北京大学国家发展研究院
National School of Development at Peking University
中国经济研究中心
China Centre for Economic Research
讨论稿系列
Working Paper Series

E2023003 2023-02-20

A Unified Nonparametric Test of Transformations on Distribution Functions with Nuisance Parameters

Zhenting Sun
China Center for Economic Research
National School of Development
Peking University

Abstract

This paper proposes a simple unified approach to testing transformations on cumulative distribution functions (CDFs) in the presence of nuisance parameters. The proposed test is constructed based on a new characterization that avoids the estimation of nuisance parameters. The critical values are obtained through a numerical bootstrap method which can easily be implemented in practice. Under suitable conditions, the proposed test is shown to be asymptotically size controlled and consistent. The local power property of the test is established. Finally, Monte Carlo simulations and an empirical study show that the test performs well on finite samples.

Keywords: Unified test, transformations on CDFs, nuisance parameters, numerical boot-strap

A Unified Nonparametric Test of Transformations on Distribution Functions with Nuisance Parameters*

Xingyu Li
National School of Development
Peking University
x.y@pku.edu.cn

Xiaojun Song†
Guanghua School of Management
Center for Statistical Science
Peking University
sxj@gsm.pku.edu.cn

Zhenting Sun‡
China Center for Economic Research
National School of Development
Peking University
zhentingsun@nsd.pku.edu.cn

August 17, 2022

Abstract

This paper proposes a simple unified approach to testing transformations on cumulative distribution functions (CDFs) in the presence of nuisance parameters. The proposed test is constructed based on a new characterization that avoids the estimation of nuisance parameters. The critical values are obtained through a numerical bootstrap method which can easily be implemented in practice. Under suitable conditions, the proposed test is shown to be asymptotically size controlled and consistent. The local power property of the test is established. Finally, Monte Carlo simulations and an empirical study show that the test performs well on finite samples.

Keywords: Unified test, transformations on CDFs, nuisance parameters, numerical bootstrap

*The authors thank Zheng Fang, Yixiao Sun, and all seminar participants for their insightful comments.
†Research supported by the National Natural Science Foundation of China (71973005)
‡Research supported by the National Natural Science Foundation of China (72103004)

1 Introduction

In many economic and financial applications, the underlying scientific problems concern whether the distributions of two samples, and more generally multiple samples, are the same, without imposing any parametric assumption on the underlying populations. Despite its central role in statistical inference and applications, testing the equality of two or multiple distributions may be restrictive in some scenarios where we would like to know if more refined structural relations exist between the populations of interest.

Of particular interest is the question whether the two or multiple distributions are simply location, scale, or location-scale transformations of each other. For illustration, we focus the discussion on the two samples problem in the following. We denote by Y and X the two random variables under study. If Y has the same distribution as $X - \theta_1$ for some unspecified location parameter $\theta_1 \in \mathbb{R}$, then Y is a location transformation of X. If Y has the same distribution as X/θ_2 for some unspecified scale parameter $\theta_2 > 0$, then Y is a scale transformation of X. If Y has the same distribution as $(X - \theta_1)/\theta_2$ for some $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$, then Y is a location-scale transformation of X. Several generalizations of the classical two samples tests to the location-scale families have been investigated in the literature. For example, Hall et al. (2013) propose an extension of the Cramér–von Mises type test based on empirical characteristic functions to examine whether the two samples come from the same location-scale family of distributions. Henze et al. (2005) and Jiménez-Gamero et al. (2017) deal with the two samples location problem using similar test statistics.

There has also been considerable attention devoted to testing location-scale transformations in the classical two samples treatment-control problem as formulated by Doksum (1974) and Lehmann and D'Abrera (1975). The null hypotheses of interest are whether the corresponding treatment induces a location shift, a scale shift, or a location-scale shift in the potential outcome distribution. Specifically, let the two random variables Y and X now represent the control and experimental outcomes from a randomized experiment, with distribution functions G and F, respectively. Then it is possible that the two distributions differ only by a location shift, so $F(x) = G(x - \theta_1)$, or that they differ by a scale shift, so $F(x) = G(x/\theta_2)$, or that they differ by a location-scale shift, so $F(x) = G((x - \theta_1)/\theta_2)$. In particular, Cox (1984) defines that there exists a "constant treatment effect" if the location shift hypothesis is satisfied. Several methods have been considered to deal with the nuisance parameters such as θ_1 and θ_2 in the above problems. Koenker and Xiao (2002) suggest tests of the location transformation and the location-scale transformation based on Khmaladze (1981)'s martingale transformation in the framework of quantile treatment effects. Ding et al. (2016) focus on the location shift hypothesis with the average treatment effect acting as a nuisance parameter (i.e., θ_1) and propose a randomization-based test for the null hypothesis of no treatment effect heterogeneity. More recently, for the same problem of testing heterogeneous treatment effects, Chung and Olivares (2021) propose a permutation test based on the Khmaladze's martingale transformation to tackle the estimated nuisance parameter, Ramirez-Cuellar (2021) suggests a new test using empirical characteristic

functions, and Chung and Olivares (2022) develop a permutation test based on a modified quantile process. Lastly, we mention that testing the location-scale shift hypothesis is interesting in the context of local quantile treatment effects; see, e.g., Melly and Wüthrich (2017) for a detailed discussion.

Transformations such as the aforementioned location and scale ones are obviously parametric transformations between two random variables. Though the present paper focuses on location and scale transformations, the proposed method can be applied to general parametric transformations. The unknown location and scale parameters act as nuisance parameters, which need to be taken into account properly in the corresponding test procedures. Classical approaches may meet theoretical difficulties as the estimation of nuisance parameters introduces extra uncertainty. In addition, these approaches usually operate in a case-dependent manner; that is, different theories and implementation procedures are required for testing different transformations. For example, the existing theories and implementation procedures for testing location, scale, or location-scale transformations are different. Even for testing the same transformation, different methods of estimating the nuisance parameters would produce different limit distributions and test results. Thus, from both theoretical and practical aspects, it is desirable to develop alternative methods that can circumvent these concerns.

The proposed approach successfully avoids the need to estimate the nuisance parameters, and thus offers a robust diagnostic device to deal with the generalized two samples and multiple samples problems. We summarize the main features of the proposed test as follows: (i) It is case-independent; (ii) it is asymptotically size controlled and consistent against a broad class of alternatives to the null while being free of the estimation of the nuisance parameters; (iii) it works for both independent (multiple) samples and paired samples; and (iv) the bootstrap test procedure is simple.

Related Literature

There exist a substantial number of tests for the problem of comparing two or multiple distributions. See, for example, Lehmann and Romano (2005) and Chen and Pokojovy (2018) for extensive reviews of them. Historically, the most popular and most commonly used two or multiple samples tests are those based on comparing empirical distribution functions. See, for example, the Kolmogorov–Smirnov, the Cramér–von Mises, and the Anderson–Darling tests by Smirnov (1939), Lehmann (1951), Rosenblatt (1952), Darling (1957), Kiefer (1959), Fisz (1960), Anderson (1962), and Scholz and Stephens (1987). Another class of tests are based on empirical characteristic functions. See, for example, Epps and Singleton (1986), Alba et al. (2001), Meintanis (2005), and Fernández et al. (2008). For more contributions on the topic of two or multiple samples problems based on various approaches, see, for example, Anderson et al. (1994), Székely et al. (2004), Zhang and Wu (2007), Martínez-Camblor and de Uña-Álvarez (2009), Bera et al. (2013), Goldman and Kaplan (2018), Chen (2020), and Song and Xiao (2022). The list is surely not exhaustive.

The critical values of our test are constructed based on the numerical bootstrap methods proposed by Hong and Li (2018) and Chen and Fang (2019b), who provide a novel methodology

for nonstandard testing issues. More discussions on this topic can be found in Dümbgen (1993), Andrews (2000), Hirano and Porter (2012), Hansen (2017), and Fang and Santos (2018). Other applications of related bootstrap methods can be found in Beare and Moon (2015), Beare and Fang (2017), Seo (2018), Beare and Shi (2019), Chen and Fang (2019a), Sun and Beare (2021), and Sun (2021).

Organization of the Paper

For independent interests, the paper discusses testing transformations on two CDFs and on multiple CDFs in two separate sections. Section 2 provides the framework and develops theoretical results for testing general parametric transformations in the context of two samples. Section 3 generalizes the results to transformations on multiple samples. Section 4 provides Monte Carlo simulation evidence to show the performance of the test on finite samples. In Section 5, we apply the proposed test to analyze age distributions. Section 6 concludes the paper. All the mathematical proofs are collected in the Online Supplementary Appendix.

Notation

Throughout the paper, all the random elements are defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let F and G be two unknown continuous CDFs on $\mathbb{R}$. Let μ be the Lebesgue measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$, where $\mathscr{B}_{\mathbb{R}}$ denotes the collection of Borel sets in $\mathbb{R}$. For an arbitrary set A, let $\ell^{\infty}(A)$ be the set of bounded real valued functions on A. Equip $\ell^{\infty}(A)$ with the supremum norm $\|\cdot\|_{\infty}$ such that $\|f\|_{\infty} = \sup_{x\in A} |f(x)|$ for every $f \in \ell^{\infty}(A)$. For a subset B of a metric space, let $\mathcal{C}(B)$ be the set of continuous real valued functions on B, and $\mathcal{C}_{\mathrm{b}}(B)$ be the set of bounded continuous functions on B, that is, $\mathcal{C}_{\mathrm{b}}(B) = \mathcal{C}(B) \cap \ell^{\infty}(B)$. Following the notation of van der Vaart and Wellner (1996), for every normed space $\mathbb{B}$ with a norm $\|\cdot\|_{\mathbb{B}}$, we define

$$\mathrm{BL}_1(\mathbb{B}) = \{\Gamma : \mathbb{B} \to \mathbb{R} : |\Gamma(a)| \leq 1 \text{ and } |\Gamma(a) - \Gamma(b)| \leq \|a - b\|_{\mathbb{B}} \text{ for all } a, b \in \mathbb{B}\}.$$

Let $\rightsquigarrow$ denote the weak convergence defined in van der Vaart and Wellner (1996, p. 4). Let $\overset{\mathbb{P}}{\rightsquigarrow}$ and $\overset{\text{a.s.}}{\rightsquigarrow}$ denote the weak convergence conditional on the sample in probability and almost surely, respectively, as defined in Kosorok (2008, pp. 19–20). For every continuous CDF $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$, let $\mathbb{W}_f$ denote a tight Borel measurable centered Gaussian process with covariance function $\mathbb{E}[\mathbb{W}_f(x_1)\mathbb{W}_f(x_2)] = f(x_1 \wedge x_2) - f(x_1) f(x_2)$ for all $x_1, x_2 \in \mathbb{R}$.

For every measure ν on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$, let $L^p(\nu)$ be the set of functions such that

$$L^p(\nu) = \left\{ f : \mathbb{R} \to \mathbb{R} : \int_{\mathbb{R}} [f(x)]^p \, \mathrm{d}\nu(x) < \infty \right\}$$

with $p \geq 1$. Equip $L^p(\nu)$ with the norm $\|\cdot\|_{L^p(\nu)}$ such that

$$\|f\|_{L^p(\nu)} = \left\{ \int_{\mathbb{R}} [f(x)]^p \, \mathrm{d}\nu(x) \right\}^{1/p}$$

for every $f \in L^p(\nu)$. Let $\mathbb{F}$ be an arbitrary vector space equipped with a norm $\|\cdot\|_{\mathbb{F}}$. For every $C \subset \mathbb{F}$ and every $\varepsilon > 0$, define the ε-neighborhood of C to be

$$C^{\varepsilon} = \left\{ g \in \mathbb{F} : \inf_{f \in C} \|f - g\|_{\mathbb{F}} \leq \varepsilon \right\}.$$

2 Transformations on Two CDFs

In this section, we consider parametric transformations between two random variables X and Y, with respective CDFs F and G. Though our empirical focus is on the location-scale transformation, for the sake of theoretical generality, we present the results for general parametric transformations that include the location-scale transformation as a special case. To begin with, let $\mathscr{G}$ be a space of functions that can be identified by a finite-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ for some $d_\theta \in \mathbb{Z}_+$, that is,

$$\mathscr{G} = \left\{ g(\cdot, \theta) : \mathbb{R} \to \mathbb{R} : \theta \in \Theta \subset \mathbb{R}^{d_\theta} \right\},$$

where $g : \mathbb{R} \times \Theta \to \mathbb{R}$ is a known function. We are interested in the hypothesis

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, F(x) = G\left(g(x, \theta)\right) \text{ for all } x \in \mathbb{R}. \tag{1}$$

The parameter θ in (1) is the nuisance parameter we need to consider in the test. A leading example of $g(x, \theta)$ is as follows.

Example 2.1: (Location-scale Transformation) Suppose that Y is equivalent to $(X - \theta_1)/\theta_2$ in distribution for some $\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}_+$. Then, the CDFs of X and Y satisfy

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - \theta_1}{\theta_2} \leq \frac{x - \theta_1}{\theta_2}\right) = \mathbb{P}\left(Y \leq \frac{x - \theta_1}{\theta_2}\right) = G\left(\frac{x - \theta_1}{\theta_2}\right).$$

Let $\Theta = [a_1, b_1] \times [a_2, b_2]$, where $-\infty < a_1 < b_1 < \infty$ and $0 < a_2 < b_2 < \infty$. In this case, the parameter $\theta = (\theta_1, \theta_2) \in \Theta$, and the function

$$g(x, \theta) = \frac{x - \theta_1}{\theta_2} \text{ for all } x \in \mathbb{R} \text{ and all } \theta \in \Theta.$$

Validating location-scale transformations is important for conducting causal inference in randomized experiments, as already discussed in the introduction. In particular, testing for the presence of idiosyncratic treatment effect heterogeneity is often achieved by testing whether the location transformation $F(x) = G(x - \theta_1)$ is satisfied for an unknown constant θ_1. That is, testing whether the distribution functions of the potential outcomes in the treatment and the control groups are shifted by θ_1; see, e.g., Ding et al. (2016), Chung and Olivares (2021), Ramirez-Cuellar (2021), and Chung and Olivares (2022). Also, testing scale transformation $F(x) = G(x/\theta_2)$ or location-scale transformation $F(x) = G((x - \theta_1)/\theta_2)$ is indispensable to the analysis of quantile treatment effects as well motivated by Koenker and Xiao (2002).

Another important economic application of location-scale transformations of random variables can be found in decision theory. Meyer (1987) shows that the location-scale transformation restriction is sufficient to ensure consistency between expected utility and moment-based rankings of random variables, and this restriction holds in many economic models. Indeed, as pointed out in Meyer (1987), the location-scale condition is empirically important and needs to be tested to explain the similarities in findings using expected utility and mean-standard deviation techniques. In practice, if we are interested in groups of individuals who share the same distributions in the same groups, the location-scale restriction can be tested using our approach.

Example 2.2: (Location-scale Transformation for Log-transformed Data) Suppose that we consider the log-transformed random variables $Z_1 = \log X$ and $Z_2 = \log Y$ and wish to test

whether Z_2 has the same distribution as $(Z_1-\theta_1)/\theta_2$ for some $\theta_1\in\mathbb{R}$ and $\theta_2\in\mathbb{R}_+$. Clearly, we can use the location-scale transformation test for Z_1 and Z_2 introduced in Example 2.1. That Z_1 and Z_2 satisfy the location-scale transformation is equivalent to that the CDFs of the original random variables X and Y satisfy

$$\begin{aligned} F(x) =& \mathbb{P}(X\le x)=\mathbb{P}\left(\frac{\log X-\theta_1}{\theta_2}\le\frac{\log x-\theta_1}{\theta_2}\right)=\mathbb{P}\left(\log Y\le\frac{\log x-\theta_1}{\theta_2}\right)\\ =&\mathbb{P}\left(Y\le\left(\frac{x}{e^{\theta_1}}\right)^{\frac{1}{\theta_2}}\right)=G\left(\left(\frac{x}{e^{\theta_1}}\right)^{\frac{1}{\theta_2}}\right)\ \text{for all } x>0. \end{aligned} \tag{2}$$

Let $\Theta=[a_1,b_1]\times[a_2,b_2]$, where $-\infty<a_1<b_1<\infty$ and $0<a_2<b_2<\infty$. In this case, the parameter $\theta=(\theta_1,\theta_2)\in\Theta$, and the function

$$g(x,\theta)=\left(\frac{x}{e^{\theta_1}}\right)^{\frac{1}{\theta_2}}\ \text{for all } x>0 \text{ and all } \theta\in\Theta. \tag{3}$$

As a result, we can also use the proposed method to test the transformation $g(x,\theta)$ defined in (3) for the original random variables. We note that log-transformed data has been widely used in economics, social psychology, biomedical science, and many other disciplines to deal with highly skewed data and make data conform to normality.

Let ν be a probability measure on $(\mathbb{R},\mathscr{B}_{\mathbb{R}})$. We first introduce the following assumptions.

Assumption 2.1: For every $\theta\in\Theta$, the function $x\mapsto g(x,\theta)$ is continuous and increasing.

Assumption 2.2: The probability measure ν on $(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ satisfies $\mu\ll\nu$, that is, if $\nu(B)=0$ for some $B\in\mathscr{B}_{\mathbb{R}}$, then $\mu(B)=0$.

Assumption 2.3: The set Θ is compact in $\mathbb{R}^{d_\theta}$.

Assumption 2.4: For every $f\in\mathcal{C}_{\mathrm{b}}(\mathbb{R})$, the map $\theta\mapsto f(g(\cdot,\theta))$, from Θ to $L^2(\nu)$, is continuous. That is, for an arbitrary fixed $\theta_0\in\Theta$ and every $\varepsilon>0$, there exists $\delta>0$ such that

$$\int_{\mathbb{R}}\left[f\left(g(x,\theta)\right)-f\left(g(x,\theta_0)\right)\right]^2\,\mathrm{d}\nu(x)<\varepsilon$$

for all $\theta\in\Theta$ with $\|\theta-\theta_0\|_2<\delta$.

Assumption 2.1 shows that we focus on the transformations that are continuous and increasing. Assumption 2.2 requires the absolute continuity of the Lebesgue measure μ with respect to the probability measure ν. For example, ν could be set to be the probability measure of a normally distributed random variable. Assumption 2.3 is a common condition on the compactness of Θ. Assumption 2.4 imposes restrictions on the structure of $\mathscr{G}$ under the measure ν. One sufficient condition for Assumption 2.4 is that $g(x,\cdot)$ is continuous on Θ for every x.

Define a function space

$$\mathbb{D}_{\mathcal{L}0}=\left\{\varphi\in\ell^\infty(\mathbb{R}\times\Theta):\theta\mapsto\varphi(\cdot,\theta),\ \text{as a map from }\Theta\text{ to }L^2(\nu),\ \text{is continuous}\right\}.$$

In the definition of $\mathbb{D}_{\mathcal{L}0}$, the continuity of the map $\theta\mapsto\varphi(\cdot,\theta)$ is understood in the sense that for every $\theta_0\in\Theta$ and every $\varepsilon>0$, there exists $\delta>0$ such that

$$\int_{\mathbb{R}}\left[\varphi(x,\theta)-\varphi\left(x,\theta_0\right)\right]^2\,\mathrm{d}\nu(x)<\varepsilon$$

for all $\theta\in\Theta$ with $\|\theta-\theta_0\|_2<\delta$. For every $f:\mathbb{R}\to\mathbb{R}$, we define a map $f\circ g:\mathbb{R}\times\Theta\to\mathbb{R}$

such that $f \circ g(x,\theta) = f(g(x,\theta))$ for every $(x,\theta) \in \mathbb{R} \times \Theta$. Define $\phi : \mathbb{R} \times \Theta \to \mathbb{R}$ with $\phi(x,\theta) = F(x) - G(g(x,\theta))$ for every $(x,\theta) \in \mathbb{R} \times \Theta$. The proposition below provides an equivalent characterization of the null hypothesis in (1). We construct the test based on this equivalent characterization. The advantage of this characterization is that it avoids the estimation of the nuisance parameter θ under the null.

Proposition 2.1: If Assumptions 2.1–2.4 hold, then the null hypothesis in (1) is equivalent to

$$\mathrm{H}_0 : \inf_{\theta \in \Theta} \int_{\mathbb{R}} [F(x) - G(g(x,\theta))]^2 \, \mathrm{d}\nu(x) = 0. \tag{4}$$

It is noteworthy that different measures ν may deliver different power properties of the test. However, searching for the optimal ν to maximize power is challenging since it may depend in a very complicated way on the data generating process. We provide an empirical way of choosing ν in practice, which is illustrated in Section 5.

2.1 Test Statistic

To construct the test statistic, we first introduce the assumptions on the samples. Suppose that $\{X_i\}_{i=1}^{n_1}$ is a random sample drawn from F, and $\{Y_i\}_{i=1}^{n_2}$ is a random sample drawn from G.

Assumption 2.5: The samples $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ satisfy one of the conditions below:

(i) Independent samples: $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ are independently and identically distributed samples, and they are independent of each other.

(ii) Matched pairs: $n_1 = n_2$ and $\{(X_i, Y_i)\}_{i=1}^{n_1}$ is independently and identically distributed. For every $i \in \{1, \ldots, n_1\}$, the two-dimensional random vector (X_i, Y_i) has a cumulative distribution function H, whose marginal distribution functions are F and G.

Assumption 2.6: The ratio $n_1/n \to \lambda \in (0,1)$ as $n \to \infty$, where $n = n_1 + n_2$.

Assumption 2.5 allows the samples to be independent of each other or matched pairs. In Assumption 2.6, n_1 and n_2 are viewed as functions of n. As $n \to \infty$, $n_1 \to \infty$ and $n_2 \to \infty$. For matched pairs, $\lambda = 1/2$ by construction.

Define a function space

$$\mathbb{D}_{\mathcal{L}} = \left\{ \varphi \in \ell^{\infty}(\mathbb{R} \times \Theta) : \int_{\mathbb{R}} [\varphi(x,\theta)]^2 \, \mathrm{d}\nu(x) < \infty \text{ for all } \theta \in \Theta \right\}.$$

Define a map $\mathcal{L}$ on $\mathbb{D}_{\mathcal{L}}$ such that $\mathcal{L}(\varphi) = \inf_{\theta \in \Theta} \int_{\mathbb{R}} [\varphi(x,\theta)]^2 \, \mathrm{d}\nu(x)$ for every $\varphi \in \mathbb{D}_{\mathcal{L}}$. Then under Assumptions 2.1–2.4, the null and the alternative hypotheses can be expressed as

$$\mathrm{H}_0 : \mathcal{L}(\phi) = 0 \text{ and } \mathrm{H}_1 : \mathcal{L}(\phi) > 0. \tag{5}$$

To test the null hypothesis in (5), we need to find an estimator for ϕ based on the samples. The cumulative distribution functions F and G can be estimated by the empirical distribution functions $\widehat{F}_{n_1}$ and $\widehat{G}_{n_2}$, respectively, such that for every $x \in \mathbb{R}$,

$$\widehat{F}_{n_1}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{1}_{(-\infty, x]}(X_i) \text{ and } \widehat{G}_{n_2}(x) = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbb{1}_{(-\infty, x]}(Y_i).$$

In the matched pairs case, the joint CDF H for (X_i, Y_i) can be estimated by

$$\widehat{H}_{n_1}(x,y) = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{1}_{(-\infty,x]\times(-\infty,y]}(X_i, Y_i)$$

for all $x, y \in \mathbb{R}$. For every $x \in \mathbb{R}$ and every $\theta \in \Theta$, let

$$\widehat{\phi}_n(x,\theta) = \widehat{F}_{n_1}(x) - \widehat{G}_{n_2}(g(x,\theta)) = \widehat{F}_{n_1}(x) - \widehat{G}_{n_2} \circ g(x,\theta), \tag{6}$$

and we set the test statistic to be $T_n\mathcal{L}(\widehat{\phi}_n)$, where $T_n = n_1 n_2/n$.

Example 2.1 (Cont.)**:** To test the location-scale transformation, classical methods would rely on the following two samples empirical process with estimated parameters:

$$\widehat{S}_n(x) = \sqrt{T_n}\left(\widehat{F}_{n_1}(x) - \widehat{G}_{n_2}\left(\frac{x-\widehat{\theta}_1}{\widehat{\theta}_2}\right)\right) \text{ for all } x \in \mathbb{R},$$

where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are suitable $\sqrt{T_n}$-consistent estimators for θ_1 and θ_2, respectively. For example, a popular choice is that $\widehat{\theta}_2 = \widehat{\sigma}_X/\widehat{\sigma}_Y$ and $\widehat{\theta}_1 = \widehat{\mu}_X - \widehat{\theta}_2\widehat{\mu}_Y$, where $\widehat{\mu}_X$ and $\widehat{\mu}_Y$ are the sample means, and $\widehat{\sigma}_X$ and $\widehat{\sigma}_Y$ are the sample standard deviations. It can be shown that, uniformly in $x \in \mathbb{R}$,

$$\begin{aligned}\widehat{S}_n(x) = &\sqrt{T_n}\left(\widehat{F}_{n_1}(x) - \widehat{G}_{n_2}\left(\frac{x-\theta_1}{\theta_2}\right)\right) + \frac{1}{\theta_2} f_G\left(\frac{x-\theta_1}{\theta_2}\right)\sqrt{T_n}\left(\widehat{\theta}_1 - \theta_1\right) \\ &+ \frac{1}{\theta_2} f_G\left(\frac{x-\theta_1}{\theta_2}\right)\frac{x-\theta_1}{\theta_2}\sqrt{T_n}\left(\widehat{\theta}_2 - \theta_2\right) + o_p(1),\end{aligned}$$

where f_G is the density function of the CDF G. The first term of the above uniform decomposition is the infeasible two samples empirical process for testing a simple null hypothesis if θ_1 and θ_2 were known. The second and third terms represent the so-called "parameter estimation uncertainty" when θ_1 and θ_2 are unspecified and need to be estimated from the data by some $\widehat{\theta}_1$ and $\widehat{\theta}_2$. Moreover, to derive the limit distribution of $\widehat{S}_n$ under the null, asymptotically linear representations of $\sqrt{T_n}(\widehat{\theta}_1 - \theta_1)$ and $\sqrt{T_n}(\widehat{\theta}_2 - \theta_2)$ are often required. Different estimators for θ_1 and θ_2 may lead to different limit distributions and thus different test results. The unknown density function f_G appearing in the decomposition may also cause technical complications. In contrast, the proposed method avoids the estimation of θ_1 and θ_2, and therefore circumvents these issues.

Lemma 2.1 establishes the weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n \to \infty$.

Lemma 2.1: Under Assumptions 2.5 and 2.6, we have

$$\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0 \text{ in } \ell^\infty(\mathbb{R}\times\Theta)$$

as $n \to \infty$, where $\mathbb{G}_0$ is a tight random element with $Var(\mathbb{G}_0(x,\theta)) = Var\{\sqrt{1-\lambda}\mathbb{W}_F(x) - \sqrt{\lambda}(\mathbb{W}_G \circ g)(x,\theta)\}$ for every $(x,\theta) \in \mathbb{R}\times\Theta$. For the independent samples case, $\mathbb{W}_F$ is independent of $\mathbb{W}_G$. For the matched pairs case, $(\mathbb{W}_F, \mathbb{W}_G)$ is jointly Gaussian with $\mathbb{E}[\mathbb{W}_F(x_1)\mathbb{W}_G(x_2)] = H(x_1,x_2) - F(x_1)G(x_2)$ for all $x_1, x_2 \in \mathbb{R}$. If, in addition, Assumption 2.4 holds, then $\mathbb{P}(\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}) = 1$.

Next, we show that the map $\mathcal{L}$ is Hadamard directionally differentiable,[1] but its Hadamard

[1]See the definition of Hadamard directional differentiability in Definition A.1.

directional derivative is degenerate under H_0. Define $\mathbb{D}_0 = \{\varphi \in \mathbb{D}_{\mathcal{L}0} : \mathcal{L}(\varphi) = 0\}$.

Lemma 2.2: If Assumptions 2.3 and 2.4 hold, then $\mathcal{L}$ is Hadamard directionally differentiable at $\phi \in \mathbb{D}_{\mathcal{L}}$ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the Hadamard directional derivative

$$\mathcal{L}'_\phi(h) = 2 \inf_{\theta \in \Theta_0(\phi)} \int_{\mathbb{R}} \phi(x,\theta) h(x,\theta) \, \mathrm{d}\nu(x) \text{ for all } h \in \mathbb{D}_{\mathcal{L}0},$$

where $\Theta_0(\phi) = \arg\min_{\theta \in \Theta} \int_{\mathbb{R}} [\phi(x,\theta)]^2 \, \mathrm{d}\nu(x)$. Moreover, if $\phi \in \mathbb{D}_0$, then the derivative $\mathcal{L}'_\phi$ is well defined on the whole of $\ell^\infty(\mathbb{R} \times \Theta)$ with $\mathcal{L}'_\phi(h) = 0$ for every $h \in \ell^\infty(\mathbb{R} \times \Theta)$.

The first order degeneracy of $\mathcal{L}$ under H_0 implies that we may need to find the second order Hadamard directional derivative[2] of $\mathcal{L}$. We assume the following conditions to guarantee the existence of the second order Hadamard directional derivative of $\mathcal{L}$.

Assumption 2.7: The function $G \circ g$ is twice differentiable with respect to θ, and the second partial derivative satisfies

$$\int_{\mathbb{R}} \sup_{\theta \in \Theta} \left\| \frac{\partial^2 (G \circ g)(z,\vartheta)}{\partial \vartheta \partial \vartheta^\intercal} \bigg|_{(z,\vartheta)=(x,\theta)} \right\|_2^2 \mathrm{d}\nu(x) < \infty, \tag{7}$$

where $\|\cdot\|_2$ denotes the ℓ^2 operator norm of a matrix.

Assumption 2.8: The set $\Theta_0 \equiv \{\theta \in \Theta : \int_{\mathbb{R}} [\phi(x,\theta)]^2 \, \mathrm{d}\nu(x) = 0\} \subset \mathrm{int}(\Theta)$, and there exist some $\kappa \in (0,1]$ and some $C > 0$ such that for all small $\varepsilon > 0$,

$$\inf_{\theta \in \Theta \setminus \Theta_0^\varepsilon} \left\{ \int_{\mathbb{R}} [\phi(x,\theta)]^2 \, \mathrm{d}\nu(x) \right\}^{1/2} \geq C\varepsilon^\kappa. \tag{8}$$

We provide Assumptions 2.7 and 2.8 following the basic idea of Chen and Fang (2019b). Assumption 2.7 requires the boundedness of the second partial derivative of $G \circ g$ in the sense of (7). Assumption 2.8 requires that the set Θ_0 is in the interior of Θ and it is well separated. The condition in (8) is similar to the partial identification assumption used in Chernozhukov et al. (2007, p. 1265). Clearly, these high level conditions exclude some empirical applications, such as the cases where F and G are CDFs of discrete random variables. But it is worth noting that these conditions are sufficient but not necessary for our results, as also mentioned by Chen and Fang (2019b). We need such high level conditions for theoretical purposes. In Section 4 for the Monte Carlo simulations, we show that our methods work well in cases where such conditions do not hold.

Lemma 2.3: If Assumptions 2.3, 2.4, 2.7, and 2.8 hold, and $\phi \in \mathbb{D}_0$, then the function $\mathcal{L}$ is second order Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the second order Hadamard directional derivative

$$\mathcal{L}''_\phi(h) = \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V(a(h))} \left\| \left[\Phi'(\theta)\right]^\intercal v + \mathscr{H}(\theta) \right\|_{L^2(\nu)}^2 \text{ for all } h \in \mathbb{D}_{\mathcal{L}0},$$

where $\Phi'(\theta) : \mathbb{R} \to \mathbb{R}^{d_\theta}$ with

$$\Phi'(\theta)(x) = -\frac{\partial (G \circ g)(z,\vartheta)}{\partial \vartheta} \bigg|_{(z,\vartheta)=(x,\theta)} \quad \text{for every } (x,\theta) \in \mathbb{R} \times \Theta,$$

$V(a) = \left\{ v \in \mathbb{R}^{d_\theta} : \|v\|_2 \leq a \right\}$ for all $a > 0$, the number $a(h) > 0$ satisfies that $Ca(h)^\kappa = 3\|h\|_\infty$

[2] See the definition of second order Hadamard directional differentiability in Definition A.2.

with C and κ defined as in Assumption 2.8, and $\mathscr{H} : \Theta \to \ell^\infty(\mathbb{R})$ with $\mathscr{H}(\theta)(x) = h(x, \theta)$ for every $(x, \theta) \in \mathbb{R} \times \Theta$.

Remark 2.1: Lemma 2.3 provides the explicit expression of the complicated second order Hadamard directional derivative of $\mathcal{L}$. We employ a numerical method that does not need to explore this function form.

With Lemma 2.3, the asymptotic null distribution of the test statistic $\mathcal{L}(\widehat{\phi}_n)$ is obtained by applying the second order delta method.

Proposition 2.2: If Assumptions 2.1–2.8 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then

$$T_n \mathcal{L}(\widehat{\phi}_n) \rightsquigarrow \mathcal{L}''_\phi(\mathbb{G}_0) \text{ as } n \to \infty.$$

2.2 The Bootstrap

The distribution of $\mathcal{L}''_\phi(\mathbb{G}_0)$ in Proposition 2.2 is unknown because both the function $\mathcal{L}''_\phi$ and the stochastic process $\mathbb{G}_0$ depend on the unknown underlying distributions F and G. Motivated by Hong and Li (2018) and Chen and Fang (2019b), we propose to approximate $\mathcal{L}''_\phi$ by a consistent estimator and approximate the distribution of $\mathbb{G}_0$ by bootstrap. We use the numerical second order Hadamard directional derivative $\widehat{\mathcal{L}}''_n$ to approximate $\mathcal{L}''_\phi$, which is defined as

$$\widehat{\mathcal{L}}''_n(h) = \frac{\mathcal{L}(\widehat{\phi}_n + \tau_n h) - \mathcal{L}(\widehat{\phi}_n)}{\tau_n^2}$$

for all $h \in \ell^\infty(\mathbb{R} \times \Theta)$, where $\{\tau_n\}$ is a sequence of tuning parameters satisfying the assumption below.

Assumption 2.9: $\{\tau_n\} \subset \mathbb{R}_+$ is a sequence of scalars such that $\tau_n \downarrow 0$ and $\tau_n \sqrt{T_n} \to \infty$ as $n \to \infty$.

Assumption 2.9 provides the rate at which $\tau_n \downarrow 0$. Under this condition, we show that $\widehat{\mathcal{L}}''_n$ approximates $\mathcal{L}''_\phi$ well in the following lemma.

Lemma 2.4: If Assumptions 2.1–2.9 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then for every sequence $\{h_n\} \subset \ell^\infty(\mathbb{R} \times \Theta)$ and every $h \in \mathbb{D}_{\mathcal{L}0}$ such that $h_n \to h$ in $\ell^\infty(\mathbb{R} \times \Theta)$ as $n \to \infty$, we have

$$\widehat{\mathcal{L}}''_n(h_n) \xrightarrow{\mathbb{P}} \mathcal{L}''_\phi(h) \text{ as } n \to \infty.$$

We approximate the distribution of $\mathbb{G}_0$ via bootstrap. Let $\{X_i^*\}_{i=1}^{n_1}$ and $\{Y_i^*\}_{i=1}^{n_2}$ be the bootstrap samples satisfying the following conditions:

(i) For independent samples: Given the raw samples $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$, the bootstrap samples $\{X_i^*\}_{i=1}^{n_1}$ and $\{Y_i^*\}_{i=1}^{n_2}$ are i.i.d. samples drawn independently from the empirical distributions $\widehat{F}_{n_1}$ and $\widehat{G}_{n_2}$, respectively.

(ii) For matched pairs: Given the raw sample $\{(X_i, Y_i)\}_{i=1}^{n_1}$, the bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=1}^{n_1}$ is an i.i.d. sample drawn from the empirical distribution $\widehat{H}_{n_1}$.

Define

$$\widehat{F}^*_{n_1}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{1}_{(-\infty,x]}\left(X_i^*\right) \text{ and } \widehat{G}^*_{n_2}(x) = \frac{1}{n_2}\sum_{i=1}^{n_2} \mathbb{1}_{(-\infty,x]}\left(Y_i^*\right)$$

for every $x \in \mathbb{R}$. Let $\widehat{\phi}^*_n(x,\theta) = \widehat{F}^*_{n_1}(x) - \widehat{G}^*_{n_2}\left(g(x,\theta)\right)$ for every $(x,\theta) \in \mathbb{R}\times\Theta$.

Lemma 2.5: If Assumptions 2.5 and 2.6 hold, then

$$\sup_{\Gamma\in \mathrm{BL}_1(\ell^\infty(\mathbb{R}\times\Theta))} \left|\mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right)\middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0\right)\right]\right| \xrightarrow{\mathbb{P}} 0,$$

and $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ is asymptotically measurable as $n \to \infty$.

With a consistent estimator $\widehat{\mathcal{L}}''_n$ for $\mathcal{L}''_\phi$ and a suitable bootstrap approximation $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ for $\mathbb{G}_0$ at hand, we can naturally approximate the distribution of $\mathcal{L}''_\phi\left(\mathbb{G}_0\right)$ by the conditional distribution of the bootstrap test statistic $\widehat{\mathcal{L}}''_n\{\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)\}$ given the raw samples.

Proposition 2.3: If Assumptions 2.1–2.9 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then

$$\sup_{\Gamma\in \mathrm{BL}_1(\mathbb{R})} \left|\mathbb{E}\left[\Gamma\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right]\right)\middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right] - \mathbb{E}\left[\Gamma\left(\mathcal{L}''_\phi\left(\mathbb{G}_0\right)\right)\right]\right| \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$.

2.3 Asymptotic Properties

Now we construct the test for the null hypothesis H_0. For a given level of significance $\alpha \in (0,1)$, define the bootstrap critical value

$$\widehat{c}_{1-\alpha,n} = \inf\left\{c \in \mathbb{R} : \mathbb{P}\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right] \le c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \ge 1-\alpha\right\}.$$

In practice, we approximate $\widehat{c}_{1-\alpha,n}$ by computing the $1-\alpha$ quantile of the n_B independently generated bootstrap test statistics, with n_B chosen as large as is computationally convenient. We reject H_0 if and only if $T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}$. The following theorem shows that the proposed test is asymptotically size controlled and consistent.

Theorem 2.1: Suppose that Assumptions 2.1–2.9 hold.

(i) If H_0 is true and the CDF of $\mathcal{L}''_\phi\left(\mathbb{G}_0\right)$ is strictly increasing and continuous at its $1-\alpha$ quantile, then

$$\lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}\right) = \alpha.$$

(ii) If H_0 is false, then

$$\lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}\right) = 1.$$

2.3.1 Local Power

In this section, we consider the local power of the test following the discussion in Chen and Fang (2019b). We first consider the independent samples case. For all n_1 and n_2, let $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ be distributed according to probability distributions P_{1n} and P_{2n}, respectively. That is, $P_{1n}(B) = \mathbb{P}(X_i \in B)$ and $P_{2n}(B) = \mathbb{P}(Y_i \in B)$ for all $B \in \mathscr{B}_\mathbb{R}$. We suppose that H_0 is false for each $\{P_{1n}, P_{2n}\}$, that is, for all $\theta \in \Theta$, $P_{1n}((-\infty,x]) \neq P_{2n}((-\infty, g(x,\theta)])$ for some $x \in \mathbb{R}$. Suppose that P_{jn} converges (in a way as described in the following assumption) to some

probability measure P_j with $j \in \{1,2\}$, and that $\{P_1, P_2\}$ satisfies H_0, that is, for some $\theta \in \Theta$, $P_1((-\infty, x]) = P_2((-\infty, g(x,\theta)])$ for all $x \in \mathbb{R}$. Under this setting, we set $F(x) = P_1((-\infty, x])$ and $G(x) = P_2((-\infty, x])$ for all $x \in \mathbb{R}$, and $\phi(x,\theta) = F(x) - G(g(x,\theta))$ for all $(x,\theta) \in \mathbb{R} \times \Theta$.

Assumption 2.10: The probability distributions P_{jn} and P_j with $j \in \{1,2\}$ satisfy that

$$\lim_{n\to\infty} \int \left(\sqrt{n_j} \left\{ \mathrm{d}P_{jn}^{1/2} - \mathrm{d}P_j^{1/2} \right\} - \frac{1}{2} v_{j0} \, \mathrm{d}P_j^{1/2} \right)^2 = 0 \tag{9}$$

for some measurable function v_{j0}, where $\mathrm{d}P_{jn}^{1/2}$ and $\mathrm{d}P_j^{1/2}$ denote the square roots of the densities of P_{jn} and P_j, respectively.

We next consider the matched pairs case. For each n, let $\{(X_i, Y_i)\}_{i=1}^{n_1}$ be distributed according to probability distribution P_n. That is, $P_n(B) = \mathbb{P}((X_i, Y_i) \in B)$ for all $B \in \mathscr{B}_{\mathbb{R}^2}$, where $\mathscr{B}_{\mathbb{R}^2}$ denotes the collection of Borel sets in $\mathbb{R}^2$. We suppose that H_0 is false for each P_n, that is, for all $\theta \in \Theta$, $P_n((-\infty, x] \times \mathbb{R}) \neq P_n(\mathbb{R} \times (-\infty, g(x,\theta)])$ for some $x \in \mathbb{R}$. Suppose that P_n converges (in a way as described in the following assumption) to some probability measure P, and that P satisfies H_0, that is, for some $\theta \in \Theta$, $P((-\infty, x] \times \mathbb{R}) = P(\mathbb{R} \times (-\infty, g(x,\theta)])$ for all $x \in \mathbb{R}$. Under this setting, we set $F(x) = P((-\infty, x] \times \mathbb{R})$ and $G(x) = P(\mathbb{R} \times (-\infty, x])$ for all $x \in \mathbb{R}$, and $\phi(x,\theta) = F(x) - G(g(x,\theta))$ for all $(x,\theta) \in \mathbb{R} \times \Theta$.

Assumption 2.11: The probability distributions P_n and P satisfy that

$$\lim_{n\to\infty} \int \left(\sqrt{n_1} \left\{ \mathrm{d}P_n^{1/2} - \mathrm{d}P^{1/2} \right\} - \frac{1}{2} v_0 \, \mathrm{d}P^{1/2} \right)^2 = 0 \tag{10}$$

for some measurable function v_0, where $\mathrm{d}P_n^{1/2}$ and $\mathrm{d}P^{1/2}$ denote the square roots of the densities of P_n and P, respectively.

Our local power results rely on Assumptions 2.10 and 2.11, which are similar to (3.10.10) in van der Vaart and Wellner (1996). For every probability measure Q and every measurable function h, we define

$$Qh = \int h \, \mathrm{d}Q. \tag{11}$$

The following proposition states formally the local power property of the test.

Proposition 2.4: Suppose that Assumptions 2.1–2.9 hold.

(i) For independent samples, if Assumption 2.10 holds, then $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0 + \psi_P$, where $\mathbb{G}_0$ is some tight random element, and

$$\psi_P(x,\theta) = \sqrt{1-\lambda} P_1 \mathbb{1}_{(-\infty,x]} v_{10} - \sqrt{\lambda} P_2 \mathbb{1}_{(-\infty, g(x,\theta)]} v_{20}$$

for all $(x,\theta) \in \mathbb{R} \times \Theta$. Let $c_{1-\alpha}$ denote the $1-\alpha$ quantile of $\mathcal{L}_\phi''(\mathbb{G}_0)$. Then it follows that

$$\liminf_{n\to\infty} \mathbb{P}\left(T_n \mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n} \right) \geq \mathbb{P}(\mathcal{L}_\phi''(\mathbb{G}_0 + \psi_P) > c_{1-\alpha}).$$

(ii) For matched pairs, if Assumption 2.11 holds, then $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0 + \psi_P$, where $\mathbb{G}_0$ is some tight random element, and

$$\psi_P(x,\theta) = \sqrt{1/2} \cdot P(\mathbb{1}_{(-\infty,x]\times\mathbb{R}} - \mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]}) v_0$$

for all $(x, \theta) \in \mathbb{R} \times \Theta$. Let $c_{1-\alpha}$ denote the $1-\alpha$ quantile of $\mathcal{L}''_\phi(\mathbb{G}_0)$. Then it follows that

$$\liminf_{n\to\infty} \mathbb{P}\left(T_n \mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}\right) \geq \mathbb{P}(\mathcal{L}''_\phi(\mathbb{G}_0 + \psi_P) > c_{1-\alpha}).$$

Proposition 2.4 follows the constructions in Theorem A.1 of Chen and Fang (2019b) and provides lower bounds for the power of the test under local perturbations to the null.

3 Transformations on Multiple CDFs

As mentioned in the introduction, the problem of comparing multiple distributions has attracted much attention since the 1950s and has continued to be an important research topic. In this section, we consider testing general parametric transformations on multiple CDFs and generalize the results in Section 2 to multiple samples. Towards this end, let $F, G_1, \ldots, G_K$ for some $K \geq 2$ be unknown continuous CDFs on $\mathbb{R}$. Let $\Theta_k \subset \mathbb{R}^{d_{\theta_k}}$ for every $k \in \{1, \ldots, K\}$ with $d_{\theta_k} \in \mathbb{Z}_+$. Let $\Theta = \Theta_1 \times \cdots \times \Theta_K$ equipped with a norm $\|\cdot\|_{K2}$ such that for every $(\theta_1, \ldots, \theta_K) \in \Theta$,

$$\|(\theta_1, \ldots, \theta_K)\|_{K2} = \left(\sum_{k=1}^K \|\theta_k\|_2^2\right)^{1/2}.$$

For every $k \in \{1, \ldots, K\}$, let $g_k : \mathbb{R} \times \Theta_k \to \mathbb{R}$ be some prespecified function. The null hypothesis of interest is

$$\mathrm{H}_0: \text{ For some } (\theta_1, \ldots, \theta_K) \in \Theta, F(x) = G_1(g_1(x, \theta_1)) = \cdots = G_K(g_K(x, \theta_K)) \text{ for all } x \in \mathbb{R}. \tag{12}$$

The parameter $(\theta_1, \ldots, \theta_K)$ in (12) is the nuisance parameter we need to take into account in the test.

Example 3.1: (Location-scale Transformations on Multiple CDFs) For every $k \in \{1, \ldots, K\}$, suppose that Y_k is equivalent to $(X - \theta_{k1})/\theta_{k2}$ in distribution for some $\theta_{k1} \in \mathbb{R}$ and $\theta_{k2} \in \mathbb{R}_+$. Then the CDFs of X and Y_k satisfy

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - \theta_{k1}}{\theta_{k2}} \leq \frac{x - \theta_{k1}}{\theta_{k2}}\right) = \mathbb{P}\left(Y_k \leq \frac{x - \theta_{k1}}{\theta_{k2}}\right) = G_k\left(\frac{x - \theta_{k1}}{\theta_{k2}}\right).$$

For every $k \in \{1, \ldots, K\}$, let $\Theta_k = [a_{k1}, b_{k1}] \times [a_{k2}, b_{k2}]$, where $-\infty < a_{k1} < b_{k1} < \infty$ and $0 < a_{k2} < b_{k2} < \infty$. Let $\Theta = \Theta_1 \times \cdots \times \Theta_K$. In this case, for every $k \in \{1, \ldots, K\}$, the parameter $\theta_k = (\theta_{k1}, \theta_{k2}) \in \Theta_k$, and the function

$$g_k(x, \theta_k) = \frac{x - \theta_{k1}}{\theta_{k2}} \text{ for all } x \in \mathbb{R} \text{ and all } \theta_k \in \Theta_k.$$

Let ν be a probability measure on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$. We now introduce the following assumptions for the transformations on multiple CDFs.

Assumption 3.1: For every $k \in \{1, \ldots, K\}$ and every $\theta_k \in \Theta_k$, the function $x \mapsto g_k(x, \theta_k)$ is continuous and increasing.

Assumption 3.2: The probability measure ν on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ satisfies $\mu \ll \nu$, that is, if $\nu(B) = 0$ for some $B \in \mathscr{B}_\mathbb{R}$, then $\mu(B) = 0$.

Assumption 3.3: The set Θ_k is compact in $\mathbb{R}^{d_{\theta_k}}$ for every $k \in \{1, \ldots, K\}$.

Assumption 3.4: For every $f \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$ and every k, the map $\theta_k \mapsto f(g_k(\cdot, \theta_k))$, from Θ_k to $L^2(\nu)$, is continuous. That is, for an arbitrary fixed $\theta_{k0} \in \Theta_k$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_{\mathbb{R}} \left[f\left(g_k(x, \theta_k)\right) - f\left(g_k(x, \theta_{k0})\right) \right]^2 \, \mathrm{d}\nu(x) < \varepsilon$$

for all $\theta_k \in \Theta_k$ with $\|\theta_k - \theta_{k0}\|_2 < \delta$.

Assumptions 3.1–3.4 are generalizations of Assumptions 2.1–2.4 in Section 2 for transformations on multiple CDFs. For every $k \in \{1, \ldots, K\}$, define a function space

$$\mathbb{D}_{\mathcal{L}k} = \left\{ \varphi_k \in \ell^\infty(\mathbb{R} \times \Theta_k) : \theta_k \mapsto \varphi_k(\cdot, \theta_k), \text{ as a map from } \Theta_k \text{ to } L^2(\nu), \text{ is continuous} \right\}.$$

Then we define $\mathbb{D}_{\mathcal{L}0} = \prod_{k=1}^K \mathbb{D}_{\mathcal{L}k}$. For every $k \in \{1, \ldots, K\}$ and every $f : \mathbb{R} \to \mathbb{R}$, we define a map $f \circ g_k : \mathbb{R} \times \Theta_k \to \mathbb{R}$ such that $f \circ g_k(x, \theta_k) = f(g_k(x, \theta_k))$ for every $(x, \theta_k) \in \mathbb{R} \times \Theta_k$. Define a map $\phi_k : \mathbb{R} \times \Theta_k \to \mathbb{R}$ for every k such that $\phi_k(x, \theta_k) = F(x) - G_k\left(g_k(x, \theta_k)\right)$ for every $(x, \theta_k) \in \mathbb{R} \times \Theta_k$. Define $\phi : \mathbb{R} \times \Theta \to \mathbb{R}^K$ such that $\phi(x, \theta) = (\phi_1(x, \theta_1), \ldots, \phi_K(x, \theta_K))$ for every $(x, \theta) \in \mathbb{R} \times \Theta$, where $\theta = (\theta_1, \ldots, \theta_K)$ and $\theta_k \in \Theta_k$ for every k. The proposition below provides an equivalent characterization of the null hypothesis in (12).

Proposition 3.1: If Assumptions 3.1–3.4 hold, then the null hypothesis in (12) is equivalent to

$$\mathrm{H}_0 : \inf_{(\theta_1, \ldots, \theta_K) \in \Theta} \int_{\mathbb{R}} \sum_{k=1}^K \left[F(x) - G_k\left(g_k(x, \theta_k)\right) \right]^2 \, \mathrm{d}\nu(x) = 0. \tag{13}$$

3.1 Test Statistic

Suppose that $\{X_i\}_{i=1}^{n_x}$ is a random sample drawn from F, and $\{Y_{ki}\}_{i=1}^{n_k}$ is a random sample drawn from G_k for every $k \in \{1, \ldots, K\}$.

Assumption 3.5: Each of the samples $\{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}$ is independently and identically distributed, and the samples $\{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}$ are jointly independent.

Assumption 3.6: The ratios $n_x/n \to \lambda_x \in (0, 1)$ and $n_k/n \to \lambda_k \in (0, 1)$ as $n \to \infty$ for every k, where $n = n_x + n_1 + \cdots + n_K$.

Assumption 3.5 requires the multiple samples to be jointly independent. In Assumption 3.6, n_x and n_k are viewed as functions of n. As $n \to \infty$, $n_x \to \infty$ and $n_k \to \infty$ for every k.

Define a function space

$$\mathbb{D}_{\mathcal{L}} = \left\{ (\varphi_1, \ldots, \varphi_K) \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k) : \int_{\mathbb{R}} \sum_{k=1}^K \left[\varphi_k(x, \theta_k) \right]^2 \, \mathrm{d}\nu(x) < \infty \text{ for all } (\theta_1, \ldots, \theta_K) \in \Theta \right\}.$$

Define a map $\mathcal{L}$ on $\mathbb{D}_{\mathcal{L}}$ such that $\mathcal{L}(\varphi) = \inf_{\theta \in \Theta} \int_{\mathbb{R}} \sum_{k=1}^K \left[\varphi_k(x, \theta_k) \right]^2 \, \mathrm{d}\nu(x)$ for every $\varphi \in \mathbb{D}_{\mathcal{L}}$ with $\varphi = (\varphi_1, \ldots, \varphi_K)$ and $\theta = (\theta_1, \ldots, \theta_K)$. Then under Assumptions 3.1–3.4, the null and the alternative hypotheses can be expressed as

$$\mathrm{H}_0 : \mathcal{L}(\phi) = 0 \text{ and } \mathrm{H}_1 : \mathcal{L}(\phi) > 0.$$

The CDFs F and G_k can be estimated by the empirical distribution functions such that for every $x \in \mathbb{R}$ and every k,

$$\widehat{F}_{n_x}(x) = \frac{1}{n_x}\sum_{i=1}^{n_x} \mathbb{1}_{(-\infty,x]}(X_i) \text{ and } \widehat{G}_{n_k}(x) = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbb{1}_{(-\infty,x]}(Y_{ki}).$$

For every $x \in \mathbb{R}$ and every $\theta \in \Theta$ with $\theta = (\theta_1, \ldots, \theta_K)$, let

$$\widehat{\phi}_{nk}(x,\theta_k) = \widehat{F}_{n_x}(x) - \widehat{G}_{n_k}(g_k(x,\theta_k)) \text{ and } \widehat{\phi}_n(x,\theta) = (\widehat{\phi}_{n1}(x,\theta_1), \ldots, \widehat{\phi}_{nK}(x,\theta_K)),$$

and set the test statistic to be $T_n\mathcal{L}(\widehat{\phi}_n)$, where $T_n = n_x \cdot \prod_{k=1}^K (n_k/n)$.

Lemma 3.1: Under Assumptions 3.5 and 3.6, we have

$$\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0 \text{ in } \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$$

as $n \to \infty$, where $\mathbb{G}_0$ is a tight random element. If, in addition, Assumption 3.4 holds, then $\mathbb{P}(\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}) = 1$.

Next, we show that the map $\mathcal{L}$ is Hadamard directionally differentiable, but its Hadamard directional derivative is also degenerate under H_0. Define $\mathbb{D}_0 = \{\varphi \in \mathbb{D}_\mathcal{L} : \mathcal{L}(\varphi) = 0\}$.

Lemma 3.2: If Assumptions 3.3 and 3.4 hold, then $\mathcal{L}$ is Hadamard directionally differentiable at $\phi \in \mathbb{D}_\mathcal{L}$ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the Hadamard directional derivative

$$\mathcal{L}'_\phi(h) = 2 \inf_{\theta \in \Theta_0(\phi)} \int_\mathbb{R} \sum_{k=1}^K \phi_k(x,\theta_k) h_k(x,\theta_k)\, \mathrm{d}\nu(x) \text{ for all } h \in \mathbb{D}_{\mathcal{L}0} \text{ with } h = (h_1, \ldots, h_K),$$

where $\Theta_0(\phi) = \arg\min_{\theta\in\Theta} \int_\mathbb{R} \sum_{k=1}^K [\phi_k(x,\theta_k)]^2 \, \mathrm{d}\nu(x)$. Moreover, if $\phi \in \mathbb{D}_0$, then the derivative $\mathcal{L}'_\phi$ is well defined on the whole of $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ with $\mathcal{L}'_\phi(h) = 0$ for every $h \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$.

We now provide high level conditions for the existence of the second order Hadamard directional derivative of $\mathcal{L}$.

Assumption 3.7: For every $k \in \{1, \ldots, K\}$, the function $G_k \circ g_k$ is twice differentiable with respect to θ_k, and the second partial derivative satisfies

$$\int_\mathbb{R} \sup_{\theta_k \in \Theta_k} \left\| \frac{\partial^2 (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k \partial \vartheta_k^\mathsf{T}} \bigg|_{(z,\vartheta_k)=(x,\theta_k)} \right\|_2^2 \mathrm{d}\nu(x) < \infty. \tag{14}$$

Assumption 3.8: The set $\Theta_0 \equiv \{\theta \in \Theta : \int_\mathbb{R} \sum_{k=1}^K [\phi_k(x,\theta_k)]^2 \, \mathrm{d}\nu(x) = 0\} \subset \mathrm{int}(\Theta)$, and there exist some $\kappa \in (0,1]$ and some $C > 0$ such that for all small $\varepsilon > 0$,

$$\inf_{\theta \in \Theta \setminus \Theta_0^\varepsilon} \left\{ \int_\mathbb{R} \sum_{k=1}^K [\phi_k(x,\theta_k)]^2 \, \mathrm{d}\nu(x) \right\}^{1/2} \geq C\varepsilon^\kappa.$$

Assumptions 3.7–3.8 are generalized versions of Assumptions 2.7–2.8 for the transformations on multiple samples. We denote $\prod_{k=1}^K L^2(\nu)$ by $L^2_K(\nu)$. Define a norm $\|\cdot\|_{L^2_K(\nu)}$ on $L^2_K(\nu)$ such that for every $\psi \in L^2_K(\nu)$ with $\psi = (\psi_1, \ldots, \psi_K)$,

$$\|\psi\|_{L^2_K(\nu)} = \left\{ \sum_{k=1}^K \|\psi_k\|^2_{L^2(\nu)} \right\}^{1/2} = \|(\|\psi_1\|_{L^2(\nu)}, \ldots, \|\psi_K\|_{L^2(\nu)})\|_2.$$

For every $\theta \in \Theta$ with $\theta = (\theta_1, \ldots, \theta_K)$, define $\Phi_k'(\theta_k) : \mathbb{R} \to \mathbb{R}^{d_{\theta_k}}$ such that

$$\Phi_k'(\theta_k)(x) = -\left.\frac{\partial (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k}\right|_{(z,\vartheta_k)=(x,\theta_k)} \quad \text{for every } x \in \mathbb{R}.$$

Let $\Phi'(\theta, v) = (\Phi_1'(\theta_1)^\mathsf{T} v_1, \ldots, \Phi_K'(\theta_K)^\mathsf{T} v_K)$ for every $\theta = (\theta_1, \ldots, \theta_K) \in \Theta$ and every $v = (v_1, \ldots, v_K) \in \prod_{k=1}^K \mathbb{R}^{d_{\theta_k}}$.

Lemma 3.3: If Assumptions 3.3, 3.4, 3.7, and 3.8 hold and $\phi \in \mathbb{D}_0$, then the function $\mathcal{L}$ is second order Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the second order Hadamard directional derivative

$$\mathcal{L}_\phi''(h) = \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V(a(h))} \left\| \Phi'(\theta, v) + \mathscr{H}(\theta) \right\|_{L_K^2(\nu)}^2 \text{ for all } h \in \mathbb{D}_{\mathcal{L}0} \text{ with } h = (h_1, \ldots, h_K),$$

where $V(a) = \{v \in \prod_{k=1}^K \mathbb{R}^{d_{\theta_k}} : \|v\|_{K2} \leq a\}$ for all $a > 0$, the number $a(h) > 0$ satisfies that $Ca(h)^\kappa = 3\{\sum_{k=1}^K \|h_k\|_\infty^2\}^{1/2}$ with C and κ defined as in Assumption 3.8, and $\mathscr{H}(\theta)(x) = (h_1(x, \theta_1), \ldots, h_K(x, \theta_K))$ for every $(x, \theta) \in \mathbb{R} \times \Theta$ with $\theta = (\theta_1, \ldots, \theta_K)$.

With Lemma 3.3, the asymptotic distribution of the test statistic $\mathcal{L}(\widehat{\phi}_n)$ under the null is obtained by applying the second order delta method.

Proposition 3.2: If Assumptions 3.1–3.8 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then

$$T_n \mathcal{L}(\widehat{\phi}_n) \rightsquigarrow \mathcal{L}_\phi''(\mathbb{G}_0) \text{ as } n \to \infty.$$

3.2 The Bootstrap

We use the numerical second order Hadamard directional derivative $\widehat{\mathcal{L}}_n''$ proposed by Hong and Li (2018) and Chen and Fang (2019b) to approximate $\mathcal{L}_\phi''$, which is defined as

$$\widehat{\mathcal{L}}_n''(h) = \frac{\mathcal{L}(\widehat{\phi}_n + \tau_n h) - \mathcal{L}(\widehat{\phi}_n)}{\tau_n^2}$$

for all $h \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$, where $\{\tau_n\}$ is a sequence of tuning parameters satisfying the assumption below.

Assumption 3.9: $\{\tau_n\} \subset \mathbb{R}_+$ is a sequence of scalars such that $\tau_n \downarrow 0$ and $\tau_n \sqrt{T_n} \to \infty$ as $n \to \infty$.

The next lemma establishes the consistency of $\widehat{\mathcal{L}}_n''$.

Lemma 3.4: If Assumptions 3.1–3.9 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then for every sequence $\{h_n\} \subset \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ and every $h \in \mathbb{D}_{\mathcal{L}0}$ such that $h_n \to h$ in $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ as $n \to \infty$, we have

$$\widehat{\mathcal{L}}_n''(h_n) \xrightarrow{\mathbb{P}} \mathcal{L}_\phi''(h) \text{ as } n \to \infty.$$

We approximate the distribution of $\mathbb{G}_0$ via bootstrap. Given the raw samples $\{\{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}\}$, let the bootstrap samples $\{\{X_i^*\}_{i=1}^{n_x}, \{Y_{1i}^*\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}^*\}_{i=1}^{n_K}\}$ be jointly independent, and drawn independently and identically from the empirical distribu-

tions $\widehat{F}_{n_x}, \widehat{G}_{n_1}, \ldots, \widehat{G}_{n_K}$, respectively. Define for every $x \in \mathbb{R}$ and every k,

$$\widehat{F}^*_{n_x}(x) = \frac{1}{n_x}\sum_{i=1}^{n_x} \mathbb{1}_{(-\infty,x]}\left(X^*_i\right) \text{ and } \widehat{G}^*_{n_k}(x) = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbb{1}_{(-\infty,x]}\left(Y^*_{ki}\right).$$

For every k, let $\widehat{\phi}^*_{nk}(x,\theta_k) = \widehat{F}^*_{n_x}(x) - \widehat{G}^*_{n_k}\left(g_k(x,\theta_k)\right)$ for every $x \in \mathbb{R}$ and every $\theta_k \in \Theta_k$. Let $\widehat{\phi}^*_n = (\widehat{\phi}^*_{n1}, \ldots, \widehat{\phi}^*_{nK})$.

Lemma 3.5: If Assumptions 3.5 and 3.6 hold, then

$$\sup_{\Gamma \in \mathrm{BL}_1\left(\prod_{k=1}^K \ell^\infty(\mathbb{R}\times\Theta_k)\right)} \left|\mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right)\middle| \{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}\right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0\right)\right]\right|$$

$\xrightarrow{\mathbb{P}} 0$, and $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ is asymptotically measurable as $n \to \infty$.

The distribution of $\mathcal{L}''_\phi(\mathbb{G}_0)$ can be approximated by the conditional distribution of the bootstrap test statistic $\widehat{\mathcal{L}}''_n\{\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)\}$ given the raw samples.

Proposition 3.3: If Assumptions 3.1–3.9 hold and H_0 is true ($\phi \in \mathbb{D}_0$), then

$$\sup_{\Gamma \in \mathrm{BL}_1(\mathbb{R})} \left|\mathbb{E}\left[\Gamma\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right]\right)\middle| \{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}\right] - \mathbb{E}\left[\Gamma\left(\mathcal{L}''_\phi(\mathbb{G}_0)\right)\right]\right|$$

$\xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

3.3 Asymptotic Properties

For a given level of significance $\alpha \in (0,1)$, define the bootstrap critical value

$$\widehat{c}_{1-\alpha,n} = \inf\left\{c \in \mathbb{R} : \mathbb{P}\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right] \le c \middle| \{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}\right) \ge 1-\alpha\right\}.$$

We reject H_0 if and only if $T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}$. The next theorem shows that the proposed test is asymptotically size controlled and consistent.

Theorem 3.1: Suppose that Assumptions 3.1–3.9 hold.

(i) If H_0 is true and the CDF of $\mathcal{L}''_\phi(\mathbb{G}_0)$ is strictly increasing and continuous at its $1-\alpha$ quantile, then

$$\lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}\right) = \alpha.$$

(ii) If H_0 is false, then

$$\lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}\right) = 1.$$

The local power results for comparisons of multiple CDFs can be obtained analogously under settings similar to those in Section 2.3.1.

4 Simulation Studies

In this section, we conduct Monte Carlo experiments to investigate the finite sample properties of the proposed test, whose critical values are obtained through a numerical bootstrap. The significance level is set to be $\alpha = 0.05$. We consider sample sizes $n_1, n_2 \in \{500, 1000, 1500\}$. In line with the previous notation, the two samples to be tested are denoted by $\{X_i\}_{i=1}^{n_1}$ and

$\{Y_i\}_{i=1}^{n_2}$, respectively. Define the following covariance matrix

$$\Sigma_3 = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}.$$

We choose the tuning parameter τ_n from the set $\{0.01, 0.02, \ldots, 0.25\}$. The warp-speed method of Giacomini et al. (2013) is employed with 1,000 Monte Carlo iterations and 1,000 bootstrap iterations in each simulation.

4.1 Parametric Transformation for Continuous Random Variables

In the first experiment, we consider continuous random variables and the location-scale transformation

$$\left\{ g(x, \theta) = \frac{x - \theta_1}{\theta_2} : -0.2 \leq \theta_1 \leq 0.2,\ 2^{-0.2} \leq \theta_2 \leq 2^{0.2} \right\}. \tag{15}$$

The probability measure ν is set to be a normal distribution with mean 0 and standard deviation $5/3$. We let $X_i \sim \mathcal{N}(0,1)$, $Z_i \sim \mathcal{N}(0,1)$, and $U_i \sim \text{Unif}[-3,3]$. For independent samples, $\{X_i\}_{i=1}^{n_1}$, $\{Z_i\}_{i=1}^{n_2}$, and $\{U_i\}_{i=1}^{n_2}$ are jointly independent; and for matched pairs, the dependence structure of (X_i, Z_i, U_i) is characterized by a Gaussian copula with the covariance matrix Σ_3. There are four data generating processes of $\{Y_i\}_{i=1}^{n_2}$:

- DGP (0) (null): $Y_i = Z_i$.
- DGP (1) (alternative): $Y_i = 0.5Z_i + 0.5U_i$.
- DGP (2) (alternative): $Y_i = 0.25Z_i + 0.75U_i$.
- DGP (3) (alternative): $Y_i = U_i$.

DGP (0) satisfies the null hypothesis, and DGPs (1)–(3) satisfy the alternative hypothesis with their "distances" to the null arranged in ascending order.

The empirical sizes and powers are displayed in Tables 2a and 2b for independent samples and matched pairs, respectively. The results show that our test has good empirical sizes and is relatively insensitive to the choice of τ_n, especially when τ_n falls in the range of $[0.07, 0.1]$. The empirical power results show that our test is powerful, and the power increases as the "distance" to the null increases or as the sample size increases.

4.2 Parametric Transformation for Discrete Random Variables

Although the cases of discrete random variables are excluded from our theoretical analysis, we use simulation evidence to show that the proposed test also works well for CDFs of discrete random variables. In the second experiment, we consider discrete random variables and the location-scale transformation $g(x, \theta)$ defined in (15). The probability measure ν is set to be a normal distribution with mean 5 and standard deviation 5. We let $X_i \sim \text{Unif}\{1, 2, \ldots, 10\}$, $U_i \sim \text{Unif}\{1, 2, \ldots, 10\}$, and $V_i \sim \text{Unif}\{1, 2, \ldots, 10\}$. For independent samples, $\{X_i\}_{i=1}^{n_1}$, $\{U_i\}_{i=1}^{n_2}$, and $\{V_i\}_{i=1}^{n_2}$ are jointly independent; and for matched pairs, the dependence structure of (X_i, U_i, V_i) is characterized by a Gaussian copula with the covariance matrix Σ_3. There are four data generating processes of $\{Y_i\}_{i=1}^{n_2}$:

- DGP (0) (null): $Y_i = U_i$.
- DGP (1) (alternative): $Y_i = 0.9U_i + 0.1V_i$.
- DGP (2) (alternative): $Y_i = 0.75U_i + 0.25V_i$.
- DGP (3) (alternative): $Y_i = 0.5U_i + 0.5V_i$.

DGP (0) satisfies the null hypothesis, and DGPs (1)–(3) satisfy the alternative hypothesis with their "distances" to the null arranged in ascending order.

The results shown in Table 1a suggest that the empirical sizes are reasonably close to the nominal size when τ_n falls in the range of $[0.05, 0.06]$ for independent samples. The results shown in Table 1b suggest that for matched pairs, the test has good empirical sizes when τ_n falls in the range of $[0.07, 0.09]$. Overall, the power performance of our test is satisfactory; as the "distance" to the null increases or as the sample size increases, the empirical power increases.

5 Empirical Application

In this section, we revisit the empirical application of Bera et al. (2013). To find evidence of any existence of the adverse selection death spiral, Bera et al. (2013) compare the state of New York, where legislation for enforcing "community rating" (premium fixed by community and not by risk category) was enacted in 1993, with Pennsylvania, where no such legislation was enacted. The authors test for differences between the age distributions of adult individuals who were covered by group insurance policies sponsored by employers with 100 or fewer employees before and after 1993. The data are from the 1987–1996 March Current Population Survey. For New York, there are 4,548 observations before 1993, and 2,517 observations after 1993. For Pennsylvania, there are 3,113 observations before 1993, and 1,875 observations after 1993. Bera et al. (2013) apply their smooth test for equality of distributions to these age data sets and conclude that the population age distributions before and after 1993 in both states are different irrespective of the community rating legislation. They find that the sources of the differences are mainly through the first (location) and second (scale) order moments, while the contributions of the third (skewness) and fourth (kurtosis) order moments are very small.

Given Bera et al. (2013)'s findings, an interesting question to ask is whether the age distributions are the same if the location and scale differences have been taken into account. In the following, we use the data sets of Bera et al. (2013) and our unified test to re-examine the differences between the age distributions. Instead of testing the equality of the distributions, we consider the null hypothesis that the age distribution after 1993 is a location, scale, or location-scale transformation of that before 1993.

For testing the location transformation, we set

$$g(x, \theta) = x - \theta \text{ with } \theta \in \Theta = [-2, 0.5].$$

For testing the scale transformation, we set

$$g(x, \theta) = x/\theta \text{ with } \theta \in \Theta = [0.5, 2].$$

For testing the location-scale transformation, we set

$$g(x, \theta) = (x - \theta_1)/\theta_2 \text{ with } \theta = (\theta_1, \theta_2) \in \Theta = [-2, 0.5] \times [0.5, 2].$$

Let $\underline{M}$ be a number slightly less than $\min_{1\leq i\leq n_1} X_i$, and $\overline{M}$ be a number slightly greater than $\max_{1\leq i\leq n_1} X_i$. We then suggest setting the measure ν as a normal distribution with mean $(\overline{M}+\underline{M})/2$ and standard deviation $(\overline{M}-\underline{M})/6$, which is $\mathcal{N}(41, 7.67^2)$ in this empirical application. We note that this choice of ν guarantees that its integration over the range of the observations of X is approximately 1.

To select the tuning parameter τ_n, we perform the following sample-based Monte Carlo simulations. Let $\mathbb{G}_{\mathrm{NY},1}$ and $\mathbb{G}_{\mathrm{NY},2}$ be the empirical age distributions before and after 1993 in New York, respectively. Let $\mathbb{G}_{\mathrm{PA},1}$ and $\mathbb{G}_{\mathrm{PA},2}$ be the empirical age distributions before and after 1993 in Pennsylvania, respectively. Let $n_{\mathrm{NY},1}$ and $n_{\mathrm{NY},2}$ be the sample sizes before and after 1993 in New York, $n_{\mathrm{PA},1}$ and $n_{\mathrm{PA},2}$ be the sample sizes before and after 1993 in Pennsylvania. For every transformation we consider, every state $j \in \{\mathrm{NY}, \mathrm{PA}\}$, and every $\ell \in \{1, 2\}$, we independently draw samples $\{X_i\}_{i=1}^{n_{j,1}}$ and $\{Y_i\}_{i=1}^{n_{j,2}}$ from $\mathbb{G}_{j,\ell}$ and perform the test using a set of tuning parameters. The significance level is set to be $\alpha = 0.05$. We repeat the above procedure 1,000 times and compute the rejection rates. Then we pick the smallest tuning parameters that yield rejection rates closest to but no larger than α. For every transformation we consider and every state $j \in \{\mathrm{NY}, \mathrm{PA}\}$, we obtain two tuning parameters, which are shown in boldface in Table 3.

We then study the power of the proposed test under the selected tuning parameters. We use the same alternative DGPs and settings as those in Section 4.1. The sample sizes are set to be those of the age data sets. For New York, $n_1 = 4548$ and $n_2 = 2517$. For Pennsylvania, $n_1 = 3113$ and $n_2 = 1875$. Table 4 shows that the selected tuning parameters yield rejection rates close to 1 in most cases.

We next use the tuning parameters selected above to perform the test for location and/or scale transformations on age distributions. Moreover, we choose the number of bootstrap samples $n_B \in \{1000, 5000, 10000\}$ to assess the stability of the test across different n_B. Table 5 reports the bootstrap p-values of the proposed test. The results suggest that the age distributions before and after 1993 are indeed different for both states, even if we have allowed for the location-scale transformation. These results do not provide strong evidence that the legislation affects the age distributions. In addition, the test results in Table 5 are quite insensitive to the number of bootstrap samples n_B. The test results have interesting implications, besides that the population age distributions before and after 1993 in both states do not belong to the same location-scale family. As a matter of fact, when the location-scale effect has been controlled for, if the distributional differences are still significant, the null hypothesis of the same location-scale family is rejected most possibly due to the differences in higher order moments, for example, skewness or kurtosis.

6 Conclusion

This paper provides a simple unified framework for testing general parametric transformations on two (and multiple) CDFs in the presence of unknown nuisance parameters. The test

is motivated from a new characterization of the null that avoids the estimation of nuisance parameters. The test is shown to possess good asymptotic properties and perform well on finite samples. Finally, we apply the proposed test with the age data sets to demonstrate its application in practice. Extensions of our unified framework to testing other problems with nuisance parameters may deserve further study.

References

Alba, M., Barrera, D., and Jiménez, M. (2001). A homogeneity test based on empirical characteristic functions. *Computational Statistics*, 16(2):255–270.

Anderson, N. H., Hall, P., and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, 50(1):41–54.

Anderson, T. W. (1962). On the distribution of the two-sample Cramér–von Mises criterion. *The Annals of Mathematical Statistics*, 33(3):1148–1159.

Andrews, D. W. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68(2):399–405.

Beare, B. K. and Fang, Z. (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electronic Journal of Statistics*, 11(2):3841–3870.

Beare, B. K. and Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, 31(3):471–492.

Beare, B. K. and Shi, X. (2019). An improved bootstrap test of density ratio ordering. *Econometrics and Statistics*, 10:9–26.

Bera, A. K., Ghosh, A., and Xiao, Z. (2013). A smooth test for the equality of distributions. *Econometric Theory*, 29(2):419–446.

Chen, Q. and Fang, Z. (2019a). Improved inference on the rank of a matrix. *Quantitative Economics*, 10(4):1787–1824.

Chen, Q. and Fang, Z. (2019b). Inference on functionals under first order degeneracy. *Journal of Econometrics*, 210(2):459–481.

Chen, S. (2020). A new distribution-free k-sample test: Analysis of kernel density functionals. *Canadian Journal of Statistics*, 48(2):167–186.

Chen, S. and Pokojovy, M. (2018). Modern and classical k-sample omnibus tests. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(1):e1418.

Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284.

Chung, E. and Olivares, M. (2021). Permutation test for heterogeneous treatment effects with a nuisance parameter. *Journal of Econometrics*, 225(2):148–174.

Chung, E. and Olivares, M. (2022). Quantile-based test for heterogeneous treatment effects.

Cox, D. R. (1984). Interaction. *International Statistical Review/Revue Internationale de Statistique*, pages 1–24.

Darling, D. A. (1957). The Kolmogorov–Smirnov, Cramér–von Mises tests. *The Annals of Mathematical Statistics*, 28(4):823–838.

Ding, P., Feller, A., and Miratrix, L. (2016). Randomization inference for treatment effect variation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):655–671.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics*, 2(2):267–277.

Dümbgen, L. (1993). On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95(1):125–140.

Epps, T. and Singleton, K. J. (1986). An omnibus test for the two-sample problem using the empirical characteristic function. *Journal of Statistical Computation and Simulation*, 26(3-4):177–203.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412.

Fernández, V. A., Gamero, M. J., and Garcia, J. M. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics & Data Analysis*, 52(7):3730–3748.

Fisz, M. (1960). On a result by M. Rosenblatt concerning the von Mises–Smirnov test. *The Annals of Mathematical Statistics*, 31(2):427–429.

Giacomini, R., Politis, D. N., and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3):567–589.

Goldman, M. and Kaplan, D. M. (2018). Comparing distributions by multiple testing across quantiles or CDF values. *Journal of Econometrics*, 206(1):143–166.

Hall, P., Lombard, F., and Potgieter, C. J. (2013). A new approach to function-based hypothesis testing in location-scale families. *Technometrics*, 55(2):215–223.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240.

Henze, N., Klar, B., and Zhu, L.-X. (2005). Checking the adequacy of the multivariate semiparametric location shift model. *Journal of Multivariate Analysis*, 93(2):238–256.

Hirano, K. and Porter, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790.

Hong, H. and Li, J. (2018). The numerical delta method. *Journal of Econometrics*, 206(2):379–394.

Jiménez-Gamero, M.-D., Alba-Fernández, M., Jodrá, P., and Barranco-Chamorro, I. (2017). Fast tests for the two-sample problem based on the empirical characteristic function. *Mathematics and Computers in Simulation*, 137:390–410.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications*, 26(2):240–257.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramér–v. Mises tests. *The Annals of Mathematical Statistics*, 30(2):420–447.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematical Statistics*, 22(2):165–179.

Lehmann, E. L. and D'Abrera, H. J. (1975). *Nonparametrics: Statistical methods based on ranks*. Holden-day.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer.

Martínez-Camblor, P. and de Uña-Álvarez, J. (2009). Non-parametric k-sample tests: Density functions vs distribution functions. *Computational Statistics & Data Analysis*, 53(9):3344–3357.

Meintanis, S. G. (2005). Permutation tests for homogeneity based on the empirical characteristic function. *Journal of Nonparametric Statistics*, 17(5):583–592.

Melly, B. and Wüthrich, K. (2017). Local quantile treatment effects. *Handbook of Quantile Regression*, pages 145–164.

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *The American Economic Review*, 77(3):421–430.

Ramirez-Cuellar, J. (2021). Testing for idiosyncratic treatment effect heterogeneity. SSRN.3946092.

Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *The Annals of Mathematical Statistics*, 23(4):617–623.

Scholz, F. W. and Stephens, M. A. (1987). K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 82(399):918–924.

Seo, J. (2018). Tests of stochastic monotonicity with improved power. *Journal of Econometrics*, 207(1):53–70.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, 2(2):3–14.

Song, X. and Xiao, Z. (2022). On smooth tests for the equality of distributions. *Econometric Theory*, 38(1):194–208.

Sun, Z. (2021). Instrument validity for heterogeneous causal effects. arXiv:2009.01995.

Sun, Z. and Beare, B. K. (2021). Improved nonparametric bootstrap tests of Lorenz dominance. *Journal of Business & Economic Statistics*, 39(1):189–199.

Székely, G. J., Rizzo, M. L., et al. (2004). Testing for equal distributions in high dimension. *InterStat*, 5(16.10):1249–1272.

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

Zhang, J. and Wu, Y. (2007). k-Sample tests based on the likelihood ratio. *Computational Statistics & Data Analysis*, 51(9):4682–4691.

Table 1: Rejection Rates for Discrete Random Variables with Location-scale Transformation

DGP	n_1	n_2	τ_n									
			0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
	500	500	0.001	0.005	0.019	0.042	0.071	0.091	0.109	0.139	0.167	0.185
	500	1,000	0.004	0.012	0.034	0.045	0.071	0.091	0.113	0.130	0.158	0.186
	500	1,500	0.006	0.020	0.030	0.050	0.067	0.094	0.122	0.150	0.178	0.189
	1,000	500	0.001	0.015	0.028	0.042	0.066	0.093	0.122	0.144	0.165	0.176
(0)	1,000	1,000	0.010	0.019	0.025	0.044	0.057	0.076	0.107	0.135	0.156	0.173
	1,000	1,500	0.014	0.021	0.029	0.045	0.061	0.094	0.120	0.142	0.164	0.179
	1,500	500	0.003	0.012	0.022	0.031	0.051	0.077	0.107	0.124	0.144	0.154
	1,500	1,000	0.011	0.023	0.041	0.051	0.073	0.099	0.115	0.141	0.154	0.160
	1,500	1,500	0.021	0.031	0.041	0.055	0.066	0.087	0.113	0.139	0.178	0.198
	500	500	0.007	0.022	0.062	0.137	0.223	0.278	0.334	0.389	0.444	0.485
	500	1,000	0.015	0.063	0.144	0.209	0.304	0.368	0.447	0.549	0.613	0.674
	500	1,500	0.025	0.082	0.170	0.265	0.375	0.459	0.519	0.583	0.637	0.669
	1,000	500	0.012	0.052	0.159	0.253	0.334	0.408	0.473	0.545	0.594	0.631
(1)	1,000	1,000	0.055	0.179	0.352	0.498	0.585	0.656	0.712	0.755	0.792	0.820
	1,000	1,500	0.128	0.280	0.470	0.618	0.723	0.782	0.839	0.888	0.908	0.921
	1,500	500	0.028	0.093	0.178	0.261	0.344	0.405	0.486	0.571	0.602	0.657
	1,500	1,000	0.143	0.330	0.501	0.633	0.719	0.799	0.851	0.877	0.896	0.904
	1,500	1,500	0.329	0.525	0.715	0.824	0.881	0.915	0.939	0.955	0.959	0.965
	500	500	0.318	0.661	0.841	0.923	0.967	0.981	0.992	0.995	0.996	0.997
	500	1,000	0.679	0.937	0.984	0.990	0.995	0.997	0.998	0.999	1.000	1.000
	500	1,500	0.811	0.962	0.988	0.994	0.999	1.000	1.000	1.000	1.000	1.000
	1,000	500	0.752	0.961	0.991	0.996	1.000	1.000	1.000	1.000	1.000	1.000
(2)	1,000	1,000	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	1,500	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	500	0.933	0.989	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	500	0.860	0.984	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	500	1,000	0.991	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1,500	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	500	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

(a) Independent Samples

DGP	n_1	n_2	τ_n									
			0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
	500	500	0.002	0.003	0.008	0.012	0.018	0.031	0.051	0.067	0.080	0.089
(0)	1,000	1,000	0.003	0.006	0.009	0.011	0.025	0.034	0.050	0.060	0.070	0.080
	1,500	1,500	0.011	0.019	0.027	0.037	0.042	0.049	0.063	0.074	0.094	0.112
	500	500	0.009	0.037	0.100	0.170	0.236	0.315	0.385	0.446	0.494	0.533
(1)	1,000	1,000	0.140	0.298	0.462	0.626	0.737	0.795	0.835	0.870	0.905	0.912
	1,500	1,500	0.497	0.737	0.890	0.934	0.963	0.980	0.986	0.992	0.994	0.996
	500	500	0.654	0.925	0.982	0.995	1.000	1.000	1.000	1.000	1.000	1.000
(2)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	500	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

(b) Matched Pairs

Table 2: Rejection Rates for Continuous Random Variables with Location-scale Transformation

DGP	n_1	n_2	τ_n									
			0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15
	500	500	0.006	0.011	0.011	0.014	0.011	0.011	0.011	0.006	0.004	0.002
	500	1,000	0.011	0.014	0.015	0.015	0.012	0.011	0.011	0.009	0.006	0.005
	500	1,500	0.020	0.025	0.027	0.026	0.020	0.019	0.011	0.009	0.009	0.009
	1,000	500	0.012	0.015	0.016	0.016	0.016	0.015	0.013	0.010	0.009	0.005
(0)	1,000	1,000	0.025	0.028	0.028	0.028	0.022	0.015	0.013	0.010	0.006	0.005
	1,000	1,500	0.023	0.027	0.027	0.027	0.028	0.026	0.021	0.016	0.013	0.010
	1,500	500	0.015	0.015	0.015	0.015	0.014	0.013	0.012	0.012	0.009	0.006
	1,500	1,000	0.026	0.032	0.034	0.034	0.031	0.026	0.025	0.021	0.020	0.010
	1,500	1,500	0.040	0.046	0.048	0.045	0.045	0.036	0.026	0.023	0.022	0.016
	500	500	0.041	0.057	0.063	0.068	0.074	0.074	0.064	0.063	0.063	0.061
	500	1,000	0.125	0.153	0.186	0.198	0.209	0.219	0.201	0.196	0.193	0.158
	500	1,500	0.147	0.185	0.208	0.222	0.223	0.219	0.208	0.188	0.174	0.154
	1,000	500	0.117	0.139	0.172	0.172	0.153	0.155	0.152	0.138	0.139	0.129
(1)	1,000	1,000	0.318	0.356	0.379	0.407	0.429	0.429	0.442	0.434	0.411	0.369
	1,000	1,500	0.421	0.467	0.476	0.522	0.534	0.535	0.553	0.535	0.527	0.481
	1,500	500	0.133	0.177	0.200	0.232	0.240	0.233	0.235	0.220	0.187	0.144
	1,500	1,000	0.487	0.550	0.593	0.613	0.605	0.598	0.598	0.601	0.598	0.543
	1,500	1,500	0.638	0.667	0.689	0.691	0.688	0.702	0.663	0.642	0.619	0.591
	500	500	0.865	0.935	0.968	0.980	0.985	0.988	0.992	0.994	0.995	0.995
	500	1,000	0.992	0.996	0.996	0.997	0.998	0.998	0.999	0.999	0.999	0.999
	500	1,500	0.996	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	500	0.987	0.993	0.995	0.996	0.999	0.999	1.000	1.000	1.000	1.000
(2)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	500	0.998	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,000	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

(a) Independent Samples

DGP	n_1	n_2	τ_n									
			0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15
	500	500	0.016	0.018	0.022	0.027	0.027	0.029	0.031	0.030	0.031	0.031
(0)	1,000	1,000	0.014	0.017	0.021	0.023	0.024	0.024	0.023	0.021	0.023	0.023
	1,500	1,500	0.030	0.042	0.045	0.048	0.046	0.046	0.046	0.045	0.042	0.037
	500	500	0.067	0.089	0.120	0.142	0.153	0.159	0.182	0.193	0.200	0.202
(1)	1,000	1,000	0.458	0.516	0.582	0.596	0.628	0.662	0.656	0.693	0.709	0.718
	1,500	1,500	0.796	0.838	0.863	0.877	0.880	0.884	0.888	0.893	0.895	0.893
	500	500	0.953	0.984	0.995	0.997	0.998	0.999	0.999	0.999	0.999	0.999
(2)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(3)	1,000	1,000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1,500	1,500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

(b) Matched Pairs

Table 3: Tuning Parameters and Rejection Rates in Sample-based Simulations

State	New York				Pennsylvania			
ℓ	1		2		1		2	
Transformation	τ_n		τ_n		τ_n		τ_n	
	0.050	0.028	0.065	0.044	0.047	0.044	0.053	0.046
	0.051	0.028	0.066	0.044	0.048	0.045	0.054	0.046
	0.052	0.029	0.067	0.044	0.049	0.045	0.055	0.046
	0.053	0.030	0.068	0.046	0.050	0.047	0.056	0.046
	0.054	0.031	0.069	0.047	0.051	0.046	**0.057**	0.047
Location	0.055	0.031	0.070	0.047	**0.052**	0.048	0.058	0.047
	0.056	0.032	**0.071**	0.048	0.053	0.051	0.059	0.051
	0.057	0.032	0.072	0.048	0.054	0.052	0.060	0.051
	0.058	0.032	0.073	0.047	0.055	0.052	0.061	0.051
	0.059	0.032	0.074	0.047	0.056	0.052	0.062	0.052
	0.060	0.031	0.075	0.046	0.057	0.052	0.063	0.052
	0.019	0.012	0.018	0.004	0.025	0.017	0.020	0.006
	0.020	0.014	0.019	0.009	0.026	0.024	0.021	0.011
	0.021	0.019	0.020	0.014	0.027	0.031	0.022	0.015
	0.022	0.026	0.021	0.023	0.028	0.033	0.023	0.019
	0.023	0.031	0.022	0.030	0.029	0.038	0.024	0.025
Scale	**0.024**	0.044	**0.023**	0.042	**0.030**	0.047	0.025	0.039
	0.025	0.063	0.024	0.059	0.031	0.055	**0.026**	0.050
	0.026	0.095	0.025	0.086	0.032	0.068	0.027	0.053
	0.027	0.118	0.026	0.117	0.033	0.076	0.028	0.062
	0.028	0.150	0.027	0.128	0.034	0.084	0.029	0.086
	0.029	0.199	0.028	0.153	0.035	0.091	0.030	0.092
	0.016	0.005	0.015	0.003	0.022	0.009	0.020	0.008
	0.017	0.005	0.016	0.004	0.023	0.015	0.021	0.011
	0.018	0.009	0.017	0.012	0.024	0.018	0.022	0.014
	0.019	0.015	0.018	0.019	0.025	0.019	0.023	0.022
	0.020	0.029	0.019	0.030	0.026	0.023	0.024	0.030
Location-scale	**0.021**	0.043	**0.020**	0.036	**0.027**	0.038	**0.025**	0.047
	0.022	0.076	0.021	0.061	0.028	0.051	0.026	0.065
	0.023	0.091	0.022	0.072	0.029	0.064	0.027	0.070
	0.024	0.126	0.023	0.085	0.030	0.092	0.028	0.088
	0.025	0.152	0.024	0.093	0.031	0.101	0.029	0.096
	0.026	0.190	0.025	0.122	0.032	0.113	0.030	0.118

Table 4: Empirical Power of the Test under the Selected Tuning Parameters

State	Transformation	τ_n	DGP (1)	DGP (2)	DGP (3)
NY	Location	0.056	0.960	1.000	1.000
NY	Location	0.071	0.982	1.000	1.000
NY	Scale	0.024	0.147	1.000	1.000
NY	Scale	0.023	0.139	1.000	1.000
NY	Location-scale	0.021	0.757	1.000	1.000
NY	Location-scale	0.020	0.729	1.000	1.000
PA	Location	0.052	0.866	1.000	1.000
PA	Location	0.057	0.888	1.000	1.000
PA	Scale	0.030	0.048	1.000	1.000
PA	Scale	0.026	0.028	1.000	1.000
PA	Location-scale	0.027	0.646	1.000	1.000
PA	Location-scale	0.025	0.592	1.000	1.000

Table 5: Bootstrap p-values of the Proposed Test for Age Distributions

State	Transformation	$n_B = 1000$		$n_B = 5000$		$n_B = 10000$	
		τ_n	p-value	τ_n	p-value	τ_n	p-value
NY	Location	0.056	0.0030	0.056	0.0052	0.056	0.0041
NY	Location	0.071	0.0050	0.071	0.0068	0.071	0.0077
NY	Scale	0.024	0.0030	0.024	0.0064	0.024	0.0089
NY	Scale	0.023	0.0100	0.023	0.0084	0.023	0.0101
NY	Location-scale	0.021	0.0290	0.021	0.0272	0.021	0.0275
NY	Location-scale	0.020	0.0500	0.020	0.0300	0.020	0.0315
PA	Location	0.052	0.0970	0.052	0.0746	0.052	0.0798
PA	Location	0.057	0.0850	0.057	0.0818	0.057	0.0798
PA	Scale	0.030	0.0040	0.030	0.0046	0.030	0.0040
PA	Scale	0.026	0.0040	0.026	0.0056	0.026	0.0081
PA	Location-scale	0.027	0.0820	0.027	0.0768	0.027	0.0795
PA	Location-scale	0.025	0.0930	0.025	0.0922	0.025	0.0952

A Unified Nonparametric Test of Transformations on Distribution Functions with Nuisance Parameters
Online Supplementary Appendix

Xingyu Li
National School of Development
Peking University
x.y@pku.edu.cn

Xiaojun Song
Guanghua School of Management
Center for Statistical Science
Peking University
sxj@gsm.pku.edu.cn

Zhenting Sun
China Center for Economic Research
National School of Development
Peking University
zhentingsun@nsd.pku.edu.cn

August 17, 2022

The supplementary appendix consists of two sections. Section A provides basic concepts and auxiliary lemmas. Section B contains the proofs for Sections 2–3 in the paper.

Appendix A Concepts and Auxiliary Results

We first introduce the Hadamard directional differentiability following Definition A.1(ii) of Chen and Fang (2019), which is equivalent to the condition (2.10) of Shapiro (2000).

Definition A.1: Let $\mathbb{H}$ and $\mathbb{K}$ be normed spaces equipped with norms $\|\cdot\|_{\mathbb{H}}$ and $\|\cdot\|_{\mathbb{K}}$, respectively, and $\mathcal{F} : \mathbb{H}_{\mathcal{F}} \subset \mathbb{H} \to \mathbb{K}$. The map $\mathcal{F}$ is said to be Hadamard directionally differentiable at $\phi \in \mathbb{H}_{\mathcal{F}}$ tangentially to a set $\mathbb{H}_0 \subset \mathbb{H}$, if there is a continuous and positively homogeneous of degree one map $\mathcal{F}'_\phi : \mathbb{H}_0 \to \mathbb{K}$ such that

$$\lim_{n\to\infty} \left\| \frac{\mathcal{F}(\phi + t_n h_n) - \mathcal{F}(\phi)}{t_n} - \mathcal{F}'_\phi(h) \right\|_{\mathbb{K}} = 0$$

holds for all sequences $\{h_n\} \subset \mathbb{H}$ and $\{t_n\} \subset \mathbb{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{H}_0$ as $n \to \infty$, and $\phi + t_n h_n \in \mathbb{H}_{\mathcal{F}}$ for all n.

For the second order Hadamard directional differentiability, we introduce Definition A.2(ii) of Chen and Fang (2019), which is equivalent to the condition (2.14) of Shapiro (2000).

Definition A.2: Let $\mathbb{H}$ and $\mathbb{K}$ be normed spaces equipped with norms $\|\cdot\|_{\mathbb{H}}$ and $\|\cdot\|_{\mathbb{K}}$, respectively, and $\mathcal{F} : \mathbb{H}_{\mathcal{F}} \subset \mathbb{H} \to \mathbb{K}$. Suppose that $\mathcal{F} : \mathbb{H}_{\mathcal{F}} \to \mathbb{K}$ is Hadamard directionally differentiable tangentially to $\mathbb{H}_0 \subset \mathbb{H}$ such that the derivative $\mathcal{F}'_\phi : \mathbb{H}_0 \to \mathbb{K}$ is well defined on $\mathbb{H}$. We say that $\mathcal{F}$ is second order Hadamard directionally differentiable at $\phi \in \mathbb{H}_{\mathcal{F}}$ tangentially to $\mathbb{H}_0$ if there is

a continuous and positively homogeneous of degree two map $\mathcal{F}''_\phi : \mathbb{H}_0 \to \mathbb{K}$ such that

$$\lim_{n\to\infty} \left\| \frac{\mathcal{F}(\phi + t_n h_n) - \mathcal{F}(\phi) - t_n \mathcal{F}'_\phi(h_n)}{t_n^2} - \mathcal{F}''_\phi(h) \right\|_{\mathbb{K}} = 0$$

holds for all sequences $\{h_n\} \subset \mathbb{H}$ and $\{t_n\} \subset \mathbb{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{H}_0$ as $n \to \infty$, and $\phi + t_n h_n \in \mathbb{H}_{\mathcal{F}}$ for all n.

To establish the weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ as $n \to \infty$, we need the following lemma.

Lemma A.1: Let $\mathcal{H} = \{h_\xi : \xi \in \Xi\}$ be a class of real valued functions indexed by Ξ. Assume that $\varphi, \varphi_1, \varphi_2, \ldots$ are random elements taking values in $\ell^\infty(\mathcal{H})$. For every $\xi \in \Xi$ and every $n \in \mathbb{Z}_+$, define $\varrho(\xi) = \varphi(h_\xi)$ and $\varrho_n(\xi) = \varphi_n(h_\xi)$. If $\varphi_n \rightsquigarrow \varphi$ in $\ell^\infty(\mathcal{H})$ as $n \to \infty$, then $\varrho_n \rightsquigarrow \varrho$ in $\ell^\infty(\Xi)$ as $n \to \infty$.

Proof of Lemma A.1: Define a map $\mathcal{I} : \ell^\infty(\mathcal{H}) \to \ell^\infty(\Xi)$ such that $\mathcal{I}(\vartheta)(\xi) = \vartheta(h_\xi)$ for every $\vartheta \in \ell^\infty(\mathcal{H})$ and every $\xi \in \Xi$. Then $\mathcal{I}$ is continuous on its domain. Indeed, for all $\vartheta_1, \vartheta_2 \in \ell^\infty(\mathcal{H})$,

$$\begin{aligned} \|\mathcal{I}(\vartheta_1) - \mathcal{I}(\vartheta_2)\|_\infty &= \sup_{\xi\in\Xi} |\mathcal{I}(\vartheta_1)(\xi) - \mathcal{I}(\vartheta_2)(\xi)| = \sup_{\xi\in\Xi} |\vartheta_1(h_\xi) - \vartheta_2(h_\xi)| \\ &\leq \sup_{h\in\mathcal{H}} |\vartheta_1(h) - \vartheta_2(h)| = \|\vartheta_1 - \vartheta_2\|_\infty . \end{aligned}$$

By Theorem 1.3.6 (continuous mapping) of van der Vaart and Wellner (1996), we have

$$\varrho_n = \mathcal{I}(\varphi_n) \rightsquigarrow \mathcal{I}(\varphi) = \varrho \text{ in } \ell^\infty(\Xi)$$

as $n \to \infty$. □

The next lemma is an analog of Lemma A.1 for the almost sure weak convergence conditional on the sample.

Lemma A.2: Let $\mathcal{H} = \{h_\xi : \xi \in \Xi\}$ be a class of real valued functions indexed by Ξ. Assume that φ is a tight random element taking values in $\ell^\infty(\mathcal{H})$, and that for every $n \in \mathbb{Z}_+$, $\mathcal{Z}_n$ is a random sample of size n and φ_n is a random element taking values in $\ell^\infty(\mathcal{H})$. For every $\xi \in \Xi$ and every $n \in \mathbb{Z}_+$, define $\varrho(\xi) = \varphi(h_\xi)$ and $\varrho_n(\xi) = \varphi_n(h_\xi)$. If $\varphi_n \overset{\text{a.s.}}{\rightsquigarrow} \varphi$ as $n \to \infty$, then $\varrho_n \overset{\text{a.s.}}{\rightsquigarrow} \varrho$ as $n \to \infty$. Moreover, if $\{\varphi_n\}$ is asymptotically measurable, then $\{\varrho_n\}$ is also asymptotically measurable.

Proof of Lemma A.2: Define a map $\mathcal{I} : \ell^\infty(\mathcal{H}) \to \ell^\infty(\Xi)$ such that $\mathcal{I}(\vartheta)(\xi) = \vartheta(h_\xi)$ for every $\vartheta \in \ell^\infty(\mathcal{H})$ and every $\xi \in \Xi$. As shown in the proof of Lemma A.1, for all $\vartheta_1, \vartheta_2 \in \ell^\infty(\mathcal{H})$,

$$\|\mathcal{I}(\vartheta_1) - \mathcal{I}(\vartheta_2)\|_\infty \leq \|\vartheta_1 - \vartheta_2\|_\infty ,$$

which implies the Lipschitz continuity of $\mathcal{I}$. The almost sure weak convergence of ϱ_n conditional on $\mathcal{Z}_n$ follows from Proposition 10.7(ii) of Kosorok (2008). The asymptotic measurability follows from the continuity of $\mathcal{I}$. □

Appendix B Proof of Main Results

B.1 Proofs for Section 2

Lemma B.1: If $\varphi_1, \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$, then $a_1\varphi_1 + a_2\varphi_2 \in \mathbb{D}_{\mathcal{L}0}$ for all $a_1, a_2 \in \mathbb{R}$, and the functions

$$\theta \mapsto \int_{\mathbb{R}} [\varphi_1(x,\theta)]^2 \, \mathrm{d}\nu(x) \text{ and } \theta \mapsto \int_{\mathbb{R}} \varphi_1(x,\theta)\varphi_2(x,\theta) \, \mathrm{d}\nu(x)$$

are continuous on Θ.

Proof of Lemma B.1: For all $\varphi_1, \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$ and all $a_1, a_2 \in \mathbb{R}$, let $M = \|\varphi_1\|_\infty \vee \|\varphi_2\|_\infty \vee 2a_1^2 \vee 2a_2^2$. By the definition of $\mathbb{D}_{\mathcal{L}0}$, for every $\theta_0 \in \Theta$ and every $\varepsilon > 0$, there exists $\delta(\theta_0, \varepsilon) > 0$ such that

$$\int_{\mathbb{R}} [\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]^2 \, \mathrm{d}\nu(x) \vee \int_{\mathbb{R}} [\varphi_2(x,\theta) - \varphi_2(x,\theta_0)]^2 \, \mathrm{d}\nu(x) < \frac{\varepsilon}{2M} \wedge \left[\frac{\varepsilon}{2M}\right]^2$$

whenever $\|\theta - \theta_0\|_2 < \delta(\theta_0, \varepsilon)$.

To show the first claim, note that

$$\begin{aligned}
&\int_{\mathbb{R}} [a_1\varphi_1(x,\theta) + a_2\varphi_2(x,\theta) - a_1\varphi_1(x,\theta_0) - a_2\varphi_2(x,\theta_0)]^2 \, \mathrm{d}\nu(x) \\
&\leq 2a_1^2 \int_{\mathbb{R}} [\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]^2 \, \mathrm{d}\nu(x) + 2a_2^2 \int_{\mathbb{R}} [\varphi_2(x,\theta) - \varphi_2(x,\theta_0)]^2 \, \mathrm{d}\nu(x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

whenever $\|\theta - \theta_0\|_2 < \delta(\theta_0, \varepsilon)$. For the second claim, we have

$$\begin{aligned}
&\left| \int_{\mathbb{R}} [\varphi_1(x,\theta)]^2 \, \mathrm{d}\nu(x) - \int_{\mathbb{R}} [\varphi_1(x,\theta_0)]^2 \, \mathrm{d}\nu(x) \right| \\
&\leq \int_{\mathbb{R}} |[\varphi_1(x,\theta) + \varphi_1(x,\theta_0)][\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]| \, \mathrm{d}\nu(x) \\
&\leq 2M \int_{\mathbb{R}} |\varphi_1(x,\theta) - \varphi_1(x,\theta_0)| \, \mathrm{d}\nu(x) \leq 2M \sqrt{\int_{\mathbb{R}} [\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]^2 \, \mathrm{d}\nu(x)} < \varepsilon
\end{aligned}$$

whenever $\|\theta - \theta_0\|_2 < \delta(\theta_0, \varepsilon)$, where the third inequality follows from the convexity of square functions and Jensen's inequality. The third claim can be proved analogously, since

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \varphi_1(x,\theta)\varphi_2(x,\theta) \, \mathrm{d}\nu(x) - \int_{\mathbb{R}} \varphi_1(x,\theta_0)\varphi_2(x,\theta_0) \, \mathrm{d}\nu(x) \right| \\
&\leq \int_{\mathbb{R}} |\varphi_1(x,\theta)[\varphi_2(x,\theta) - \varphi_2(x,\theta_0)] + \varphi_2(x,\theta_0)[\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]| \, \mathrm{d}\nu(x) \\
&\leq M \int_{\mathbb{R}} |\varphi_1(x,\theta) - \varphi_1(x,\theta_0)| \, \mathrm{d}\nu(x) + M \int_{\mathbb{R}} |\varphi_2(x,\theta) - \varphi_2(x,\theta_0)| \, \mathrm{d}\nu(x) \\
&\leq M \sqrt{\int_{\mathbb{R}} [\varphi_1(x,\theta) - \varphi_1(x,\theta_0)]^2 \, \mathrm{d}\nu(x)} + M \sqrt{\int_{\mathbb{R}} [\varphi_2(x,\theta) - \varphi_2(x,\theta_0)]^2 \, \mathrm{d}\nu(x)} < \varepsilon
\end{aligned}$$

whenever $\|\theta - \theta_0\|_2 < \delta(\theta_0, \varepsilon)$, where the third inequality follows from the convexity of square functions and Jensen's inequality. □

Proof of Proposition 2.1: If $F(x) = G(g(x,\theta))$ for all $x \in \mathbb{R}$ with some $\theta \in \Theta$, then (4) holds trivially.

Next, we show that (4) implies (1). Recall that μ is the Lebesgue measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Since $G \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$, Assumption 2.4 implies that $G \circ g \in \mathbb{D}_{\mathcal{L}0}$ and hence $\phi \in \mathbb{D}_{\mathcal{L}0}$ by Lemma B.1. Also, by Lemma B.1, the function $\theta \mapsto \int_{\mathbb{R}} [F(x) - G(g(x,\theta))]^2 \, \mathrm{d}\nu(x)$ is continuous on Θ. By Assumption

2.3, there exists $\theta_0 \in \Theta$ such that

$$\int_{\mathbb{R}} [F(x) - G(g(x,\theta_0))]^2 \, \mathrm{d}\nu(x) = \inf_{\theta\in\Theta} \int_{\mathbb{R}} [F(x) - G(g(x,\theta))]^2 \, \mathrm{d}\nu(x) = 0. \tag{B.1}$$

Define $A = \{x \in \mathbb{R} : F(x) \neq G(g(x,\theta_0))\}$. Then (B.1) implies that $\nu(A) = 0$ by Proposition 2.16 of Folland (1999). By the assumption that $\mu \ll \nu$, $\mu(A) = 0$. We now claim that $A = \varnothing$. Otherwise, there is an $x_0 \in \mathbb{R}$ such that $F(x_0) \neq G(g(x_0,\theta_0))$. Since both F and G are continuous and $g(\cdot,\theta_0)$ is continuous, there exists $\delta > 0$ such that $F(x) \neq G(g(x,\theta_0))$ for all $x \in [x_0, x_0+\delta]$. This contradicts $\mu(A) = 0$. Therefore, we have $F(x) = G(g(x,\theta_0))$ for all $x \in \mathbb{R}$. □

Lemma B.2: Under Assumptions 2.5 and 2.6, we have

$$\lim_{n\to\infty} \sup_{(x,\theta)\in\mathbb{R}\times\Theta} \left|\widehat{\phi}_n(x,\theta) - \phi(x,\theta)\right| = 0 \text{ almost surely.}$$

Proof of Lemma B.2: By Theorem 19.1 of van der Vaart (1998), we have

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\widehat{F}_{n_1}(x) - F(x)\right| = 0 \text{ and } \lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\widehat{G}_{n_2}(x) - G(x)\right| = 0 \text{ almost surely.}$$

Note that for every $(x,\theta) \in \mathbb{R}\times\Theta$,

$$\left|\widehat{G}_{n_2}(g(x,\theta)) - G(g(x,\theta))\right| \leq \sup_{z\in\mathbb{R}} \left|\widehat{G}_{n_2}(z) - G(z)\right|,$$

which implies

$$\lim_{n\to\infty} \sup_{(x,\theta)\in\mathbb{R}\times\Theta} \left|\widehat{G}_{n_2}(g(x,\theta)) - G(g(x,\theta))\right| = 0 \text{ almost surely.}$$

Then the desired result follows from the definitions of $\widehat{\phi}_n$ and ϕ. □

Proof of Lemma 2.1: Firstly, we show the weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ for the independent samples case. By Theorem 19.3 of van der Vaart (1998), we have

$$\sqrt{n_1}(\widehat{F}_{n_1} - F) \rightsquigarrow \mathbb{W}_F \text{ in } \ell^\infty(\mathbb{R}) \text{ and } \sqrt{n_2}(\widehat{G}_{n_2} - G) \rightsquigarrow \mathbb{W}_G \text{ in } \ell^\infty(\mathbb{R})$$

as $n \to \infty$, where $\mathbb{W}_F$ and $\mathbb{W}_G$ are independent of each other. Define two classes of indicator functions

$$\mathcal{G}_1 = \left\{\mathbb{1}_{(-\infty,x]} : x \in \mathbb{R}\right\} \text{ and } \mathcal{G}_2 = \left\{\mathbb{1}_{(-\infty,g(x,\theta)]} : (x,\theta) \in \mathbb{R}\times\Theta\right\}.$$

Let $\widehat{\mathcal{Y}}_{n_2}$ be a stochastic process and $\mathcal{Y}$ be a real valued function such that

$$\widehat{\mathcal{Y}}_{n_2}(f) = \frac{1}{n_2}\sum_{i=1}^{n_2} f(Y_i) \text{ and } \mathcal{Y}(f) = \mathbb{E}[f(Y_i)]$$

for every measurable f. By Example 2.5.4 of van der Vaart and Wellner (1996), $\mathcal{G}_1$ is a Donsker class. Therefore, $\sqrt{n_2}(\widehat{\mathcal{Y}}_{n_2} - \mathcal{Y}) \rightsquigarrow \mathbb{Y}$ in $\ell^\infty(\mathcal{G}_1)$ as $n \to \infty$, where $\mathbb{Y}$ is a tight measurable centered Gaussian process with $\mathbb{E}[\mathbb{Y}(f_1)\mathbb{Y}(f_2)] = \mathcal{Y}(f_1 f_2) - \mathcal{Y}(f_1)\mathcal{Y}(f_2)$ for all $f_1, f_2 \in \mathcal{G}_1$. Since $\mathcal{G}_2 \subset \mathcal{G}_1$, it follows that for every $h \in \mathcal{C}_\mathrm{b}(\ell^\infty(\mathcal{G}_2))$, $h \in \mathcal{C}_\mathrm{b}(\ell^\infty(\mathcal{G}_1))$ and

$$\mathbb{E}[h(\sqrt{n_2}(\widehat{\mathcal{Y}}_{n_2} - \mathcal{Y}))] \to \mathbb{E}[h(\mathbb{Y})],$$

which implies $\sqrt{n_2}(\widehat{\mathcal{Y}}_{n_2} - \mathcal{Y}) \rightsquigarrow \mathbb{Y}$ in $\ell^\infty(\mathcal{G}_2)$ as $n \to \infty$. It is easy to show that $\widehat{G}_{n_2} \circ g(x,\theta) = \widehat{\mathcal{Y}}_{n_2}(\mathbb{1}_{(-\infty,g(x,\theta)]})$ and $G \circ g(x,\theta) = \mathcal{Y}(\mathbb{1}_{(-\infty,g(x,\theta)]})$ for every $(x,\theta) \in \mathbb{R}\times\Theta$. Define a random element $W \in \ell^\infty(\mathbb{R}\times\Theta)$ such that $W(x,\theta) = \mathbb{Y}(\mathbb{1}_{(-\infty,g(x,\theta)]})$ for all $(x,\theta) \in \mathbb{R}\times\Theta$. By Lemma A.1, $\sqrt{n_2}(\widehat{G}_{n_2} \circ g - G \circ g) \rightsquigarrow W$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n \to \infty$. By the independence between $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$, Assumption 2.6 of this paper, and Example 1.4.6 of van der Vaart and Wellner

(1996), we have the joint weak convergence

$$\begin{bmatrix} \sqrt{T_n}\left(\widehat{F}_{n_1} - F\right) \\ \sqrt{T_n}\left(\widehat{G}_{n_2} \circ g - G \circ g\right) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \sqrt{1-\lambda}\mathbb{W}_F \\ \sqrt{\lambda}W \end{bmatrix} \text{ in } \ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R} \times \Theta)$$

as $n \to \infty$, where $\mathbb{W}_F$ and W are independent of each other. Define $\mathbb{A} = \ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R} \times \Theta)$, and define the norm $\|\cdot\|_{\mathbb{A}}$ with $\|(f,h)\|_{\mathbb{A}} = \|f\|_\infty + \|h\|_\infty$ for every $(f,h) \in \mathbb{A}$. Let $\mathcal{I} : \mathbb{A} \to \ell^\infty(\mathbb{R} \times \Theta)$ be such that $\mathcal{I}(f,h)(x,\theta) = f(x) - h(x,\theta)$ for every $(f,h) \in \mathbb{A}$ and every $(x,\theta) \in \mathbb{R} \times \Theta$. Note that

$$\|\mathcal{I}(f_1,h_1) - \mathcal{I}(f_2,h_2)\|_\infty = \sup_{(x,\theta)\in\mathbb{R}\times\Theta} |f_1(x) - h_1(x,\theta) - f_2(x) + h_2(x,\theta)|$$

$$\leq \sup_{x\in\mathbb{R}} |f_1(x) - f_2(x)| + \sup_{(x,\theta)\in\mathbb{R}\times\Theta} |h_1(x,\theta) - h_2(x,\theta)| = \|(f_1,h_1) - (f_2,h_2)\|_{\mathbb{A}}$$

for all $(f_1,h_1),(f_2,h_2) \in \mathbb{A}$, and therefore $\mathcal{I}$ is continuous. The weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ to some tight $\mathbb{G}_0 = \sqrt{1-\lambda}\mathbb{W}_F - \sqrt{\lambda}W$ follows from Theorem 1.3.6 (continuous mapping) of van der Vaart and Wellner (1996). Furthermore,

$$\mathbb{E}\{[(\mathbb{W}_G \circ g)(x_1,\theta_1)][(\mathbb{W}_G \circ g)(x_2,\theta_2)]\} = G[g(x_1,\theta_1) \wedge g(x_2,\theta_2)] - G[g(x_1,\theta_1)]G[g(x_2,\theta_2)]$$

$$= \mathcal{Y}\left(\mathbb{1}_{(-\infty,g(x_1,\theta_1)]}\mathbb{1}_{(-\infty,g(x_2,\theta_2)]}\right) - \mathcal{Y}\left(\mathbb{1}_{(-\infty,g(x_1,\theta_1)]}\right)\mathcal{Y}\left(\mathbb{1}_{(-\infty,g(x_2,\theta_2)]}\right) = \mathbb{E}[W(x_1,\theta_1)W(x_2,\theta_2)]$$

for all $(x_1,\theta_1),(x_2,\theta_2) \in \mathbb{R} \times \Theta$, which verifies the variance $Var(\mathbb{G}_0(x,\theta))$ for every $(x,\theta) \in \mathbb{R} \times \Theta$.

Define a metric ρ_2 on $\mathcal{G}_1$ such that for all $x_1, x_2 \in \mathbb{R}$,

$$\rho_2\left(\mathbb{1}_{(-\infty,x_1]}, \mathbb{1}_{(-\infty,x_2]}\right) = \mathbb{E}\left[\left(\mathbb{Y}(\mathbb{1}_{(-\infty,x_1]}) - \mathbb{Y}(\mathbb{1}_{(-\infty,x_2]})\right)^2\right]^{1/2}.$$

By the discussion in Example 1.5.10 of van der Vaart and Wellner (1996), $\mathbb{Y}$ almost surely has a ρ_2-uniformly continuous path on $\mathcal{G}_1$. Also, for all $x_1, x_2 \in \mathbb{R}$,

$$\rho_2\left(\mathbb{1}_{(-\infty,x_1]}, \mathbb{1}_{(-\infty,x_2]}\right)^2 = \mathbb{E}\left[\left(\mathbb{Y}(\mathbb{1}_{(-\infty,x_1]}) - \mathbb{Y}(\mathbb{1}_{(-\infty,x_2]})\right)^2\right]$$

$$= \mathbb{E}\left[\mathbb{Y}(\mathbb{1}_{(-\infty,x_1]})^2\right] + \mathbb{E}\left[\mathbb{Y}(\mathbb{1}_{(-\infty,x_2]})^2\right] - 2\mathbb{E}\left[\mathbb{Y}(\mathbb{1}_{(-\infty,x_1]})\mathbb{Y}(\mathbb{1}_{(-\infty,x_2]})\right]$$

$$= \mathcal{Y}((\mathbb{1}_{(-\infty,x_1]} - \mathbb{1}_{(-\infty,x_2]})^2) - (\mathcal{Y}(\mathbb{1}_{(-\infty,x_1]}) - \mathcal{Y}(\mathbb{1}_{(-\infty,x_2]}))^2.$$

This implies that

$$\rho_2\left(\mathbb{1}_{(-\infty,x_k]}, \mathbb{1}_{(-\infty,x]}\right) \to 0 \text{ as } x_k \to x.$$

Let $\tilde{W}(x) = \mathbb{Y}(\mathbb{1}_{(-\infty,x]})$ for all $x \in \mathbb{R}$. Then almost surely, $\tilde{W}$ is bounded and has a continuous path on $\mathbb{R}$. Note that $W(x,\theta) = \tilde{W} \circ g(x,\theta)$ for all $(x,\theta) \in \mathbb{R} \times \Theta$. Then by Assumption 2.4, $\mathbb{P}(\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}) = 1$.

Secondly, we show the weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ for the matched pairs case. Let $h_{x,\theta}(v) = \mathbb{1}_{(-\infty,x]\times\mathbb{R}}(v_1,v_2) - \mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]}(v_1,v_2)$ for every $(x,\theta) \in \mathbb{R} \times \Theta$ and every $v = (v_1,v_2) \in \mathbb{R}^2$. Define four classes of functions on $\mathbb{R}^2$:

$$\mathcal{H}_1 = \left\{\mathbb{1}_{(-\infty,x]\times\mathbb{R}} : x \in \mathbb{R}\right\}, \mathcal{H}_2 = \left\{-\mathbb{1}_{\mathbb{R}\times(-\infty,x]} : x \in \mathbb{R}\right\}, \mathcal{H}_+ = \{h_1 + h_2 : h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\},$$

$$\text{and } \mathcal{H} = \left\{\mathbb{1}_{(-\infty,x]\times\mathbb{R}} - \mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]} : (x,\theta) \in \mathbb{R} \times \Theta\right\} \subset \mathcal{H}_+.$$

Let $V_i = (X_i, Y_i)$ for $i = 1, 2, \ldots, n_1$. Define $\widehat{\mathcal{V}}_{n_1}$ to be a stochastic process and $\mathcal{V}$ to be a real

valued function with

$$\widehat{\mathcal{V}}_{n_1}(f) = \frac{1}{n_1}\sum_{i=1}^{n_1} f(V_i) \text{ and } \mathcal{V}(f) = \mathbb{E}[f(V_i)]$$

for every measurable f. By Example 2.5.4 of van der Vaart and Wellner (1996), both $\mathcal{H}_1$ and $\mathcal{H}_2$ are Donsker classes. Note that $\sup_{f\in\mathcal{H}_1\cup\mathcal{H}_2}|\mathcal{V}(f)| \leq 1$, then by Example 2.10.7 and Theorem 2.10.1 of van der Vaart and Wellner (1996), $\mathcal{H}_+$ and $\mathcal{H}$ are Donsker classes. Let $\mathcal{H}_U = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_+$. By Example 2.10.7 of van der Vaart and Wellner (1996), $\mathcal{H}_U$ is a Donsker class. Therefore, $\sqrt{n_1}(\widehat{\mathcal{V}}_{n_1} - \mathcal{V}) \rightsquigarrow \mathbb{V}$ in $\ell^\infty(\mathcal{H}_U)$ as $n \to \infty$, where $\mathbb{V}$ is a centered Gaussian process with $\mathbb{E}[\mathbb{V}(f_1)\mathbb{V}(f_2)] = \mathcal{V}(f_1 f_2) - \mathcal{V}(f_1)\mathcal{V}(f_2)$ for all $f_1, f_2 \in \mathcal{H}_U$.

Under the assumption of matched pairs,

$$\widehat{\phi}_n(x,\theta) = \frac{1}{n_1}\sum_{i=1}^{n_1}\left[\mathbb{1}_{(-\infty,x]}(X_i) - \mathbb{1}_{(-\infty,g(x,\theta)]}(Y_i)\right] = \frac{1}{n_1}\sum_{i=1}^{n_1} h_{x,\theta}(V_i) = \widehat{\mathcal{V}}_{n_1}(h_{x,\theta}),$$

$$\text{and } \phi(x,\theta) = \mathbb{E}\left[\mathbb{1}_{(-\infty,x]}(X_i) - \mathbb{1}_{(-\infty,g(x,\theta)]}(Y_i)\right] = \mathbb{E}[h_{x,\theta}(V_i)] = \mathcal{V}(h_{x,\theta})$$

for all $(x,\theta) \in \mathbb{R}\times\Theta$. Define random elements $W_n, W \in \ell^\infty(\mathbb{R}\times\Theta)$ with $W_n(x,\theta) = \sqrt{n_1}(\widehat{\mathcal{V}}_{n_1}(h_{x,\theta}) - \mathcal{V}(h_{x,\theta}))$ and $W(x,\theta) = \mathbb{V}(h_{x,\theta})$ for every $(x,\theta)\in\mathbb{R}\times\Theta$. Similarly to the proof for the independent samples case, we can show that $\sqrt{n_1}(\widehat{\mathcal{V}}_{n_1} - \mathcal{V}) \rightsquigarrow \mathbb{V}$ in $\ell^\infty(\mathcal{H})$ as $n\to\infty$ because $\mathcal{H}\subset\mathcal{H}_U$. By Lemma A.1, $W_n \rightsquigarrow W$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n\to\infty$. Thus, it follows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\ell^\infty(\mathbb{R}\times\Theta)$ with $\mathbb{G}_0 = \sqrt{1/2}W$ as $n\to\infty$. Furthermore,

$$\begin{aligned}
&\mathbb{E}\{[\mathbb{W}_F(x_1) - \mathbb{W}_G\circ g(x_1,\theta_1)][\mathbb{W}_F(x_2) - \mathbb{W}_G\circ g(x_2,\theta_2)]\}\\
&= \mathbb{E}[\mathbb{W}_F(x_1)\mathbb{W}_F(x_2)] - \mathbb{E}[\mathbb{W}_F(x_1)\mathbb{W}_G\circ g(x_2,\theta_2)]\\
&\quad - \mathbb{E}[\mathbb{W}_F(x_2)\mathbb{W}_G\circ g(x_1,\theta_1)] + \mathbb{E}[\mathbb{W}_G\circ g(x_1,\theta_1)\mathbb{W}_G\circ g(x_2,\theta_2)]\\
&= F(x_1\wedge x_2) - F(x_1)F(x_2) - H(x_1, g(x_2,\theta_2)) + F(x_1)G(g(x_2,\theta_2))\\
&\quad - H(x_2, g(x_1,\theta_1)) + F(x_2)G(g(x_1,\theta_1)) + G(g(x_1,\theta_1)\wedge g(x_2,\theta_2))\\
&\quad - G(g(x_1,\theta_1))G(g(x_2,\theta_2))\\
&= \mathbb{E}\left\{\left[\mathbb{1}_{(-\infty,x_1]}(X_i) - \mathbb{1}_{(-\infty,g(x_1,\theta_1)]}(Y_i)\right]\left[\mathbb{1}_{(-\infty,x_2]}(X_i) - \mathbb{1}_{(-\infty,g(x_2,\theta_2)]}(Y_i)\right]\right\}\\
&\quad - \mathbb{E}\left[\mathbb{1}_{(-\infty,x_1]}(X_i) - \mathbb{1}_{(-\infty,g(x_1,\theta_1)]}(Y_i)\right]\mathbb{E}\left[\mathbb{1}_{(-\infty,x_2]}(X_i) - \mathbb{1}_{(-\infty,g(x_2,\theta_2)]}(Y_i)\right]\\
&= \mathcal{V}(h_{x_1,\theta_1}h_{x_2,\theta_2}) - \mathcal{V}(h_{x_1,\theta_1})\mathcal{V}(h_{x_2,\theta_2}) = \mathbb{E}[W(x_1,\theta_1)W(x_2,\theta_2)]
\end{aligned}$$

for all $(x_1,\theta_1), (x_2,\theta_2)\in\mathbb{R}\times\Theta$, which verifies the variance $Var(\mathbb{G}_0(x,\theta))$ for every $(x,\theta)\in\mathbb{R}\times\Theta$. Let $A\subset\ell^\infty(\mathcal{H}_U)$ be the collection of all functions $f\in\ell^\infty(\mathcal{H}_U)$ such that $f(h_1+h_2) = f(h_1) + f(h_2)$ for all $h_1\in\mathcal{H}_1$ and all $h_2\in\mathcal{H}_2$. Let $f_k\in A$ such that $f_k\to f$ for some $f\in\ell^\infty(\mathcal{H}_U)$. Then we have that for all $h_1\in\mathcal{H}_1$ and all $h_2\in\mathcal{H}_2$,

$$f(h_1+h_2) = \lim_{k\to\infty} f_k(h_1+h_2) = \lim_{k\to\infty}\{f_k(h_1) + f_k(h_2)\} = f(h_1) + f(h_2).$$

This implies that A is closed. Since $\sqrt{n_1}(\widehat{\mathcal{V}}_{n_1} - \mathcal{V}) \rightsquigarrow \mathbb{V}$ in $\ell^\infty(\mathcal{H}_U)$ and $\sqrt{n_1}(\widehat{\mathcal{V}}_{n_1} - \mathcal{V})\in A$, then by Theorem 1.3.4(iii) (portmanteau) of van der Vaart and Wellner (1996), $\mathbb{V}\in A$ almost surely. This implies $W(x,\theta) = \mathbb{V}(\mathbb{1}_{(-\infty,x]\times\mathbb{R}}) + \mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]})$ almost surely.

Define a metric ρ_2 on $\mathcal{H}_U$ such that for all $h_1, h_2\in\mathcal{H}_U$,

$$\rho_2(h_1,h_2) = \mathbb{E}\left[(\mathbb{V}(h_1) - \mathbb{V}(h_2))^2\right]^{1/2}.$$

By the discussion in Example 1.5.10 of van der Vaart and Wellner (1996), $\mathbb{V}$ almost surely has a ρ_2-uniformly continuous path on $\mathcal{H}_U$. Also, for all $x_1, x_2 \in \mathbb{R}$,

$$\begin{aligned}
&\rho_2\left(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]}, -\mathbb{1}_{\mathbb{R}\times(-\infty,x_2]}\right)^2 = \mathbb{E}\left[\left(\mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]}) - \mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_2]})\right)^2\right] \\
=& \mathbb{E}\left[\mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]})^2\right] + \mathbb{E}\left[\mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_2]})^2\right] - 2\mathbb{E}\left[\mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]})\mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_2]})\right] \\
=& \mathcal{V}((\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]} - \mathbb{1}_{\mathbb{R}\times(-\infty,x_2]})^2) - (\mathcal{V}(\mathbb{1}_{\mathbb{R}\times(-\infty,x_1]}) - \mathcal{V}(\mathbb{1}_{\mathbb{R}\times(-\infty,x_2]}))^2.
\end{aligned}$$

This implies that

$$\rho_2\left(-\mathbb{1}_{\mathbb{R}\times(-\infty,x_k]}, -\mathbb{1}_{\mathbb{R}\times(-\infty,x]}\right) \to 0 \text{ as } x_k \to x.$$

Let $\tilde{W}(x) = \mathbb{V}(-\mathbb{1}_{\mathbb{R}\times(-\infty,x]})$ for all $x \in \mathbb{R}$. Then almost surely, $\tilde{W}$ is bounded and has a continuous path on $\mathbb{R}$. Note that $W(x,\theta) = \mathbb{V}(\mathbb{1}_{(-\infty,x]\times\mathbb{R}}) + \tilde{W} \circ g(x,\theta)$ for all $(x,\theta) \in \mathbb{R}\times\Theta$. Then by Assumption 2.4, $\mathbb{P}(\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}) = 1$. □

Proof of Lemma 2.2: Define a map $\mathcal{S} : \mathbb{D}_{\mathcal{L}} \to \ell^\infty(\Theta)$ such that for every $\varphi \in \mathbb{D}_{\mathcal{L}}$ and every $\theta \in \Theta$,

$$\mathcal{S}(\varphi)(\theta) = \int_{\mathbb{R}} [\varphi(x,\theta)]^2 \, \mathrm{d}\nu(x).$$

We show that the Hadamard directional derivative of $\mathcal{S}$ at $\phi \in \mathbb{D}_{\mathcal{L}}$ is

$$\mathcal{S}'_\phi(h)(\theta) = \int_{\mathbb{R}} 2\phi(x,\theta)h(x,\theta) \, \mathrm{d}\nu(x) \text{ for all } h \in \mathbb{D}_{\mathcal{L}0}.$$

Because $F, G \in \mathcal{C}_b(\mathbb{R})$, by Assumption 2.4 and Lemma B.1, $\mathcal{S}(\phi) \in \mathcal{C}(\Theta)$. Indeed, for all sequences $\{h_n\}_{n=1}^\infty \subset \ell^\infty(\mathbb{R}\times\Theta)$ and $\{t_n\}_{n=1}^\infty \subset \mathbb{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_{\mathcal{L}0}$ as $n \to \infty$, and $\phi + t_n h_n \in \mathbb{D}_{\mathcal{L}}$ for all n, we have that $M = \sup_{n\in\mathbb{Z}_+} \|h_n\|_\infty < \infty$, and

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{\mathcal{S}(\phi + t_n h_n)(\theta) - \mathcal{S}(\phi)(\theta)}{t_n} - \mathcal{S}'_\phi(h)(\theta)\right| \\
=& \sup_{\theta\in\Theta}\left|\int_{\mathbb{R}} t_n h_n^2(x,\theta) + 2\phi(x,\theta)\left[h_n(x,\theta) - h(x,\theta)\right] \, \mathrm{d}\nu(x)\right| \\
\leq& \int_{\mathbb{R}} t_n M^2 + 2\|\phi\|_\infty \|h_n - h\|_\infty \, \mathrm{d}\nu(x) = t_n M^2 + 2\|\phi\|_\infty \|h_n - h\|_\infty \to 0,
\end{aligned}$$

since $t_n \downarrow 0$ and $h_n \to h$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n \to \infty$.

Define a function $\mathcal{R}$ such that for every $\psi \in \mathcal{C}(\Theta)$, $\mathcal{R}(\psi) = \inf_{\theta\in\Theta}\psi(\theta)$. By Lemma S.4.9 of Fang and Santos (2019), $\mathcal{R}$ is Hadamard directionally differentiable at every $\psi \in \mathcal{C}(\Theta)$ tangentially to $\mathcal{C}(\Theta)$ with the Hadamard directional derivative

$$\mathcal{R}'_\psi(f) = \inf_{\theta\in\Theta_0^*(\psi)} f(\theta) \text{ for all } f \in \mathcal{C}(\Theta),$$

where $\Theta_0^*(\psi) = \arg\min_{\theta\in\Theta}\psi(\theta)$.

Note that $\mathcal{L}(\varphi) = \mathcal{R}[\mathcal{S}(\varphi)] = \mathcal{R}\circ\mathcal{S}(\varphi)$ for every $\varphi \in \mathbb{D}_{\mathcal{L}}$. By Proposition 3.6(i) of Shapiro (1990), $\mathcal{L}$ is Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the Hadamard directional derivative

$$\mathcal{L}'_\phi(h) = \mathcal{R}'_{\mathcal{S}(\phi)}\left[\mathcal{S}'_\phi(h)\right] = \inf_{\theta\in\Theta_0^*(\mathcal{S}(\phi))}\int_{\mathbb{R}} 2\phi(x,\theta)h(x,\theta) \, \mathrm{d}\nu(x) \text{ for all } h \in \mathbb{D}_{\mathcal{L}0}.$$

Since $\Theta_0^*(\mathcal{S}(\phi)) = \arg\min_{\theta\in\Theta}\int_{\mathbb{R}}[\phi(x,\theta)]^2 \, \mathrm{d}\nu(x)$, the desired result follows.

Now we turn to the degeneracy of $\mathcal{L}'_\phi$ under the condition that $\phi \in \mathbb{D}_0$. If $\phi \in \mathbb{D}_0$, for every

$\theta \in \Theta_0(\phi)$, we have

$$\int_{\mathbb{R}} [\phi(x,\theta)]^2 \, \mathrm{d}\nu(x) = 0,$$

and consequently $\phi(x,\theta) = 0$ holds for ν-almost every x. Therefore, $\mathcal{L}'_\phi(h) = 0$ for every $h \in \ell^\infty(\mathbb{R}\times\Theta)$ whenever $\phi \in \mathbb{D}_0$. □

Proof of Lemma 2.3: Define $\Phi : \Theta \to L^2(\nu)$ such that $\Phi(\theta)(x) = \phi(x,\theta)$ for every $(x,\theta) \in \mathbb{R}\times\Theta$. Then it is easy to show that

$$\mathcal{L}(\phi) = \inf_{\theta\in\Theta} \int_{\mathbb{R}} [\phi(x,\theta)]^2 \, \mathrm{d}\nu(x) = \inf_{\theta\in\Theta} \|\Phi(\theta)\|^2_{L^2(\nu)} = 0,$$

and $\Theta_0(\phi) = \{\theta \in \Theta : \|\Phi(\theta)\|_{L^2(\nu)} = 0\} = \Theta_0$. Consider all sequences $\{t_n\}_{n=1}^\infty \subset \mathbb{R}_+$ and $\{h_n\}_{n=1}^\infty \subset \ell^\infty(\mathbb{R}\times\Theta)$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_{\mathcal{L}0}$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n\to\infty$, and $\phi + t_n h_n \in \mathbb{D}_{\mathcal{L}}$ for all n. For notational simplicity, define $\mathscr{H}_n : \Theta \to L^2(\nu)$ for every $n \in \mathbb{Z}_+$ such that $\mathscr{H}_n(\theta)(x) = h_n(x,\theta)$ for every $(x,\theta) \in \mathbb{R}\times\Theta$, and define $\mathscr{H} : \Theta \to L^2(\nu)$ such that $\mathscr{H}(\theta)(x) = h(x,\theta)$ for every $(x,\theta) \in \mathbb{R}\times\Theta$.

Since $h_n \to h \in \mathbb{D}_{\mathcal{L}0}$ in $\ell^\infty(\mathbb{R}\times\Theta)$, it follows that $\|h\|_\infty \vee \sup_{n\in\mathbb{Z}_+} \|h_n\|_\infty = M_1$ for some $M_1 < \infty$. Then we have that

$$\begin{aligned}
&|\mathcal{L}(\phi + t_n h_n) - \mathcal{L}(\phi + t_n h)| = \left| \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}_n(\theta)\|^2_{L^2(\nu)} - \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|^2_{L^2(\nu)} \right| \\
&= \left| \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}_n(\theta)\|_{L^2(\nu)} + \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L^2(\nu)} \right| \\
&\quad \cdot \left| \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}_n(\theta)\|_{L^2(\nu)} - \inf_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L^2(\nu)} \right| \\
&\leq \left| \inf_{\theta\in\Theta_0(\phi)} \|\Phi(\theta) + t_n\mathscr{H}_n(\theta)\|_{L^2(\nu)} + \inf_{\theta\in\Theta_0(\phi)} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L^2(\nu)} \right| \\
&\quad \cdot \left(t_n \sup_{\theta\in\Theta} \|\mathscr{H}_n(\theta) - \mathscr{H}(\theta)\|_{L^2(\nu)} \right) \\
&\leq 2M_1 t_n^2 \|h_n - h\|_\infty = o\left(t_n^2\right),
\end{aligned}$$

where the first inequality follows from the Lipschitz continuity of the supremum map and the triangle inequality, and the second inequality follows from the fact that $\Phi(\theta) = 0$ ν-almost everywhere for every $\theta \in \Theta_0(\phi)$.

Then for the h, pick an $a(h) > 0$ such that $Ca(h)^\kappa = 3\|h\|_\infty$, where C and κ are defined as in Assumption 2.8. For sufficiently large $n \in \mathbb{Z}_+$ such that $t_n^\kappa \geq t_n$, we have that

$$\begin{aligned}
&\inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L^2(\nu)} \\
&\geq \inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta)\|_{L^2(\nu)} + \inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \left[-t_n \|\mathscr{H}(\theta)\|_{L^2(\nu)} \right] \\
&= \inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta)\|_{L^2(\nu)} - \sup_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} t_n \|\mathscr{H}(\theta)\|_{L^2(\nu)} \\
&\geq C\left(a(h)t_n\right)^\kappa - t_n \sup_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\mathscr{H}(\theta)\|_{L^2(\nu)} \geq 3\|h\|_\infty t_n^\kappa - t_n\|h\|_\infty \\
&> t_n \inf_{\theta\in\Theta_0(\phi)} \|\mathscr{H}(\theta)\|_{L^2(\nu)} = \inf_{\theta\in\Theta_0(\phi)} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L^2(\nu)} \geq \sqrt{\mathcal{L}(\phi + t_n h)},
\end{aligned} \tag{B.2}$$

where the second inequality follows from Assumption 2.8.

By Lemma B.1 and the fact that $\phi \in \mathbb{D}_{\mathcal{L}0}$ and $h \in \mathbb{D}_{\mathcal{L}0}$, the map $\theta \mapsto \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|^2_{L^2(\nu)}$

is continuous at every $\theta \in \Theta$ for every $n \in \mathbb{Z}_+$. Since Θ and $\Theta_0(\phi)^{a(h)t_n}$ are compact sets in $\mathbb{R}^{d_\theta}$, it follows that

$$\begin{aligned}\mathcal{L}(\phi + t_n h) &= \min_{\theta\in\Theta} \|\Phi(\theta) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)} \\ &= \min\left\{ \inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)}, \min_{\theta\in\Theta\cap\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)} \right\}.\end{aligned}$$

This, together with (B.2), implies that

$$\mathcal{L}(\phi + t_n h) = \min_{\theta\in\Theta\cap\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)}.$$

For every $a > 0$, let $V(a) = \left\{v \in \mathbb{R}^{d_\theta} : \|v\|_2 \le a\right\}$. For every $\theta \in \Theta_0(\phi)$ and every $a > 0$, define

$$V_n(a, \theta) = \{v \in V(a) : \theta + t_n v \in \Theta\}.$$

It is easy to show that (with the compactness of $\Theta_0(\phi)$)

$$\bigcup_{\theta\in\Theta_0(\phi)} \bigcup_{v\in V_n(a(h),\theta)} \{\theta + t_n v\} = \Theta \cap \Theta_0(\phi)^{a(h)t_n}.$$

Therefore,

$$\mathcal{L}(\phi + t_n h) = \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta + t_n v)\|^2_{L^2(\nu)}.$$

Note that $0 \in V_n(a(h), \theta)$. Then for every $\theta_0 \in \Theta_0(\phi)$,

$$\begin{aligned}&\left|\mathcal{L}(\phi + t_n h) - \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)}\right| \\ =& \left| \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta + t_n v)\|_{L^2(\nu)} \right. \\ &\left. + \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta)\|_{L^2(\nu)} \right| \\ &\cdot \left| \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta + t_n v)\|_{L^2(\nu)} \right. \\ &\left. - \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta)\|_{L^2(\nu)} \right| \\ \le& 2\|\Phi(\theta_0) + t_n \mathscr{H}(\theta_0)\|_{L^2(\nu)} \sup_{\theta\in\Theta_0(\phi)} \sup_{v\in V_n(a(h),\theta)} t_n \|\mathscr{H}(\theta + t_n v) - \mathscr{H}(\theta)\|_{L^2(\nu)} \\ \le& 2t_n^2 \|h\|_\infty \sup_{\theta_1,\theta_2\in\Theta:\|\theta_1-\theta_2\|_2\le a(h)t_n} \|\mathscr{H}(\theta_1) - \mathscr{H}(\theta_2)\|_{L^2(\nu)} = o(t_n^2),\end{aligned}$$

where the last equality follows from the definition of $\mathbb{D}_{\mathcal{L}0}$ and the compactness of Θ.

For every $\theta \in \Theta$, define $\Phi'(\theta) : \mathbb{R} \to \mathbb{R}^{d_\theta}$ such that

$$\Phi'(\theta)(x) = -\left.\frac{\partial(G \circ g)(z, \vartheta)}{\partial\vartheta}\right|_{(z,\vartheta)=(x,\theta)} \quad \text{for every } x \in \mathbb{R}.$$

Using an argument similar to the previous result, we have

$$\begin{aligned}&\left| \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta)\|^2_{L^2(\nu)} \right. \\ &\left. - \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \left\|\Phi(\theta) + t_n \left[\Phi'(\theta)\right]^\intercal v + t_n \mathscr{H}(\theta)\right\|^2_{L^2(\nu)} \right| \\ \le& 2t_n^2 \|h\|_\infty \sup_{\theta\in\Theta_0(\phi)} \sup_{v\in V_n(a(h),\theta)} \left\| \frac{\Phi(\theta + t_n v) - \Phi(\theta)}{t_n} - \left[\Phi'(\theta)\right]^\intercal v \right\|_{L^2(\nu)}.\end{aligned}$$

Then Assumption 2.7 implies that for all $\theta \in \Theta_0(\phi)$ and all $v \in V_n(a(h), \theta)$,

$$\begin{aligned}
&\left\| \frac{\Phi(\theta + t_n v) - \Phi(\theta)}{t_n} - \left[\Phi'(\theta)\right]^{\mathsf{T}} v \right\|_{L^2(\nu)}^2 \\
&= \int_{\mathbb{R}} \left[\frac{G(g(x, \theta + t_n v)) - G(g(x,\theta))}{t_n} - \left(\left. \frac{\partial (G \circ g)(z, \vartheta)}{\partial \vartheta} \right|_{(z,\vartheta)=(x,\theta)} \right)^{\mathsf{T}} v \right]^2 \mathrm{d}\nu(x) \\
&= \int_{\mathbb{R}} \left[\frac{t_n}{2} v^{\mathsf{T}} \left(\left. \frac{\partial^2 (G \circ g)(z, \vartheta)}{\partial \vartheta \partial \vartheta^{\mathsf{T}}} \right|_{(z,\vartheta)=(x,\theta + t_n^*(x) v)} \right) v \right]^2 \mathrm{d}\nu(x) \\
&\leq \frac{a(h)^4 t_n^2}{4} \int_{\mathbb{R}} \sup_{\theta^* \in \Theta} \left\| \left. \frac{\partial^2 (G \circ g)(z, \vartheta)}{\partial \vartheta \partial \vartheta^{\mathsf{T}}} \right|_{(z,\vartheta)=(x,\theta^*)} \right\|_2^2 \mathrm{d}\nu(x) = O(t_n^2),
\end{aligned}$$

where $0 \leq t_n^*(x) \leq t_n$ for all x and all n, and the last inequality follows from the property of the ℓ^2 operator norm. Then it follows that

$$\sup_{\theta \in \Theta_0(\phi)} \sup_{v \in V_n(a(h), \theta)} \left\| \frac{\Phi(\theta + t_n v) - \Phi(\theta)}{t_n} - \left[\Phi'(\theta)\right]^{\mathsf{T}} v \right\|_{L^2(\nu)} = o(1).$$

Since $\Theta_0(\phi) \subset \mathrm{int}(\Theta)$, for sufficiently large n, we have $V_n(a(h), \theta) = V(a(h))$. Combining the above results yields

$$\left| \mathcal{L}(\phi + t_n h_n) - t_n^2 \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V(a(h))} \left\| \left[\Phi'(\theta)\right]^{\mathsf{T}} v + \mathscr{H}(\theta) \right\|_{L^2(\nu)}^2 \right| = o\left(t_n^2\right).$$

This completes the proof. □

Proof of Proposition 2.2: Note that both $\ell^\infty(\mathbb{R} \times \Theta)$ and $\mathbb{R}$ are normed spaces. By Lemma 2.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$. Lemma 2.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\ell^\infty(\mathbb{R} \times \Theta)$ as $n \to \infty$ and $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. Therefore, Assumptions 2.1(i), 2.1(ii), 2.2(i), and 2.2(ii) of Chen and Fang (2019) are satisfied. The desired result follows from Theorem 2.1 of Chen and Fang (2019), the facts that $\mathcal{L}(\phi) = 0$ and $\mathcal{L}'_\phi(h) = 0$ for all $h \in \ell^\infty(\mathbb{R} \times \Theta)$ whenever $\phi \in \mathbb{D}_0$, and that $(\widehat{\phi}_n - \phi) \in \ell^\infty(\mathbb{R} \times \Theta)$ for every $n \in \mathbb{Z}_+$. □

Proof of Lemma 2.4: Note that both $\ell^\infty(\mathbb{R} \times \Theta)$ and $\mathbb{R}$ are normed spaces, and by Lemma 2.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at $\phi \in \mathbb{D}_0$ tangentially to $\mathbb{D}_{\mathcal{L}0}$. By Lemma 2.2, $\mathcal{L}'_\phi(h) = 0$ for all $h \in \ell^\infty(\mathbb{R} \times \Theta)$ whenever $\phi \in \mathbb{D}_0$. Lemma 2.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\ell^\infty(\mathbb{R} \times \Theta)$ as $n \to \infty$, where $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. Therefore, Assumptions 2.1, 2.2(i), 2.2(ii), and 3.5 of Chen and Fang (2019) hold, and the desired result follows from Proposition 3.1 of Chen and Fang (2019). □

Proof of Lemma 2.5: Firstly, we prove the results for the independent samples case. Define

$$\mathcal{F} = \left\{ \mathbb{1}_{(-\infty, x]} : x \in \mathbb{R} \right\} \text{ and } \mathcal{G} = \left\{ \mathbb{1}_{(-\infty, g(x,\theta)]} : (x, \theta) \in \mathbb{R} \times \Theta \right\}.$$

Define $\widehat{\mathcal{X}}_{n_1}$, $\widehat{\mathcal{Y}}_{n_2}$, $\mathcal{X}$, and $\mathcal{Y}$ as

$$\widehat{\mathcal{X}}_{n_1}(f) = \frac{1}{n_1} \sum_{i=1}^{n_1} f(X_i), \widehat{\mathcal{Y}}_{n_2}(f) = \frac{1}{n_2} \sum_{i=1}^{n_2} f(Y_i), \mathcal{X}(f) = \mathbb{E}[f(X_i)], \text{ and } \mathcal{Y}(f) = \mathbb{E}[f(Y_i)]$$

for all measurable f. Let $(W_{11}, \ldots, W_{1n_1})$ and $(W_{21}, \ldots, W_{2n_2})$ be two independent random vectors of multinomial weights independent of $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$. Define $\widehat{\mathcal{X}}_{n_1}^*$ and $\widehat{\mathcal{Y}}_{n_2}^*$ to be

the bootstrap versions of $\widehat{\mathcal{X}}_{n_1}$ and $\widehat{\mathcal{Y}}_{n_2}$, respectively, with

$$\widehat{\mathcal{X}}^*_{n_1}(f) = \frac{1}{n_1}\sum_{i=1}^{n_1} f\left(X_i^*\right) = \frac{1}{n_1}\sum_{i=1}^{n_1} W_{1i} f\left(X_i\right) \text{ and } \widehat{\mathcal{Y}}^*_{n_2}(f) = \frac{1}{n_2}\sum_{i=1}^{n_2} f\left(Y_i^*\right) = \frac{1}{n_2}\sum_{i=1}^{n_2} W_{2i} f\left(Y_i\right)$$

for every measurable f. By Example 2.5.4 of van der Vaart and Wellner (1996), the class $\mathcal{F}$ is Donsker. Because $\mathcal{G} \subset \mathcal{F}$, by Theorem 2.10.1 of van der Vaart and Wellner (1996), the class $\mathcal{G}$ is also Donsker. Therefore,

$$\sqrt{n_1}\left(\widehat{\mathcal{X}}_{n_1} - \mathcal{X}\right) \rightsquigarrow \mathbb{X} \text{ in } \ell^\infty(\mathcal{F}) \text{ and } \sqrt{n_2}\left(\widehat{\mathcal{Y}}_{n_2} - \mathcal{Y}\right) \rightsquigarrow \mathbb{Y} \text{ in } \ell^\infty(\mathcal{G})$$

as $n \to \infty$, where $\mathbb{X}$ and $\mathbb{Y}$ are independent centered Gaussian processes with $\mathbb{E}\left[\mathbb{X}\left(f_1\right)\mathbb{X}\left(f_2\right)\right] = \mathcal{X}\left(f_1 f_2\right) - \mathcal{X}\left(f_1\right)\mathcal{X}\left(f_2\right)$ and $\mathbb{E}\left[\mathbb{Y}\left(h_1\right)\mathbb{Y}\left(h_2\right)\right] = \mathcal{Y}\left(h_1 h_2\right) - \mathcal{Y}\left(h_1\right)\mathcal{Y}\left(h_2\right)$ for all $f_1, f_2 \in \mathcal{F}$ and all $h_1, h_2 \in \mathcal{G}$. Moreover, because $\mathcal{F}$ and $\mathcal{G}$ are classes of indicator functions, we have

$$\mathcal{X}\left[\sup_{f\in\mathcal{F}}\left(f - \mathcal{X}(f)\right)^2\right] \leq 1 \text{ and } \mathcal{Y}\left[\sup_{h\in\mathcal{G}}\left(h - \mathcal{Y}(h)\right)^2\right] \leq 1.$$

By Theorem 2.7 of Kosorok (2008), it follows that

$$\sqrt{n_1}\left(\widehat{\mathcal{X}}^*_{n_1} - \widehat{\mathcal{X}}_{n_1}\right) \overset{\text{a.s.}}{\rightsquigarrow} \mathbb{X} \text{ and } \sqrt{n_2}\left(\widehat{\mathcal{Y}}^*_{n_2} - \widehat{\mathcal{Y}}_{n_2}\right) \overset{\text{a.s.}}{\rightsquigarrow} \mathbb{Y}$$

as $n \to \infty$.

It is easy to show that

$$\widehat{F}_{n_1}(x) = \widehat{\mathcal{X}}_{n_1}\left(\mathbb{1}_{(-\infty,x]}\right), \left(\widehat{G}_{n_2} \circ g\right)(x,\theta) = \widehat{\mathcal{Y}}_{n_2}\left(\mathbb{1}_{(-\infty,g(x,\theta)]}\right),$$

$$\widehat{F}^*_{n_1}(x) = \widehat{\mathcal{X}}^*_{n_1}\left(\mathbb{1}_{(-\infty,x]}\right), \text{ and } \left(\widehat{G}^*_{n_2} \circ g\right)(x,\theta) = \widehat{\mathcal{Y}}^*_{n_2}\left(\mathbb{1}_{(-\infty,g(x,\theta)]}\right)$$

for every $x \in \mathbb{R}$ and every $\theta \in \Theta$. Define $W_F(x) = \mathbb{X}(\mathbb{1}_{(-\infty,x]})$ and $W(x,\theta) = \mathbb{Y}(\mathbb{1}_{(-\infty,g(x,\theta)]})$ for every $x \in \mathbb{R}$ and every $\theta \in \Theta$. By Lemma A.2, we have that

$$\sqrt{n_1}\left(\widehat{F}^*_{n_1} - \widehat{F}_{n_1}\right) \overset{\text{a.s.}}{\rightsquigarrow} W_F \text{ and } \sqrt{n_2}\left(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g\right) \overset{\text{a.s.}}{\rightsquigarrow} W. \tag{B.3}$$

For notational simplicity, let $\mathcal{Z}_n = \left\{\left\{X_i\right\}_{i=1}^{n_1}, \left\{Y_i\right\}_{i=1}^{n_2}\right\}$ and $\mathbb{A} = \ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R}\times\Theta)$. By the independence between the two weight vectors, we have that for all bounded, nonnegative, Lipschitz functions Γ_1 and Γ_2 on $\ell^\infty(\mathbb{R})$ and $\ell^\infty(\mathbb{R}\times\Theta)$, respectively,

$$\begin{aligned}&\mathbb{E}\left[\Gamma_1\left(\sqrt{n_1}\left(\widehat{F}^*_{n_1} - \widehat{F}_{n_1}\right)\right)\Gamma_2\left(\sqrt{n_2}\left(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g\right)\right)\middle|\mathcal{Z}_n\right]\\ =&\mathbb{E}\left[\Gamma_1\left(\sqrt{n_1}\left(\widehat{F}^*_{n_1} - \widehat{F}_{n_1}\right)\right)\middle|\mathcal{Z}_n\right]\cdot\mathbb{E}\left[\Gamma_2\left(\sqrt{n_2}\left(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g\right)\right)\middle|\mathcal{Z}_n\right].\end{aligned}$$

Then with the independence of W_F and W, by Example 1.4.6 of van der Vaart and Wellner (1996) and Assumption 2.6 of this paper,

$$\sup_{\Gamma\in\text{BL}_1(\mathbb{A})}\left|\mathbb{E}\left[\Gamma\left(\begin{bmatrix}\sqrt{T_n}\left(\widehat{F}^*_{n_1} - \widehat{F}_{n_1}\right)\\ \sqrt{T_n}\left(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g\right)\end{bmatrix}\right)\middle|\mathcal{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\begin{bmatrix}\sqrt{1-\lambda}W_F\\ \sqrt{\lambda}W\end{bmatrix}\right)\right]\right| \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$.

Define a map $\mathcal{I} : \mathbb{A} \to \ell^\infty(\mathbb{R}\times\Theta)$ such that $\mathcal{I}\left(f,h\right)(x,\theta) = f(x) - h(x,\theta)$ for every $(f,h) \in \mathbb{A}$ and every $(x,\theta) \in \mathbb{R}\times\Theta$. As shown in the proof of Lemma 2.1,

$$\left\|\mathcal{I}\left(f_1,h_1\right) - \mathcal{I}\left(f_2,h_2\right)\right\|_\infty \leq \left\|\left(f_1,h_1\right) - \left(f_2,h_2\right)\right\|_{\mathbb{A}}$$

for all $\left(f_1,h_1\right), \left(f_2,h_2\right) \in \mathbb{A}$, where $\left\|(f,h)\right\|_{\mathbb{A}} = \left\|f\right\|_\infty + \left\|h\right\|_\infty$ for every $(f,h) \in \mathbb{A}$. This implies the Lipschitz continuity of $\mathcal{I}$. By the proof similar to that of Proposition 10.7(ii) of Kosorok (2008), we can show that

$$\sup_{\Gamma\in\text{BL}_1(\ell^\infty(\mathbb{R}\times\Theta))}\left|\mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right)\middle|\mathcal{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\tilde{\mathbb{G}}_0\right)\right]\right| \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$, where $\tilde{\mathbb{G}}_0 = \sqrt{1-\lambda} W_F - \sqrt{\lambda} W$. By the properties of W_F and W, it can be verified that $\tilde{\mathbb{G}}_0$ is equivalent to $\mathbb{G}_0$ in law. The desired result follows from Lemma 1.9.2(i) of van der Vaart and Wellner (1996).

Because $\mathcal{F}$ and $\mathcal{G}$ are both Donsker, by Theorem 2.6 of Kosorok (2008), $\sqrt{n_1}(\widehat{\mathcal{X}}^*_{n_1} - \widehat{\mathcal{X}}_{n_1})$ and $\sqrt{n_2}(\widehat{\mathcal{Y}}^*_{n_2} - \widehat{\mathcal{Y}}_{n_2})$ are asymptotically measurable. By Lemma A.2, $\sqrt{n_1}(\widehat{F}^*_{n_1} - \widehat{F}_{n_1})$ and $\sqrt{n_2}(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g)$ are asymptotically measurable. By (B.3) and the asymptotic measurability of $\sqrt{n_1}(\widehat{F}^*_{n_1} - \widehat{F}_{n_1})$ and $\sqrt{n_2}(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g)$, we can show that $\sqrt{n_1}(\widehat{F}^*_{n_1} - \widehat{F}_{n_1})$ and $\sqrt{n_2}(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g)$ are asymptotically tight. Then by Lemma 1.4.4 of van der Vaart and Wellner (1996), $(\sqrt{n_1}(\widehat{F}^*_{n_1} - \widehat{F}_{n_1}), \sqrt{n_2}(\widehat{G}^*_{n_2} \circ g - \widehat{G}_{n_2} \circ g))$ is asymptotically measurable. The asymptotic measurability of $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ follows from the continuity of $\mathcal{I}$.

Secondly, we prove the results for the matched pairs case. In the exposition below, $\{V_i\}_{i=1}^{n_1}$, $h_{x,\theta}$, $\mathcal{H}$, $\widehat{\mathcal{V}}_{n_1}$, $\mathcal{V}$, and $\mathbb{V}$ are defined as in the second part of the proof of Lemma 2.1. Let $(W_1, \ldots, W_{n_1})$ be a random vector of multinomial weights independent of $\{(X_i, Y_i)\}_{i=1}^{n_1}$. Define $\widehat{\mathcal{V}}^*_{n_1}$ as

$$\widehat{\mathcal{V}}^*_{n_1}(f) = \frac{1}{n_1} \sum_{i=1}^{n_1} f(V_i^*) = \frac{1}{n_1} \sum_{i=1}^{n_1} W_i f(V_i)$$

for every measurable f, where $V_i^* = (X_i^*, Y_i^*)$ for $i = 1, \ldots, n_1$. Because $-1 \leq f \leq 1$ holds for all $f \in \mathcal{H}$, we have

$$\mathcal{V}\left[\sup_{f \in \mathcal{H}} (f - \mathcal{V}(f))^2\right] \leq 4.$$

As shown in the second part of the proof of Lemma 2.1, the class $\mathcal{H}$ is Donsker. By Theorem 2.7 of Kosorok (2008), we have

$$\sqrt{n_1}\left(\widehat{\mathcal{V}}^*_{n_1} - \widehat{\mathcal{V}}_{n_1}\right) \overset{\text{a.s.}}{\rightsquigarrow} \mathbb{V}.$$

By Theorem 2.6 of Kosorok (2008), $\sqrt{n_1}(\widehat{\mathcal{V}}^*_{n_1} - \widehat{\mathcal{V}}_{n_1})$ is asymptotically measurable as $n \to \infty$. Under the assumption of matched pairs,

$$\widehat{\phi}^*_n(x, \theta) = \frac{1}{n_1} \sum_{i=1}^{n_1} \left[\mathbb{1}_{(-\infty, x]}(X_i^*) - \mathbb{1}_{(-\infty, g(x,\theta)]}(Y_i^*)\right] = \frac{1}{n_1} \sum_{i=1}^{n_1} h_{x,\theta}(V_i^*) = \widehat{\mathcal{V}}^*_{n_1}(h_{x,\theta})$$

for every $(x, \theta) \in \mathbb{R} \times \Theta$. The results follow from Lemma A.2 and the fact that $T_n = n_1/2$. □

Proof of Proposition 2.3: Note that both $\ell^\infty(\mathbb{R} \times \Theta)$ and $\mathbb{R}$ are normed spaces, and by Lemma 2.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at $\phi \in \mathbb{D}_0$ tangentially to $\mathbb{D}_{\mathcal{L}0}$. Lemma 2.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\ell^\infty(\mathbb{R} \times \Theta)$ as $n \to \infty$ and $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. By Lemma B.1, $\mathbb{D}_{\mathcal{L}0}$ is closed under vector addition, that is, $\varphi_1 + \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$ whenever $\varphi_1, \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$. By construction, the random weights used to construct the bootstrap samples are independent of the data set and $\widehat{\phi}^*_n$ is a measurable function of the random weights. By Lemma 2.5,

$$\sup_{\Gamma \in \mathrm{BL}_1(\ell^\infty(\mathbb{R} \times \Theta))} \left|\mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right) \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right] - \mathbb{E}\left[\Gamma(\mathbb{G}_0)\right]\right| \xrightarrow{\mathbb{P}} 0,$$

and $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ is asymptotically measurable as $n \to \infty$. Lemma 2.4 establishes the consistency of $\widehat{\mathcal{L}}''_n$ for $\mathcal{L}''_\phi$. Therefore, Assumptions 2.1(i), 2.1(ii), 2.2, 3.1, 3.2, and 3.4 of Chen and Fang (2019) are satisfied, and the result follows from Theorem 3.3 of Chen and Fang (2019). □

Proof of Theorem 2.1: (i). Let Ψ be the cumulative distribution function of $\mathcal{L}''_\phi(\mathbb{G}_0)$ and $c_{1-\alpha}$ be the $1-\alpha$ quantile for $\mathcal{L}''_\phi(\mathbb{G}_0)$. Define

$$\widehat{\Psi}_n(c) = \mathbb{P}\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right] \leq c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right)$$

for every $n \in \mathbb{Z}_+$ and every $c \in \mathbb{R}$. Let $C_\Psi \subset \mathbb{R}$ be the set of continuity points of Ψ, and $\mathbb{L}(\mathbb{R})$ be the set of all Lipschitz continuous functions $\Gamma : \mathbb{R} \to [0,1]$. For every $\Gamma \in \mathbb{L}(\mathbb{R})$, let $M = 1 \vee L_\Gamma$, where L_Γ is the Lipschitz constant of Γ. Then $\Gamma/M \in \mathrm{BL}_1(\mathbb{R})$, and by Proposition 2.3,

$$\mathbb{E}\left[\Gamma\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right]\right) \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right] \xrightarrow{\mathbb{P}} \mathbb{E}\left[\Gamma\left(\mathcal{L}''_\phi(\mathbb{G}_0)\right)\right] \tag{B.4}$$

as $n \to \infty$ if H_0 is true. By Lemma 10.11(i) of Kosorok (2008), we have $\widehat{\Psi}_n(c) \xrightarrow{\mathbb{P}} \Psi(c)$ for every $c \in C_\Psi$. Because Ψ is strictly increasing and continuous at $c_{1-\alpha}$ and a cumulative distribution function has at most countably many discontinuity points, for every $\varepsilon > 0$, there exist $a_1, a_2 \in C_\Psi$ such that $a_1 < c_{1-\alpha} < a_2$, $|a_1 - c_{1-\alpha}| < \varepsilon$, and $|a_2 - c_{1-\alpha}| < \varepsilon$. Let

$$\delta = \frac{1}{2}\left[\left|\Psi(a_1) - (1-\alpha)\right| \wedge \left|\Psi(a_2) - (1-\alpha)\right|\right].$$

From the definition of $\widehat{c}_{1-\alpha,n}$, it follows that

$$\begin{aligned}
\mathbb{P}\left(|\widehat{c}_{1-\alpha,n} - c_{1-\alpha}| > \varepsilon\right) &\leq \mathbb{P}\left(\widehat{c}_{1-\alpha,n} < a_1\right) + \mathbb{P}\left(\widehat{c}_{1-\alpha,n} > a_2\right) \\
&\leq \mathbb{P}\left(\widehat{\Psi}_n(a_1) \geq 1-\alpha\right) + \mathbb{P}\left(\widehat{\Psi}_n(a_2) < 1-\alpha\right) \\
&\leq \mathbb{P}\left(\left|\widehat{\Psi}_n(a_1) - \Psi(a_1)\right| > \delta\right) + \mathbb{P}\left(\left|\widehat{\Psi}_n(a_2) - \Psi(a_2)\right| > \delta\right),
\end{aligned}$$

and the last line converges to 0 since $\widehat{\Psi}_n(a_1) \xrightarrow{\mathbb{P}} \Psi(a_1)$ and $\widehat{\Psi}_n(a_2) \xrightarrow{\mathbb{P}} \Psi(a_2)$ as $n \to \infty$. This implies that $\widehat{c}_{1-\alpha,n} \xrightarrow{\mathbb{P}} c_{1-\alpha}$ as $n \to \infty$.

By Proposition 2.2 of this paper, if H_0 is true ($\phi \in \mathbb{D}_0$), then $T_n\mathcal{L}(\widehat{\phi}_n) \rightsquigarrow \mathcal{L}''_\phi(\mathbb{G}_0)$ as $n \to \infty$. By Lemma 2.8(i) of van der Vaart (1998), $T_n\mathcal{L}(\widehat{\phi}_n) - \widehat{c}_{1-\alpha,n} \rightsquigarrow \mathcal{L}''_\phi(\mathbb{G}_0) - c_{1-\alpha}$ as $n \to \infty$. Since the cumulative distribution function of $\mathcal{L}''_\phi(\mathbb{G}_0)$ is continuous and strictly increasing at $c_{1-\alpha}$, the cumulative distribution function of $\mathcal{L}''_\phi(\mathbb{G}_0) - c_{1-\alpha}$ is continuous at 0 and $\mathbb{P}(\mathcal{L}''_\phi(\mathbb{G}_0) - c_{1-\alpha} > 0) = \alpha$. By Lemma 2.2(i) (portmanteau) of van der Vaart (1998), we have

$$\lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}\left(\widehat{\phi}_n\right) > \widehat{c}_{1-\alpha,n}\right) = \lim_{n\to\infty} \mathbb{P}\left(T_n\mathcal{L}\left(\widehat{\phi}_n\right) - \widehat{c}_{1-\alpha,n} > 0\right) = \mathbb{P}\left(\mathcal{L}''_\phi(\mathbb{G}_0) - c_{1-\alpha} > 0\right) = \alpha.$$

(ii). For all $\theta \in \Theta$ and all $\phi_1, \phi_2 \in \mathbb{D}_\mathcal{L}$,

$$\begin{aligned}
&\left|\int_{\mathbb{R}} [\phi_1(x,\theta)]^2 \, \mathrm{d}\nu(x) - \int_{\mathbb{R}} [\phi_2(x,\theta)]^2 \, \mathrm{d}\nu(x)\right| \\
&\leq \int_{\mathbb{R}} \left|[\phi_1(x,\theta) + \phi_2(x,\theta)]\,[\phi_1(x,\theta) - \phi_2(x,\theta)]\right| \mathrm{d}\nu(x) \leq \left(\|\phi_1\|_\infty + \|\phi_2\|_\infty\right)\|\phi_1 - \phi_2\|_\infty.
\end{aligned}$$

This implies that

$$\left|\mathcal{L}(\phi_1) - \mathcal{L}(\phi_2)\right| \leq \left(\|\phi_1\|_\infty + \|\phi_2\|_\infty\right)\|\phi_1 - \phi_2\|_\infty.$$

Therefore, the function $\varphi \mapsto \mathcal{L}(\varphi)$ is continuous. If H_0 is false, then by Lemma B.2 of this paper, Theorem 1.9.5 (continuous mapping), and Lemma 1.9.2(iii) of van der Vaart and Wellner (1996), we have $\mathcal{L}(\widehat{\phi}_n) \to \mathcal{L}(\phi) > 0$ almost surely as $n \to \infty$. Thus, $1/(\tau_n T_n \mathcal{L}(\widehat{\phi}_n)) \to 0$ almost surely. By definition, for every $h \in \ell^\infty(\mathbb{R} \times \Theta)$,

$$\left|\tau_n\widehat{\mathcal{L}}''_n(h)\right| = \left|\frac{\mathcal{L}\left(\widehat{\phi}_n + \tau_n h\right) - \mathcal{L}\left(\widehat{\phi}_n\right)}{\tau_n}\right| \leq \sup_{\theta\in\Theta}\left|\frac{1}{\tau_n}\int_{\mathbb{R}} \left\{2\tau_n h(x,\theta)\,\widehat{\phi}_n(x,\theta) + \tau_n^2 h^2(x,\theta)\right\} \mathrm{d}\nu(x)\right|$$

$$\leq 2\|h\widehat{\phi}_n\|_\infty + \tau_n \left\|h^2\right\|_\infty .$$

Define $\widehat{\mathcal{L}}_{bn}(h) = 2\|h\widehat{\phi}_n\|_\infty + \tau_n\|h^2\|_\infty$ for every $h \in \ell^\infty(\mathbb{R}\times\Theta)$. In the proof of Lemma 2.5, we have shown that for almost every sequence $\{X_i\}_{i=1}^\infty$ and every sequence $\{Y_i\}_{i=1}^\infty$, $\sqrt{T_n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ converges weakly to $\mathbb{G}_0$. By Lemma B.2, $\widehat{\phi}_n \to \phi$ almost surely. Thus, for almost every sequence $\{X_i\}_{i=1}^\infty$ and every sequence $\{Y_i\}_{i=1}^\infty$, $\widehat{\mathcal{L}}_{bn}(\sqrt{T_n}(\widehat{\phi}_n^* - \widehat{\phi}_n))$ converges weakly to $2\left\|\mathbb{G}_0\cdot\phi\right\|_\infty$ by Lemmas 1.9.2(i) and 1.10.2(iii), Example 1.4.7, and Theorem 1.3.6 (continuous mapping) of van der Vaart and Wellner (1996). Also, we have that

$$\left|\tau_n\widehat{\mathcal{L}}_n''\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right)\right| \leq \widehat{\mathcal{L}}_{bn}\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right).$$

By definition,

$$\widehat{c}_{1-\alpha,n} = \inf\left\{c\in\mathbb{R} : \mathbb{P}\left(\widehat{\mathcal{L}}_n''\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \leq c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \geq 1-\alpha\right\}.$$

Then it follows that

$$\begin{aligned}
\tau_n\widehat{c}_{1-\alpha,n} &= \inf\left\{\tau_n c\in\mathbb{R} : \mathbb{P}\left(\tau_n\widehat{\mathcal{L}}_n''\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \leq \tau_n c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \geq 1-\alpha\right\} \\
&= \inf\left\{c\in\mathbb{R} : \mathbb{P}\left(\tau_n\widehat{\mathcal{L}}_n''\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \leq c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \geq 1-\alpha\right\} \\
&\leq \inf\left\{c\in\mathbb{R} : \mathbb{P}\left(\widehat{\mathcal{L}}_{bn}\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \leq c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \geq 1-\alpha\right\}.
\end{aligned}$$

Define

$$\widehat{c}'_{1-\alpha,n} = \inf\left\{c\in\mathbb{R} : \mathbb{P}\left(\widehat{\mathcal{L}}_{bn}\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \leq c \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right) \geq 1-\alpha\right\}.$$

By proof similar to that of (i), we can show that $\widehat{c}'_{1-\alpha,n} \xrightarrow{\mathbb{P}} c'_{1-\alpha}$, where $c'_{1-\alpha}$ is the $1-\alpha$ quantile of $2\left\|\mathbb{G}_0\cdot\phi\right\|_\infty$.[1] Then it follows that

$$\mathbb{P}(T_n\mathcal{L}(\widehat{\phi}_n) > \widehat{c}_{1-\alpha,n}) = \mathbb{P}\left(\tau_n\widehat{c}_{1-\alpha,n} \Big/ \left[\tau_n T_n\mathcal{L}(\widehat{\phi}_n)\right] < 1\right) \geq \mathbb{P}\left(\widehat{c}'_{1-\alpha,n} \Big/ \left[\tau_n T_n\mathcal{L}(\widehat{\phi}_n)\right] < 1\right) \to 1.$$

This completes the proof. □

Proof of Proposition 2.4: For notational simplicity, we mainly consider the matched pairs case. The proof for independent samples can be achieved analogously. By (11), we have that for every n, $F_n(x) = \mathbb{P}(X_i \leq x) = P_n 1_{(-\infty,x]\times\mathbb{R}}$ and $G_n(x) = \mathbb{P}(Y_i \leq x) = P_n 1_{\mathbb{R}\times(-\infty,x]}$. The subscript n in F_n and G_n indicates that F_n and G_n may change as n increases. Let $\mathcal{P}$ be the set of all probability measures on $\mathscr{B}_{\mathbb{R}^2}$, where $\mathscr{B}_{\mathbb{R}^2}$ is the collection of Borel sets in $\mathbb{R}^2$. Define four classes of functions on $\mathbb{R}^2$:

$$\mathcal{H}_1 = \left\{\mathbb{1}_{(-\infty,x]\times\mathbb{R}} : x\in\mathbb{R}\right\}, \mathcal{H}_2 = \left\{-\mathbb{1}_{\mathbb{R}\times(-\infty,x]} : x\in\mathbb{R}\right\}, \mathcal{H}_+ = \left\{h_1+h_2 : h_1\in\mathcal{H}_1, h_2\in\mathcal{H}_2\right\},$$

and $\mathcal{H} = \left\{\mathbb{1}_{(-\infty,x]\times\mathbb{R}} - \mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]} : (x,\theta)\in\mathbb{R}\times\Theta\right\} \subset \mathcal{H}_+$.

It can be shown that $\mathcal{H}_1$ and $\mathcal{H}_2$ are both VC classes with VC indices $V(\mathcal{H}_1) = V(\mathcal{H}_2) = 2$. Let $N(\varepsilon, \mathcal{H}_j, L^r(Q))$ denote the covering number under the $L^r(Q)$ norm for $\mathcal{H}_j$ with $j\in\{1,2\}$ and all $\varepsilon > 0$. By Theorem 2.6.7 of van der Vaart and Wellner (1996) with envelop function $F = 1$ and $r \geq 1$, we have that for each j and every probability measure Q,

$$N(\varepsilon, \mathcal{H}_j, L^r(Q)) \leq K_j 2(16e)^2(1/\varepsilon)^2$$

for universal constants $K_1, K_2 \geq 1$ and every $\varepsilon\in(0,1)$. Then we have that

$$N(\varepsilon, \mathcal{H}_+, L^r(Q)) \leq N(\varepsilon/2, \mathcal{H}_1, L^r(Q))\cdot N(\varepsilon/2, \mathcal{H}_2, L^r(Q)) = K_1K_2 4(16e)^4(4/\varepsilon)^4. \tag{B.5}$$

[1] Here, we implicitly assume that the CDF of $2\|\mathbb{G}_0\cdot\phi\|_\infty$ is continuous and strictly increasing at $c'_{1-\alpha}$.

By the construction of $\mathcal{H}_+$, $F = 1$ is a measurable envelop function such that $\int F^2 \, \mathrm{d}Q < \infty$ and $\lim_{M\to\infty} \sup_{Q\in\mathcal{P}} \int F^2 \cdot 1\{F > M\} \, \mathrm{d}Q = 0$. For all $Q \in \mathcal{P}$ and all $\varepsilon \geq 2$,

$$N(\varepsilon \|F\|_{L^2(Q)}, \mathcal{H}_+, L^2(Q)) = 1.$$

Let $\mathcal{Q}$ denote the set of finitely discrete probability measures. Then we have that

$$\int_0^\infty \sup_{Q\in\mathcal{Q}} \sqrt{\log N(\varepsilon \|F\|_{L^2(Q)}, \mathcal{H}_+, L^2(Q))} \, \mathrm{d}\varepsilon = \int_0^2 \sup_{Q\in\mathcal{Q}} \sqrt{\log N(\varepsilon \|F\|_{L^2(Q)}, \mathcal{H}_+, L^2(Q))} \, \mathrm{d}\varepsilon < \infty.$$

Note that the set of rational numbers is dense in $\mathbb{R}$. Following the strategy of the proof of Lemma C.5 in Sun (2021), it can be shown that $\mathcal{H}_+$ is Donsker and pre-Gaussian uniformly in $\mathcal{P}$ by Theorem 2.8.3 of van der Vaart and Wellner (1996). By the construction of $\mathcal{H}$ with $\mathcal{H} \subset \mathcal{H}_+$, we can show that $\mathcal{H}$ is Donsker and pre-Gaussian uniformly in $\mathcal{P}$.

Let $\mathcal{Q}_n$ denote the collection of all possible realizations of empirical measures of n observations. Then by (B.5), it can be shown that $\sup_{Q\in\mathcal{Q}_n} \log N(\varepsilon \|F\|_{L^1(Q)}, \mathcal{H}_+, L^1(Q)) = o(n)$ with the envelop function $F = 1$. By Theorem 2.8.1 of van der Vaart and Wellner (1996), $\mathcal{H}_+$ is Glivenko–Cantelli uniformly in $\mathcal{P}$. By the construction of $\mathcal{H}$ with $\mathcal{H} \subset \mathcal{H}_+$, we can show that $\mathcal{H}$ is Glivenko–Cantelli uniformly in $\mathcal{P}$.

Let $\widehat{P}_n$ denote the empirical probability measure of P_n such that for every measurable function h,

$$\widehat{P}_n h = \frac{1}{n_1} \sum_{i=1}^{n_1} h(X_i, Y_i).$$

Also, we let $\widehat{P}_n^*$ denote the bootstrap empirical probability measure of P_n such that for every measurable function h,

$$\widehat{P}_n^* h = \frac{1}{n_1} \sum_{i=1}^{n_1} h(X_i^*, Y_i^*).$$

Since P_n may change as $n \to \infty$, we now define $\phi_n(x, \theta) = F_n(x) - G_n(g(x, \theta)) = P_n h_{x,\theta}$, where $h_{x,\theta} = \mathbb{1}_{(-\infty,x]\times\mathbb{R}} - \mathbb{1}_{\mathbb{R}\times(-\infty,g(x,\theta)]} \in \mathcal{H}$. Note that $\phi(x, \theta) = F(x) - G(g(x, \theta)) = P h_{x,\theta}$ for every $(x, \theta) \in \mathbb{R} \times \Theta$. Also, we note that $\widehat{\phi}_n(x, \theta) = \widehat{P}_n h_{x,\theta}$ and $\widehat{\phi}_n^*(x, \theta) = \widehat{P}_n^* h_{x,\theta}$ for every (x, θ).

Since $\sup_{h\in\mathcal{H}} |Ph| < \infty$ and $\sup_{h\in\mathcal{H}} |P_n h^2| < \infty$, under Assumption 2.11, by Theorem 3.10.12 of van der Vaart and Wellner (1996) we have that $\sqrt{n_1}(\widehat{P}_n - P)$ converges under P_n in distribution in $\ell^\infty(\mathcal{H})$ to the process $h \mapsto \mathbb{G}(h) + Phv_0$, where $\mathbb{G}$ is a tight Brownian bridge. Also, by Theorem 3.10.12 of van der Vaart and Wellner (1996), $\sup_{h\in\mathcal{H}} |\sqrt{n_1}(P_n - P)h - Phv_0| \to 0$.

Following the strategy of the proof of Lemma C.16 in Sun (2021), we can show that $\sqrt{T_n}(\widehat{P}_n^* - \widehat{P}_n) \overset{\text{a.s.}}{\rightsquigarrow} \sqrt{1/2} \cdot \mathbb{G}$ under Assumption 2.11, and that $\sqrt{T_n}(\widehat{P}_n^* - \widehat{P}_n)$ is asymptotically measurable. Since $\widehat{\phi}_n^*(x, \theta) - \widehat{\phi}_n(x, \theta) = (\widehat{P}_n^* - \widehat{P}_n)h_{x,\theta}$ for every (x, θ), by Lemma A.2 we have that

$$\sup_{\Gamma\in \mathrm{BL}_1(\ell^\infty(\mathbb{R}\times\Theta))} \left| \mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n \right) \right) \middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2} \right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0 \right) \right] \right| \xrightarrow{\mathbb{P}} 0,$$

where $\mathbb{G}_0(x, \theta) = \sqrt{1/2} \cdot \mathbb{G}(h_{x,\theta})$ for every $(x, \theta) \in \mathbb{R} \times \Theta$, and $\sqrt{T_n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ is asymptotically measurable as $n \to \infty$. Since $\widehat{\phi}_n(x, \theta) - \phi(x, \theta) = (\widehat{P}_n - P)h_{x,\theta}$ for every (x, θ), by Lemma A.1 we have that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0 + \psi_P$, where $\psi_P(x, \theta) = \sqrt{1/2} \cdot Ph_{x,\theta}v_0$ for all (x, θ). Since $\mathbb{G}$ is tight, we can show that $\mathbb{G}_0 + \psi_P$ is tight. Note that P satisfies H_0. Then by a proof similar to that of Lemma 2.4, we can show that for every sequence $\{h_n\} \subset \ell^\infty(\mathbb{R} \times \Theta)$ and every $h \in \mathbb{D}_{\mathcal{L}0}$

such that $h_n \to h$ in $\ell^\infty(\mathbb{R}\times\Theta)$ as $n\to\infty$, we have

$$\widehat{\mathcal{L}}''_n(h_n) \xrightarrow{\mathbb{P}} \mathcal{L}''_\phi(h) \text{ as } n\to\infty.$$

Next, by a proof similar to that of Proposition 2.3, we have that

$$\sup_{\Gamma\in \mathrm{BL}_1(\mathbb{R})} \left| \mathbb{E}\left[\Gamma\left(\widehat{\mathcal{L}}''_n\left[\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right]\right)\middle| \{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\right] - \mathbb{E}\left[\Gamma\left(\mathcal{L}''_\phi(\mathbb{G}_0)\right)\right]\right| \xrightarrow{\mathbb{P}} 0$$

as $n\to\infty$.

Following a strategy similar to that of the proof of Theorem 2.1, we can show that $\widehat{c}_{1-\alpha,n} \xrightarrow{\mathbb{P}} c_{1-\alpha}$ and $T_n\mathcal{L}(\widehat{\phi}_n) \rightsquigarrow \mathcal{L}''_\phi(\mathbb{G}_0+\psi_P)$ as $n\to\infty$.[2] The result follows from Lemma 1.10.2(iii), Example 1.4.7, and Theorems 1.3.6 and 1.3.4(ii) of van der Vaart and Wellner (1996). □

B.2 Proofs for Section 3

Lemma B.3: For every $k\in\{1,\dots,K\}$, if $\varphi_1,\varphi_2\in\mathbb{D}_{\mathcal{L}k}$, then $a_1\varphi_1+a_2\varphi_2\in\mathbb{D}_{\mathcal{L}k}$ for all $a_1,a_2\in\mathbb{R}$, and the functions

$$\theta_k \mapsto \int_{\mathbb{R}} [\varphi_1(x,\theta_k)]^2 \,\mathrm{d}\nu(x) \text{ and } \theta_k \mapsto \int_{\mathbb{R}} \varphi_1(x,\theta_k)\varphi_2(x,\theta_k)\,\mathrm{d}\nu(x)$$

are continuous at every $\theta_k\in\Theta_k$.

Proof of Lemma B.3: The proof is similar to that of Lemma B.1. □

Proof of Proposition 3.1: If $F(x) = G_k(g_k(x,\theta_k))$ for all $x\in\mathbb{R}$ with some $\theta_k\in\Theta_k$ for all $k\in\{1,\dots,K\}$, then (13) holds trivially.

Next, we show that (13) implies (12). Recall that μ is the Lebesgue measure on $(\mathbb{R},\mathscr{B}_\mathbb{R})$. Since $G_k\in\mathcal{C}_\mathrm{b}(\mathbb{R})$, Assumption 3.4 implies that $G_k\circ g_k\in\mathbb{D}_{\mathcal{L}k}$ and hence $\phi_k\in\mathbb{D}_{\mathcal{L}k}$. By Lemma B.3, the function $\theta_k\mapsto\int_\mathbb{R}[F(x)-G_k(g_k(x,\theta_k))]^2\,\mathrm{d}\nu(x)$ is continuous on Θ_k. Thus, the function $(\theta_1,\dots,\theta_K)\mapsto\int_\mathbb{R}\sum_{k=1}^K[F(x)-G_k(g_k(x,\theta_k))]^2\,\mathrm{d}\nu(x)$ is continuous on Θ. By Assumption 3.3, there exists $\theta_0\in\Theta$ with $\theta_0=(\theta_{01},\dots,\theta_{0K})$ such that

$$\int_\mathbb{R}\sum_{k=1}^K[F(x)-G_k(g_k(x,\theta_{0k}))]^2\,\mathrm{d}\nu(x) = \inf_{(\theta_1,\dots,\theta_K)\in\Theta}\int_\mathbb{R}\sum_{k=1}^K[F(x)-G_k(g_k(x,\theta_k))]^2\,\mathrm{d}\nu(x) = 0. \tag{B.6}$$

Define $A_k=\{x\in\mathbb{R}: F(x)\neq G_k(g_k(x,\theta_{0k}))\}$ for every $k\in\{1,\dots,K\}$. Then (B.6) implies that $\nu(A_k)=0$ by Proposition 2.16 of Folland (1999). By the assumption that $\mu\ll\nu$, $\mu(A_k)=0$. We now claim that $A_k=\varnothing$. Otherwise, there is an $x_0\in\mathbb{R}$ such that $F(x_0)\neq G_k(g_k(x_0,\theta_{0k}))$. Since both F and G_k are continuous and $g_k(\cdot,\theta_{0k})$ is continuous, there exists $\delta>0$ such that $F(x)\neq G_k(g_k(x,\theta_{0k}))$ for all $x\in[x_0,x_0+\delta]$. This contradicts $\mu(A_k)=0$. Therefore, we have $F(x)=G_k(g_k(x,\theta_{0k}))$ for all $x\in\mathbb{R}$ and all k. □

Lemma B.4: Under Assumptions 3.5 and 3.6, we have

$$\lim_{n\to\infty}\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\left\|\widehat{\phi}_n(x,\theta)-\phi(x,\theta)\right\|_2 = 0 \text{ almost surely.}$$

Proof of Lemma B.4: By Theorem 19.1 of van der Vaart (1998) and Assumption 3.6, we

[2] Similarly to Theorem 2.1, here we implicitly assume that the cumulative distribution function of $\mathcal{L}''_\phi(\mathbb{G}_0)$ is strictly increasing and continuous at its $1-\alpha$ quantile $c_{1-\alpha}$.

have

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|\widehat{F}_{n_x}(x)-F(x)|=0 \text{ almost surely,}$$

$$\text{and } \lim_{n\to\infty}\sup_{x\in\mathbb{R}}|\widehat{G}_{n_k}(x)-G_k(x)|=0 \text{ almost surely for every } k.$$

Note that for every $(x,\theta_k)\in\mathbb{R}\times\Theta_k$,

$$\left|\widehat{G}_{n_k}\left(g_k(x,\theta_k)\right)-G_k\left(g_k(x,\theta_k)\right)\right|\leq\sup_{z\in\mathbb{R}}\left|\widehat{G}_{n_k}(z)-G_k(z)\right|,$$

which implies

$$\lim_{n\to\infty}\sup_{(x,\theta_k)\in\mathbb{R}\times\Theta_k}\left|\widehat{G}_{n_k}\left(g_k(x,\theta_k)\right)-G_k\left(g_k(x,\theta_k)\right)\right|=0 \text{ almost surely.}$$

Then the desired result follows from the definitions of $\widehat{\phi}_n$ and ϕ. □

Proof of Lemma 3.1: By Theorem 19.3 of van der Vaart (1998), we have

$$\sqrt{n_x}(\widehat{F}_{n_x}-F)\rightsquigarrow\mathbb{W}_F \text{ in } \ell^\infty(\mathbb{R}), \text{ and for all } k\in\{1,\ldots,K\}, \sqrt{n_k}(\widehat{G}_{n_k}-G_k)\rightsquigarrow\mathbb{W}_{G_k} \text{ in } \ell^\infty(\mathbb{R})$$

as $n\to\infty$, where $\mathbb{W}_F,\mathbb{W}_{G_1},\ldots,\mathbb{W}_{G_K}$ are jointly independent. Define classes of indicator functions

$$\mathcal{G}_0=\left\{\mathbb{1}_{(-\infty,x]}:x\in\mathbb{R}\right\} \text{ and } \mathcal{G}_k=\left\{\mathbb{1}_{(-\infty,g_k(x,\theta_k)]}:(x,\theta_k)\in\mathbb{R}\times\Theta_k\right\} \text{ for all } k.$$

Let $\widehat{\mathcal{Y}}_{n_k}$ be a stochastic process and $\mathcal{Y}_k$ be a real valued function such that

$$\widehat{\mathcal{Y}}_{n_k}(f)=\frac{1}{n_k}\sum_{i=1}^{n_k}f\left(Y_{ki}\right) \text{ and } \mathcal{Y}_k(f)=\mathbb{E}\left[f\left(Y_{ki}\right)\right]$$

for all measurable f. By Example 2.5.4 of van der Vaart and Wellner (1996), $\mathcal{G}_0$ is a Donsker class. Therefore, $\sqrt{n_k}(\widehat{\mathcal{Y}}_{n_k}-\mathcal{Y}_k)\rightsquigarrow\mathbb{Y}_k$ in $\ell^\infty(\mathcal{G}_0)$ as $n\to\infty$, where $\mathbb{Y}_k$ is a tight measurable centered Gaussian process. Since $\mathcal{G}_k\subset\mathcal{G}_0$, it follows that for every $h\in\mathcal{C}_{\mathrm{b}}(\ell^\infty(\mathcal{G}_k))$, $h\in\mathcal{C}_{\mathrm{b}}(\ell^\infty(\mathcal{G}_0))$ and

$$\mathbb{E}[h(\sqrt{n_k}(\widehat{\mathcal{Y}}_{n_k}-\mathcal{Y}_k))]\to\mathbb{E}[h(\mathbb{Y}_k)],$$

which implies that $\sqrt{n_k}(\widehat{\mathcal{Y}}_{n_k}-\mathcal{Y}_k)\rightsquigarrow\mathbb{Y}_k$ in $\ell^\infty(\mathcal{G}_k)$ as $n\to\infty$. It is easy to show that $\widehat{G}_{n_k}\circ g_k(x,\theta_k)=\widehat{\mathcal{Y}}_{n_k}(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]})$ and $G_k\circ g_k(x,\theta_k)=\mathcal{Y}_k(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]})$ for every $(x,\theta_k)\in\mathbb{R}\times\Theta_k$. Define a random element $W_k\in\ell^\infty(\mathbb{R}\times\Theta_k)$ such that $W_k(x,\theta_k)=\mathbb{Y}_k(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]})$ for all $(x,\theta_k)\in\mathbb{R}\times\Theta_k$. By Lemma A.1, $\sqrt{n_k}(\widehat{G}_{n_k}\circ g_k-G_k\circ g_k)\rightsquigarrow W_k$ in $\ell^\infty(\mathbb{R}\times\Theta_k)$ as $n\to\infty$. Let $\lambda_{-x}=\prod_{k=1}^K\lambda_k$ and $\lambda_{-k}=(\lambda_x\cdot\prod_{j=1}^K\lambda_j)/\lambda_k$. By the joint independence of the samples, Assumption 3.6 of this paper, and Example 1.4.6 of van der Vaart and Wellner (1996), we have the joint weak convergence

$$\begin{bmatrix}\sqrt{T_n}\left(\widehat{F}_{n_x}-F\right)\\ \sqrt{T_n}\left(\widehat{G}_{n_1}\circ g_1-G_1\circ g_1\right)\\ \vdots\\ \sqrt{T_n}\left(\widehat{G}_{n_K}\circ g_K-G_K\circ g_K\right)\end{bmatrix}\rightsquigarrow\begin{bmatrix}\sqrt{\lambda_{-x}}\mathbb{W}_F\\ \sqrt{\lambda_{-1}}W_1\\ \vdots\\ \sqrt{\lambda_{-K}}W_K\end{bmatrix} \text{ in } \ell^\infty(\mathbb{R})\times\ell^\infty(\mathbb{R}\times\Theta_1)\times\cdots\times\ell^\infty(\mathbb{R}\times\Theta_K)$$

as $n\to\infty$, where $\mathbb{W}_F,W_1,\ldots,W_K$ are jointly independent. Define

$$\mathbb{A}=\ell^\infty(\mathbb{R})\times\ell^\infty(\mathbb{R}\times\Theta_1)\times\cdots\times\ell^\infty(\mathbb{R}\times\Theta_K) \text{ and } \mathbb{B}=\ell^\infty(\mathbb{R}\times\Theta_1)\times\cdots\times\ell^\infty(\mathbb{R}\times\Theta_K).$$

Define the norms $\|\cdot\|_{\mathbb{A}}$ and $\|\cdot\|_{\mathbb{B}}$ on $\mathbb{A}$ and $\mathbb{B}$, respectively, with $\|(f,h_1,\ldots,h_K)\|_{\mathbb{A}}=\|f\|_\infty+\sum_{k=1}^K\|h_k\|_\infty$ for every $(f,h_1,\ldots,h_K)\in\mathbb{A}$ and $\|(h_1,\ldots,h_K)\|_{\mathbb{B}}=\sum_{k=1}^K\|h_k\|_\infty$ for every

$(h_1, \ldots, h_K) \in \mathbb{B}$. Let $\mathcal{I} : \mathbb{A} \to \mathbb{B}$ be such that

$$\mathcal{I}(f, h_1, \ldots, h_K)(x, \theta) = (f(x) - h_1(x, \theta_1), \ldots, f(x) - h_K(x, \theta_K))$$

for every $(f, h_1, \ldots, h_K) \in \mathbb{A}$ and every $(x, \theta) \in \mathbb{R} \times \Theta$ with $\theta = (\theta_1, \ldots, \theta_K)$ and $\Theta = \Theta_1 \times \cdots \times \Theta_K$. Note that

$$\left\|\mathcal{I}\left(f', h_1', \ldots, h_K'\right) - \mathcal{I}(f, h_1, \ldots, h_K)\right\|_{\mathbb{B}} = \sum_{k=1}^{K} \sup_{(x,\theta_k) \in \mathbb{R} \times \Theta_k} \left|f'(x) - h_k'(x, \theta_k) - f(x) + h_k(x, \theta_k)\right|$$

$$\leq K \sup_{x \in \mathbb{R}} \left|f'(x) - f(x)\right| + \sum_{k=1}^{K} \sup_{(x,\theta_k) \in \mathbb{R} \times \Theta_k} \left|h_k'(x, \theta_k) - h_k(x, \theta_k)\right|$$

for all $(f', h_1', \ldots, h_K'), (f, h_1, \ldots, h_K) \in \mathbb{A}$, and therefore $\mathcal{I}$ is continuous. The weak convergence of $\sqrt{T_n}(\widehat{\phi}_n - \phi)$ to a tight random element $\mathbb{G}_0 = \mathcal{I}(\sqrt{\lambda_{-x}}\mathbb{W}_F, \sqrt{\lambda_{-1}}W_1, \ldots, \sqrt{\lambda_{-K}}W_K)$ follows from Theorem 1.3.6 (continuous mapping) of van der Vaart and Wellner (1996). Furthermore, by the proof similar to that of Lemma 2.1, $\mathbb{P}(\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}) = 1$. □

Proof of Lemma 3.2: Define a map $\mathcal{S} : \mathbb{D}_{\mathcal{L}} \to \ell^\infty(\Theta)$ such that for every $\varphi \in \mathbb{D}_{\mathcal{L}}$ and every $\theta \in \Theta$ with $\varphi = (\varphi_1, \ldots, \varphi_K)$ and $\theta = (\theta_1, \ldots, \theta_K)$,

$$\mathcal{S}(\varphi)(\theta) = \int_{\mathbb{R}} \sum_{k=1}^{K} [\varphi_k(x, \theta_k)]^2 \, \mathrm{d}\nu(x).$$

We show that the Hadamard directional derivative of $\mathcal{S}$ at $\phi \in \mathbb{D}_{\mathcal{L}}$ is

$$\mathcal{S}_\phi'(h)(\theta) = \int_{\mathbb{R}} 2 \sum_{k=1}^{K} \phi_k(x, \theta_k) h_k(x, \theta_k) \, \mathrm{d}\nu(x) \text{ for all } h \in \mathbb{D}_{\mathcal{L}0} \text{ with } h = (h_1, \ldots, h_K).$$

Because $F, G_k \in \mathcal{C}_{\mathrm{b}}(\mathbb{R})$, by Assumption 3.4 and Lemma B.3, $\mathcal{S}(\phi) \in \mathcal{C}(\Theta)$. Indeed, for all sequences $\{h_n\}_{n=1}^{\infty} \subset \prod_{k=1}^{K} \ell^\infty(\mathbb{R} \times \Theta_k)$ with $h_n = (h_{n1}, \ldots, h_{nK})$ and $\{t_n\}_{n=1}^{\infty} \subset \mathbb{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_{\mathcal{L}0}$ as $n \to \infty$ with $h = (h_1, \ldots, h_K)$, and $\phi + t_n h_n \in \mathbb{D}_{\mathcal{L}}$ for all n, we have that $M = \max_{k \in \{1,\ldots,K\}} \sup_{n \in \mathbb{Z}_+} \|h_{nk}\|_\infty < \infty$, and

$$\begin{aligned}
&\sup_{\theta \in \Theta} \left| \frac{\mathcal{S}(\phi + t_n h_n)(\theta) - \mathcal{S}(\phi)(\theta)}{t_n} - \mathcal{S}_\phi'(h)(\theta) \right| \\
&= \sup_{\theta \in \Theta} \left| \sum_{k=1}^{K} \int_{\mathbb{R}} t_n h_{nk}^2(x, \theta_k) + 2\phi_k(x, \theta_k) \left[h_{nk}(x, \theta_k) - h_k(x, \theta_k)\right] \, \mathrm{d}\nu(x) \right| \\
&\leq \sum_{k=1}^{K} \int_{\mathbb{R}} t_n M^2 + 2 \|\phi_k\|_\infty \|h_{nk} - h_k\|_\infty \, \mathrm{d}\nu(x) \to 0,
\end{aligned}$$

since $t_n \downarrow 0$ and $h_n \to h$ in $\prod_{k=1}^{K} \ell^\infty(\mathbb{R} \times \Theta_k)$ as $n \to \infty$.

Define a function $\mathcal{R}$ such that for every $\psi \in \mathcal{C}(\Theta)$, $\mathcal{R}(\psi) = \inf_{\theta \in \Theta} \psi(\theta)$. By Lemma S.4.9 of Fang and Santos (2019), $\mathcal{R}$ is Hadamard directionally differentiable at every $\psi \in \mathcal{C}(\Theta)$ tangentially to $\mathcal{C}(\Theta)$ with the Hadamard directional derivative

$$\mathcal{R}_\psi'(f) = \inf_{\theta \in \Theta_0^*(\psi)} f(\theta) \text{ for all } f \in \mathcal{C}(\Theta),$$

where $\Theta_0^*(\psi) = \arg\min_{\theta \in \Theta} \psi(\theta)$.

Note that $\mathcal{L}(\varphi) = \mathcal{R}[\mathcal{S}(\varphi)] = \mathcal{R} \circ \mathcal{S}(\varphi)$ for every $\varphi \in \mathbb{D}_{\mathcal{L}}$. By Proposition 3.6(i) of Shapiro (1990), $\mathcal{L}$ is Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the Hadamard

directional derivative

$$\mathcal{L}'_\phi(h) = \mathcal{R}'_{\mathcal{S}(\phi)}\left[\mathcal{S}'_\phi(h)\right] = \inf_{\theta\in\Theta_0^*(\mathcal{S}(\phi))}\int_{\mathbb{R}} 2\sum_{k=1}^{K}\phi_k(x,\theta_k)h_k(x,\theta_k)\,\mathrm{d}\nu(x) \text{ for all } h\in\mathbb{D}_{\mathcal{L}0}$$

with $h=(h_1,\ldots,h_K)$.

Since $\Theta_0^*(\mathcal{S}(\phi)) = \arg\min_{\theta\in\Theta}\int_{\mathbb{R}}\sum_{k=1}^{K}\left[\phi_k(x,\theta_k)\right]^2\,\mathrm{d}\nu(x)$, the desired result follows.

Now we turn to the degeneracy of $\mathcal{L}'_\phi$ under the condition that $\phi\in\mathbb{D}_0$. If $\phi\in\mathbb{D}_0$, for every $\theta\in\Theta_0(\phi)$ with $\theta=(\theta_1,\ldots,\theta_K)$, we have

$$\int_{\mathbb{R}}\sum_{k=1}^{K}\left[\phi_k(x,\theta_k)\right]^2\,\mathrm{d}\nu(x) = 0,$$

and consequently $\phi_k(x,\theta_k)=0$ holds for ν-almost every x and every k. Therefore, $\mathcal{L}'_\phi(h)=0$ for every $h\in\prod_{k=1}^{K}\ell^\infty(\mathbb{R}\times\Theta_k)$ whenever $\phi\in\mathbb{D}_0$. □

Proof of Lemma 3.3: For every k, define $\Phi_k:\Theta_k\to L^2(\nu)$ such that $\Phi_k(\theta_k)(x)=\phi_k(x,\theta_k)$ for every $(x,\theta_k)\in\mathbb{R}\times\Theta_k$. Define $\Phi:\Theta\to\prod_{k=1}^{K}L^2(\nu)$ such that for every $\theta\in\Theta$ with $\theta=(\theta_1,\ldots,\theta_K)$, $\Phi(\theta)=(\Phi_1(\theta_1),\ldots,\Phi_K(\theta_K))$. Then it is easy to show that

$$\mathcal{L}(\phi)=\inf_{\theta\in\Theta}\int_{\mathbb{R}}\sum_{k=1}^{K}\left[\phi_k(x,\theta_k)\right]^2\,\mathrm{d}\nu(x)=\inf_{\theta\in\Theta}\sum_{k=1}^{K}\left\|\Phi_k(\theta_k)\right\|^2_{L^2(\nu)}=\inf_{\theta\in\Theta}\left\|\Phi(\theta)\right\|^2_{L^2_K(\nu)}=0,$$

and $\Theta_0(\phi)=\{\theta\in\Theta:\sum_{k=1}^{K}\left\|\Phi_k(\theta_k)\right\|^2_{L^2(\nu)}=0\}=\Theta_0$. Consider all sequences $\{t_n\}_{n=1}^{\infty}\subset\mathbb{R}_+$ and $\{h_n\}_{n=1}^{\infty}\subset\prod_{k=1}^{K}\ell^\infty(\mathbb{R}\times\Theta_k)$ such that $t_n\downarrow 0$, $h_n\to h\in\mathbb{D}_{\mathcal{L}0}$ as $n\to\infty$, and $\phi+t_nh_n\in\mathbb{D}_{\mathcal{L}}$ for all n, where $h_n=(h_{n1},\ldots,h_{nK})$ and $h=(h_1,\ldots,h_K)$. For notational simplicity, for every k and every n, define $\mathscr{H}_{nk}:\Theta_k\to L^2(\nu)$ such that $\mathscr{H}_{nk}(\theta_k)(x)=h_{nk}(x,\theta_k)$ for every $(x,\theta_k)\in\mathbb{R}\times\Theta_k$, and define $\mathscr{H}_k:\Theta_k\to L^2(\nu)$ such that $\mathscr{H}_k(\theta_k)(x)=h_k(x,\theta_k)$ for every $(x,\theta_k)\in\mathbb{R}\times\Theta_k$. For every $\theta\in\Theta$ with $\theta=(\theta_1,\ldots,\theta_K)$, let $\mathscr{H}_n(\theta)=(\mathscr{H}_{n1}(\theta_1),\ldots,\mathscr{H}_{nK}(\theta_K))$ and $\mathscr{H}(\theta)=(\mathscr{H}_1(\theta_1),\ldots,\mathscr{H}_K(\theta_K))$. Since $h_n\to h\in\mathbb{D}_{\mathcal{L}0}\subset\prod_{k=1}^{K}\ell^\infty(\mathbb{R}\times\Theta_k)$, it follows that $\max_{k\in\{1,\ldots,K\}}(\|h_k\|_\infty\vee\sup_{n\in\mathbb{Z}_+}\|h_{nk}\|_\infty)=M_1$ for some $M_1<\infty$. Then we have that

$$\begin{aligned}
&\left|\mathcal{L}\left(\phi+t_nh_n\right)-\mathcal{L}\left(\phi+t_nh\right)\right| = \left|\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}_n(\theta)\right\|^2_{L^2_K(\nu)}-\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}(\theta)\right\|^2_{L^2_K(\nu)}\right| \\
&= \left|\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}_n(\theta)\right\|_{L^2_K(\nu)}+\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}(\theta)\right\|_{L^2_K(\nu)}\right| \\
&\quad\cdot\left|\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}_n(\theta)\right\|_{L^2_K(\nu)}-\inf_{\theta\in\Theta}\left\|\Phi(\theta)+t_n\mathscr{H}(\theta)\right\|_{L^2_K(\nu)}\right| \\
&\le \left|\inf_{\theta\in\Theta_0(\phi)}\left\|\Phi(\theta)+t_n\mathscr{H}_n(\theta)\right\|_{L^2_K(\nu)}+\inf_{\theta\in\Theta_0(\phi)}\left\|\Phi(\theta)+t_n\mathscr{H}(\theta)\right\|_{L^2_K(\nu)}\right| \\
&\quad\cdot\left(t_n\sup_{\theta\in\Theta}\left\|\mathscr{H}_n(\theta)-\mathscr{H}(\theta)\right\|_{L^2_K(\nu)}\right) \\
&= O\left(t_n^2\left\{\sum_{k=1}^{K}\left\|h_{nk}-h_k\right\|^2_\infty\right\}^{1/2}\right)=o\left(t_n^2\right),
\end{aligned}$$

where the inequality follows from the Lipschitz continuity of the supremum map and the triangle inequality, and the third equality follows from the fact that $\Phi(\theta)=0$ ν-almost everywhere for every $\theta\in\Theta_0(\phi)$.

Then for the h, pick an $a(h)>0$ such that $Ca(h)^\kappa=3(\sum_{k=1}^{K}\|h_k\|^2_\infty)^{1/2}$, where C and κ are

defined as in Assumption 3.8. For sufficiently large $n \in \mathbb{Z}_+$ such that $t_n^\kappa \geq t_n$, we have that

$$
\begin{aligned}
&\inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)} \\
\geq &\inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta)\|_{L_K^2(\nu)} + \inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \left[-t_n \|\mathscr{H}(\theta)\|_{L_K^2(\nu)}\right] \\
= &\inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta)\|_{L_K^2(\nu)} - \sup_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} t_n \|\mathscr{H}(\theta)\|_{L_K^2(\nu)} \\
\geq &\, C\,(a(h)t_n)^\kappa - t_n \sup_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\mathscr{H}(\theta)\|_{L_K^2(\nu)} \geq 3\left(\sum_{k=1}^K \|h_k\|_\infty^2\right)^{1/2} t_n^\kappa - t_n\left(\sum_{k=1}^K \|h_k\|_\infty^2\right)^{1/2} \\
> &\, t_n \inf_{\theta\in\Theta_0(\phi)} \|\mathscr{H}(\theta)\|_{L_K^2(\nu)} = \inf_{\theta\in\Theta_0(\phi)} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)} \geq \sqrt{\mathcal{L}\,(\phi + t_n h)},
\end{aligned} \tag{B.7}
$$

where the second inequality follows from Assumption 3.8.

By Lemma B.3 and the fact that $\phi \in \mathbb{D}_{\mathcal{L}0}$ and $h \in \mathbb{D}_{\mathcal{L}0}$, the map $\theta \mapsto \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)}^2$ is continuous at every $\theta \in \Theta$ for every $n \in \mathbb{Z}_+$. Since Θ and $\Theta_0(\phi)^{a(h)t_n}$ are compact sets in $\prod_{k=1}^K \mathbb{R}^{d_{\theta_k}}$, it follows that

$$
\begin{aligned}
&\mathcal{L}\,(\phi + t_n h) = \min_{\theta\in\Theta} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)}^2 \\
= &\min\left\{\inf_{\theta\in\Theta\setminus\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)}^2,\ \min_{\theta\in\Theta\cap\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)}^2\right\}.
\end{aligned}
$$

This, together with (B.7), implies that

$$
\mathcal{L}\,(\phi + t_n h) = \min_{\theta\in\Theta\cap\Theta_0(\phi)^{a(h)t_n}} \|\Phi(\theta) + t_n\mathscr{H}(\theta)\|_{L_K^2(\nu)}^2.
$$

For every $a > 0$, let $V(a) = \{v \in \prod_{k=1}^K \mathbb{R}^{d_{\theta_k}} : \|v\|_{K2} \leq a\}$. For every $\theta \in \Theta_0(\phi)$ and every $a > 0$, define

$$
V_n(a,\theta) = \{v \in V(a) : \theta + t_n v \in \Theta\}.
$$

It is easy to show that (with the compactness of $\Theta_0(\phi)$)

$$
\bigcup_{\theta\in\Theta_0(\phi)} \bigcup_{v\in V_n(a(h),\theta)} \{\theta + t_n v\} = \Theta \cap \Theta_0(\phi)^{a(h)t_n}.
$$

Therefore,

$$
\mathcal{L}\,(\phi + t_n h) = \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta + t_n v)\|_{L_K^2(\nu)}^2.
$$

Note that $0 \in V_n(a(h),\theta)$. Then for every $\theta_0 \in \Theta_0(\phi)$,

$$
\begin{aligned}
&\left|\mathcal{L}\,(\phi + t_n h) - \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta)\|_{L_K^2(\nu)}^2\right| \\
= &\left|\inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta + t_n v)\|_{L_K^2(\nu)}\right. \\
&\left. + \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta)\|_{L_K^2(\nu)}\right| \\
&\cdot \left|\inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta + t_n v)\|_{L_K^2(\nu)}\right. \\
&\left. - \inf_{\theta\in\Theta_0(\phi)} \inf_{v\in V_n(a(h),\theta)} \|\Phi\,(\theta + t_n v) + t_n\mathscr{H}\,(\theta)\|_{L_K^2(\nu)}\right| \\
\leq &\, 2\,\|\Phi\,(\theta_0) + t_n\mathscr{H}\,(\theta_0)\|_{L_K^2(\nu)} \sup_{\theta\in\Theta_0(\phi)} \sup_{v\in V_n(a(h),\theta)} t_n \|\mathscr{H}\,(\theta + t_n v) - \mathscr{H}(\theta)\|_{L_K^2(\nu)}
\end{aligned}
$$

$$\leq 2t_n^2 \left\{ \sum_{k=1}^{K} \|h_k\|_\infty^2 \right\}^{1/2} \sup_{\theta_1,\theta_2 \in \Theta : \|\theta_1 - \theta_2\|_{K2} \leq a(h) t_n} \|\mathscr{H}(\theta_1) - \mathscr{H}(\theta_2)\|_{L_K^2(\nu)} = o(t_n^2),$$

where the last equality follows from the definition of $\mathbb{D}_{\mathcal{L}0}$ and the compactness of Θ.

For every $\theta \in \Theta$ with $\theta = (\theta_1, \ldots, \theta_K)$, define $\Phi_k'(\theta_k) : \mathbb{R} \to \mathbb{R}^{d_{\theta_k}}$ such that

$$\Phi_k'(\theta_k)(x) = - \left. \frac{\partial (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k} \right|_{(z,\vartheta_k)=(x,\theta_k)} \quad \text{for every } x \in \mathbb{R}.$$

For every $\theta = (\theta_1, \ldots, \theta_K)$ and every $v = (v_1, \ldots, v_K)$, let

$$\Phi'(\theta, v)(x) = (\Phi_1'(\theta_1)(x)^\mathsf{T} v_1, \ldots, \Phi_K'(\theta_K)(x)^\mathsf{T} v_K)$$

for all x. Using an argument similar to the previous result, we have

$$\left| \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V_n(a(h),\theta)} \|\Phi(\theta + t_n v) + t_n \mathscr{H}(\theta)\|_{L_K^2(\nu)}^2 \right.$$
$$\left. - \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V_n(a(h),\theta)} \left\|\Phi(\theta) + t_n \Phi'(\theta, v) + t_n \mathscr{H}(\theta)\right\|_{L_K^2(\nu)}^2 \right|$$
$$\leq 2O(t_n^2) \sup_{\theta \in \Theta_0(\phi)} \sup_{v \in V_n(a(h),\theta)} \left\{ \sum_{k=1}^{K} \left\| \frac{\Phi_k(\theta_k + t_n v_k) - \Phi_k(\theta_k)}{t_n} - [\Phi_k'(\theta_k)]^\mathsf{T} v_k \right\|_{L^2(\nu)}^2 \right\}^{1/2}.$$

For every $\theta \in \Theta_0(\phi)$ and every $v \in V_n(a(h), \theta)$, Assumption 3.7 implies that

$$\left\| \frac{\Phi_k(\theta_k + t_n v_k) - \Phi_k(\theta_k)}{t_n} - \left[\Phi_k'(\theta_k)\right]^\mathsf{T} v_k \right\|_{L^2(\nu)}^2$$
$$= \int_{\mathbb{R}} \left[\frac{G_k(g_k(x, \theta_k + t_n v_k)) - G_k(g_k(x, \theta_k))}{t_n} - \left(\left. \frac{\partial (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k} \right|_{(z,\vartheta_k)=(x,\theta_k)} \right)^\mathsf{T} v_k \right]^2 \mathrm{d}\nu(x)$$
$$= \int_{\mathbb{R}} \left[\frac{t_n}{2} v_k^\mathsf{T} \left(\left. \frac{\partial^2 (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k \partial \vartheta_k^\mathsf{T}} \right|_{(z,\vartheta_k)=(x,\theta_k + t_{kn}^*(x) v_k)} \right) v_k \right]^2 \mathrm{d}\nu(x)$$
$$\leq \frac{a(h)^4 t_n^2}{4} \int_{\mathbb{R}} \sup_{\theta_k^* \in \Theta_k} \left\| \left. \frac{\partial^2 (G_k \circ g_k)(z, \vartheta_k)}{\partial \vartheta_k \partial \vartheta_k^\mathsf{T}} \right|_{(z,\vartheta_k)=(x,\theta_k^*)} \right\|_2^2 \mathrm{d}\nu(x) = O(t_n^2),$$

where $0 \leq t_{kn}^*(x) \leq t_n$ for all x, all n, and all k, and the last inequality follows from the property of the ℓ^2 operator norm. Then it follows that

$$\sup_{\theta \in \Theta_0(\phi)} \sup_{v \in V_n(a(h),\theta)} \left\{ \sum_{k=1}^{K} \left\| \frac{\Phi_k(\theta_k + t_n v_k) - \Phi_k(\theta_k)}{t_n} - \Phi_k'(\theta_k)^\mathsf{T} v_k \right\|_{L^2(\nu)}^2 \right\}^{1/2} = o(1).$$

Since $\Theta_0(\phi) \subset \mathrm{int}(\Theta)$, for sufficiently large n, we have $V_n(a(h), \theta) = V(a(h))$. Combining the above results yields

$$\left| \mathcal{L}(\phi + t_n h_n) - t_n^2 \inf_{\theta \in \Theta_0(\phi)} \inf_{v \in V(a(h))} \left\|\Phi'(\theta, v) + \mathscr{H}(\theta)\right\|_{L_K^2(\nu)}^2 \right| = o\left(t_n^2\right).$$

This completes the proof. □

Proof of Proposition 3.2: Note that both $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ and $\mathbb{R}$ are normed spaces. By Lemma 3.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at ϕ tangentially to $\mathbb{D}_{\mathcal{L}0}$. Lemma 3.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ as $n \to \infty$ and $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. Therefore, Assumptions 2.1(i), 2.1(ii), 2.2(i), and 2.2(ii) of Chen and Fang (2019) are satisfied. The desired result follows from Theorem 2.1 of Chen and Fang (2019), the fact that $\mathcal{L}(\phi) = 0$ and $\mathcal{L}_\phi'(h) = 0$ for all $h \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ whenever $\phi \in \mathbb{D}_0$,

and that $(\widehat{\phi}_n - \phi) \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ for every $n \in \mathbb{Z}_+$. □

Proof of Lemma 3.4: Note that both $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ and $\mathbb{R}$ are normed spaces, and by Lemma 3.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at $\phi \in \mathbb{D}_0$ tangentially to $\mathbb{D}_{\mathcal{L}0}$. By Lemma 3.2, $\mathcal{L}'_\phi(h) = 0$ for all $h \in \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ whenever $\phi \in \mathbb{D}_0$. Lemma 3.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ as $n \to \infty$ and $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. Therefore, Assumptions 2.1, 2.2(i), 2.2(ii), and 3.5 of Chen and Fang (2019) hold, and the desired result follows from Proposition 3.1 of Chen and Fang (2019). □

Proof of Lemma 3.5: Define

$$\mathcal{F} = \left\{\mathbb{1}_{(-\infty,x]} : x \in \mathbb{R}\right\} \text{ and } \mathcal{G}_k = \left\{\mathbb{1}_{(-\infty,g_k(x,\theta_k)]} : (x,\theta_k) \in \mathbb{R} \times \Theta_k\right\} \text{ for every } k.$$

Define $\widehat{\mathcal{X}}_{n_x}$, $\widehat{\mathcal{Y}}_{n_k}$, $\mathcal{X}$, and $\mathcal{Y}_k$ as

$$\widehat{\mathcal{X}}_{n_x}(f) = \frac{1}{n_x}\sum_{i=1}^{n_x} f\left(X_i\right), \widehat{\mathcal{Y}}_{n_k}(f) = \frac{1}{n_k}\sum_{i=1}^{n_k} f\left(Y_{ki}\right), \mathcal{X}(f) = \mathbb{E}\left[f\left(X_i\right)\right], \text{ and } \mathcal{Y}_k(f) = \mathbb{E}\left[f\left(Y_{ki}\right)\right]$$

for all measurable f. Let $\{W_{xi}\}_{i=1}^{n_x}, \{W_{1i}\}_{i=1}^{n_1}, \ldots, \{W_{Ki}\}_{i=1}^{n_K}$ be jointly independent random vectors of multinomial weights that are independent of $\{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}$. Define $\widehat{\mathcal{X}}^*_{n_x}$ and $\widehat{\mathcal{Y}}^*_{n_k}$ to be the bootstrap versions of $\widehat{\mathcal{X}}_{n_x}$ and $\widehat{\mathcal{Y}}_{n_k}$, respectively, with

$$\widehat{\mathcal{X}}^*_{n_x}(f) = \frac{1}{n_x}\sum_{i=1}^{n_x} f\left(X^*_i\right) = \frac{1}{n_x}\sum_{i=1}^{n_x} W_{xi} f\left(X_i\right) \text{ and } \widehat{\mathcal{Y}}^*_{n_k}(f) = \frac{1}{n_k}\sum_{i=1}^{n_k} f\left(Y^*_{ki}\right) = \frac{1}{n_k}\sum_{i=1}^{n_k} W_{ki} f\left(Y_{ki}\right)$$

for every measurable f. By Example 2.5.4 of van der Vaart and Wellner (1996), the class $\mathcal{F}$ is Donsker. Because $\mathcal{G}_k \subset \mathcal{F}$ for every k, by Theorem 2.10.1 of van der Vaart and Wellner (1996), the class $\mathcal{G}_k$ is also Donsker. Therefore,

$$\sqrt{n_x}\left(\widehat{\mathcal{X}}_{n_x} - \mathcal{X}\right) \rightsquigarrow \mathbb{X} \text{ in } \ell^\infty(\mathcal{F}) \text{ and } \sqrt{n_k}\left(\widehat{\mathcal{Y}}_{n_k} - \mathcal{Y}_k\right) \rightsquigarrow \mathbb{Y}_k \text{ in } \ell^\infty(\mathcal{G}_k)$$

as $n \to \infty$, where $\mathbb{X}, \mathbb{Y}_1, \ldots, \mathbb{Y}_K$ are jointly independent centered Gaussian processes. Moreover, because $\mathcal{F}$ and $\mathcal{G}_k$ are classes of indicator functions, we have that

$$\mathcal{X}\left[\sup_{f\in\mathcal{F}}\left(f - \mathcal{X}(f)\right)^2\right] \leq 1 \text{ and } \mathcal{Y}_k\left[\sup_{h\in\mathcal{G}_k}\left(h - \mathcal{Y}_k(h)\right)^2\right] \leq 1.$$

By Theorem 2.7 of Kosorok (2008), it follows that

$$\sqrt{n_x}\left(\widehat{\mathcal{X}}^*_{n_x} - \widehat{\mathcal{X}}_{n_x}\right) \overset{\text{a.s.}}{\rightsquigarrow} \mathbb{X} \text{ and } \sqrt{n_k}\left(\widehat{\mathcal{Y}}^*_{n_k} - \widehat{\mathcal{Y}}_{n_k}\right) \overset{\text{a.s.}}{\rightsquigarrow} \mathbb{Y}_k$$

as $n \to \infty$.

It is easy to show that

$$\widehat{F}_{n_x}(x) = \widehat{\mathcal{X}}_{n_x}\left(\mathbb{1}_{(-\infty,x]}\right), \left(\widehat{G}_{n_k} \circ g_k\right)(x,\theta_k) = \widehat{\mathcal{Y}}_{n_k}\left(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]}\right),$$

$$\widehat{F}^*_{n_x}(x) = \widehat{\mathcal{X}}^*_{n_x}\left(\mathbb{1}_{(-\infty,x]}\right), \text{ and } \left(\widehat{G}^*_{n_k} \circ g_k\right)(x,\theta_k) = \widehat{\mathcal{Y}}^*_{n_k}\left(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]}\right)$$

for every $x \in \mathbb{R}$, every $\theta_k \in \Theta_k$, and every k. Define $W_F(x) = \mathbb{X}(\mathbb{1}_{(-\infty,x]})$ and $W_k(x,\theta_k) = \mathbb{Y}_k(\mathbb{1}_{(-\infty,g_k(x,\theta_k)]})$ for every $x \in \mathbb{R}$ and every $\theta_k \in \Theta_k$. By Lemma A.2, we have that

$$\sqrt{n_x}\left(\widehat{F}^*_{n_x} - \widehat{F}_{n_x}\right) \overset{\text{a.s.}}{\rightsquigarrow} W_F \text{ and } \sqrt{n_k}\left(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k\right) \overset{\text{a.s.}}{\rightsquigarrow} W_k. \tag{B.8}$$

For simplicity, let $\mathcal{Z}_n = \{\{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K}\}$, $\mathbb{A} = \ell^\infty(\mathbb{R}) \times \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$, and $\mathbb{B} = \prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$. Define norms $\|\cdot\|_{\mathbb{A}}$ and $\|\cdot\|_{\mathbb{B}}$ on $\mathbb{A}$ and $\mathbb{B}$, respectively, such that for

every $(f,h) \in \mathbb{A}$ with $h = (h_1, \ldots, h_K)$ and every $w \in \mathbb{B}$ with $w = (w_1, \ldots, w_K)$,

$$\|(f,h)\|_{\mathbb{A}} = \|f\|_\infty + \sum_{k=1}^K \|h_k\|_\infty \text{ and } \|w\|_{\mathbb{B}} = \sum_{k=1}^K \|w_k\|_\infty.$$

By the joint independence of the weight vectors, we have that for all bounded, nonnegative, Lipschitz functions Γ_x on $\ell^\infty(\mathbb{R})$ and Γ_k on $\ell^\infty(\mathbb{R} \times \Theta_k)$,

$$\mathbb{E}\left[\Gamma_x\left(\sqrt{n_x}\left(\widehat{F}^*_{n_x} - \widehat{F}_{n_x}\right)\right)\prod_{k=1}^K \Gamma_k\left(\sqrt{n_k}\left(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k\right)\right)|\mathcal{Z}_n\right]$$

$$= \mathbb{E}\left[\Gamma_x\left(\sqrt{n_x}\left(\widehat{F}^*_{n_x} - \widehat{F}_{n_x}\right)\right)|\mathcal{Z}_n\right] \cdot \prod_{k=1}^K \mathbb{E}\left[\Gamma_k\left(\sqrt{n_k}\left(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k\right)\right)|\mathcal{Z}_n\right].$$

Let $\lambda_{-x} = \prod_{k=1}^K \lambda_k$ and $\lambda_{-k} = (\lambda_x \cdot \prod_{j=1}^K \lambda_j)/\lambda_k$. Then with the joint independence of the random elements $\{W_F, W_1, \ldots, W_K\}$, by Example 1.4.6 of van der Vaart and Wellner (1996) and Assumption 3.6 of this paper,

$$\sup_{\Gamma \in \mathrm{BL}_1(\mathbb{A})} \left| \mathbb{E}\left[\Gamma\left(\begin{bmatrix} \sqrt{T_n}\left(\widehat{F}^*_{n_x} - \widehat{F}_{n_x}\right) \\ \sqrt{T_n}\left(\widehat{G}^*_{n_1} \circ g_1 - \widehat{G}_{n_1} \circ g_1\right) \\ \vdots \\ \sqrt{T_n}\left(\widehat{G}^*_{n_K} \circ g_K - \widehat{G}_{n_K} \circ g_K\right) \end{bmatrix}\right)\middle| \mathcal{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\begin{bmatrix} \sqrt{\lambda_{-x}} W_F \\ \sqrt{\lambda_{-1}} W_1 \\ \vdots \\ \sqrt{\lambda_{-K}} W_K \end{bmatrix}\right)\right] \right| \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$.

Define a map $\mathcal{I} : \mathbb{A} \to \mathbb{B}$, such that

$$\mathcal{I}(f,h)(x,\theta) = (f(x) - h_1(x,\theta_1), \ldots, f(x) - h_K(x,\theta_K))$$

for every $(f,h) \in \mathbb{A}$ and every $(x,\theta) \in \mathbb{R} \times \Theta$ with $h = (h_1, \ldots, h_K)$ and $\theta = (\theta_1, \ldots, \theta_K)$. It is easy to show the Lipschitz continuity of $\mathcal{I}$. By the proof similar to that of Proposition 10.7(ii) of Kosorok (2008), we can show that

$$\sup_{\Gamma \in \mathrm{BL}_1\left(\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)\right)} \left|\mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}^*_n - \widehat{\phi}_n\right)\right)\middle| \mathcal{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\tilde{\mathbb{G}}_0\right)\right]\right| \xrightarrow{\text{a.s.}} 0$$

as $n \to \infty$, where $\tilde{\mathbb{G}}_0 = \mathcal{I}(\sqrt{\lambda_{-x}} W_F, \sqrt{\lambda_{-1}} W_1, \ldots, \sqrt{\lambda_{-K}} W_K)$. By the properties of the random elements $\{W_F, W_1, \ldots, W_K\}$, it can be verified that $\tilde{\mathbb{G}}_0$ is equivalent to $\mathbb{G}_0$ in law. The desired result follows from Lemma 1.9.2(i) of van der Vaart and Wellner (1996).

Because $\mathcal{F}$ and $\mathcal{G}_k$ are Donsker, by Theorem 2.6 of Kosorok (2008), $\sqrt{n_x}(\widehat{\mathcal{X}}^*_{n_x} - \widehat{\mathcal{X}}_{n_x})$ and $\sqrt{n_k}(\widehat{\mathcal{Y}}^*_{n_k} - \widehat{\mathcal{Y}}_{n_k})$ (for every k) are asymptotically measurable. By Lemma A.2, $\sqrt{n_x}(\widehat{F}^*_{n_x} - \widehat{F}_{n_x})$ and $\sqrt{n_k}(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k)$ are asymptotically measurable. By (B.8) and the asymptotic measurability of $\sqrt{n_x}(\widehat{F}^*_{n_x} - \widehat{F}_{n_x})$ and $\sqrt{n_k}(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k)$, we can show that $\sqrt{n_x}(\widehat{F}^*_{n_x} - \widehat{F}_{n_x})$ and $\sqrt{n_k}(\widehat{G}^*_{n_k} \circ g_k - \widehat{G}_{n_k} \circ g_k)$ are asymptotically tight. Then by Lemmas 1.4.3 and 1.4.4 of van der Vaart and Wellner (1996),

$$(\sqrt{n_x}(\widehat{F}^*_{n_x} - \widehat{F}_{n_x}), \sqrt{n_1}(\widehat{G}^*_{n_1} \circ g_1 - \widehat{G}_{n_1} \circ g_1), \ldots, \sqrt{n_K}(\widehat{G}^*_{n_K} \circ g_K - \widehat{G}_{n_K} \circ g_K))$$

is asymptotically measurable. The asymptotic measurability of $\sqrt{T_n}(\widehat{\phi}^*_n - \widehat{\phi}_n)$ follows from the continuity of $\mathcal{I}$. □

Proof of Proposition 3.3: Note that both $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ and $\mathbb{R}$ are normed spaces, and by Lemma 3.3, the map $\mathcal{L}$ is second order Hadamard directionally differentiable at $\phi \in \mathbb{D}_0$ tangentially to $\mathbb{D}_{\mathcal{L}0}$. Lemma 3.1 shows that $\sqrt{T_n}(\widehat{\phi}_n - \phi) \rightsquigarrow \mathbb{G}_0$ in $\prod_{k=1}^K \ell^\infty(\mathbb{R} \times \Theta_k)$ as $n \to \infty$

and $\mathbb{G}_0$ is tight with $\mathbb{G}_0 \in \mathbb{D}_{\mathcal{L}0}$ almost surely. By Lemma B.3, $\mathbb{D}_{\mathcal{L}0}$ is closed under vector addition, that is, $\varphi_1 + \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$ whenever $\varphi_1, \varphi_2 \in \mathbb{D}_{\mathcal{L}0}$. By construction, the random weights used to construct the bootstrap samples are independent of the data set and $\widehat{\phi}_n^*$ is a measurable function of the random weights. By Lemma 3.5,

$$\sup_{\Gamma \in \mathrm{BL}_1\left(\prod_{k=1}^K \ell^\infty(\mathbb{R}\times\Theta_k)\right)} \left| \mathbb{E}\left[\Gamma\left(\sqrt{T_n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right) \middle| \{X_i\}_{i=1}^{n_x}, \{Y_{1i}\}_{i=1}^{n_1}, \ldots, \{Y_{Ki}\}_{i=1}^{n_K} \right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0\right)\right] \right|$$
$$\xrightarrow{\mathbb{P}} 0,$$

and $\sqrt{T_n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ is asymptotically measurable as $n \to \infty$. Lemma 3.4 establishes the consistency of $\widehat{\mathcal{L}}_n''$ for $\mathcal{L}_\phi''$. Therefore, Assumptions 2.1(i), 2.1(ii), 2.2, 3.1, 3.2, and 3.4 of Chen and Fang (2019) are satisfied, and the result follows from Theorem 3.3 of Chen and Fang (2019). □

Proof of Theorem 3.1: Under Assumptions 3.1–3.9, with Propositions 3.2 and 3.3, the desired results can be proved by arguments similar to those in the proof of Theorem 2.1. □

References

Chen, Q. and Fang, Z. (2019). Inference on functionals under first order degeneracy. *Journal of Econometrics*, 210(2):459–481.

Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487.

Shapiro, A. (2000). Statistical inference of stochastic optimization problems. In Uryasev, S. P., editor, *Probabilistic Constrained Optimization: Methodology and Applications*, pages 282–307. Springer.

Sun, Z. (2021). Instrument validity for heterogeneous causal effects. arXiv:2009.01995.

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

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