + 0.15 \cdot Z - 0.54 \cdot Z^2 - 2.42 \cdot Z^3 - 1.82 \cdot Z^4 + U & \text{if } Z < 0 \end{cases} ,$$

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.2^2)$. We use this design to examine the power of our test as the functional forms are different on either side of the threshold.

Model Q: Following Shen and Zhang (2014), let

$$W = \begin{cases} 9.81 - 0.14 \cdot Z - 0.05 \cdot Z^2 + 0.02 \cdot X^3 + 2 \cdot U & \text{if } Z \geq 0 \\ 9.81 - 0.14 \cdot Z - 0.05 \cdot Z^2 + 0.02 \cdot X^3 + 1 \cdot U & \text{if } Z < 0 \end{cases} ,$$

where $Z \sim \mathcal{B}(2,2) - 1$ and $U \sim \mathcal{N}(0, 0.2^2)$. The motivation for this design is twofold. First, similar to the previous design it examines how our test performs in terms of power. Second, it highlights a model where our test is expected to be more powerful than the common practice of testing the continuity of means.

We compare the performance of our test to the one proposed by Shen and Zhang (2014), termed **SZ**. This test uses a Kolmogorov Smirnov type test statistic and traditional asymptotic arguments that involve a bandwidth parameter going to zero. To be specific, **SZ** rejects the null at 5% significance level when

$$A \left(N \cdot \tilde{f}/2 \right)^2 \sup_z \left| \tilde{H}_n^-(w) - \tilde{H}_n^+(w) \right| > 1.3581 , \tag{20}$$

where A is constant based on the choice of kernel, $\tilde{f}$ is the estimated density of Z at zero using the whole sample, and $\tilde{H}_n^-(w)$ and $\tilde{H}_n^+(w)$ are estimated cdfs of W at $Z = 0$ using local linear

	n	Perm			NR Perm			SZ		
		25	50	$0.5 \cdot n$	25	50	$0.5 \cdot n$	h^*	h_{SZ}	h^{**}
	500	4.80	6.90	4.80	4.80	6.90	4.80	3.30	3.85	4.70
A	1000	5.15	5.25	5.25	5.15	5.25	5.25	3.70	4.65	4.90
	5000	5.15	5.40	8.05	5.15	5.40	8.05	4.30	5.10	5.75
	6500	5.05	5.00	8.55	5.05	5.00	8.55	4.75	5.70	5.40
	500	4.50	4.00	4.50	4.45	4.00	4.45	3.10	3.95	7.90
B	1000	5.15	4.65	4.65	5.15	4.65	4.65	4.20	3.85	4.85
	5000	5.15	5.35	4.90	5.15	5.30	4.90	4.05	4.55	4.80
	6500	5.00	4.70	4.55	5.00	4.70	4.55	4.25	4.25	4.60
	500	4.50	4.00	4.50	4.45	4.00	4.45	3.15	4.00	4.90
C	1000	5.15	4.65	4.65	5.15	4.65	4.65	3.40	3.55	5.20
	5000	5.15	5.35	4.90	5.15	5.30	4.90	4.15	5.00	5.45
	6500	5.00	4.70	4.55	5.00	4.70	4.55	4.45	4.45	5.30
	500	31.65	92.80	31.65	31.55	92.80	31.55	7.95	12.45	18.25
D	1000	13.20	54.70	54.70	13.20	54.65	54.65	13.45	20.90	28.15
	5000	5.70	8.55	99.55	5.70	8.55	99.55	16.10	29.65	44.40
	6500	5.50	6.05	100.00	5.50	6.05	100.00	15.55	29.25	44.45

Table 1: Rejection probabilities under the null hypothesis

estimators. Shen and Zhang (2014) propose using a rule of thumb bandwidth h_{SZ} derived from Imbens and Kalyanaraman (2012). To measure robustness against bandwidth choice, we also report rejection probabilities for $0.75 \cdot h_{SZ}$ and $1.25 \cdot h_{SZ}$.

We also report rejection probabilities for an equality of means test for model Q. The test is performed using Calonico et al. (2014) and bandwidth is chosen using Imbens and Kalyanaraman (2012), termed **CCT** and h_{IK} respectively. The bandwidth choice for the bias estimate was also taken to be h_{IK}.

Table 1 reports the size results and Table 2 the power results of our test along with **SZ**. The test is conducted at a 5% significance level. 2,000 Monte Carlo simulations and 999 random permutations are performed. We use varying sample sizes of $n \in \{500, 1000, 5000, 6500\}$, where the final one is motivated by our empirical application in Section 6. Results for both the randomized and non randomized (NR) version of our permutation test are tabulated. Finally, we report three values of the tuning parameter q,

$$q \in \{25, 50, 0.05 \cdot n\} , \tag{21}$$

where the first two are fixed and the last one grows at rate n. The latter is not allowed by our asymptotic framework, and so we include it here as a robustness check.

	n	Perm			NR Perm			SZ			CCT
		25	50	$0.05 \cdot n$	25	50	$0.05 \cdot n$	h^*	h_{SZ}	h^{**}	h_{IK}
	500	24.15	38.25	24.15	24.00	38.25	24.00	16.70	23.65	30.10	–
P	1000	25.70	42.55	42.55	25.65	42.55	42.55	31.20	39.65	49.35	–
	5000	24.10	44.15	98.25	24.10	44.15	98.25	80.60	89.15	93.35	–
	6500	25.35	47.80	99.35	25.20	47.80	99.35	88.15	95.40	97.35	–
	500	17.00	48.35	17.00	16.95	48.35	16.95	11.65	16.60	22.70	6.15
Q	1000	17.70	47.30	47.30	17.60	47.30	47.30	27.35	39.45	51.05	5.00
	5000	19.25	49.70	100.00	19.25	49.70	100.00	97.60	99.80	99.95	4.85
	6500	18.75	49.10	100.00	18.70	49.10	100.00	99.55	100.00	100.00	5.25

Table 2: Rejection probabilities under the alternative hypothesis

There are a number of important features that arise from a visual inspection of Table 1. First, across all designs our permutation test seems to be more robust to the choice of q than **SZ** is to the choice of bandwidth. In fact, in most designs the null rejection probability of our test is closer to the nominal level than those of **SZ**. Second, models B and C illustrate that when the conditional cdf of W is continuous at $Z = 0$ and also in a neighborhood of $Z = 0$, our test works very well across the board and for all values of q, including $0.05 \cdot n$. Finally, model D is a difficult case where the conditional mean of W exhibits a high first-order derivative at the threshold, see Figure 4. Even for such a difficult model (see Kamat, 2015, for a formal treatment of why this case is expected to introduce size distortions in finite samples) we can see that the permutation test controls size well for n sufficiently large, as long as q is fixed, as it is the case in the asymptotic framework of Section 4. This, however, does not happen for the **SZ** test, where the rejection probabilities could reach 44% even when $n = 6500$. This last design also serves to illustrate that choosing q really big (note that $q = 325$ when $n = 6500$) does not necessarily deliver a test with good size controls. However, even for $q = 100$ (not reported), the rejection probabilities under the null in model D get close to 5% for $n = 6500$, which is again a case consistent with our asymptotic framework.

The power results in Table 2 also provide several important insights. First, it clearly highlights that the asymptotic power of our test depends on q. A higher choice of q results in higher power for any sample size. Second, model Q demonstrates that our test is more powerful than testing equality of means, which as expected, continues to give rejection probabilities of 5%. Third, it shows that allowing the choice of q to grow with sample size can result in power comparable to **SZ**, which uses traditional asymptotic arguments. However, this would come as possible sacrifice in size control and so it illustrate the size-power trade-off involved in the choice of q.

We conclude this section by noting that the new permutation test seems to deliver a better size control than test **SZ**, and improved power over simply testing the equality of means. However, its power could be lower than that of test **SZ** in some designs and some sample sizes. In the next section we show that out test can have non-trivial power in empirically relevant settings.

6 Empirical Application

In this section we reevaluate the validity of the design in Lee (2008), which has been used to illustrate recent methodological advancements in RDD; see Imbens and Kalyanaraman (2012). In this influential application of RDD, 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 time t. The assignment rule then takes a threshold value of zero that determines the treatment of incumbency to the Democratic candidate, which is used to study their election outcomes in time $t+1$. Required for the application of our test, five predetermined variables are present that contain electoral information on the Democrat runner and the opposition in time $t-1$ and t. These variables in particular are already determined by the time of the election in t. One of these variables is continuous and the remaining are discrete. The total number of observations is 6,559 with 2,740 below the threshold. The dataset is publicly available at `http://economics.mit.edu/faculty/angrist/data1/mhe`.

The predetermined variables consist of past electoral outcomes for the Democrat runner and opposition such as vote share, election wins and experience. It is easy to justify that these variables determine the quality of the candidates and hence might have discontinuously higher values for the incumbents. However, as noted in Section 2, application of RDD requires the conditional distribution of the perdetermined unobservables and observables to be continuous at the threshold. To verify this claim, Lee examines the implication if there are discontinuities in means of the predetermined variables. Local linear regressions are performed with observations in different margins around the threshold. To account for the discrete covariates, conditional probabilities are first estimated and then used in the local linear regression. It is worth emphasizing that this step is not present in our test, which allows for discrete covariates; see Section 4.2. The estimates and graphical illustration of the conditional means are used to conclude that there are no discontinuities at the threshold and the design is valid. Here, we frame the validity of the design in terms of the hypotheses in (3) and use the newly developed permutation test to the covariates in the dataset.

Table 3 reports the p-values of our test applied to the different covariates, where the number of random permutations performed are 999, as explained in Remark 3.1. Application of our test requires a choice of q that denotes the number of closest observations used from either side of the threshold. For the robustness of our conclusion, we report results for a range of choices of $q \in \{25, 50, 100\}$ that are consistent with the values we reported in Section 5. Our results show that the null hypothesis of continuity of the conditional distributions of the covariates at the threshold is rejected for several of the covariates at 5% significance level, in contrast to the results reported by Lee (2008). Table 3 additionally reports the maximum neighborhood of Z used (i.e., the

Variable	q		
	25	50	100
Democrat vote share election $t-1$	2.40	0.80	7.11
Democrat win election $t-1$	0.10	0.40	2.00
Democrat political experience t	6.61	0.80	0.30
Opposition political experience t	4.60	3.80	3.50
Democrat electoral experience t	14.11	15.12	6.51
Opposition electoral experience t	52.15	14.21	3.50
All covariates	24.12	20.12	20.92
Neighbourhood of observations	[-0.0043,0.0036]	[-0.0100,0.0087]	[-0.0193,0.0174]

Table 3: Test results with p-value (in %) for covariates in Lee (2008)

(a) $q = 25$ (b) $q = 50$ (c) $q = 100$

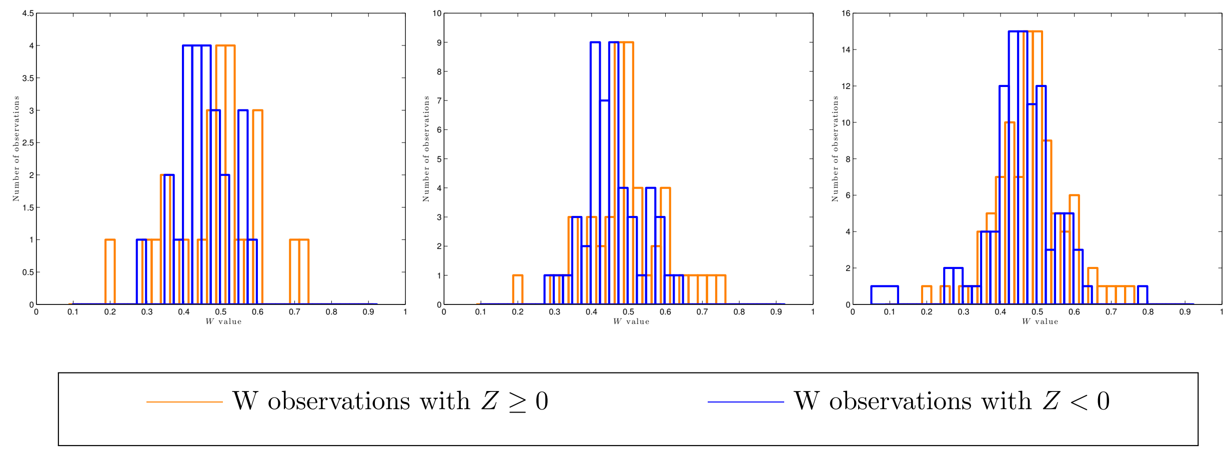

Figure 7: Histograms for *Democrat vote share election* $t-1$ values for the two samples determined by Z, the margin of victory in election t

maximum and minimum value of the $2q$ closest order statistics) to highlight the size of the window around the threshold that we effectively use. To better understand our results, we graphically illustrate in Figure 7 the distribution of the two samples for the continuous covariate *Democrat vote share election* $t-1$. A visual inspection of the difference in the two distributions suggests that for candidates close to the threshold, the incumbents might have discontinuously higher vote share in the past election. This reasoning is statistically confirmed by our test. Finally, when we test the null hypothesis that the distribution of the entire vector of covariates does not change discontinuously at the threshold (row labeled *All covariates*), then the p-values are above 20% for all values of q, leading to no rejections of this joint null hypothesis.

7 Concluding Remarks

In this paper we propose an asymptotically valid permutation test for a testable implication of the identification assumption in the regression discontinuity design (RDD). The asymptotic framework for our test is based on the simple intuition that observations close to the threshold are *approximately* randomly assigned to either side of it when the null hypothesis holds. This allows us to permute these observations to conduct an approximately valid test. Formally, we exploit the framework, with novel additions, from Canay et al. (2014), which first developed the insight of approximating randomization tests in this manner. Our results also represent a novel application of induced order statistics to frame our problem, and we present a result about induced order statistics that may be of independent interest.

A final important aspect we would like to highlight of our test is its simplicity. The test only requires computing two empirical cdf's for the induced order statistic, and does not involve kernels, local polynomials, bias correction, or bandwidth choices. We are currently working on a STATA package that will permit practitioners to effortlessly implement the test we propose in this paper.

A Proof of Theorem 4.1

First, note that the joint distribution of the induced order statistics $W^-_{n,[q]},\ldots,W^-_{n,[1]},W^+_{n,[1]},\ldots,W^+_{n,[q]}$ are conditionally independent given $(Z_1,\ldots,Z_n)$, with conditional cdfs

$$H_0(w|Z^-_{n,(q)}),\ldots,H_0(w|Z^-_{n,(1)}),H_1(w|Z^+_{n,(1)}),\ldots,H_1(w|Z^+_{n,(q)})\ .$$

A proof of this result can be found in Bhattacharya (1974, Lemma 1). Now let $\mathcal{A}=\sigma(Z_1,\ldots,Z_n)$ be the sigma algebra generated by $(Z_1,\ldots,Z_n)$. It follows that

$$\begin{aligned}
\Pr\left\{\bigcap_{j=1}^{q}\{W^-_{n,[j]}\le w^-_j\}\bigcap_{j=1}^{q}\{W^+_{n,[j]}\le w^+_j\}\right\} &= E\left[\Pr\left\{\bigcap_{j=1}^{q}\{W^-_{n,[j]}\le w^-_j\}\bigcap_{j=1}^{q}\{W^+_{n,[j]}\le w^+_j\}|\mathcal{A}\right\}\right]\\
&= E\left[\Pr\left\{\bigcap_{j=1}^{q}\{W_{n,[j]}(0)\le w^-_j\}\bigcap_{j=1}^{q}\{W_{n,[j]}(1)\le w^+_j\}|\mathcal{A}\right\}\right]\\
&= E\left[\Pi_{j=1}^{q}H_0(w^-_j|Z^-_{n,(j)})\cdot\Pi_{j=1}^{q}H_1(w^+_j|Z^+_{n,(j)})\right]\ .
\end{aligned}$$

The first and second equalities follow from the law of iterated expectations and (2), and the last equality follows from the conditional independence of the induced order statistics.

Let $f_{n,(q^-,\ldots,q^+)}(z_{q^-},\ldots,z_{q^+})$ denote the joint density of

$$Z^-_{n,(q)}\le\cdots\le Z^-_{n,(1)}<0\le Z^+_{n,(1)}\le\cdots\le Z^+_{n,(q)}\ ,$$

so that we can write the last term in the previous display as

$$\int_0^\infty\int_0^{z_{q^+}}\cdots\int_0^{z_{(q-1)^-}}\Pi_{j=1}^{q}H_0(w^-_j|z_{j^-})\cdot\Pi_{j=1}^{q}H_1(w^+_j|z_{j^+})f_{n,(q^-,\ldots,q^+)}(z_{q^-},\ldots,z_{q^+})dz_{q^-},\ldots,dz_{q^+}\ .$$

By Assumption 4.2, the integrand term

$$\Pi_{j=1}^{q}H_0(w^-_j|z_{j^-})\cdot\Pi_{j=1}^{q}H_1(w^+_j|z_{j^+})$$

is a bounded continuous function of $(z_{q^-},\ldots,z_{q^+})$. Suppose that the order statistics $Z^-_{n,(j)}$ and $Z^+_{n,(q)}$, for $j\in\{1,\ldots,q\}$, converge in distribution to a degenerate distribution with mass at $(0,0,\ldots,0)$. It would then follow from the definition of weak convergence (see van der Vaart, 1998, Lemma 2.2) that

$$\lim_{n\to\infty}E\left[\Pi_{j=1}^{q}H_0(w^-_j|z_{j^-})\cdot\Pi_{j=1}^{q}H_1(w^+_j|z_{j^+})\right]=E\left[\Pi_{j=1}^{q}H_0(w^-_j|0)\cdot\Pi_{j=1}^{q}H_1(w^+_j|0)\right]\ .$$

Hence, it is sufficient to prove that for any given $j\in\{1,\ldots,q\}$, $Z^-_{n,(j)}=o_p(1)$ and $Z^+_{n,(q)}=o_p(1)$. We prove $Z^+_{n,(q)}=o_p(1)$ by complete induction, and omit the other proof as the result follows from similar arguments.

Take $j = 1$ and $\epsilon > 0$ such that $F^+(\epsilon) \in (0,1)$, where

$$F^+(z) = P\{Z_i \le z | Z_i \ge 0\}$$

denotes the marginal cdf of Z conditional on it being non negative. Assumption 4.1 ensures that such ϵ exists. Note that

$$1 - F^+_{n,(1)}(\epsilon) \equiv \Pr\{Z_{n,(1)} > \epsilon\} = \Pr\{Z_i > \epsilon \text{ for all } i | Z_i \ge 0\} = [1 - F^+(\epsilon)]^n \to 0 \ , \tag{22}$$

so that $Z_{n,(1)} = o_p(1)$. Now let $F^+_{n,(j)}(x)$ denote the cdf of $Z_{n,(j)}$, which is given by

$$\begin{aligned} F^+_{n,(j)}(x) &= \Pr\{Z_{n,(j)} \le x\} \\ &= \Pr\{ \text{ at least } j \text{ of the } Z_i \text{ are less than or equal to } x | Z_i \ge 0\} \\ &= \sum_{i=j}^{n} \binom{n}{i} [F^+(x)]^i [1 - F^+(x)]^{n-i} \\ &= F^+_{n,(j+1)}(x) + \binom{n}{j} [F^+(x)]^j [1 - F^+(x)]^{n-j} \ , \end{aligned}$$

so that we can write

$$1 - F^+_{n,(j+1)}(x) = 1 - F_{n,(j)}(x) - \binom{n}{j} [F^+(x)]^j [1 - F^+(x)]^{n-j} \text{ for } j \in \{1, \dots, q-1\} \ . \tag{23}$$

It follows from (22) that $1 - F^+_{n,(1)}(\epsilon) \to 0$ for any $\epsilon > 0$ as $n \to \infty$. In order to complete the proof we assume that $1 - F^+_{n,(j)}(\epsilon) \to 0$ for $j \in \{1, \dots, q-1\}$ and show that this implies that $1 - F^+_{n,(j+1)}(\epsilon) \to 0$. By (23) this is equivalent to showing that

$$\binom{n}{j} [F^+(\epsilon)]^j [1 - F^+(\epsilon)]^{n-j} \to 0 \ .$$

To this end, note that

$$\binom{n}{j} [F^+(\epsilon)]^j [1 - F^+(\epsilon)]^{n-j} \le n^j [1 - F^+(\epsilon)]^{n-j} = \left[e^{\frac{j \log n}{n-j}} [1 - F^+(\epsilon)] \right]^{n-j} \to 0 \ , \tag{24}$$

where the convergence follows after noticing that there exists $N \in \mathbf{R}$ such that $e^{\frac{j \log n}{n-j}} [1 - F^+(\epsilon)] < 1$ for all $n > N$ and any $j \in \{1, \dots, q-1\}$. The result follows. ■

B Proof of Theorem 4.2

Part 1.

Continuous case: Let $\{P_n \in \mathbf{P}_{n,0} : n \ge 1\}$ be given. By Assumption 4.5(i) and the Almost Sure Representation Theorem (c.f van der Vaart, 1998, Theorem 2.19), there exists $\tilde{S}_n$, $\tilde{S}$, and $U \sim U(0,1)$, defined on a common probability space $(\Omega, \mathcal{A}, P)$, such that

$$\tilde{S}_n \to \tilde{S} \text{ w.p.1} \ ,$$

$\tilde{S}_n \stackrel{d}{=} S_n,\ \tilde{S} \stackrel{d}{=} S,$ and $U \perp (\tilde{S}_n, \tilde{S})$. Consider the permutation test based on $\tilde{S}_n$, this is,

$$\tilde{\phi}(\tilde{S}_n, U) \equiv \begin{cases} 1 & T(\tilde{S}_n) > T^{(k)}(\tilde{S}_n) \text{ or } T(\tilde{S}_n) = T^{(k)}(\tilde{S}_n) \text{ and } U < a(\tilde{S}_n) \\ 0 & T(\tilde{S}_n) < T^{(k)}(\tilde{S}_n) \end{cases} .$$

Denote the randomization test based on $\tilde{S}$ by $\tilde{\phi}(\tilde{S}, U)$, where the same uniform variable U is used in $\tilde{\phi}(\tilde{S}_n, U)$ and $\tilde{\phi}(\tilde{S}, U)$.

Since $\tilde{S}_n \stackrel{d}{=} S_n$, it follows immediately that $E_{P_n}[\phi(S_n)] = E_P[\tilde{\phi}(\tilde{S}_n, U)]$. In addition, since $\tilde{S} \stackrel{d}{=} S$, Assumption 4.5(ii) implies that $E_P[\tilde{\phi}(\tilde{S}, U)] = \alpha$ by the usual arguments behind randomization tests, see Lehmann and Romano (2005, Chapter 15). It therefore suffices to show

$$E_P[\tilde{\phi}(\tilde{S}_n, U)] \to E_P[\tilde{\phi}(\tilde{S}, U)] \ . \tag{25}$$

In order to show (33), let E_n be the event where the ordered values of $\{S_j : 1 \le j \le 2q\}$ and $\{S_{n,j} : 1 \le j \le 2q\}$ correspond to the same permutation π of $\{1, \ldots, 2q\}$, i.e., if $S_{\pi(j)} = S_k$ then $S_{n,\pi(j)} = S_{n,k}$ for $1 \le j \le 2q$ and $1 \le k \le 2q$. We first claim that $I\{E_n\} \to 1$ w.p.1. To see this, note that Assumption 4.5(iii) and $\tilde{S} \stackrel{d}{=} S$ imply that

$$\tilde{S}_{(1)}(\omega) < \tilde{S}_{(2)}(\omega) < \cdots < \tilde{S}_{(2q)}(\omega) \tag{26}$$

for all ω in a set with probability one under P. Moreover, since $\tilde{S}_n \to \tilde{S}$ w.p.1, there exists a set Ω^* with $P\{\Omega^*\} = 1$ such that both (34) and $\tilde{S}_n(\omega) \to \tilde{S}(\omega)$ hold for all $\omega \in \Omega^*$. For all ω in this set, let $\pi(1,\omega), \ldots, \pi(2q,\omega)$ be the permutation that delivers the order statistics in (34). It follows that for any $\omega \in \Omega^*$ and any $j \in \{1, \ldots, 2q-1\}$, if $\tilde{S}_{\pi(j,\omega)}(\omega) < \tilde{S}_{\pi(j+1,\omega)}(\omega)$ then

$$\tilde{S}_{n,\pi(j,\omega)}(\omega) < \tilde{S}_{n,\pi(j+1,\omega)}(\omega) \text{ for } n \text{ sufficiently large } \ . \tag{27}$$

We can therefore conclude that

$$I\{E_n\} \to 1 \ w.p.1 \ ,$$

which proves the first claim.

We now prove (33) in two steps. First, we note that

$$E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n\}] = E_P[\tilde{\phi}(\tilde{S}, U)I\{E_n\}] \ . \tag{28}$$

This is true because, on the event E_n, the rank statistics in (13) of the vectors $\tilde{S}_n^\pi$ and $\tilde{S}^\pi$ coincide for all $\pi \in \mathbf{G}$, and by Assumption 4.5(iv), the test statistic $T(S)$ only depends on the order of the observations, leading to $\tilde{\phi}(\tilde{S}_n, U) = \tilde{\phi}(\tilde{S}, U)$ on E_n. Second, since $I\{E_n\} \to 1$ w.p.1 it follows that $\tilde{\phi}(\tilde{S}, U)I\{E_n\} \to \tilde{\phi}(\tilde{S}, U)$ w.p.1 and $\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\} \to 0$ w.p.1. We can therefore use (38) and invoke the dominated convergence theorem to conclude that,

$$\begin{aligned} E_P[\tilde{\phi}(\tilde{S}_n, U)] &= E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\ &= E_P[\tilde{\phi}(\tilde{S}, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\ &\to E_P[\tilde{\phi}(\tilde{S}, U)] \ . \end{aligned}$$

This completes the proof of the first part of the statement of the theorem for the continuous case.

Discrete case: The proof for the discrete setting is similar to the continuous one with few intuitive differences. We reproduce it here for completeness.

Let $\{P_n \in \mathbf{P}_{n,0} : n \geq 1\}$ be given. By Assumption 4.6(i) and the Almost Sure Representation Theorem (c.f van der Vaart, 1998, Theorem 2.19), there exists $\tilde{S}_n$, $\tilde{S}$, and $U \sim U(0,1)$, defined on a common probability space $(\Omega, \mathcal{A}, P)$, such that

$$\tilde{S}_n \to \tilde{S} \text{ w.p.1 },$$

$\tilde{S}_n \stackrel{d}{=} S_n$, $\tilde{S} \stackrel{d}{=} S$, and $U \perp (\tilde{S}_n, \tilde{S})$. Consider the permutation test based on $\tilde{S}_n$, this is,

$$\tilde{\phi}(\tilde{S}_n, U) \equiv \begin{cases} 1 & T(\tilde{S}_n) > T^{(k)}(\tilde{S}_n) \text{ or } T(\tilde{S}_n) = T^{(k)}(\tilde{S}_n) \text{ and } U < a(\tilde{S}_n) \\ 0 & T(\tilde{S}_n) < T^{(k)}(\tilde{S}_n) \end{cases} .$$

Denote the randomization test based on $\tilde{S}$ by $\tilde{\phi}(\tilde{S}, U)$, where the same uniform variable U is used in $\tilde{\phi}(\tilde{S}_n, U)$ and $\tilde{\phi}(\tilde{S}, U)$.

Since $\tilde{S}_n \stackrel{d}{=} S_n$, it follows immediately that $E_{P_n}[\phi(S_n)] = E_P[\tilde{\phi}(\tilde{S}_n, U)]$. In addition, since $\tilde{S} \stackrel{d}{=} S$, Assumption 4.6(ii) implies that $E_P[\tilde{\phi}(\tilde{S}, U)] = \alpha$ by the usual arguments behind randomization tests, see Lehmann and Romano (2005, Chapter 15). It therefore suffices to show

$$E_P[\tilde{\phi}(\tilde{S}_n, U)] \to E_P[\tilde{\phi}(\tilde{S}, U)] . \tag{29}$$

In order to show (29), let E_n be the event where $\tilde{S}_n = \tilde{S}$. We first claim that $I\{E_n\} \to 1$ w.p.1. To see this, note that Assumption 4.6(iii), both $\tilde{S}$ and $\tilde{S}_n$ are discrete random variables taking values in $\mathcal{S}$ with $|\mathcal{S}| = m$. Moreover, since $\tilde{S}_n \to \tilde{S}$ w.p.1, there exists a set Ω^* with $P\{\Omega^*\} = 1$ such that $\tilde{S}_n(\omega) \to \tilde{S}(\omega)$ hold for all $\omega \in \Omega^*$. It follows that for any $\omega \in \Omega^*$ and any $j \in \{1, \ldots, 2q\}$,

$$\tilde{S}_{n,j}(\omega) = \tilde{S}_j(\omega) \text{ for } n \text{ sufficiently large } , \tag{30}$$

which follows from the fact that both S and S_n are discretely distributed. We can therefore conclude that

$$I\{E_n\} \to 1 \; w.p.1 ,$$

which proves the first claim.

We now prove (29) in two steps. First, we note that

$$E_P[\tilde{\phi}(\tilde{S}_n, U) I\{E_n\}] = E_P[\tilde{\phi}(\tilde{S}, U) I\{E_n\}] . \tag{31}$$

This is true because, on the event E_n, $\tilde{S}_n^\pi$ and $\tilde{S}^\pi$ coincide for all $\pi \in \mathbf{G}$, leading to $\tilde{\phi}(\tilde{S}_n, U) = \tilde{\phi}(\tilde{S}, U)$ on E_n. Second, since $I\{E_n\} \to 1$ $w.p.1$ it follows that $\tilde{\phi}(\tilde{S}, U) I\{E_n\} \to \tilde{\phi}(\tilde{S}, U)$ w.p.1

and $\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\} \to 0$ w.p.1. We can therefore use (31) and invoke the dominated convergence theorem to conclude that,

$$
\begin{aligned}
E_P[\tilde{\phi}(\tilde{S}_n, U)] &= E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\
&= E_P[\tilde{\phi}(\tilde{S}, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\
&\to E_P[\tilde{\phi}(\tilde{S}, U)] \ .
\end{aligned}
$$

This completes the proof for the discrete case and the first part of the statement of the theorem.

Part 2.

Let $\{P_n \in \mathbf{P}_{n,0} : n \geq 1\}$ be given and note that by Theorem 4.1 it follows that

$$
\begin{aligned}
S_n &= (S_{n,1}, \ldots, S_{n,2q}) = (W^-_{n,[1]}, \ldots, W^-_{n,[q]}, W^+_{n,[1]}, \ldots, W^+_{n,[q]}) \\
&\xrightarrow{d} (S_1, \ldots, S_{2q}) \ ,
\end{aligned}
$$

where $(S_1, \ldots, S_{2q})$ are i.i.d. with cdf $H(w|0) = H_0(w|0) = H_1(w|0)$. The conditions in Assumption 4.5.(i)-(ii) immediately follow as $(S_1, \ldots, S_{2q}) \stackrel{d}{=} (S_{\pi(1)}, \ldots, S_{\pi(2q)})$ for any $\pi \in \mathbf{G}$. Assumption 4.5.(iii) follows the fact that $(S_1, \ldots, S_{2q})$ are i.i.d. with cdf $H(w|0)$, where $H(w|0)$ is absolutely continuous by Assumption 4.3. Similarly, Assumption 4.6.(iii) follows the fact that $(S_1, \ldots, S_{2q})$ are i.i.d. with cdf $H(w|0)$, where $H(w|0)$ is discretely distributed by Assumption 4.4.

We are left to prove that the test statistic in (10) satisfies Assumption 4.5.(iv). To show this, note that $T(S)$ as in (10) admits the alternative representation

$$
T(S) = \frac{1}{q} T^*(S) - \frac{4q^2 - 1}{12q} \ ,
$$

where

$$
T^*(S) = \frac{1}{q} \sum_{i=1}^{q} (R_i^* - i)^2 + \frac{1}{q} \sum_{j=1}^{q} (R_{q+j}^* - j)^2 \ ,
$$

$R_1^* < R_2^* < \cdots < R_q^*$ denote the increasingly ordered ranks $R_1, \ldots, R_q$ of the first q variables in S, and $R_{q+1}^* < \cdots < R_{2q}^*$ are the increasingly ordered ranks $R_{q+1}, \ldots, R_{2q}$ of the last q values in S. It follows immediately that this test statistic satisfies Assumption 4.5.(iv). This completes the proof of the second part of the statement of the theorem. ■

C The multidimensional case

In this appendix we discuss the case where $W(0)$ and $W(1)$ are K-dimensional vectors. The test statistic in (10) and the test construction in (12) immediately apply to this case where W is a vector consisting of a combination of discrete and continuously distributed random variables. Here we show that the permutation test for this setting is also asymptotically valid. We first state the primitive conditions required to prove this.

Assumption C.1. *The random variable Z has a continuous distribution with cdf F and density f such that $f(z) > 0$ for all z in a neighborhood of zero.*

Assumption C.2. *The conditional cdfs $H_0(w|z)$ and $H_1(w|z)$ are continuous functions of z.*

Assumption C.3. *The random vectors $W(0)$ and $W(1)$ have each component $W_k(0)$ and $W_k(1)$ either continuously distributed or discretely distributed with $m_k \in \mathbf{N}$ points of support.*

Assumption C.1-C.2 are the same as Assumption 4.1-4.2, which are required for Theorem 4.1 to hold. Moreover, we note that Assumption C.3 is a stronger assumption than required and our result in Theorem C.1 will also hold when only assuming certain components are absolutely continuous conditional on $Z = 0$. However, we do require, similar to Assumption 4.4, that the discrete components are discrete for all Z. We thus state Assumption C.3 as it is for expositional purposes.

We formalize the high level assumptions required for the validity of the permutation test for the vector case in the following assumption.

Assumption C.4. *If $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$, then*

(i) $S_n = S_n(X^{(n)}) \xrightarrow{d} S$ *under* P_n.

(ii) $S^\pi \stackrel{d}{=} S$ *for all* $\pi \in \mathbf{G}$.

(iii) *S is a random vector such that each component S_k is either absolutely continuously distributed taking values in $\mathcal{S}_k \subseteq \mathbf{R}$ or discretely distributed taking values in $\mathcal{S}_k \subseteq \mathbf{R}$ with $|\mathcal{S}_k| = m_k$. For each discrete component, the corresponding $S_{n,k}$ is also discretely distributed with the same support for all $n \geq 1$.*

(iv) *$T : \mathcal{S} \to \mathbf{R}$ is invariant to rank with respect to each absolutely continuous component, i.e., it only depends on the order of the elements of each continuous component.*

The above assumption in essentially Assumption 4.5 and 4.6 applied individually to each component of the vector. We now formalize our result for the vector case in Theorem C.1, which shows that the permutation test defined in (12) leads to a test that is asymptotically level α whenever Assumption C.4 holds. In addition, the same theorem also shows that Assumption C.1-C.3 are sufficient primitive conditions for the asymptotic validity of our test.

Theorem C.1. *Suppose $X^{(n)} \sim P_n \in \mathbf{P}_n$ and consider the problem of testing (3). Let $S_n : \mathcal{X}_n \to \mathcal{S}$, $T : \mathcal{S} \to \mathbf{R}$ and $\mathbf{G} : \mathcal{S} \to \mathcal{S}$ be such that Assumption C.4 holds. Then, for any $\alpha \in (0,1)$, $\phi(S_n)$ defined in (12) satisfies*

$$E_{P_n}[\phi(S_n)] \to \alpha \tag{32}$$

as $n \to \infty$ whenever $P_n \in \mathbf{P}_{n,0}$ for all $n \geq 1$. Moreover, if $T : \mathcal{S} \to \mathbf{R}$ is the Cramér Von Mises test statistic in (10) and Assumptions C.1-C.3 hold, then Assumption C.4 also holds and (32) follows.

C.1 Proof of Theorem C.1

Part 1.

Let $\{P_n \in \mathbf{P}_{n,0} : n \geq 1\}$ be given. By Assumption C.4(i) and the Almost Sure Representation Theorem (c.f van der Vaart, 1998, Theorem 2.19), there exists $\tilde{S}_n$, $\tilde{S}$, and $U \sim U(0,1)$, defined on a common probability space $(\Omega, \mathcal{A}, P)$, such that

$$\tilde{S}_n \to \tilde{S} \text{ w.p.1 },$$

$\tilde{S}_n \stackrel{d}{=} S_n,\ \tilde{S} \stackrel{d}{=} S,$ and $U \perp (\tilde{S}_n, \tilde{S})$. Consider the permutation test based on $\tilde{S}_n$, this is,

$$\tilde{\phi}(\tilde{S}_n, U) \equiv \begin{cases} 1 & T(\tilde{S}_n) > T^{(k)}(\tilde{S}_n) \text{ or } T(\tilde{S}_n) = T^{(k)}(\tilde{S}_n) \text{ and } U < a(\tilde{S}_n) \\ 0 & T(\tilde{S}_n) < T^{(k)}(\tilde{S}_n) \end{cases} .$$

Denote the randomization test based on $\tilde{S}$ by $\tilde{\phi}(\tilde{S}, U)$, where the same uniform variable U is used in $\tilde{\phi}(\tilde{S}_n, U)$ and $\tilde{\phi}(\tilde{S}, U)$.

Since $\tilde{S}_n \stackrel{d}{=} S_n$, it follows immediately that $E_{P_n}[\phi(S_n)] = E_P[\tilde{\phi}(\tilde{S}_n, U)]$. In addition, since $\tilde{S} \stackrel{d}{=} S$, Assumption C.4(ii) implies that $E_P[\tilde{\phi}(\tilde{S}, U)] = \alpha$ by the usual arguments behind randomization tests, see Lehmann and Romano (2005, Chapter 15). It therefore suffices to show

$$E_P[\tilde{\phi}(\tilde{S}_n, U)] \to E_P[\tilde{\phi}(\tilde{S}, U)] \ . \tag{33}$$

Before we show (33), we introduce the additional notation to easily refer to the different components of the vectors S_j and $S_{n,j}$. Let the first K^c elements of S_j and $S_{n,j}$ for $j \in \{1, \dots, 2q\}$ denote the absolutely continuous components, where each component is denoted by $S^c_{k,j}$ and $S^c_{n,k,j}$ for $1 \leq k \leq K^c$. Let the remaining subvector S^d_j and $S^d_{n,j}$ of dimension $K^d = K - K^c$ for $j \in \{1, \dots, 2q\}$ denote the discrete component of S_j and $S_{n,j}$, and $(s^*_1, \dots, s^*_m)$ denote its points of support. Using this notation, we can partition S_j and $S_{n,j}$ as (S^c_j, S^d_j) and $(S^c_{n,j}, S^d_{n,j})$, respectively.

In order to show (33), let E_n be the event where the following holds. First, the ordered values of each continuous component $\{S^c_{k,j} : 1 \leq j \leq 2q\}$ and $\{S^c_{n,k,j} : 1 \leq j \leq 2q\}$ correspond to the same permutation π_k of $\{1, \dots, 2q\}$ for $1 \leq k \leq K^c$, i.e., if $S^c_{k,\pi_k(j)} = S^c_{k,l}$ then $S^c_{n,k,\pi_k(j)} = S^c_{n,k,l}$

for $1 \leq j, l \leq 2q$ and $1 \leq k \leq K^c$. Second, the discrete subvectors $\{S_j^d : 1 \leq j \leq 2q\}$ and $\{S_{n,j}^d : 1 \leq j \leq 2q\}$ coincide, i.e., $S_j^d = S_{n,j}^d$ for $1 \leq j \leq 2q$.

We first claim that $I\{E_n\} \to 1$ w.p.1. To see this, note that Assumption C.4(iii) and $\tilde{S} \stackrel{d}{=} S$ imply that for all ω in a set with probability one under P we have for each continuous component k of S that

$$\tilde{S}_{k,(1)}^c(\omega) < \tilde{S}_{k,(2)}^c(\omega) < \cdots < \tilde{S}_{k,(2q)}^c(\omega) \ , \tag{34}$$

and for the discrete subvector of $\tilde{S}$ that

$$\tilde{S}_j^d(\omega) = s_l^* \ , \tag{35}$$

for $1 \leq j \leq 2q$ and some $1 \leq l \leq m$. Moreover, since $\tilde{S}_n \to \tilde{S}$ w.p.1, there exists a set Ω^* with $P\{\Omega^*\} = 1$ such that (34), (35) and $\tilde{S}_n(\omega) \to \tilde{S}(\omega)$ hold for all $\omega \in \Omega^*$. For all ω in this set, let $\pi_k(1,\omega), \ldots, \pi_k(2q,\omega)$ be the permutation that delivers the order statistics in (34) for the k^{th} continuous component. It follows that for any $\omega \in \Omega^*$ and any $j \in \{1, \ldots, 2q-1\}$, if for any continuous component k we have $\tilde{S}_{k,\pi_k(j,\omega)}^c(\omega) < \tilde{S}_{k,\pi_k(j+1,\omega)}^c(\omega)$ then

$$\tilde{S}_{n,k,\pi_k(j,\omega)}^c(\omega) < \tilde{S}_{n,k,\pi_k(j+1,\omega)}^c(\omega) \text{ for } n \text{ sufficiently large} \ , \tag{36}$$

and moreover, if for the discrete subvector we have $\tilde{S}_j^d(\omega) = s_l^*$ then

$$\tilde{S}_{n,j}^d(\omega) = s_l^* \text{ for } n \text{ sufficiently large} \ , \tag{37}$$

which follows from the fact that both $\{S_j^d : 1 \leq j \leq 2q\}$ and $\{S_{n,j}^d : 1 \leq j \leq 2q\}$ are discretely distributed. We can therefore conclude that

$$I\{E_n\} \to 1 \ w.p.1 \ ,$$

which proves the first claim.

We now prove (33) in two steps. First, we note that

$$E_P[\tilde{\phi}(\tilde{S}_n, U) I\{E_n\}] = E_P[\tilde{\phi}(\tilde{S}, U) I\{E_n\}] \ . \tag{38}$$

This is true because, on the event E_n, the following two hold. First, for each continuous component the rank statistics in (13) of the vectors $\tilde{S}_{n,k}^{c,\pi}$ and $\tilde{S}_k^{c,\pi}$ coincide for $1 \leq k \leq K^c$ and for all $\pi \in \mathbf{G}$. Then we have by Assumption C.4(iv) that the test statistic $T(S)$ only depends on the order of the elements of each continuous component. Second, the discrete subvectors $\tilde{S}_n^{d,\pi}$ and $\tilde{S}^{d,\pi}$ coincide for all $\pi \in \mathbf{G}$. These two properties in turn result in, on the event E_n, $T(\tilde{S}_n^\pi)$ equaling $T(\tilde{S}^\pi)$ for all $\pi \in \mathbf{G}$, which leads to $\tilde{\phi}(\tilde{S}_n, U) = \tilde{\phi}(\tilde{S}, U)$ on E_n.

Then for the second step in proving (33), since $I\{E_n\} \to 1$ w.p.1 it follows that $\tilde{\phi}(\tilde{S}, U) I\{E_n\} \to \tilde{\phi}(\tilde{S}, U)$ w.p.1 and $\tilde{\phi}(\tilde{S}_n, U) I\{E_n^c\} \to 0$ w.p.1. We can therefore use (38) and invoke the dominated convergence theorem to conclude that,

$$\begin{aligned}E_P[\tilde{\phi}(\tilde{S}_n, U)] &= E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\ &= E_P[\tilde{\phi}(\tilde{S}, U)I\{E_n\}] + E_P[\tilde{\phi}(\tilde{S}_n, U)I\{E_n^c\}] \\ &\to E_P[\tilde{\phi}(\tilde{S}, U)] \ .\end{aligned}$$

This completes the proof of the first part of the statement of the theorem for the multidimensional case.

Part 2.

Let $\{P_n \in \mathbf{P}_{n,0} : n \geq 1\}$ be given and note that by Theorem 4.1 it follows that

$$\begin{aligned}S_n &= (S_{n,1}, \ldots, S_{n,2q}) = (W^-_{n,[1]}, \ldots, W^-_{n,[q]}, W^+_{n,[1]}, \ldots, W^+_{n,[q]}) \\ &\xrightarrow{d} (S_1, \ldots, S_{2q}) \ ,\end{aligned}$$

where $(S_1, \ldots, S_{2q})$ are i.i.d. with cdf $H(w|0) = H_0(w|0) = H_1(w|0)$. The conditions in Assumption C.4.(i)-(ii) immediately follow as $(S_1, \ldots, S_{2q}) \stackrel{d}{=} (S_{\pi(1)}, \ldots, S_{\pi(2q)})$ for any $\pi \in \mathbf{G}$. Assumption C.4.(iii) also follows immediately by Assumption C.3. Finally, to show Assumption C.4.(iv) we first demonstrate that the test statistic in (10) admits an alternate representation. By Assumption C.3, let without loss of generality the first K^c components be absolutely continuous and the rest be discrete. Denote by S_i^d the discrete subvector of S_i and by

$$R_{k,i} = \sum_{j=1}^{2q} I\{S^c_{k,j} \leq S^c_{k,i}\} \ ,$$

the rank of the k^{th} continuous component of S_i for $1 \leq i \leq 2q$ and $1 \leq k \leq K^c$. Finally, the test statistic can be rewritten in the following alternate representation

$$T(S) = \frac{1}{2q}\sum_{j=1}^{2q}\left(\frac{1}{q}\sum_{i=1}^{q}\left[I\{S_i^d \leq S_j^d\}\prod_{k=1}^{K^c} 1\{R_{k,i} \leq R_{k,j}\}\right] - \frac{1}{q}\sum_{i=q+1}^{2q}\left[I\{S_i^d \leq S_j^d\}\prod_{k=1}^{K^c}\{R_{k,i} \leq R_{k,j}\}\right]\right)^2 \ .$$

The above representation follows from first rewriting

$$I\{S_i \leq S_j\} = I\{S_i^d \leq S_j^d\}\prod_{k=1}^{K^c} I\{S^c_{k,i} \leq S^c_{k,j}\} \ ,$$

and then noticing that for $1 \leq k \leq K^c$

$$I\{S^c_{k,i} \leq S^c_{k,j}\} = I\{R_{k,i} \leq R_{k,j}\} \ .$$

This representation illustrates that for the absolutely continuous components the test statistic only depends on their individual orderings. It then follows immediately that this test statistic satisfies Assumption C.4.(iv). This completes the proof of the second part of the statement of the theorem for the multidimensional case. ■

References

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

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

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2014). Randomization tests under an approximate symmetry assumption. Manuscript.

Card, D., Dobkin, C. and Maestas, N. (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare. *American Economic Review*, **98** 2242–58.

Cattaneo, M. D., Frandsen, B. R. and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference*, **3** 1 – 24.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, **41** 484–507.

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

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica*, **78** 1863–1903. URL http://dx.doi.org/10.3982/ECTA8121.

Frandsen, B. R., Frlich, M. and Melly, B. (2012). Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics*, **168** 382 – 395.

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. URL http://www.jstor.org/stable/2692190.

Hajek, J., Sidak, Z. and Sen, P. K. (1999). *Theory of rank tests*. 2nd ed. Academic press.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics*, **23** pp. 169–192. URL http://www.jstor.org/stable/2236445.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, **79** 933–959.

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

Kamat, V. (2015). Nonparametric inference in the regression discontinuity design. ArXiv:1505.06483.

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, URL http://www.sciencedirect.com/science/article/pii/S0304407607001121.

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.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142** 698 – 714. The regression discontinuity design: Theory and applications, URL http://www.sciencedirect.com/science/article/pii/S0304407607001133.

Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics*, **17** 141–159. URL http://dx.doi.org/10.1214/aos/1176347007.

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association*, **85** pp. 686–692. URL http://www.jstor.org/stable/2290003.

Shen, S. and Zhang, X. (2014). Distributional tests for regression discontinuity: Theory and empirical examples.

Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review*, **99** 179–215.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, **32** pp. 711–754.