

Testing Conditional Stochastic Dominance at Target Points*

Federico A. Bugni

Department of Economics
Northwestern University

federico.bugni@northwestern.edu

Ivan A. Canay

Department of Economics
Northwestern University

iacanay@northwestern.edu

Deborah Kim

Department of Economics
University of Warwick

deborah.kim@warwick.ac.uk

November 18, 2025

Abstract

This paper introduces a novel test for conditional stochastic dominance at specific values of the conditioning covariates, referred to as target points. The test is relevant for analyzing treatment effects, discrimination, and income inequality. We propose a Kolmogorov–Smirnov-type test statistic based on induced order statistics from independent samples. A key feature is its data-independent critical value, which eliminates the need for resampling techniques such as the bootstrap, as well as kernel smoothing or parametric assumptions. Instead, the test relies on a tuning parameter to select relevant observations. We establish that the induced order statistics converge to independent draws from the true conditional distributions and that the test is asymptotically of level α under weak regularity conditions. Monte Carlo simulations confirm the strong finite-sample performance of our method.

KEYWORDS: stochastic dominance, regression discontinuity design, induced order statistics, rank tests, permutation tests.

JEL classification codes: C12, C14.

*We thank Xinran Li, David Kaplan, Matias Cattaneo, and participants at several conferences and seminars for helpful comments and discussion.

1 Introduction

The concept of stochastic dominance has long been central to numerous areas of applied research. This paper examines a specific aspect of stochastic dominance: testing conditional stochastic dominance (CSD) at specific values of the conditioning covariate, referred to as target points. Such conditional comparisons are crucial in many contexts, including evaluating treatment effects in social programs within a regression discontinuity design, analyzing economic disparities across demographic groups, and investigating potential discrimination in decision-making processes.

Unconditional stochastic dominance methods, which analyze entire distributions, have been extensively studied and widely applied in the literature, with foundational contributions dating back to Hodges (1958) and McFadden (1989) and more recent developments in Abadie (2002); Barrett and Donald (2003); Linton et al. (2005), and Linton et al. (2010), among others. However, in many empirical settings, the primary interest lies in dominance conditional on a subset of the population defined by specific characteristics or values of a conditioning variable. For instance, in regression discontinuity designs, the nature of the methodology necessitates comparing outcome distributions conditional on the cutoff of the running variable (Donald et al. (2012); Shen and Zhang (2016); Goldman and Kaplan (2018); Qu and Yoon (2019)). Likewise, in wage discrimination studies, researchers may seek to compare wage distributions across demographic groups while controlling for observed skill levels (Becker (1957); Canay et al. (2024); Bharadwaj et al. (2024)).

The primary goal of this paper is to test whether the conditional cumulative distribution function (CDF) of one variable stochastically dominates that of the other at specific values of a conditioning variable. Formally, we consider the null hypothesis

$$H_0 : F_Y(t|z) \leq F_X(t|z) \quad \text{for all } (t, z) \in \mathbf{R} \times \mathcal{Z} ,$$

against the alternative that there exists some (t, z) for which the reversed inequality holds strictly. Here, $F_Y(\cdot|z)$ and $F_X(\cdot|z)$ represent the conditional CDFs of the random variables Y and X , respectively, given $Z = z$. Importantly, we focus on situations where the set of target values, \mathcal{Z} , is not the entire support of Z , but rather a finite collection of points (including the case of a singleton). To address this testing problem, we propose a novel procedure that leverages induced order statistics based on independent samples from Y and X . Our test statistic, a Kolmogorov–Smirnov-type measure, captures the maximal deviation between the empirical CDFs of the two samples, conditional on observations near the target points. Crucially, the critical value we propose is derived in a deterministic, non-data-dependent manner once a tuning parameter is accounted for, ensuring computational simplicity.

Our contributions in this paper are both methodological and theoretical. First, we introduce a novel test for CSD at target points, particularly suited for settings where researchers seek to compare distributions conditional on covariates at specific values of the conditioning covariates. The proposed test exploits induced order statistics, leveraging observations closest to the target

conditioning point to construct empirical CDFs that form the basis of our test statistic. Unlike traditional methods, our approach neither relies on kernel smoothing nor imposes parametric assumptions on conditional distributions. However, it requires a tuning parameter, which serves a role analogous to bandwidth selection in nonparametric estimation.

Second, we establish the asymptotic validity of our test under two alternative asymptotic frameworks. In the first framework, the numbers of effective “local” observations, q_y and q_x , are held fixed as the sample size $n \rightarrow \infty$. In this setting, we show that the test statistic converges to a limiting experiment in which the induced order statistics behave as independent draws from the true conditional CDFs at the target point. This convergence yields a feasible critical value that remains valid without relying on resampling methods such as the bootstrap. Our regularity conditions in this fixed- q framework are mild: the conditional CDFs at the target point may contain finitely many discontinuities; both Y and X may be continuous, discrete, or mixed; and we require only that the conditional CDFs $F_Y(t | z)$ and $F_X(t | z)$ be continuous in z at the target point, uniformly in t . This is a weaker condition than the smoothness assumptions, such as twice differentiability in z , typically imposed in nonparametric methods. In the second framework, the numbers of effective observations q_y and q_x are allowed to diverge as $n \rightarrow \infty$. Building on recent results on convergence rates for induced order statistics (Bugni et al., 2025), we derive an explicit rate at which the rejection probability of our test approaches its nominal level as a function of q_y , q_x , and n . This second framework complements the fixed- q results by demonstrating that the test continues to control size when the number of effective observations increases, and it provides practical guidance for the data-dependent selection of the tuning parameters q_y and q_x .

Third, we show that the proposed critical value aligns with the one obtained from a permutation-based approach when the random variables Y and X are both continuous, thus establishing a natural connection between our method and the broader literature on permutation-based inference. To the best of our knowledge, this result provides the first formal justification for the validity of permutation-based inference in testing unconditional stochastic dominance. We demonstrate that the critical value of our test cannot be improved when both Y and X are continuous. However, this result does not extend to the case when either Y or X is discretely distributed. For this latter case, we introduce a *refined* critical value, which is typically smaller than the default one we propose and only a function of the support points for Y and X . This refinement enhances power relative to the default critical value, though it comes at the cost of increased computational complexity.

Finally, we examine the finite-sample performance of our test through Monte Carlo simulations and present an empirical application to illustrate its implementation. Both exercises suggest that the data-dependent rule we propose for selecting the key tuning parameters performs well in practice.

Our work contributes to the extensive literature on stochastic dominance testing, building on seminal contributions such as Anderson (1996); Davidson and Duclos (2000); Abadie (2002); Barrett and Donald (2003); Linton et al. (2005), and Linton et al. (2010). These studies examine the null

hypothesis of unconditional stochastic dominance and predominantly rely on asymptotic arguments and resampling techniques. Our approach to testing CSD differs in that, in the limit experiment, the conditional testing problem simplifies to a finite-sample unconditional testing problem. In the context of CSD testing, prior research—including [Delgado and Escanciano \(2013\)](#); [Gonzalo and Olmo \(2014\)](#); [Chang et al. \(2015\)](#), and [Andrews and Shi \(2017\)](#)—evaluate stochastic dominance over a range or the entire support of a continuous conditioning variable. Our work diverges from this literature by testing CSD at specific target points. A distinct line of research, including [Donald et al. \(2012\)](#); [Shen and Zhang \(2016\)](#); [Goldman and Kaplan \(2018\)](#), and [Qu and Yoon \(2019\)](#), studies CSD within regression discontinuity designs, where dominance is defined conditional on cutoffs. All of these methods assume continuity of conditional distributions, limiting their applicability in empirical settings featuring discrete mass points. Our method relaxes this constraint and accommodates distributions with finitely many discontinuities. The resulting test is novel, computationally simple, and valid across a broader class of distributions.

Our work closely aligns with the well-established literature on testing the equality of two continuous distributions. Foundational contributions by [Gnedenko and Korolyuk \(1951\)](#), [Korolyuk \(1955\)](#), and [Blackman \(1956\)](#) established that the finite-sample distribution of the two-sample one-sided Kolmogorov–Smirnov test statistic is pivotal under the null, deriving closed-form expressions under various simplifying assumptions. Later research by [Hodges \(1958\)](#); [Hájek and Šidák \(1967\)](#), and [Durbin \(1973\)](#) developed methods to approximate this finite-sample distribution. Although our null hypothesis differs, our critical value coincides with the corresponding quantile of this distribution. Notably, [Hodges \(1958\)](#) and [McFadden \(1989\)](#) proposed that the two-sample one-sided Kolmogorov–Smirnov test, originally designed for testing equality of two distributions, could be adapted to test stochastic dominance under continuity assumptions, though without formal proof. We provide a rigorous justification for this claim.

The remainder of the paper is organized as follows. [Section 2](#) defines the testing problem and introduces notation. [Section 3](#) presents the induced order statistics that form the basis of our procedure and introduces the proposed test. [Section 4](#) establishes its asymptotic validity under two alternative asymptotic frameworks. [Section 5](#) discusses several refinements and extensions, including results on rates of convergence, a data-dependent rule for selecting the tuning parameters, and a refinement of the test that increases power when Y and X are discrete. Importantly, [Section 5](#) also provides formal results on the validity of permutation tests for stochastic dominance with continuously distributed random variables, offering novel arguments for the validity of permutation-based inference in this setting. [Section 6](#) evaluates the finite-sample performance of our test through Monte Carlo simulations. Finally, [Section 7](#) concludes. All proofs are contained in the Appendix.

2 Testing problem

Let (Y, Z) and (X, Z) denote random pairs, each supported on $\mathbf{R} \times \mathbf{R}$. Let P_Y and P_X denote the joint distributions of (Y, Z) and (X, Z) , respectively. For $t \in \mathbf{R}$ and z in the support of Z , define the conditional distribution functions

$$F_Y(t|z) := P_Y\{Y \leq t \mid Z = z\} \quad \text{and} \quad F_X(t|z) := P_X\{X \leq t \mid Z = z\} .$$

We are interested in testing the null hypothesis:

$$H_0 : F_Y(t|z) \leq F_X(t|z) \text{ for all } (t, z) \in \mathbf{R} \times \mathcal{Z} \tag{2.1}$$

versus the alternative hypothesis:

$$H_1 : F_Y(t|z) > F_X(t|z) \text{ for some } (t, z) \in \mathbf{R} \times \mathcal{Z} ,$$

where $\mathcal{Z} = \{z_1, \dots, z_L\}$ is a finite set of target values of the conditioning variable. The case where $L = 1$ and Z is a continuously distributed random variable is both simpler and particularly relevant in empirical applications. To minimize notational clutter, we focus on this case for most of the remainder of the paper, with Section C.3 addressing the case where $L > 1$.

To test this hypothesis, we assume the analyst observes two independent samples:

$$\{(Y_i, Z_i) : 1 \leq i \leq n_y\} \quad \text{and} \quad \{(X_j, Z_j) : 1 \leq j \leq n_x\} . \tag{2.2}$$

We refer to the first sample as the Y -sample, which consists of n_y i.i.d. draws from P_Y , and to the second sample as the X -sample, which consists of n_x i.i.d. draws from P_X . Independence of the samples implies that the data-generating distribution is the product measure

$$P := P_Y \otimes P_X .$$

All probability and expectation statements in what follows are taken with respect to P unless otherwise noted. We denote by \mathbf{P} the class of data-generating distributions P satisfying the regularity conditions imposed later, and let

$$\mathbf{P}_0 := \{P \in \mathbf{P} : (2.1) \text{ holds}\} \tag{2.3}$$

denote the subset of distributions $P \in \mathbf{P}$ satisfying the null hypothesis in (2.1).

3 Test based on induced ordered statistics

Let the observed data be those given in (2.2). Let (q_y, q_x) be two small positive integers (relative to (n_y, n_x)) and consider the point $z_0 \in \mathcal{Z}$. The test we propose is based on the following two samples:

- The q_y values of $\{Y_i : 1 \leq i \leq n_y\}$ associated with the q_y values of $\{Z_i : 1 \leq i \leq n_y\}$ closest to z_0 , and
- The q_x values of $\{X_j : 1 \leq j \leq n_x\}$ associated with the q_x values of $\{Z_j : 1 \leq j \leq n_x\}$ closest to z_0 .

To define these samples formally, we introduce g -order statistics for the conditioning variable Z , where $g(Z) := |Z - z_0|$; see Reiss (1989, Section 2.1) and Kaufmann and Reiss (1992). For any two values $z, z' \in \mathcal{Z}$, we define the ordering \leq_g as follows:

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

This defines a g -ordering on the set \mathcal{Z} . The g -order statistics $Z_{g,(i)}$ are then the values of Z ordered according to this criterion:

$$Z_{g,(1)} \leq_g Z_{g,(2)} \leq_g \cdots \leq_g Z_{g,(n)}.$$

If there are ties in the g -ordering, they can be resolved arbitrarily—for example, by relying on the original sample order.

We then take the values of $\{Y_i : 1 \leq i \leq n_y\}$ associated with the q_y smallest g -ordered statistics of Z in the Y -sample, denoted by

$$Y_{n_y,[1]}, Y_{n_y,[2]}, \dots, Y_{n_y,[q_y]}. \quad (3.1)$$

That is, $Y_{n_y,[j]} = Y_k$ if $Z_{g,(j)} = Z_k$ for $k = 1, \dots, q_y$. Similarly, we take the values of $\{X_i : 1 \leq i \leq n_x\}$ associated with the q_x smallest g -ordered statistics of Z in the X -sample, denoted by

$$X_{n_x,[1]}, X_{n_x,[2]}, \dots, X_{n_x,[q_x]}. \quad (3.2)$$

The random variables in (3.1) and (3.2) are referred to as *induced order statistics* or *concomitants of order statistics*; see David and Galambos (1974); Bhattacharya (1974); Canay and Kamat (2018). Intuitively, we view these samples as independent samples of Y and X , conditional on Z being ‘close’ to z_0 . A key feature of our test is that it relies solely on these induced order statistics. Letting $n := \min\{n_y, n_x\}$ and $q := q_y + q_x$, the effective (pooled) sample is

$$S_n = (S_{n,1}, \dots, S_{n,q}) := (Y_{n_y,[1]}, \dots, Y_{n_y,[q_y]}, X_{n_x,[1]}, \dots, X_{n_x,[q_x]}). \quad (3.3)$$

It is also important to note that the first q_y elements of S_n are associated with the Y -sample, while the remaining q_x elements come from the X -sample.

Having defined the induced order statistics, we now define our test statistic as

$$T(S_n) = \sup_{t \in \mathbf{R}} \left(\hat{F}_{n,Y}(t) - \hat{F}_{n,X}(t) \right) = \max_{k \in \{1, \dots, q_y\}} \left(\hat{F}_{n,Y}(S_{n,k}) - \hat{F}_{n,X}(S_{n,k}) \right), \quad (3.4)$$

where the empirical CDFs are

$$\hat{F}_{n,Y}(t) := \frac{1}{q_y} \sum_{j=1}^{q_y} I\{S_{n,j} \leq t\} \quad \text{and} \quad \hat{F}_{n,X}(t) := \frac{1}{q_x} \sum_{j=1}^{q_x} I\{S_{n,q_y+j} \leq t\}.$$

The test statistic in (3.4) is a two-sample one-sided Kolmogorov–Smirnov (KS) test statistic, see Hajek et al. (1999, p. 99), and the test we propose rejects the null hypothesis in (2.1) when $T(S_n)$ exceeds a critical value, defined next.

To introduce the critical value, let $\alpha \in (0, 1)$ be given and $\{U_j : 1 \leq j \leq q\}$ be a sequence of uniform random variables i.i.d., that is, $U_j \sim U[0, 1]$. Define $\Delta(u)$ as

$$\Delta(u) := \frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\}, \quad (3.5)$$

and

$$c_\alpha(q_y, q) := \inf_{x \in \mathbf{R}} \left\{ P \left\{ \sup_{u \in (0,1)} \Delta(u) \leq x \right\} \geq 1 - \alpha \right\}. \quad (3.6)$$

The test we propose for the null hypothesis in (2.1) is

$$\phi(S_n) := I\{T(S_n) > c_\alpha(q_y, q)\}. \quad (3.7)$$

We reiterate that (3.7) corresponds to our test with a single target point ($L = 1$). Section C.3 extends this framework to the general case with multiple target points ($L > 1$). Furthermore, in addition to our default critical value in (3.6), we provide a refined critical value tailored for discrete Y and X with finite support in Section 5.2.

The choice of the two-sample one-sided KS test statistic in (3.4) is crucial for accommodating cases where Y and X are discrete or mixed. While one could construct analogues of our test in (3.7) using alternative statistics commonly employed in the stochastic dominance literature—such as one-sided versions of the Cramér–von Mises or Anderson–Darling statistics—our asymptotic validity result does not generally extend to these alternatives. See the discussion following Theorem 2.

Remark 1. *From a computational perspective, a notable feature of the test $\phi(S_n)$ is that the supremum over \mathbf{R} in (3.4) can be replaced by a maximum over the set $\{1, \dots, q_y\}$. This simplification arises because the KS test statistic increases only when evaluated at points corresponding to the Y -sample, which are the first q_y elements in the vector S_n . Consequently, the supremum in $c_\alpha(q_y, q)$*

can also be replaced with a maximum over $\{1, \dots, q_y\}$. This allows us to compute the critical value $c_\alpha(q_y, q)$ with arbitrary accuracy through simulation. ■

Remark 2. The values of (q_y, q_x) are tuning parameters chosen by the researcher. Section 4 shows that the test is asymptotically valid when (q_y, q_x) are either held fixed or allowed to diverge slowly with the sample size. In Section 5.1, we develop data-dependent guidelines for selecting (q_y, q_x) that are compatible with these asymptotic requirements, and we assess their practical performance through simulations in Section 6 and in the empirical application in Section C.4. ■

Remark 3. A natural application of our test is in the context of a sharp regression discontinuity design, where an outcome \tilde{Y} depends on a running variable \tilde{Z} , and the point of interest is the cutoff z_0 . The observed sample in this case is $\{(\tilde{Y}_i, \tilde{Z}_i) : 1 \leq i \leq n\}$, and the two samples needed for implementing our test are the following:

$$\begin{aligned} \{(Y_i, Z_i) : 1 \leq i \leq n_y\} &:= \{(\tilde{Y}_i, \tilde{Z}_i) : 1 \leq i \leq n_y \text{ such that } Z_i \leq z_0\} \text{ and} \\ \{(X_i, Z_i) : 1 \leq i \leq n_x\} &:= \{(\tilde{Y}_i, \tilde{Z}_i) : 1 \leq i \leq n_x \text{ such that } Z_i > z_0\}. \end{aligned}$$

Importantly, this formulation shows that the point z_0 can be either an interior or a boundary point in its support. We apply our test to this setting in our Monte Carlo simulations in Section 6 and our empirical application in Section C.4. ■

4 Asymptotic framework and formal results

In this section, we derive the asymptotic properties of the test in (3.7) using two alternative asymptotic frameworks. The first one requires $q := q_y + q_x$ to be fixed as $n \rightarrow \infty$ and represents a finite sample situation where the effective number of observations used by the test is too small to credibly invoke approximations that require a “large” value of q . The second framework requires $q \rightarrow \infty$ slowly as $n \rightarrow \infty$, 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 .

4.1 Asymptotic results for fixed q

In this section, we examine the asymptotic properties of the test in (3.7) within a framework where $q := q_y + q_x$ is fixed and $n := \min\{n_y, n_x\} \rightarrow \infty$. We first derive the asymptotic properties of induced order statistics in (3.3), and then present our main theorem.

We start by deriving a result on the induced order statistics collected in the vector S_n in (3.3). To do so, we make the following assumptions.

Assumption 1. For any $\varepsilon > 0$ and $z \in \mathcal{Z}$, $P\{Z \in (z - \varepsilon, z + \varepsilon)\} > 0$.

Assumption 2. For any $z \in \mathcal{Z}$ and sequence $z_k \rightarrow z$, $\sup_{t \in \mathbf{R}} |F_Y(t|z_k) - F_Y(t|z)| \rightarrow 0$ and $\sup_{t \in \mathbf{R}} |F_X(t|z_k) - F_X(t|z)| \rightarrow 0$.

Assumption 1 requires that the distribution of Z is locally dense at each of the points in \mathcal{Z} . This includes the case where Z has a mass point at $z \in \mathcal{Z}$. Assumption 2 is a smoothness assumption required to guarantee that conditioning on observations close to z is informative about the distribution conditional on $Z = z$.

Theorem 1. Let Assumptions 1 and 2 hold. Then,

$$S_n \xrightarrow{d} S = (S_1, \dots, S_q), \quad (4.1)$$

where for any $s := (s_1, \dots, s_q) \in \mathbf{R}^q$, the random vector S satisfies

$$P\{S \leq s\} = \prod_{j=1}^{q_y} F_Y(s_j|z_0) \cdot \prod_{j=q_y+1}^q F_X(s_j|z_0).$$

Theorem 1 is a special case of Theorem 3 in the appendix with $L = 1$, which generalizes Canay and Kamat (2018, Theorem 4.1) to accommodate multiple conditioning values. It establishes that the limiting distribution of the induced order statistics in the vector S_n is such that the elements of the vector, denoted by S , are mutually independent. Specifically, the first q_y elements of this vector follow the distribution $F_Y(\cdot|z_0)$, while the remaining q_x elements follow $F_X(\cdot|z_0)$. The proof leverages the fact that the induced order statistics S_n in (3.3) are conditionally independent given (Z_1, \dots, Z_n) , with conditional CDFs

$$F_Y(\cdot|Z_{n_y,(1)}), \dots, F_Y(\cdot|Z_{n_y,(q_y)}), F_X(\cdot|Z_{n_x,(1)}), \dots, F_X(\cdot|Z_{n_x,(q_x)}).$$

The result then follows by showing that $Z_{n_y,(j)} \xrightarrow{p} z_0$ and $Z_{n_x,(j)} \xrightarrow{p} z_0$ for all $j \in \{1, \dots, q\}$, and invoking standard properties of weak convergence. Theorem 1 plays a crucial role in our asymptotic validity result presented in Theorem 2.

In addition to Assumptions 1 and 2, we also require that, conditional on $Z = z$, the random variables Y and X have distributions with a finite number of discontinuity points. To state this assumption formally, let $\mathcal{D}_Y(z)$ and $\mathcal{D}_X(z)$ denote the sets of discontinuity points of the CDFs of $Y|Z = z$ and $X|Z = z$, respectively.

Assumption 3. For any $z \in \mathcal{Z}$, $|\mathcal{D}_Y(z)|$ and $|\mathcal{D}_X(z)|$ are finite.

It is important to note that Assumption 3 allows both Y and X to be continuous, discrete, or mixed random variables. However, it excludes cases where these variables have countably many discontinuities conditional on $z \in \mathcal{Z}$. We also point out that Theorem 1 does not require Assumption 3.

We now formalize our main result in Theorem 2, which shows that the test defined in (3.7) is asymptotically level α under the assumptions we just introduced. Below, we denote by $E_P[\cdot]$ the expected value with respect to the distribution $P \in \mathbf{P}$.

Theorem 2. *Let \mathbf{P} the space of distributions that satisfy Assumptions 1, 2, and 3 hold, and let \mathbf{P}_0 be as in (2.3). Let $\alpha \in (0, 1)$ be given, T be the KS test statistic in (3.4), and $\phi(\cdot)$ be the test in (3.7). Then,*

$$\limsup_{n \rightarrow \infty} E_P[\phi(S_n)] \leq \alpha \quad \text{for all } P \in \mathbf{P}_0 . \quad (4.2)$$

Theorem 2 establishes the asymptotic validity of the test in (3.7). There are three main reasons why the inequality in (C-2) may be strict, resulting in the limiting rejection probability strictly below α . First, for distributions P in the interior of \mathbf{P}_0 , where the inequality in (2.1) holds strictly for some $t \in \mathbf{R}$, the test is expected to reject with probability less than α , with a magnitude depending on the ‘distance’ of P from the boundary of \mathbf{P}_0 . Second, in cases where Y or X are not continuously distributed, the critical value defined in (3.6) serves as an upper bound for the desired quantile, as discussed further in Section 5.2. Finally, the test statistic $\Delta(u)$ in (3.5) is discretely distributed, taking only a limited number of distinct values. Consequently, the achieved significance level

$$\bar{\alpha} := P \left\{ \sup_{u \in (0,1)} \Delta(u) > c_\alpha(q_y, q) \right\} \quad (4.3)$$

satisfies $\bar{\alpha} \leq \alpha$ by definition, but may be strictly less than α . Whether $\bar{\alpha} = \alpha$ occurs or not depends on whether the critical value $c_\alpha(q_y, q)$ aligns exactly with one of the discrete jumps in the CDF of $\sup_{u \in (0,1)} \Delta(u)$, which depends on α , q_y , and q .

We derive Theorem 2 by linking the weak convergence of the induced order statistics S_n to the limit variable S in (4.1) with the finite-sample validity of the test $\phi(S)$ in the limit experiment. This connection becomes nontrivial when the data are not continuously distributed. The KS statistic plays a central role in addressing these challenges. First, our proof leverages the fact that the rank of the induced order statistics is preserved as the sample size grows. The KS statistic, being rank-based, ensures that the rejection rate of $\phi(S_n)$ converges to that of $\phi(S)$ in the limit experiment. Second, the structure of the KS statistic implies that our test controls size in the limit experiment over our class of null distributions—including those that are discrete or mixed—thereby establishing asymptotic validity. By contrast, as shown in Section C.2 in the appendix, analogous results generally fail when using the one-sided versions of the Cramér–von Mises or Anderson–Darling statistics, which are commonly used in the stochastic dominance testing. Nonetheless, in the special case where the data are continuously distributed, we show in Theorem 6 that the Cramér–von Mises-based test remains valid.

Remark 4. *Goldman and Kaplan (2018) develop an inference framework for the two-sample one-sided hypothesis test in (2.1), interpreting it as a multiple-testing procedure and with applications to the regression discontinuity design. As in this section, they adopt an asymptotic framework with*

fixed q and construct critical values by simulating i.i.d. $U(0, 1)$ random variables given a test statistic. Our approach to stochastic dominance testing, however, differs from theirs in several important respects. First, we focus on the KS statistic, whereas they advocate for a different statistic based on the so-called Dirichlet approach, which they argue may offer power advantages. This distinction allows us to accommodate discrete or mixed distributions, while their analysis is restricted to continuously distributed data; see Appendix C.2 for details.¹ Second, we also derive results under an asymptotic framework where $q \rightarrow \infty$ as $n \rightarrow \infty$. This extension allows us to obtain data-dependent rules for choosing the tuning parameters that satisfy the required rate-of-convergence conditions. ■

4.2 Asymptotic results for large q

In this section, we examine the asymptotic properties of the test in (3.7) within a framework where $q := q_y + q_x \rightarrow \infty$ as $n \rightarrow \infty$. Our results here follow from the recently derived rates of convergence for induced order statistics in Bugni et al. (2025), and so we keep the discussion brief.

The result in Theorem 1 establishes that the limiting distribution of the induced order statistics in the vector S_n is such that the elements of the limit vector, denoted by S , are mutually independent. Thus, S_n serves as an “approximation” to S , and while the infeasible test $\phi(S)$ would be ideal, in practice we must work with $\phi(S_n)$. Since $\phi \leq 1$, the variational characterization of total variation implies

$$|E_P[\phi(S_n)] - E_P[\phi(S)]| \leq \text{TV}(\mathcal{L}(S_n), \mathcal{L}(S)) , \quad (4.4)$$

where $\mathcal{L}(\cdot)$ denotes the distribution of a random vector and TV denotes total variation distance. Hence, a bound on the rate of convergence of $\text{TV}(\mathcal{L}(S_n), \mathcal{L}(S))$ would immediately yield a rate at which the rejection probability $E_P[\phi(S_n)]$ is asymptotically bounded by α , since in our setting $E_P[\phi(S)] \leq \alpha$. Unfortunately, Theorem 1 is silent about the rate for the convergence of S_n to S .

Theorem 5 in Appendix C.1 leverages the results in Bugni et al. (2025) to show that, under assumptions slightly stronger than those stated in Section 4.1, it follows that

$$\text{TV}(\mathcal{L}(S_n), \mathcal{L}(S)) = O\left(\frac{q_y^3}{n_y^2} + \frac{q_x^3}{n_x^2}\right) , \quad (4.5)$$

and so

$$E_P[\phi(S_n)] \leq \alpha + O\left(\frac{q_y^3}{n_y^2} + \frac{q_x^3}{n_x^2}\right) . \quad (4.6)$$

Hence, the test defined in (3.7) is asymptotically level α under those same assumptions, provided that $q_y = o(n_y^{2/3})$ and $q_x = o(n_x^{2/3})$. We rely on these rates of convergence in the next section to derive data-dependent rules of thumb for selecting the two tuning parameters.

¹Although Goldman and Kaplan (2018, p. 146) state that their test remains asymptotically valid—albeit conservative—for discrete data, they do not provide a formal proof. By contrast, Theorem 2 establishes this property for the KS statistic. Interestingly, asymptotic validity does not extend to other test statistics: Appendix C.2 shows that it fails for tests based on the Cramér–von Mises and Anderson–Darling statistics.

Remark 5. *Theorem 5 in Appendix C.1 also shows that $\phi(S_n)$ in (3.7) is consistent for any fixed alternative $P \in \mathbf{P}_1$ when $q \rightarrow \infty$ at the specified rates. ■*

Remark 6. *In a framework where $q \rightarrow \infty$, one could alternatively define a test using a critical value based on the $1 - \alpha$ quantile of the limiting distribution of the (scaled) KS statistic in (3.4). Denote this limiting critical value by $c_\alpha(\infty)$. Numerical evaluations show that our scaled critical value is smaller than the limiting critical value: $\sqrt{q_y q_x / q} c_\alpha(q_y, q) \leq c_\alpha(\infty)$ for $\alpha \in \{0.1, 0.05, 0.01\}$ and $(q_y, q_x) \in \{10, 20, \dots, 500\}^2$, with strict inequality in most cases. For example, when $\alpha = 5\%$ and $(q_y, q_x) = (70, 70)$ —a value consistent with our simulations—we obtain $\sqrt{q_y q_x / q} c_\alpha(q_y, q) = 1.1832 < 1.2239 = c_\alpha(\infty)$. The smaller critical value associated with our test translates into an improvement in power of roughly 4% under a local-power approximation for location-shift alternatives. We omit the details for brevity, but they are available upon request. ■*

5 Discussion and extensions

5.1 Data-dependent choice of tuning parameters

We now discuss the practical considerations for implementing our test. We propose a data-dependent method for two tuning parameters q_y and q_x , drawing on arguments from [Armstrong and Kolesár \(2018\)](#), similar to the approach used by [Bugni and Canay \(2021\)](#). Importantly, the analysis from these arguments is consistent with the rates of convergence implied by (4.5). Concretely, this method leverages a bias-variance trade-off inherent in the estimation of the conditional CDFs used in the test statistic for $\phi(S_n)$, within an asymptotic framework where q can grow slowly with n . Our goal is to provide practical guidance for choosing these tuning parameters based on the data, rather than claiming optimality or even validity of any sort. We examine the performance of this rule via Monte Carlo simulations in Section 6 and use it in the empirical application in Section C.4.

We propose choosing q_x and q_y using the following data-dependent rules:

$$q_y^* := n_y^{1/2} \left(\frac{4 \cdot \phi_{\mu_Z, \sigma_Z}^2(z_0)}{\frac{2}{\sigma_Z} \frac{1}{\sqrt{2\pi e}} + \frac{|\rho_Y|}{\sigma_Z \sqrt{1-\rho_Y^2}} \frac{1}{\sqrt{2\pi}}} \right)^{2/3} \quad (5.1)$$

and

$$q_x^* := n_x^{1/2} \left(\frac{4 \cdot \phi_{\mu_Z, \sigma_Z}^2(z_0)}{\frac{2}{\sigma_Z} \frac{1}{\sqrt{2\pi e}} + \frac{|\rho_X|}{\sigma_Z \sqrt{1-\rho_X^2}} \frac{1}{\sqrt{2\pi}}} \right)^{2/3}, \quad (5.2)$$

where $\mu_Z := E[Z]$, $\sigma_Z^2 := \text{Var}[Z]$, $\phi_{\mu_Z, \sigma_Z}(\cdot)$ denotes the probability density function of a normal distribution with mean μ_Z and variance σ_Z^2 , and ρ_Y and ρ_X are the correlation coefficients between Y and Z , and X and Z , respectively.

To provide some intuition as to why this rule of thumb may be reasonable, assume that the random variable Z is continuous with a density function $f_Z(\cdot)$ satisfying

$$|f_Z(z_1) - f_Z(z_2)| \leq C_Z |z_1 - z_2| \text{ for a Lipschitz constant } C_Z < \infty ,$$

and any values $z_1, z_2 \in \mathcal{Z}$. In addition, suppose that the conditional CDF of Y satisfies

$$\left| \frac{\partial F_Y(t|z)}{\partial z} \right| \leq C_Y \quad \text{for a constant } C_Y < \infty ,$$

and that the conditional CDF of X satisfies the same condition with a constant C_X . It can be shown that the standardized bias B_{n_y, q_y} associated with the estimator of the conditional CDF $F_Y(\cdot|z_0)$ satisfies

$$|B_{n_y, q_y}| \leq \frac{q_y^{3/2}}{n_y} \frac{2C_Z + C_Y}{4f_Z^2(z_0)} . \quad (5.3)$$

Let t^* denote the right-hand side of (5.3). Solving for q_y , we obtain

$$q_y = n_y^{2/3} (t^*)^{2/3} \left(\frac{4f_Z^2(z_0)}{2C_Z + C_Y} \right)^{2/3} .$$

The proposed data-dependent rule in (5.1) and (5.2) can be viewed as under-smoothed approximations of these values, where the unknown Lipschitz constants are approximated by the working model $Z \sim N(\mu_Z, \sigma_Z)$. This guarantees that $q_y = o(n_y^{2/3})$, which is the condition discussed in Section 4.2 to obtain (4.5). The constant multiplying $n_y^{1/2}$ in (5.1) is intuitive for two reasons. First, it reflects that a steeper density at z_0 , or a steeper derivative of the conditional CDFs at z_0 calls for smaller q_y . In such cases, nearby observations provide a poorer approximation of the quantities at z_0 . Since the maximum slope is determined by the constants C_Z and C_Y , the rule is inversely proportional to these constants. Second, the rule accounts for low density at z_0 . When $f_Z(z_0)$ is small, the q_y nearest observations are likely to be farther from z_0 , again requiring smaller q_y . While one could replace the normality assumption with a nonparametric estimator of $f_Z(\cdot)$, it is unfortunately impossible to adaptively choose C_Z and C_Y for testing (2.1) (see, e.g., [Armstrong and Kolesár, 2018](#)). Since any data-dependent rule for q must reference C_Z and C_Y , we prioritize simplicity and use normality for both $f_Z(\cdot)$ and the associated constants.

5.2 Refined critical value for discrete data

While our default critical value in (3.6) is valid as long as the distributions of random variables Y and X have finitely many discontinuities, it is possible to construct a smaller *refined* critical value when both variables are discretely distributed with a limited number of support points. Our test with the refined critical value still maintains the asymptotic validity, though it comes at the cost of additional computational complexity.

To motivate the refined critical value, let \mathbf{Y} and \mathbf{X} denote the support of Y and X , respectively. Our proof for the asymptotic validity (specifically, Theorem 4 in the appendix) relies on the following inequalities:

$$P \left\{ \sup_{u \in \mathcal{U}} \Delta(u) > c_\alpha(q_y, q) \right\} \leq P \left\{ \sup_{u \in (0,1)} \Delta(u) > c_\alpha(q_y, q) \right\} \leq \alpha.$$

where $\mathcal{U} := \cup_{t \in \mathbf{Y}} \{u = F_X(t|z_0)\}$ is the set of values that $F_X(t|z_0)$ takes as t varies over \mathbf{Y} and $\Delta(u)$ is given in (3.5). Once we replace the set \mathcal{U} with the interval $(0, 1)$, our default critical value $c_\alpha(q_y, q)$, defined as a quantile of $\sup_{u \in (0,1)} \Delta(u)$ in (3.6), ensures the probability is bounded below α . However, replacing the set \mathcal{U} with $(0, 1)$ could be unnecessarily conservative when \mathcal{U} contains only a few points—either because \mathbf{Y} contains only a few points or because $F_X(t|z_0)$ takes few distinct values as t varies. In fact, the cardinality of the set \mathcal{U} is determined by the smaller support size of Y and X .

To define our refined critical value, let r denote the smaller support size of Y and X ,

$$r := \min\{|\mathbf{Y}|, |\mathbf{X}|\},$$

and let \mathbf{U}_r denote the collection of all ordered r -tuples of distinct points in $(0,1)$,

$$\mathbf{U}_r := \{(u_1, u_2, \dots, u_r) \in (0, 1)^r : u_1 < u_2 < \dots < u_r\}.$$

We denote an arbitrary element of \mathbf{U}_r by \mathcal{U}_r . Our refined critical value is defined as

$$c_\alpha^r(q_y, q) := \inf_{x \in \mathbf{R}} \left\{ \inf_{\mathcal{U}_r \in \mathbf{U}_r} P \left\{ \sup_{u \in \mathcal{U}_r} \Delta(u) \leq x \right\} \geq 1 - \alpha \right\}, \quad (5.4)$$

with $\Delta(u)$ as in (3.5), and the refined test for the null hypothesis in (2.1) when either Y or X is discretely distributed with a limited number of support points is thus

$$\phi^r(S_n) := I\{T(S_n) > c_\alpha^r(q_y, q)\}. \quad (5.5)$$

Section C.3 presents the general version of this test for the case $L > 1$.

The power gains of using $c_\alpha^r(q_y, q)$ over $c_\alpha(q_y, q)$ are most pronounced when r is small. Our numerical analysis shows the largest gains for $r \leq 10$. Thus, this refinement is most effective for discrete data with a limited number of support points, rather than all discrete settings. Moreover, the computational cost of $c_\alpha^r(q_y, q)$ increases with r , and so as r grows, the gains diminish while the cost rises.

We propose to compute $c_\alpha^r(q_y, q)$ numerically, by solving the following optimization problem:

$$c_\alpha^r(q_y, q) = \min \left\{ x \in [c_{\text{lb}}, c_{\text{ub}}] \cap \mathbf{T} : \inf_{\mathcal{U}_r \in \mathbf{U}_r} P \left\{ \max_{u \in \mathcal{U}_r} \Delta(u) \leq x \right\} \geq 1 - \alpha \right\}, \quad (5.6)$$

where \mathbf{T} is the support of $\Delta(u)$ in (3.5). Here, $c_{\text{ub}} = c_\alpha(q_y, q)$ and c_{lb} is given by

$$c_{\text{lb}} := \min_{x \in \mathbf{T}} \left\{ P \left\{ \max_{u \in \left\{ \frac{1}{1+r}, \frac{2}{1+r}, \dots, \frac{r}{1+r} \right\}} \Delta(u) \leq x \right\} \geq 1 - \alpha \right\} .$$

The fact that $c_\alpha(q_y, q)$ provides a valid upper bound is unsurprising given the preceding discussion. On the other hand, c_{lb} serves as a valid lower bound because $\left\{ \frac{1}{1+r}, \frac{2}{1+r}, \dots, \frac{r}{1+r} \right\}$ is a specific element in \mathbf{U}_r . In our numerical evaluations, we often found that $c_\alpha^r(q_y, q) = c_{\text{lb}}$, but not always. This indicates that the additional optimization in (5.6) cannot be generally avoided. For modest values of r and q , however, this optimization step is computationally straightforward, primarily due to the relatively small number of points typically found in $[c_{\text{lb}}, c_{\text{ub}}] \cap \mathbf{T}$.

Remark 7. *The support \mathbf{T} of $\Delta(u)$ is a discrete subset of $[-1, 1]$ and can be easily enumerated for modest values of q . Specifically, the support has cardinality bounded by $(q_y + 1)(q_x + 1)$, and it is independent of the realizations of the random variables as well as the specific value that $u \in (0, 1)$ takes. ■*

5.3 Properties of our test in the limit experiment

In this section, we study the properties of the test $\phi(\cdot)$ in (3.7) in the limit experiment associated with the asymptotic framework in Section 4.1. By Theorem 1, this test is equivalent to $\phi(S)$, where

$$S = (S_1, \dots, S_q), \quad S_j \sim F_Y(\cdot | z_0) \text{ for } j \leq q_y \text{ and } S_j \sim F_X(\cdot | z_0) \text{ for } j > q_y . \quad (5.7)$$

In words, in the limit experiment, we observe one random sample of size q_y from the distribution $F_Y(\cdot | z_0)$ and the other independent random sample of size q_x from the distribution $F_X(\cdot | z_0)$. The KS test statistic in (3.4) is a function of S , and the critical value $c_\alpha(q_y, q)$ remains unchanged. We begin our discussion by focusing on the case where S is *continuously distributed*.

The finite-sample properties of the two-sample one-sided KS statistic have been extensively studied in the literature of testing equality of two continuous (unconditional) distributions. Early works established that the test statistic's finite-sample distribution is pivotal under the null and developed algorithms for its computation (Gnedenko and Korolyuk (1951); Korolyuk (1955); Blackman (1956); Hodges (1958); Hájek and Šidák (1967), and Durbin (1973)). Our critical value is obtained from this pivotal finite-sample distribution, despite the different null hypothesis.

This connection to the literature of testing equality of two distributions arises from the observation that the distribution $F_Y(\cdot | z_0) = F_X(\cdot | z_0)$ is the least favorable within the set of null distributions \mathbf{P}_0 in (2.3) satisfying stochastic dominance, a point first made by Lehmann (1951) and later reiterated by Hodges (1958); McFadden (1989), and Goldman and Kaplan (2018). We define the subset of continuous distributions in \mathbf{P}_0 that satisfy $F_Y(\cdot | z_0) = F_X(\cdot | z_0)$ as \mathbf{P}_0^* , and denote a generic element in \mathbf{P}_0^* by P^* . Although these papers studied the distribution of KS statistic

under P^* , they did not provide a formal proof that P^* determines the size of the test under the null hypothesis in (2.1). For completeness, we formally state this result in Lemma 1 below, and provide its proof.

Lemma 1. *Let $\mathbf{P}_0^* \subset \mathbf{P}_0$ be the subset of distributions P^* of the random variable S in (5.7), such that for a continuous CDF F , $S_j \sim F$ for all $j = 1, \dots, q$. Let $\phi(\cdot)$ be the test defined in (3.7). Then, for any $P^* \in \mathbf{P}_0^*$, we have*

$$\sup_{P \in \mathbf{P}_0} E_P[\phi(S)] = E_{P^*}[\phi(S)] = \bar{\alpha} ,$$

where $\bar{\alpha} \leq \alpha$ is defined in (4.3).

Lemma 1 shows that when $S \sim P^* \in \mathbf{P}_0^*$, our test is ‘exact’ in that it achieves the closest possible rejection rate to α , defined as $\bar{\alpha}$ in (4.3). In this case, we have

$$T(S) \stackrel{d}{=} T(U) \quad \text{where} \quad \{U_j \sim U[0, 1] : 1 \leq j \leq q\} \text{ are i.i.d.}$$

This demonstrates that the analytical (finite-sample) critical value $c_\alpha(q_y, q)$ for the one-sided KS test can be accurately approximated by simulating uniform random variables. However, when $S \sim P \in \mathbf{P}_0$ is such that $P\{S_i = S_j : i \neq j\} > 0$, the connection $T(S) \stackrel{d}{=} T(U)$ breaks down. In this case, $c_\alpha(q_y, q)$ is no longer the finite-sample *analytical* quantile of $T(S)$, but rather a valid upper bound.

When S is continuously distributed, the proposed test $\phi(S)$ is equivalent to a non-randomized permutation test. This connection establishes the validity of permutation tests for testing stochastic dominance. While Hodges (1958) and McFadden (1989) suggested that a permutation test could be used for this purpose, they did not provide a formal justification. To the best of our knowledge, our proof of this result is novel.

To formally define a permutation test, we introduce the following notation. Let \mathbf{G} denote the set of all permutations $\pi = (\pi(1), \dots, \pi(q))$ of $\{1, \dots, q\}$. The permuted values of S are given by

$$S^\pi = (S_{\pi(1)}, \dots, S_{\pi(q)}) .$$

The (non-randomized) permutation test is then defined as follows:

$$\begin{aligned} \phi^P(S) &:= I\{T(S) > c_\alpha^P(S)\} \\ c_\alpha^P(S) &:= \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I\{T(S^\pi) \leq x\} \geq 1 - \alpha \right\} . \end{aligned} \quad (5.8)$$

It follows from standard arguments (see, e.g., Lehmann and Romano (2005, Ch. 15)) that when S is invariant to permutations, i.e., $S \stackrel{d}{=} S^\pi$, the randomized version of the test $\phi^P(S)$ is exact in finite samples. However, under the null hypothesis in (2.1), we have that $S \not\stackrel{d}{=} S^\pi$ for some $P \in \mathbf{P}_0$, and so

invariance (or the so-called randomization hypothesis) fails. Therefore, the traditional finite-sample arguments for validity no longer apply. Alternative arguments that claim validity of permutation tests when invariance does not hold typically require $q \rightarrow \infty$; see [Chung and Romano \(2013\)](#); [Canay et al. \(2017\)](#), and [Bugni et al. \(2018\)](#), among others. In our current setting, where q is fixed, such arguments do not apply.

We contribute to this literature by demonstrating that, when S is continuously distributed, our test is equivalent to a non-randomized permutation test. The formal statement of this result follows.

Lemma 2. *For any random variable $\tilde{S} \in \mathbf{R}^q$ with $P\{\tilde{S}_i \neq \tilde{S}_j : i \neq j\} = 1$, we have*

$$P\{c_\alpha^p(\tilde{S}) = c_\alpha(q_y, q)\} = 1 . \quad (5.9)$$

Moreover, (5.9) no longer holds if \tilde{S} is such that $P\{\tilde{S}_i = \tilde{S}_j : i \neq j\} > 0$.

Lemma 2 shows that our data-independent critical value in (3.6) is equivalent to the critical value of a permutation test when the random variable \tilde{S} has no ties. Lemma 2 immediately implies when S_n in (3.3) is continuously distributed, we have

$$\phi^p(S_n) \stackrel{a.s.}{=} \phi(S_n) . \quad (5.10)$$

It follows from (5.10) and Theorem 2 that, when S is continuously distributed, a non-randomized permutation test controls the limiting rejection probability under the null hypothesis in (2.1). Importantly, this result holds even though invariance does not hold for all $P \in \mathbf{P}_0$.

Remark 8. *Lemmas 1 and 2 establish the validity of permutation tests for testing stochastic dominance in finite-sample settings. This result illustrates an instance where permutation tests can provide finite-sample valid inference even in settings where the randomization hypothesis does not hold. The only similar result we are aware of is that of [Caughey et al. \(2023\)](#), who consider a design-based framework with the null hypothesis $\tau_i \leq 0$, where τ_i represents a unit-level treatment effect. Like our work, their result is valid under the condition that the random variables are continuously distributed or that a random tie-breaking rule is applied to handle ties. ■*

The results in Lemmas 1 and 2 reveal interesting and novel connections between our test and classical arguments involving finite-sample critical values and permutation tests. However, these results critically depend on the random variable S being continuously distributed and do not extend to cases where S is discretely distributed.

When S is discrete and ties occur with positive probability, i.e., $P\{S_i = S_j\} > 0 : i \neq j$, the finite-sample distribution of the KS test statistic $T(S)$ depends on the number and location of these ties. A natural approach to handle ties is to redefine the test statistic to randomly break them, effectively making the test a randomized one. While this would allow us to establish an analog of

Lemma 2 for discrete data, we do not pursue such an extension, as randomized tests are rarely used in practice. Despite our best efforts, we were unable to demonstrate that a permutation test could control size under the null hypothesis of stochastic dominance in (2.1) when S is discrete, without relying on random tie-breaking rules.

6 Simulations

In this section, we evaluate the finite-sample performance of the test in (3.7) for $L = 1$ or the test proposed in Section C.3 for $L > 1$ through a simulation study. We present a variety of data-generating processes to illustrate both the strengths and potential limitations of our test.

We consider seven distinct designs with four cases, (a) to (d), in each design as follows:

- **Case (a):** the null hypothesis holds with equality and $L = 1$
- **Case (b):** the null hypothesis holds with strict inequality and $L = 1$
- **Case (c):** the null hypothesis holds with equality and $L = 2$
- **Case (d):** the null hypothesis is violated and $L = 1$

The first three designs are based on the following location–scale model:

$$Y = \mu_Y(Z) + \sigma_Y(Z)U \quad \text{and} \quad X = \mu_X(Z) + \sigma_X(Z)V, \quad (6.1)$$

where U and V are random variables with specified distributions, and the conditioning variable Z follows a non-negative Beta(2, 2) distribution. Under this location–scale model, whenever $\sigma_Y(z_\ell) = \sigma_X(z_\ell)$, the null hypothesis in (2.1) holds as long as

$$\mu_Y(z_\ell) \geq \mu_X(z_\ell). \quad (6.2)$$

Design 1 satisfies (6.2) for all $z \in (0, 1)$. Design 2 satisfies (6.2) at $z_\ell = 0.5$, but violates the inequality for $z > 0.5$, which may affect the performance of our test in finite samples due to its reliance on induced order statistics. Design 3 is such that U and V are $U[0, 1]$, which guarantees that (2.1) holds even when $\sigma_Y(z_\ell) < \sigma_X(z_\ell)$, provided $\mu_Y(z_\ell) - \mu_X(z_\ell) \geq \sigma_X(z_\ell) - \sigma_Y(z_\ell)$.

Design 4 is a slight variation of the location-scale model to accommodate a regression discontinuity design (RDD):

$$Y = \mu_Y(Z) + \sigma_Y(Z)U \quad \text{if } Z \geq 0, \quad \text{and} \quad X = \mu_X(Z) + \sigma_X(Z)V \quad \text{if } Z < 0. \quad (6.3)$$

Following Shen and Zhang (2016), we consider the case where U and V are independent $N(0, 1)$ random variables, $Z \sim 2\text{Beta}(2, 2) - 1$, and both $\mu_Y(Z)$ and $\mu_X(Z)$ are defined in terms of the

function

$$\mu(z) = 0.61 - 0.02z + 0.06z^2 + 0.17z^3 . \quad (6.4)$$

A key feature shared by Designs 1 through 4 is that both X and Y are continuously distributed; a feature not shared by the next three designs.

Design 5 is such that U and V follow log-normal distributions with the bottom 20% censored. This setup reflects features of wage distributions, which often exhibit a point mass at the minimum wage. Design 6 defines conditional probabilities $P\{X = k|Z\}$ and $P\{Y = k|Z\}$ as

$$P\{X = k|Z\} = \frac{e^{\theta_k^x(\frac{3}{2}-Z)}}{\sum_{j=1}^3 e^{\theta_j^x(\frac{3}{2}-Z)}} \quad \text{and} \quad P\{Y = k|Z\} = \frac{e^{\theta_k^y(\frac{3}{2}-Z)}}{\sum_{j=1}^3 e^{\theta_j^y(\frac{3}{2}-Z)}} \quad \text{for } k = 1, 2, 3,$$

where, for $\mu_Y(z) \in [-1, 1]$,

$$\theta_1^y = \theta_1^x - \mu_Y(z), \quad \theta_2^y = \theta_2^x + \mu_Y(z), \quad \text{and} \quad \theta_3^y = \theta_3^x .$$

The parameter $\theta_k^x = (-0.5, -1.5, -2)$ controls the baseline log-odds of each category for X , while the factor $(3/2 - Z)$ introduces a monotonic dependence on Z . Finally, Design 7 considers

$$X|Z \sim B\left([25Z], \frac{1}{2}\right) \quad \text{and} \quad Y|Z \sim B\left([25Z] + \mu_Y(Z), \frac{1}{2}\right) ,$$

where $B(\cdot, \cdot)$ denotes a binomial distribution and $[x]$ represents the nearest integer to x .

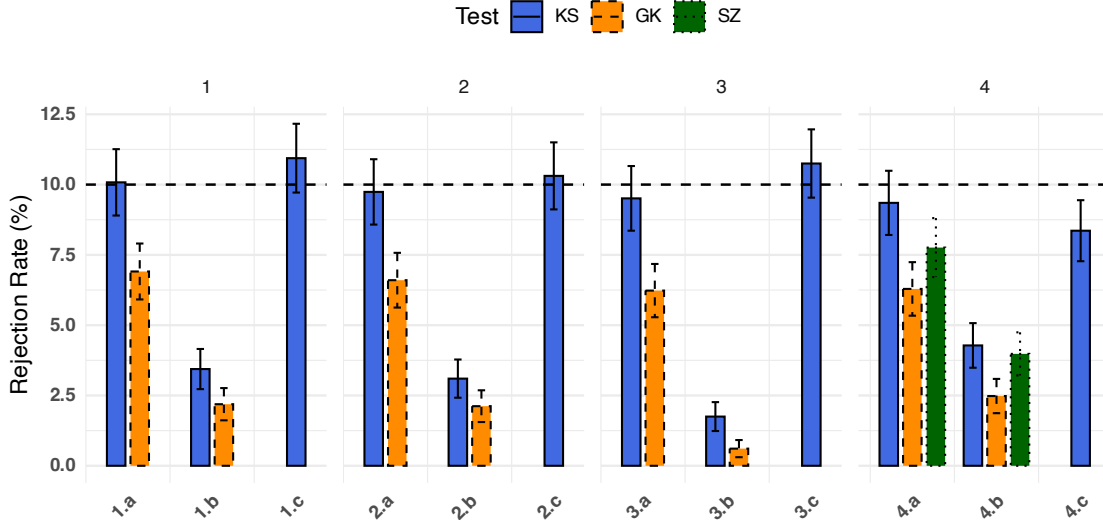


Figure 1: Rejection rates under H_0 for Designs 1-4 in cases (a)-(c): $n = 1,000$, $\alpha = 10\%$, $MC = 10,000$.

The parameter values for all Designs are reported in Appendix C.5, where we also report the mean values of q_y^* and q_x^* across simulations.

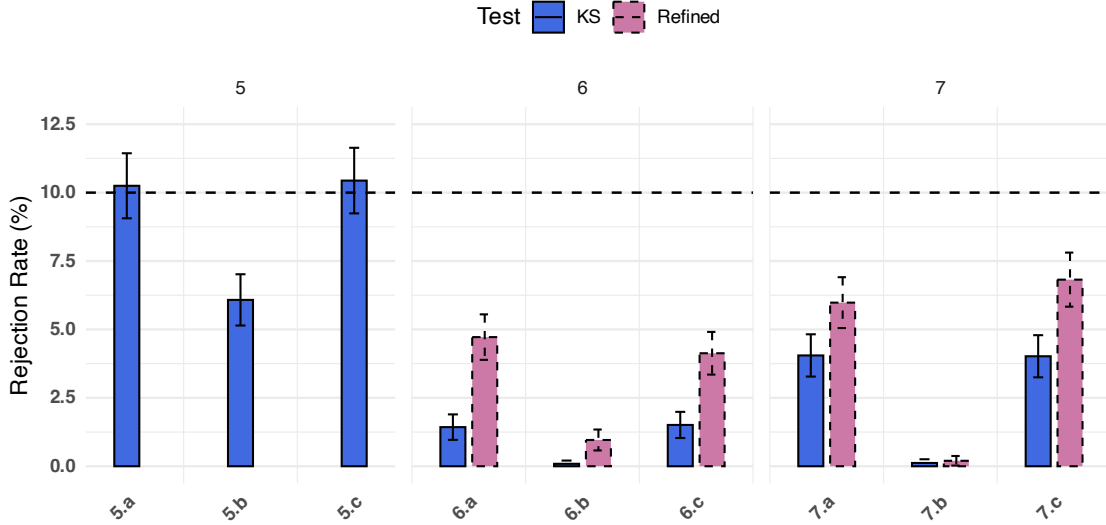


Figure 2: Rejection rates under H_0 for Designs 5-7 in cases (a)-(c): $n = 1,000$, $\alpha = 10\%$, $MC = 10,000$.

We report results for sample size $n = 1,000$ and nominal level $\alpha = 10\%$, based on 10,000 Monte Carlo simulations to test the null hypothesis in (2.1). To implement the test $\phi(\cdot)$ in (3.7), denoted by ‘KS’ in the figures, we select the tuning parameters (q_y, q_x) using the data-dependent rules in (5.1) and (5.2). For Designs 1-3 with $L = 1$, we compare our test with the one proposed in Goldman and Kaplan (2018, GK). For the RDD Design 7 with $L = 1$, we compare our test with the method in Shen and Zhang (2016, SZ).² For the mixed and discrete designs, we present results for our test in (3.7) as well as for its refined version in (5.5).

Figure 1 reports rejection probabilities under the null hypothesis for the continuously distributed designs. When the data-generating process satisfies the null in (2.1) with equality (case (a)), the KS test’s rejection probabilities closely align with the nominal level and consistently outperform the GK and SZ tests. When the null holds with strict inequality (case (b)), rejection rates fall below α , consistent with our critical value serving as a valid upper bound for the true quantile. Case (c), where $L = 2$, shows behavior similar to case (a), where $L = 1$. Overall, the KS test exhibits excellent size control.

Figure 2 reports rejection probabilities under the null hypothesis for the mixed and discretely distributed designs. When the data-generating process satisfies the null in (2.1) with equality, the KS test we propose in (3.7) works well when the data is mixed, but it is conservative when the data is discrete with few support points. Designs 6 and 7 illustrate that the refined critical value $c_\alpha^r(q, q_y)$ in (5.4) offers a more accurate approximation of the true quantile of the KS test statistic, though it may still be somewhat conservative.

²SZ is implemented using the authors’ rule of thumb, whereas for GK we use our own rule of thumb, as the original paper does not provide guidance on how to select their tuning parameter. Note that neither of these tests is defined when $L > 1$.

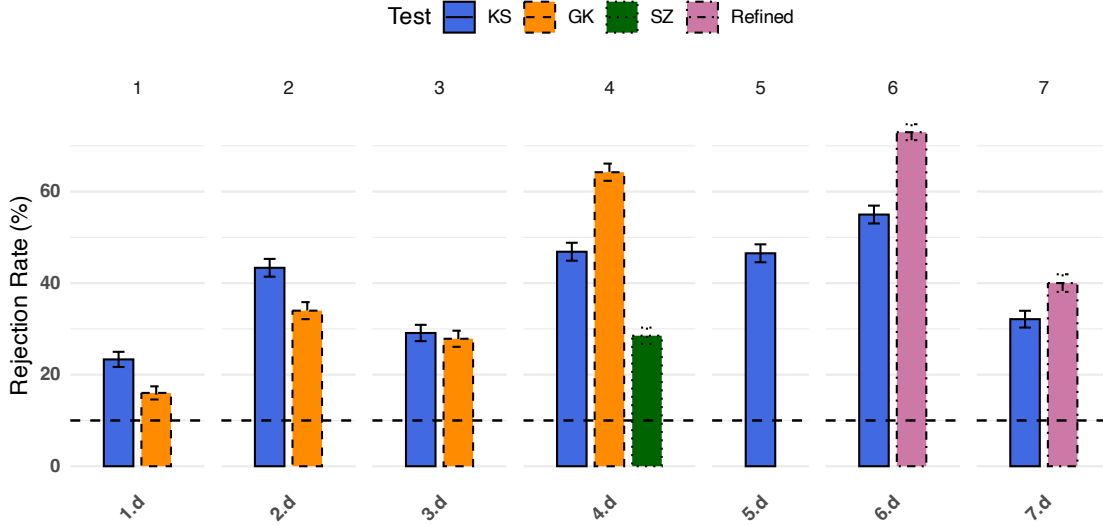


Figure 3: Rejection rates under H_1 for Designs 1-7: $n = 1,000$, $\alpha = 10\%$, $MC = 10,000$.

Figure 3 reports rejection probabilities under the alternative hypothesis for all designs. The power of the KS test is similar to that of GK and could be above, below, or roughly the same. The power of the KS test is much higher than that of SZ for this design. This is worth noting given that the KS test uses about 90 observations (for both q_y and q_x) while the SZ test uses 276 effective observations on each side of the threshold. The power results also demonstrate that the refined critical value $c_\alpha^r(q, q_y)$ in (5.4) enhances power in discrete cases.

7 Concluding remarks

This paper introduces a novel test for conditional stochastic dominance (CSD) at target points, offering a flexible, nonparametric approach that avoids kernel smoothing while ensuring computational efficiency. By leveraging induced order statistics, our method constructs empirical CDFs using observations closest to the target conditioning point. We establish the asymptotic properties of our test, demonstrating its validity under weak regularity conditions, and derive a critical value that eliminates the need for resampling techniques such as the bootstrap. Additionally, we extend our framework to better handle discrete data, proposing a refined critical value that enhances the power of the test with minimal additional information. Monte Carlo simulations align with our theoretical results and suggest that our test performs well in finite samples, making our test readily applicable to empirical research in economics, finance, and public policy.

An important feature of our test is its simplicity. Once the key tuning parameters are computed, the test only requires a standard test statistic with a deterministic critical value, without the need for kernels, local polynomials, bias correction, or bandwidth selection. Furthermore, our test

admits a clear interpretation in the limit experiment, which allows us to connect it with classical analytical critical values and permutation-based tests. In this sense, our findings contribute to the broader literature on stochastic dominance testing by refining conditional inference methods and establishing new links between permutation-based and rank-based approaches. One open question we did not address in this paper concerns the validity of permutation-based tests for the hypothesis of stochastic dominance when both random variables, Y and X , are discrete. Despite attempts to formalize this result, we were unable to prove or disprove it. Extensive Monte Carlo simulations (not reported here) suggest that the test may be valid, and this is an area we plan to explore further.

Appendix A Proof of the main results

A.1 Proof of Theorem 2

By Theorem 1,

$$S_n \xrightarrow{d} S = (S_1, \dots, S_q)$$

where the elements of S are independent, and $S_j \sim F_Y(\cdot|z_0)$ for $j = 1, \dots, q_y$ and $S_j \sim F_X(\cdot|z_0)$ for $j = q_y + 1, \dots, q$. By the almost-sure representation theorem, we have a sequence of random vectors $\{\tilde{S}_n : 1 \leq n \leq \infty\}$ and a random vector \tilde{S} defined on a common probability space $(\Omega, \mathcal{A}, \tilde{P})$ such that

$$\tilde{S}_n \stackrel{d}{=} S_n, \quad \tilde{S} \stackrel{d}{=} S, \quad \text{and} \quad \tilde{S}_n \xrightarrow{a.s.} \tilde{S}.$$

Let $R(s)$ denote the rank of s , which maps s to a permutation of $\{1, 2, \dots, q\}$. Define the event that the rank of the two vectors coincides as follows,

$$F_n = \{R(\tilde{S}_n) = R(\tilde{S})\}.$$

To reach the conclusion, it suffices to show that

$$\tilde{P}\{F_n\} \rightarrow 1. \tag{A-1}$$

To see this, consider the following argument,

$$E_P[\phi(S_n)] \stackrel{(1)}{=} E_{\tilde{P}}[\phi(\tilde{S}_n)] \stackrel{(2)}{=} E_{\tilde{P}}[\phi(\tilde{S})I\{F_n\} + \phi(\tilde{S}_n)I\{F_n^c\}] \stackrel{(3)}{\rightarrow} E_{\tilde{P}}[\phi(\tilde{S})] \stackrel{(4)}{\leq} \alpha,$$

where (1) holds by $\tilde{S}_n \stackrel{d}{=} S_n$, (2) by the fact that T is invariant to rank-preserving transformations, (3) by (A-1) and $\phi(\cdot) \in \{0, 1\}$, and (4) by $P \in \mathbf{P}_0$ and Theorem 4.

We devote the remainder of the proof to establishing (A-1). Define $\mathcal{D} = \mathcal{D}_X(z_0) \cup \mathcal{D}_Y(z_0)$, where $\mathcal{D}_X(z_0)$ and $\mathcal{D}_Y(z_0)$ denote the sets of discontinuity points as specified in Assumption 3. For any $\varepsilon > 0$, let

$$E_1(\varepsilon) := \{ \{|\tilde{S}_i - \tilde{S}_j| > \varepsilon\} \cup \{\tilde{S}_i = \tilde{S}_j \in \mathcal{D}\} : i \neq j = 1, \dots, q \},$$

$$E_{n,2}(\varepsilon) := \{ |\tilde{S}_{n,k} - \tilde{S}_k| < \varepsilon/2 : k = 1, \dots, q \},$$

$$E_{n,3} := \{ \{ \tilde{S}_k \in \mathcal{D} \} \subseteq \{ \tilde{S}_k = \tilde{S}_{n,k} \} : k = 1, \dots, q \} .$$

Observe that

$$E_1(\varepsilon) \cap E_{n,2}(\varepsilon) \cap E_{n,3} \subseteq F_n. \quad (\text{A-2})$$

To establish this, consider the following argument. For any $i, j = 1, \dots, q$, there are three possible cases: (i) $\tilde{S}_i < \tilde{S}_j$, (ii) $\tilde{S}_i > \tilde{S}_j$, or (iii) $\tilde{S}_i = \tilde{S}_j$. First, consider case (i), where $\tilde{S}_i < \tilde{S}_j$. Under $E_1(\varepsilon)$, this implies $\tilde{S}_i < \tilde{S}_j - \varepsilon$. Under $E_{n,2}(\varepsilon)$, we have $\tilde{S}_{n,i} - \varepsilon/2 < \tilde{S}_i$ and $\tilde{S}_j < \tilde{S}_{n,j} + \varepsilon/2$. Combining these inequalities yields $\tilde{S}_{n,i} < \tilde{S}_{n,j}$, as required. Case (ii) follows identically by reversing the roles of i and j . Finally, consider case (iii), where $\tilde{S}_i = \tilde{S}_j$. Under $E_1(\varepsilon)$, this implies $\tilde{S}_i = \tilde{S}_j \in \mathcal{D}$. By $E_{n,3}$, it follows that $\tilde{S}_{n,i} = \tilde{S}_{n,j} \in \mathcal{D}$. Since this argument holds for all $i, j = 1, \dots, q$, we conclude that F_n follows, as desired.

By (A-2), (A-1) follows that there exists $\varepsilon > 0$ such that

$$\tilde{P}\{E_1(\varepsilon) \cap E_{n,2}(\varepsilon) \cap E_{n,3}\} \rightarrow 1. \quad (\text{A-3})$$

For arbitrary $\delta > 0$, (A-3) follows from finding $\varepsilon = \varepsilon(\delta) > 0$ and $N(\delta)$ such that $\tilde{P}\{E_1(\varepsilon) \cap E_{n,2}(\varepsilon) \cap E_{n,3}\} \geq 1 - \delta$ for all $n \geq N(\delta)$. Let $\varepsilon_1 = \inf\{\|\tilde{d} - d\|/2 : d < \tilde{d} \in \mathcal{D}\} > 0$. By Lemma 4, $\exists \varepsilon_2 > 0$ such that, for $i \neq j = 1, \dots, q$,

$$\begin{aligned} \tilde{P}\{\{|\tilde{S}_i - \tilde{S}_j| < \varepsilon_2\} \cap \{\tilde{S}_i, \tilde{S}_j \in \mathcal{D}^c\}\} &< \delta/(9q(q-1)) , \\ \tilde{P}\{\cup_{d \in \mathcal{D}}\{|\tilde{S}_i - d| < \varepsilon_2\} \cap \{\tilde{S}_i \in \mathcal{D}^c\}\} &< \delta/(9q(q-1)) . \end{aligned}$$

Finally, set $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\} > 0$ for the remainder of the proof. By elementary arguments, it suffices to show that: (i) $\tilde{P}\{E_1(\varepsilon)^c\} \leq \delta/3$, (ii) $\exists N_2(\delta) \in \mathbf{N}$ s.t. $\tilde{P}\{E_{n,2}(\varepsilon)^c\} \leq \delta/3$ for all $n \geq N_2(\delta)$, and (iii) $\exists N_3(\delta) \in \mathbf{N}$ s.t. $\tilde{P}\{E_{n,3}(\varepsilon)^c\} \leq \delta/3$ for all $n \geq N_3(\delta)$. We divide the rest of the proof into three results.

First, we show that $\tilde{P}\{E_1(\varepsilon)^c\} \leq \delta/3$. To this end, pick $i \neq j = 1, \dots, q$ arbitrarily. Note that

$$\tilde{P}\{\{|\tilde{S}_i - \tilde{S}_j| < \varepsilon\} \cap \{\tilde{S}_i = \tilde{S}_j \in \mathcal{D}\}^c\} \stackrel{(1)}{\leq} \delta/(3q(q-1)) ,$$

where (1) holds by $i \neq j = 1, \dots, q$, \tilde{S}_i and \tilde{S}_j being identically distributed, and $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. From here, we conclude that

$$\tilde{P}\{E_1(\varepsilon)^c\} \leq \sum_{i \neq j} \tilde{P}\{\{|\tilde{S}_i - \tilde{S}_j| < \varepsilon\} \cap \{\tilde{S}_i = \tilde{S}_j \in \mathcal{D}\}^c\} \leq \delta/3 ,$$

as desired. Second, we show that $\exists N_2(\delta) \in \mathbf{N}$ such that $\tilde{P}\{E_{n,2}(\varepsilon)^c\} \leq \delta/3$ for all $n \geq N_2(\delta)$. To see this, note that

$$\tilde{P}\{E_{n,2}(\varepsilon)^c\} \leq \sum_{k=1}^q P\{|\tilde{S}_{n,k} - \tilde{S}_k| > \varepsilon/2\} . \quad (\text{A-4})$$

By $\tilde{S}_n \xrightarrow{a.s.} \tilde{S}$, $\exists N_2(\delta)$ such that the right-hand side is less than $\delta/3$, as desired. Finally, we show that $\exists N_3(\delta) \in \mathbf{N}$ such that $\tilde{P}\{E_{n,3}(\varepsilon)^c\} \leq \delta/3$ for all $n \geq N_3(\delta)$. To see this, note that

$$\begin{aligned} \tilde{P}\{E_{n,3}(\varepsilon)^c\} &= \tilde{P}\{\exists k = 1, \dots, q : \{ \{ \tilde{S}_k \in \mathcal{D} \} \subseteq \{ \tilde{S}_k = \tilde{S}_{n,k} \} \}^c\} \\ &\leq \sum_{k=1}^q \tilde{P}\{\{ \tilde{S}_k \in \mathcal{D} \} \cap \{ \tilde{S}_k \neq \tilde{S}_{n,k} \} \} . \end{aligned}$$

By Lemma 3, $\exists N_3(\delta)$ such that the right-hand side is less than $\delta/3$, as desired. This completes the proof of (A-1) and the theorem.

A.2 Proof of Lemma 1

Note that

$$E_{P^*}[\phi(S)] \stackrel{(1)}{\leq} \sup_{P \in \mathbf{P}_0} E_P[\phi(S)] \stackrel{(2)}{\leq} \bar{\alpha},$$

where (1) holds by $P^* \in \mathbf{P}_0$ and (2) by Theorem 4. To complete the proof, it suffices to show that $E_{P^*}[\phi(S)] = \bar{\alpha}$. To this end, consider the following argument:

$$\begin{aligned} & E_{P^*}[\phi(S)] \\ & \stackrel{(1)}{=} P^* \left\{ \sup_{t \in \mathbf{R}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{S_j \leq t\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{S_j \leq t\} \right) > c_\alpha(q_y, q) \right\} \\ & \stackrel{(2)}{=} P^* \left\{ \sup_{t \in \mathbf{R}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq F(t)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq F(t)\} \right) > c_\alpha(q_y, q) \right\} \\ & \stackrel{(3)}{=} P^* \left\{ \sup_{u \in (0,1)} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\} \right) > c_\alpha(q_y, q) \right\} \\ & \stackrel{(4)}{=} \bar{\alpha}, \end{aligned} \tag{A-5}$$

where (1) holds by (3.4), (2) holds by Pollard (2002, Eq. 36) and the same arguments used in the proof of Theorem 4, (3) follows from the continuity of F guaranteeing that

$$\sup_{t \in \mathbf{R}} \Delta(F(t)) = \sup_{u \in (0,1)} \Delta(u)$$

for $\Delta(u) := \frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\}$ as defined in (3.5), and (4) by definition of $\bar{\alpha}$ in (4.3).

A.3 Proof of Lemma 2

Let $Q := \{Q_1, \dots, Q_q\} = \{1, 2, \dots, q\}$ and denote by $Q^\pi := \{Q_{\pi(1)}, Q_{\pi(2)}, \dots, Q_{\pi(q)}\}$ the permutation $\pi = (\pi(1), \pi(2), \dots, \pi(q))$ of Q . Let $R(s)$ denote the rank of s , which maps s to a permutation of Q . Since the KS statistic $T(\cdot)$ in (3.4) is a rank statistic, it follows that for any s

$$T(s) = T^*(R(s)), \tag{A-6}$$

where T^* is a known function; see Hajek et al. (1999, page 99). That is, the KS test statistic depends on S only through $R(S)$. Define

$$c_\alpha^P := \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I\{T(Q^\pi) \leq x\} \geq 1 - \alpha \right\}. \tag{A-7}$$

We divide the rest of the argument into four steps.

Step 1. For any $s \in \mathbf{R}^q$ with $s_i \neq s_j$ for $i \neq j$, and $c_\alpha^p(s)$ as in (5.8),

$$c_\alpha^p(s) = c_\alpha^p.$$

To establish this, consider the following derivation,

$$\begin{aligned} c_\alpha^p(s) &= \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T(s^\pi) \leq x\} \geq 1 - \alpha \right\} \\ &\stackrel{(1)}{=} \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(R(s^\pi)) \leq x\} \geq 1 - \alpha \right\} \\ &\stackrel{(2)}{=} \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*((Q^{\bar{\pi}})^\pi) \leq x\} \geq 1 - \alpha \right\} \\ &\stackrel{(3)}{=} \inf_{x \in \mathbf{R}} \left\{ \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(Q^\pi) \leq x\} \geq 1 - \alpha \right\} \\ &\stackrel{(4)}{=} c_\alpha^p. \end{aligned} \tag{A-8}$$

Here, (1) follows by (A-6), (2) follows since $s_i \neq s_j$ for $i \neq j$ implies that $R(s^\pi) = (R(s))^\pi$ and $R(s) = Q^{\bar{\pi}(s)}$ for some $\bar{\pi}(s) \in \mathbf{G}$, (3) by $\mathbf{G} = \tilde{\mathbf{G}} := \{\pi \circ \bar{\pi}(s) : \pi \in \mathbf{G}\}$, which follows from the fact that \mathbf{G} is a group, and (4) by (A-7).

Step 2. $c_\alpha^p = c_\alpha(q_y, q)$, where $c_\alpha(q_y, q)$ is defined in (3.6).

Let $\{U_i : 1 \leq i \leq q\}$ be i.i.d. with $U_i \sim U(0, 1)$ and let $\hat{\pi}$ be a uniformly chosen permutation from \mathbf{G} , independent of U . Note that $c_\alpha(q_y, q)$ is the $(1 - \alpha)$ -quantile of $T(U)$ and, by (A-7), c_α^p is the $(1 - \alpha)$ -quantile of the CDF $\frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(Q^\pi) \leq x\}$. The desired result then follows from noting that $\frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(Q^\pi) \leq x\}$ is the CDF of $T(U)$, as we show next.

Let $E := \{U_i \neq U_j \text{ for } i \neq j\}$. For any $x \in \mathbf{R}$, our desired result follows from this derivation:

$$\begin{aligned} P \{T(U) \leq x\} &\stackrel{(1)}{=} P \{T(U^{\hat{\pi}}) \leq x\} \\ &\stackrel{(2)}{=} P \{ \{T(U^{\hat{\pi}}) \leq x\} \cap E \} \\ &\stackrel{(3)}{=} \int_{s \in E} \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T(u^\pi) \leq x\} dP_U^*(u) \\ &\stackrel{(4)}{=} \int_{u \in E} \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(Q^\pi) \leq x\} dP_U^*(u) \\ &\stackrel{(5)}{=} \frac{1}{|\mathbf{G}|} \sum_{\pi \in \mathbf{G}} I \{T^*(Q^\pi) \leq x\}. \end{aligned}$$

Here, (1) holds by $U \stackrel{d}{=} U^{\hat{\pi}}$, (2) and (5) by $P\{E\} = 1$, (3) by $\hat{\pi} \perp U$ and $\hat{\pi}$ uniformly chosen in \mathbf{G} , and (4) by repeating the arguments used to derive (A-8).

Step 3. By Step 1, $\{S_i \neq S_j \text{ for } i \neq j\} \subseteq \{c_\alpha^p(S) = c_\alpha^p\}$, and so $P \{c_\alpha^p(S) = c_\alpha^p\} = 1$ holds by our assumption. By Step 2, $c_\alpha^p = c_\alpha(q_y, q)$. The desired result follows from combining these points.

Step 4. To show the last statement, consider $S = \{1 : 1 \leq j \leq q\}$. Then, $T(S^\pi) = T(S) = 0$ for all $\pi \in \mathbf{G}$, and so $c_\alpha^p(S) = 0$. On the other hand, Step 2 implies $c_\alpha^p = c_\alpha(q_y, q)$, which are positive for typical values of (q_y, q, α) . For example, $q_y = 1$, $q = 2$, and $\alpha = 0.1$ yield $c_\alpha^p = c_\alpha(q_y, q) = 0.5$.

Appendix B Supporting technical results

For any $\varepsilon > 0$, we use $o_\varepsilon(1)$ to denote an expression that converges to zero as $\varepsilon \rightarrow 0$. Analogously, for any $n \in \mathbf{N}$, we use $o_n(1)$ to denote an expression that converges to zero as $n \rightarrow \infty$.

B.1 Auxiliary theorems

Theorem 3. *Let Assumptions 1 and 2 hold with $\mathcal{Z} = \{z_1, z_2, \dots, z_L\}$, and let S_n^ℓ be defined as in (C-11). Then,*

$$(S_n^{1'}, \dots, S_n^{L'}) \xrightarrow{d} (S^{1'}, \dots, S^{L'}) , \quad (\text{B-1})$$

where, for any $(s_1^\ell, \dots, s_{q_y}^\ell, s_{q_y+1}^\ell, \dots, s_{q^\ell}^\ell) \in \mathbf{R}^{q^\ell}$ for each $\ell = 1, \dots, L$ with $q^\ell = q_y^\ell + q_x^\ell$, $(S^{1'}, \dots, S^{L'})$ has the following distribution:

$$P \left\{ \bigcap_{\ell=1}^L \bigcap_{i=1}^{q^\ell} \{S_i^\ell \leq s_i^\ell\} \right\} = \prod_{\ell=1}^L \prod_{i=1}^{q_y^\ell} F_Y(s_i^\ell | z_\ell) \prod_{j=1}^{q_x^\ell} F_X(s_{j+q_y^\ell}^\ell | z_\ell) ,$$

Proof. For each $\ell = 1, \dots, L$, let M_y^ℓ denote the subset of the indices $i = 1, \dots, n_y$ corresponding to the q_y^ℓ first g -order statistics $(Z_{\ell,(1)}, \dots, Z_{\ell,(q_y^\ell)})$, and let M_x^ℓ denote the subset of the indices $j = 1, \dots, n_x$ corresponding to the q_x^ℓ first g -order statistics $(Z_{\ell,(1)}, \dots, Z_{\ell,(q_x^\ell)})$. Let E_n denote the following event:

$$E_n = E_{n_y, y} \cap E_{n_x, x} \text{ where } E_{n_y, y} := \left\{ \bigcap_{\ell=1}^L M_y^\ell = \emptyset \right\} \text{ and } E_{n_x, x} := \left\{ \bigcap_{\ell=1}^L M_x^\ell = \emptyset \right\} .$$

In words, E_n means that the subsets of the data used in each of the L tests have no observations in common. We begin by showing that

$$P\{E_n\} \rightarrow 1 . \quad (\text{B-2})$$

Since the two sample are independent, $P\{E_n\} = P\{E_{n_y, y}\}P\{E_{n_x, x}\}$ and so we only prove $P\{E_{n_y, y}\} \rightarrow 1$ as the other case is analogous.

Let $\varepsilon := \frac{1}{2} \min \{|z_\ell - z_{\ell'}| : \ell \neq \ell', \ell, \ell' = 1, \dots, L\} > 0$. For each $\ell = 1, \dots, L$, let $B_{\ell, y} = \sum_{i=1}^{n_y} I\{|Z_i - z_\ell| \leq \varepsilon\}$, and note that

$$\bigcap_{\ell=1}^L \{B_{\ell, y} \geq q_y^\ell\} \subseteq E_n .$$

From here, we have that (B-2) follows if we show that

$$P\{B_{\ell, y} < q_y^\ell\} \rightarrow 0 \text{ for all } \ell = 1, \dots, L . \quad (\text{B-3})$$

By Assumption 1 and $\max_{\ell=1,\dots,L} q^\ell$ being bounded, $\exists N(\varepsilon)$ s.t. for all $n_y \geq N(\varepsilon)$,

$$0 < \frac{1}{2} P\{|Z_i - z_\ell| \leq \varepsilon\} \leq P\{|Z_i - z_\ell| \leq \varepsilon\} - \max_{\ell=1,\dots,L} \frac{q_y^\ell}{n_y}. \quad (\text{B-4})$$

Then, for all $n_y \geq N(\varepsilon)$, we have

$$\begin{aligned} P\{B_{\ell,y} < q_y^\ell\} &= P\{B_{\ell,y}/n_y - P\{|Z_i - z_\ell| \leq \varepsilon\} < q_y^\ell/n_y - P\{|Z_i - z_\ell| \leq \varepsilon\}\} \\ &\stackrel{(1)}{\geq} P\{B_{\ell,y}/n_y - P\{|Z_i - z_\ell| \leq \varepsilon\} > \frac{1}{2} P\{|Z_i - z_\ell| \leq \varepsilon\}\} \stackrel{(2)}{\rightarrow} 0, \end{aligned}$$

as desired, where (1) holds by (B-4) and (2) holds by the LLN as $n_y \rightarrow \infty$, as $B_{\ell,y} = \sum_{i=1}^{n_y} I\{|Z_i - z_\ell| \leq \varepsilon\} \sim Bi(n_y, P\{|Z_i - z_\ell| \leq \varepsilon\})$.

We are now ready to prove the desired result. For any $(s_1^\ell, \dots, s_{q_y^\ell}^\ell, s_{q_y^\ell+1}^\ell, \dots, s_{q_x^\ell}^\ell) \in \mathbf{R}^{q^\ell}$ for each $\ell = 1, \dots, L$ with $q^\ell = q_y^\ell + q_x^\ell$, we have

$$\begin{aligned} &P\left\{\bigcap_{\ell=1}^L \bigcap_{i=1}^{q^\ell} \{S_{n,i}^\ell \leq s_i^\ell\}\right\} \\ &\stackrel{(1)}{=} E\left[P\left\{\bigcap_{\ell=1}^L \left\{\bigcap_{i=1}^{q_y^\ell} \{Y_{n,[i]}^\ell \leq s_i^\ell\} \cap \bigcap_{j=1}^{q_x^\ell} \{X_{n,[j]}^\ell \leq s_{j+q_y^\ell}^\ell\}\right\} \middle| \mathcal{A}\right\}\right] \\ &\stackrel{(2)}{=} E\left[\prod_{\ell=1}^L P\left\{\left\{\bigcap_{i=1}^{q_y^\ell} \{Y_{n,[i]}^\ell \leq s_i^\ell\} \cap \bigcap_{j=1}^{q_x^\ell} \{X_{n,[j]}^\ell \leq s_{j+q_y^\ell}^\ell\}\right\} \middle| \mathcal{A}\right\} I\{E_n\}\right] + o_n(1) \\ &\stackrel{(3)}{=} E\left[\prod_{\ell=1}^L P\left\{\left\{\bigcap_{i=1}^{q_y^\ell} \{Y_{n,[i]}^\ell \leq s_i^\ell\} \cap \bigcap_{j=1}^{q_x^\ell} \{X_{n,[j]}^\ell \leq s_{j+q_y^\ell}^\ell\}\right\} \middle| \mathcal{A}\right\}\right] + o_n(1) \\ &\stackrel{(4)}{=} E\left[\prod_{\ell=1}^L \prod_{i=1}^{q_y^\ell} F_Y(s_i^\ell | Z_{n_y,(i)}^\ell) \prod_{j=1}^{q_x^\ell} F_X(s_{j+q_y^\ell}^\ell | Z_{n_x,(j)}^\ell)\right], \end{aligned} \quad (\text{B-5})$$

where (1) holds by the LIE with \mathcal{A} equal to the sigma-algebra generated by the Z observations from both samples, (2) by (B-2) and the fact E_n implies that the subsets of the data used in each of the L tests have no observations in common, so they are independent conditional on \mathcal{A} , (3) by (B-2), and (4) by repeating the arguments in the proof of Theorem 1.

Next, we show that

$$Z_{n_y,(i)}^\ell \xrightarrow{P} z_\ell \text{ for all } i = 1, \dots, q_y \text{ and } \ell = 1, \dots, L, \quad (\text{B-6})$$

$$Z_{n_x,(j)}^\ell \xrightarrow{P} z_\ell \text{ for all } j = 1, \dots, q_x \text{ and } \ell = 1, \dots, L. \quad (\text{B-7})$$

We only show (B-6), as (B-7) can be shown analogously. To this end, fix $\ell = 1, \dots, L$ arbitrarily. We prove the result by complete induction on $i = 1, \dots, q_y$. Take $i = 1$ and fix $\varepsilon > 0$ arbitrarily. Then,

$$\begin{aligned} P\{|Z_{n_y,(1)}^\ell - z_\ell| < \varepsilon\} &= P\{\text{at least 1 of } \{Z_i : 1 \leq i \leq n_y\} \text{ is s.t. } \{|Z_i - z_\ell| < \varepsilon\}\} \\ &\stackrel{(1)}{=} \sum_{u=1}^{n_y} \binom{n_y}{u} P\{|Z - z_\ell| < \varepsilon\}^u [1 - P\{|Z - z_\ell| < \varepsilon\}]^{n_y-u} \end{aligned}$$

$$\stackrel{(2)}{=} 1 - P\{|Z - z_\ell| \geq \varepsilon\}^{n_y} \stackrel{(3)}{\rightarrow} 1, \quad (\text{B-8})$$

as desired, where (1) holds by the fact that $\{Z_i : 1 \leq i \leq n_y\}$ are identically distributed, (2) by the Binomial Theorem, and (3) by Assumption 1. For the inductive step, we assume $Z_{n_y, (j)}^\ell - z_\ell = o_p(1)$ for $j \in \{1, \dots, q_y - 1\}$, and prove that $Z_{n_y, (j+1)}^\ell - z_\ell = o_p(1)$. For this, consider the following derivation,

$$\begin{aligned} & P\{|Z_{n_y, (j)}^\ell - z_\ell| < \varepsilon\} \\ &= P\{\text{at least } j \text{ of the } \{Z_i : 1 \leq i \leq n_y\} \text{ are s.t. } \{|Z_i - z_\ell| < \varepsilon\}\} \\ &\stackrel{(1)}{=} \sum_{u=j}^{n_y} \binom{n_y}{u} P\{|Z - z_\ell| < \varepsilon\}^u P\{|Z - z_\ell| \geq \varepsilon\}^{n_y-u} \\ &\stackrel{(2)}{=} P\{|Z_{n_y, (j+1)}^\ell - z_\ell| < \varepsilon\} + \binom{n_y}{j} P\{|Z - z_\ell| < \varepsilon\}^j P\{|Z - z_\ell| \geq \varepsilon\}^{n_y-j}, \end{aligned}$$

where (1) holds by the fact that $\{Z_i : 1 \leq i \leq n_y\}$ are identically distributed and (2) by the Binomial Theorem. The desired result then follows from assumption that $Z_{n_y, (j)}^\ell - z_\ell = o_p(1)$ and the following derivation

$$\begin{aligned} & \binom{n_y}{j} P\{|Z - z_\ell| < \varepsilon\}^j [1 - P\{|Z - z_\ell| < \varepsilon\}]^{n_y-j} \\ & \leq n_y^j P\{|Z - z_\ell| \geq \varepsilon\}^{n_y-j} = \left[\exp^{\frac{j \ln n_y}{n_y-j}} P\{|Z - z_\ell| \geq \varepsilon\} \right]^{n_y-j} \stackrel{(1)}{\rightarrow} 0, \end{aligned}$$

where (1) follows from Assumption 1 and noticing that $\exists N$ s.t. $\exp^{\frac{j \ln n_y}{n_y-j}} P\{|Z - z_\ell| \geq \varepsilon\} \leq P\{|Z - z_\ell| \geq \varepsilon\} < 1$ for all $n_y > N$ and any $j \in \{1, \dots, q_y - 1\}$.

By (B-5), (B-6), and Assumption 2, we have

$$\lim_{n \rightarrow \infty} P \left\{ \bigcap_{\ell=1}^L \bigcap_{i=1}^{q_\ell} \{S_{n,i}^\ell \leq s_i^\ell\} \right\} = \prod_{\ell=1}^L \prod_{i=1}^{q_\ell} F_Y(s_i^\ell | z_\ell) \prod_{j=1}^{q_x} F_X(s_{j+q_y}^\ell | z_\ell). \quad (\text{B-9})$$

By definition of convergence in distribution, (B-9) implies (B-1), as desired. ■

Theorem 4. *Let S be the random variable in Theorem 1. Then, for any $P \in \mathbf{P}_0$ and $\alpha \in (0, 1)$, we obtain*

$$E_P[\phi(S)] \leq \bar{\alpha} \leq \alpha,$$

where

$$\bar{\alpha} := P \left\{ \sup_{u \in (0,1)} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\} \right) > c_\alpha(q_y, q) \right\}. \quad (\text{B-10})$$

Moreover, the first inequality becomes an equality under P such that $F_Y(t|z_0) = F_X(t|z_0)$ for all $t \in \mathbf{R}$ and these are continuous functions of $t \in \mathbf{R}$.

Proof. Recall that $S = (S_1, \dots, S_{q_y}, S_{q_y+1}, \dots, S_q)$ are independent, and such that $S_j \sim F_Y(t|z_0)$ for all $j = 1, \dots, q_y$ and $S_{q_y+j} \sim F_X(t|z_0)$ for all $j = 1, \dots, q_x$. Denote by $Q_Y(\cdot|z_0)$ and $Q_X(\cdot|z_0)$ the quantile

functions of $F_Y(t|z_0)$ and $F_X(t|z_0)$. Consider the following argument:

$$\begin{aligned}
& E_P[\phi(S)] \\
&= P \left\{ \max_{k \in \{1, \dots, q_y\}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{S_j \leq S_k\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{S_j \leq S_k\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(1)}{=} P \left\{ \sup_{t \in \mathbf{Y}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{Q_Y(U_j|z_0) \leq t\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{Q_X(U_j|z_0) \leq t\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(2)}{=} P \left\{ \sup_{t \in \mathbf{Y}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq F_Y(t|z_0)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq F_X(t|z_0)\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(3)}{\leq} P \left\{ \sup_{t \in \mathbf{Y}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq F_X(t|z_0)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq F_X(t|z_0)\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(4)}{=} P \left\{ \sup_{u \in \mathcal{U}} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(5)}{\leq} P \left\{ \sup_{u \in (0,1)} \left(\frac{1}{q_y} \sum_{j=1}^{q_y} I\{U_j \leq u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq u\} \right) > c_\alpha(q_y, q) \right\} \\
&\stackrel{(6)}{=} \bar{\alpha} ,
\end{aligned}$$

where (1) follows from the quantile transformation and the fact that replacing $\max_{k \in \{1, \dots, q_y\}}$ with $\sup_{t \in \mathbf{Y}}$ does not affect the magnitude of the test statistic, (2) follows from Pollard (2002, Eq. 36), (3) from $P \in \mathbf{P}_0$, (4) from a simple change of variables and $\mathcal{U} := \cup_{t \in \mathbf{Y}} \{u = F_X(t|z_0)\}$, (5) from $\mathcal{U} \subseteq (0, 1)$, and (6) from the definition of $c_\alpha(q_y, q)$ and $\bar{\alpha}$, and the fact that $\{U_j : j = 1, \dots, q\}$ are i.i.d. distributed as $U(0, 1)$.

To conclude the proof, it suffices to show that: (i) inequality (3) holds as an equality under the condition $F_Y(t | z_0) = F_X(t | z_0)$ for all $t \in \mathbf{R}$, and (ii) inequality (5) holds as an equality when these functions are continuous in t . The first claim is immediate. For the second, it follows from the fact that the continuity of the CDFs implies $\mathcal{U} = (0, 1)$. By elementary properties of CDFs, $\lim_{t \rightarrow -\infty} F_X(t | z_0) = 0$ and $\lim_{t \rightarrow \infty} F_X(t | z_0) = 1$. By the intermediate value theorem, for any $u \in (0, 1)$, there exists $t \in \mathbf{R}$ such that $u = F_X(t | z_0)$, implying $u \in \mathcal{U}$, as desired. ■

B.2 Auxiliary lemmas

Lemma 3. *Suppose that $S_n \sim P$ with P satisfying Assumptions 1, 2, and 3 hold. Consider a sequence of random vectors $\{\tilde{S}_n : 1 \leq n < \infty\}$ and a random vector \tilde{S} defined on a common probability space $(\Omega, \mathcal{A}, \tilde{P})$ such that*

$$\tilde{S}_n \stackrel{d}{=} S_n, \quad \tilde{S} \stackrel{d}{=} S, \quad \text{and} \quad \tilde{S}_n \xrightarrow{a.s.} \tilde{S}. \quad (\text{B-11})$$

Then, for any $j \in \{1, \dots, q\}$,

$$\tilde{P}\{\tilde{S}_{n,j} \neq \tilde{S}_j, \tilde{S}_j \in \mathcal{D}\} = o_n(1).$$

Proof. We focus on an arbitrary $j \in \{1, \dots, q_y\}$. The argument for $j \in \{q_y + 1, \dots, q\}$ follows analogously

by replacing Y with X . By Lemma 6, it suffices to show that

$$\sup_{t \in \mathbb{R}} |F_{\tilde{S}_{n,j}}(t) - F_{\tilde{S}_j}(t)| = o_n(1), \quad (\text{B-12})$$

where $F_{\tilde{S}_{n,j}}$ and $F_{\tilde{S}_j}$ denote the distribution functions of $\tilde{S}_{n,j}$ and \tilde{S}_j in $(\Omega, \mathcal{A}, \tilde{P})$.

By the proof in Theorem 3 (with $L = 1$ and $\mathcal{Z} = \{z_0\}$), we have $F_{S_j}(t) = F_Y(t|z_0)$ and $F_{S_{n,j}}(t) = E[F_Y(t|Z_{n_y,(j)})]$ where $Z_{n_y,(j)} \xrightarrow{P} z_0$. By these and (B-11), (B-12) follows from

$$\sup_{t \in \mathbb{R}} |E[F_Y(t|Z_{n_y,(j)})] - F_Y(t|z_0)| = o_n(1). \quad (\text{B-13})$$

For any $\varepsilon > 0$,

$$\begin{aligned} E[F_Y(t|Z_{n_y,(j)})] &= \int_{|z-z_0| \leq \varepsilon} F_Y(t|z) dP_{Z_{n_y,(j)}}(z) + \int_{|z-z_0| > \varepsilon} F_Y(t|z) dP_{Z_{n_y,(j)}}(z) \\ &\leq \int_{|z-z_0| \leq \varepsilon} F_Y(t|z) dP_{Z_{n_y,(j)}}(z) + P(|Z_{n_y,(j)} - z_0| > \varepsilon) \\ &\stackrel{(1)}{=} \int_{|z-z_0| \leq \varepsilon} F_Y(t|z) dP_{Z_{n_y,(j)}}(z) + o_n(1), \end{aligned}$$

where (1) holds by $Z_{n_y,(j)} \xrightarrow{P} z_0$. Then,

$$\sup_{t \in \mathbb{R}} |E[F_Y(t|Z_{n_y,(j)})] - F_Y(t|z_0)| \leq \sup_{t \in \mathbb{R}} \sup_{|z-z_0| \leq \varepsilon} |F_Y(t|z) - F_Y(t|z_0)| + o_n(1).$$

Fix $\delta > 0$ arbitrarily. By Assumption 2, $\exists \varepsilon > 0$ such that $\sup_{t \in \mathbb{R}} \sup_{|z-z_0| \leq \varepsilon} |F_Y(t|z) - F_Y(t|z_0)| < \delta/2$. For all large enough n_y , the right-hand side is bounded above by δ . Since the choice of δ was arbitrary, (B-13) follows. ■

Lemma 4. *Let V_1 and V_2 be independent random variables that are discontinuous at a finite set of points \mathcal{D}_1 and \mathcal{D}_2 , respectively. Then, for any $\delta > 0$, $\exists \varepsilon > 0$ small enough s.t.*

$$\begin{aligned} P\{\cup_{d \in \mathcal{D}_1} \{|V_1 - d| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\}\} &< \delta, \\ P\{\{|V_1 - V_2| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\} \cap \{V_2 \in \mathcal{D}_2^c\}\} &< \delta. \end{aligned}$$

Proof. Set $\bar{\varepsilon} := \{\min |d - \tilde{d}| : d, \tilde{d} \in \mathcal{D}_1 \cap d \neq \tilde{d}\} > 0$. For any $\varepsilon \in (0, \bar{\varepsilon}/2)$,

$$\begin{aligned} &P\{\cup_{d \in \mathcal{D}_1} \{|V_1 - d| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\}\} \\ &\leq \sum_{d \in \mathcal{D}_1} P\{\{|V_1 - d| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\}\} \\ &\leq_{(1)} \sum_{d \in \mathcal{D}_1} P\{V_1 \in (d - \varepsilon, d) \cup (d, d + \varepsilon)\} \stackrel{(2)}{=} o_\varepsilon(1), \end{aligned}$$

where (1) holds by $\varepsilon \in (0, \bar{\varepsilon}/2)$ and (2) by the fact that $(d - \varepsilon, d) \cap (d, d + \varepsilon)$ has no discontinuities in the CDF. The right-hand side is less than δ by making ε arbitrarily small, as desired. Also, for any $\varepsilon \in (0, \bar{\varepsilon}/2)$,

$$P\{\{|V_1 - V_2| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\} \cap \{V_2 \in \mathcal{D}_2^c\}\}$$

$$\stackrel{(1)}{=} \int_{\mathcal{D}_2^c} P\{\{|V_1 - v_2| < \varepsilon\} \cap \{V_1 \in \mathcal{D}_1^c\}\} dP_{V_2}(v_2) \stackrel{(2)}{=} o_\varepsilon(1) ,$$

where (1) holds by $V_1 \perp V_2$, and (2) holds by Lemma 5 and the dominated convergence theorem. The right-hand side is less than δ by making ε arbitrarily small, as desired. ■

Lemma 5. *Consider a random variable V whose CDF is discontinuous at a finite set of points \mathcal{D} . Then, for any $v \in \mathbf{R}$,*

$$P\{\{|V - v| < \varepsilon\} \cap \{V \in \mathcal{D}^c\}\} = o_\varepsilon(1) .$$

Proof. Set $\bar{\varepsilon} := \{\min |d - \tilde{d}| : d, \tilde{d} \in \mathcal{D}_1 \cap d \neq \tilde{d}\} > 0$. Fix $\varepsilon < \bar{\varepsilon}/2$. There are two possibilities: $v \in \mathcal{D}$ or $v \notin \mathcal{D}$. First, consider $v \in \mathcal{D}$. In this case,

$$P\{|V - v| < \varepsilon\} \cap \{V \in \mathcal{D}^c\} \leq P\{V \in (v - \varepsilon, v) \cap (v, v + \varepsilon)\} \stackrel{(1)}{=} o_\varepsilon(1) ,$$

where (1) holds because $(d - \varepsilon, d) \cap (d, d + \varepsilon)$ has no mass points. Second, consider $v \notin \mathcal{D}$. Then,

$$P\{\{|V - v| < \varepsilon\} \cap \{V \in \mathcal{D}^c\}\} \leq F_V(v + \varepsilon) - F_V(v - \varepsilon) \stackrel{(2)}{=} o_\varepsilon(1) ,$$

where (1) by the fact that $(v - \varepsilon, v + \varepsilon]$ has no mass points, so F_V is continuous on that interval. ■

Lemma 6. *Let $\{V_n : n \geq 1\}$ be a sequence of random variables that satisfy $V_n \xrightarrow{P} V$, where V is a random variable whose CDF is discontinuous at a finite set of points \mathcal{D} . Furthermore, assume $\sup_{t \in \mathbf{R}} |F_{V_n}(t) - F_V(t)| \rightarrow 0$. Then,*

$$P\{\{V \in \mathcal{D}\} \cap \{V_n \neq V\}\} \rightarrow 0 .$$

Proof. Fix $\delta > 0$ arbitrarily. It suffices to find $N = N(\delta)$ such that $P\{V \in \mathcal{D}, V_n \neq V\} < \delta$ for all $n > N$.

Set $\bar{\varepsilon} := \{\min |d - \tilde{d}| : d, \tilde{d} \in \mathcal{D} \cap d \neq \tilde{d}\} > 0$. For any $\varepsilon \in (0, \bar{\varepsilon}/2)$, consider the following argument.

$$\begin{aligned} & P\{\{V \in \mathcal{D}\} \cap \{V_n \neq V\}\} \\ &= \sum_{d \in \mathcal{D}} P\{\{V = d\} \cap \{V_n \neq d\}\} \\ &\stackrel{(1)}{=} \sum_{d \in \mathcal{D}} P\{\{V = d\} \cap \{V_n \neq d\} \cap \{|V_n - d| < \varepsilon\}\} + o_n(1) \\ &\leq \sum_{d \in \mathcal{D}} P\{V_n \in (d - \varepsilon, d) \cap (d, d + \varepsilon)\} + o_n(1) \\ &\stackrel{(2)}{=} \sum_{d \in \mathcal{D}} P\{V \in (d - \varepsilon, d) \cap (d, d + \varepsilon)\} + o_n(1) \\ &\stackrel{(3)}{=} o_\varepsilon(1) + o_n(1) , \end{aligned}$$

where (1) holds by $V_n \xrightarrow{P} V$, (2) by $\sup_{t \in \mathbf{R}} |F_{V_n}(t) - F_V(t)| = o_n(1)$, and (3) by the fact that $(d - \varepsilon, d) \cap (d, d + \varepsilon)$ has no discontinuities in the CDF. For all large enough n and small enough ε , the right-hand side is bounded by δ , as desired. ■

Appendix C Extensions and Supplementary Analyses

C.1 Asymptotics for large q

In this section, we invoke two results from [Bugni et al. \(2025\)](#) to show (4.6). To properly use these results, we need to introduce some basic notation and two assumptions.

Let P_W denote the distribution of (W, Z) , with W standing for either Y or X . Assume P_W is dominated by the product measure $d\mu := \nu(dw) \otimes dz$, where dz is Lebesgue measure on \mathbf{R} and ν is a fixed σ -finite measure on \mathbf{R} . We denote the corresponding joint density by $f(w, z)$ and the marginal density of Z by

$$f_Z(z) := \int f(w, z) \nu(dw) .$$

With this notation in place, we introduce our next assumption.

Assumption 4. *The random variable Z has a density $f_Z(z)$ with respect to Lebesgue measure. In addition, $f_Z(z)$ satisfies $|f_Z(z) - f_Z(z_\ell)| \leq C_Z |z - z_\ell|$ for a constant $C_Z < \infty$ and all $z_\ell \in \mathcal{Z}$.*

Denote the density of the conditional distribution of W given $Z = z$ by

$$p_z(w) := \frac{f(w, z)}{f_Z(z)} .$$

We say that $p_z(w)$ is differentiable in quadratic mean at z_0 if there exists a measurable score $\dot{\ell}_{z_0} : \mathbf{R} \rightarrow \mathbf{R}$ with

$$\int \dot{\ell}_{z_0}^2(w) p_{z_0}(w) \nu(dw) < \infty ,$$

such that, as $t \rightarrow 0$,

$$\int \left[\sqrt{p_{z_0+t}(w)} - \sqrt{p_{z_0}(w)} - \frac{1}{2} (t \cdot \dot{\ell}_{z_0}(w)) \sqrt{p_{z_0}(w)} \right]^2 \nu(dw) = o(t^2) .$$

We now state our last assumption of this section.

Assumption 5. *The densities $p_z(y)$ and $p_z(x)$ are differentiable in quadratic mean at all $z_\ell \in \mathcal{Z}$.*

Finally, for $r > 0$ let $B_r := \{z : |z - z_0| < r\}$, and define the ‘‘local’’ conditional distribution of W given $Z \in B_r$ by $P_{W,r}^*\{A\} := P_W\{W \in A \mid Z \in B_r\}$. The corresponding density is

$$h_r(w) := \frac{\int_{B_r} f(w, z) dz}{\int_{B_r} f_Z(z) dz} .$$

With a slight abuse of notation, we write $P_{W,0}^*$ for the conditional distribution of W given $Z = z_0$.

Recall that the vector of induced order statistics S_n in [Theorem 1](#) is defined as

$$S_n := (S_n^Y, S_n^X) = (Y_{n_y,[1]}, \dots, Y_{n_y,[q_y]}, X_{n_x,[1]}, \dots, X_{n_x,[q_x]}) , \quad (\text{C-1})$$

while $S := (S^Y, S^X)$ is a $q = q_y + q_x$ -dimensional vector of i.i.d. random variables with distribution $P_{Y,0}^* \otimes P_{X,0}^*$. Under [Assumptions 4 and 5](#), [Theorem 2](#) in [Bugni et al. \(2025\)](#) shows that, uniformly over $n \in \mathbf{N}$, $q_y \in$

$\{1, 2, \dots, n_y\}$, and $q_x \in \{1, 2, \dots, n_x\}$

$$\text{TV}(\mathcal{L}(S_n^Y), \mathcal{L}(S^Y)) = O(q_y(q_y/n_y)^a) \quad \text{and} \quad \text{TV}(\mathcal{L}(S_n^X), \mathcal{L}(S^X)) = O(q_x(q_x/n_x)^a)$$

provided that $\text{TV}(P_{Y,r}^*, P_{Y,0}^*) = O(r^a)$ and $\text{TV}(P_{X,r}^*, P_{X,0}^*) = O(r^a)$ for some $a > 0$. In turn, Theorem 3 in Bugni et al. (2025) establishes that both of these conditions hold for $a = 2$ under the same assumptions. Putting these results together leads to the following theorem.

Theorem 5. *Let \mathbf{P} be the class of distributions $P := P_Y \otimes P_X$ satisfying Assumptions 4 and 5, and let \mathbf{P}_0 be as in (2.3). Let $\alpha \in (0, 1)$ be given, T be the KS test statistic in (3.4), and $\phi(\cdot)$ be the test in (3.7). Then,*

$$E_P[\phi(S_n)] \leq \alpha + O\left(\frac{q_y^3}{n_y^2} + \frac{q_x^3}{n_x^2}\right), \quad (\text{C-2})$$

whenever $P \in \mathbf{P}_0$. If we let $n := \min\{n_y, n_x\}$ and $\frac{q_y}{q_y + q_x} \rightarrow c \in (0, 1)$ as $n \rightarrow \infty$, then for any $P' \in \mathbf{P}_1$

$$\lim_{n \rightarrow \infty} E_{P'}[\phi(S_n)] = 1. \quad (\text{C-3})$$

It is worth noting that Assumption 4 is stronger than Assumption 1, and Assumption 5 is stronger than Assumption 2. This is natural: the stronger assumptions deliver explicit rates of convergence, whereas the weaker assumptions ensure convergence but do not characterize its rate. Note also that (C-3) ensures that $\phi(S_n)$ is consistent under fixed alternatives, which necessarily requires an asymptotic framework where q diverges. For an in-depth discussion of the roles of these assumptions, as well as related rate results under weaker versions of Assumption 5, we refer the reader to Bugni et al. (2025).

C.2 Results for other test statistics

Our paper proposes the test in (3.7) based on the one-sided KS test statistic in (3.4) and a critical value in (3.6) constructed by simulating i.i.d. uniform random variables applied to this statistic. In this section, we investigate the properties of analogous tests that replace the KS statistic with other commonly used statistics in the stochastic dominance literature, such as the Cramér–von Mises (CvM) and Anderson–Darling (AD) statistics.

The main takeaway of this section is that the validity of our approach does not extend to tests that replace the KS statistic with the CvM or AD statistics. Specifically, we provide examples of data-generating processes that satisfy the null hypothesis in (2.1) and the test overrejects. Our examples involve discretely distributed data, which is allowed by our assumptions. That said, the CvM-based version of our test remains valid under the assumption of continuously distributed data, as we show later.

For concreteness, we now define the CvM and AD analogs of our test for the null hypothesis in (2.1). Given the pooled sample S_n in (3.3), the one-sided CvM test statistic is given by

$$T_{\text{CvM}}(S_n) = \int \left(\hat{F}_{n,Y}(s) - \hat{F}_{n,X}(s) \right)^+ d\hat{F}_{n,S}(s) = \frac{1}{q} \sum_{j=1}^q \left(\hat{F}_{n,Y}(S_{n,j}) - \hat{F}_{n,X}(S_{n,j}) \right)^+, \quad (\text{C-4})$$

where $\hat{F}_{n,S}(s)$ denotes the empirical CDF of S_n and $x^+ = \max\{x, 0\}^2$. In turn, the one-sided AD test

statistic is given by

$$T_{\text{AD}}(S_n) = \int \frac{(\hat{F}_{n,Y}(s) - \hat{F}_{n,X}(s))^+}{\hat{F}_{n,S}(s)(1 - \hat{F}_{n,S}(s))} d\hat{F}_{n,S}(s) = \frac{1}{q} \sum_{j=1}^q \frac{(\hat{F}_{n,Y}(S_{n,j}) - \hat{F}_{n,X}(S_{n,j}))^+}{\hat{F}_{n,S}(S_{n,j})(1 - \hat{F}_{n,S}(S_{n,j}))}. \quad (\text{C-5})$$

These statistics align with their standard textbook definitions, as discussed in [Hajek et al. \(1999, p. 101\)](#). For both of these statistics, we can define the analog test as in our paper. For the CvM statistic, the corresponding test is

$$\phi_{\text{CvM}}(S) = I\{T_{\text{CvM}}(S) > c_{\alpha, \text{CvM}}(q_y, q)\}. \quad (\text{C-6})$$

where $c_{\alpha, \text{CvM}}(q_y, q) = \inf_{x \in \mathbf{R}} \{P\{T_{\text{CvM}}(U) \leq x\} \geq 1 - \alpha\}$ and $U = (U_1, \dots, U_q)$ is a vector of i.i.d. $U(0, 1)$. The testing procedure based on the AD statistic is defined analogously.

We are now ready to argue that the CvM and AD analogs of our test are invalid under our assumptions. Since the CvM and AD statistics are both rank statistics, we can repeat arguments in [Theorem 2](#) to show that, for any $P \in \mathbf{P}_0$,

$$\lim_{n \rightarrow \infty} E_P[\phi_{\text{CvM}}(S_n)] = E_{\tilde{P}}[\phi_{\text{CvM}}(\tilde{S})] \quad \text{and} \quad \lim_{n \rightarrow \infty} E_P[\phi_{\text{AD}}(S_n)] = E_{\tilde{P}}[\phi_{\text{AD}}(\tilde{S})].$$

To prove that these tests are invalid, it then suffices to find any $P \in \mathbf{P}_0$ such that $E_{\tilde{P}}[\phi_{\text{CvM}}(\tilde{S})] > \alpha$ and $E_{\tilde{P}}[\phi_{\text{AD}}(\tilde{S})] > \alpha$. To this end, we consider $q_y = 2, q_x = 1, \{Y|Z = z_0\}$ and $\{X|Z = z_0\}$ distributed according to *Bernoulli*(0.5). For $\alpha = 5\%$, we get $E_{\tilde{P}}[\phi_{\text{CvM}}(\tilde{S})] \approx 12.5\%$ and $E_{\tilde{P}}[\phi_{\text{AD}}(\tilde{S})] \approx 12.5\%$, as desired.

A notable feature of the examples above is that the data are discrete. This naturally raises the question of whether the validity of these tests can be recovered in settings with continuously distributed data. While we were unable to establish this result for the AD test, we were able to do so for the CvM test.

Theorem 6. *Let \mathbf{P} the space of distributions that satisfy [Assumptions 1, 2](#), and $Y|Z = z_0$ and $X|Z = z_0$ are continuously distributed, and let \mathbf{P}_0 be the subset of \mathbf{P} such that [\(2.1\)](#) holds. Let $\alpha \in (0, 1)$ be given, T_{CvM} be the CvM test statistic in [\(C-4\)](#), and $\phi_{\text{CvM}}(\cdot)$ be the test in [\(C-6\)](#). Then,*

$$\limsup_{n \rightarrow \infty} E_P[\phi_{\text{CvM}}(S_n)] \leq \alpha \quad (\text{C-7})$$

whenever $P \in \mathbf{P}_0$.

Proof. This result follows from repeating arguments that prove [Theorem 2](#). The only difference in the proof is that [Theorem 4](#) is replaced by [Theorem 7](#). ■

The previous theorem relies on the following auxiliary result.

Theorem 7. *Let S be the random variable in [Theorem 1](#) and let \mathbf{P}_0 be as in [Theorem 6](#). Then, for any $P \in \mathbf{P}_0$ and $\alpha \in (0, 1)$, we obtain*

$$E_P[\phi_{\text{CvM}}(S)] \leq \bar{\alpha}_{\text{CvM}} \leq \alpha,$$

where

$$\bar{\alpha}_{\text{CvM}} = P\{T_{\text{CvM}}(U) > c_{\alpha, \text{CvM}}(q_y, q)\}.$$

Moreover, the first inequality becomes an equality under P such that $F_Y(t|z_0) = F_X(t|z_0)$ for all $t \in \mathbf{R}$.

Proof. For any $t \in \mathbf{R}$, consider the CDF

$$F_M(t) = \frac{q_Y}{q} F_{Y|Z=z_0}(t) + \frac{q_X}{q} F_{X|Z=z_0}(t),$$

and let Q_M denote its corresponding quantile function. Let Q_Y and Q_X are the quantiles functions associated with $F_Y(\cdot|z_0)$ and $F_X(\cdot|z_0)$, respectively.

Since $Y|Z = z_0$ and $X|Z = z_0$ are continuously distributed, F_M is continuous. In turn, this implies that Q_M is strictly increasing. By $F_{Y|Z=z_0}(t) \leq F_{X|Z=z_0}(t)$ for all $t \in \mathbf{R}$, we also have

$$Q_Y(u) \geq Q_M(u) \geq Q_X(u) \quad \forall u \in (0, 1). \quad (\text{C-8})$$

For brevity notation, we denote the critical value as $c_\alpha \equiv c_{\alpha, \text{CvM}}(q_y, q)$. Then,

$$\begin{aligned} & E_P[\phi_{\text{CvM}}(S)] \\ &= P \left\{ \frac{1}{q} \sum_{u=1}^q \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{S_i \leq S_u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{S_j \leq S_u\} \right)^+ > c_\alpha \right\} \\ &\stackrel{(1)}{=} P \left\{ \left[\begin{aligned} & \frac{1}{q} \sum_{u=1}^{q_y} \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{Q_Y(U_i) \leq Q_Y(U_u)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{Q_X(U_j) \leq Q_Y(U_u)\} \right)^+ + \\ & \frac{1}{q} \sum_{u=q_y+1}^q \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{Q_Y(U_i) \leq Q_X(U_u)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{Q_X(U_j) \leq Q_X(U_u)\} \right)^+ \end{aligned} \right] > c_\alpha \right\} \\ &\stackrel{(2)}{\leq} P \left\{ \left[\begin{aligned} & \frac{1}{q} \sum_{u=1}^{q_y} \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{Q_M(U_i) \leq Q_M(U_u)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{Q_M(U_j) \leq Q_M(U_u)\} \right)^+ + \\ & \frac{1}{q} \sum_{u=q_y+1}^q \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{Q_M(U_i) \leq Q_M(U_u)\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{Q_M(U_j) \leq Q_M(U_u)\} \right)^+ \end{aligned} \right] > c_\alpha \right\} \\ &\stackrel{(3)}{=} P \left\{ \left[\begin{aligned} & \frac{1}{q} \sum_{u=1}^{q_y} \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{U_i \leq U_u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq U_u\} \right)^+ + \\ & \frac{1}{q} \sum_{u=q_y+1}^q \left(\frac{1}{q_y} \sum_{i=1}^{q_y} I\{U_i \leq U_u\} - \frac{1}{q_x} \sum_{j=q_y+1}^q I\{U_j \leq U_u\} \right)^+ \end{aligned} \right] > c_\alpha \right\} \\ &\stackrel{(4)}{=} P \{T_{\text{CvM}}(U) > c_\alpha\} \stackrel{(5)}{=} \bar{\alpha}_{\text{CvM}} \stackrel{(6)}{\leq} \alpha, \quad (\text{C-9}) \end{aligned}$$

where (1) holds by the quantile transformation with (U_1, \dots, U_q) i.i.d. $U(0, 1)$, (2) by

$$\{Q_Y(U_a) \leq Q_X(U_b)\} \subseteq \{Q_M(U_a) \leq Q_M(U_b)\} \subseteq \{Q_X(U_a) \leq Q_Y(U_b)\} \quad (\text{C-10})$$

for all $a, b = 1, \dots, q$, (3) by the fact that Q_M is strictly increasing, and (4), (5), and (6) by the definitions of T_{CvM} and $\bar{\alpha}_{\text{CvM}}$. We now show that (C-10) holds. For any $a, b = 1, \dots, q$, suppose that $Q_Y(U_a) \leq Q_X(U_b)$. By this and (C-8), we have that $Q_M(U_a) \leq Q_Y(U_a) \leq Q_X(U_b) \leq Q_M(U_b)$, which implies that $Q_M(U_a) \leq Q_M(U_b)$, as desired by the first inclusion. In turn, by this and (C-8), we have that $Q_X(U_a) \leq Q_M(U_a) \leq Q_M(U_b) \leq Q_Y(U_b)$, as desired by the second inclusion.

Since $P \in \mathbf{P}_0$ was arbitrary, (C-9) implies that $\sup_{P \in \mathbf{P}_0} E_P[\phi_{\text{CvM}}(S)] \leq \bar{\alpha}$. Furthermore, under $F_Y(\cdot|z_0) = F_X(\cdot|z_0)$, we have that $Q_Y = Q_X = Q_M$, and so the inequality (2) in (C-9) holds with equality, as desired. ■

C.3 Multiple target points

Throughout the paper, we have focused on testing the null hypothesis (2.1) with $L = 1$ to emphasize the main ideas and maintain clarity. This section describes how our method extends when $L > 1$. In this setting, we address the intersection nature of the null hypothesis by using a maximum test.

To formally define the test, we need to update our notation to explicitly reflect dependence on both ℓ and α . Consider the point $z_\ell \in \mathcal{Z}$ for some $\ell \in \{1, \dots, L\}$. Let $g_\ell(Z) := |Z - z_\ell|$ and, for any two values $z, z' \in \mathcal{Z}$, define the ordering $z \leq_\ell z'$ as follows:

$$z \leq_\ell z' \quad \text{if and only if} \quad g_\ell(z) \leq g_\ell(z') .$$

For each ℓ , this leads to g -order statistics $Z_{\ell,(i)}$ in each sample, as well as induced order statistics for Y and X , that we denote by

$$Y_{n_y,[1]}^\ell, Y_{n_y,[2]}^\ell, \dots, Y_{n_y,[q_y]}^\ell \quad \text{and} \quad X_{n_x,[1]}^\ell, X_{n_x,[2]}^\ell, \dots, X_{n_x,[q_x]}^\ell .$$

As with $L = 1$, any ties in the g -ordering can be resolved arbitrarily. Finally, if we let $n := n_y + n_x$ and $q := q_y + q_x$, the effective (pooled) sample is given by

$$S_n^\ell = (S_{n,1}^\ell, \dots, S_{n,q}^\ell) := (Y_{n_y,[1]}^\ell, \dots, Y_{n_y,[q_y]}^\ell, X_{n_x,[1]}^\ell, \dots, X_{n_x,[q_x]}^\ell) . \quad (\text{C-11})$$

Our test is entirely based on S_n^ℓ , with the default test statistic being

$$T(S_n^\ell) = \sup_{t \in \mathbf{R}} \left(\hat{F}_{n,Y}^\ell(t) - \hat{F}_{n,X}^\ell(t) \right) = \max_{k \in \{1, \dots, q_y\}} \left(\hat{F}_{n,Y}^\ell(S_{n,k}^\ell) - \hat{F}_{n,X}^\ell(S_{n,k}^\ell) \right) ,$$

where the empirical CDFs are,

$$\hat{F}_{n,Y}^\ell(t) := \frac{1}{q_y} \sum_{j=1}^{q_y} I\{S_{n,j}^\ell \leq t\} \quad \text{and} \quad \hat{F}_{n,X}^\ell(t) := \frac{1}{q_x} \sum_{j=1}^{q_x} I\{S_{n,q_y+j}^\ell \leq t\} .$$

In order to define our test below in (C-12), we first introduce $\phi_\ell(\alpha)$, where

$$\phi_\ell(S_n^\ell, \alpha) := I\{T(S_n^\ell) > c_\alpha(q_y, q)\} .$$

The test $\phi_\ell(S_n^\ell, \alpha)$ depends on the point z_ℓ , or equivalently the index ℓ , in two ways. First, the random variable S_n^ℓ depends on z_ℓ through the induced order statistics, meaning the test statistic $T(S_n^\ell)$ is influenced by the choice of the target point z_ℓ . Second, the critical value $c_\alpha(q_y, q)$ is a function of (q_y, q) , which, through the data-dependent rules introduced in Section 5.1, implicitly depends on z_ℓ . The dependence on α is directly evident from the definition of the critical value $c_\alpha(q_y, q)$.

To test the null hypothesis in (2.1) when $L > 1$, we propose the following test:

$$\phi(S_n^1, S_n^2, \dots, S_n^L) := \max_{\ell \in \{1, \dots, L\}} \phi_\ell \left(S_n^\ell, 1 - (1 - \alpha)^{1/L} \right) . \quad (\text{C-12})$$

In other words, the test $\phi(S_n^1, S_n^2, \dots, S_n^L)$ is the maximum of the L individual tests, each computed at a target point z_ℓ . However, each of these individual tests is performed at a level of $1 - (1 - \alpha)^{1/L}$, rather than at the nominal level α . Since \mathcal{Z} consists of a finite number of points, the asymptotic validity of the test $\phi(\alpha)$

follows from the validity of each individual test ϕ_ℓ and the fact that they are asymptotically independent. We formalize this result in Theorem 3.

C.4 An empirical illustration

In this section, we apply our methodology to assess the distributional effect of retirement on household expenditures. Our empirical analysis follows Battistin et al. (2009) and Shen and Zhang (2016) (SZ, henceforth). As in these studies, we use data from the 1993–2004 Italian Survey on Household Income and Wealth (SHIW), a repeated cross-section survey that covers roughly 8,000 households across more than 3,000 municipalities; [dataset] Shen and Zhang (2016).

We use the SHIW data to evaluate whether retirement shifts the distribution of household expenditures downward. As SZ explains, this hypothesis can be examined by framing the problem as a regression discontinuity design, comparing the distributions of household expenditures at ages just before and just after retirement. To formulate our problem, let Z denote the age of the household head, z_0 the retirement age, Y household expenditure before retirement, and X household expenditure after retirement. Under continuity of the conditional distributions in the retirement age of the household head, the test of a negative distributional effect of retirement can be formulated as the following CSD testing problem:

$$H_0 : F_Y(t|z_0) \leq F_X(t|z_0) \text{ for all } t \in \mathbf{R} \quad \text{versus} \quad H_1 : F_Y(t|z_0) > F_X(t|z_0) \text{ for some } t \in \mathbf{R} . \quad (\text{C-13})$$

We note that H_0 in (C-13) is a special case of (2.1) with $\mathcal{Z} = \{z_0\}$. Following the logic of regression discontinuity designs, conditioning on $Z = z_0$ is key to controlling for confounding factors.

Table 1 reports the results of implementing the test in (C-13) using both our methodology (KS) and the procedure proposed by SZ. Following SZ, we consider expenditures on nondurables and food, and we report results for four samples: the full sample, the subsample of households with wives eligible for retirement, the subsample of households with wives not yet eligible for retirement, and the subsample of households consisting of single men. The table presents the p-values for both tests and their corresponding effective sample sizes. The tests only reject H_0 in (C-13) for the subsample of households in which the wife is eligible for retirement. For this group, SZ rejects H_0 for both food and nondurable expenditures (at the 1% and 5% levels, respectively), whereas our test (KS) rejects only for food expenditures (at the 10% level). Consistent with our simulation evidence, the SZ test uses substantially larger effective sample sizes than our KS test.

| sample (n) | expenditure type | p-value (KS) | q (KS) | p-value (SZ) | q (SZ) |
|----------------------------------|---------------------|-----------------|-------------|-----------------|-------------|
| all households (26,914) | nondurable | 22.82 | 43 | 13.69 | 8,739 |
| | food | 86.96 | 42 | 71.74 | 8,739 |
| wife already eligible (4,798) | nondurable | 99.29 | 20 | 66.72 | 1,822 |
| | food | 99.34 | 20 | 99.69 | 1,414 |
| wife ineligible (7,305) | nondurable | 25.71 | 22 | 0.15*** | 1,460 |
| | food | 9.64* | 22 | 3.94** | 1,460 |
| single men (9,001) | nondurable | 37.07 | 25 | 65.82 | 2,884 |
| | food | 23.58 | 25 | 31.19 | 2,537 |

Table 1: Empirical results of our KS test and the SZ test using the SHIW data for each sample and expenditure type. q indicates the effective sample size. Significant at *10%, **5%, and ***1%.

C.5 Details on the Monte Carlo Simulations

The parameter values used in the simulations of Section 6 are reported in the following tables. The parameters for Designs 1-4 are summarized in Table 2, those for Designs 5-7 in Table 3, and the parameter values for case (d) - power results - in Table 4. Table 5 reports the mean values of q_y^* and q_x^* across simulations.

| Design | $\mu_Y(z)$ | $\mu_X(z)$ | $\sigma_Y(Z)$ | $\sigma_X(Z)$ | U, V | z_ℓ |
|--------|----------------|----------------------------|---------------|---------------|-----------|------------|
| 1a | z | z | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 1b | $1.05z$ | z | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 1c | z | z | z^2 | z^2 | $N(0, 1)$ | 0.25, 0.75 |
| 2a | z | $z^2 + 0.25$ | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 2b | $1.05z$ | $0.5z + 0.25$ | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 2c | z | $z - (z - 0.25)(z - 0.75)$ | z^2 | z^2 | $N(0, 1)$ | 0.25, 0.75 |
| 3a | z | z | z^2 | z^2 | $U[0, 1]$ | 0.5 |
| 3b | $z + 0.1z^2$ | z | $0.95z^2$ | z^2 | $U[0, 1]$ | 0.5 |
| 3c | z | z | z^2 | z^2 | $U[0, 1]$ | 0.25, 0.75 |
| 4a | $\mu(z)$ | $\mu(z)$ | 1 | 1 | $N(0, 1)$ | 0 |
| 4b | $\mu(z) + 0.1$ | $\mu(z)$ | 1 | 1 | $N(0, 1)$ | 0 |
| 4c | $\mu(z)$ | $\mu(z)$ | 1 | 1 | $N(0, 1)$ | -0.5, 0.5 |

Table 2: Parameter values for the continuously distributed designs. Designs 1 to 3 are based on the location-scale model in (6.1), while Designs 4 is based on the RDD model in (6.3).

| Design | $\mu_Y(z)$ | $\mu_X(z)$ | $\sigma_Y(Z)$ | $\sigma_X(Z)$ | U, V | z_ℓ |
|--------|------------|------------|---------------|---------------|----------|------------|
| 5a | 0 | 0 | z^2 | z^2 | C-logN | 0.5 |
| 5b | 0 | 0 | $1.05z^2$ | z^2 | C-logN | 0.5 |
| 5c | 0 | 0 | z^2 | z^2 | C-logN | 0.25, 0.75 |
| 6a | 0 | – | – | – | discrete | 0.5 |
| 6b | 1 | – | – | – | discrete | 0.5 |
| 6c | 0 | – | – | – | discrete | 0.25, 0.75 |
| 7a | 0 | – | – | – | Binomial | 0.5 |
| 7b | 1 | – | – | – | Binomial | 0.5 |
| 7c | 0 | – | – | – | Binomial | 0.25, 0.75 |

Table 3: Parameter values for the discrete and mixed designs. Designs 5 is based on the location-scale model in (6.1), while Designs 6 and 7 are discrete designs as described in the main text.

| Design | $\mu_Y(z)$ | $\mu_X(z)$ | $\sigma_Y(Z)$ | $\sigma_X(Z)$ | U, V | z_ℓ |
|--------|------------|---------------|---------------|---------------|-----------|----------|
| 1d | $0.95z$ | z | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 2d | z | $0.6z + 0.25$ | z^2 | z^2 | $N(0, 1)$ | 0.5 |
| 3d | z | z | $0.90z^2$ | z^2 | $U[0, 1]$ | 0.5 |
| 4d | $\mu(z)$ | $\mu(z)$ | $0.5 + z^2$ | 1 | $N(0, 1)$ | 0 |
| 5d | 0 | 0 | $0.90z^2$ | z^2 | C-logN | 0.5 |
| 6d | -1/2 | - | - | - | discrete | 0.5 |
| 7d | -1 | - | - | - | Binomial | 0.5 |

Table 4: Parameter values for all designs under the alternative hypothesis H_1 .

| Design | 1 | | | 2 | | | 3 | | | 4 | | |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | a | b | c | a | b | c | a | b | c | a | b | c |
| q_y^* | 79.54 | 78.63 | 34.84 | 79.51 | 79.53 | 34.85 | 45.48 | 44.04 | 20.05 | 45.28 | 45.29 | 19.96 |
| q_x^* | 79.52 | 79.52 | 34.86 | 79.72 | 89.57 | 34.92 | 45.49 | 45.48 | 20.05 | 46.14 | 46.20 | 20.35 |
| Design | 5 | | | 6 | | | 7 | | | | | |
| | a | b | c | a | b | c | a | b | c | | | |
| q_y^* | 82.97 | 82.99 | 36.35 | 96.74 | 102.1 | 44.66 | 59.65 | 60.46 | 26.20 | - | - | - |
| q_x^* | 82.99 | 82.98 | 36.37 | 96.75 | 96.78 | 42.34 | 59.64 | 59.61 | 26.20 | - | - | - |

Table 5: Mean values of q_y^* and q_x^* across simulations: $n = 1,000$, $\alpha = 10\%$, $MC = 10,000$.

References

- ABADIE, A. (2002): “Bootstrap tests for distributional treatment effects in instrumental variable models,” *Journal of the American statistical Association*, 97, 284–292.
- ANDERSON, G. (1996): “Nonparametric tests of stochastic dominance in income distributions,” *Econometrica: Journal of the Econometric Society*, 1183–1193.
- ANDREWS, D. W. AND X. SHI (2017): “Inference based on many conditional moment inequalities,” *Journal of Econometrics*, 196, 275–287.
- ARMSTRONG, T. B. AND M. KOLESÁR (2018): “Optimal inference in a class of regression models,” *Econometrica*, 86, 655–683.
- BARRETT, G. F. AND S. G. DONALD (2003): “Consistent tests for stochastic dominance,” *Econometrica*, 71, 71–104.
- BATTISTIN, E., A. BRUGIAVINI, E. RETTORE, AND G. WEBER (2009): “The retirement consumption puzzle: evidence from a regression discontinuity approach,” *American Economic Review*, 99, 2209–2226.
- BECKER, G. S. (1957): “The Economics of Discrimination,” *University of Chicago Press Economics Books*.
- BHARADWAJ, P., R. DEB, AND L. RENO (2024): “Statistical Discrimination and the Distribution of Wages,” Tech. rep., National Bureau of Economic Research.

- BHATTACHARYA, P. (1974): “Convergence of sample paths of normalized sums of induced order statistics,” *The Annals of Statistics*, 1034–1039.
- BLACKMAN, J. (1956): “An extension of the Kolmogorov distribution,” *The Annals of Mathematical Statistics*, 513–520.
- BUGNI, F. A. AND I. A. CANAY (2021): “Testing Continuity of a Density via g-order statistics in the Regression Discontinuity Design,” *Journal of Econometrics*, 221, 138–159.
- BUGNI, F. A., I. A. CANAY, AND D. KIM (2025): “On the Rate of Convergence of Induced Ordered Statistics and their Applications,” Tech. rep., Northwestern University.
- BUGNI, F. A., I. A. CANAY, AND A. M. SHAIKH (2018): “Inference under Covariate Adaptive Randomization,” *Journal of the American Statistical Association*, 113, 1784–1796.
- CANAY, I. A. AND V. KAMAT (2018): “Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design,” *The Review of Economic Studies*, 85, 1577–1608.
- CANAY, I. A., M. MOGSTAD, AND J. MOUNTJOY (2024): “On the use of outcome tests for detecting bias in decision making,” *Review of Economic Studies*, 91, 2135–2167.
- CANAY, I. A., J. P. ROMANO, AND A. M. SHAIKH (2017): “Randomization Tests under an Approximate Symmetry Assumption,” *Econometrica*, 85, 1013–1030.
- CAUGHEY, D., A. DAFOE, X. LI, AND L. MIRATRIX (2023): “Randomisation inference beyond the sharp null: bounded null hypotheses and quantiles of individual treatment effects,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85, 1471–1491.
- CHANG, M., S. LEE, AND Y.-J. WHANG (2015): “Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements,” *The Econometrics Journal*, 18, 307–346.
- CHUNG, E. AND J. P. ROMANO (2013): “Exact and asymptotically robust permutation tests,” *The Annals of Statistics*, 41, 484–507.
- [DATASET] SHEN, S. AND X. ZHANG (2016): “Replication data for: Distributional Tests for Regression Discontinuity: Theory and Empirical Examples,” Harvard Dataverse, version 1.1.
- DAVID, H. AND J. GALAMBOS (1974): “The asymptotic theory of concomitants of order statistics,” *Journal of Applied Probability*, 762–770.
- DAVIDSON, R. AND J.-Y. DUCLOS (2000): “Statistical inference for stochastic dominance and for the measurement of poverty and inequality,” *Econometrica*, 68, 1435–1464.
- DELGADO, M. A. AND J. C. ESCANCIANO (2013): “Conditional stochastic dominance testing,” *Journal of Business & Economic Statistics*, 31, 16–28.

- DONALD, S. G., Y.-C. HSU, AND G. F. BARRETT (2012): “Incorporating covariates in the measurement of welfare and inequality: methods and applications,” *The Econometrics Journal*, 15, C1–C30.
- DURBIN, J. (1973): *Distribution theory for tests based on the sample distribution function*, SIAM.
- GNEDENKO, B. V. AND V. S. KOROLYUK (1951): “On the maximum divergence of two empirical distributions (in Russian),” *DAN*, 80, 525.
- GOLDMAN, M. AND D. M. KAPLAN (2018): “Comparing distributions by multiple testing across quantiles or CDF values,” *Journal of Econometrics*, 206, 143–166.
- GONZALO, J. AND J. OLMO (2014): “Conditional stochastic dominance tests in dynamic settings,” *International Economic Review*, 55, 819–838.
- HÁJEK, J., Z. SIDÁK, AND P. K. SEN (1999): *Theory of rank tests*, Academic press, 2nd edition ed.
- HODGES, J. (1958): “The significance probability of the Smirnov two-sample test,” *Arkiv för matematik*, 3, 469–486.
- HÁJEK, J. AND Z. ŠIDÁK (1967): *Theory of Rank Tests*, New York: Academic Press.
- KAUFMANN, E. AND R.-D. REISS (1992): “On conditional distributions of nearest neighbors,” *Journal of Multivariate Analysis*, 42, 67–76.
- KOROLYUK, V. S. (1955): “On the discrepancy of empiric distributions for the case of two independent samples,” *Izv. Akad. Nauk SSSR Ser. Mat.*, 19, 81–96.
- LEHMANN, E. L. (1951): “Consistency and unbiasedness of certain nonparametric tests,” *The annals of mathematical statistics*, 165–179.
- LEHMANN, E. L. AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*, Springer, New York, 3rd ed.
- LINTON, O., E. MAASOUMI, AND Y.-J. WHANG (2005): “Consistent testing for stochastic dominance under general sampling schemes,” *The Review of Economic Studies*, 72, 735–765.
- LINTON, O., K. SONG, AND Y.-J. WHANG (2010): “An improved bootstrap test of stochastic dominance,” *Journal of Econometrics*, 154, 186–202.
- MCFADDEN, D. (1989): “Testing for stochastic dominance,” in *Studies in the economics of uncertainty: In honor of Josef Hadar*, Springer, 113–134.
- POLLARD, D. (2002): *A User’s Guide to Measure Theoretic Probability*, Cambridge University Press, New York.

- QU, Z. AND J. YOON (2019): “Uniform inference on quantile effects under sharp regression discontinuity designs,” *Journal of Business & Economic Statistics*, 37, 625–647.
- REISS, R.-D. (1989): *Approximate distributions of order statistics: with applications to nonparametric statistics*, Springer-Verlag, New York.
- SHEN, S. AND X. ZHANG (2016): “Distributional Tests for Regression Discontinuity: Theory and Empirical Examples,” *Review of Economics and Statistics*.