A Consistent Test for Stochastic Dominance Relations under Multi-way Clustering *

Yaqian Wu[a], Qingliang Fan[b], and Yu-Chin Hsu[c]

[a]School of Economics, Huazhong University of Science and Technology, Wuhan, China
[b]Department of Economics, The Chinese University of Hong Kong, Hong Kong
[c]Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University; CRETA, National Taiwan University, Taiwan

July 17, 2024

Abstract

Accounting for multi-way clustering is important when we make statistical inference. This paper constructs a new Kolmogorov-Smirnov test for stochastic dominance under multi-way clustered sampled data. We show that our test controls the size well asymptotically and is consistent against fixed alternatives. Our test is also robust to the degenerate cases. The simulation results support our theoretical findings. In an empirical study, we test the stochastic dominance relations between American income distributions in 2018 and 2022, considering the two-way clustering of states and industries.

Keywords: Stochastic dominance; Multi-way clustering; Bootstrap; Test consistency.
JEL codes: C12 C38.

*Emails: wuyaqian0818@163.com (Y. Wu), michaelqfan@gmail.com (Q. Fan), ychsu.econ@gmail.com (Y.-C. Hsu). Fan acknowledges financial support from the Research Grants Council of Hong Kong, GRF-14500223, 14501024.

1 Introduction

Stochastic dominance (SD) is an ordering rule of distribution functions, which has been explored in the fields of economics, finance, and insurance. Existing literature on stochastic dominance tests mainly focuses on the i.i.d. or weakly dependent data such as Anderson (1996), Davidson and Duclos (2000), McFadden (1989), Barrett and Donald (2003), Linton et al. (2005) and Donald and Hsu (2016) (DH hereafter). For a comprehensive review of the stochastic dominance tests, see Whang (2019). However, in real-world applications, the observations may exhibit multi-way clustering, i.e., the dependence of units along multiple cross-sectional dimensions. For instance, the wages of two individuals may exhibit correlation if they reside in the same metropolitan area or are employed in the same industrial sector. Overlooking such inherent correlations can result in issues with inference, specifically regarding size and power. Therefore, it is crucial to take into account the multi-way clustering in the tests. This paper proposes a new Kolmogorov-Smirnov (KS) test for stochastic dominance to handle multi-way clustered sampled data.

The wage data mentioned above is an example of two-way clustering. There has been a growing interest among economists in the study of multi-way cluster sampled data, such as network data, matched employer-employee data, matched student-teacher data, scanner data indexed by stores and products. In the realm of statistical inference, Cameron et al. (2011) propose a robust inference method tailored for multi-way clustered data. Menzel (2021) conducts analyses on the bootstrap validity under multi-way clustered sampled data. Davezies et al. (2021) develope an empirical process theory under multi-way clustered sampling, which can be applied to a large class of models. MacKinnon et al. (2021) introduce a wild bootstrap method and develope the asymptotic inference methods for multi-way clustering. Chiang et al. (2022) investigate double/debiased machine learning under multi-way clustered sampling environments. Chiang et al. (2023) propose a novel method of algorithmic subsampling for multi-way cluster-dependent data. Furthermore, MacKinnon et al. (2023a) offer a guide to practitioners for cluster-robust inference methods. MacKinnon et al. (2023b) propose a test for the correct level of clustering in regression models. Abadie et al. (2023) provide guidance on when and how to adjust standard errors for clustering using a novel framework for clustered inference on average treatment effects. Related to this practically important literature, we focus on the test for stochastic dominance under multi-way clustered data.

Building upon DH, who use a recentering method to improve the power of the test without resorting to the least favorable configuration, we propose a stochastic dominance test with improved size and power properties for multi-way clustering.[1] We first construct a KS test statistic tailored for multi-way clustering regarding the first-order stochastic dominance and show its limit distribution under the null. Second, we compute the critical values by adding the recentering function to the simulated processes. It is shown that our test can control the size well asymptotically and is consistent against fixed alternatives. Also, our test is robust to the degenerate cases which include the case where the data have no cross-sectional dependence. The simulation results demonstrate the good size and power performances of the proposed test. We also apply our test to assess the stochastic dominance relations between American income distributions in 2018 and 2022.

The remainder of the paper is organized as follows. Section 2 begins with a description of the multi-way clustered sampled data, followed by a detailed explanation of the testing procedure and the establishment of the asymptotic properties. Section 3 collects Monte Carlo simulation results to evaluate the finite sample performance. Section 4 provides several extensions of our test. In Section 5, we test stochastic dominance relations between American income distributions in 2018 and 2022. Finally, Section 6 concludes the paper. All technical proofs are relegated to the Appendix.

Notations. Throughout the paper, $\overset{d}{=}$ denotes identical distribution, $\xrightarrow{a.s.}$ denotes the convergence almost surely, $\xrightarrow{p}$ denotes the convergence in probability, $\xrightarrow{d}$ denotes the convergence in distribution, $\Rightarrow$ denotes the weak convergence, $\overset{p}{\Rightarrow}$ denotes the conditional weak conver-

[1] The recentering method is similar to various methods used in different papers such as Hansen (2005), Linton et al. (2010), Andrews and Soares (2010), Andrews and Shi (2013) and Andrews and Shi (2014).

gence in probability.[2] For any real numbers set $A = \{a_1, \cdots, a_n\}$, $\underline{A} = \min\{a_1, \cdots, a_n\}$.

2 A Test for Stochastic Dominance under Multi-way Clustering

We first introduce multi-way clustering in Section 2.1, then describe the null hypothesis and test statistic in Section 2.2. Section 2.3 develops the bootstrap procedure to simulate critical values. Section 2.4 establishes the asymptotic size and power properties of our test.

2.1 Multi-way Clustering

Multi-way clustering means that the observations may be correlated when they share the same cluster along multiple dimensions. A *cell* is the intersection of clusters in different dimensions. For simplicity, we first consider the two-way clustering where each cell contains one observation. We index the two dimensions by $i \in \{1, \cdots, N_1\}$ and $j \in \{1, \cdots, N_2\}$, where N_1 and N_2 are numbers of observations for cluster dimensions i and j, respectively. The cells are indexed by (i, j), and the random sample corresponding to cell (i, j) is denoted by Y_{ij}. This data structure is depicted in Table 1. Using again the example of individual wages, the first cluster dimension is the area of residence and the second cluster dimension is the industrial sector. The two-way clustering is defined formally in Assumption 1.

Table 1: Two-way clustering

	1	2	$\cdots$	N_2
1	Y_{11}	Y_{12}	$\cdots$	Y_{1N_2}
2	Y_{21}	Y_{22}	$\cdots$	Y_{2N_2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
N_1	$Y_{N_1 1}$	$Y_{N_1 2}$	$\cdots$	$Y_{N_1 N_2}$

[2]See van der Vaart and Wellner (1996, Chapter 3.6) for a precise definition of conditional weak convergence in probability.

Assumption 1 (Two-way clustering)**.** *Assume that*

1. $\{Y_{ij} : i = 1, \ldots, N_1 \text{ and } j = 1, \ldots, N_2\}$ *is separately exchangeable. Namely, for any two tuple permutations* (π_1, π_2),

$$\{Y_{ij}\}_{i,j} \overset{d}{=} \left\{Y_{\pi_1(i)\pi_2(j)}\right\}_{i,j};$$

2. *For any* $c_1, c_2 \geq 1$, $\{Y_{ij}\}_{i \leq c_1, j \leq c_2}$ *is independent of* $\{Y_{ij}\}_{i \geq c_1+1, j \geq c_2+1}$.

Assumption 1.1 states that the DGP is invariant by a relabelling of each dimension of clustering, that is, the joint distribution of the data remains identical under any two possible permutations of the labels. More generally, the data of all rows or columns are assumed to have the same distribution. For example, (Y_{11}, Y_{13}) has the same distribution as (Y_{22}, Y_{24}), and (Y_{11}, Y_{31}) has the same distribution as (Y_{22}, Y_{42}), but (Y_{11}, Y_{12}) does not have the same distribution as (Y_{11}, Y_{22}). This assumption is natural in many settings, except for time series. Assumption 1.2 imposes that any two blocks on the diagonal that do not overlap are independent. In particular, Y_{11} and Y_{22} are assumed to be independent, while Y_{11} and Y_{12} are not. Combined with Assumption 1.1, it implies that cells do not share the same rows or columns are mutually independent. Assumption 1 can also be found in Menzel (2021), Davezies et al. (2021), MacKinnon et al. (2021), Chiang et al. (2022) and Chiang et al. (2023).

By Aldous-Hoover-Kallenberg representation (Aldous, 1981; Hoover, 1979; Kallenberg, 1989), under Assumption 1, the random varaible Y_{ij} can be represented by $Y_{ij} = g\left(V_{i0}, V_{0j}, V_{ij}\right)$ for some Borel-measurable function g, where $\{V_{i0}\}_i, \{V_{0j}\}_j$, and $\{V_{ij}\}_{i,j}$ are mutually independent, and each of $\{V_{i0}\}_i, \{V_{0j}\}_j$, and $\{V_{ij}\}_{i,j}$ are i.i.d. We can interpret V_{i0}, V_{0j} and V_{ij} as ith clustering effect, jth clustering effect, and idiosyncratic effect, respectively. The data generating processes considered in our simulation studies is consistent with this representation.

2.2 Stochastic Dominance

Let X and Y be two random variables, and their CDF's are F_X and F_Y, respectively.

Assumption 2. *Suppose that*

1. $\mathcal{Z} = [0, \bar{z}]$, *where* $\bar{z} < \infty$;
2. F_X *and* F_Y *are continuous functions on* $\mathcal{Z}$ *such that* $F_X(z) = F_Y(z) = 0$ *iff* $z = 0$, *and* $F_X(z) = F_Y(z) = 1$ *iff* $z = \bar{z}$.

Assumption 2 requires $F_X(z)$ and $F_Y(z)$ to be continuous functions on $\mathcal{Z}$, which is generally assumed in the literature such as Barrett and Donald (2003) and DH. For simplicity, we assume the supports for $F_X(z)$ and $F_Y(z)$ are the same, but in practice they are allowed to be different as discussed in DH. To test if Y first-order stochastic dominates (SD1) X, we formulate our hypotheses as:

$$H_0 : F_Y(z) \leq F_X(z) \text{ for all } z \in \mathcal{Z}; \tag{1}$$

$$H_1 : F_Y(z) > F_X(z) \text{ for some } z \in \mathcal{Z}. \tag{2}$$

Let $\{Y_{ij} : i = 1, \ldots, N_1 \text{ and } j = 1, \ldots, N_2\}$ with sample size $N_1 N_2$ and $\{X_{ij} : i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2\}$ with sample size $M_1 M_2$ be random samples from distributions with CDF's F_Y and F_X, respectively. We make the following assumption regarding the sampling process.

Assumption 3. *Suppose that*

1. $\{Y_{ij} : i = 1, \ldots, N_1$ *and* $j = 1, \ldots, N_2\}$ *and* $\{X_{ij} : i = 1, \ldots, M_1$ *and* $j = 1, \ldots, M_2\}$ *satisfy Assumption 1, and are independent.*
2. *Denote* $\underline{N} = \min\{N_1, N_2\}$, $\underline{M} = \min\{M_1, M_2\}$, $\underline{N} \to \infty$, $\underline{M} \to \infty$. *For* $k = 1, 2$, $\underline{N}/N_k \to \beta_{y,k} \geq 0$, $\underline{M}/M_k \to \beta_{x,k} \geq 0$.

3. $\underline{M}$ *is a function of* $\underline{N}$ *satisfying that* $\underline{N}/\left(\underline{N}+\underline{M}\right) \to \lambda \in (0,1)$ *when* $\underline{N} \to \infty$*.*

Assumption 3.2 specifies the asymptotic framework where the sample size grows across all dimensions. The condition that $\underline{N}/N_k \to \beta_{y,k} \geq 0$ is a mild one since it allows for different convergence rates along the different clustering dimensions. Assumption 3.3 requires that $\underline{N}$ and $\underline{M}$ grow at the same rate. Henceforth, when taking limits as $\underline{N} \to \infty$, we assume that $\underline{M}(\underline{N})$ also goes to infinity.

Let the estimators for $F_Y(z)$ and $F_X(z)$ be

$$\widehat{F}_Y(z) = \frac{1}{N_1 N_2} \sum_{i,j} \mathbf{1}\left\{Y_{ij} \leq z\right\}, \quad \widehat{F}_X(z) = \frac{1}{M_1 M_2} \sum_{i,j} \mathbf{1}\left\{X_{ij} \leq z\right\},$$

where, $\mathbf{1}\{\cdot\}$ denotes the indicator function. We define KS test statistic tailored for two-way clustering as

$$\widehat{T} = \sqrt{\frac{\underline{M}\cdot\underline{N}}{\underline{M}+\underline{N}}} \sup_{z\in\mathcal{Z}} \left(\widehat{F}_Y(z) - \widehat{F}_X(z)\right).$$

Define $K(z_1, z_2) = \lambda K_x(z_1, z_2) + (1-\lambda) K_y(z_1, z_2)$ with

$$K_x\left(z_1, z_2\right) = \beta_{x,1}\mathrm{Cov}\left(\mathbf{1}\left\{X_{11} \leq z_1\right\}, \mathbf{1}\left\{X_{12} \leq z_2\right\}\right) + \beta_{x,2}\mathrm{Cov}\left(\mathbf{1}\left\{X_{11} \leq z_1\right\}, \mathbf{1}\left\{X_{21} \leq z_2\right\}\right),$$
$$K_y\left(z_1, z_2\right) = \beta_{y,1}\mathrm{Cov}\left(\mathbf{1}\left\{Y_{11} \leq z_1\right\}, \mathbf{1}\left\{Y_{12} \leq z_2\right\}\right) + \beta_{y,2}\mathrm{Cov}\left(\mathbf{1}\left\{Y_{11} \leq z_1\right\}, \mathbf{1}\left\{Y_{21} \leq z_2\right\}\right),$$

which are the covariance kernel of the limiting processes of $\sqrt{\underline{M}}(\widehat{F}_X(\cdot) - F_X(\cdot))$ and $\sqrt{\underline{N}}(\widehat{F}_Y(\cdot) - F_Y(\cdot))$, respectively.

Assumption 4. *Suppose that at lease one of* $K_x(z_1, z_2)$ *and* $K_y(z_1, z_2)$ *is a non-zero function.*

Assumption 4 implies that multi-way clustering presents in at least one of X and Y samples, i.e., we consider the non-degenerate case. Let $\mathcal{Z}^* = \{z \in \mathcal{Z} | F_Y(z) = F_X(z)\}$, we give the the limit distribution of $\widehat{T}$ in the following lemma.

Lemma 2.1. *Suppose that Assumptions 1-4 hold. Then,*

1. *Under H_0 in* (1), $\widehat{T} \xrightarrow{d} \sup_{z \in \mathcal{Z}^*} \mathbb{G}_K(z)$, *where,* $\mathbb{G}_K$ *is a mean zero Gaussian process with covariance kernel* $K(z_1, z_2)$.

2. *Under H_1 in* (2), $\widehat{T} \xrightarrow{p} \infty$.

The first part of Lemma 2.1 states that those z's with $F_Y(z) < F_X(z)$ will not contribute to the null distribution asymptotically. This result is similar to those in Barrett and Donald (2003) and DH.

2.3 Simulated Process

In practice, the distribution of $\mathbb{G}_K$ is non-pivotal since it depends on the true CDFs. Hence, the critical values cannot be tabulated directly. We refer to the pigeonhole bootstrap proposed by Davezies et al. (2021) to approximate the distribution of $\mathbb{G}_K$. The principle of the bootstrap for Y is:

1. N_1 elements are sampled with replacement and equal probability in the set $\{1, \ldots, N_1\}$. For each i in this set, let $W_i^{Y,1}$ denote the number of times i is selected this way.

2. N_2 elements are sampled with replacement and equal probability in the set $\{1, \ldots, N_2\}$. For each j in this set, let $W_j^{Y,2}$ denote the number of times j is selected this way.

3. Cell (i, j) is then selected $W_{ij}^Y = W_i^{Y,1} W_j^{Y,2}$ times in the bootstrap sample. By construction, any bootstrap sample consists of exactly $N_1 N_2$ cells.

Also, the bootstrap samples for X can be obtained by the similarly way. Then, we define the bootstrap counterpart of the empirical process $\mathbb{G}_K$ as:

$$\widehat{\mathbb{G}}_K(z) = \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \widehat{D}_{\underline{N}}(z),$$

where,

$$\widehat{D}_{\underline{N}}(z) = \left(\frac{1}{N_1 N_2} \sum_{ij} W_{ij}^Y \mathbf{1}\{Y_{ij} \leq z\} - \widehat{F}_Y(z) \right) - \left(\frac{1}{M_1 M_2} \sum_{ij} W_{ij}^X \mathbf{1}\{X_{ij} \leq z\} - \widehat{F}_X(z) \right).$$

The critical values constructed directly by $\widehat{\mathbb{G}}_K$ may lead to an asymptotically conservative test. Here, we use the recentering method in DH to overcome this problem. Define $\mu(z) = \min\{F_Y(z) - F_X(z), 0\}$. Obviously, we have $\mu(z) = F_Y(z) - F_X(z)$ under H_0. Therefore, the limit null distribution of $\widehat{T}$ is the same as that of

$$\widehat{T}_R = \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \sup_{z \in \mathcal{Z}} \left[\left(\widehat{F}_Y(z) - \widehat{F}_X(z) \right) - (F_Y(z) - F_X(z)) + \mu(z) \right].$$

We introduce the sample analogue of $\mu(z)$ in the resampling procedure to compute the critical values.

Assumption 5. *Let $\delta_{\underline{N}}$ be a sequence of negative numbers such that $\lim_{\underline{N} \to \infty} \delta_{\underline{N}} = -\infty$ and $\lim_{\underline{N} \to \infty} \underline{N}^{-1/2} \delta_{\underline{N}} = 0$,*

Define the recentering function as

$$\widehat{\mu}_{\underline{N}}(z) = \left(\widehat{F}_Y(z) - \widehat{F}_X(z) \right) \cdot \mathbf{1} \left\{ \sqrt{\underline{N}} \left(\widehat{F}_Y(z) - \widehat{F}_X(z) \right) < \delta_{\underline{N}} \right\}.$$

Let α be the nominal level, define the critical values $\widehat{c}_{1-\alpha,\eta}$ as

$$\widehat{c}_{1-\alpha,\eta} = \max \left\{ \widehat{c}_{1-\alpha,\underline{N}}, \eta \right\},$$
$$\widehat{c}_{1-\alpha,\underline{N}} = \sup_c \left\{ P_b \left(\sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \sup_{z \in \mathcal{Z}} \left(\widehat{D}_{\underline{N}}(z) + \widehat{\mu}_{\underline{N}}(z) \right) \leq c \right) \leq 1 - \alpha \right\},$$

where, η is an arbitrarily small positive number, P_b is the probability with respect to bootstrap sampling. We choose $\eta = 0$ and $\delta_{\underline{N}} = -0.1\sqrt{\log\log(\underline{N} + \underline{M})}$ in the simulations.

2.4 Asymptotic Size and Power Properties

This section investigates the pointwise asymptotic size and power properties of our test. Firstly, we consider the asymptotic size of the test.

Theorem 2.2. *Suppose that Assumptions 1-5 hold, and $\alpha < 1/2$. If we reject H_0 when $\widehat{T} > \widehat{c}_{1-\alpha,\eta}$, then under H_0 in* (1), *the following statements are true:*

1. *If $F_Y(z) = F_X(z)$ and $K(z,z) > 0$ for some $z \in (0, \bar{z})$, then $\lim_{\eta \to 0} \lim_{\underline{N} \to \infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) = \alpha$;*

2. *Otherwise, $\lim_{\underline{N} \to \infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) = 0$.*

The first part of Theorem 2.2 shows that the asymptotic size will be close to the pre-specified significance level α when the null is at a point on the boundary such that $F_Y(z) = F_X(z)$ for some $z \in (0, \bar{z})$ and $K(z,z) > 0$ for these z's. The second part of this theorem asserts that the asymptotic size will be equal to zero when either (a), $F_Y(z) = F_X(z)$ for some $z \in (0, \bar{z})$ but $K(z,z) = 0$ for these z's, or (b), the null is an interior point such that $F_Y(z) < F_X(z)$ for all $z \in (0, \bar{z})$. As we show in Appendix C, in both two cases, the limiting distribution of test statistic $\widehat{T}$ is degenerate at 0, while the critical value $\widehat{c}_{1-\alpha,\eta}$ converges to a small positive number η, which substantiates the result.

Theorem 2.3. *Suppose that Assumptions 1-5 hold, and $\alpha < 1/2$. Then, under the fixed alternative hypothesis defined in* (2), $\lim_{\underline{N} \to \infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) = 1$.

Theorem 2.3 shows that our test is consistent against any fixed alternatives. The second part of Lemma 2.1 gives the divergence of the test statistic $\widehat{T}$ under a fixed alternative, while the critical value $\widehat{c}_{1-\alpha,\eta}$ is bounded in probability. Thus, we get the result in this theorem.

Remark 1 (Degenerate Case)**.** It is worth mentioning that our test is robust against the potential degeneracy. Degeneracy refers to the situation where asymptotic covariance kernels K_x and K_y are both zero functions when no units share the same cluster, such as the i.i.d. sampled data. Suppose that both K_x and K_y are zero functions. In this case, by Theorem 3.4 of Davezies et al. (2021), we can show that under the null, the test statistic $\widehat{T}$ will converge to zero and the the critical value $\widehat{c}_{1-\alpha,\eta}$ converges to a small positive number η, so the asymptotic size is zero for any fixed $\eta > 0$. On the other hand, under any fixed alternative, the test statistic $\widehat{T}$ will diverge to infinity and the critical value still converges to a small positive number η, so we obtain the consistency of our test. Even though our test is robust to the degenerate cases, the cost of such robustness is that our test is more conservative and less powerful. For example, when the data are i.i.d., the tests proposed in DH are less conservative and more powerful than ours and we can see this result in the simulations.

Remark 2 (Heterogeneous Sample Size in Different Cells)**.** Although the current section predominantly discuss the scenario where each cell contains a single observation, the theoretical results presented herein are also applicable to scenarios with heterogeneous sample sizes across different cells. A detailed exposition on managing such heterogeneity is postponed to Section 4.2. Our empirical data embodies a two-way clustering framework, wherein different cells exhibit distinct sample size.

3 Simulations

In this section, we run simulations to examine the finite-sample size and power properties of our test. We denote our test as "WFH" and compare our test with the methods in DH. The algorithm of DH test is briefly introduced in the Appendix D. For our method, we pick $\delta_{\underline{N}} = -0.1\sqrt{\log\log(\underline{N}+\underline{M})}$ and $\eta = 0$. We consider the two-way balanced design $M_1 = M_2 = N_1 = N_2 = \underline{M} = \underline{N}$ with one observation per cell. We consider the two-way clustered sampled data (non-degenerate cases) in the first two examples (where the DGP is the same as in Section 2.1) and the degenerate cases in the last two example. In each simulation case, the rejection rate is estimated by 500 Monte-Carlo simulations and in each Monte-Carlo repetition, the p-values of various tests are approximated based on 500 bootstrapped samples. The significance level α is set at 5%.

Example 1: We consider non-degenerate two-way clustered sampled data case. We first consider the size cases. Each X_{ij} is drawn in a standard Gaussian distribution, but the variance due to cell shocks only represents 60% of the total variance, whereas row and column shocks represent 20% of the variance each. Let $\mu_y \geq 0$; we generate the samples Y_{ij} in a similar way.

$$
\begin{aligned}
X_{ij} =& \frac{1}{\sqrt{5}} \left(U_{i0} + U_{0j} + \sqrt{3} U_{ij} \right), \quad U_{i0}, U_{0j}, U_{ij} \sim \mathcal{N}\left(0, I_3\right), \\
Y_{ij} =& \frac{1}{\sqrt{5}} \left(V_{i0} + V_{0j} + \sqrt{3} V_{ij} \right), \quad V_{i0}, V_{0j}, V_{ij} \sim \mathcal{N}\left(\mu_y, I_3\right),
\end{aligned}
$$

where, I_3 is a 3×3 identity matrix. When $\mu_y = 0$, $F_Y(z) = F_X(z)$ for all z. When $\mu_y > 0$, $F_Y(z) \leq F_X(z)$ for all z and the equality holds only when $F_Y(z) = F_X(z) = 0$ and $F_Y(z) = F_X(z) = 1$, which implies that Y SD1 X, and we use this example to illustrate the size properties of our test. We consider μ_y from $\{0, 0.1, 0.2\}$ and vary $\underline{N}$ from $\{20, 30, 50, 60\}$. The smaller the μ_y is, the closer the distribution of X and Y is. We report the rejection rates for each example and μ_y in Table 2.

Table 2: Rejection rates of Example 1

		$\underline{N}$			
μ_y		20	30	50	60
0	WFH	0.032	0.038	0.040	0.045
	DH	0.328	0.356	0.374	0.386
0.1	WFH	0	0	0	0
	DH	0.108	0.072	0.048	0.038
0.2	WFH	0	0	0	0
	DH	0.062	0.050	0.002	0

Table 2 illustrates that the DH test tends to be over-sized under the two-way clustering, especially when the distributions of X and Y are identical, which emphasizes the necessity for proposing a new test that can handle multi-way clustered sampled data. When $F_Y(z) = F_X(z)$ for all z, the size of our test approaches the pre-specified significance level 5% as the sample size increases. When $F_Y(z) = F_X(z)$ holds only at the points such that $F_Y(z) = F_X(z) = 0$ and $F_Y(z) = F_X(z) = 1$, the sizes of our method remain at 0, which is consistent with Theorem 2.2.

Example 2: We consider non-degenerate two-way clustered sampled data case. We consider the power cases. Let $\sigma_y > 0$, we generate the samples Y_{ij} by

$$Y_{ij} = \frac{1}{\sqrt{5}}\left(V_{i0} + V_{0j} + \sqrt{3}V_{ij}\right), \quad V_{i0}, V_{0j}, V_{ij} \sim \mathcal{N}\left(0, \sigma_y I_3\right).$$

where, I_3 is a 3×3 identity matrix. We set σ_y from $\{0.5, 1.5, 1.75, 2\}$ and vary $\underline{N}$ from $\{20, 30, 50, 60\}$. When $\sigma_y < 1$, $F_Y(z) < F_X(z)$ for all $z < 0$ and $F_Y(z) > F_X(z)$ for all $z > 0$. When $\sigma_y > 1$, $F_Y(z) > F_X(z)$ for all $z < 0$ and $F_Y(z) < F_X(z)$ for all $z > 0$. The closer the σ is to 1, the closer the distribution of X and Y is. We set the significance level as 5% and report the rejection rates for each example and each σ_y in Table 3. It is observed that the power of our test approaches 1 as the sample size increases. The power of DH test consistently exceeds that of ours, attributable to the fact that their test statistic is inherently larger and has a more rapid growth rate as the sample size grows. Furthermore, the greater the distance between the distributions of X and Y, the higher the power.

Table 3: Rejection rates of Example 2

		$\underline{N}$			
σ_y		20	30	50	60
0.5	WFH	0.432	0.796	0.992	1
	DH	0.994	1	1	1
1.5	WFH	0.118	0.232	0.516	0.603
	DH	0.786	0.952	0.998	1
1.75	WFH	0.254	0.518	0.872	0.966
	DH	0.944	1	1	1
2	WFH	0.408	0.796	0.992	1
	DH	0.988	0.998	1	1

Example 3: In this case, we consider the degenerate case when the null is true. We generate the i.i.d. sampled data as in DH. Given a constant $l \in (0, 1)$, the two samples are generated

by

$$
\begin{aligned}
X_{ij} &= 1(\mathcal{U}_{ij} \leq l)\frac{\mathcal{U}_{ij}^2}{l} + 1(\mathcal{U}_{ij} > l)\mathcal{U}_{ij}, \\
Y_{ij} &= \mathcal{V}_{ij},
\end{aligned}
$$

where, $\mathcal{U}_{ij}$ and $\mathcal{V}_{ij}$ are independent uniform distributions over $[0, 1]$. In this case, $F_Y(z) \leq F_X(z)$ for all z and $F_Y(z) < F_X(z)$ for $z \in (0, l)$. Thus, this case is used to compare the size properties of different methods in the degenerate case. We consider $l \in \{0.1, 0.5\}$ and $\underline{N} \in \{20, 30\}$. The rejection rates are reported in the upper panel in Table 4. As depicted in Table 4, the sizes of our test are always 0 under the degenerate case, aligning with the theoretical results detailed in the Remark 1 of Section 2.4.

Example 4: In this case, the null is false and the sampled data is i.i.d., and we aim to compare the power properties of our test and DH under the degeneracy. Given $l \in (0, 1)$, the two samples are generated by

$$
\begin{aligned}
X_{ij} &= \mathbf{1}\left\{\mathcal{U}_{ij} \leq l\right\}\frac{\mathcal{U}_{ij}^2}{l} + \mathbf{1}\left\{\mathcal{U}_{ij} > l\right\}\mathcal{U}_{ij}, \\
Y_{ij} &= \mathbf{1}\left\{\mathcal{V}_{ij} \leq l\right\}\mathcal{V}_{ij} + \mathbf{1}\left\{\mathcal{V}_{ij} > l\right\}\left(l + \frac{(\mathcal{V}_{ij} - l)^2}{1 - l}\right),
\end{aligned}
$$

where, $\mathcal{U}_{ij}$ and $\mathcal{V}_{ij}$ are independent uniform distributions over $[0, 1]$. In this case, $F_Y(z) < F_X(z)$ for $z \in (0, l)$ and $F_Y(z) > F_X(z)$ for $z \in (l, 1)$. We consider $l \in \{0.1, 0.5\}$ and $\underline{N} \in \{20, 30\}$. The rejection rates are reported in the lower panel of Table 4. It is observed that the power of our test tends to 1 as the sample size increases, which implies the consistency of our test. The power of DH test consistently exceeds ours. The results in Example 3 and Example 4 support the fact that under i.i.d. sampling and the null, our test will be more conservative than DH, and under alternatives, even our test is still consistent, but ours is usually less powerful than DH.

Table 4: Rejection rates of Example 3 and Example 4

		N			
		20		30	
		$l = 0.1$	0.5	0.1	0.5
Example 3	WFH	0	0	0	0
	DH	0.066	0.026	0.042	0.044
Example 4	WFH	1	0.604	1	0.986
	DH	1	0.986	1	1

4 Extensions

In this section, we extend our test to higher order stochastic dominance and general K-way clustered sampled data.

4.1 Testing for Higher Order Stochastic Dominance

To give a general form of various orders of stochastic dominance, let $\mathscr{F}_q(\cdot; F)$ be the function that integrates the function F to order $q-1$, so that

$$\begin{aligned}\mathscr{F}_1(z;F) =& F(z),\ \mathscr{F}_2(z;F) = \int_0^z F(t)dt = \int_0^z \mathscr{F}_1(t;F)dt,\ \cdots,\\ \mathscr{F}_q(z;F) =& \int_0^z \mathscr{F}_{q-1}(t;F)dt.\end{aligned}$$

Then stochastic dominance of order q (SDq) of Y over X corresponds to $\mathscr{F}_q(z;F_Y) \leq \mathscr{F}_q(z;F_X)$ for all z. To test if Y SDq X, the hypotheses can be formulated as

$$\begin{aligned}H_0^q &: \mathscr{F}_q(z;F_Y) \leq \mathscr{F}_q(z;F_X) \quad \text{for all } z \in \mathcal{Z},\\ H_1^q &: \mathscr{F}_q(z;F_Y) > \mathscr{F}_q(z;F_X) \quad \text{for some } z \in \mathcal{Z}.\end{aligned}$$

It is well known that SDq implies SD$(q+1)$, but the converse is not true. The test statistic for testing the higher order stochastic dominance can be written as:

$$\widehat{T}_q = \sqrt{\frac{\underline{M}\cdot\underline{N}}{\underline{M}+\underline{N}}}\sup_{z\in\mathcal{Z}}\left(\mathscr{F}_q(z;\widehat{F}_Y)-\mathscr{F}_q(z;\widehat{F}_X)\right),$$

where, one can show that

$$\mathscr{F}_q\left(z;\widehat{F}_Y\right) = \frac{1}{N_1N_2}\sum_{ij}\mathscr{F}_q\left(z;\mathbf{1}\left\{Y_{ij}\le z\right\}\right) = \frac{1}{N_1N_2}\sum_{ij}\frac{1}{(q-1)!}\mathbf{1}\left\{Y_{ij}\le z\right\}\left(z-Y_{ij}\right)^{q-1}.$$

Define the recentering function as $\mu_q = \min\left\{\mathscr{F}_q\left(z;F_Y\right)-\mathscr{F}_q\left(z;F_X\right),0\right\}$ and

$$\widehat{\mu}_{q,\underline{N}} = \left(\mathscr{F}_q(z;\widehat{F}_Y)-\mathscr{F}_q(z;\widehat{F}_X)\right)\mathbf{1}\left\{\sqrt{\underline{N}}\left(\mathscr{F}_q(z;\widehat{F}_Y)-\mathscr{F}_q(z;\widehat{F}_X)\right)<\delta_{\underline{N}}\right\}.$$

Let $\widehat{D}_{q,\underline{N}} = \mathscr{F}_q\left(z;\widehat{D}_{\underline{N}}(z)\right)$ and α be the significance level. Define the critical value as

$$\begin{aligned}\widehat{c}_{q,1-\alpha,\eta} &= \max\left\{\widehat{c}_{q,1-\alpha,\underline{N}},\eta\right\},\\ \widehat{c}_{q,1-\alpha,\underline{N}} &= \sup\left\{c \mid P_b\left(\sqrt{\frac{\underline{M}\cdot\underline{N}}{\underline{M}+\underline{N}}}\sup_{z\in\mathcal{Z}}\left(\widehat{D}_{q,\underline{N}}(z)+\widehat{\mu}_{q,\underline{N}}(z)\right)\le c\right)\le 1-\alpha\right\}.\end{aligned}$$

Then, we reject the null hypothesis H_0^q when $\widehat{T}_q > \widehat{c}_{q,1-\alpha,\underline{N}}$.

Define $K_q(z_1,z_2) = \lambda K_{q,x}(z_1,z_2)+(1-\lambda)K_{q,y}(z_1,z_2)$ with

$$\begin{aligned}K_{q,x}\left(z_1,z_2\right) = \frac{1}{[(q-1)!]^2}&\left\{\beta_{x,1}\mathrm{Cov}\left(\mathbf{1}\left\{X_{11}\le z_1\right\}\left(z_1-X_{11}\right)^{q-1},\mathbf{1}\left\{X_{12}\le z_2\right\}\left(z_2-X_{12}\right)^{q-1}\right)\right.\\ &\left.+\beta_{x,2}\mathrm{Cov}\left(\mathbf{1}\left\{X_{11}\le z_1\right\}\left(z_1-X_{11}\right)^{q-1},\mathbf{1}\left\{X_{21}\le z_2\right\}\left(z_2-X_{21}\right)^{q-1}\right)\right\},\\ K_{q,y}\left(z_1,z_2\right) = \frac{1}{[(q-1)!]^2}&\left\{\beta_{y,1}\mathrm{Cov}\left(\mathbf{1}\left\{Y_{11}\le z_1\right\}\left(z_1-Y_{11}\right)^{q-1},\mathbf{1}\left\{Y_{12}\le z_2\right\}\left(z_2-Y_{12}\right)^{q-1}\right)\right.\\ &\left.+\beta_{y,2}\mathrm{Cov}\left(\mathbf{1}\left\{Y_{11}\le z_1\right\}\left(z_1-Y_{11}\right)^{q-1},\mathbf{1}\left\{Y_{21}\le z_2\right\}\left(z_2-Y_{21}\right)^{q-1}\right)\right\},\end{aligned}$$

which are covariance kernels of the limiting processes of $\sqrt{\underline{M}}[\mathscr{F}_q(z;\widehat{F}_X(\cdot))-\mathscr{F}_q(z;F_X(\cdot))]$

and $\sqrt{\underline{N}}[\mathscr{F}_q(z;\widehat{F}_Y(\cdot))-\mathscr{F}_q(z;F_Y(\cdot))]$, respectively.

Assumption 6. *Suppose that at lease one of $K_{q,x}(z_1,z_2)$ and $K_{q,y}(z_1,z_2)$ is a non-zero function.*

Similar to the proofs of Theorems 2.2 and 2.3, we can show that our test for higher order stochastic dominance have the same size and consistency properties as the test for SD1.

Theorem 4.1. *Suppose that Assumptions 1-3 and 5-6 hold, and $\alpha < 1/2$. If we reject H_0 when $\widehat{T}_q > \widehat{c}_{q,1-\alpha,\eta}$, then the following statements are true:*

1. *Under H_0^q, if $\mathscr{F}_q(z;F_Y) = \mathscr{F}_q(z;F_X)$ and $K_q(z,z) > 0$ for some $z \in (0,\bar{z})$, then $\lim_{\eta\to 0}\lim_{\underline{N}\to\infty} P\left(\widehat{T}_q > \widehat{c}_{q,1-\alpha,\eta}\right) = \alpha$; otherwise, $\lim_{\underline{N}\to\infty} P\left(\widehat{T}_q > \widehat{c}_{q,1-\alpha,\eta}\right) = 0$ for any fixed $\eta > 0$.*
2. *Under H_1^q, $\lim_{\eta\to 0}\lim_{\underline{N}\to\infty} P\left(\widehat{T}_q > \widehat{c}_{q,1-\alpha,\eta}\right) = 1$.*

4.2 Stochastic Dominance under K-way Clustering

We can extend our test to K-way ($K > 2$) clustering where the numbers of observations are heterogeneous across cells. Here, we index cells by the K-tuple $\boldsymbol{j} = (j_1,\cdots,j_K)$ for $j_k = 1,\cdots,N_k$, where N_k is the number of clusters in the sample for dimension k, $k = 1,\cdots,K$. For each cell, the number of observations within cell is denoted by $n_{\boldsymbol{j}}$, the random sample corresponding to unit ℓ in cell $\boldsymbol{j}$ is then denoted $Y_{\ell,\boldsymbol{j}}$, $\ell = 1,\cdots,n_{\boldsymbol{j}}$. We let $\boldsymbol{j} \geq \mathbf{1}$ to mean that $j_k \geq 1$ for all $k = 1,\cdots,K$. The key assumptions of this K-way clustering are the following.

Assumption 7. *Assume that:*

1. *The array $\left(n_{\boldsymbol{j}},(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right)_{\boldsymbol{j}\geq\mathbf{1}}$ is separately exchangeable. Namely, for any K-tuple of permutations $(\pi_1,\ldots,\pi_K)$,*

$$\left(n_{\boldsymbol{j}},(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right)_{\boldsymbol{j}\geq\mathbf{1}} \stackrel{d}{=} \left(n_{\pi_1(j_1),\ldots,\pi_K(j_K)},\left(Y_{\ell,\pi_1(j_1),\ldots,\pi_K(j_K)}\right)_{\ell\geq 1}\right)_{\boldsymbol{j}\geq\mathbf{1}}.$$

2. *For any* $\boldsymbol{c} \geq \mathbf{1}$, $\left(n_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right)_{\mathbf{1}\leq \boldsymbol{j}\leq \boldsymbol{c}}$ *is independent of* $\left(n_{\boldsymbol{j}'}, (Y_{\ell,\boldsymbol{j}'})_{\ell\geq 1}\right)_{\boldsymbol{j}'\geq \boldsymbol{c}+\mathbf{1}}$.

3. $E\left(n_{\mathbf{1}}\right) > 0$.

Assumption 7 is a generalization of Assumption 1. Assumption 7.3 only excludes arrays that are almost surely empty. It is worth pointing out that Assumption 7.2 does not impose any restriction on the distribution of $\left(n_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right)$. Hence, the dependence between $n_{\boldsymbol{j}}$ and the $(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}$, and the dependence between the $(Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}$ within cell $\boldsymbol{j}$, are left unrestricted. This implies, for instance, that conditional on $n_{\boldsymbol{j}}$, the correlation between $Y_{\ell,\boldsymbol{j}}$ and $Y_{\ell',\boldsymbol{j}}$ may vary with $N_{\boldsymbol{j}}$. In this sense, we allow for cluster heterogeneity, as defined by Carter et al. (2017). Also, $Y_{\ell,\boldsymbol{j}}$ may have a different distribution from $Y_{\ell',\boldsymbol{j}}$, for $\ell \neq \ell'$.

Suppose we have another data set denoted as $\left(m_{\boldsymbol{i}}, (X_{\ell,\boldsymbol{i}})_{\ell\geq 1}\right)_{\boldsymbol{i}\geq \mathbf{1}}$, where, $\boldsymbol{i} = (i_1, \cdots, i_K)$ for $i_k = 1, \cdots, M_k$ is the cell index, M_k denotes the number of clusters in the sample for dimension k, $k = 1, \cdots, K$, $m_{\boldsymbol{i}}$ is the number of observations within cell $\boldsymbol{i}$.

Assumption 8. *Assume that:*

1. $\left(n_{\boldsymbol{j}}, (Y_{\ell,\boldsymbol{j}})_{\ell\geq 1}\right)_{\boldsymbol{j}\geq \mathbf{1}}$ *and* $\left(m_{\boldsymbol{i}}, (X_{\ell,\boldsymbol{i}})_{\ell\geq 1}\right)_{\boldsymbol{i}\geq \mathbf{1}}$ *are two independent random samples from distributions with CDF's* F_Y *and* F_X*, respectively and satify Assumption 7.*

2. *Define* $\underline{N} = \min\{N_1, \cdots, N_K\} \to \infty$ *and* $\underline{M} = \min\{M_1, \cdots, M_K\} \to \infty$ *for all* $k = 1 \ldots K$, $\underline{N}/N_k \to \beta_{y,k} \geq 0$, $\underline{M}/M_k \to \beta_{x,k} \geq 0$.

3. $\underline{M}$ *is a function of* $\underline{N}$ *satisfying that* $\underline{N}/\left(\underline{N} + \underline{M}\right) \to \lambda \in (0, 1)$ *when* $\underline{N} \to \infty$.

We consider the unit-level distributions and estimate them by

$$\widehat{F}_Y(z) = \frac{\sum_{\boldsymbol{j}} \sum_{\ell=1}^{n_{\boldsymbol{j}}} \mathbf{1}\left\{Y_{\ell,\boldsymbol{j}} \leq z\right\}}{\sum_{\boldsymbol{j}} n_{\boldsymbol{j}}}, \quad \widehat{F}_X(z) = \frac{\sum_{\boldsymbol{i}} \sum_{\ell=1}^{m_{\boldsymbol{i}}} \mathbf{1}\left\{X_{\ell,\boldsymbol{i}} \leq z\right\}}{\sum_{\boldsymbol{i}} m_{\boldsymbol{i}}}.$$

Then, to test if Y SD1 X, we use the KS test statistic:

$$\widehat{T} = \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \sup_{z\in\mathcal{Z}} \widehat{F}_Y(z) - \widehat{F}_X(z).$$

The testing procedure for first-order stochastic dominance under K-way $(K > 2)$ clustering is almost the same as in the previous two-way clustering case. Here, we only point out what is different between the two. First, under K-way clustering, the covariance kernel is $K(z_1, z_2) = \lambda K_x(z_1, z_2) + (1-\lambda)K_y(z_1, z_2)$ with

$$K_x(z_1, z_2) = \sum_{i=1}^{K} \beta_{x,k} \mathrm{Cov}\left(\sum_{\ell=1}^{m_{\mathbf{1}}} \mathbf{1}\left\{X_{\ell,\mathbf{1}} \le z_1\right\}, \sum_{\ell=1}^{m_{\mathbf{1}}} \mathbf{1}\left\{X_{l,\mathbf{2}_i} \le z_2\right\}\right),$$
$$K_y(z_1, z_2) = \sum_{i=1}^{K} \beta_{y,k} \mathrm{Cov}\left(\sum_{\ell=1}^{n_{\mathbf{1}}} \mathbf{1}\left\{Y_{\ell,\mathbf{1}} \le z_1\right\}, \sum_{\ell=1}^{n_{\mathbf{1}}} \mathbf{1}\left\{Y_{l,\mathbf{2}_i} \le z_2\right\}\right),$$

where $\mathbf{2}_i$ is the k-tuple with 2 in each entry except entry i, which is 1. Second, under K-way clustering, the principle of the pigeonhole bootstrap for Y is:

1. For each $k \in \{1, \ldots, K\}$, N_k elements are sampled with replacement and equal probability in the set $\{1, \ldots, N_k\}$. For each j_k in this set, let $W_{j_k}^{Y,k}$ denote the number of times j_k is selected this way.

2. Cell $\boldsymbol{j} = (j_1, \ldots, j_K)$ is then selected $W_{\boldsymbol{j}}^{Y} = \prod_{k=1}^{K} W_{j_k}^{Y,k}$ times in the bootstrap sample. By construction, any bootstrap sample consists of exactly $\Pi_{k=1}^{K} N_k$ cells.

By similar arguments as the two-way clustering case, we can show that, under Assumptions 2, 4, 5, 7 and 8, the results in Section 2.4 hold under K-way clustering. It is also possible to extend to test higher order stochastic dominance under K-way clustering. In addition, Remark 1 after Theorem 2.3 applies here.

5 Empirical Analysis

In this section, we revisit a long-standing empirical question in labor economics, namely, the stochastic dominance relations between income distributions. The wages of two individuals may be correlated because they live in the same metropolitan area or work in the same industrial sector. There are abundant studies using empirical data that are clustered at industrial or geographical levels (Hersch, 1998; Kolko and Neumark, 2010; Cameron et al., 2011; MacKinnon et al., 2023a).

The individual income data from the American Community Survey (ACS) is obtained from IPUMS (Ruggles et al., 2024) for the years 2018 and 2022, which covers all 50 U.S. states and the District of Columbia, as well as 269 industries. The number of individuals is heterogeneous across the intersections of states and industries. For each individual, we record the type of industry in which they reported an occupation, the specific state in which they were employed, and their total pre-tax wage and salary income for the previous year. We also restrict attention to individuals who identify as black. Table 5 gives basic descriptive statistics for these data. Figure 1 plots the empirical CDF for the income data. We compare the income distributions in 2018 and 2022 using our method.

Table 5: Description statistics

		2018	2022
Income	Sample	78740	75585
	Mean	4.19	4.92
	s.d.	4.79	5.65
	Median	3.00	3.60
State	Num	51	51
	s.d.	1660.68	1607.70
Industry	Num	269	269
	s.d.	464.61	446.72

NOTE: the income is measured in 10,000 US dollars, the “Num” in the panel “State” is the number of different states, the “s.d.” in the panel “State” is the standard deviation for the number of individuals in different states. Similarly, the “Num” in the panel “Industry” is the number of different industries, the “s.d.” in the panel “Industry” is the standard deviation for the number of individuals in different industries.

We set $\delta_{\underline{N}} = -0.1\sqrt{\log\log 102} \approx -0.12$, $\eta = 0$ and approximate the p-value by 500 replications. We conduct first, second, and third-order stochastic dominance tests and summarize the p-values in Table 6. The panel labeled “2022 versus 2018” presents the p-values for testing whether the 2022 income distribution stochastically dominates the 2018 income distribution for the specific order, and the other panel presents the results for opposite hypotheses. In Table 6, we find that, for the black population, our test indicates that the income distribution of 2022 stochastically dominates that of 2018 for the orders 1-3.[3]

Table 6: p values of stochastic dominance test

	2022 versus 2018			2018 versus 2022		
	SD1	SD2	SD3	SD1	SD2	SD3
WFH	0.594	0.474	0.490	0.004	0.002	0.002

6 Conclusion

This paper constructs a new KS test for stochastic dominance to handle multi-way clustered sampled data. We show that our test has the correct size control asymptotically and is consistent against fixed alternatives. The results from simulations illustrate the good size and power performances of our test. The empirical study suggests that the American income distribution of 2022 stochastically dominates that of 2018.

[3]We have also conducted DH tests, the p values for "2022 versus 2018" are 0.016, 0.210 and 0.326, respectively. The p values for "2018 versus 2022" are all 0 (rounded to 3 decimal places). The smaller p values again indicate the findings in the simulations. Under multi-way clustering, the DH tests tend to be oversized under the null and more powerful under alternative.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis 11*(4), 581–598.

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica: Journal of the Econometric Society 64*(5), 1183–1193.

Andrews, D. W. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica 81*(2), 609–666.

Andrews, D. W. and X. Shi (2014). Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics 179*(1), 31–45.

Andrews, D. W. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Barrett, G. F. and S. G. Donald (2003). Consistent tests for stochastic dominance. *Econometrica 71*(1), 71–104.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Carter, A. V., K. T. Schnepel, and D. G. Steigerwald (2017). Asymptotic behavior of at-test robust to cluster heterogeneity. *Review of Economics and Statistics 99*(4), 698–709.

Chiang, H. D., K. Kato, Y. Ma, and Y. Sasaki (2022). Multiway cluster robust double/debiased machine learning. *Journal of Business & Economic Statistics 40*(3), 1046–1056.

Chiang, H. D., J. Li, and Y. Sasaki (2023). Algorithmic subsampling under multiway clustering. *Econometric Theory*, 1–44.

Davezies, L., X. D'Haultfoeuille, and Y. Guyonvarch (2021). Empirical process results for exchangeable arrays. *The Annals of Statistics 49*(2), 845 – 862.

Davidson, R. and J.-Y. Duclos (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica 68*(6), 1435–1464.

Donald, S. and Y.-C. Hsu (2016). Improving the power of tests of stochastic dominance. *Econometric Reviews 35*, 1–33.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics 23*(4), 365–380.

Hersch, J. (1998). Compensating differentials for gender-specific job injury risks. *The American Economic Review 88*(3), 598–607.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Working paper*.

Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. *Journal of Multivariate Analysis 30*(1), 137–154.

Kolko, J. and D. Neumark (2010). Does local business ownership insulate cities from economic shocks? *Journal of Urban Economics 67*(1), 103–115.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies 72*(3), 735–765.

Linton, O., K. Song, and Y.-J. Whang (2010). An improved bootstrap test of stochastic dominance. *Journal of Econometrics 154*(2), 186–202.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2021). Wild bootstrap and asymptotic inference with multiway clustering. *Journal of Business & Economic Statistics 39*(2), 505–519.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023a). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics 232*(2), 272–299.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023b). Testing for the appropriate level of clustering in linear regression models. *Journal of Econometrics 235*(2), 2027–2056.

McFadden, D. (1989). Testing for stochastic dominance. In *Studies in the economics of uncertainty: In honor of Josef Hadar*, pp. 113–134. Springer.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica 89*(5), 2143–2188.

Ruggles, S., S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rodgers, and M. Schouweiler (2024). Ipums usa: Version 15.0 [dataset]. minneapolis, mn: Ipums.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer.

Whang, Y.-J. (2019). *Econometric analysis of stochastic dominance*. Cambridge University Press.

A Auxiliary Lemmas

Before giving the formal proof for the theorems in this paper, we present a few auxiliary lemmas. Define $\widehat{A}_{\underline{N}}(z) = \widehat{F}_Y(z) - \widehat{F}_X(z)$, $A(z) = F_Y(z) - F_X(z)$ and $\mathcal{Z}^+ = \{z \mid A(z) \geq 0\}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space equipped with probability distribution $\mathbb{P}$. For a random variable V, let $V(\omega)$ be the conditional distribution of V conditional on $\omega \in \Omega$. For example, $\widehat{\mu}_{\underline{N}}(z)(\omega)$ is the realization of $\widehat{\mu}_{\underline{N}}(z)$ when $\omega \in \Omega$. Also, $\widehat{D}_{\underline{N}}(z)(\omega)$ denotes the conditional distribution of $\widehat{D}_{\underline{N}}(z)$ conditional on $\omega \in \Omega$.

Firstly, define $\widehat{T}_0(z) = \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M}+\underline{N}}} \left[\widehat{F}_Y(z) - \widehat{F}_X(z) - (F_Y(z) - F_X(z))\right]$. Lemma A.1 give the limit distribution of $\widehat{T}_0$.

Lemma A.1. *Under Assumptions 1-3, we have*

$$\widehat{T}_0 \Rightarrow \mathbb{G}_K,$$

where, $\mathbb{G}_K$ *is a mean zero Gaussian process with covariance kernel* K *satisfying*

$$K(z_1, z_2) = \lambda K_x(z_1, z_2) + (1-\lambda) K_y(z_1, z_2),$$

with

$$\begin{aligned}
K_x(z_1, z_2) =& \beta_{x,1} Cov(\mathbf{1}\{X_{11} \leq z_1\}, \mathbf{1}\{X_{12} \leq z_2\}) + \beta_{x,2} Cov(\mathbf{1}\{X_{11} \leq z_1\}, \mathbf{1}\{X_{21} \leq z_2\}), \\
K_y(z_1, z_2) =& \beta_{y,1} Cov(\mathbf{1}\{Y_{11} \leq z_1\}, \mathbf{1}\{Y_{12} \leq z_2\}) + \beta_{y,2} Cov(\mathbf{1}\{Y_{11} \leq z_1\}, \mathbf{1}\{Y_{21} \leq z_2\}).
\end{aligned}$$

Proof. Davezies et al. (2021) have shown the uniform laws of large numbers and the central limit theorems for separately exchangeable arrays under some mild conditions in their Theorem 3.4. Here, we only need to verify whether their conditions are satisfied in our setting. First, when we consider two-way clustering, our Assumption 1 coincides with their Assumption 6. Second, when we estimate empirical CDFs, we consider a function class of indicator functions, $\mathcal{F} = \left\{f_t = 1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\right\}$, we know that the envelop function of $\mathcal{F}$

is $F(x) = 1$ and the bracketing number of $\mathcal{F}$ is $N_{[]}(\epsilon, \mathcal{F}, L_2(Q)) \leq \frac{2}{\epsilon^2}$ for any probability measure Q, which implies the Assumptions 2-4 in Davezies et al. (2021). Thus, we have

$$\begin{gathered} \sup_{z \in \mathcal{Z}} \mid \widehat{F}_Y(z) - F_Y(z) \mid \xrightarrow{p} 0, \quad \sup_{z \in \mathcal{Z}} \mid \widehat{F}_X(z) - F_X(z) \mid \xrightarrow{p} 0, \\ \sqrt{\underline{N}} \left(\widehat{F}_Y - F_Y \right) \Rightarrow \mathbb{G}_Y, \quad \sqrt{\underline{M}} \left(\widehat{F}_X - F_X \right) \Rightarrow \mathbb{G}_X, \end{gathered} \tag{3}$$

where, $\mathbb{G}_Y$ is a mean zero Gaussian process with kernel K_y satisfying

$$K_y(z_1, z_2) = \beta_{y,1} \mathrm{Cov}\left(\mathbf{1}\{Y_{11} \leq z_1\}, \mathbf{1}\{Y_{12} \leq z_2\}\right) + \beta_{y,2} \mathrm{Cov}\left(\mathbf{1}\{Y_{11} \leq z_1\}, \mathbf{1}\{Y_{21} \leq z_2\}\right),$$

and $\mathbb{G}_X$ is a mean zero Gaussian process with kernel K_x satisfying

$$K_x(z_1, z_2) = \beta_{x,1} \mathrm{Cov}\left(\mathbf{1}\{X_{11} \leq z_1\}, \mathbf{1}\{X_{12} \leq z_2\}\right) + \beta_{x,2} \mathrm{Cov}\left(\mathbf{1}\{X_{11} \leq z_1\}, \mathbf{1}\{X_{21} \leq z_2\}\right).$$

By the independence between samples $\{Y_{ij} : i = 1, \ldots, N_1$ and $j = 1, \ldots, N_2\}$ and $\{X_{ij} : i = 1, \ldots, M_1$ and $j = 1, \ldots, M_2\}$, we have

$$\widehat{T}_0 = \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \left[\widehat{F}_Y - \widehat{F}_X - (F_Y - F_X) \right] \Rightarrow \mathbb{G}_K = \sqrt{1 - \lambda}\, \mathbb{G}_Y - \sqrt{\lambda}\, \mathbb{G}_X,$$

where, $\mathbb{G}_K$ is a mean zero Gaussian process with covariance kernel K satisfying

$$K(z_1, z_2) = \lambda K_x(z_1, z_2) + (1 - \lambda) K_y(z_1, z_2).$$

$\square$

Lemma A.2 shows that the simulated processes $\widehat{\mathbb{G}}_K$ converges weakly in probability to the process $\mathbb{G}_K$.

Lemma A.2. *Suppose that Assumptions 1-3 hold. Then*

$$\widehat{\mathbb{G}}_K \overset{p}{\Rightarrow} \mathbb{G}_K.$$

Proof. As discussed in the proof of Lemma A.1, our Assumptions 1-3 imply the Assumptions 2-4 and 6 in Davezies et al. (2021). So, based on their Theorem 3.4.3, we have,

$$\widehat{\mathbb{G}}_X \overset{p}{\Rightarrow} \mathbb{G}_X, \quad \widehat{\mathbb{G}}_Y \overset{p}{\Rightarrow} \mathbb{G}_Y.$$

Then, by the independence between samples $\{Y_{ij} : i = 1, \ldots, N_1 \text{ and } j = 1, \ldots, N_2\}$ and $\{X_{ij} : i = 1, \ldots, M_1 \text{ and } j = 1, \ldots, M_2\}$, we have

$$\begin{aligned}\widehat{\mathbb{G}}_K(z) =& \sqrt{\frac{\underline{M} \cdot \underline{N}}{\underline{M} + \underline{N}}} \left[\left(\frac{1}{N_1 N_2} \sum_{ij} W_{ij}^Y \mathbf{1}\{Y_{ij} \le z\} - \widehat{F}_Y(z) \right) - \left(\frac{1}{M_1 M_2} \sum_{ij} W_{ij}^X \mathbf{1}\{X_{ij} \le z\} - \widehat{F}_X(z) \right) \right] \\ =& \sqrt{1-\lambda}\widehat{\mathbb{G}}_Y - \sqrt{\lambda}\widehat{\mathbb{G}}_X \overset{p}{\Rightarrow} \sqrt{1-\lambda}\mathbb{G}_Y - \sqrt{\lambda}\mathbb{G}_X = \mathbb{G}_K.\end{aligned}$$

□

Lemma A.3 shows that the uniform convergence of $\widehat{\mu}_{\underline{N}}$. Lemmas A.2 - A.5 are used to the proof of Theorem 2.2.

Lemma A.3. *Suppose that Assumptions 1-5 hold. Then*

$$\sup_{z \in \mathcal{Z}} |\widehat{\mu}_{\underline{N}}(z) - \mu(z)| \overset{p}{\to} 0.$$

Lemma A.4. *Suppose that Assumptions 1-5 hold. Then*

$$\sup_{z \in \mathcal{Z}^+} \mathbf{1}\left\{ \underline{N}\widehat{A}_{\underline{N}}(z) < \delta_{\underline{N}} \right\} \overset{p}{\to} 0.$$

Lemma A.5. *Let ℓ_{N_1}, ℓ_{N_2}, ℓ_{M_1}, and ℓ_{M_2} be any subsequences of N_1, N_2, M_1 and M_2, respectively. Denote $\underline{\ell_N} = \min\{\ell_{N_1}, \ell_{N_2}\}$ and $\underline{\ell_M} = \min\{\ell_{M_1}, \ell_{M_2}\}$. If the following statements hold:*

1. $\sqrt{\underline{\ell_M} \cdot \underline{\ell_N} / \left(\underline{\ell_M} + \underline{\ell_N}\right)}\, \widehat{D}_{\underline{\ell_N}}(z) \overset{a.s.}{\Rightarrow} \mathbb{G}_K(z)$;
2. $\sup_{z \in \mathcal{Z}} \left| \widehat{\mu}_{\underline{\ell_N}}(z) - \mu(z) \right| \overset{a.s.}{\to} 0$;
3. $P\left(\omega \,|\sup_{z \in \mathcal{Z}^+} |\, \widehat{\mu}_{\underline{\ell_N}}(z)(\omega) \,| = 0 \text{ eventually } \right) = 1$;

then

$$P\left(\omega \mid \sup_{z\in\mathcal{Z}} \sqrt{\frac{\underline{\ell_M}\cdot\underline{\ell_N}}{\underline{\ell_M}+\underline{\ell_N}}}\left(\widehat{D}_{\underline{\ell_N}}(z)(\omega)+\widehat{\mu}_{\underline{\ell_N}}(z)(\omega)\right) \xrightarrow{d} \sup_{z\in\mathcal{Z}^+} \mathbb{G}_K(z)\right)=1.$$

Define $\widehat{c}_{1-\alpha,\underline{\ell_N}}$ *as*

$$\widehat{c}_{1-\alpha,\underline{\ell_N}}=\sup\left\{c \mid P^u\left(\sqrt{\frac{\underline{\ell_M}\cdot\underline{\ell_N}}{\underline{\ell_M}+\underline{\ell_N}}}\left(\widehat{D}_{\underline{\ell_N}}(z)+\widehat{\mu}_{\underline{\ell_N}}(z)\right)\leq c\right)\leq 1-\alpha\right\},$$

then $\widehat{c}_{1-\alpha,\underline{\ell_N}} \xrightarrow{a.s.} c^+_{1-\alpha}$*, where* $c^+_{1-\alpha}$ *is the* $(1-\alpha)$*th quantile of* $\sup_{z\in\mathcal{Z}^+}\mathbb{G}_K(z)$.

The proof of Lemmas A.3-A.5 are similar to DH.

B Proof of Lemma 2.1

For the first part, denote $H_{\widehat{T}}(a)$ as the CDF of $\widehat{T}$ and $H_{\mathbb{G}}(a)$ as that of $\sup_{z\in\mathcal{Z}^*}\mathbb{G}_K(z)$. The proof of $\widehat{T}\xrightarrow{d}\sup_{z\in\mathcal{Z}^*}\mathbb{G}_K(z)$ is equivalent to show that

$$\lim_{\underline{N}\to\infty} H_{\widehat{T}}(a)=H_{\mathbb{G}}(a), \tag{4}$$

for each continuous point a of $H_{\mathbb{G}}(a)$. Here, we focus on the proof of (4). Firstly, for all $a<0$, we have

$$H_{\widehat{T}}(a)=H_{\mathbb{G}}(a)=0, \tag{5}$$

since the fact that $\widehat{T}\geq 0$ and $\sup_{z\in\mathcal{Z}^*}\mathbb{G}_K(z)\geq 0$. Then, for all $a>0$, following the proof of Proposition 1 (A) in Barrett and Donald (2003), we have

$$\lim_{\underline{N}\to\infty} P(\widehat{T}\leq a)=P(\sup_{z\in\mathcal{Z}^*}\mathbb{G}_K(z)\leq a), \tag{6}$$

that is, $\lim_{\underline{N}\to\infty} H_{\widehat{T}}(a)=H_{\mathbb{G}}(a)$ for all $a>0$. Next, we will talk about the continuity of $H_{\mathbb{G}}(a)$ at $a=0$. If $H_{\mathbb{G}}(a)$ is not continuous at $a=0$, then the result holds directly. When

$H_{\mathbb{G}}(a)$ is continuous at $a = 0$, $H_{\mathbb{G}}(0) = 0$ since $H_{\mathbb{G}}(a) = 0$ for all $a < 0$. Therefore, the only task left is to show that $\lim_{\underline{N}\to\infty} H_{\widehat{T}}(0) = 0$. By the nondecreasing property of CDF, we have $0 \le H_{\widehat{T}}(0) \le H_{\widehat{T}}(\epsilon)$ for all $\epsilon > 0$. Consequently,

$$0 \le \limsup_{\underline{N}\to\infty} H_{\widehat{T}}(0) \le \limsup_{\underline{N}\to\infty} H_{\widehat{T}}(\epsilon) = H_{\mathbb{G}}(\epsilon),$$

where the last equality follows from the fact that $\limsup_{\underline{N}\to\infty} H_{\widehat{T}}(\epsilon) = H_{\mathbb{G}}(\epsilon)$ for all $\epsilon > 0$. Since ϵ can be arbitrarily small, we have $\lim_{\epsilon\to 0} H_{\mathbb{G}}(\epsilon) = H_{\mathbb{G}}(0) = 0$. Also, $\liminf_{\underline{N}\to\infty} H_{\widehat{T}}(0) \ge 0$ because $H_{\widehat{T}}(0) \ge 0$ for all $\underline{N}$. These imply that

$$\lim_{\underline{N}\to\infty} H_{\widehat{T}}(0) = 0. \tag{7}$$

Combine (5)-(7), $\widehat{T} \xrightarrow{d} \sup_{z\in\mathcal{Z}^*} \mathbb{G}_K(z)$ is thus proved.

For the second part, we note that if the alternative H_1 is true, then there is some z', say $z' \in \mathcal{Z}$ such that $F_Y(z') - F_X(z') = \rho > 0$ for some ρ. The results in (3) imply that $\widehat{F}_Y(z') - \widehat{F}_X(z') \xrightarrow{p} \rho$. Then the result in second part follows by

$$\widehat{T} \ge \sqrt{\frac{\underline{N}\cdot\underline{M}}{\underline{N}+\underline{M}}}\left(\widehat{F}_Y(z') - \widehat{F}_X(z')\right) \xrightarrow{p} \infty.$$

$\square$

C Proof of Theorems

Proof of Theorem 2.2. Firstly, for any sequence a_n, $a_n \xrightarrow{p} a$ is equivalent to show that, for any subsequence k_n of n, there is a further subsequence ℓ_n of k_n such that $a_{\ell_n} \xrightarrow{a.s.} a$.

Define $c^+_{1-\alpha}$ as the $(1-\alpha)$th quantile of $\sup_{z\in\mathcal{Z}^+} \mathbb{G}_K(z)$. We want to show that $\widehat{c}_{1-\alpha,\underline{N}} \xrightarrow{p} c^+_{1-\alpha}$, which is equivalent to showing that, for any subsequences $k_N = \{k_{N_1}, k_{N_2}\}$ of $N = \{N_1, N_2\}$ and $k_M = \{k_{M_1}, k_{M_2}\}$ of $M = \{M_1, M_2\}$, there are further subsequences $\ell_N =$

$\{\ell_{N_1}, \ell_{N_2}\}$ of k_N and $\ell_M = \{\ell_{M_1}, \ell_{M_2}\}$ of k_M such that $\widehat{c}_{1-\alpha,\underline{\ell_N}} \overset{a.s.}{\to} c^+_{1-\alpha}$, where, $\underline{\ell_N} = \min\{\ell_{N_1}, \ell_{N_2}\}$, $\underline{\ell_M} = \min\{\ell_{M_1}, \ell_{M_2}\}$.

Now we focus on the proof of $\widehat{c}_{1-\alpha,\underline{\ell_N}} \overset{a.s.}{\to} c^+_{1-\alpha}$. By Lemma A.2, Lemma A.3 and Lemma A.4, we have

$$\sqrt{\underline{N} \cdot \underline{M}/(\underline{N} + \underline{M})}\widehat{D}_{\underline{N}}(z) \overset{p}{\Rightarrow} \mathbb{G}_K(z),$$

$$\sup_{z \in \mathcal{Z}} \mid \widehat{\mu}_{\underline{N}}(z) - \mu(z) \mid \overset{p}{\to} 0,$$

$$\sup_{z \in \mathcal{Z}^+} \mathbf{1}\left\{\sqrt{\underline{N}}\widehat{A}_{\underline{N}}(z) < \delta_{\underline{N}}\right\} \overset{p}{\to} 0.$$

Then, the equivalent statement of convergence in probability implies that,

$$\sqrt{\underline{\ell_M} \cdot \underline{\ell_N}/\left(\underline{\ell_M} + \underline{\ell_N}\right)}\widehat{D}_{\underline{\ell_N}} \overset{a.s.}{\Rightarrow} \mathbb{G}_K(z), \tag{8}$$

$$\sup_{z \in \mathcal{Z}} \left|\widehat{\mu}_{\underline{\ell_N}}(z) - \mu(z)\right| \overset{a.s.}{\to} 0, \tag{9}$$

$$\sup_{z \in \mathcal{Z}^+} \mathbf{1}\left\{\sqrt{\underline{N}}\widehat{A}_{\underline{\ell_N}}(z) < \delta_{\underline{\ell_N}}\right\} \overset{a.s.}{\to} 0. \tag{10}$$

Note that the indicator function takes a value of 0 and 1 only, so (10) implies that

$$P\left(\omega \left| \sup_{z \in \mathcal{Z}^+} \mathbf{1}\left\{\sqrt{\underline{N}}\widehat{A}_{\underline{\ell_N}}(z)(\omega) < \delta_{\underline{\ell_N}}\right\}\right| = 0 \text{ eventually}\right) = 1,$$

which further implies that

$$P\left(\omega \left| \sup_{z \in \mathcal{Z}^+}\right| \widehat{\mu}_{\underline{\ell_N}}(z)(\omega) \mid= 0 \text{ eventually}\right) = 1, \tag{11}$$

because the definition of $\widehat{\mu}_{\underline{\ell_N}}(z)$, $\widehat{\mu}_{\underline{\ell_N}}(z) = \widehat{A}_{\underline{\ell_N}}(z) \cdot \mathbf{1}\left\{\sqrt{\underline{N}}\widehat{A}_{\underline{\ell_N}}(z) < \delta_{\underline{\ell_N}}\right\}$. Combine (8), (9) and (11), Lemma A.5 implies $\widehat{c}_{1-\alpha,\underline{\ell_N}} \overset{a.s.}{\to} c^+_{1-\alpha}$, and it follows that $\widehat{c}_{1-\alpha,\underline{N}} \overset{p}{\to} c^+_{1-\alpha}$.

For the first part, we have $c^+_{1-\alpha} > 0$ because the limiting null distribution is nondegenerate. Thus, we have $\widehat{c}_{1-\alpha,\eta} \xrightarrow{p} c^+_{1-\alpha} > 0$ when η is small enough. Hence,

$$\begin{aligned}
\lim_{\underline{N}\to\infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) &= \lim_{\underline{N}\to\infty} P\left(\sup_{z\in\mathcal{Z}^*} \mathbb{G}_K(z) > \widehat{c}_{1-\alpha,\eta}\right) \\
&= \lim_{\underline{N}\to\infty} P\left(\sup_{z\in\mathcal{Z}^+} \mathbb{G}_K(z) > \widehat{c}_{1-\alpha,\eta}\right) \\
&= \lim_{\underline{N}\to\infty} P\left(\sup_{z\in\mathcal{Z}^+} \mathbb{G}_K(z) > c^+_{1-\alpha}\right) = \alpha,
\end{aligned}$$

where, the second equality holds because $\mathcal{Z}^* = \mathcal{Z}^+$ under H_0, which gives the result when η is small enough.

For the second part, we need to consider two cases: (a) $F_Y(z) = F_X(z)$ for some $z \in (0, \bar{z})$, but for these z's, $K(z, z) = 0$ as well, and (b) $F_Y(z) < F_X(z)$ for all $z \in (0, \bar{z})$. In both two cases, the limiting distributions of $\sup_{z\in\mathcal{Z}^+} \mathbb{G}_K(z)$ and $\sup_{z\in\mathcal{Z}^*} \mathbb{G}_K(z)$ are degenerate at 0. This degeneracy in the first case arises because $K(z, z) = 0$ for $z \in \mathcal{Z}^*$ and is attributed to the fact that $\mathcal{Z}^* = \mathcal{Z}^+ = \emptyset$ in the second case. Then, $c^+_{1-\alpha} = 0$, $\widehat{c}_{1-\alpha,\eta} = \max\left\{\widehat{c}_{1-\alpha,\underline{N}}, \eta\right\} \xrightarrow{p} \eta > 0$ and $\widehat{T} \xrightarrow{p} 0$. These imply,

$$\lim_{\underline{N}\to\infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) = 0.$$

□

Proof of Theorem 2.3. Under a fixed alternative hypothesis, we have $\widehat{T} \xrightarrow{p} \infty$ and $\widehat{c}_{1-\alpha,\eta}$ is bounded in probability. Hence, $\lim_{n\to\infty} P\left(\widehat{T} > \widehat{c}_{1-\alpha,\eta}\right) = 1$. □

D A Brief Introduction to DH Test

In this section, we give a brief description for the algorithm of DH test. Denote N and M are the total sample sizes for Y and X, respectively. When testing the first order stochastic dominance, the KS test statistic used in DH is

$$\widehat{S} = \sqrt{\frac{MN}{M+N}} \sup_{z\in\mathcal{Z}} \left(\widehat{F}_Y(z) - \widehat{F}_X(z)\right),$$

where, $\widehat{F}_Y(z)$ and $\widehat{F}_Y(z)$ are the same with our paper. DH introduce three methods to simulate the limit process, here, we only use the bootstrap method with separate samples. Draw a random sample of size N from $\{Y_{ij}\}$ to form $\widehat{F}_{Y,b}$ and a random sample of size M from $\{X_{ij}\}$ to form $\widehat{F}_{X,b}$. Define

$$\widehat{D}_N(z) = \left(\widehat{F}_{Y,b}(z) - \widehat{F}_Y(z)\right) - \left(\widehat{F}_{X,b}(z) - \widehat{F}_X(z)\right).$$

Let $\delta_N = -0.1\sqrt{\log\log(N+M)}$, define the recentering function as

$$\widehat{\mu}_N(z) = (F_Y(z) - F_X(z)) \cdot 1\left(\sqrt{N}\,(F_Y(z) - F_X(z)) < \delta_N\right).$$

Let α be the nominal level, calculate the critical values as

$$\begin{aligned}
\widetilde{c}_{1-\alpha,\eta} &= \max\left\{\widetilde{c}_{1-\alpha,N}, \eta\right\}, \\
\widetilde{c}_{1-\alpha,N} &= \sup_{c}\left\{P_b\left(\sqrt{\frac{MN}{M+N}}\sup_{z\in\mathcal{Z}}\left(\widehat{D}_N(z) + \widehat{\mu}_N(z)\right) \le c\right) \le 1-\alpha\right\}.
\end{aligned}$$

We reject H_0 when $\widehat{S} > \widetilde{c}_{1-\alpha,\eta}$. The extension to higher order stochastic dominance is straightforward as delineated in Section 4.1.

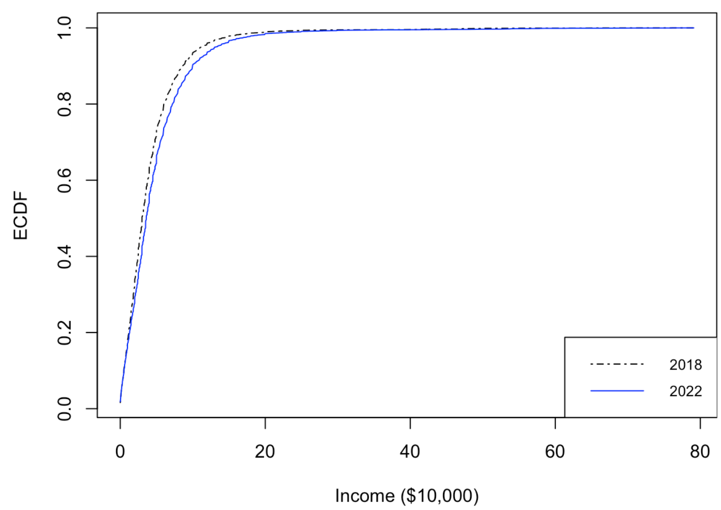

Figure 1: The empirical CDFs of income in 2018 and 2022