Economics Working Paper Series

2020 – 05

A stochastic dominance test under survey nonresponse with an application to comparing trust levels in Lebanese public institutions

Ali Fakih, Paul Makdissi, Walid Marrouch, Rami V. Tabri, and Myra Yazbeck

May 2020

A stochastic dominance test under survey nonresponse with an application to comparing trust levels in Lebanese public institutions*

Ali Fakih† Lebanese American University

Paul Makdissi‡ University of Ottawa

Walid Marrouch§ Lebanese American University

Rami V. Tabri¶ University of Sydney

Myra Yazbeck‖ University of Ottawa

May 23, 2020

Abstract

Stochastic dominance comparisons of distributions based on ordinal data arise in many areas of economics. This paper develops a testing procedure for such comparisons under survey sampling from large finite populations with nonresponse using the worst-case bounds of the distributions. The advantage of using these bounds in distributional comparisons is that conclusions are robust to the nature of the nonresponse-generating mechanism. While these bounds on the distributions are often too wide in practice, we show that they can be informative for distributional comparisons in an empirical analysis. This paper examines the dynamics of trust in Lebanese public institutions using the 2013 *World Values Survey* as well as the 2016 and 2018 waves of the *Arab Barometer*, and finds convincing evidence of a decrease in confidence in most public institutions between 2013 and 2016.

JEL Classification: C12;C14

Keywords: Empirical Likelihood; Stochastic Dominance Test; Ordinal Variables; Survey Nonresponse

1 Introduction

Distributional comparisons of ordinal variables is a frequently encountered topic in many areas of economics. Examples of ordinal data are self-reported Likert-type scale responses like health status in the *National Health Interview Survey*, happiness levels in the *General Social Survey*, and trust levels in public institutions in the *World Values Survey*. The difficulty arising with such comparisons is that the absence of a

*This paper was previously circulated under the title "Confidence in public institutions and the run up to the October 2019 uprising in Lebanon" as IZA Discussion Paper No. 13104.

† Department of Economics, Lebanese American University, Lebanon.

‡ Department of Economics, University of Ottawa, Ottawa, Canada

§ Department of Economics, Lebanese American University, Lebanon.

¶ Corresponding author. School of Economics, University of Sydney, Sydney, Australia; Email: rami.tabri@sydney.edu.au

‖ Department of Economics, University of Ottawa, Ottawa, Canada.

numerical scale of such variables prevents the use of traditional tools for distributional analysis. First-order stochastic dominance has been proposed to circumvent this difficulty since the comparisons based upon it are scale-independent (e.g., Allison and Foster, 2004). Additionally, second-order stochastic dominance has been proposed to compare such distributions in terms of inequality. As a consequence, stochastic dominance orderings have been, and continue to be, one of the most widely used analytical tools for comparing such distributions. For example, Madden (2009) compares mental stress levels across populations, Allison and Foster (2004) propose a method of health status comparisons, Dutta and Foster (2013) extend Allison and Foster's approach to happiness status comparisons, and Jenkins (2019) establishes that a version of generalized Lorenz dominance comparisons of such distributions is equivalent to comparisons based on classes of ordinal data inequality indices put forward by Cowell and Flachaire (2017).

In practice, empirical researchers perform 'dominance checks' by visually inspecting graphs of stochastic dominance curves (e.g., Jenkins, 2020). While it is one thing to observe that the ordinates for one distribution lie above that for another, it still has to be established that such dominance is statistically significant because of error due to sampling. There is a substantial body of literature on tests for stochastic dominance – see, for example, the recent book by Whang (2019) and the references therein. To the best of our knowledge, all of these tests, however, are not applicable to data generated from surveys with complex designs (e.g., involving stratification and clustering). The reason is that these tests apply only to data generated from a simple random sample or time series, which is rarely the case for data from surveys like the aforementioned examples. Furthermore, these tests are only valid when the data do not contain missing values, which is generally unachievable in practice, as nonresponse is a universal problem in all self-reported surveys.

This paper has two objectives. The first is methodological: to develop a statistical test for first-order stochastic dominance of ordinal distributions that applies under survey nonresponse, using distributional bounds. The standard practice with nonresponse has been to use weights and imputations to implement assumptions that nonresponse is conditionally random; however, these assumptions are unverifiable in practice and may yield biased inferences due to misspecification (see, for example, Manski, 2016). Our approach circumvents this difficulty by comparing the worst-case bounds of distributions, because it achieves robustness to the nature of the nonresponse-generating process. The testing procedure is based on the test of Davidson and Duclos (2013) for first-order stochastic dominance. Their test uses the method of empirical likelihood (Owen, 2001) and is based on the random sampling assumption without missing data problems. We adapt their procedure to the context of survey sampling from large finite populations using the method of pseudo-empirical likelihood proposed by Chen and Sitter (1999), and use the worst-case bounds to account for survey nonresponse.

The test posits a null hypothesis of nondominance and its negation as the alternative hypothesis, which is dominance. This formulation of the testing problem is advantageous because in practice one wishes to find convincing evidence for dominance, and rejecting the null in this formulation with a small significance level provides this evidence. The pseudo-empirical likelihood-ratio statistic we propose is similar to the test statistic in Davidson and Duclos (2013), and we establish that the procedure which compares it to a critical value drawn from the chi-square distribution with one degree of freedom is asymptotically valid with uniformity. An appealing feature of this testing procedure is its computational simplicity; however, it is conservative. Thus, a rejection of the null using such a critical value provides very convincing evidence for the alternative hypothesis of dominance.

The idea of using bounds to account for data problems started with the seminal paper of Horowitz and Manski (1995). Since then, there has been a growing influential literature on partial identification that has shaped empirical practice (e.g., Haile and Tamer, 2003, Canay, 2010, Andrews and Soares, 2010, Shi, 2015, and Andrews and Shi, 2017 among others). Our work is related to various papers in that literature, most closely to those involving distributional analyses using bounds when data contain missing values, such as Blundell et al. (2007), and Lee (2009). These papers focus on refining the worst-case bounds of continuous distributions using economic theory and the development of testable implications based on them, under random sampling. By contrast, our focus is on comparisons of ordinal distributions under survey sampling with nonresponse, using the worst-case bounds of the distributions.

Other testing procedures that posit a null of nondominance are Berger (1988) and Alvarez-Esteban et al. (2017). Similar to other tests in the literature, both focus on continuous distributions and the random sampling assumption without missing data problems. An important difference between those tests and the test of Davidson and Duclos (2013) is that the latter applies to both continuous and discrete random variables.

The second objective of this paper is an empirical analysis on the dynamics of trust in Lebanese public institutions leading to the October 2019 uprising. The uprising turned into multiple social protests demanding social justice, the replacement of the corrupted political elite, and the end to the sectarian political system that shaped the last three decades of the post 1975-1990 civil war era. Caught by surprise, many politicians have argued that this uprising does not reflect a real desire for change since, very recently (in May 2018), a general election was held and the Lebanese voted the same political class back into power. Considering that (i) the voter turnout was very low[1] and (ii) the lack of space to express distrust of political elites is significant, it is interesting to investigate whether the electoral outcome reflects the population's views.

The data we use are on trust (or confidence)[2] levels from the 2013 *World Values Survey* as well as the 2016 and 2018 waves of the *Arab Barometer*. We use these datasets to explore how trust in Lebanese public institutions has been changing over time.[3] Given that there is a high incidence of nonresponse for some of these variables in the datasets, it is essential to appropriately account for nonresponse in the comparisons. We show the worst-case bounds can be informative about changes in the distribution of trust levels over time and find evidence on the change in salience of trust in political institutions, using the proposed test. This finding is important as it points to the instrumental role of analyzing information on people's perceptions (e.g., trust, confidence and satisfaction) of public policy in developing countries where such information is discounted and, in some cases, fully rejected by political elites.

The rest of this paper is organized as follows. Section 2 presents the testing problem of interest. Section 3 introduces the testing procedure, and Section 4 presents the asymptotic framework and uniform validity result. Section 5 presents the empirical analysis on dynamics of trust in Lebanese public institutions, and Section 6 concludes. All proofs are relegated to the Appendix.

[1] Baalbaki (2018).
[2] We use the words "trust" and "confidence" interchangeably.
[3] Earlier versions of the Arab Barometer do not include questions on trust in public institutions.

2 Testing Problem and Worst-case Bounds

Consider an ordinal variable measured on a scale characterized by the numerical labels $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$, where $-\infty < s_1 < s_2 < \cdots < s_K < +\infty$, and let F_A and F_B be two cumulative distribution functions (CDFs) supported on $\mathcal{S}$. Additionally, let $\mathcal{S}^\circ = \mathcal{S} - \{s_K\}$. We say that F_B *strictly* dominates F_A, stochastically, at first-order, if $F_B(s) < F_A(s)$ for each $s \in \mathcal{S}^\circ$. Note that we exclude the support point $\{s_K\}$ as $F_B(s_K) = F_A(s_K) = 1$ because F_B and F_A are CDFs.

Following Davidson and Duclos (2013), the testing problem of interest takes the following form:

$$H_0 : \max_{s \in \mathcal{S}^\circ} (F_B(s) - F_A(s)) \geq 0 \quad \text{Vs.} \quad H_1 : F_B(s) < F_A(s)\ \forall s \in \mathcal{S}^\circ. \tag{2.1}$$

The null hypothesis states that F_B does not strictly dominate F_A, stochastically, at first-order. The alternative hypothesis is the negation of the null. We formulate the null and alternative hypotheses as in (2.1) since we would like to use the data to provide strong evidence for dominance: $F_B(s) < F_A(s)\ \forall s \in \mathcal{S}^\circ$.

Let X_A and X_B be random variables with distributions F_A and F_B, respectively. Survey nonresponse in the setup of this paper means that

$$\text{the practitioner observes } = \begin{cases} X_Y & \text{if } D_Y = 1, \\ \text{missing value code} & \text{if } D_Y = 0, \end{cases}$$

where D_Y is a 0/1 binary random variable indicating an individual's response to the survey in the reference population $Y \in \{A, B\}$. Consequently, F_A and F_B are not necessarily point-identified unless we are prepared to make strong unverifiable assumptions about the nonresponse-generating process. Our approach circumvents the imposition of such assumptions by using the worst-case bounds on these CDFs, as they are robust to the nature of the nonresponse-generating process. Using the Law of Total Probability, Manski (1994) showed that for each $s \in \mathcal{S}$

$$F_Y(s) = \text{Prob}\,[D_Y = 1]\ \text{Prob}\,[X_Y \leq x \mid D_Y = 1] + \text{Prob}\,[X_Y \leq s \mid D_Y = 0]\ \text{Prob}\,[D_Y = 0]\,, \tag{2.2}$$

holds, and noted that $0 \leq \text{Prob}\,[X_Y \leq s \mid D_Y = 0] \leq 1$ for each $s \in \mathcal{S}$ because it is a conditional CDF. Now substituting these inequalities into (2.2) yields the worst-case bounds: $\underline{F}_Y(s) \leq F_Y(s) \leq \overline{F}_Y(s)$ for each $s \in \mathcal{S}$, where

$$\overline{F}_Y(s) = \text{Prob}\,[D_Y = 1]\ \text{Prob}\,[X_Y \leq s \mid D_Y = 1] + \text{Prob}\,[D_Y = 0] \quad \text{and}$$
$$\underline{F}_Y(s) = \text{Prob}\,[D_Y = 1]\ \text{Prob}\,[X_Y \leq s \mid D_Y = 1] \quad \forall s \in \mathcal{S}.$$

As these bounds are CDFs and depend only on observed values, our approach to testing H_0 compares the worst-case lower bound of F_A with the worst-case upper bound of F_B. In particular, we consider the following testing problem:

$$H_0^1:\ \max_{s\in\mathcal{S}^\circ}\left(\overline{F}_B(s)-\underline{F}_A(s)\right)\geq 0 \quad \text{Vs.} \quad H_1^1:\overline{F}_B(s)<\underline{F}_A(s)\ \forall s\in\mathcal{S}^\circ. \tag{2.3}$$

Rejecting H_0^1 in favor of H_1^1 in (2.3) implies rejection of H_0 in favor of H_1 in (2.1), since $F_B(s)\leq\overline{F}_B(s)<\underline{F}_A(s)\leq F_A(s)\ \forall s\in\mathcal{S}$.

3 Testing Procedure

We adapt the testing procedure of Davidson and Duclos (2013) for the test problem (2.3). To that end, we use the method of pseudo-empirical likelihood put forward by Chen and Sitter (1999). We have access to two independent survey-samples from potentially different designs: $\{X_{Y,i},D_{Y,i},W_{Y,i}\}_{i=1}^{n_Y}$ for $Y\in\{A,B\}$, where the $W_{Y,i}$ are the survey design weights which have no missing values and n_Y is the sample size from population Y. The surveys weights have been scaled so that $\sum_i W_{Y,i}=n_Y$, for each $Y\in\{A,B\}$. To obtain a sample from $\overline{F}_B$ using the sample $\{X_{B,i},D_{B,i}\}_{i=1}^{n_B}$, replace each missing value in it with the smallest value in $\mathcal{S}$ (i.e., s_1). Similarly, modifying the sample $\{X_{A,i},D_{A,i}\}_{i=1}^{n_A}$ by replacing each missing value in it with the largest value in $\mathcal{S}$ (i.e., s_K) yields a sample from $\underline{F}_A$. Specifically, the samples are

$$\overline{X}_{B,i}=\begin{cases}X_{B,i} & \text{if } D_{B,i}=1,\\ s_1 & \text{if } D_{B,i}=0,\end{cases}\ \forall i=1,\dots,n_B \quad \text{and} \quad \underline{X}_{A,j}=\begin{cases}X_{A,j} & \text{if } D_{A,j}=1,\\ s_K & \text{if } D_{A,j}=0\end{cases}\ \forall j=1,\dots,n_A.$$

Following Davidson and Duclos (2013), the test focuses on the boundary of the null hypothesis H_0^1 in (2.3). A pair of populations in the boundary of H_0^1 has at least one $s\in\mathcal{S}^\circ$ such that $\overline{F}_B(s)=\underline{F}_A(s)$. Thus, only one such $s\in\mathcal{S}^\circ$ is required. To maximize the pseudo-empirical likelihood function (PELF) under the constraint of the boundary of the null, we begin by computing for each $s\in\mathcal{S}^\circ$ the maximum PELF when imposing $\overline{F}_B(s)=\underline{F}_A(s)$. We then choose the value $s\in\mathcal{S}^\circ$ which gives the greatest value of the constrained PELF. For a given $s\in\mathcal{S}^\circ$, the constraint we impose is

$$0=\sum_i\sum_j p_j^A p_i^B H_{i,j}(s), \tag{3.1}$$

where $H_{ij}(s)=W_{B,i}W_{A,j}h(\overline{X}_{B,i},\underline{X}_{A,j},s)\ \forall i,j$, with $h(Z_1,Z_2,z)=1\left[Z_1\leq z\right]-1\left[Z_2\leq z\right]$, and $\{p_1^Y,\dots,p_{n_Y}^Y\}$ for $Y\in\{A,B\}$ are probability masses on the samples. The maximisation of the PELF corresponding to the restriction (3.1) is

$$\max_{p_1^Y,\dots,p_{n_Y}^Y:Y\in\{A,B\}}\sum_{Y\in\{A,B\}}\sum_i W_{Y,i}\log p_i^Y \text{ subject to } p_i^Y>0\ \forall(i,Y),\ \sum_i p_i^Y W_{Y,i}=1\ \forall Y, \tag{3.2}$$

and the equality constraint (3.1). We denote by $L_R(s)$ the maximal value of this optimization problem. The unconstrained estimators of $\overline{F}_B$ and $\underline{F}_A$ are the Hájek CDF estimators

$$\hat{\overline{F}}_B(s) = n_B^{-1}\sum_i W_{B,i}1\left[\overline{X}_{B,i} \leq s\right] \quad \text{and} \quad \hat{\underline{F}}_A(s) = n_A^{-1}\sum_i W_{A,i}1\left[\underline{X}_{A,i} \leq s\right]\ \forall s \in \mathcal{S}, \tag{3.3}$$

respectively. The pseudo-empirical likelihood-ratio test statistic for testing H_0^1 is thus defined as

$$LR^{(A,B)} = \begin{cases} \min_{s\in\mathcal{S}^\circ} 2(L_{UR} - L_R(s)) & \text{if } \hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ, \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

where $L_{UR} = \sum_{Y\in\{A,B\}}\sum_i W_{Y,i}\log(1/n_Y)$ is the unconstrained maximum values of the PELF. The formulation of $LR^{(A,B)}$ as in (3.4) implements the procedure if we observe the sample satisfies $\hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ$; that is, dominance in the sample, holds. Otherwise, $LR^{(A,B)} = 0$ and we do not reject H_0^1. This formulation of the test statistic follows the prescription described in Section 6 of Davidson and Duclos (2013).

The decision rule is to reject H_0^1 if and only if $LR^{(A,B)} > c(\alpha)$, where $c(\alpha)$ is the $1-\alpha$ quantile from the χ_1^2 distribution. The next section develops the asymptotic framework and uniform validity theory for inference using this decision rule.

4 Uniform Asymptotic Framework and Validity

We start with a brief description of the probability space of the survey designs with finite populations. Denote by $\mathcal{P}_{Y,N_Y} = \{\{X_{Y,i}, D_{Y_i}\}, i = 1,2,3,\ldots,N_Y\}$ the set of vectors of population Y with total N_Y, for $Y = A, B$. Given a sampling scheme on population $\mathcal{P}_{Y,N_Y}$, let $\mathcal{U}_{N_Y} = \{U_Y : U_Y \subset \{1,2,\ldots,N_Y\}\}$ denote the set of all possible samples under the scheme, for $Y = A, B$. The sigma-algebra generated by $\mathcal{U}_{N_Y}$ is denoted as σ_{N_Y} for $Y = A, B$. Additionally, let $\mathcal{M}_{Y,N_Y} = \{\mathcal{P}_{Y,N_Y} : X_{Y,i} \in \mathcal{S}$ and $D_{Y_i} \in \{0,1\}\ \forall i = 1,\ldots,N_Y\}$, for $Y = A, B$. The set $\mathcal{M}_{Y,N_Y}$ consists of all finite populations Y with total N_Y. Following Rubin-Bleuer and Kratina (2005), the probability sampling design associated with a given sampling scheme on $\mathcal{P}_{Y,N_Y}$ is the function $P_Y : \sigma_{N_Y} \times \mathcal{M}_{Y,N_Y} \to [0,1]$ such that (i) for all $U_Y \in \mathcal{U}_{N_Y}$, $P_Y(U_Y,\cdot)$ is Borel measurable in $\mathcal{M}_{Y,N_Y}$, and (ii) for all $\mathcal{P}_{Y,N_Y} \in \mathcal{M}_{Y,N_Y}$, $P_Y(\cdot,\mathcal{P}_{Y,N_Y})$ is a probability measure on $\mathcal{U}_{N_Y}$, for $Y = A, B$. We say $(\mathcal{U}_{N_Y}, \sigma_{N_Y}, P_Y)$ is a design probability space, where $P_Y(U_Y,\cdot) > 0$ for any $U_Y \in \mathcal{U}_{N_Y}$, $\sum_{U_Y\in\mathcal{U}_{N_Y}} P_Y(U_Y,\cdot) = 1$, for $Y = A, B$. The design probability space for sampling with two independent survey samples on the two populations is $(\mathcal{U}_{N_A}\times\mathcal{U}_{N_B}, \sigma_{N_A}\times\sigma_{N_B}, P)$, with the joint probability set function P being the product of P_A and P_B. In this framework, the sample size from the given scheme on population $\mathcal{P}_{Y,N_Y}$ is thus $n_Y = \sum_{i\in U_Y}$, and is a random variable. In this formulation, the scaled design weights, $W_{Y,i}$, are proportional to the first-order inclusion probabilities $\pi_{Y,i} = \sum_{U_Y\in\mathcal{U}_{N_Y}:i\in U_Y} P_Y(U_Y,\mathcal{P}_{Y,N_Y})$ and satisfy $\frac{W_{Y,i}}{n_Y} = \frac{\pi_{Y,i}^{-1}}{\sum_{i\in U_Y}\pi_{Y,i}^{-1}}$, for $Y = A, B$.

Now let $\Pi = \{\mathcal{P}_{A,N_A}, \mathcal{P}_{B,N_B}\}$, and denote by Π_0 the true population. Then the set of all finite populations of totals N_A and N_B is thus $\mathcal{M}_{N_A,N_B} = \{\Pi = \{\mathcal{P}_{A,N_A}, \mathcal{P}_{B,N_B}\} : \mathcal{P}_{Y,N_Y} \in \mathcal{M}_{Y,N_Y}, Y \in \{A,B\}\}$.

The model of the null hypothesis is given by

$$\mathcal{M}^0_{N_A,N_B} = \left\{ \Pi \in \mathcal{M}_{N_A,N_B} : \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_B(s) - \underline{F}_A(s) \right) \geq 0 \right\},$$

so that the assertion Π_0 satisfies H_0^1 in (2.3) is equivalent to $\Pi_0 \in \mathcal{M}^0_{N_A,N_B}$. The size of the test for H_0^1 in (2.3) is defined as $\sup_{\Pi\in\mathcal{M}^0_{N_A,N_B}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right]$, where the notation $E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right]$ denotes the expected value of the statistic $1[LR^{(A,B)} > c(\alpha)]$, taken over all samples possible under the designs for the finite population Π. In our framework the only source of randomness is from sampling according to the surveys' probability designs on the different populations (i.e., the probability set functions P_A and P_B).

We shall approximate the size of the test using its asymptotic size when the populations' totals tend to infinity. In that direction, we embed $\mathcal{M}^0_{N_A,N_B}$ in to a hypothetical sequence $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ that satisfies enough restrictions so that

$$\limsup_{N_A,N_B\to+\infty} \sup_{\Pi\in\mathcal{M}^0_{N_A,N_B}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right] \leq \alpha, \tag{4.1}$$

holds. The approach to proving (4.1) uses a characterization of it in terms of sequences of finite populations, $\{\Pi_{N_A,N_B} = \{\mathcal{P}_{A,N_A}, \mathcal{P}_{B,N_B}\}, N_A, N_B = 1, 2, \ldots\}$, where $\Pi_{N_A,N_B} \in \mathcal{M}^0_{N_A,N_B}$ for all N_A and N_B, and the asymptotic distribution of any sample quantity is its limiting distribution along this hypothetical infinite sequence. As the only source of randomness is from sampling according to the probability designs of the surveys, the asymptotic distribution of $\{LR^{(A,B)} \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ is calculated from the sequence of survey-samples selected from the sequence of finite populations $\{\Pi_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$. Note that $LR^{(A,B)} \mid \Pi_{N_A,N_B}$ means the statistic, $LR^{(A,B)}$, is a function of the survey samples selected from population Π_{N_A,N_B}.

Next, we describe the conditions on the surveys' designs for obtaining (4.1). For each N_A and N_B, denote by F_{A,N_A} and F_{B,N_B} the population CDFs of the ordinal variable under $\mathcal{P}_{A,N_A}$ and $\mathcal{P}_{B,N_B}$, respectively. Additionally, let $\overline{F}_{B,N_B}$ and $\underline{F}_{A,N_A}$ denote the worst-case upper and lower bounds on F_{B,N_B} and F_{A,N_A}, respectively. For a given sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, the conditions we impose on the designs of the surveys are given by the following assumption.

Assumption 1. (i) $E[n_Y \mid \mathcal{P}_{Y,N_Y}] \to +\infty$ as $N_Y \to +\infty$, for $Y = A, B$.

(ii) $\max_{s\in\mathcal{S}^\circ} \left|\hat{\overline{F}}_B(s) - \overline{F}_{B,N_B}(s)\right| \mid \mathcal{P}_{B,N_B} \xrightarrow{P} 0$ and $\max_{s\in\mathcal{S}^\circ} \left|\hat{\underline{F}}_A(s) - \underline{F}_{A,N_A}(s)\right| \mid \mathcal{P}_{A,N_A} \xrightarrow{P} 0$ as $N_A, N_B \to +\infty$.

(iii) For each $s \in \mathcal{S}^\circ$, VAR $\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right) > 0$ for $N_A, N_B = 1, 2, \ldots$

(iv) $\dfrac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}} \mid \Pi_{N_A,N_B} \xrightarrow{d} N(0,1)$ as $N_A, N_B \to +\infty$, uniformly in $s \in \mathcal{S}^\circ$.

(v) $\max_{s\in\mathcal{S}^\circ} |Deff(s)-1| \mid \Pi_{N_A,N_B} \xrightarrow{P} 0$ as $N_A, N_B \to +\infty$, where $Deff(s)$ is equal to

$$\frac{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}{n_A^{-1}\sum_{i\in U_A}\frac{W_{A,i}}{n_A}\left(1[\underline{X}_{A,i}\le s] - \hat{\underline{F}}_{A,N_A}(s)\right)^2 + n_B^{-1}\sum_{i\in U_B}\frac{W_{B,i}}{n_B}\left(1[\overline{X}_{B,i}\le s] - \hat{\overline{F}}_{B,N_B}(s)\right)^2}.$$

(vi) The above conditions hold for all subsequences $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ in place of $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$,

The conditions of Assumption 1 are versions of commonly used large-sample properties in the partial identification and survey sampling literatures for developing asymptotic results (e.g, Wu and Rao, 2006, and Andrews and Soares, 2010). These conditions are mild and satisfied by many survey designs. Condition (i) specifies the means of the sample sizes diverge with the populations' totals. Condition (ii) imposes design-based consistency of the Hájek estimators $\{\hat{\overline{F}}_B \mid \mathcal{P}_{B,N_B}\}_{N_B=1}^{+\infty}$ and $\{\hat{\underline{F}}_A \mid \mathcal{P}_{A,N_A}\}_{N_A=1}^{+\infty}$ that holds with uniformity over $s\in\mathcal{S}^\circ$. Condition (iii) requires the design variances of $\{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ to be positive. Condition (iv) requires that a design-based uniform central limit theorem holds for an appropriately standardized version of $\{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$. To elucidate, let $\Psi(\cdot)_{s,N_A,N_B}$ denote the CDF of the statistic in Condition (vi), then this condition imposes $\lim_{N_A,N_B\to+\infty}\sup_{s\in\mathcal{S}^\circ}\sup_{v\in\mathbb{R}}|\Psi(v)_{s,N_A,N_B} - \Phi(v)| = 0$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Condition (v) is a design-based uniform consistency requirement of the Hájek-type estimator

$$n_A^{-1}\sum_{i\in U_A}\frac{W_{A,i}}{n_A}\left(1[\underline{X}_{A,i}\le s] - \hat{\underline{F}}_{A,N_A}(s)\right)^2 + n_B^{-1}\sum_{i\in U_B}\frac{W_{B,i}}{n_B}\left(1[\overline{X}_{B,i}\le s] - \hat{\overline{F}}_{B,N_B}(s)\right)^2$$

for the design variance $\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$. Their ratio is related to the so-called *design effect* introduced by Wu and Rao (2006), and Condition (v) states that there is no design effect of the estimators $\{\hat{\overline{F}}_B \mid \mathcal{P}_{B,N_B}\}_{N_B=1}^{+\infty}$ and $\{\hat{\underline{F}}_A \mid \mathcal{P}_{A,N_A}\}_{N_A=1}^{+\infty}$. This condition is only imposed for simplicity as we discuss below after the statements of formal results. Condition (vi) is important for establishing (4.1) using its characterization via Theorem 1 below.

The sequence of null models for developing (4.1) is demarcated by the following definition.

Definition 1. The sequence $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1,2,\ldots\}$ is such that every sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, where $\Pi_{N_A,N_B}\in\mathcal{M}^0_{N_A,N_B}$ for all N_A and N_B, satisfies Assumption 1.

We have the following characterization of (4.1) for $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1,2,\ldots\}$ satisfying Definition 1.

Theorem 1. Suppose $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1,2,\ldots\}$ satisfies Definition 1. Let $\mathbb{W}$ denote the set of all sequences of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ where $\Pi_{N_A,N_B}\in\mathcal{M}^0_{N_A,N_B}$ for all N_A and N_B, and satisfies Assumption 1, and let $\alpha\in(0,1)$. Then, condition (4.1) is equivalent to

$$\limsup_{N_A,N_B\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] \le \alpha \quad \forall\,\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}\in\mathbb{W}. \tag{4.2}$$

Proof. See Appendix B.1. □

An important distinction between our framework and that used in the literature on inference for finite populations is that we develop the behavior of the test over a set of sequences of finite populations, whereas that literature's focus has been on a single sequence of that sort (e.g., Wu and Rao, 2006). The result of Theorem 1 shows that this distinction is analogous to the difference between uniform and pointwise asymptotics in the partial identification literature.

Establishing uniform asymptotic validity of the testing procedure, as in (4.2), is crucial for reliable inference in large finite populations, because the limiting distribution of the test statistic is discontinuous as a function of the underlying sequence of finite populations. Discontinuities of this type can create asymptotic size problems that are analogous to those that arise with parameters that are near a boundary (e.g., Andrews and Guggenberger, 2009). The next result establishes the uniform asymptotic validity of the testing procedure.

Theorem 2. Let $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$, $\mathbb{W}$ and α be as in Theorem 1. Then (4.2) holds.

Proof. See Appendix B.2. □

The proposed test is, in fact, asymptotically conservative. As $LR^{(A,B)} \leq 2(L_{UR} - L_R(s))$ holds for any $s \in \mathcal{S}^\circ$, we establish (4.2) by showing the limiting null distribution of $\{2(L_{UR}-L_R(s)) \mid \Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1}$ along $\{\Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1} \in \mathbb{W}$ drifting to/on the boundary of H^1_0 where the rejection probability is highest, is χ^2_1. Hence, a rejection of H^1_0 based on this test using a small significance level constitutes very strong evidence in favor of H^1_1, and hence, is very strong evidence in favor of H_1 defined in (2.1). It should be noted that Condition (v) is important in deriving this limiting distribution theory (see, for example, the proof of Lemma C.1 in the Appendix), and is only imposed for simplicity. When this condition does not hold, an adjustment of the statistics $\{LR^{(A,B)} \mid \Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1}$ is required for the result of Theorem 2 to go through. Specifically, the adjustment $\{LR^{(A,B)}/Deff(s) \mid \Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1}$ so that $\{2(L_{UR} - L_R(s))/Deff(s) \mid \Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1}$ has the same limiting distribution theory along the aforementioned sequences $\{\Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1} \in \mathbb{W}$. The implementation of this adjustment in practice is to plug in a consistent estimator of $Deff(s)$. The form of $Deff(s)$ in Condition (v) implies that estimators of it can be obtained by plugging in an estimator of the design-variance VAR $\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$, and estimators of this design-variance using data from complex surveys are abundant and well established.

The next section presents an empirical application of the methodology this paper develops. The above results are empirically relevant as the surveys we use are nationally representative of finite populations whose totals, N_A and N_B, are in the millions.

5 Empirical Application

To compare the patterns in the average confidence in Lebanese public institutions we use data from the *World Values Survey* (2013) and the *Arab Barometer* (2016 and 2018). Both surveys are nationally representative and report the level of confidence in courts, government, parliament, political parties and police force. Although there is a subtle difference between the words "trust" and "confidence" in English, both words are

Table 1: Nonresponse Probability Estimates

Year	Courts	Government	Parliament	Political Party	Police Force
2013	0.047	0.18	0.07	0.06	0.035
2016	0.004	0.0008	0.005	0.007	0.003
2018	0.0008	0.0025	0.0008	0.0058	0.0008

translated to “thiqa” in the Arabic questionnaire of both surveys, and we use the words “trust” and “confidence” interchangeably. Trust level data from these surveys is ordinal with four categories of responses: *“not at all”*, *“not very much”*, *“a lot”*, and *“a great deal”* to which a numerical scale is applied such that the variable takes integer values from 1 to 4, respectively. In terms of our notation, we have $\mathcal{S} = \{1, 2, 3, 4\}$. Denoting the CDFs of trust levels in 2013, 2016, and 2018, by F_{13}, F_{16} and F_{18}, respectively, for a given institution, the foregoing investigation of interest can be formulated as inference for the first-order (strict) stochastic dominance chain:

$$F_{13}(s) < F_{16}(s) < F_{18}(s) \text{ for } s = 1, 2, 3. \tag{5.1}$$

The datasets’ sizes are $n_{13} = n_{16} = 1,200$ and $n_{18} = 2,400$. However, the question on trust in “Political Parties” was only asked on a subsample of size 1,215 in 2018. Table 1 reports the nonresponse frequencies. This frequency for trust in the “Government” in 2013 survey is quite large (18%), and the rest are all less than 10%. While it is tempting to consider particular assumptions/models to explain these nonresponse frequencies, they are unverifiable in practice and may yield biased inferences. Using the worst-case bounds, as we propose, permits the entire spectrum of models for these frequencies in inference. This approach is especially useful when this frequency is large, like with “Government”, as there can be a diversity of explanations for it including fear of retaliation from public authorities. Our approach circumvents this difficulty by comparing the CDFs F_{13}, F_{16} and F_{18}, using their worst-case bounds, because it achieves robustness to the nature of the nonresponse-generating process. Denote the worst-case upper bounds on F_{13} and F_{16} by $\overline{F}_{13}$ and $\overline{F}_{16}$, respectively, and the worst-case lower bounds on F_{16} and F_{18} by $\underline{F}_{16}$ and $\underline{F}_{18}$, respectively. If we can establish

$$\overline{F}_{13}(s) < \underline{F}_{16}(s) \quad \text{and} \quad \overline{F}_{16}(s) < \underline{F}_{18}(s) \quad \text{for } s = 1, 2, 3, \tag{5.2}$$

then the chain of inequalities (5.1) holds, because the CDFs and their worst-case bounds would satisfy

$$F_{13}(s) \leq \overline{F}_{13}(s) < \underline{F}_{16}(s) \leq F_{16}(s) \leq \overline{F}_{16}(s) < \underline{F}_{18}(s) \leq F_{18}(s) \quad \text{for } s = 1, 2, 3.$$

A straightforward extension of the proposed testing procedure applies to testing the inequalities in (5.2)

Table 2: Realised values of test statistics and decisions based on 5% significance level

null hypothesis	Courts	Government	Parliament	Political Party	Police Force
H_0^2	0 Do not Reject	0 Do not Reject	0 Do not Reject	0 Do not Reject	0 Do not Reject
$H_0^{2,1}$	66.7 Reject	21.66 Reject	10.37 Reject	17.22 Reject	0 Do not Reject

under the alternative. The null hypothesis in this setting is

$$H_0^2 : \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_{13}(s) - \underline{F}_{16}(s)\right) \geq 0 \text{ or } \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_{16}(s) - \underline{F}_{18}(s)\right) \geq 0,$$

By treating H_0^2 as the union of two sub-hypotheses $H_0^{2,1} : \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_{13}(s) - \underline{F}_{16}(s)\right) \geq 0$ and $H_0^{2,2} : \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_{16}(s) - \underline{F}_{18}(s)\right) \geq 0$, an intersection-union procedure applies to this testing problem. The test statistic is $LR = \min\left\{LR^{(16,13)}, LR^{(18,16)}\right\}$ where $LR^{(16,13)}$ and $LR^{(18,16)}$ are the test statistics corresponding to the test problems $H_0^{2,i}$ versus $H_1^{2,i}$: not $H_0^{2,i}$, for $i = 1, 2$. The decision rule is to reject H_0^2 if and only if $LR > c(\alpha)$, where $c(\alpha)$ is the $1 - \alpha$ quantile from the χ_1^2 distribution. By using $LR^{(16,13)}$ $LR^{(18,16)}$ as the building blocks for LR, a straightforward extension of the framework of the previous section can be used to show that this testing problem is asymptotically valid with uniformity.

The first row of Table 2 reports the realized values of LR and the conclusions of the hypothesis tests of H_0^2. Unfortunately, there is no evidence at the 5% level to reject H_0^2 across all institutions. As the realized values of the test statistic are all equal to zero, it follows that this conclusion also holds for any significance level. Therefore, there is no evidence for the chain (5.2). This finding is mainly driven by the comparisons between $\overline{F}_{16}$ and $\underline{F}_{18}$. For example, the right panel of Figure 1 reports $\hat{\overline{F}}_{16}$ and $\hat{\underline{F}}_{18}$ for trust in the government. From the figure, it is clear that the null hypothesis $H_0^{2,2}$ holds in the sample, implying that $LR^{(18,16)} = 0$, and consequently, $LR = 0$. We obtain similar results for trust in the other public institutions, indicating that we cannot robustly rank the trust levels in any public institutions between 2016 and 2018.

Contrastingly, dominance holds in the sample between 2013 and 2016 across all public institutions. For example, the left panel of Figure 1 reports $\hat{\overline{F}}_{13}$ and $\hat{\underline{F}}_{16}$ for trust in the "Government", showing that $\hat{\overline{F}}_{13}(s) < \hat{\underline{F}}_{16}(s)\ \forall s \in \mathcal{S}^\circ$. The second row of Table 2 reports the realized values of $LR^{(16,13)}$ with the conclusion of the hypothesis testing of $H_0^{2,1}$ for each institution. For confidence in courts, government, parliament and political parties, we reject the null hypothesis $H_0^{2,1}$ in favour of the alternative $\overline{F}_{13}(s) < \underline{F}_{16}(s)\ \forall s \in \mathcal{S}^\circ$. It is also worth noting that for these variables, $H_0^{2,1}$ is also rejected at the 1% level.[4] However, $H_0^{2,1}$ cannot be rejected for trust in the police force.

We do not find evidence that confidence levels in public institutions have been diminishing or increasing between 2016 and 2018. However, the results provide very strong evidence that trust levels in the Lebanese courts, government, parliament and political parties have dropped between 2013 and 2016, as the dominance

[4]The critical value in this case is approximately 6.63.

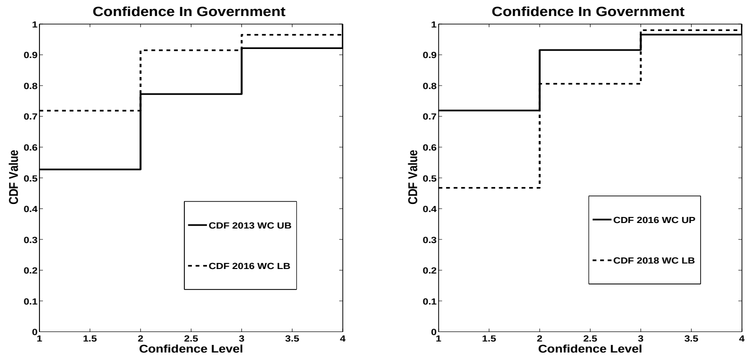

Figure 1: Trust in the government for the period 2013-2016 and 2016-2018

orderings are statistically significant at the 1% level for these institutions. Given that the test employs a conservative critical value, these findings are quite powerful. Furthermore, these conclusions are robust to any assumptions on (i) the numerical scale applied to the trust level categories, and (ii) the nonresponse-generating process. Consequently, an analyst cannot obtain a different result with any set of reasonable assumptions on the numerical scale and nonresponse mechanism.

6 Conclusion

This paper extends the testing procedure of the Davidson and Duclos (2013) for first-order stochastic dominance to the empirically relevant setting of survey sampling for ordinal data and uses worst-case bounds to account for nonresponse. Additionally, it presents evidence on erosion of trust in public institutions in Lebanon occurring prior to October 2019, showing that the worst-case bounds can be informative in practice. A direction of future research is to extend the proposed method to second-order stochastic dominance as it is useful for robust comparisons based on classes of ordinal inequality indices (e.g., Jenkins, 2019).

References

Allison, R. and J. E. Foster (2004). Measuring health inequality using qualitative data. *Journal of Health Economics 23*(3), 505 – 524.

Alvarez-Esteban, P. C., E. del Barrio, J. A. Cuesta-Albertos, and C. Matran (2017, 08). Models for the assessment of treatment improvement: The ideal and the feasible. *Statist. Sci. 32*(3), 469–485.

Andrews, D. W. and P. Guggenberger (2009, 6). Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. *Econometric Theory 25*, 669–709.

Andrews, D. W. and X. Shi (2017). Inference based on many conditional moment inequalities. *Journal of Econometrics 196*(2), 275 – 287.

Andrews, D. W. K. and G. Soares (2010). Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection. *Econometrica 78*(1), 119–157.

Baalbaki, N. (2018). Are the Lebanese happy? corruption and resilience in the light of the parliamentary elections. https://lb.boell.org/en/2018/03/27/are-lebanese-happy-corruption-and-resilience-light-parliamentary-elections.

Berger, R. L. (1988). A nonparametric, intersection-union test for stochastic order. In *Statistical decision theory and related topics, IV, Vol. 2 (West Lafayette, Ind., 1986)*, pp. 253–264. Springer, New York.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica 75*(2), 323–363.

Canay, I. A. (2010). EL Inference for Partially Identified Models:Large Deviations Optimality and Bootstrap Validity. *Journal of Econometrics 156*(2), 408–425.

Chen, J. and R. R. Sitter (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica 9*, 385–406.

Cowell, F. A. and E. Flachaire (2017). Inequality with ordinal data. *Economica 84*(334), 290–321.

Davidson, R. and J.-Y. Duclos (2013). Testing for Restricted Stochastic Dominance. *Econometric Reviews 32*(1), 84–125.

Dutta, I. and J. Foster (2013). Inequality of happiness in the u.s.: 1972–2010. *Review of Income and Wealth 59*(3), 393–415.

Haile, P. and E. Tamer (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy 111*(1), 1–51.

Horowitz, J. L. and C. F. Manski (1995). Identification and robustness with contaminated and corrupted data. *Econometrica 63*(2), 281–302.

Jenkins, S. P. (2019, November). Inequality comparisons with ordinal data. IZA Discussion Paper No.12811.

Jenkins, S. P. (2020, March). Comparing distributions of ordinal data. IZA Discussion Paper No.13057.

Lee, D. S. (2009, 07). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies 76*(3), 1071–1102.

Madden, D. (2009). Mental stress in ireland, 1994–2000: a stochastic dominance approach. *Health Economics 18*(10), 1202–1217.

Manski, C. F. (1994). *The selection problem*, Volume 1 of *Econometric Society Monographs*, pp. 143–170. Cambridge University Press.

Manski, C. F. (2016). Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics 191*(2), 293 – 301.

Owen, A. (2001). *Empirical Likelihood*, Volume 92 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Rubin-Bleuer, S. and I. S. Kratina (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics 33*(6), 2789–2810.

Shi, X. (2015). Model selection tests for moment inequality models. *Journal of Eonometrics 187*, 1–17.

Whang, Y.-J. (2019). *Econometric Analysis of Stochastic Dominance: Concepts, Methods, Tools, and Applications*. Themes in Modern Econometrics. Cambridge University Press.

Wu, C. and J. N. K. Rao (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 34*(3), 359–375.

A Supplementary Material: Outline

This Appendix provides supplementary material to this paper. It is organized as follows.

- Appendix B presents the proofs of Theorems 1 and 2.
- Appendix C presents technical lemmas that are used in the proof of Theorem 2.

B Proofs: Theorems 1 and 2

B.1 Theorem 1

Proof. We first show (4.1) $\implies$ (4.2). The proof proceeds by the direct method. Suppose that (4.1) holds, and let $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$. Then

$$E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] \leq \sup_{\Pi \in \mathcal{M}^0_{N_A,N_B}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right] \ \forall N_A, N_B, \tag{B.1}$$

holds. Taking the limit superiors on both sides of the inequality (B.1) implies the inequality (4.2). As the sequence $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$ was arbitrary, this inequality holds for all such sequences.

Now we shall prove the reverse direction:(4.2) $\implies$ (4.1). The proof proceeds by contraposition. Suppose that (4.1) does not hold, i.e.,

$$\limsup_{N_A,N_B \to +\infty} \sup_{\Pi \in \mathcal{M}^0_{N_A,N_B}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right] > \alpha. \tag{B.2}$$

Then we have to construct a sequence $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$ such that

$$\limsup_{N_A,N_B \to +\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] > \alpha$$

to prove the result. To that end,the condition (B.2) implies the largest subsequential limit of the sequence $\left\{\sup_{\Pi \in \mathcal{M}^0_{N_A,N_B}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right]\right\}_{N_A,N_B=1}^{+\infty}$ exceeds α. Thus, there is a sequence $\{N_{A,m}, N_{B,m}\}_{m=1}^{+\infty}$ such that the limit of $\left\{\sup_{\Pi \in \mathcal{M}^0_{N_{A,m},N_{B,m}}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right]\right\}_{m=1}^{+\infty}$ exceeds α,; e.g., the limit is equal to $\alpha + \nu$ where $\nu > 0$. Now let $\epsilon > 0$ be such that $\nu > \epsilon > 0$. For each m there exists $\Pi'_{N_{A,m},N_{B,m}} \in \mathcal{M}^0_{N_{A,m},N_{B,m}}$ such that

$$E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi'_{N_{A,m},N_{B,m}}\right] > \sup_{\Pi \in \mathcal{M}^0_{N_{A,m},N_{B,m}}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right] - \epsilon. \tag{B.3}$$

Now taking limit superior of both sides of (B.3) with respect to m, yields

$$\limsup_{m\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi'_{N_{A,m},N_{B,m}}\right] > \limsup_{m\to+\infty} \sup_{\Pi\in\mathcal{M}^0_{N_{A,m},N_{B,m}}} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right] - \epsilon$$
$$> \alpha + \nu - \epsilon > \alpha.$$

Thus, we have constructed a sequence of populations $\left\{\Pi'_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty} \in \mathbb{W}$ with the desired property. This concludes the proof. $\square$

B.2 Theorem 2

Proof. The proof proceeds by the direct method. Let $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$. Denote by β the event of dominance in the sample. i.e., $\hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ$. From the definition of the test statistic, the largest subsequential limit of the sequence $\{E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right]\}_{N_A,N_B=1}^{+\infty}$ is along subsequences of populations that give the event β the highest probability and the largest values of the statistic on that event. Otherwise, the test statistics are more likely to equal to zero, implying the rejection probability would be smaller.

Let $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ denote a subsequence of $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$. Note that for each $s \in \mathcal{S}^\circ$

$$\frac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_{A,m},N_{B,m}}\right)}} \mid \Pi_{N_{A,m},N_{B,m}}$$

is equal to

$$\frac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) - \left(\overline{F}_{B,m}(s) - \underline{F}_{A,m}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_{A,m},N_{B,m}}\right)}} \mid \Pi_{N_{A,m},N_{B,m}} + \frac{\overline{F}_{B,m}(s) - \underline{F}_{A,m}(s)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_{A,m},N_{B,m}}\right)}}. \tag{B.4}$$

Therefore, by Assumption 1, $\lim_{m\to+\infty} E[1[\beta] \mid \Pi_{N_{A,m},N_{B,m}}] = \text{Prob}\left[Z_s < 0\ \forall s \in \mathcal{S}^\circ\right]$ where $Z \overset{d}{\sim} MVN(c, \Omega)$, where $c = [c_1, c_2, \ldots, c_{K-1}]$ with

$$c_j = \lim_{m\to+\infty} \frac{\overline{F}_{B,m}(s_j) - \underline{F}_{A,m}(s_j)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s_j) - \hat{\underline{F}}_A(s_j) \mid \Pi_{N_{A,m},N_{B,m}}\right)}} \quad \text{for } j = 1, \ldots, K-1,$$

and Ω is limiting covariance matrix of the first terms in (B.4). From this limiting multivariate normality, the event has highest probability along sequences where the vector c has the following structure: there exists a unique i such that $c_i = 0$ and $c_j = -\infty$ for all $j \in \{1, 2, \ldots, K-1\} - \{i\}$. Furthermore, this configuration of the vector c happens to also give the largest values of the test statistic as it corresponds to populations closest to H_1^1 that satisfy the restrictions of the null hypothesis. Such populations are either on or get close to the boundary of the null model $\mathcal{M}^0_{N_A,N_B}$, when N_A and N_B are large. Note that the boundary of $\mathcal{M}^0_{N_A,N_B}$ is defined as

$$\partial\mathcal{M}^0_{N_A,N_B} = \left\{\Pi \in \mathcal{M}^0_{N_A,N_B} : \exists s \in \mathcal{S}^\circ \text{ such that } \overline{F}_B(s) = \underline{F}_A(s)\right\}.$$

Let $\mathcal{C}$ denote the set of all vectors c, where one component is equal to zero and the rest equal to $-\infty$. There are two cases for the sequences to consider: infinitely many elements of $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ either satisfy (i) $\Pi_{N_A,N_B} \in \partial\mathcal{M}^0_{N_A,N_B}$, or (ii) $\Pi_{N_A,N_B} \in \mathcal{M}^0_{N_A,N_B} - \partial\mathcal{M}^0_{N_A,N_B}$. In both of these cases we shall confine our analysis to subsequences $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ that have the limiting structure on $c \in \mathcal{C}$. Note that the second case corresponds to subsequences that are in the interior of the model of the null which are drifting to the boundary.

We start with case (i) by considering a subsequence of populations $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ that satisfies $\Pi_{N_{A,m},N_{B,m}} \in \partial\mathcal{M}^0_{N_{A,m},N_{B,m}}$ for all m. Any subsequence of this sort has at least one component of c equal to zero. In fact this property is enough to obtain the desired result in this case. Let $S^\circ_{N_{A,m},N_{B,m}} = \{s \in \mathcal{S}^\circ : \overline{F}_{B,N_{B,m}}(s) = \underline{F}_{A,N_{A,m}}(s)\}$ and $s_{\text{min}} = \min S^\circ_{N_{A,m},N_{B,m}}$. Then from the definition of the test statistic

$$E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] \leq E\left[1[2(L_{UR} - L_R(s_{\text{min}})) > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right]$$

holds. By taking the limit superior of both sides of this inequality, we obtain

$$\begin{aligned}\limsup_{m\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] &\leq \limsup_{m\to+\infty} E\left[1[2(L_{UR} - L_R(s_{\text{min}})) > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] \\ &= \alpha,\end{aligned}$$

where the equality follows from applying Lemma C.1 to the sequence $\{2(L_{UR}-L_R(s_{\text{min}})) \mid \Pi_{N_{A,m},N_{B,m}}\}_{m=1}^{+\infty}$. Specifically, Lemma C.1 establishes the limiting distribution of this sequence of statistics is equal to χ^2_1. The above derivations covers the case of subsequences whose corresponding vector $c \in \mathcal{C}$.

Next, we focus on case (ii). Let $c \in \mathcal{C}$ be such that $c_e = 0$ and $c_j = -\infty$ for all $j \neq e$. Then, from the definition of the test statistic

$$E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] \leq E\left[1[2(L_{UR} - L_R(s_e)) > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right],$$

holds. By taking the limit superior of both sides of this inequality, we obtain

$$\begin{aligned}\limsup_{m\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] &\leq \limsup_{m\to+\infty} E\left[1[2(L_{UR} - L_R(s_e)) > c(\alpha)] \mid \Pi_{N_{A,m},N_{B,m}}\right] \\ &= \alpha,\end{aligned}$$

where the equality follows from applying Lemma C.2 to the sequence $\{2(L_{UR}-L_R(s_e)) \mid \Pi_{N_{A,m},N_{B,m}}\}_{m=1}^{+\infty}$. This concludes the proof. □

C Technical Lemmas

Before presenting the technical results, we briefly develop the Lagrangian method for the PELF (3.2) and its first order conditions because the technical results are based upon it. Given a population Π, we have access to two independent survey samples from it given by $U_A \in \mathcal{U}_A$ and $U_B \in \mathcal{U}_B$. The maximisation of the PELF corresponding to the restriction (3.1) at $s \in \mathcal{S}^\circ$ is thus

$$\max_{p_1^Y,\ldots,p_{n_Y}^Y : Y\in\{A,B\}} \sum_{Y\in\{A,B\}} \sum_i W_{Y,i} \log p_i^Y \text{ subject to } p_i^Y > 0,\ \sum_i p_i^Y W_{Y,i} = 1\ \forall i \in U_Y, Y = A, B,$$

$$\text{and } 0 = \sum_{i\in U_B} \sum_{j\in U_A} p_j^A p_i^B H_{i,j}(s).$$

The Lagrangian is

$$\mathcal{L} = \sum_{Y\in\{A,B\}} \left[\sum_{i\in U_Y} W_{Y,i} \log p_i^Y + \gamma_Y \left(1 - \sum_{i\in U_Y} p_i^Y W_{Y,i} \right) \right] - \lambda \sum_{i\in U_B} \sum_{j\in U_A} p_j^A p_i^B H_{i,j}(s) \tag{C.1}$$

in which γ_A, γ_B and λ are the multiplies on the constraints. The first order conditions are

$$0 = \frac{\partial \mathcal{L}}{\partial p_i^B} \quad \forall i \in U_B \iff \frac{W_{B,i}}{p_i^B} = W_{B,i}\gamma_Y + \sum_{j\in U_A} p_j^A H_{i,j}(s) \quad \forall i \in U_B \tag{C.2}$$

$$0 = \frac{\partial \mathcal{L}}{\partial p_j^Y} \quad \forall j \in U_A \iff \frac{W_{A,i}}{p_i^A} = W_{A,i}\gamma_Y + \sum_{i\in U_B} p_i^B H_{i,j}(s) \quad \forall j \in U_A \tag{C.3}$$

$$0 = \frac{\partial \mathcal{L}}{\partial \gamma_Y} \quad Y = A, B \iff \sum_{i\in U_Y} p_i^Y W_{Y,i} = 1 \quad Y = A, B \tag{C.4}$$

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda} \iff \sum_{i\in U_B} \sum_{j\in U_A} p_j^A p_i^B H_{i,j}(s) = 0. \tag{C.5}$$

By multiplying (C.2) by p_i^B and summing over $i \in U_B$, one obtains $\gamma_B = n_B$ upon using the constraints. A similar argument can be used to show that $\gamma_A = n_A$ by multiplying (C.3) by p_j^A and summing over $j \in U_A$. This yields

$$p_i^B = \left[n_B + \lambda \sum_{j\in U_A} p_j^A W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\text{min}}) \right]^{-1} \qquad \forall i \in U_B,$$

$$p_j^A = \left[n_A + \lambda \sum_{i\in U_B} p_i^B W_{B,i} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\text{min}}) \right]^{-1} \qquad \forall j \in U_A.$$

C.1 Results

Lemma C.1. Let $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ and $\mathbb{W}$ be as in Theorem 1, and let $\alpha \in (0, 1)$. Given a sequence $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$, suppose that there is a subsequence of $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ such that

$\Pi_{N_{A,m},N_{B,m}} \in \partial\mathcal{M}^0_{N_{A,m},N_{B,m}}$ for all m. Furthermore, let $S^\circ_{N_{A,m},N_{B,m}} = \{s \in \mathcal{S}^\circ : \overline{F}_{B,N_{B,m}}(s) = \underline{F}_{A,N_{A,m}}(s)\}$ and $s_{\min} = \min S^\circ_{N_{A,m},N_{B,m}}$. Then, $2(L_{UR} - L_R(s_{\min})) \mid \Pi_{N_{A,m},N_{B,m}} \xrightarrow{d} \chi^2_1$ as $m \to +\infty$.

Proof. We have the survey samples $\{X_{Y,i}, D_{Y,i}, W_{Y,i} : i \in U_Y\}$ for $Y = A, B$, from population $\Pi_{N_{A,m},N_{B,m}}$. For notational simplicity, we shall drop "$\mid \Pi_{N_{A,m},N_{B,m}}$" from the notation, as it is clear that we are taking limits along $\{\Pi_{N_{A,m},N_{B,m}}\}_{m=1}^{+\infty}$. Let $S^\circ_{N_{A,m},N_{B,m}} = \{s \in \mathcal{S}^\circ : \overline{F}_{B,N_{B,m}}(s) = \underline{F}_{A,N_{A,m}}(s)\}$ and $s_{\min} = \min S^\circ_{N_{A,m},N_{B,m}}$. Note that $s_{\min}$ is non-random and changes with the population sequence.

The corresponding samples we use in the testing procedure are $\{\overline{X}_{B,i}, W_{B,i} : i \in U_B\}$ and $\{\underline{X}_{A,j}, W_{A,j} : j \in U_A\}$. Using Lagrange multipliers for deriving the solutions to the PELF problem (3.2), we find

$$p_i^B = \left[n_B + \lambda \sum_{j \in U_A} p_j^A W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min})\right]^{-1} \quad \forall i \in U_B, \tag{C.6}$$

$$p_j^A = \left[n_A + \lambda \sum_{i \in U_B} p_j^B W_{B,i} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min})\right]^{-1} \quad \forall j \in U_A. \tag{C.7}$$

where λ is defined by $\sum_i \sum_j p_j^A p_i^B W_{B,i} W_{B,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min}) = 0$.

We simplify our notation by matching it to that in Section 11.4 of Owen (2001) and follow his derivations. In doing so, we shall also drop the index sets for summations: U_A for j and U_B for i. Introduce the terms

$$\bar{H}_{i\bullet}(s_{\min}) = \frac{1}{n_A} \sum_j W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min}), \quad \tilde{H}_{i\bullet}(s_{\min}) = \sum_j p_j^A W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min}) \tag{C.8}$$

$$\bar{H}_{\bullet j}(s_{\min}) = \frac{1}{n_B} \sum_i W_{B,i} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min}), \quad \tilde{H}_{\bullet j}(s_{\min}) = \sum_i p_i^B W_{B,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min}), \tag{C.9}$$

$H_{ij}(s_{\min}) = W_{B,i} W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\min})$, and $\bar{H}_{\bullet\bullet}(s_{\min}) = \hat{\overline{F}}_B(s_{\min}) - \hat{\underline{F}}_A(s_{\min})$. Then,

$$p_i^B = \frac{1}{n_B}\left[1 - \left(\frac{\lambda \tilde{H}_{i\bullet}(s_{\min})}{n_B}\right) + \left(\frac{\lambda \tilde{H}_{i\bullet}(s_{\min})}{n_B}\right)^2 - \left(\frac{\lambda \tilde{H}_{i\bullet}(s_{\min})}{n_B}\right)^3 + \cdots\right] \ \forall i$$

$$p_j^A = \frac{1}{n_A}\left[1 - \left(\frac{\lambda \tilde{H}_{\bullet j}(s_{\min})}{n_A}\right) + \left(\frac{\lambda \tilde{H}_{\bullet j}(s)}{n_A}\right)^2 - \left(\frac{\lambda \tilde{H}_{\bullet j}(s_{\min})}{n_A}\right)^3 + \cdots\right] \ \forall j.$$

Substituting these values into $\sum_i \sum_j p_j^A p_i^B H_{ij}(s_{\min}) = 0$, we obtain

$$\begin{aligned} 0 = \bar{H}_{\bullet\bullet}(s_{\min}) &- \lambda\left[\frac{\sum_i \sum_j H_{ij}(s_{\min}) \tilde{H}_{i\bullet}(s_{\min})}{n_B^2 n_A} + \frac{\sum_i \sum_j H_{ij}(s_{\min}) \tilde{H}_{\bullet j}(s_{\min})}{n_A^2 n_B}\right] \\ &+ \lambda^2\left[\frac{\sum_i \sum_j H_{ij}(s_{\min}) \tilde{H}^2_{i\bullet}(s_{\min})}{n_B^3 n_A} + \frac{\sum_i \sum_j H_{ij}(s_{\min}) \tilde{H}^2_{\bullet j}(s_{\min})}{n_A^3 n_B} + \frac{\sum_i \sum_j H_{ij}(s_{\min}) \tilde{H}_{i\bullet}(s_{\min}) \tilde{H}_{\bullet j}(s_{\min})}{n_B^2 n_A^2}\right] + \cdots \end{aligned} \tag{C.10}$$

The equality (C.10) is equivalent to a convergent alternating series in λ, with the modulus of the coefficients of λ^ℓ bounded by ℓ, for $\ell = 1, 2, \ldots$. Consequently, $|\lambda| < 1$. For large m, condition (i) of Assumption 1 implies that we can ignoring higher-order terms in λ to find $\lambda \overset{a}{=} \bar{H}_{\bullet\bullet}(s_{\text{min}})/D(s_{\text{min}})$, where $\overset{a}{=}$ denotes asymptotic equivalence as $m \to +\infty$, and

$$D(s_{\text{min}}) = \frac{1}{n_B^2 n_A} \sum_i \sum_j H_{ij}(s_{\text{min}}) \bar{H}_{i\bullet}(s_{\text{min}}) + \frac{1}{n_B n_A^2} \sum_i \sum_j H_{ij}(s_{\text{min}}) \bar{H}_{\bullet i}(s_{\text{min}}) \tag{C.11}$$

$$D(s_{\text{min}}) = \frac{1}{n_B} \sum_i \frac{W_{B,i}}{n_B} \bar{H}_{i\bullet}^2(s_{\text{min}}) + \frac{1}{n_A} \sum_j \frac{W_{A,i}}{n_A} \bar{H}_{\bullet j}^2(s_{\text{min}}). \tag{C.12}$$

In finding this $D(s_{\text{min}})$, the term

$$\tilde{H}_{i\bullet}(s_{\text{min}}) = \bar{H}_{i\bullet}(s_{\text{min}}) - \frac{\lambda}{n_A^2} \sum_j W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s_{\text{min}}) \tilde{H}_{\bullet j}(s_{\text{min}}) \tag{C.13}$$

has been replaced by $\bar{H}_{i\bullet}$ and $\tilde{H}_{\bullet j}$ has been replaced by $\bar{H}_{\bullet j}$, with the differences being absorbed into the coefficient of λ^2.

Now keeping up to order λ^2 in the profile PELF and using a standard expansion of the logarithm function $\log(1 + x)$ for $|x| < 1$, we find

$$\begin{aligned}
2(L_{UR} - L_R(s_{\text{min}})) &= 2 \sum_i W_{B,i} \log\left(1 + \frac{\lambda \tilde{H}_{i\bullet}(s_{\text{min}})}{n_B}\right) + 2 \sum_j W_{A,i} \log\left(1 + \frac{\lambda \tilde{H}_{\bullet j}(s_{\text{min}})}{n_A}\right) \\
&\overset{a}{=} 2 \sum_i W_{B,i} \left[\frac{\lambda \tilde{H}_{i\bullet}(s_{\text{min}})}{n_B} - \frac{1}{2}\left(\frac{\lambda \tilde{H}_{i\bullet}(s_{\text{min}})}{n_B}\right)^2\right] \\
&\quad + 2 \sum_j W_{A,i} \left[\frac{\lambda \tilde{H}_{\bullet j}(s_{\text{min}})}{n_A} - \frac{1}{2}\left(\frac{\lambda \tilde{H}_{\bullet j}(s_{\text{min}})}{n_A}\right)^2\right].
\end{aligned}$$

Replacing $\tilde{H}(s_{\text{min}})$'s by corresponding $\bar{H}(s_{\text{min}})$'s and keeping terms to order λ^2, we get

$$\begin{aligned}
2(L_{UR} - L_R(s_{\text{min}})) &\overset{a}{=} 2 \sum_i \frac{W_{B,i} \lambda \bar{H}_{i\bullet}(s_{\text{min}})}{n_B} - \frac{2\lambda^2}{n_A} \sum_j \frac{W_{A,i}}{n_A} \bar{H}_{\bullet j}^2(s_{\text{min}}) - \sum_i W_{B,i} \left(\frac{\lambda \bar{H}_{i\bullet}(s_{\text{min}})}{n_B}\right)^2 \\
&\quad + 2 \sum_j \frac{W_{A,j} \lambda \bar{H}_{\bullet j}(s_{\text{min}})}{n_A} - \frac{2\lambda^2}{n_B} \sum_i \frac{W_{B,i}}{n_B} \bar{H}_{i\bullet}^2(s_{\text{min}}) - \sum_j W_{A,j} \left(\frac{\lambda \bar{H}_{\bullet j}(s_{\text{min}})}{n_A}\right)^2 \\
&= 4\lambda \bar{H}_{\bullet\bullet}(s_{\text{min}}) - 3\lambda^2 D(s_{\text{min}}) \\
&\overset{a}{=} \bar{H}_{\bullet\bullet}^2(s_{\text{min}})/D(s_{\text{min}}) \\
&= \left(\frac{\bar{H}_{\bullet\bullet}(s_{\text{min}})}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_{\text{min}}) \mid \Pi_{N_{A,m}, N_{B,m}})}}\right)^2 \frac{\text{VAR}(\bar{H}_{\bullet\bullet}(s_{\text{min}}) \mid \Pi_{N_{A,m}, N_{B,m}})}{D(s_{\text{min}})}
\end{aligned}$$

by Condition (iii) of Assumption 1. Now Conditions (ii) and (vi) along this subsequence in the model of the null imply that $\bar{H}_{\bullet\bullet}(s_{\min}) \xrightarrow{P} 0$ as $m \to +\infty$, so that under the conditions of Assumption 1,

$$\left(\frac{\bar{H}_{\bullet\bullet}(s_{\min})}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_{\min}) \mid \Pi_{N_{A,m},N_{B,m}})}} \right)^2 \xrightarrow{d} \chi_1^2,$$

as $m \to +\infty$, holds, along this subsequence in the model of the null. Thus, to complete the argument, we need to show that

$$\text{VAR}(\bar{H}_{\bullet\bullet}(s_{\min}) \mid \Pi_{N_{A,m},N_{B,m}})/D(s_{\min}) \xrightarrow{P} 1,$$

as $m \to +\infty$, holds, under Assumption 1. Unpacking the form of $D(s_{\min})$ reveals that it is equal to

$$\begin{aligned} D(s_{\min}) &= \frac{1}{n_A} \sum_{i \in U_A} \frac{W_{A,i}}{n_A} \left(1[\underline{X}_{A,i} \leq s_{\min}] - \hat{\overline{F}}_{B,N_B}(s_{\min}) \right)^2 \\ &\quad + \frac{1}{n_B} \sum_{i \in U_B} \frac{W_{B,i}}{n_B} \left(1[\overline{X}_{B,i} \leq s_{\min}] - \hat{\underline{F}}_{A,N_A}(s_{\min}) \right)^2 \\ &= \frac{1}{n_A} \sum_{i \in U_A} \frac{W_{A,i}}{n_A} \left(1[\underline{X}_{A,i} \leq s_{\min}] - \hat{\underline{F}}_{A,N_A}(s_{\min}) + \bar{H}_{\bullet\bullet}(s_{\min}) \right)^2 \\ &\quad + \frac{1}{n_B} \sum_{i \in U_B} \frac{W_{B,i}}{n_B} \left(1[\overline{X}_{B,i} \leq s_{\min}] - \hat{\overline{F}}_{B,N_B}(s_{\min}) + \bar{H}_{\bullet\bullet}(s_{\min}) \right)^2. \end{aligned}$$

Now Conditions (ii) and (vi) along this subsequence in the model of the null imply that $\bar{H}_{\bullet\bullet}(s_{\min}) \xrightarrow{P} 0$ as $m \to +\infty$, so that

$$\begin{aligned} D(s_{\min}) &\overset{a}{=} \frac{1}{n_A} \sum_{i \in U_A} \frac{W_{A,i}}{n_A} \left(1[\underline{X}_{A,i} \leq s_{\min}] - \hat{\underline{F}}_{A,N_A}(s_{\min}) \right)^2 \\ &\quad + \frac{1}{n_B} \sum_{i \in U_B} \frac{W_{B,i}}{n_B} \left(1[\overline{X}_{B,i} \leq s_{\min}] - \hat{\overline{F}}_{B,N_B}(s_{\min}) \right)^2. \end{aligned}$$

Consequently, $\text{VAR}(\bar{H}_{\bullet\bullet}(s_{\min}) \mid \Pi_{N_{A,m},N_{B,m}})/D(s_{\min})$ is asymptotically equivalent to

$$\frac{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B} \right)}{n_A^{-1} \sum_{i \in U_A} \frac{W_{A,i}}{n_A} \left(1[\underline{X}_{A,i} \leq s] - \hat{\underline{F}}_{A,N_A}(s) \right)^2 + n_B^{-1} \sum_{i \in U_B} \frac{W_{B,i}}{n_B} \left(1[\overline{X}_{B,i} \leq s] - \hat{\overline{F}}_{B,N_B}(s) \right)^2}.$$

and we can apply Conditions (v) and (iv) to show that it converges to unity along the subsequence in the model of the null hypothesis. This concludes the proof. □

For the next result, we need some additional notation. Given a sequence, $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ and a subsequence of it, $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$, define

$$c_j = \lim_{m\to+\infty} \frac{\overline{F}_{B,m}(s_j) - \underline{F}_{A,m}(s_j)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s_j) - \hat{\underline{F}}_A(s_j) \mid \Pi_{N_{A,m},N_{B,m}}\right)}} \quad \text{for } j = 1, \ldots, K-1. \tag{C.14}$$

Conditions on these limits is how we demarcate the subsequences we consider.

Lemma C.2. Let $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ and $\mathbb{W}$ be as in Theorem 1, and let $\alpha \in (0,1)$. Given a sequence $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty} \in \mathbb{W}$, suppose that there is a subsequence $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ such that (i) $\Pi_{N_{A,m},N_{B,m}} \in \mathcal{M}^0_{N_A,N_B} - \partial\mathcal{M}^0_{N_{A,m},N_{B,m}}$ for all m, and (ii) $c_e = 0$ and $c_j = -\infty$ for all $j \neq e$, where these constants are defined in (C.14). Then, $2(L_{UR} - L_R(s_e)) \mid \Pi_{N_{A,m},N_{B,m}} \xrightarrow{d} \chi^2_1$ as $m \to +\infty$.

Proof. We have the survey samples $\{X_{Y,i}, D_{Y,i}, W_{Y,i} : i \in U_Y\}$ for $Y = A, B$, from population $\Pi_{N_{A,m},N_{B,m}}$. For notational simplicity, we shall drop "$\mid \Pi_{N_{A,m},N_{B,m}}$" from the notation, as it is clear that we are taking limits along $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$. Let $s_e \in \mathcal{S}^\circ$ correspond to the constant c_e. Note that s_e is non-random and does not change with the population sequence.

We follow steps identical to those in Lemma C.1 to obtain

$$2(L_{UR} - L_R(s_e)) \overset{a}{=} \left(\frac{\bar{H}_{\bullet\bullet}(s_e) - (\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e)) + (\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e))}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})}}\right)^2 * \text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})/D(s_e)$$

Noting that $c_e = 0$ means

$$0 = c_e = \lim_{m\to+\infty} \frac{\overline{F}_{B,m}(s_e) - \underline{F}_{A,m}(s_e)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s_e) - \hat{\underline{F}}_A(s_e) \mid \Pi_{N_{A,m},N_{B,m}}\right)}},$$

the second term under the square in

$$\left(\frac{\bar{H}_{\bullet\bullet}(s_e) - (\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e))}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})}} + \frac{\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e)}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})}}\right)^2$$

does not affect the limiting behavior of $2(L_{UR} - L_R(s_e))$. Consequently, to complete the arguments, The conditions of Assumption 1 and $c_e = 0$ can be combined in steps similar to those of Lemma C.1 to deduce that

$$\left(\frac{\bar{H}_{\bullet\bullet}(s_e) - (\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e)) + (\overline{F}_{B,N_{B,m}}(s_e) - \underline{F}_{A,N_{A,m}}(s_e))}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})}}\right)^2 \mid \Pi_{N_{A,m},N_{B,m}} \xrightarrow{d} \chi^2_1$$

$$\text{and} \quad \text{VAR}(\bar{H}_{\bullet\bullet}(s_e) \mid \Pi_{N_{A,m},N_{B,m}})/D(s_e) \mid \Pi_{N_{A,m},N_{B,m}} \xrightarrow{P} 1,$$

as $m \to +\infty$, hold. We omit these details for brevity. $\square$