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

June 2021

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

June 7, 2021

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, and School of Economics, University of Queensland, Australia

numerical scale of such variables prevents the use of traditional tools for distributional analysis.[1] 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 either perform 'dominance checks' by visually inspecting graphs of stochastic dominance curves (e.g., Jenkins, 2020), or test for stochastic dominance (e.g. Martinez-Marquina et al., 2019, González and Moral, 2019 and Aryal and Gabrielli, 2013) using a statistical procedure. 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

[1] To see why standard distributional analysis does not apply, consider the following simple numerical example involving the mean. Suppose that $x_1 = (2, 3, 2, 1)$ and $x_2 = (3, 0, 0, 5)$ are distributions of two populations consisting of 8 respondents each, whose responses are for the 4 ordinal categories {"not at all", "not very much", "a lot", "a great deal"}. Given the scale $c = [1, 2, 3, 4]$ the mean of x_1 is higher than the mean of x_2 $(2.3 > 1.8)$, and given the scale $c' = [1, 10, 11, 12]$ the ranking is reversed $(6.3 < 6.6)$.

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.

We also derive the asymptotic power of this test for local and non-local alternatives. Under reasonable assumptions, the test has asymptotic power equal to 1 against (i) distant local and (ii) non-local alternatives. Additionally, the test is asymptotically biased against alternatives that converge to the boundary of the null hypothesis at a sufficiently fast rate. This result is not surprising, as the test is not asymptotically similar on the boundary of the null hypothesis.

The idea of using bounds to account for data problems is not new, but gained popularity 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 Kaur et al. (1994). 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 procedures put forward by Alvarez-Esteban et al. (2017) also focus on a null of non-dominance, but for relaxed versions of stochastic dominance. However, their methods also rely on the data not suffering from missingness.

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[2] 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)[3] levels from the 2016 and 2018 waves of the *Arab Barometer*

[2] Baalbaki (2018).

[3] We use the words "trust" and "confidence" interchangeably.

as well as the 2013 *World Values Survey*. We use these datasets to explore how trust in Lebanese public institutions has been changing over time.[4] 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 asymptotic power properties of the test. Section 6 discusses the scope of our main results and implications for empirical practice. Section 7 presents the empirical analysis on dynamics of trust in Lebanese public institutions, and Section 8 concludes. All proofs are relegated to the Appendix.

2 Setup

Consider an ordinal variable measured on a scale characterized by the numerical labels $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$, where $-\infty < s_1 < s_2 < \cdots < s_K < +\infty$. Survey samples on the ordinal variable of interest are drawn from fixed finite populations according to the survey designs, which can have missing values due to non-response. For $Y \in \{A, B\}$, 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 finite population Y with total N_Y, where $X_{Y,i} \in \mathcal{S}$ and $D_{Y,i}$ 0/1 binary variables indicating an individual's response to the survey in the reference population for each i. In this setup,

$$\text{the practitioner observes} \begin{cases} X_{Y,i} & \text{if } D_{Y,i} = 1 \\ \text{missing value code} & \text{if } D_{Y,i} = 0, \end{cases}$$

when individual i is sampled from population $\mathcal{P}_{Y,N_Y}$, for $Y \in \{A, B\}$.

An objective of this paper is to compare the finite population distributions of the ordinal variable arising from $\mathcal{P}_{A,N_A}$ and $\mathcal{P}_{B,N_B}$. For each $Y \in \{A, B\}$, the cumulative distribution function (CDF) of the ordinal variable under $\mathcal{P}_{Y,N_Y}$ is defined as $F_{Y,N_Y}(s) = N_Y^{-1} \sum_{j=1}^{N_Y} 1[X_{Y,j} \leq s]$ for each $s \in \mathcal{S}$. We say that F_{B,N_B} *strictly* dominates F_{A,N_A}, stochastically, at first-order, if $F_{B,N_B}(s) < F_{A,N_A}(s)$ for each $s \in \mathcal{S}^\circ$, where $\mathcal{S}^\circ = \mathcal{S} - \{s_K\}$. Note that we have excluded the support point s_K as $F_{B,N_B}(s_K) = F_{A,N_A}(s_K) = 1$ because F_{B,N_B} and F_{A,N_A} are CDFs. We would like to infer that $F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$, holds, using the survey samples. For this reason, we follow Davidson and Duclos (2013) by positing the testing problem

$$H_0 : \max_{s \in \mathcal{S}^\circ} (F_{B,N_B}(s) - F_{A,N_A}(s)) \geq 0 \quad \text{Vs.} \quad H_1 : F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ. \tag{2.1}$$

The null hypothesis states that F_{B,N_B} does not strictly dominate F_{A,N_A}, stochastically, at first-order. The

[4] Earlier versions of the Arab Barometer do not include questions on trust in public institutions.

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 $F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$.

With this objective, the contrasts $F_{B,N_B}(s_i) - F_{A,N_A}(s_i)$ for $i = 1, \ldots, K-1$, are a natural estimand for developing a testing procedure for the testing problem (2.1). This estimand can be defined using the notion of estimating functions (Godambe and Thompson, 2009). Let $\theta_{N_A,N_B} = [\theta_{N_A,N_B}(1), \ldots, \theta_{N_A,N_B}(s_{K-1})]$ be vector of size $K-1$ such that $\theta_{N_A,N_B}(i) \in [-1, 1]$ for each i, which is defined by the following estimating functions

$$g(X_A, X_B, \theta_{N_A,N_B}(i)) = 1\,[X_B \leq s_i] - 1\,[X_A \leq s_i] - \theta_{N_A,N_B}(i), \quad \forall i = 1, \ldots, K-1. \tag{2.2}$$

Thus, if we observed the finite populations $\mathcal{P}_{A,N_A}$ and $\mathcal{P}_{B,N_B}$, then

$$\frac{1}{N_A}\frac{1}{N_B}\sum_{\ell=1}^{N_A}\sum_{j=1}^{N_B} g\left(X_{A,\ell}, X_{B,j}, \theta_{N_A,N_B}(i)\right) = 0, \quad \forall i = 1, \ldots, K-1, \tag{2.3}$$

are the so-called *census estimating equations*. These equations target this estimand because $\theta_{N_A,N_B}(i) = F_{B,N_B}(s_i) - F_{A,N_A}(s_i)$ for each i, is their solution. A design-based estimator of θ_{N_A,N_B} can be defined as the solution of the sample estimating equation

$$\sum_{\ell \in U_A}\sum_{j \in U_B} \frac{g\left(X_{A,\ell}, X_{B,j}, \theta_{N_A,N_B}(i)\right)}{\pi_{A,\ell}\,\pi_{B,j}} = 0, \quad \forall i = 1, \ldots, K-1,$$

where $U_Y \subset \{1, \ldots, N_Y\}$ is a survey sample from population $\mathcal{P}_{Y,N_Y}$ and $\pi_{Y,e} = \text{Prob}\,[e \in U_Y]$ for $Y \in \{A, B\}$, are the inclusion probabilities of the samples. The resulting estimator has the form

$$\hat{\theta}_{N_A,N_B}(i) = \hat{F}_{B,N_B}(s_i) - \hat{F}_{A,N_A}(s_i)\ \ \forall i = 1, \ldots, K-1, \quad \text{where}$$
$$\hat{F}_Y(s) = \frac{\sum_{j \in U_Y} \pi_{Y,j}^{-1}\,1\,[X_{Y,j} \leq s]}{\sum_{j \in U_Y} \pi_{Y,j}^{-1}} \quad \forall Y \in \{A, B\}$$

are the Hájek estimators of the CDFs F_{B,N_B} and F_{A,N_A}. In practice, the inclusion probabilities $\pi_{Y,e}$ are included in the surveys' datasets in the form of survey *design* weights $\{W_{Y,i}, i \in U_Y\}$, which do not include any missing values, and satisfy

$$\frac{W_{Y,i}}{n_Y} = \frac{\pi_{Y,i}^{-1}}{\sum_{i \in U_Y} \pi_{Y,i}^{-1}} \tag{2.4}$$

where $n_Y = \sum_{i \in U_Y}$. These weights enable consistent estimation of θ_{N_A,N_B} using $\hat{\theta}_{N_A,N_B}$ when $D_{A,\ell} = D_{B,j} = 1$ for all ℓ and j (i.e., when survey samples are always complete). Consequently, this estimator can be used in a testing procedure for the testing problem (2.1). However, as nonresponse is a feature of survey datasets in practice, which is not necessarily ignorable, this classical estimating equations approach to developing a testing procedure for the testing problem (2.1) is not feasible, because θ_{N_A,N_B} is partially identified.

For inference under survey nonresponse, it is productive to firstly consider the identified set of θ_{N_A,N_B}. The practitioner can only learn the distribution of the ordinal variable for individuals who respond to the surveys. These distributions are

$$F_{Y,N_Y,1}(s) = \frac{\sum_{j=1}^{N_Y} 1\left[X_{Y,j} \leq s,\ D_{Y,j} = 1\right]}{\sum_{j=1}^{N_Y} D_{Y,j}} \quad \forall s \in \mathcal{S} \text{ and } forall Y \in \{A, B\}. \tag{2.5}$$

Also, the practitioner can learn the fraction of response: $\sum_{j=1}^{N_Y} D_{Y,j}/N_Y$ for $Y = A, B$. Following Manski (1994), an application of the Law of Total Probability to the CDF F_{Y,N_Y} reveals that it can be decomposed as

$$F_{Y,N_Y}(s) = F_{Y,N_Y,1}(s)\left(N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}\right) + F_{Y,N_Y,0}(s)\left(1 - N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}\right), \quad \text{where} \tag{2.6}$$

$$F_{Y,N_Y,0}(s) = \frac{\sum_{j=1}^{N_Y} 1\left[X_{Y,j} \leq s,\ D_{Y,j} = 0\right]}{1 - N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}} \quad \forall s \in \mathcal{S}. \tag{2.7}$$

Manski (1994) also noted that $0 \leq F_{Y,N_Y,0}(s) \leq 1$ for each $s \in \mathcal{S}$, because it is a CDF. Substituting these inequalities into (2.6) yields the worst-case bounds: $\underline{F}_{Y,N_Y}(s) \leq F_{Y,N_Y}(s) \leq \overline{F}_{Y,N_Y}(s)$ for each $s \in \mathcal{S}$, where

$$\overline{F}_{Y,N_Y}(s) = F_{Y,N_Y,1}(s)\left(N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}\right) + \left(1 - N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}\right) \quad \text{and} \tag{2.8}$$

$$\underline{F}_{Y,N_Y}(s) = \begin{cases} F_{Y,N_Y,1}(s)\left(N_Y^{-1}\sum_{j=1}^{N_Y} D_{Y,j}\right), & s \in \mathcal{S}^\circ, \\ 1, & s = s_K. \end{cases} \tag{2.9}$$

These bounds exhaust all the information on F_{Y,N_Y} in the absence of any information on nonresponse from the practitioner's perspective. Consequently, the identified set of θ_{N_A,N_B} based on the populations $\mathcal{P}_{A,N_A}$ and $\mathcal{P}_{B,N_B}$ is given by

$$\Theta_I = \left\{\theta \in [-1,1]^{K-1} : \underline{F}_{B,N_B}(s_i) - \overline{F}_{A,N_A}(s_i) \leq \theta(i) \leq \overline{F}_{B,N_B}(s_i) - \underline{F}_{A,N_A}(s_i),\ \forall i\right\}.$$

This set is a singleton when either $D_{A,\ell} = D_{B,j} = 1$ for all ℓ and j, which is the complete data scenario, or when $F_{Y,N_Y,1}(s) = F_{Y,N_Y,0}(s)$ for each $s \in \mathcal{S}$ and $Y \in \{A, B\}$, which corresponds to nonresponse being ignorable. Without assuming anything about the nature of nonresponse, Θ_I is not necessarily a singleton, which renders θ_{N_A,N_B} as partially identified.

We adopt a bounds approach in developing a testing procedure for inferring $F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$. In particular, we consider the following testing problem:

$$H_0^1 : \max_{s \in \mathcal{S}^\circ}\left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right) \geq 0 \quad \text{Vs.} \quad H_1^1 : \overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s) < 0\ \forall s \in \mathcal{S}^\circ. \tag{2.10}$$

Rejecting H_0^1 in favor of H_1^1 in (2.10) implies rejection of H_0 in favor of H_1 in (2.1), since $F_{B,N_B}(s) \leq \overline{F}_{B,N_B}(s) < \underline{F}_{A,N_A}(s) \leq F_{A,N_A}(s)\ \forall s \in \mathcal{S}$. The natural estimand in the testing problem (2.10) are the contrasts $\overline{\theta}_{N_A,N_B}(i) = \overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s_i)$ for each i. It is defined by the estimating functions

$$g(\underline{X}_A, \overline{X}_B, \overline{\theta}_{N_A,N_B}(i)) = 1\left[\overline{X}_B \leq s_i\right] - 1\left[\underline{X}_A \leq s_i\right] - \overline{\theta}_{N_A,N_B}(i), \quad \forall i = 1, \ldots, K-1, \tag{2.11}$$

which are of the same form as the estimating functions in (2.2), but with X_A and X_B replaced by

$$\underline{X}_A = \begin{cases} X_A & \text{if } D_A = 1, \\ s_K & \text{if } D_A = 0, \end{cases} \quad \text{and} \quad \overline{X}_B = \begin{cases} X_B & \text{if } D_B = 1, \\ s_1 & \text{if } D_B = 0, \end{cases}$$

respectively. This adjustment of X_A and X_B essentially replaces the finite populations $\mathcal{P}_{A,N_A}$ and $\mathcal{P}_{B,N_B}$ with $\{\{\underline{X}_{A,i}, D_{A,i}\}, i = 1, 2, 3, \ldots, N_A\}$ and $\{\{\overline{X}_{B,i}, D_{B,i}\}, i = 1, 2, 3, \ldots, N_B\}$, respectively. The advantage of this adjustment is that the CDF bounds (2.8) and (2.9), have the following representation in terms of the populations $\{\{\underline{X}_{A,i}, D_{A,i}\}, i = 1, 2, 3, \ldots, N_A\}$ and $\{\{\overline{X}_{B,i}, D_{B,i}\}, i = 1, 2, 3, \ldots, N_B\}$: $\underline{F}_{A,N_A}(s) = N_A^{-1} \sum_{j=1}^{N_A} 1\left[\underline{X}_{A,i} \leq s\right]$ and $\overline{F}_{B,N_B}(s) = N_B^{-1} \sum_{j=1}^{N_B} 1\left[\overline{X}_{B,i} \leq s\right]$, so that $\overline{\theta}_{N_A,N_B}$ is the solution of the census estimating equations

$$\frac{1}{N_A}\frac{1}{N_B}\sum_{\ell=1}^{N_A}\sum_{j=1}^{N_B} g\left(\underline{X}_{A,\ell}, \overline{X}_{B,j}, \overline{\theta}_{N_A,N_B}(i)\right) = 0, \quad \forall i = 1, \ldots, K-1. \tag{2.12}$$

A design-based estimator of $\overline{\theta}_{N_A,N_B}$ can be defined as the solution of the sample version of this estimating equation

$$\frac{1}{n_A}\frac{1}{n_B}\sum_{\ell \in U_A}\sum_{j \in U_B} W_{A,\ell} W_{B,j} g\left(\underline{X}_{A,\ell}, \overline{X}_{B,j}, \overline{\theta}_{N_A,N_B}(i)\right) = 0, \quad \forall i = 1, \ldots, K-1,$$

where, as above, U_Y is a survey sample from population $\mathcal{P}_{Y,N_Y}$ and $\{W_{Y,i}, i \in U_Y\}$ are the survey design weights for $Y \in \{A, B\}$ that satisfy (2.4). While it is a common practice to use survey weights that are adjusted for selection/missing data in estimating a finite population characteristic, it should be noted that such weights (i) are functions of data, and hence, are random variables, and (ii) depend on a specification of the nonresponse-generating process, which can be misspecified. Consequently, statistical procedures that employ adjusted weights should appropriately account for the sampling variability due the adjustment, and the error due to mispecification of the nonresponse model. Otherwise, their use in empirical practice can result in biased inferences.

By contrast, the design weights are non-random, but are usually scaled as in (2.4). Additionally, they enable consistent estimation of $\overline{\theta}_{N_A,N_B}$. This combination of the worst-case bounds and the design weights enables the practitioner to account for selection of any kind in a testing procedure for the test problem (2.10). The next section describes this testing procedure.

3 Testing Procedure

We adapt the testing procedure of Davidson and Duclos (2013) for the test problem (2.10). 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: for each $Y \in \{A, B\}$, $\{X_{Y,i}, D_{Y,i}, W_{Y,i} : i \in U_Y\}$ where $U_Y \subset \{1, \ldots, N_Y\}$, and the $W_{Y,i}$ are the survey design weights which have no missing values. The surveys' design weights have been scaled as in (2.4), which implies that $\sum_{i\in U_Y} W_{Y,i} = \sum_{i\in U_Y} = n_Y$, holds for each $Y \in \{A, B\}$.

The first step is to adjust the survey samples to obtain samples from $\underline{F}_{A,N_A}$ and $\overline{F}_{B,N_B}$. Using the sample $\{X_{B,i}, D_{B,i}, i \in U_B\}$, replace each missing value in it with the smallest value in $\mathcal{S}$ (i.e., s_1). Similarly, modifying the sample $\{X_{A,i}, D_{A,i}, i \in U_A\}$ 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

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

Following Davidson and Duclos (2013), the test focuses on the boundary of the null hypothesis H_0^1 in (2.10). A pair of populations in the boundary of H_0^1 has at least one $s \in \mathcal{S}^\circ$ such that $\overline{F}_{B,N_B}(s) = \underline{F}_{A,N_A}(s)$. 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 $\overline{F}_{B,N_B}(s) = \underline{F}_{A,N_A}(s)$. 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

$$0 = \sum_{i\in U_B} \sum_{j\in U_A} p_j^A p_i^B H_{i,j}(s), \tag{3.1}$$

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

$$\begin{aligned} &\max_{p_i^Y:\, i\in U_Y,\, Y\in\{A,B\}} \sum_{Y\in\{A,B\}} \sum_{i\in U_Y} W_{Y,i} \log p_i^Y \quad \text{subject to} \\ &p_i^Y > 0,\ \sum_{i\in U_Y} p_i^Y W_{Y,i} = 1 \quad \forall i \in U_Y \text{ and } Y \in \{A, B\},\ \text{and } 0 = \sum_{i\in U_B} \sum_{j\in U_A} p_j^A p_i^B H_{i,j}(s). \end{aligned} \tag{3.2}$$

We denote by $L_R(s)$ the maximal value of this optimization problem. The unconstrained estimators of $\overline{F}_{B,N_B}$ and $\underline{F}_{A,N_A}$ are the Hájek CDF estimators

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

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 $L_{UR} = \sum_{Y \in \{A,B\}} \sum_{i \in U_Y} W_{Y,i} \log(1/n_Y)$ is the unconstrained maximum value 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 two sections develop the asymptotic properties of this testing procedure under null and alternative hypotheses for large finite populations.

4 Asymptotic Null Properties

This section develops the uniform asymptotic validity of the proposed testing procedure as the sizes of the finite populations diverge (i.e., as $N_A, N_B \to +\infty$). Establishing validity with uniformity is crucial for reliable inference in large finite populations, because the asymptotic 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). Because the survey sampling schemes can be complex (e.g., involving multiple stages and stratification), we present a detailed analysis of this asymptotic null property.

We start with a brief description of the probability space of the survey designs with finite populations. Recall that $\mathcal{P}_{Y,N_Y} = \{\{X_{Y,i}, D_{Y,i}\}, i = 1, 2, 3, \ldots, N_Y\}$ is the set of vectors of population Y with total N_Y, for $Y \in \{A, B\}$. 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 $Y \in \{A, B\}$. The sigma-algebra generated by $\mathcal{U}_{N_Y}$ is denoted as σ_{N_Y} for $Y \in \{A, B\}$. Additionally, let $\mathcal{M}_{Y,N_Y} = \{\mathcal{P}_{Y,N_Y} : X_{Y,i} \in \mathcal{S} \text{ and } D_{Y_i} \in \{0, 1\}\ \forall i = 1, \ldots, N_Y\}$, for $Y \in \{A, B\}$. 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 $Y \in \{A, B\}$. 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 $Y \in \{A, B\}$. 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 (2.4) for $Y \in \{A, B\}$.

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

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

so that the assertion Π_0 satisfies H_0^1 in (2.10) is equivalent to $\Pi_0 \in \mathcal{M}^0_{N_A,N_B}$. The size of the test for H_0^1 in (2.10) 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 $E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi\right]$ 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 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 $Y \in \{A, B\}$.

(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) $\dfrac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s) \right)}{\sqrt{\mathrm{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B} \right)}} \mid \Pi_{N_A,N_B} \xrightarrow{d} N(0, 1)$ 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

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

(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

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

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, and Zhao et al., 2020). 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. 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 is $\{LR^{(A,B)}/Deff(s) \mid \Pi_{N_A,N_B}\}^{+\infty}_{N_A,N_B=1}$ 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.

5 Asymptotic Power Properties

This section develops the asymptotic power properties of the proposed testing procedure. In particular, we develop results on the test's consistency and its behavior under alternatives that converge to the boundary of H^1_0. The model of the *alternative* hypothesis is given by

$$\mathcal{M}^1_{N_A,N_B} = \left\{\Pi \in \mathcal{M}_{N_A,N_B} : \max_{s\in\mathcal{S}^\circ} \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right) < 0\right\},$$

so that the assertion Π_0 satisfies H^1_1 in (2.10) is equivalent to $\Pi_0 \in \mathcal{M}^1_{N_A,N_B}$. Thus, the power of the test is defined as $E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_0\right]$, where $\Pi_0 \in \mathcal{M}^1_{N_A,N_B}$. We shall describe the properties of the test when the populations' totals tend to infinity (i.e., as $N_A, N_B \to +\infty$). In that direction, we embed $\mathcal{M}^1_{N_A,N_B}$ into a hypothetical sequence $\{\mathcal{M}^1_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ and consider sequences of finite populations

$\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ such that $\Pi_{N_A,N_B} \in \mathcal{M}_{N_A,N_B}^1$ for each N_A and N_B, which satisfy certain restrictions described below. In this framework, asymptotic power of the test along the sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ is given by the limit

$$\lim_{N_A,N_B\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right]. \tag{5.1}$$

Because we are employing a fixed critical value, the asymptotic behavior of $\{LR^{(A,B)} \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ steers the limit (5.1).

In addition to Assumption 1, we impose the following conditions on $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, which are helpful for deriving the asymptotic behavior of $\{LR^{(A,B)} \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ under the alternative.

Assumption 2. Let $K_{N_A,N_B} = \frac{E[n_B|\mathcal{P}_{B,N_B}]\,E[n_A|\mathcal{P}_{A,N_A}]}{E[n_B|\mathcal{P}_{B,N_B}]+E[n_A|\mathcal{P}_{A,N_A}]}$, for each N_A and N_B.

(i) $\frac{K_{N_A,N_B}}{n_B} - \frac{K_{N_A,N_B}}{E[n_B|\mathcal{P}_{B,N_B}]} \mid \mathcal{P}_{B,N_B} \xrightarrow{P} 0$ and $\frac{K_{N_A,N_B}}{n_A} - \frac{K_{N_A,N_B}}{E[n_A|\mathcal{P}_{A,N_A}]} \mid \mathcal{P}_{A,N_A} \xrightarrow{P} 0$, as $N_A, N_B \to +\infty$.

(ii) For each $s \in \mathcal{S}^\circ$, $\lim_{N_A,N_B\to+\infty} \text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right) = 0$ and $\lim_{N_A,N_B\to+\infty} K_{N_A,N_B}\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$ exists and is positive.

The conditions of Assumption 2 are commonly used restrictions on the sampling designs of the surveys (see, e.g., Zhao et al., 2020). Because we have two populations of potentially different sizes, we need to introduce a quantity that stands in for a 'common' population size that can be used to describe the rate of decrease of the design-variance $\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$ as $N_A, N_B \to +\infty$. Note that the design-variance decreasing with the population sizes is a common structure of survey designs (see, e.g., Fuller, 2009) and Assumption 2 formalises this structure within our framework. Condition (i) of Assumption 2 imposes a condition on the way in which the sample sizes grow relative to their population sizes. It is a strengthening of Part (i) of Assumption 1, as that assumption only implies the divergence of the sample sizes n_A and n_B. Condition (ii) of Assumption 2 imposes a rate of decrease on the design-variance $\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$ as $N_A, N_B \to +\infty$ with respect to the 'common' population size K_{N_A,N_B}.

We assess the asymptotic power properties of the test against a given sequence of finite populations in the alternative hypothesis by deriving the value of the limit in (5.1) for that sequence. This entails characterising the asymptotic behavior of $\{LR^{(A,B)} \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ in terms of the asymptotic behavior of $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, for each $s \in \mathcal{S}^\circ$, under different conditions on

$$\tau(s) = \lim_{N_A,N_B\to+\infty} \frac{\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}} \mid \Pi_{N_A,N_B}. \tag{5.2}$$

Under Assumption 2, sequences in the alternative hypothesis satisfy $\tau(s) \in \{0, -\infty\}$ for each $s \in \mathcal{S}^\circ$, because (i) $(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s))$ is uniformly bounded for all N_A and N_B, and for all $s \in \mathcal{S}^\circ$, and

(ii) VAR $\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$ is decreasing to zero at the rate $[K_{N_A,N_B}]^{-1/2}$, for each $s \in \mathcal{S}^\circ$. The technical results describing the asymptotic behavior of $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ in these different scenarios are given by Lemmas D.1 and D.2, which are relegated to the Appendix D for ease of exposition.

The next result presents the asymptotic power properties of the test.

Theorem 3. Let $\{\mathcal{M}^1_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ be as described above, $\alpha \in (0,1)$, and τ be given by (5.2). Suppose that the sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ is such that $\Pi_{N_A,N_B} \in \mathcal{M}^1_{N_A,N_B}$ for each N_A and N_B, and satisfies Assumptions 1 and 2. The following statements hold.

1. If $\tau(s) = -\infty$ for each $s \in \mathcal{S}^\circ$, then $\lim_{N_A,N_B \to +\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] = 1$.

2. Suppose $\exists s \in \mathcal{S}^\circ$ such that $\tau(s) = 0$, and let $S_0 = \{s \in \mathcal{S}^\circ : \tau(s) = 0\}$. Then

$$\lim_{N_A,N_B \to +\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] = \text{Prob}\left[Z_s^2 > c(\alpha),\ Z_s < 0,\ s \in S_0\right], \tag{5.3}$$

where $\{Z_s,\ s \in S_0\} \overset{d}{\sim} MVN(\mathbf{0}, \Xi)$ and Ξ is the limiting covariance matrix of the terms

$$\frac{\hat{\overline{F}}_{B,N_B}(s) - \hat{\underline{F}}_{A,N_A}(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}} \mid \Pi_{N_{A,N_A},N_{B,N_B}},\ s \in S_0,\ \text{as } N_A, N_B \to +\infty.$$

Proof. See Appendix B.3. ■

At this point, we describe the kinds of sequences of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ covered by Theorem 3. Regarding Part 1 of Theorem 3, the scenarios in which $\tau(s) = -\infty$ for each $s \in \mathcal{S}^\circ$, there are essentially two classes of alternatives that are demarcated by our assumptions. The first class of examples of such a sequence of populations are those that satisfy

$$\limsup_{N_A,N_B \to +\infty} \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right) \mid \Pi_{N_A,N_B} < 0,\ \forall s \in \mathcal{S}^\circ.$$

In this case, the decreasing design-variance (i.e., Condition (ii) of Assumption 2) is driving the limit (5.1), for each $s \in \mathcal{S}^\circ$. The kinds of sequences of populations captured by this class are those in the alternative whose subsequences *do not* tend to the boundary of the null hypothesis, as $N_A, N_B \to +\infty$. A second class of examples are those sequences that have a subsequence $\left\{\Pi_{N_{A,m},N_{B,m}}\right\}_{m=1}^{+\infty}$ where for at least one $s \in \mathcal{S}^\circ$, $(\overline{F}_{B,N_{B,m}}(s) - \underline{F}_{A,N_{A,m}}(s)) \mid \Pi_{N_{A,m},N_{B,m}} \uparrow 0$ as $m \to +\infty$, and

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

These are sequences of finite populations in the alternative that have an accumulation point on the boundary of the null hypothesis, with subsequences converging to it at a slower rate than the rate $\left[K_{N_{A,m},N_{B,m}}\right]^{-1/2}$.

Regarding Part 2 of Theorem 3, which covers scenarios where $\exists s \in \mathcal{S}^\circ$ such that $\tau(s) = 0$, the sequences $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ converge to the boundary of H_0^1. This convergence also occurs at a rate faster than $[K_{N_A,N_B}]^{-1/2}$, which is why $\tau(s) = 0$. If $S_0 = \{s\}$, then the limiting local power function satisfies

$$\lim_{N_A,N_B\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] = \text{Prob}\left[Z_s^2 > c(\alpha),\ Z_s < 0\right] \leq \text{Prob}\left[Z_s^2 > c(\alpha)\right] = \alpha,$$

as $Z_s^2 \overset{d}{\sim} \chi_1^2$. If $|S_0| > 1$, then we can also conclude

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

because Prob $\left[Z_s^2 > c(\alpha),\ Z_s < 0,\ s \in S_0\right] \leq$ Prob $\left[Z_{s\prime}^2 > c(\alpha)\right] = \alpha$, for $s' \in S_0$. These derivations show the test may be asymptotically biased against such alternatives. This finding is not surprising, as the test is *not* asymptotically similar on the boundary of H_0^1. We also note that if the CDFs are close for more than one s, then this may exacerbate the asymptotic bias.

6 Discussion

This section presents a discussion of the scope of our main results and implications for empirical practice. The statistical procedure of this paper is designed to infer $F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$, on the basis of two independent survey samples and under survey nonresponse of any kind. This robustness with respect to the nature of the nonresponse-generating process is achieved by using an inferential approach that ranks the worst-case bounds of these CDFs with a rejection event that implies $F_{B,N_B}(s) < F_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$; i.e., the testing problem (2.10). However, this robustness comes at the expense of an ambiguity arising under the null hypothesis H_0^1 in (2.10). Specifically, under the null, there exists an $s^* \in \mathcal{S}^\circ$ such that $\overline{F}_{B,N_B}(s^*) \geq \underline{F}_{A,N_A}(s^*)$, and this condition does not preclude that $F_{B,N_B}(s) < F_{A,N_A}(s)$, holds, at each $s \in \mathcal{S}^\circ$. The uninformative nature of H_0^1 is a consequence of circumventing assumptions about the nature of the nonresponse-generating process, which are unverifiable in practice. If one fails to reject this null hypothesis in practice, then, unfortunately, one cannot conclude anything informative about the ranking of the CDFs, F_{B,N_B} and F_{A,N_A}, using their worst-case bounds. In such a situation, we recommend empirical researchers perform a sensitivity analysis of this empirical conclusion (i.e., non-rejection of H_0^1) with respect to plausible assumptions on the nonresponse-generating process.

In line with the partial identification approach in econometrics (e.g., Tamer, 2010), a sensitivity analysis entails executing the testing procedure under different assumptions on the nonresponse-generating process, in order to study the impact of those assumptions on inference. The virtue of this type of analysis is that it establishes, in a transparent way, clear links between empirical outcomes and different assumptions made on the nonresponse-generating process. Such an analysis would reveal non-trivial links between assumptions on the nonresponse-generating process and inferences made, when in the sample, the bounds $\underline{F}_{A,N_A}$ and $\overline{F}_{B,N_B}$ are not so far apart and the estimates of the nonresponse fractions, $\left(1 - \sum_{j=1}^{N_Y} D_{Y,j}/N_Y\right)$ for $Y \in \{A, B\}$, are not too small. For concreteness, we discuss two examples and illustrate their usage in the empirical application of Section 7. It should be noted, though, that there is a vast middle ground between the setup that makes no assumption on nonresponse, which is the framework of our paper, and the practice of positing point-identifying assumptions on the CDFs F_{B,N_B} and F_{A,N_A}.

Example 1. For population $Y \in \{A, B\}$, dominant selection posits a weak first-order stochastic dominance relation between the CDFs $F_{Y,N_Y,1}$ and $F_{Y,N_Y,0}$, which are defined in (2.5) and (2.7), respectively. That is, $F_{Y,N_Y,1}$ and $F_{Y,N_Y,0}$ are the CDFs of the ordinal variable for the subpopulations of $\mathcal{P}_{Y,N_Y}$ corresponding to responders and nonresponders to the survey, respectively. Consider, for instance,

$$F_{Y,N_Y,1}(s) \leq F_{Y,N_Y,0}(s) \quad \forall s \in \mathcal{S}^{\circ}. \tag{6.1}$$

In words, these inequalities state for population Y that responders of the survey are more likely to report larger values than nonresponders on the scale $\mathcal{S}$. This assumption may be plausible in practice, and we consider it in the empirical application on trust levels in Lebanese public institutions. Substituting the inequalities (6.1) into the representation of F_{Y,N_Y} described in (2.6) for $Y = A$, refines the worst-case lower bound $\underline{F}_{A,N_A}$ to $F_{A,N_A,1}$. Consequently, one would consider the following testing problem:

$$H_0^2 : \max_{s \in \mathcal{S}^{\circ}} \left(\overline{F}_{B,N_B}(s) - F_{A,N_A,1}(s) \right) \geq 0 \quad \text{Vs.} \quad H_1^2 : \overline{F}_{B,N_B}(s) - F_{A,N_A,1}(s) < 0 \; \forall s \in \mathcal{S}^{\circ}. \tag{6.2}$$

Rejecting H_0^2 in favor of H_1^2 in (6.2) implies rejection of H_0 in favor of H_1 in (2.1), since $F_{B,N_B}(s) \leq \overline{F}_{B,N_B}(s) < F_{A,N_A,1}(s) \leq F_{A,N_A}(s) \; \forall s \in \mathcal{S}$. The natural estimand in this testing problem is the contrast $\overline{F}_{B,N_B}(s) - F_{A,N_A,1}(s) \; \forall s \in \mathcal{S}^{\circ}$. It is defined as the $(K-1)$-dimensional vector γ_{N_A,N_B} through the estimating functions

$$h\left(\overline{X}_B, X_A, \beta_{N_A,N_B}(i); \delta\right) = 1\left[\overline{X}_B \leq s_i\right] - \delta\, 1\left[X_A \leq s_i, D_A = 1\right] - \gamma_{N_A,N_B}(i), \; \forall i = 1, \ldots, K-1,$$

where $\delta = \left[N_A^{-1} \sum_{j=1}^{N_A} D_{A,j}\right]^{-1}$ is a nuisance parameter. Thus, we can use the census estimating equation corresponding to these estimating functions to find an estimator of γ_{N_A,N_B}.

Example 2. A practitioner may find it credible to consider structure on the nonresponse propensities conditional on the outcome, as in Section 4.2 of Manski (2016). In the framework of this paper, for $Y \in \{A, B\}$, this assumption models

$$R_Y(s) = \frac{N_Y^{-1} \sum_{i=1}^{N_Y} 1\left[D_{Y,i} = 0, X_{Y,i} \leq s\right]}{N_Y^{-1} \sum_{j=1}^{N_Y} 1\left[X_{Y,i} \leq s\right]}$$

as lying in a neighborhood of the nonresponse fraction $\left(1 - \sum_{j=1}^{N_Y} D_{Y,j}/N_Y\right)$, for each $s \in \mathcal{S}^{\circ}$. Specifically, one can consider numbers α_s and β_s, for $s \in \mathcal{S}^{\circ}$, such that $0 \leq \alpha_s \leq 1 \leq \beta_s$, $\alpha_{s_K} = \beta_{s_K} = 1$, and

$$\alpha_s \left(1 - \sum_{j=1}^{N_Y} D_{Y,j}/N_Y\right) \leq R_Y(s) \leq \beta_s \left(1 - \sum_{j=1}^{N_Y} D_{Y,j}/N_Y\right), \quad \forall s \in \mathcal{S}^{\circ}. \tag{6.3}$$

Note that these conditions cover the assumption of ignorable nonresponse as the limiting case where $\alpha_s = \beta_s = 1$ for each $s \in \mathcal{S}^\circ$. Furthermore, they can capture forms of monotone selection by selecting β_s and α_s to be decreasing in s. For brevity, we only present the bounds on F_{Y,N_Y} under this assumption and relegate their derivation to Appendix E. For each $s \in \mathcal{S}^\circ$, $\overline{F}^M_{Y,N_Y}(s) \leq F_{Y,N_Y}(s) \leq \underline{F}^M_{Y,N_Y}(s)$ where

$$\overline{F}^M_{Y,N_Y}(s) = \overline{\kappa}(s)\, F_{Y,N_Y,1}(s) \quad \text{and} \quad \underline{F}^M_{Y,N_Y}(s) = \underline{\kappa}(s)\, F_{Y,N_Y,1}(s), \tag{6.4}$$

with

$$\overline{\kappa}(s) = \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y} + \left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right) \left(\frac{\beta_s \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \beta_s \left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right) \quad \text{and} \tag{6.5}$$

$$\underline{\kappa}(s) = \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y} + \left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right) \left(\frac{\alpha_s \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \alpha_s \left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right). \tag{6.6}$$

A practitioner may find it plausible to posit such an assumption and perform a sensitivity analysis with respect to different values of $\{\alpha_s, \beta_s : s \in \mathcal{S}^\circ\}$ in combination with the bounds (6.4). For example, for a given set of values for those parameters, she may find this assumption plausible for population B, leading to a refinement of the worst-case bound $\overline{F}_{B,N_B}$, given by $\overline{F}^M_{B,N_B}$. Consequently, she would consider the following testing problem:

$$H_0^3 : \max_{s \in \mathcal{S}^\circ} \left(\overline{F}^M_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right) \geq 0 \quad \text{Vs.} \quad H_1^3 : \overline{F}^M_{B,N_B}(s) - \underline{F}_{A,N_A}(s) < 0 \; \forall s \in \mathcal{S}^\circ. \tag{6.7}$$

Rejecting H_0^3 in favor of H_1^3 in (6.7) implies rejection of H_0 in favor of H_1 in (2.1), since $F_{B,N_B}(s) \leq \overline{F}^M_{B,N_B}(s) < \underline{F}_{A,N_A}(s) \leq F_{A,N_A}(s)\ \forall s \in \mathcal{S}$. The natural estimand in this testing problem is the contrast $\overline{F}^M_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\ \forall s \in \mathcal{S}^\circ$. It is defined as the $(K-1)$-dimensional vector ψ_{N_A,N_B} through the estimating functions

$$w\left(X_B, \underline{X}_A, \psi_{N_A,N_B}(i); \upsilon\, \overline{\kappa}(s_i)\right) = \upsilon\, \overline{\kappa}(s_i)\, 1\left[X_B \leq s_i, D_B = 1\right] - 1\left[\underline{X}_A \leq s_i\right] - \psi_{N_A,N_B}(i), \tag{6.8}$$

for each $i = 1, \ldots, K-1$, where $\overline{\kappa}$ is described in (6.5) and $\upsilon = \left[N_B^{-1} \sum_{j=1}^{N_B} D_{B,j}\right]^{-1}$. The nuisance parameter in this case is the vector $[\upsilon\, \overline{\kappa}(s_1), \ldots, \upsilon\, \overline{\kappa}(s_{K-1})]$. Thus, we can use the census estimating equation corresponding to these estimating functions to find an estimator of ψ_{N_A,N_B}.

As might be expected, our empirical likelihood testing procedure can be extended to the setup covering the above examples by using "plug-in" estimators of the nuisance parameter. Developing a general theory for such extensions requires appropriate modifications to Assumptions 1 and 2 to account for (i) using different bounds than the worst-case bounds, and (ii) the nuisance parameters. Developing this general theory goes beyond the intended scope of this paper and are left for subsequent research.

This paper is part of the literature on inference for parameters defined by moment inequalities. It is a vast literature, and the majority of those tests apply to testing problems in which the null and alternative hypotheses are the reverse of the ones in our testing problem (2.1). Specifically, those tests apply to the following testing problem:

$$H_0^5 : \max_{s \in \mathcal{S}^\circ} \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s) \right) \leq 0 \quad \text{Vs.} \quad H_1^5 : \exists s \in \mathcal{S}^\circ \text{ such that } \overline{F}_{B,N_B}(s) > \underline{F}_{A,N_A}(s), \quad (6.9)$$

where H_0^5 in (6.9) states $\overline{F}_B$ dominates $\underline{F}_A$, stochastically and weakly, at the first-order, and H_1^5 is the negation of H_0^5 (i.e., non-dominance). Examples of tests that apply to the testing problem (6.9) include Andrews and Soares (2010) and Canay (2010) among many other tests. The testing problems, (6.9) and (2.1), are complementary in that positing a null of dominance cannot be used to infer dominance but can serve to infer non-dominance, while positing a null of non-dominance cannot serve to infer non-dominance but can lead to inferring dominance (Davidson and Duclos, 2013, p. 87). The reason is that failure to reject the null hypothesis, in general, cannot be taken as strong confirmation that the null hypothesis holds, unless test power is sufficiently high. Therefore, how one sets up the null and alternative hypotheses in their testing problem depends on the type of inference one wishes to make and the error one wishes to guard against. In this paper, we wish to infer strict dominance; consequently, we posit a null of non-dominance.

There aren't many testing procedures in the moment inequalities literature that apply to the testing problem (2.1). We only know of one such test other than the one put forward by Davidson and Duclos (2013). It is the test proposed by Kaur et al. (1994), which is based on the minimum t-statistic. Their test applies to second-order stochastic dominance with complete data, and Davidson and Duclos (2013) have shown how to adapt their test to the case of testing for first-order stochastic dominance. Furthermore, Davidson and Duclos establish the asymptotic local equivalence of their empirical likelihood statistic with the squared t-statistics under the null, at each point in the interior of the union of the populations' supports (Davidson and Duclos, 2013, Theorem 1 and its corollary) where the two CDFs are equal. Lemmas C.1 and C.2 in Appendix C establish an analogous result under the null hypothesis H_0^1 in (2.10), as $2(L_{UR} - L_R(s))$ ends up being asymptotically equivalent to the square of a t-statistic, given by

$$\left[\frac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s) \right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B} \right)}} \right]^2,$$

at any point $s \in \mathcal{S}^\circ$ such that $\overline{F}_{B,N_B}(s) = \underline{F}_{A,N_A}(s)$. A similar result holds under local alternatives, given by Lemma D.2 in Appendix D, which is used to develop Part 2 of Theorem 3. These technical results imply that our empirical likelihood test and the test based on the minimum of the squared t-statistics are asymptotically locally equivalent.

The next section presents an empirical application of the methodology this paper develops. The theory 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.

Table 1: Estimates of Nonresponse Fractions (i.e., $1 - N_Y^{-1} \sum_{j=1}^{N_Y} D_{Y,j}$)

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

7 Empirical Application

This section applies the methodology of this paper to the comparison of trust levels in Lebanese public institutions. It is organised as follows. Section 7.1 describes the data, and reports the identified sets of the trust distributions for the institutions we consider. Section 7.2 applies the proposed methodology using this data, and Section 7.3 illustrates the practicality of a sensitivity analysis using the nonresponse mechanisms described in Examples 1 and 2 of Section 6.

7.1 Description of the Data

We would like to use waves 3 - 5 from the *Arab Barometer* survey to compare the patterns in the average confidence in Lebanese public institutions in the run up to the October 2019 mass protests. These waves correspond to the periods 2013, 2016, and 2018. Unfortunately, data on the trust variables we consider are not available in wave 3, but are included in waves 4 and 5. For this reason, we also use data from the *World Values Survey* for 2013, as the question on trust for the Lebanese institutions we have considered are identical in the two surveys.

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 translated to "thiqa" in the Arabic questionnaire of both surveys, and we thus 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 $\mathcal{S} = \{1, 2, 3, 4\}$.

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 estimates of the nonresponse fractions. This estimate for trust in the "Government" in the 2013 survey is quite large (18%), and the rest are all less than 10%. Consequently, without any additional information on nonresponse, the finite population CDFs of the trust variables are not point-identified. While it is tempting to consider particular assumptions/models to explain these nonresponse estimates, they are unverifiable in practice and may yield biased inferences. Instead, we examine the information about those CDFs without imposing such

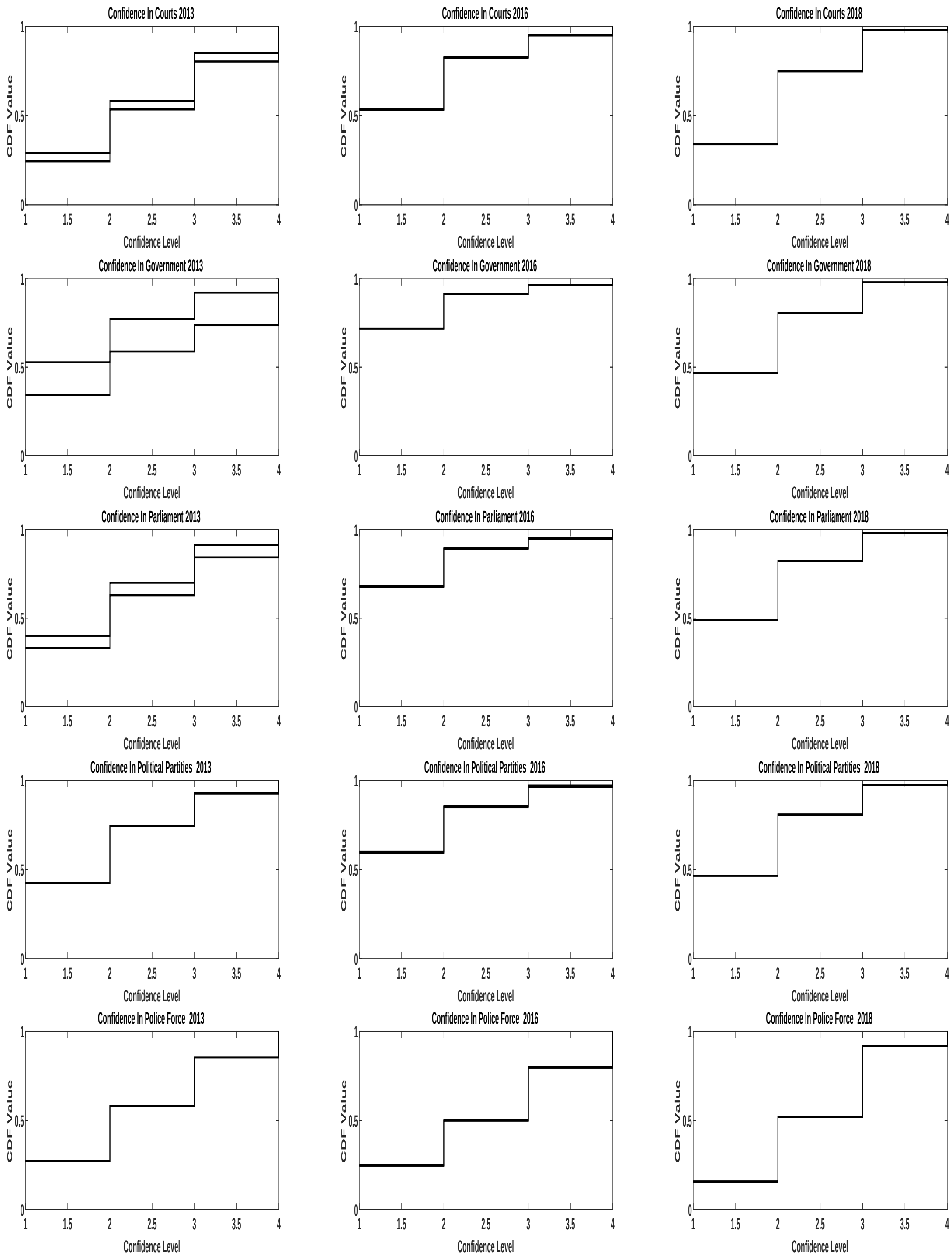

Figure 1: Estimated Identified Sets of Trust Distributions

assumptions on nonresponse. We represent this information graphically in Figure 1, which reports estimates of the identified sets of those CDFs based on the worst-case bounds described in (2.8) and (2.9).

The informativeness of the identified sets in Figure 1 is driven by the corresponding sizes of the nonresponse fraction estimates reported in Table 1. From the decomposition (2.6), the lower the estimates of the nonresponse fractions, the tighter are the bounds in Figure 1. Furthermore, the higher the estimates of the nonresponse fractions, the wider are the bounds in the figure. Because the estimates of the nonresponse fractions are quite small for all institutions in 2016 and 2018 (i.e., $< 1\%$), their corresponding bounds in Figure 1 are quite tight.[5] By contrast, the 2013 estimates of the nonresponse fractions are orders of magnitude higher than their 2016 and 2018 counterparts, which yield wider, and hence, less informative bounds. Figure 1 reveals that except for "Police Force", the estimates of the identified sets reveal that there is much distrust of all of the Lebanese institutions during the period 2013-2018, as the probability mass at $s_1 = 1$ based on the lower worst-case bound is either in the neighborhood of 50% or larger for those institutions. This empirical finding on "Police Force" is not unusual as the Internal Security Forces, which is the title of the police force in Lebanon, has been involved in trust building programs with local communities during this period (Slachmuijlder, 2015).

7.2 Analysis

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:

$$F_{13}(s) < F_{16}(s) < F_{18}(s) \text{ for } s = 1, 2, 3. \tag{7.1}$$

Using the worst-case bounds in inference, as we propose, permits the entire spectrum of models of nonresponse. This approach is especially useful when the incidence of nonresponse 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

$$\overline{F}_{13}(s) < \underline{F}_{16}(s) \quad \text{and} \quad \overline{F}_{16}(s) < \underline{F}_{18}(s) \quad \text{for } s = 1, 2, 3, \tag{7.2}$$

then the chain of inequalities (7.1) 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 (7.2)

[5] These bounds are so close that they appear to coincide. We would have to radically increase the resolution on their y-axes to visualise the bounds as two CDFs.

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

null hypothesis	Courts	Government	Parliament	Political Party	Police Force
H_0^6	0 Do not Reject	0 Do not Reject	0 Do not Reject	0 Do not Reject	0 Do not Reject
$H_0^{6,1}$	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

$$H_0^6 : \max_{s\in\mathcal{S}^\circ}\left(\overline{F}_{13}(s) - \underline{F}_{16}(s)\right) \geq 0 \text{ or } \max_{s\in\mathcal{S}^\circ}\left(\overline{F}_{16}(s) - \underline{F}_{18}(s)\right) \geq 0,$$

By treating H_0^6 as the union of two sub-hypotheses $H_0^{6,1} : \max_{s\in\mathcal{S}^\circ}\left(\overline{F}_{13}(s) - \underline{F}_{16}(s)\right) \geq 0$ and $H_0^{6,2} : \max_{s\in\mathcal{S}^\circ}\left(\overline{F}_{16}(s) - \underline{F}_{18}(s)\right) \geq 0$, 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 $H_0^{6,i}$ versus $H_1^{6,i}$: not $H_0^{6,i}$, for $i = 1, 2$. The decision rule is to reject H_0^6 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.[6]

The first row of Table 2 reports the realized values of LR and the conclusions of the hypothesis tests of H_0^6. Unfortunately, there is no evidence at the 5% level to reject H_0^6 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 (7.2). This finding is mainly driven by the comparisons between $\overline{F}_{16}$ and $\underline{F}_{18}$. For example, the right panel of Figure 2 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 $H_0^{6,2}$ 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 2 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 $H_0^{6,1}$ for each institution. For confidence in courts, government, parliament and political parties, we reject the null hypothesis $H_0^{6,1}$ 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, $H_0^{6,1}$ is also rejected at the 1% level.[7] However, $H_0^{6,1}$ 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

[6]No multiplicity adjustment is required to asymptotically control the level of this intersection-union testing procedure, as it is invariant to a Bonferroni correction.

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

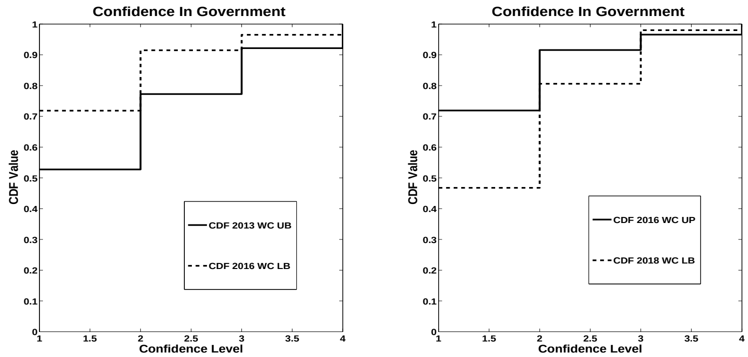

Figure 2: Trust in the government for the periods 2013-2016 and 2016-2018

courts, government, parliament and political parties have dropped between 2013 and 2016, as the dominance 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.

7.3 Sensitivity Analysis Illustration

Now we illustrate the force of a sensitivity analysis in this application using assumptions on nonresponse described in Examples 1 and 2 of Section 6. We focus on trust in "Government" where one is interested in inferring that $F_{13}(s) < F_{18}(s)$, holds, for each $s \in \{1, 2, 3\}$, because there is a high incidence of nonresponse with Government in 2013 (≈18%, see Table 1), which is in contrast to the years 2016 and 2018 where the nonresponse rates are substantially smaller than in 2013.

We begin by considering the situation where we are agnostic about the nature of nonresponse, leading us to the testing problem

$$H_0^7 : \max_{s \in \{1,2,3\}} \left(\overline{F}_{13}(s) - \underline{F}_{18}(s)\right) \geq 0 \quad \text{Vs.} \quad H_1^7 : \overline{F}_{13}(s) < \underline{F}_{18}(s) \; \forall s \in \{1, 2, 3\}. \tag{7.3}$$

The left panel in Figure 3 reports the Hájek estimates of $\overline{F}_{13}$ and $\overline{F}_{18}$, showing that

$$\max_{s \in \{1,2,3\}} \left(\hat{\overline{F}}_{13}(s) - \hat{\overline{F}}_{18}(s)\right) \geq 0,$$

holds. Since $\hat{\underline{F}}_{18}(s) \leq \hat{\overline{F}}_{18}(s)$ holds for each $s \in \mathcal{S}$, it follows that

$$\max_{s \in \{1,2,3\}} \left(\hat{\overline{F}}_{13}(s) - \hat{\underline{F}}_{18}(s)\right) \geq \max_{s \in \{1,2,3\}} \left(\hat{\overline{F}}_{13}(s) - \hat{\overline{F}}_{18}(s)\right) \geq 0, \tag{7.4}$$

holds. We thus conclude that H_0^7 holds in the sample, and hence, fail to reject this null hypothesis. Consequently, we face an ambiguous situation where it is still possible that $F_{13}(s) < F_{18}(s)$, holds, for each

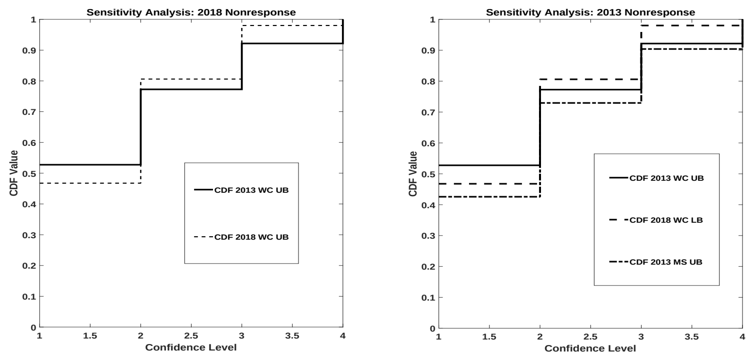

Figure 3: Trust in the Lebanese Government: period 2013-2018.

$s \in \{1, 2, 3\}$, under H_0^7. The inequalities (7.4) also indicate that a sensitivity analysis of this empirical conclusion with respect to assumptions on nonresponse for the 2018 population would not yield different conclusions. This includes, for instance, dominant selection discussed in Example 1, where it is plausible that in 2018, nonresponders are more likely to have lower levels of confidence/trust in the Lebanese Government in comparison to responders.

Next, we consider sensitivity with respect to assumptions on nonresponse for the 2013 population, based on the assumption in Example 2 of Section 6, using a specific value of $(\beta_1, \beta_2, \beta_3, \beta_4)$, because $\overline{F}^M_{B,N_B}$ depends only on those quantities. In practice, one can consider different values of that vector which are also plausible in performing a sensitivity analysis, and for illustration we consider only one value of this vector. Suppose that the value $(\beta_1, \beta_2, \beta_3, \beta_4) = (1.05, 1.05, 1, 1)$ is plausible in practice. Based on the discussion in Example 2, this view results in the use of the CDF $\overline{F}^M_{13}$, defined in (6.4) but with $B = 13$, instead of $\overline{F}_{13}$. Consequently, we consider the testing problem (6.7), with $A = 18$. The right panel in Figure 3 reports the Hájek estimates of $\overline{F}_{13}$, $\underline{F}_{18}$, and $\overline{F}^M_{13}$ showing that dominance holds in the sample, i.e., $\max_{s\in\{1,2,3\}} \left(\hat{\overline{F}}^M_{13}(s) - \hat{\underline{F}}_{18}(s) \right) < 0$, so that we need to implement the pseudo-empirical likelihood test. Appendix E describes how to implement the test for this testing problem. It must be modified on account of the presence of a nuisance parameter in the estimating function. The realised value of the test statistic, described in (E.6) of the Appendix, is $LR^{(13,18)} = 4.8906$, which exceeds the 5% χ^2_1 critical value $c(0.05) = 3.84$. Hence, we reject H_0^3 in (6.7) at the 5% significance level in favour of H_1^3, implying that $F_{13}(s) < F_{18}(s)$, holds, for each $s \in \{1, 2, 3\}$ under this particular form of monotone sample selectivity. It should be noted that we have not provided a proof of the asymptotic validity of this adjustment of our testing procedure. However, we expect that it is true under appropriate modifications of the conditions of Assumption 1. Developing the asymptotic validity of this modified testing procedure goes beyond the intended scope of the paper and is left for a future project.

8 Conclusion

This paper extends the testing procedure of 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.

There are a number of directions of future research. First, it would be interesting to extend our framework to incorporate bounds on the ordinal variables' distributions arising from plausible assumptions on the nonresponse-generating process. This extension would permit the implementation of a sensitivity analysis of empirical conclusions with respect to such assumptions, as we have illustrated in our paper. The main challenge with this extension is to develop the asymptotic theory for the pseudo-empirical likelihood-ratio testing procedure when the estimating functions under consideration include nuisance parameters. Second, it would be interesting to also 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). Third, while our focus has been on testing for stochastic dominance, the finite population/design-based framework we develop can be extended to a variety of settings, such as hypothesis testing for the sign of an average treatment effect using worst-case treatment effect bounds, based on survey samples with missing data.

9 Acknowledgement

We are grateful to two referees and a co-editor for their speedy reports containing thoughtful and constructive remarks. Rami V. Tabri also thanks Matthew Joel Elias for helpful comments and feedback.

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.

Aryal, G. and M. F. Gabrielli (2013). Testing for collusion in asymmetric first-price auctions. *International Journal of Industrial Organization 31*(1), 26–35.

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.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley Series in Survey Methodology. John Wiley & Sons, Inc.

Godambe, V. and M. E. Thompson (2009). Chapter 26 - estimating functions and survey sampling. In C. Rao (Ed.), *Handbook of Statistics*, Volume 29 of *Handbook of Statistics*, pp. 83–101. Elsevier.

González, X. and M. J. Moral (2019). Effects of antitrust prosecution on retail fuel prices. *International Journal of Industrial Organization 67*, 102537.

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.

Kaur, A., B. Prakasa Rao, and H. Singh (1994). Testing for Second Order Stochastic Domiance of Two Distributions. *Econometric Theory 10*, 849–866.

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.

Martinez-Marquina, A., M. Niederle, and E. Vespa (2019). Failures in contingent reasoning: The role of uncertainty. *American Economic Review 109*(10), 3437 – 3474.

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.

Slachmuijlder, L. (2015). Lebanon: Building trust between police and local communities. In *Local Ownership in Security: Case Studies of Peacebuilding Approaches*, pp. 68–70. Alliance for Peacebuilding, GPPAC, Kroc Institute.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics 2*(1), 167–195.

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.

Zhao, P., D. Haziza, and C. Wu (2020). Survey weighted estimating equation inference with nuisance functionals. *Journal of Econometrics 216*(2), 516–536.

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, 2 and 3.
- Appendix C presents technical lemmas that are used in the proof of Theorem 2.
- Appendix D presents technical lemmas that are used in the proof of Theorem 3.
- Appendix E presents technical details and results for the discussion in Section 6.

B Proofs: Theorems 1, 2 and 3

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

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

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. ■

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. ■

B.3 Proof of Theorem 3

Proof. **Part 1**. The proof proceeds by the direct method. As in the proof of Theorem 2, denote by β the event of dominance in the sample. i.e., $\hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ$. By arguments identical to those in that theorem, we have $\lim_{N_A,N_B\to+\infty} E[1[\beta] \mid \Pi_{N_A,N_B}] = \text{Prob}\,[Z_s + \tau(s) < 0\ \forall s \in \mathcal{S}^\circ]$ where $Z \overset{d}{\sim} MVN(\mathbf{0},\Omega)$, and Ω is limiting covariance matrix of the first terms in (B.4). Since $\tau(s) = -\infty$ for each $s \in \mathcal{S}^\circ$, this limit must be equal to unity. Consequently, the event β occurs with probability tending to unity as $N_A, N_B \to +\infty$. This means that $LR^{(A,B)} = \min_{s\in\mathcal{S}^\circ} 2(L_{UR} - L_R(s))$, with probability tending to unity as $N_A, N_B \to +\infty$.

Next, Fix an $s \in \mathcal{S}^\circ$. Under the conditions of the theorem, we can apply Lemma D.1 to the sequence $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ to conclude that $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \overset{P}{\longrightarrow} +\infty$ as $N_A, N_B \to +\infty$. This argument applies to each $s \in \mathcal{S}^\circ$, so that this divergence of $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ holds for all $s \in \mathcal{S}^\circ$. Consequently, the form of $LR^{(A,B)}$, which is given by (3.4), implies that this divergence yields $LR^{(A,B)} \mid \Pi_{N_A,N_B} \overset{P}{\longrightarrow} +\infty$. Since the critical value $c(\alpha)$ is fixed and finite, it follows that $\lim_{N_A,N_B\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] = 1$. This concludes the proof.

Part 2. The proof proceeds by the direct method. Fix an $s \in S_0$. Under the conditions of the theorem, we can apply Lemma D.2 to conclude that $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \overset{d}{\longrightarrow} \chi_1^2$, as $N_A, N_B \to +\infty$. This argument applies to each $s \in S_0$, and note that the random variables $\{2(L_{UR} - L_R(s)),\ s \in S_0\}$ are statistically dependent for each N_A and N_B. From the proof of Lemma D.2, the statistics $\{2(L_{UR} - L_R(s)),\ s \in S_0\}$ are asymptotically equivalent to the squares of

$$\frac{\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}},\ s \in S_0,$$

whose asymptotic distribution is $MVN(\mathbf{0},\Xi)$ by Assumption 1, where Ξ is their limiting covariance matrix. Consequently, from the form of $LR^{(A,B)}$,

$$LR^{(A,B)} > c(\alpha) \iff 2(L_{UR} - L_R(s)) > c(\alpha)\ \forall s \in S_0 \text{ and } \hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ,$$

asymptotically, as Lemma D.1 establishes that $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \overset{P}{\longrightarrow} +\infty$ for $s \notin S_0$. As in Part 1 above, β denotes the event of dominance in the sample. i.e., $\hat{\overline{F}}_B(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ$. By arguments similar to those in the proof of Theorem 2, we have

$$\lim_{N_A,N_B\to+\infty} E[1[\beta] \mid \Pi_{N_A,N_B}] = \text{Prob}\,[Z_s + \tau(s) < 0,\ s \in \mathcal{S}^\circ] = \text{Prob}\,[Z_s < 0,\ s \in S_0]\,,$$

where $\{Z_s,\ s \in S_0\} \overset{d}{\sim} MVN(\mathbf{0},\Xi)$. Whence,

$$\lim_{N_A,N_B\to+\infty} E\left[1[LR^{(A,B)} > c(\alpha)] \mid \Pi_{N_A,N_B}\right] = \text{Prob}\left[Z_s^2 > c(\alpha),\ Z_s < 0,\ s \in S_0\right],$$

This concludes the proof. ∎

C Technical Lemmas for Theorem 2

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

$$\max_{p_i^Y:\, i \in U_Y,\, Y \in \{A,B\}} \sum_{Y \in \{A,B\}} \sum_{i \in U_Y} W_{Y,i} \log p_i^Y \quad \text{subject to}$$
$$p_i^Y > 0, \sum_{i \in U_Y} p_i^Y W_{Y,i} = 1 \quad \forall i \in U_Y \text{ and } Y = A, B, \text{ and } 0 = \sum_{i \in U_B} \sum_{j \in U_A} p_j^A p_i^B H_{i,j}(s).$$

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

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

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

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

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

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

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

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 $s_{\text{min}} = \min S^\circ_{N_{A,m},N_{B,m}}$. Then, $2(L_{UR} - L_R(s_{\text{min}})) \mid \Pi_{N_{A,m},N_{B,m}} \xrightarrow{d} \chi^2_1$ 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 $\{\Pi_{N_{A,m},N_{B,m}}\}^{+\infty}_{m=1}$. 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 $s_{\text{min}} = \min S^\circ_{N_{A,m},N_{B,m}}$. 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

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

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

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

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

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

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

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

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

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

Substituting these values into $\sum_i \sum_j p^A_j p^B_i H_{ij}(s_{\text{min}}) = 0$, we obtain

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

The equality (C.10) is equivalent to a convergent series in λ. Ignoring higher-order terms in λ to find

$\lambda \overset{a}{=} \bar{H}_{\bullet\bullet}(s_{\min})/D(s_{\min})$, where $\overset{a}{=}$ denotes asymptotic equivalence as $m \to +\infty$, and

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

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

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

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

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

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

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

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

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,

$$\left(\frac{\bar{H}_{\bullet\bullet}(s_{\min})}{\sqrt{\mathrm{VAR}(\bar{H}_{\bullet\bullet}(s_{\min}) \mid \Pi_{N_{A,m},N_{B,m}})}}\right)^2 \xrightarrow{d} \chi^2_1,$$

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

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

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

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

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

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

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

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

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

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

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

Lemma C.2. Let $\{\mathcal{M}^0_{N_A,N_B}, N_A, N_B = 1,2,\dots\}$ 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. ■

D Technical Lemmas for Theorem 3

The first result describes the asymptotic behavior of $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, for a given $s \in \mathcal{S}^\circ$, when the limit in (5.2) equals negative infinity.

Lemma D.1. Let $\{\mathcal{M}^1_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ be as described above, and $s \in \mathcal{S}^\circ$. Furthermore let $\tau(s)$ be given by (5.2). For each sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ such that $\Pi_{N_A,N_B} \in \mathcal{M}^1_{N_A,N_B}$ for each N_A and N_B, satisfying Assumptions 1 and 2, and $\tau(s) = -\infty$, we have $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \xrightarrow{P} +\infty$, as $N_A, N_B \to +\infty$.

Proof. The proof proceeds by the direct method. We have the survey samples $\{X_{Y,i}, D_{Y,i}, W_{Y,i} : i \in U_Y\}$ for $Y = A, B$, from population Π_{N_A,N_B}. For notational simplicity, we shall drop "$\mid \Pi_{N_A,N_B}$" from the notation, as it is clear that we are taking limits along $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$. Note that $s \in \mathcal{S}^\circ$ is non-random and does not change with the population sequence.

We follow steps identical to those in Lemma C.1 of the manuscript's appendix to obtain

$$2(L_{UR} - L_R(s)) = 2\sum_i W_{B,i} \log\left(1 + \frac{\lambda \tilde{H}_{i\bullet}(s)}{n_B}\right) + 2\sum_j W_{A,i} \log\left(1 + \frac{\lambda \tilde{H}_{\bullet j}(s)}{n_A}\right),$$

where $\lambda \overset{a}{=} \bar{H}_{\bullet\bullet}(s_{\text{min}})/D(s)$. Note that by duality arguments,

$$2(L_{UR} - L_R(s)) = \max_{\ell \in \mathbb{R}} \left[2\sum_i W_{B,i} \log\left(1 + \frac{\ell \acute{H}_{i\bullet}(s;\ell)}{n_B}\right) + 2\sum_j W_{A,i} \log\left(1 + \frac{\ell \acute{H}_{\bullet j}(s;\ell)}{n_A}\right)\right]$$

where

$$\acute{H}_{i\bullet}(s;\ell) = \sum_j p_j^A(\ell) W_{A,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s), \quad \acute{H}_{\bullet j}(s;\ell) = \sum_i p_i^B(\ell) W_{B,j} h(\overline{X}_{B,i}, \underline{X}_{A,j}, s),$$

and the $p_i^B(\ell)$ and $p_j^A(\ell)$ solve the system

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

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

Consequently,

$$2(L_{UR}-L_R(s)) \geq 2\sum_i W_{B,i}\log\left(1+\frac{\ell\acute{H}_{i\bullet}(s;\ell)}{n_B}\right)+2\sum_j W_{A,i}\log\left(1+\frac{\ell\acute{H}_{\bullet j}(s;\ell)}{n_A}\right), \tag{D.1}$$

for any $\ell \in \mathbb{R}$. We shall pick ℓ based on the sequence in (5.2), but suffice it for now, we just need $\ell = o_P(n_Y)$ as $N_A, N_B \to +\infty$, for $Y \in \{A, B\}$. This property ensures that

$$\frac{\ell\acute{H}_{i\bullet}(s;\ell)}{n_B}, \frac{\ell\acute{H}_{\bullet j}(s;\ell)}{n_A} \in (-1,1)$$

for each i and j, holds in large population sizes, since $\acute{H}_{i\bullet}(s;\ell), \acute{H}_{\bullet j}(s;\ell) \in (-1,1)$ is always true for each i and j. The importance of this point is that we can now legitimately apply the same expansion of $\log(1+x)$ for $|x|<1$, as in Lemma C.1, but to the right side of (D.1) and keep the terms up to order ℓ^2. In doing so, we obtain

$$\begin{aligned}
2(L_{UR}-L_R(s)) &\geq 2\sum_i W_{B,i}\log\left(1+\frac{\ell\acute{H}_{i\bullet}(s;\ell)}{n_B}\right)+2\sum_j W_{A,i}\log\left(1+\frac{\ell\acute{H}_{\bullet j}(s;\ell)}{n_A}\right), \\
&\stackrel{a}{=} 2\sum_i W_{B,i}\left[\frac{\ell\acute{H}_{i\bullet}(s;\ell)}{n_B}-\frac{1}{2}\left(\frac{\ell\acute{H}_{i\bullet}(s;\ell)}{n_B}\right)^2\right] \\
&\quad + 2\sum_j W_{A,i}\left[\frac{\ell\acute{H}_{\bullet j}(s;\ell)}{n_A}-\frac{1}{2}\left(\frac{\ell\acute{H}_{\bullet j}(s;\ell)}{n_A}\right)^2\right].
\end{aligned}$$

Noting that

$$\acute{H}_{i\bullet}(s;\ell) = \bar{H}_{i\bullet}(s) - \frac{\ell}{n_A^2}\sum_j W_{A,j}h(\overline{X}_{B,i}, \underline{X}_{A,j}, s)\acute{H}_{\bullet j}(s;\ell) \quad \text{and} \tag{D.2}$$

$$\acute{H}_{\bullet j}(s;\ell) = \bar{H}_{\bullet j}(s) - \frac{\ell}{n_B^2}\sum_i W_{B,i}h(\overline{X}_{B,i}, \underline{X}_{A,j}, s)\acute{H}_{i\bullet}(s;\ell), \tag{D.3}$$

the choice of $\ell = o_P(n_Y)$ for $Y \in \{A, B\}$, also allows us to replace $\acute{H}(s;\ell)$'s by corresponding $\bar{H}(s)$'s, to obtain

$$\begin{aligned}
2(L_{UR}-L_R(s)) &\stackrel{a}{\geq} 2\sum_i \frac{W_{B,i}\ell\bar{H}_{i\bullet}(s)}{n_B} - \frac{2\ell^2}{n_A}\sum_j \frac{W_{A,i}}{n_A}\bar{H}_{\bullet j}^2(s) - \sum_i W_{B,i}\left(\frac{\ell\bar{H}_{i\bullet}(s)}{n_B}\right)^2 \\
&\quad + 2\sum_j \frac{W_{A,j}\ell\bar{H}_{\bullet j}(s)}{n_A} - \frac{2\ell^2}{n_B}\sum_i \frac{W_{B,i}}{n_B}\bar{H}_{i\bullet}^2(s) - \sum_j W_{A,j}\left(\frac{\ell\bar{H}_{\bullet j}(s)}{n_A}\right)^2 \\
&= 4\ell\bar{H}_{\bullet\bullet}(s) - 3\frac{\ell^2}{n_A}\sum_j \frac{W_{A,i}}{n_A}\bar{H}_{\bullet j}^2(s) - 3\frac{\ell^2}{n_B}\sum_i \frac{W_{B,i}}{n_B}\bar{H}_{i\bullet}^2(s).
\end{aligned} \tag{D.4}$$

By setting ℓ as

$$\ell = \frac{-1}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}}, \tag{D.5}$$

the last two terms on the right side of (D.4) are $O_P(1)$ because

$$\sum_j \frac{W_{A,i}}{n_A} \bar{H}^2_{\bullet j}(s), \sum_i \frac{W_{B,i}}{n_B} \bar{H}^2_{i\bullet}(s) \in [0,1] \quad \text{with probability } 1$$

and

$$\frac{\ell^2}{n_Y} = \frac{1}{n_y \, \text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)} = \frac{K_{N_A,N_B}/n_Y}{K_{N_A,N_B} \, \text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}$$

is bounded in probability by Assumption 2, for $Y = A, B$. That is, Condition (ii) of that assumption ensures the denominator has a positive and finite limit, and Condition (i) of that assumption implies that the numerator satisfies

$$\frac{K_{N_A,N_B}}{n_Y} \overset{a}{=} \frac{K_{N_A,N_B}}{E[n_Y|\mathcal{P}_{Y,N_Y}]} \leq 1,$$

which holds for $Y = A, B$.

Now we shall focus on the first term on the right side of (D.4): $4\ell\bar{H}_{\bullet\bullet}(s)$, but with ℓ given by (D.5). We have

$$4\ell\bar{H}_{\bullet\bullet}(s) = 4\ell\left(\bar{H}_{\bullet\bullet}(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right)\right) + 4\ell\left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right).$$

Substituting out ℓ yields

$$4\ell\bar{H}_{\bullet\bullet}(s) = -4\left(\frac{\bar{H}_{\bullet\bullet}(s) - \left(\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}}\right) + \frac{4\left(\underline{F}_{A,N_A}(s) - \overline{F}_{B,N_B}(s)\right)}{\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}}. \tag{D.6}$$

The first term on the right side of (D.6) converges in distribution to $N(0,16)$ by Part (iv) of Assumption 2, and the second term diverges to $-\tau(s) = +\infty$. Consequently, $2(L_{UR} - L_R(s))$ diverges to $+\infty$ as $N_A, N_B \to +\infty$.

Finally, all we need to complete the proof is to establish that the choice of ℓ in (D.5) satisfies $\ell = o_P(n_Y)$

as $N_A, N_B \to +\infty$, under Assumptions 1 and 2. To that end, consider the case $Y = A$. Then note that

$$\left|\frac{\ell}{n_A}\right| = \frac{1}{n_A\sqrt{\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}} = \frac{\sqrt{K_{N_A,N_B}/n_A}\,(1/\sqrt{n_A})}{\sqrt{K_{N_A,N_B}\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}}$$
$$\overset{a}{=} \frac{\sqrt{K_{N_A,N_B}/E[n_A|\mathcal{P}_{A,N_A}]}\,(1/\sqrt{n_A})}{\sqrt{K_{N_A,N_B}\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)}} \tag{D.7}$$

by Part (i) of Assumption 2. Since

$$\lim_{N_A,N_B\to+\infty}\sqrt{\frac{K_{N_A,N_B}}{E[n_A \mid \mathcal{P}_{A,N_A}]}} = \lim_{N_A,N_B\to+\infty}\sqrt{\frac{E[n_B \mid \mathcal{P}_{B,N_B}]}{E[n_B \mid \mathcal{P}_{B,N_B}] + E[n_A \mid \mathcal{P}_{A,N_A}]}}$$
$$\leq \limsup_{N_A,N_B\to+\infty}\sqrt{\frac{E[n_B \mid \mathcal{P}_{B,N_B}]}{E[n_B \mid \mathcal{P}_{B,N_B}] + E[n_A \mid \mathcal{P}_{A,N_A}]}} \leq 1,$$

and Part (ii) of Assumption 2 sets $\lim_{N_A,N_B\to+\infty} K_{N_A,N_B}\text{VAR}\left(\hat{\overline{F}}_B(s) - \hat{\underline{F}}_A(s) \mid \Pi_{N_A,N_B}\right)$ to be positive and finite, the expression (D.7) converges to zero in probability, because Part (i) of Assumption 1 implies that

$$\frac{1}{\sqrt{n_A}} \mid \mathcal{P}_{A,N_A} \xrightarrow{P} 0.$$

Therefore, $\ell = o_P(n_A)$ as $N_A, N_B \to +\infty$, and an identical argument applies to the case $Y = B$, which we omit for brevity. This concludes the proof. ■

The next result describes the asymptotic behavior of $\{2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$, for a given $s \in \mathcal{S}^\circ$, when the limit in (5.2) does not diverge to $-\infty$.

Lemma D.2. Let $\{\mathcal{M}^1_{N_A,N_B}, N_A, N_B = 1, 2, \ldots\}$ be as described above, and $\tau(\cdot)$ be given by (5.2). Suppose that the sequence of finite populations $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$ is such that $\Pi_{N_A,N_B} \in \mathcal{M}^1_{N_A,N_B}$ for each N_A and N_B, and satisfies Assumptions 1 and 2. Let $s \in \mathcal{S}^\circ$ be such that $\tau(s) = 0$ for this sequence. Then $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \xrightarrow{d} \chi^2_1$, as $N_A, N_B \to +\infty$.

Proof. The proof proceeds by the direct method. We have the survey samples $\{X_{Y,i}, D_{Y,i}, W_{Y,i} : i \in U_Y\}$ for $Y = A, B$, from population Π_{N_A,N_B}. For notational simplicity, we shall drop "$\mid \Pi_{N_A,N_B}$" from the notation, as it is clear that we are taking limits along $\{\Pi_{N_A,N_B}\}_{N_A,N_B=1}^{+\infty}$.

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

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

Now the conditions of Assumption 1 (in the manuscript), Assumption 2, and $\tau(s) = 0$ imply

$$\left(\frac{\bar{H}_{\bullet\bullet}(s) - (\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s)) + (\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s))}{\sqrt{\text{VAR}(\bar{H}_{\bullet\bullet}(s) \mid \Pi_{N_A,N_B})}} \right)^2 \mid \Pi_{N_A,N_B} \xrightarrow{d} Z_s^2$$

as $N_A, N_B \to +\infty$, hold, where $Z_s \sim N(0,1)$. All that remains is to complete the proof is to establish that

$$\text{and} \quad \text{VAR}(\bar{H}_{\bullet\bullet}(s) \mid \Pi_{N_A,N_B})/D(s) \mid \Pi_{N_A,N_B} \xrightarrow{P} 1,$$

as $N_A, N_B \to +\infty$.

Unpacking the form of $D(s)$ reveals that it is equal to

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

where $\gamma(s) = \lim_{N_A,N_B \to +\infty} (\overline{F}_{B,N_B}(s) - \underline{F}_{A,N_A}(s))$. Now Conditions (ii) and (vi) of Assumption 1 along this sequence in the alternative hypothesis imply that $(\bar{H}_{\bullet\bullet}(s) - \gamma(s)) \xrightarrow{P} 0$ as $N_A, N_B \to +\infty$, and $\tau(s) = 0$ implies that $\gamma(s) = 0$, so that

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

Consequently, $\text{VAR}(\bar{H}_{\bullet\bullet}(s) \mid \Pi_{N_A,N_B})/D(s)$ is asymptotically equivalent to the design-effect, $Deff(s)$, which is defined in Condition (v) of Assumption 1. Therefore, this ratio converges in probability to 1. Thus, $2(L_{UR} - L_R(s)) \mid \Pi_{N_A,N_B} \xrightarrow{d} \chi_1^2$, as $N_A, N_B \to +\infty$.

■

E Example 2 in Section 6

This section develops the bounds in (6.4), and describes the testing procedure for the test problem (6.7). The point of departure is following representation of $F_{A,N_A,0}$, which holds for each for each $s \in \mathcal{S}^\circ$,

$$F_{A,N_A,0}(s) = \left[\frac{\sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}\right] \left[\frac{R_Y(s)}{1 - R_Y(s)}\right] F_{A,N_A,1}(s) \quad \text{where} \tag{E.1}$$

$$R_Y(s) = \frac{N_Y^{-1} \sum_{i=1}^{N_Y} 1\left[D_{Y,i} = 0, X_{Y,i} \leq s\right]}{N_Y^{-1} \sum_{j=1}^{N_Y} 1\left[X_{Y,i} \leq s\right]}. \tag{E.2}$$

By substituting the inequalities (6.3) into this representation, we deduce the following bounds:

$$\begin{aligned}
F_{A,N_A,0}(s) &\leq \left[\frac{\sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}\right] \left[\frac{\beta_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}{1 - \beta_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right] F_{A,N_A,1}(s) \\
&= \left[\frac{\beta_s \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \beta_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right] F_{A,N_A,1}(s), \quad \text{and} \\
F_{A,N_A,0}(s) &\geq \left[\frac{\sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}\right] \left[\frac{\alpha_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}{1 - \alpha_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right] F_{A,N_A,1}(s) \\
&= \left[\frac{\alpha_s \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}}{1 - \alpha_s\left(1 - \sum_{j=1}^{N_Y} \frac{D_{Y,j}}{N_Y}\right)}\right] F_{A,N_A,1}(s).
\end{aligned}$$

Now we can substitute the aforementioned bounds into the representation (2.6) to obtain bounds on the CDF F_{Y,N_Y}, given by (6.4).

Next, we describe the steps for implementing our testing procedure but for the test problem (6.7). The testing procedure we have developed, which is based on the worst-case bounds, must be adjusted to now account for the presence of a nuisance parameter $\upsilon\,\overline{\kappa}$. In particular, for a given $s \in \mathcal{S}^\circ$, the equality restriction in the pseudo-empirical likelihood problem now becomes

$$0 = \sum_{\left\{i \in U_B : D_{B,i} = 1\right\}} \sum_{j \in U_A} p_j^A p_i^B \hat{W}_{B,i} W_{A,j}\, w\left(X_{B,i}, \underline{X}_{A,j}, 0; \hat{\overline{\kappa}}(s)\right) \quad \text{where} \tag{E.3}$$

$$\hat{\overline{\kappa}}(s) = \left(n_B^{-1} \sum_{i \in U_B} W_{B,i} 1\left[D_{B,i} = 1\right] + \left(n_B^{-1} \sum_{i \in U_B} W_{B,i} 1\left[D_{B,i} = 0\right]\right) \frac{\frac{\beta_s}{n_B} \sum_{i \in U_B} W_{B,i} 1\left[D_{B,i} = 0\right]}{1 - \frac{\beta_s}{n_B} \sum_{i \in U_B} W_{B,i} 1\left[D_{B,i} = 0\right]}\right) \tag{E.4}$$

is the plug-in estimator of $\overline{\kappa}(s)$, the estimating function w is given by (6.8), and the survey weights $\{\hat{W}_{B,i} : i \in U_B \text{ and } D_{B,i} = 1\}$ are normalised versions of their counterparts in $\{W_{B,i} :, i \in U_B \text{ and } D_{B,i} = 1\}$, so that

$$\sum_{\{i\in U_B:D_{B,i}=1\}} \hat{W}_{B,i} = \sum_{i\in U_B} W_{B,i}1\left[D_{B,i}=1\right]. \tag{E.5}$$

Note that summing over $\{i \in U_B : D_{B,i} = 1\}$ in (E.3) along with the normalisation (E.5) evacuates the parameter $\upsilon = \left[N_B^{-1}\sum_{j=1}^{N_B} D_{B,j}\right]^{-1}$ from the estimating equation (E.3). Hence, the maximisation of the PELF corresponding to the restriction (E.3) is

$$\begin{aligned}
\max_{\left\{p_i^B: i\in U_B \,\&\, D_{B,i}=1, \text{ and } p_j^A: j\in U_A\right\}} \sum_{\{i\in U_B:D_{B,i}=1\}} \hat{W}_{B,i}\log p_i^B + \sum_{j\in U_A} W_{A,j}\log p_j^A \quad \text{subject to} \\
p_i^B > 0\ \forall i, \sum_{\{i\in U_B:D_{B,i}=1\}} p_i^B\hat{W}_{B,i} = 1,\ p_j^A > 0\ \forall j, \sum_{j\in U_A} p_i^A W_{A,j} = 1, \\
\text{and } 0 = \sum_{\{i\in U_B:D_{B,i}=1\}}\sum_{j\in U_A} p_j^A p_i^B \hat{W}_{B,i}W_{A,j}\, w\left(X_{B,i}, \underline{X}_{A,j}, 0; \hat{\overline{\kappa}}(s)\right).
\end{aligned}$$

Let $L_R(s)$ denote the maximal value of this optimization problem. The unconstrained estimators of $\overline{F}_{B,N_B}^M$ and $\underline{F}_{A,N_A}$ are the Hájek estimators

$$\hat{\overline{F}}_B^M(s) = \frac{\sum_{\{i\in U_B:D_{B,i}=1\}} \hat{W}_{B,i}\,\hat{\overline{\kappa}}(s)\,1\left[\overline{X}_{B,i}\le s\right]}{\sum_{i\in U_B} W_{B,i}1\left[D_{B,i}=1\right]} \text{ and } \hat{\underline{F}}_A(s) = n_A^{-1}\sum_{j\in U_A} W_{A,j}1\left[\underline{X}_{A,j}\le s\right], \forall s\in\mathcal{S},$$

respectively. The pseudo-empirical likelihood-ratio test statistic in this case is thus defined as

$$LR^{(A,B)} = \begin{cases} \min_{s\in\mathcal{S}^\circ} 2(L_{UR} - L_R(s)) & \text{if } \hat{\overline{F}}_B^M(s) < \hat{\underline{F}}_A(s)\ \forall s\in\mathcal{S}^\circ, \\ 0 & \text{otherwise,} \end{cases} \tag{E.6}$$

where

$$L_{UR} = \sum_{\{i\in U_B:D_{B,i}=1\}} \hat{W}_{B,i}\log(1/n_{B,1}) + \sum_{j\in U_A} W_{A,j}\log(1/n_A) \text{ with } n_{B,1} = \sum_{i\in U_B} W_{B,i}\,1\left[D_{B,i}=1\right],$$

is the unconstrained maximum value of the PELF. The formulation of $LR^{(A,B)}$ implements the procedure if we observe the sample satisfies $\hat{\overline{F}}_B^M(s) < \hat{\underline{F}}_A(s)\ \forall s \in \mathcal{S}^\circ$; that is, dominance in the sample, holds. Otherwise, $LR^{(A,B)} = 0$ and we do not reject H_0^3 in (6.7).

The decision rule is to reject H_0^3 if and only if $LR^{(A,B)} > c(\alpha)$, where $c(\alpha)$ is the $1-\alpha$ quantile from the χ_1^2 distribution. On the asymptotic validity of this procedure, we expect that the same line of arguments as in proof of Theorem 1 to be used, with appropriate adjustments, to account for the presence of the nuisance parameter $\overline{\kappa}(\cdot)$. This is because the convergence rate of $\hat{\overline{\kappa}}(\cdot)$ to $\overline{\kappa}(\cdot)$ is typically of the order root-n; see for example, Zhao et al. (2020). Developing the asymptotic validity of this testing procedure goes beyond the intended scope of the paper and is left for a future project.