Testing functional inequalities conditional on estimated functions*

Andres Aradillas-Lopez†

June 11, 2022

Abstract

Testing functional inequalities is a common problem in econometrics, often derived from restrictions of economic models. In many applications, these functionals may be conditioned on a rich collection of covariates X, with continuous support. In such cases, to apply nonparametric methods, researchers may wish to aggregate X into a lower-dimensional collection of functions. Motivated by this problem, we introduce a test for functional inequalities conditional on estimated functions. We focus on cases where the conditioning functions have a parametric functional form and are indexed by a finite-dimensional parameter vector which is estimated in a first step. Our proposed tests are based on one-sided Cramér–von Mises (CvM) statistics where violations to the inequalities are measured through a tuning parameter converging to zero. This will allow our test-statistics to adapt asymptotically to the measure of the contact sets (the set of values of conditioning variables where the inequalities are binding). A regularization of the asymptotic standard error of our test-statistics yields asymptotically pivotal properties. In Monte Carlo experiments, our test displays good power properties, capable of detecting violations to the inequalities that occur with very low probability.
Keywords: Functional inequalities, nonparametric tests, conditional moments, curse of dimensionality.
JEL classification: C1, C12, C14.

1 Introduction

A commonly encountered problem in econometrics involves inequalities of conditional moments. These functional inequalities can arise as testable implications of economic models. If these functionals are conditioned on a vector of observable covariates X, they can be estimated nonparamet-

*A previous version of this paper was titled "Nonparametric Tests for Conditional Affiliation in Auctions and Other Models"

†Department of Economics, Pennsylvania State University, University Park, PA 16802, United States. Email: aaradill@psu.edu

rically; however, in many applications X may contain a rich number of continuous variables, making nonparametric estimators susceptible to the curse of dimensionality. In such cases, researchers may opt to condition on a lower-dimensional index or *conditioning function* of X. Often, these conditioning functions have a parametric form indexed by a finite-dimensional parameter θ, and can be expressed as $g(X,\theta)$. For example, consider the conjecture that Y_1 first-order stochastically dominates Y_2 conditional on X. This would be evaluated by testing $E[\mathbb{1}\{Y_1 \leq t\} - \mathbb{1}\{Y_2 \leq t\} \mid X] \leq 0$ w.p.1 $\forall\ t$. If X includes a rich collection of continuous random variables, to mitigate the curse of dimensionality, researchers may prefer to focus on testing the condition $E[\mathbb{1}\{Y_1 \leq t\} - \mathbb{1}\{Y_2 \leq t\} \mid g(X,\theta^*)] \leq 0$ w.p.1 $\forall\ t$, where θ^* may be unknown but can be consistently estimated with an estimator $\widehat{\theta}$. Conditioning on $g(X,\theta^*)$ instead of X could be justified by an exclusion restriction if the researcher assumes that $Y_1, Y_2|X \sim Y_1, Y_2|g(X,\theta^*)$. Alternatively, it may be justified by iterated expectation arguments or simply by the need to mitigate the curse of dimensionality. In this paper we take as given that the goal of the researcher is to do a test conditional on $g(X,\theta^*)$, and we develop a test where θ^* is replaced with a first-step estimator $\widehat{\theta}$. The study of how to choose $g(X,\theta^*)$ "optimally" is outside the scope of this paper.

To our knowledge, this appears to be the first paper explicitly devoted to the study of testing functional inequalities conditional on estimated functions. For the reasons outlined above, this is a relevant problem, particularly for practitioners. Our method is an extension of the type of one-sided Cramér–von Mises (CvM) tests proposed in Aradillas-López, Gandhi, and Quint (2016) to the case of estimated conditioning functions. Violations to the inequalities are measured through a tuning parameter b_n converging to zero. This will allow or test-statistic to adapt asymptotically to the measure of the contact sets (the set of values of conditioning variables where the inequalities are binding). A regularization of the asymptotic standard error of our test-statistic will yield asymptotically pivotal properties. Existing methods for conditional moment inequalities (CMIs) include, among others, Andrews and Shi (2013), Lee, Song, and Whang (2013), Lee, Song, and Whang (2018), Armstrong (2015), Armstrong (2014), Chetverikov (2017), Armstrong and Chan (2016) and Armstrong (2018). However, none of the existing procedures considers the case where the conditioning variable is a function that is estimated in a first step, and the asymptotic properties of these and other existing methods in this case have not been characterized. The goal of this paper is to contribute to fill in this gap in the literature.

The paper proceeds as follows. Section 2 provides the setup and the description of the type of functional inequalities this paper studies. Section 3 describes the econometric tests and their asymptotic properties. Section 4 discusses how to extend our methodology to the problem of testing multiple inequalities. Section 5 includes results from Monte Carlo experiments. Section 6 concludes. The online **Appendix**[1] describes the steps of the econometric proofs for our main results along with additional results referenced throughout the paper. Step-by-step derivations

[1] Available online at http://www.personal.psu.edu/aza12/condit-ineq-functions-appendix.pdf

and the full details of the proofs are included in the **Econometric Supplement** of the paper[2]

2 Setup and description of the inequalities to be tested

2.1 Some preliminaries

We will describe our model and results in the context of a space of distributions $\mathcal{F}$ which contains the distribution that generated the data observed. We will describe conditions that yield asymptotic properties that hold uniformly over $\mathcal{F}$. Our model includes a triple of random variables (Y, X, Z) whose distinct roles will be made clear below. We observe a sample $(Y_i, X_i, Z_i)_{i=1}^n$ of independent observations of a distribution $F \in \mathcal{F}$. We will let $\mathcal{S}_\xi$ denote the support of a r.v ξ. As we shall describe below, Z can include elements from (Y, X) and we can have $Z = (Y, X)$. We will group $V \equiv (Y, X) \cup Z$, and we will maintain the assumption that $\mathcal{S}_V$ is the same for all $F \in \mathcal{F}$. Throughout, we will indicate functionals of the distribution F by including the subscript F except when this results in cumbersome notation. In every case, the exposition will make clear which objects are functionals of F. We will maintain the assumption that the space of distributions $\mathcal{F}$ satisfies the following compactness feature. For any measurable set S,

$$\sup_{F \in \mathcal{F}} P_F(S) = p \quad \Longrightarrow \quad \exists\ F^* \in \mathcal{F} : P_{F^*}(S) = p.$$

Following convention, we will use the following terminology for a given sequence $\{\xi_n\}$,

(i) $\xi_n = o_p(n^\lambda)$ *uniformly over* $\mathcal{F}$ if,

$$\sup_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) \longrightarrow 0 \quad \forall\ c > 0.$$

(ii) $\xi_n = O_p(n^\lambda)$ *uniformly over* $\mathcal{F}$ if, for any $\varepsilon > 0$ there exist a finite $\Delta_\varepsilon > 0$ and $n_\varepsilon \in \mathbb{N}$ such that

$$\sup_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > \Delta_\varepsilon\right) < \varepsilon \quad \forall\ n \geq n_\varepsilon.$$

Following convention, we say that $s_n = O(n^\gamma)$ for a deterministic sequence s_n if for some $\Delta > 0$, $\exists\ n_0$ such that $\|n^{-\gamma}s_n\| < \Delta$ for all $n \geq n_0$. Thus, we say $\sup_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) = O(n^\gamma)$ for for a given $c > 0$, if $\exists\ n_0$ and $\Delta > 0$ such that $n^{-\gamma} \sup_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) < \Delta \quad \forall\ n \geq n_0$.

[2] Available online at http://www.personal.psu.edu/aza12/condit-ineq-functions-supplement.pdf

2.2 Components of the models studied here

2.2.1 Conditioning functions

We have a collection of D real-valued, parametric *conditioning functions* $(g_d)_{d=1}^{D}$ whose arguments are X and a parameter $\theta \in \mathbb{R}^k$. We will group

$$\underbrace{g(X,\theta)}_{D\times 1} \equiv \underbrace{(g_1(X,\theta), g_2(X,\theta), \ldots, g_D(X,\theta))'}_{D \text{ conditioning functions}}.$$

For example, we may have $g(X,\theta) \equiv (X_1'\theta_1, \ldots, X_D'\theta_D)'$, with $X \equiv \cup_{d=1}^{D} X_d$ and $\theta \equiv (\theta_1', \ldots, \theta_D')'$.

2.2.2 Conditional functionals and index parameters

The model includes a collection of P known real-valued functions $(S_p)_{p=1}^{P}$, which depend on Y, and on an *index parameter* $t \in \mathcal{T} \subseteq \mathbb{R}^{d_t}$, where $\mathcal{T}$ is a known, pre-specified, bounded subset of $\mathbb{R}^{d_t}$. For each $x \in \mathcal{S}_X$, $t \in \mathcal{T}$, $\theta \in \Theta$ and $F \in \mathcal{F}$, we define the following functionals,

$$\Gamma_{p,F}(x,t,\theta) = E_F\left[S_p(Y,t) \,\middle|\, g(X,\theta) = g(x,\theta)\right], \quad \text{with}$$
$$\underbrace{\Gamma_F(x,t,\theta)}_{P\times 1} \equiv \Big(\Gamma_{1,F}(x,t,\theta), \ldots, \Gamma_{P,F}(x,t,\theta)\Big)'$$

Models without index parameters t will be a a special case of our general setup and methodology.

2.2.3 Functional inequalities to test

The last basic component is real-valued, known, real-valued transformation $\mathcal{B}$ of the vector of functionals $\Gamma_F(\cdot)$, and a model that predicts,

$$\mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \;\; F\text{–a.e } x \in \mathcal{S}_X, \; \forall \; t \in \mathcal{T}. \tag{1}$$

The parameter value θ_F^* is unknown but will be assumed to be the probability-limit of an estimator $\widehat{\theta}$ whose properties will be described below. We will let Θ denote the parameter space for θ_F^*, assumed to be the same for all $F \in \mathcal{F}$. Going back to the case of linear indices, $g(X,\theta) \equiv \left(X_1'\theta_1, \ldots, X_D'\theta_D\right)'$, the parameter value θ_F^* could be, e.g, the probability-limit of a least-squares estimator in a linear regression model or a semiparametric estimator for a single index (Han (1987), Powell, Stock, and Stoker (1989), Ichimura (1993)) or a multiple index model (Ichimura and Lee (1991), Picone and Butler (2000), Donkers and Schafgans (2008)). We focus for now on the case where $\mathcal{B}$ is a scalar, but we will show how to extend our approach to the case of multiple inequalities in Section 4.

When would inequalities of the type (1) **arise?**

Restrictions of the form (1) could arise in cases where we start with functional inequalities conditional on X which are then reduced to inequalities conditional on $g(X,\theta_F^*)$, either by iterated expectations arguments, by the assumption of exclusion restrictions or simply by a desire to mitigate the curse of dimensionality. Three possible instances are described next.

(A) Iterated expectations when the mapping $\mathcal{B}$ is linear.- Let α be a vector of known constants and consider an econometric model that predicts,

$$\alpha' E_F[S(Y,t)|X] \leq 0 \; F\text{–a.s}, \forall\, t \in \mathcal{T}. \tag{2}$$

By iterated expectations, $\alpha' E_F[S(Y,t)|\varphi(X)] \leq 0$ F–a.s, $\forall\, t \in \mathcal{T}$, for any measurable φ. Thus, (1) arises as a special case, where $\varphi(X) \equiv g(X,\theta_F^*)$ and $\mathcal{B}(\Gamma) \equiv \alpha'\Gamma$.

(B) An exclusion restriction of the form $Y|X \sim Y|g(X,\theta_F^*)$ is assumed.- Let $\Upsilon_{p,F}(x,t) \equiv E_F[S_p(Y,t)|X = x]$, and $\Upsilon_F(x,t) \equiv (\Upsilon_{1,F}(x,t),\ldots,\Upsilon_{P,F}(x,t))'$, and consider a model that predicts,

$$\mathcal{B}(\Upsilon_F(x,t)) \leq 0 \;\text{ for } F\text{–a.e } x \in \mathcal{S}_X \text{ and } \forall\, t \in \mathcal{T}. \tag{3}$$

In this case, (3) turns into our inequality (1) if we assume the exclusion restriction $Y|X \sim Y|g(X,\theta_F^*)$.

(C) Dimension reduction.- Researchers may focus on (1) directly when X includes a large collection of covariates which are aggregated into $g(X,\theta_F^*)$ to mitigate the curse of dimensionality without necessarily assuming the exclusion restriction $Y|X \sim Y|g(X,\theta_F^*)$.

Remark 1 The problem of how to choose the conditioning functions $g(X,\theta_F^*)$ "optimally" is outside the scope of this paper. We will take the choice of $g(X,\theta_F^*)$ as given. ■

2.2.4 Examples

Examples of economic models with testable implications described as (1) include the following.

Example 1: First order stochastic dominance

Suppose Y_1, Y_2 have common support and we have an economic model that predicts the first-order stochastic dominance restriction $F_{Y_1|X}(\cdot|X) \succsim_{FOSD} F_{Y_2|X}(\cdot|X)$ F–a.s. This relation implies the inequality $F_{Y_1|X}(t|X) \leq F_{Y_2|X}(t|X)$ F–a.s, $\forall\, t$. Let $Y \equiv (Y_1,Y_2)$ and $S(Y,t) \equiv \mathbb{1}\{Y_1 \leq t\} - \mathbb{1}\{Y_2 \leq t\}$. Then, $E_F[S(Y,t)|X] \leq 0$ F–a.s, $\forall\, t$, which is a special case of (2). Thus, (1) follows by iterated expectations, with $\mathcal{B}$ being the identity function.

Example 2: Second order stochastic dominance

Consider now a second-order stochastic dominance restriction $F_{Y_1|X}(\cdot|X) \gtrsim_{SOSD} F_{Y_2|X}(\cdot|X)$ $F-$a.s. This requires $\int_{-\infty}^{t} F_{Y_1|X}(v|x)dv \leq \int_{-\infty}^{t} F_{Y_2|X}(v|x)dv$ for $F-$a.e $x \in \mathcal{S}_X$ and $\forall\ t$. Note that $\int_{-\infty}^{t} \mathbb{1}\{\xi \leq v\}\, dv = \max\{t-\xi,\, 0\}$. Thus, for $\ell = 1, 2$,

$$\begin{aligned} E_F[\max\{t-Y_\ell,\, 0\}|X=x] &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{t} \mathbb{1}\{y \leq v\}\, dv\right) f_{Y_\ell|X}(y|x)dy \\ &= \int_{-\infty}^{t}\left(\int_{-\infty}^{\infty} \mathbb{1}\{y \leq v\}\, f_{Y_\ell|X}(y|x)dy\right)dv = \int_{-\infty}^{t} F_{Y_\ell|X}(v|x)dv \end{aligned}$$

Denote $S(Y,t) \equiv \max\{t-Y_1,\, 0\} - \max\{t-Y_2,\, 0\}$. The model predicts $E_F[S(Y,t)|X] \leq 0$ $F-$a.s, $\forall\ t$. As in the FOSD example, this is a special case of (2). Thus, (1) follows by iterated expectations, with $\mathcal{B}$ being the identity function.

Example 3: Covariance restrictions

There exist economic models that yield restrictions of the general form

$$Cov\Big(\eta_1(Y,t), \eta_2(Y,t)\Big|X\Big) \leq 0 \ F\text{–a.s}, \ \forall\ t \in \mathcal{T}. \tag{4}$$

where η_1 and η_2 are known functions. As shown in Aradillas-López and Gandhi (2016), covariance restrictions of this form arise in incomplete information games with ordinal action spaces when we conjecture that an aggregate index of the actions of player p's opponents, $\varphi(Y_{-p})$ (e.g, $\varphi(Y_{-p}) = \sum_{q \neq p} Y_q$) is a strategic substitute for Y_p, the action of player p. Under payoff shape restrictions described by the authors, (4) must hold for $\eta_1(Y,t) = \mathbb{1}\{Y_p \geq t\}$ and $\eta_2(Y,t) = \varphi(Y_{-p})$. Here, t denotes a particular element in the action space of player p, and $\mathcal{T}$ denotes the entire action space. Let $S_1(Y,t) \equiv \eta_1(Y,t)\cdot\eta_2(Y,t)$, $S_2(Y,t) \equiv \eta_1(Y,t)$ and $S_3(Y,t) \equiv \eta_2(Y,t)$. Then, (4) implies $E_F[S_1(Y,t)|X] - E_F[S_2(Y,t)|X]\cdot E_F[S_3(Y,t)|X] \leq 0$ $F-$a.s, $\forall\ t \in \mathcal{T}$. In this case, (1) arises if we assume the exclusion restriction $Y|X \sim Y|g(X,\theta_F^*)$. The transformation $\mathcal{B}$ is $\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3) = \Gamma_1 - \Gamma_2\cdot\Gamma_3$.

Example 4: Affiliation

Let $Y \equiv (Y_1,\ldots,Y_L) \in \mathbb{R}^L$. Following convention, let $a \vee b \equiv \max\{a,b\}$ and $a \wedge b \equiv \min\{a,b\}$ (element-wise). Take $\delta \in \mathbb{R}_+^L$, $u \in \mathbb{R}^L$, and let $G_F(u,\delta|X) = P_F\Big(u-\delta \leq Y \leq u+\delta\Big|X\Big)$. Take $t_1, t_2 \in \mathbb{R}^L$ and $t_3 \in \mathbb{R}_+^L$, group $t \equiv (t_1', t_2', t_3')'$ and define,

$$\tau_F(t,X) = G_F(t_1,t_3|X)\cdot G_F(t_2,t_3|X) - G_F(t_1 \vee t_2, t_3|X)\cdot G_F(t_1 \wedge t_2, t_3|X).$$

Using the definition in Milgrom and Weber (1982, Lemma 1), the elements of Y are *affiliated conditional on* X if and only if $\tau_F(t,X) \leq 0$ $F-$a.s, $\forall\ t \in \mathbb{R}^L \times \mathbb{R}^L \times \mathbb{R}_+^L$. Let $S_1(Y,t) \equiv \mathbb{1}\{t_1 - t_3 \leq Y \leq$

$t_1+t_3\}$, $S_2(Y,t) \equiv \mathbb{1}\{t_2-t_3 \le Y \le t_2+t_3\}$, $S_3(Y,t) \equiv \mathbb{1}\{t_1 \vee t_2-t_3 \le Y \le t_1 \vee t_2+t_3\}$, and $S_4(Y,t) \equiv \mathbb{1}\{t_1 \wedge t_2-t_3 \le Y \le t_1 \wedge t_2+t_3\}$. Assuming $Y|X \sim Y|g(X,\theta_F^*)$, then (1) arises with $\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4) \equiv \Gamma_1 \cdot \Gamma_2 - \Gamma_3 \cdot \Gamma_4$.

Example 5: Conditional moment inequalities (CMI)

Our methodology will include CMI models that predict $E[S(Y)|X] \le 0$ F–a.s (without an index parameter t) as special cases. These models are a special case of (2), and (1) follows by iterated expectations, with $\mathcal{B}$ being the identity function.

2.2.5 An estimator for θ_F^*

The value of θ_F^* is unknown but point-identified and we have at hand an estimator $\widehat{\theta}$, which is a function of $(Z_i)_{i=1}^n$. *The only role of* $(Z_i)_{i=1}^n$ *in our model is in the construction of* $\widehat{\theta}$. Z can have elements in common with (Y,X) and we can have $Z=(Y,X)$. Our results can be readily extended to the case where $\widehat{\theta}$ is obtained from a separate, auxiliary sample, and this would simplify the asymptotic analysis. We focus on the case where $\widehat{\theta}$ is obtained from the same sample because it represents the most common scenario. The relevant properties of $\widehat{\theta}$ will be described next.

Assumption 1 *θ_F^* is an element of the parameter space $\Theta \subseteq \mathbb{R}^k$ for every $F \in \mathcal{F}$ and we also have $\widehat{\theta} \in \Theta$. The parameter space Θ will be taken to be a bounded and convex subset of $\mathbb{R}^k$. The estimator satisfies the following linear representation,*

$$\widehat{\theta} = \theta_F^* + \frac{1}{n}\sum_{i=1}^n \psi_F^\theta(Z_i) + \varepsilon_n^\theta,$$

where ε_n^θ is such that there exists a $\tau > 0$ such that

$$\left\|\varepsilon_n^\theta\right\| = o_p\left(\frac{1}{n^{1/2+\tau}}\right) \text{ uniformly over } \mathcal{F}\text{, i.e, } \sup_{F\in\mathcal{F}} P_F\left(n^{1/2+\tau}\cdot\left\|\varepsilon_n^\theta\right\| \ge \delta\right) \longrightarrow 0 \ \forall\ \delta > 0$$

and

$$\sup_{F\in\mathcal{F}} P_F\left(\left\|\varepsilon_n^\theta\right\| \ge c\right) = O\left(\frac{1}{(r_n\cdot c)^q}\right) \quad \forall\ c > 0$$

for some integer $q \ge 2$ and a sequence $r_n \longrightarrow \infty$. The integer q and the sequence r_n are such that $\exists\ \overline{\delta} > 0$ such that $n^{1/2+\overline{\delta}}/r_n^q \longrightarrow 0$

Examples of estimators that satisfy Assumption 1

In Appendix A4[3] we describe conditions under which two examples of estimators satisfy the restrictions in Assumption 1. The examples we include there are **OLS** (Appendix A4.1) and a **semi-**

[3] Available online at `http://www.personal.psu.edu/aza12/condit-ineq-functions-appendix.pdf`

parametric multiple index estimator (Appendix A4.2). We include two additional examples: **GMM** and **density-weighted average derivatives** in the Econometric Supplement of the paper[4] (Section S5). In all cases, the conditions needed to satisfy the restrictions in Assumption 1 involve commonly assumed existence-of-moments and Jacobian-invertibility conditions.

2.3 The goal of this paper: a test for (1)

Our goal is to construct a test for the functional inequality (1) over a prespecified testing range of values of x.

2.3.1 A target testing range

Let $\mathcal{X} \subset \mathcal{S}_X$ denote a *prespecified*, compact subset of the support of X, and let $\mathcal{G}$ be another prespecified, compact subset of $\mathbb{R}^D$. Let,

$$\mathcal{X}_F^* = \{x \in \mathcal{S}_X : x \in \mathcal{X} \text{ and } g(x, \theta_F^*) \in \mathcal{G}\}$$

$\mathcal{X}_F^*$ constitutes our *target testing range* for the functional inequalities (1). $\mathcal{X}_F^*$ will be assumed to satisfy conditions that yield uniform asymptotic properties for our nonparametric estimators.

2.3.2 Weight functions

Following our choice of $\mathcal{G}$, we introduce a collection of *weight functions* $(\omega_p)_{p=1}^P$. Each ω_p is a mapping $\omega_p : \mathbb{R}^D \to \mathbb{R}$, satisfying $\omega_p(g) \geq 0 \ \forall \ g \in \mathbb{R}^D$, and $\omega_p(g) > 0 \iff g \in \mathcal{G}$. We will define

$$\begin{aligned} Q_{p,F}(x,t,\theta_F^*) &\equiv \Gamma_{p,F}(x,t,\theta_F^*) \cdot \omega_p(g(x,\theta_F^*)), \\ Q_F(x,t,\theta_F^*) &\equiv (Q_{1,F}(x,t,\theta_F^*), \ldots, Q_{P,F}(x,t,\theta_F^*))', \\ \omega(g(x,\theta_F^*)) &\equiv (\omega_1(g(x,\theta_F^*)), \ldots, \omega_P(g(x,\theta_F^*)))'. \end{aligned} \tag{5}$$

2.3.3 A maintained assumption about the mapping $\mathcal{B}$ and the weight functions

The rest of the paper will focus on cases where the transformation $\mathcal{B}$ enables us to design weight functions $(\omega_p)_{p=1}^P$ such that,

$$\begin{aligned} &\mathcal{B}(\Gamma_1 \cdot \omega_1, \ldots, \Gamma_P \cdot \omega_P) = \mathcal{B}(\Gamma_1, \ldots, \Gamma_P) \cdot \mathcal{H}(\omega_1, \ldots, \omega_P), \\ &\text{where } \begin{cases} \mathcal{H}(\cdot) \geq 0, \\ \mathcal{H}(\omega_1, \ldots, \omega_P) > 0 \iff \omega_p > 0 \ \forall \ p. \end{cases} \end{aligned} \tag{6}$$

[4]Available online at `http://www.personal.psu.edu/aza12/condit-ineq-functions-supplement.pdf`

Therefore, it follows that

$$\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) = \mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right)\cdot\mathcal{H}\left(\omega(g(x,\theta_F^*))\right),$$
$$\text{where}\begin{cases}\mathcal{H}(\cdot)\geq 0,\\ \mathcal{H}(\omega(g(x,\theta_F^*))>0 \iff g(x,\theta_F^*)\in\mathcal{G}.\end{cases} \tag{7}$$

The condition in (5) **restricts the type of functional inequalities that we will study here**. It includes all linear transformations $\mathcal{B}$, but it can also include nonlinear transformations. In particular, it includes all the examples in Section 2.2.4.

Examples 1, 2 and 5 (FOSD, SOSD and CMI): Here, we have $\mathcal{B}(\Gamma)=\Gamma$ and the condition in (6) is satisfied trivially, since $\mathcal{B}(\Gamma\cdot\omega)=\Gamma\cdot\omega=\mathcal{B}(\Gamma)\cdot\mathcal{H}(\omega)$, with $\mathcal{H}(\omega)\equiv\omega$. ■

Example 3 (Covariance restrictions): We now have $P=3$ and $\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3)=\Gamma_1-\Gamma_2\Gamma_3$. Take any $\omega_1>0$ and let $\omega_2=\omega_3=\omega_1^{1/2}$. We have $\mathcal{B}(\Gamma_1\cdot\omega_1,\Gamma_2\cdot\omega_2,\Gamma_3\cdot\omega_3)=\Gamma_1\omega_1-\Gamma_2\omega_1^{1/2}\Gamma_3\omega_1^{1/2}=(\Gamma_1-\Gamma_2\cdot\Gamma_3)\cdot\omega_1=\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3)\cdot\mathcal{H}(\omega)$. Thus, the condition in (6) holds with $\mathcal{H}(\omega)\equiv\omega_1$. ■

Example 4 (Affiliation): In this example we have $P=4$ and let $\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4)=\Gamma_1\Gamma_2-\Gamma_3\Gamma_4$. Take any $\omega_1>0$ and set $\omega_2=\omega_3=\omega_4=\omega_1$. Then, $\mathcal{B}(\Gamma_1\cdot\omega_1,\Gamma_2\cdot\omega_2,\Gamma_3\cdot\omega_3,\Gamma_4\cdot\omega_4)=\Gamma_1\omega_1\Gamma_2\omega_1-\Gamma_3\omega_1\Gamma_4\omega_1=(\Gamma_1\Gamma_2-\Gamma_3\Gamma_4)\cdot\omega_1^2=\mathcal{B}(\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4)\cdot\mathcal{H}(\omega)$, and condition in (6) holds with $\mathcal{H}(\omega)\equiv\omega_1^2$. ■

2.3.4 Our test

We will construct a test for the inequality

$$\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right)\leq 0 \;\text{ for } F\text{-a.e } x\in\mathcal{X} \text{ and } \forall\; t\in\mathcal{T}. \tag{8}$$

From (7), testing (8) is equivalent to testing the restriction,

$$\mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right)\leq 0 \;\text{ for } F\text{-a.e } x\in\mathcal{X}_F^* \text{ and } \forall\; t\in\mathcal{T} \tag{9}$$

The goal if this paper is to test (9) by constructing a test for (8).

2.3.5 Two types of population statistics

Let $(a)_+\equiv a\vee 0$. Note from (7) that

$$\left(\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right)\right)_+ = \left(\mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right)\right)_+\mathcal{H}(\omega(g(x,\theta_F^*)) \tag{10}$$

We will focus on one-sided Cramér–von Mises (CvM) population statistics of the type studied in Aradillas-López, Gandhi, and Quint (2016), and we consider two cases.

1.– **When we integrate out the index parameter t with respect to the distribution of an observable covariate:** The researcher observes a r.v t_i whose distribution is used as the target weight function to integrate out the index parameter t. t_i can be one of the elements already included in V_i, otherwise we will expand our definition of V_i to $V_i \equiv (Y_i, X_i, Z_i, t_i)$. Let $\phi(x,t)$ be a pre-specified function satisfying $\phi(x,t) \geq 0$ for all (x,t) and $\phi(x,t) > 0$ if and only if $(x,t) \in \mathcal{X} \times \mathcal{T}$. Let $F_{X,t}$ denote the joint distribution of (X,t). In this case we will focus on the following population statistic,

$$\begin{aligned} T_{1,F} &\equiv E_{F_{X,t}}\left[(\mathcal{B}(Q_F(X,t,\theta_F^*)))_+ \phi(X,t)\right] \\ &= \int_{x,t} (\mathcal{B}(\Gamma_F(x,t,\theta_F^*)))_+ \mathcal{H}(\omega(g(x,\theta_F^*))\phi(x,t)dF_{X,t}(x,t) \end{aligned} \tag{11A}$$

where the last equality follows from (10). Our weight functions ensure that we remain within our target testing range $\mathcal{X}_F^*$. By construction, $T_{1,F} \geq 0$ and $T_{1,F} = 0$ if and only if $\mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0$ $F_{X,t}$–a.e $(x,t) \in \mathcal{X}_F^* \times \mathcal{T}$.

2.– **When we integrate out the index parameter t using a pre-specified weight function:** In this case we pre-specify a weight function $d\mathcal{W}$ for t, satisfying $d\mathcal{W}(t) \geq 0$ for all t and $d\mathcal{W}(t) > 0$ if and only if $t \in \mathcal{T}$. For simplicity, normalize $\int_{t\in\mathcal{T}} d\mathcal{W}(t) = 1$. Let $\phi(x)$ be now a function only of x satisfying $\phi(x) \geq 0$ for all x and $\phi(x) > 0$ if and only if $x \in \mathcal{X}$. For a given $t \in \mathcal{T}$ let

$$\begin{aligned} T_{0,F}(t) &\equiv E_{F_X}\left[(\mathcal{B}(Q_F(X,t,\theta_F^*)))_+ \phi(X)\right] \\ &= \int_{x} (\mathcal{B}(\Gamma_F(x,t,\theta_F^*)))_+ \mathcal{H}(\omega(g(x,\theta_F^*)))\phi(x)dF_X(x). \end{aligned}$$

In this case we focus on the following population statistic,

$$\begin{aligned} T_{2,F} &\equiv \int_t T_{0,F}(t)d\mathcal{W}(t) \\ &= \int_t \left(E_{F_X}\left[(\mathcal{B}(\Gamma_F(X,t,\theta_F^*)))_+ \mathcal{H}(\omega(g(X,\theta_F^*)))\phi(X)\right]\right)d\mathcal{W}(t) \end{aligned} \tag{11B}$$

By construction, $T_{2,F} \geq 0$ and $T_{2,F} = 0$ if and only if $\mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0$ for $F_X \otimes \mathcal{W}$–a.e $(x,t) \in \mathcal{X}_F^* \times \mathcal{T}$.

Remark 2 Models without index parameters t constitute a special case of (11B).

3 Proposed econometric tests and their asymptotic properties

Our test statistics will be based on estimators of $T_{1,F}$ and $T_{2,F}$ where we replace θ_F^* with $\widehat{\theta}$.

3.1 Notational definitions of some key functionals

We will focus on the case where the conditioning functions $g(X,\theta_F^*)$ are jointly continuously distributed[5] with joint density function denoted by $f_g(\cdot)$. For a given $x \in \mathcal{S}_X$ and $t \in \mathcal{T}$, we will define $R_{p,F}(x,t,\theta_F^*) \equiv \Gamma_{p,F}(x,t,\theta_F^*) \cdot \omega_p(g(x,\theta_F^*)) \cdot f_g(g(x,\theta_F^*))$. Then, $Q_{p,F}(x,t,\theta_F^*) = \frac{R_{p,F}(x,t,\theta_F^*)}{f_g(g(x,\theta_F^*))}$.

3.2 Estimators of the functionals involved

We will construct kernel-based estimators. Let $K : \mathbb{R}^D \to \mathbb{R}$ be a kernel function and let $h_n \longrightarrow 0$ be a nonnegative bandwidth sequence (their properties will be described below). For any pair x_1, x_2 and $\theta \in \Theta$, denote $\Delta g(x_1,x_2,\theta) \equiv g(x_1,\theta) - g(x_2,\theta)$. For a given x and $\theta \in \Theta$, define

$$\widehat{f_g}(g(x,\theta)) \equiv \frac{1}{n \cdot h_n^D} \sum_{i=1}^{n} K\left(\frac{\Delta g(X_i,x,\theta)}{h_n}\right).$$

$\widehat{f_g}(g(x,\widehat{\theta}))$ will be our estimator for $f_g(g(x,\theta_F^*))$. Next, let

$$\widehat{R}_p(x,t,\theta) \equiv \frac{1}{n \cdot h_n^D} \sum_{i=1}^{n} S_p(Y_i,t)\omega_p(g(X_i,\theta))K\left(\frac{\Delta g(X_i,x,\theta)}{h_n}\right),$$

be our estimator for $R_{p,F}(x,t,\theta_F^*)$. Accordingly, our estimator for $Q_{p,F}(x,t,\theta_F^*)$ will be

$$\widehat{Q}_p(x,t,\widehat{\theta}) \equiv \frac{\widehat{R}_p(x,t,\widehat{\theta})}{\widehat{f_g}(g(x,\widehat{\theta}))}, \quad \text{with} \quad \widehat{Q}(x,t,\widehat{\theta}) \equiv \left(\widehat{Q}_1(x,t,\widehat{\theta}),\ldots,\widehat{Q}_P(x,t,\widehat{\theta})\right)' \tag{12}$$

3.3 Our estimators for $T_{1,F}$ and $T_{2,F}$

Letting $\widehat{Q}(x,t,\widehat{\theta})$ be as described in (12), our test-statistics for $T_{1,F}$ and $T_{2,F}$ will be,

$$\begin{aligned}
\widehat{T}_1 &\equiv \frac{1}{n}\sum_{i=1}^{n} \mathcal{B}\left(\widehat{Q}(X_i,t_i,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t_i,\widehat{\theta})\right) \geq -b_n\right\}\phi(X_i,t_i), \\
\widehat{T}_2 &\equiv \int_t \left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}\phi(X_i)\right) d\mathcal{W}(t) \equiv \int_t \widehat{T}_0(t)d\mathcal{W}(t), \qquad (13)\\
&\text{where} \quad \widehat{T}_0(t) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}\phi(X_i),
\end{aligned}$$

and where $b_n \longrightarrow 0$ is a nonnegative sequence with properties described in Assumption 4 below.

[5]Note that this only presupposes that a subset of elements in X are continuously distributed and it allows for some of its elements to be discrete random variables.

3.4 Asymptotic properties of the estimators $\widehat{T_1}$ and $\widehat{T_2}$

In this section we will enumerate assumptions that will result in an asymptotic linear representation result for our estimators. Even though the assumptions are technical in nature, we will summarize intuitively how each one contributes to our main asymptotic result.

3.4.1 Asymptotic properties of $\widehat{Q}(x,t,\widehat{\theta}) - Q_F(x,t,\theta_F^*)$

For a given $F \in \mathcal{F}$ we will let $\mathcal{S}_{g,F}$ denote the support of $g(X,\theta_F^*)$. That is,

$$\mathcal{S}_{g,F} = \Big\{ g \equiv (g_1,\ldots,g_D) \in \mathbb{R}^D : g = g(x,\theta_F^*) \text{ for some } x \in \mathcal{S}_X \Big\}. \tag{14}$$

Next, let $\mathcal{S}_{g,F}^{\mathcal{X}}$ denote the restriction of $\mathcal{S}_{g,F}$ over the testing range $\mathcal{X}$. That is,

$$\mathcal{S}_{g,F}^{\mathcal{X}} = \Big\{ g \equiv (g_1,\ldots,g_D) \in \mathbb{R}^D : g = g(x,\theta_F^*) \text{ for some } x \in \mathcal{X} \Big\}. \tag{15}$$

Assume that both $\mathcal{S}_{g,F}^{\mathcal{X}}$ and $\mathcal{G}$ are subsets of $int(\mathcal{S}_{g,F})$ (the interior of the support $\mathcal{S}_{g,F}$) for each $F \in \mathcal{F}$. Our first assumption involves smoothness of functionals involved in the estimation of $\widehat{Q}(x,t,\widehat{\theta})$, and of the conditioning functions g. These conditions are similar to smoothness restrictions assumed in other nonparametric problems.

Assumption 2 ***(Smoothness I)*** *The conditioning functions $g(X,\theta_F^*)$ are jointly continuously distributed, with joint density function denoted by $f_g(\cdot)$. There exist constants $\underline{f}_g > 0$, $\overline{f}_g < \infty$ and $\overline{\Gamma} < \infty$ such that, for each $F \in \mathcal{F}$,*

$$\inf_{x\in\mathcal{X}} f_g(g(x,\theta_F^*)) \geq \underline{f}_g, \quad \sup_{x\in\mathcal{X}} f_g(g(x,\theta_F^*)) \leq \overline{f}_g, \text{ and } \sup_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left|\Gamma_{p,F}(x,t,\theta_F^*)\right| \leq \overline{\Gamma}, \quad p = 1,\ldots,P.$$

And there exists a constant $\overline{C}_0 < \infty$ such that, for each $d = 1,\ldots,D$ and $\ell = 1,\ldots,k$,

$$\sup_{x\in\mathcal{X}} \left| \frac{\partial g_d(x,\theta_F^*)}{\partial \theta_\ell} \right| \leq \overline{C}_0 \quad \forall\, F \in \mathcal{F}.$$

Let $int(A)$ denote the interior of the set A. Let $\mathcal{S}_{g,F}$ and $\mathcal{S}_{g,F}^{\mathcal{X}}$ be as defined in (14) *and* (15)*. Then, $\mathcal{S}_{g,F}^{\mathcal{X}} \subset int\left(\mathcal{S}_{g,F}\right)$ for each $F \in \mathcal{F}$. Furthermore, there exists a $c > 0$ such that if we define*

$$\overline{\mathcal{S}}_{g,F}^{\mathcal{X}} = \Big\{ u \equiv (u_1,\ldots,u_D) \in \mathbb{R}^D\text{: } g - c \leq u \leq g + c \text{ (element-wise) for some } g \in \mathcal{S}_{g,F}^{\mathcal{X}} \Big\}$$

then, $\overline{\mathcal{S}}_{g,F}^{\mathcal{X}} \in int\left(\mathcal{S}_{g,F}\right)$ for each $F \in \mathcal{F}$. In addition, the following properties hold over $\overline{\mathcal{S}}_{g,F}^{\mathcal{X}}$.

For a given $g \equiv (g_1, \ldots, g_D) \in \mathcal{S}_{g,F}$, *and each* $\ell = 1, \ldots, k$ *and* $d = 1, \ldots, D$, *let*

$$\Omega_{f_g}^{d,\ell}(g) = E_F\left[\frac{\partial g_d(X, \theta_F^*)}{\partial \theta_\ell}\middle| g(X, \theta_F^*) = g\right].$$

There exists an integer M *such that, for each* $\ell = 1, \ldots, k$ *and* $d = 1, \ldots, D$ *and every* $1 \le m \le M+1$ *and* $(j_1, \ldots, j_D)$ *such that* $\sum_{d=1}^D j_d = m$, *both*

$$\frac{\partial^m f_g(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}} \quad \text{and} \quad \frac{\partial^m \Omega_{f_g}^{d,\ell}(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}$$

are well-defined for any $(g_1, \ldots, g_D) \in \overline{\mathcal{S}}_{g,F}^{\mathcal{X}}$ *and every* $F \in \mathcal{F}$. *Furthermore, there exists a constant* $\overline{C}_1 < \infty$ *such that, for each* $d = 1, \ldots, D$ *and* $\ell = 1, \ldots, k$, *and for all* $1 \le m \le M+1$ *and* $(j_1, \ldots, j_D)$ *such that* $\sum_{d=1}^D j_d = m$,

$$\sup_{(g_1, \ldots, g_D) \in \overline{\mathcal{S}}_{g,F}^{\mathcal{X}}} \left|\frac{\partial^m f_g(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}\right| \le \overline{C}_1 \quad \text{and} \quad \sup_{(g_1, \ldots, g_D) \in \overline{\mathcal{S}}_{g,F}^{\mathcal{X}}} \left|\frac{\partial^m \Omega_{f_g}^{d,\ell}(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}\right| \le \overline{C}_1 \quad \forall\ F \in \mathcal{F}$$

The pre-specified weight function $\omega_p(\cdot)$ *is bounded above by a constant* $\overline{\omega}$ *and it is designed such that, for any* $g \equiv (g_1, \ldots, g_D) \in \mathbb{R}^D$ *and all* $(j_1, \ldots, j_D)$ *such that* $\sum_{d=1}^D j_d = m$, *and* $1 \le m \le M+1$ *(with* M *being the integer described above),*

$$\frac{\partial^m \omega_p(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}$$

is well defined. Furthermore, there exists a finite constant $\overline{C}_\omega$ *such that*

$$\sup_{(g_1, \ldots, g_D) \in \overline{\mathcal{S}}_{g,F}^{\mathcal{X}}} \left|\frac{\partial^m \omega_p(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}\right| \le \overline{C}_\omega,$$

for all $(j_1, \ldots, j_D)$ *such that* $\sum_{d=1}^D j_d = m$, *and* $1 \le m \le M+1$. *Next let,*

$$\begin{aligned}
\Omega_{R_p,0}(g,t) &= E_F\left[S_p(Y,t)\middle| g(X, \theta_F^*) = g\right], \\
\Omega_{R_p,1}^{d,\ell}(g,t) &= E_F\left[S_p(Y,t)\omega_p(g(X, \theta_F^*))\frac{\partial g_d(X, \theta_F^*)}{\partial \theta_\ell}\middle| g(X, \theta_F^*) = g\right], \\
\Omega_{R_p,2}(g,t) &= E_F\left[S_p(Y,t)\omega_p(g(X, \theta_F^*))\middle| g(X, \theta_F^*) = g\right], \\
\Omega_{R_p,3}^{\ell}(g,t) &= E_F\left[S_p(Y,t)\frac{\partial \omega_p(g(X, \theta_F^*))}{\partial \theta_\ell}\middle| g(X, \theta_F^*) = g\right].
\end{aligned}$$

(note that $\Omega_{R_p,0}\Big(g(x,\theta_F^*),t\Big)=\Gamma_{p,F}(x,t,\theta_F^*)$). There exists a constant $\overline{C}_2<\infty$ such that, for each p, ℓ and d, and for all $1\leq m\leq M+1$ and $(j_1,\ldots,j_D)$ such that $\sum_{d=1}^{D} j_d=m$,

$$\left.\begin{array}{ll}\sup\limits_{\substack{(g_1,\ldots,g_D)\in\overline{\mathcal{S}}_{g,F}^{\mathcal{X}}\\ t\in\mathcal{T}}}\left|\dfrac{\partial^m\Omega_{R_p,0}(g_1,\ldots,g_D,t)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right|\leq\overline{C}_2, & \sup\limits_{\substack{(g_1,\ldots,g_D)\in\overline{\mathcal{S}}_{g,F}^{\mathcal{X}}\\ t\in\mathcal{T}}}\left|\dfrac{\partial^m\Omega^{d,\ell}_{R_p,1}(g_1,\ldots,g_D,t)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right|\leq\overline{C}_2,\\ \sup\limits_{\substack{(g_1,\ldots,g_D)\in\overline{\mathcal{S}}_{g,F}^{\mathcal{X}}\\ t\in\mathcal{T}}}\left|\dfrac{\partial^m\Omega_{R_p,2}(g_1,\ldots,g_D,t)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right|\leq\overline{C}_2, & \sup\limits_{\substack{(g_1,\ldots,g_D)\in\overline{\mathcal{S}}_{g,F}^{\mathcal{X}}\\ t\mathcal{T}}}\left|\dfrac{\partial^m\Omega^{\ell}_{R_p,3}(g_1,\ldots,g_D,t)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right|\leq\overline{C}_2\end{array}\right\}\ \forall\ F\in\mathcal{F}$$

■

The smoothness conditions in Assumption 2, combined with bias-reduction properties of the kernels and bandwidths will make the bias of our estimators disappear asymptotically at the appropriate rate. Our next set of assumptions involves *manageability* conditions for empirical processes (see Pollard (1990, Definition 7.9) or Andrews (1994, Assumption A, Theorem 1)) that are relevant to our problem. These manageability properties will allow us to invoke the maximal inequality results in Sherman (1994). We will focus on *Euclidean* classes of functions, whose properties for empirical processes and U-processes were studied, for example, in Pollard (1984), Nolan and Pollard (1987), Pakes and Pollard (1989), Pollard (1990), Sherman (1994) and Andrews (1994). We proceed first by stating the definition of Euclidean classes of functions from Definitions 1 and 3 in Sherman (1994).

Euclidean classes of functions (Definitions 1 and 3 in Sherman (1994))

Let $\mathcal{T}$ be a space and d be a pseudometric defined on $\mathcal{T}$. For each $\varepsilon>0$, define the packing number *$D(\varepsilon,d,\mathcal{T})$ to be the largest number D for which there exist points $m_1,\ldots,m_D$ in $\mathcal{T}$ such that $d(m_i,m_j)>\varepsilon$ for each $i\neq j$. Packing numbers are a measure of how big $\mathcal{T}$ is with respect to d. Let $\mathscr{G}$ be a class of functions on $\mathbb{R}$. We say that G is an* envelope *for $\mathscr{G}$ is $\sup_{\mathscr{G}}|g(\cdot)|\leq G(\cdot)$. Let μ be a measure on $\mathcal{S}_Z^k$ and denote $\mu h\equiv\int h(z_1,\ldots,z_k)d\mu(z_1,\ldots,z_k)$. We say that the class of functions $\mathscr{G}$ is Euclidean (A,V) for the envelope G if, for any measure μ such that $\mu G^2<\infty$, we have $D(x,d_\mu,\mathscr{G})\leq Ax^{-V},\quad 0<x\leq 1$, where, for $g_1,g_2\in\mathscr{G}$, $d_\mu(g_1,g_2)=(\mu|g_1-g_2|^2/\mu G^2)^{1/2}$. The constants A and V must not depend on μ.*

Examples of Euclidean classes of functions

Examples of Euclidean classes of functions can be found, e.g, Pollard (1984), Nolan and Pollard (1987), Pakes and Pollard (1989), Pollard (1990), Sherman (1994) and Andrews (1994). They encompass a vast collection of functions that appear in many econometric models. A partial list of examples is the following.

• (Pakes and Pollard (1989, Lemma 2.13)) Let $\mathscr{G} = \{g(\cdot,t) : t \in T\}$ be a class of functions on $\mathscr{X}$ indexed by a bounded subset T of $\mathbb{R}^d$. If there exists an $\alpha > 0$ and a $\phi(\cdot) \geq 0$ such that $|g(x,t) - g(x,t')| \leq \phi(x) \cdot \|t - t'\|^\alpha$ for $x \in \mathscr{X}$ and $t, t' \in T$. Then $\mathscr{G}$ is Euclidean for the envelope $G \equiv |g(\cdot,t_0)| + M\phi(\cdot)$, where $t_0 \in T$ is an arbitrary point and $M \equiv (2\sqrt{d}\sup_T \|t - t_0\|)^\alpha$.

• (Nolan and Pollard (1987, Lemma 22), Pakes and Pollard (1989, Example 10)) Let $\lambda(\cdot)$ be a real-valued function of bounded variation on $\mathbb{R}$. The class $\mathscr{G}$ of all functions on $\mathbb{R}^d$ of the form $x \to \lambda(\alpha' x + \beta)$, with α ranging over $\mathbb{R}^d$ and β ranging over $\mathbb{R}$ is Euclidean for the constant envelope $G \equiv \sup|\lambda|$.

• (Pakes and Pollard (1989, Example 2.9)) Let $\{g_1, \ldots, g_k\}$ be a finite set of functions on $\mathscr{X}$. For each $0 < M < \infty$, let $\mathscr{G}_M$ denote the class of all linear combinations $\sum_i \alpha_i g_i(\cdot)$ with $\sum_i |\alpha_i| \leq M$. The class $\mathscr{G}_M$ is Euclidean for the envelope $G \equiv M \cdot \max_i |g_i|$.

• (Pakes and Pollard (1989, Lemma 2.12)) Let g be a real-valued function on a set $\mathscr{X}$ and define subgraph$(g) = \{(x,s) \in \mathscr{X} \otimes \mathbb{R} : 0 < s < g(x) \;\text{ or }\; 0 > s > g(x)\}$. If $\{\text{subgraph}(g) : g \in \mathscr{G}\}$ is a VC class of sets, then $\mathscr{G}$ is Euclidean for every envelope.

• (Pakes and Pollard (1989, p.1033) and a consequence of the previous example) A class $\mathscr{G}$ of indicator functions over a class of sets $\mathscr{D}$ is Euclidean for the envelope $G \equiv 1$ if and only if $\mathscr{D}$ is a VC class of sets.

• (Andrews (1994, Section 4)) From the above examples, the Type I, II and III classes of functions described in Andrews (1994) are special cases of Euclidean classes of functions.

Assumption 3 ***(Manageability I)*** *Let q be the integer described in Assumption 1. There exists a non-negative function $H_1(\cdot)$ on $\mathcal{S}_X$ and a $\overline{\mu}_{H_1} < \infty$ such that $E_F[H_1(X)^{4q}] \leq \overline{\mu}_{H_1}$ for all $F \in \mathcal{F}$, and the following conditions are satisfied,*

(i) For each conditioning function g_d, we have $\sup_{\theta\in\Theta} |g_d(x,\theta)| \leq H_1(x) \;\forall\; x \in \mathcal{S}_X$.

(ii) For F−a.e $x \in \mathcal{S}_X$, each conditioning function $g_d(x,\theta)$ is twice-continuously differentiable with respect to θ and, for each $\{\ell, m\} \in 1, \ldots, k$ and,

$$\left.\begin{array}{l} \left|\frac{\partial g_d(x,\theta)}{\partial\theta_\ell} - \frac{\partial g_d(x,\theta')}{\partial\theta_\ell}\right| \leq H_1(x) \cdot \|\theta - \theta'\| \\ \left|\frac{\partial^2 g_d(x,\theta)}{\partial\theta_\ell\partial\theta_m} - \frac{\partial^2 g_d(x,\theta')}{\partial\theta_\ell\partial\theta_m}\right| \leq H_1(x) \cdot \|\theta - \theta'\| \end{array}\right\} \;\forall\; x \in \mathcal{S}_X \;\text{ and }\; \theta, \theta' \in \Theta.$$

(iii) For each $p = 1, \ldots, P$, the class of functions

$$\mathscr{S}_p = \left\{ m : \mathcal{S}_Y \longrightarrow \mathbb{R} : m(y) = S_p(y,t) \;\text{ for some } t \in \mathcal{T} \right\}$$

is Euclidean for an envelope $\overline{S}(Y)$ *that satisfies* $E_F[\overline{S}(Y)^{4q}] \leq \overline{\mu}_{\overline{S}} < \infty$ *for all* $F \in \mathcal{F}$.

(iv) There exist constants $\overline{A}_1$ *and* $\overline{V}_1$ *such that, for each* $F \in \mathcal{F}$, *each* $\ell = 1,\ldots,k$ *and* $p = 1,\ldots,P$, *the following classes of functions are Euclidean* $(\overline{A}_1, \overline{V}_1)$,

$$\mathscr{M}_{p,F} = \left\{ m : \mathcal{S}_X \longrightarrow \mathbb{R} : m(x) = \Gamma_{p,F}(x,t,\theta^*_F) \text{ for some } t \in \mathcal{T} \right\},$$
$$\mathscr{R}^{\ell}_{p,F} = \left\{ m : \mathcal{X} \longrightarrow \mathbb{R} : m(x) = \Xi_{R_p}(x,t,\theta^*_F) \text{ for some } t \in \mathcal{T} \right\},$$

for an envelope $\overline{G}(X)$ *that satisfies* $E_F[\overline{G}(X)^{4q}] \leq \overline{\mu}_{\overline{G}} < \infty$ *for all* $F \in \mathcal{F}$.

■

By Assumptions 2 and 3, there exists a constant C such that

$$\left.\begin{array}{l} \left| \frac{\partial \omega_p(x,\theta)}{\partial \theta_\ell} - \frac{\partial \omega_p(x,\theta')}{\partial \theta_\ell} \right| \leq C \cdot H_1(x) \cdot \|\theta - \theta'\| \\ \left| \frac{\partial^2 \omega_p(x,\theta)}{\partial \theta_\ell \partial \theta_m} - \frac{\partial^2 \omega_p(x,\theta')}{\partial \theta_\ell \partial \theta_m} \right| \leq C \cdot H_1(x) \cdot \|\theta - \theta'\| \end{array} \right\} \quad \forall\, x \in \mathcal{S}_X \text{ and } \theta, \theta' \in \Theta.$$

for each $\ell, m \in 1,\ldots,k$. From here, the conditions in Assumption 3 and Lemma 2.13 in Pakes and Pollard (1989) imply that, for each $d = 1,\ldots,D$ and $\ell, m \in 1,\ldots,k$, the classes of functions

$$\mathscr{H}^{\ell}_{1d} = \left\{ m : \mathcal{S}_X \longrightarrow \mathbb{R}\text{: } m(x) = \frac{\partial \Delta g_d(x,s,\theta)}{\partial \theta_\ell} \text{ for some } s \in \mathcal{S}_X,\, \theta \in \Theta \right\},$$
$$\mathscr{H}^{\ell,m}_{2d} = \left\{ m : \mathcal{S}_X \longrightarrow \mathbb{R}\text{: } m(x) = \frac{\partial^2 \Delta g_d(x,s,\theta)}{\partial \theta_\ell \partial \theta_m} \text{ for some } s \in \mathcal{S}_X,\, \theta \in \Theta \right\},$$
$$\mathscr{H}^{\ell}_{3} = \left\{ m : \mathcal{S}_X \longrightarrow \mathbb{R}\text{: } m(x) = \frac{\partial \omega_p(x,s,\theta)}{\partial \theta_\ell} \text{ for some } s \in \mathcal{S}_X,\, \theta \in \Theta \right\},$$
$$\mathscr{H}^{\ell,m}_{4} = \left\{ m : \mathcal{S}_X \longrightarrow \mathbb{R}\text{: } m(x) = \frac{\partial^2 \omega_p(x,s,\theta)}{\partial \theta_\ell \partial \theta_m} \text{ for some } s \in \mathcal{S}_X,\, \theta \in \Theta \right\}$$

are Euclidean for an envelope $G_1(x)$ that satisfies $E_F[G_1(X)^{4q}] \leq \overline{\mu}_{G_1} < \infty$ for all $F \in \mathcal{F}$, with q being the integer described in Assumption 1. These classes of functions will span various empirical processes that are relevant to our problem. The next building block towards our main result involves the properties of our kernel functions and tuning parameters.

Assumption 4 ***(Kernels and bandwidths)*** *In what follows, let* M *be the integer described in Assumption 2.*

(i) The kernel $K : \mathbb{R}^D \longrightarrow \mathbb{R}$ *is a multiplicative kernel of the form* $K(\psi) = \prod_{d=1}^{D} \kappa(\psi_d)$ *(with* $\psi \equiv (\psi_1,\ldots,\psi_D)'$*), where* $\kappa(\cdot)$ *is a bias-reducing kernel of order* M *with support of the form* $[-S,S]$ *(i.e,*

$\kappa(S)=\kappa(-S)=0$, $\kappa(v)=0\ \forall\ v\notin(-S,S)$, with $\int_{-S}^{S}\kappa(v)dv=1$, $\int_{-S}^{S}v^j\kappa(v)dv=0$ for $j=1,\ldots,M-1$ and $\int_{-S}^{S}|v|^M\kappa(v)dv<\infty$) and symmetric around zero (i.e, $\kappa(v)=\kappa(-v)$ for all v). $\kappa(\cdot)$ is twice continuously differentiable, and we will denote $\kappa^{(1)}(v)\equiv\frac{d\kappa(v)}{dv}$ and $\kappa^{(2)}(v)\equiv\frac{d^2\kappa(v)}{dv^2}$. The kernel κ as well as its first two derivatives are bounded, with $|\kappa(\cdot)|\le\overline{\kappa}$, $\left|\kappa^{(1)}(\cdot)\right|\le\overline{\kappa}$ and $\left|\kappa^{(2)}(\cdot)\right|\le\overline{\kappa}$ for a constant $\overline{\kappa}<\infty$. Note that since $\kappa(\cdot)$ is symmetric around zero, $\kappa^{(1)}(\cdot)$ is antisymmetric around zero, satisfying $\kappa^{(1)}(v)=-\kappa^{(1)}(-v)\ \forall\ v$.

(ii) $\kappa(\cdot)$, $\kappa^{(1)}(\cdot)$ and $\kappa^{(2)}(\cdot)$ are functions of bounded variation and, for each $d=1,\ldots,D$, the following classes of functions are Euclidean for the constant envelope $\overline{\kappa}$,

$$\mathscr{M}_1^d=\Big\{m:\mathcal{S}_X\longrightarrow\mathbb{R}\text{: } m(x)=\kappa\Big(\alpha\cdot g_d(x,\theta)+\beta\cdot g_d(s,\theta)\Big)\text{ for some } s\in\mathcal{S}_X,\ \theta\in\Theta,\ \alpha,\beta\in\mathbb{R}\Big\},$$
$$\mathscr{M}_2^d=\Big\{m:\mathcal{S}_X\longrightarrow\mathbb{R}\text{: } m(x)=\kappa^{(1)}\Big(\alpha\cdot g_d(x,\theta)+\beta\cdot g_d(s,\theta)\Big)\text{ for some } s\in\mathcal{S}_X,\ \theta\in\Theta,\ \alpha,\beta\in\mathbb{R}\Big\},$$
$$\mathscr{M}_3^d=\Big\{m:\mathcal{S}_X\longrightarrow\mathbb{R}\text{: } m(x)=\kappa^{(2)}\Big(\alpha\cdot g_d(x,\theta)+\beta\cdot g_d(s,\theta)\Big)\text{ for some } s\in\mathcal{S}_X,\ \theta\in\Theta,\ \alpha,\beta\in\mathbb{R}\Big\},$$

(iii) Let $\tau>0$ be as described in Assumption 1. The bandwidth sequence h_n is such that there exists $0<\epsilon<(\tau\wedge 1/2)$ such that $n^{1/2-\epsilon}\cdot\left(h_n^{2D}\wedge h_n^{D+2}\right)\longrightarrow\infty$ and $n^{1/2+\epsilon}\cdot h_n^M\longrightarrow 0$.

(iv) The bandwidth $b_n\longrightarrow 0$ used in the construction of $\widehat{T_1}$ and $\widehat{T_2}$ satisfies $\left(n^{1/2}\wedge r_n\right)\cdot h_n^{D+1}\cdot b_n\longrightarrow\infty$ and $n^{1/2+\delta_0}\cdot b_n^2\longrightarrow 0$ for some $\delta_0>0$.

We have provided a partial list of known classes of Euclidean functions above. For example, if $g(x,\theta)=x'\theta$, the Euclidean property in part (ii) of Assumption 4 follows directly from Lemma 22 in Nolan and Pollard (1987).

3.4.2 A linear representation result for $\widehat{T_1}-T_{1,F}$ and $\widehat{T_2}-T_{2,F}$

Two key results

We will proceed under the assumption that the transformation $\mathcal{B}$ satisfies the following smoothness and regularity conditions.

Assumption 5 ***(Smoothness II)*** *The following conditions hold for the transformation $\mathcal{B}$.*

(i) There exist finite constants $M_1>0$ and $M_2>0$ such that, for all $(x,t)\in\mathcal{X}\times\mathcal{T}$ and $F\in\mathcal{F}$,

$$\left\|\widehat{Q}(x,t,\widehat{\theta})-Q_F(x,t,\theta_F^*)\right\|\le M_2$$
$$\Longrightarrow\quad\left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta}))-\mathcal{B}(Q_F(x,t,\theta_F^*))\right|\le M_1\cdot\left\|\widehat{Q}(x,t,\widehat{\theta})-Q_F(x,t,\theta_F^*)\right\|$$

(ii) $\mathcal{B}(\cdot)$ *is twice continuously differentiable, with*

$$\underbrace{\nabla_Q \mathcal{B}(Q)}_{1\times P} \equiv \left(\frac{\partial \mathcal{B}(Q)}{\partial Q_1}, \ldots, \frac{\partial \mathcal{B}(Q)}{\partial Q_P}\right), \quad \underbrace{\nabla_{QQ'} \mathcal{B}(Q)}_{P\times P} \equiv \begin{pmatrix} \frac{\partial^2 \mathcal{B}(Q)}{\partial Q_1^2} & \cdots & \frac{\partial^2 \mathcal{B}(Q)}{\partial Q_1 \partial Q_P} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{B}(Q)}{\partial Q_P \partial Q_1} & \cdots & \frac{\partial^2 \mathcal{B}(Q)}{\partial Q_P^2} \end{pmatrix}$$

(iii) Let $\overline{Q} \equiv \overline{\Gamma} \cdot \overline{\omega}$, *where* $\overline{\Gamma}$ *and* $\overline{\omega}$ *are the constants described in Assumption 2. There exist finite constants* $C_Q > 0$ *and* $\overline{H}_Q > 0$, *such that*

$$||Q|| \leq \overline{Q} + C_Q \quad \Longrightarrow \quad \left\|\nabla_Q \mathcal{B}(Q)\right\| \leq \overline{H}_Q \ \textit{and} \ \left\|\nabla_{QQ'} \mathcal{B}(Q)\right\| \leq \overline{H}_Q \qquad \blacksquare$$

In Appendix A5 we show that the examples presented in Section 2.2.4 satisfy the restrictions in Assumption 5. Combining Assumptions 1-5, Proposition S1 and Section S3.5 in the online Econometric Supplement show that,

$$\sup_{F\in\mathcal{F}} P_F\left(\sup_{(x,t)\in\mathcal{X}} \left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) - \mathcal{B}(Q_F(x,t,\theta_F^*))\right| \geq b_n\right) \longrightarrow 0, \tag{16}$$

The next step towards our result analyzes the asymptotic properties of the term $\mathbb{1}\{\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) \geq -b_n\} - \mathbb{1}\{\mathcal{B}(Q_F(x,t,\theta_F^*)) \geq 0\}$. To this end, we introduce the following manageability assumption.

Assumption 6 ***(Manageability II)*** *For a given* $F \in \mathcal{F}$ *and* $b \in \mathbb{R}$, *define the following class of sets,* $\mathcal{C}_F(b) = \left\{(x,t) \in \mathcal{X} \times \mathcal{T} : \mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) \geq b\right\}$. *There exists* $\overline{D}_1^{VC} < \infty$ *and* $b_0 > 0$ *such that, for each* $F \in \mathcal{F}$, *the class of sets* $\mathscr{S}_F = \left\{\mathcal{C}_F(b) \text{ for some } b \in [-b_0, b_0]\right\}$ *is a VC class of sets with VC dimension bounded above by* $\overline{D}_1^{VC}$.

By the properties of VC classes of sets (see, e.g, Pakes and Pollard (1989, Lemma 2.5)), and the Euclidean properties of indicator functions over VC classes of sets (see Pakes and Pollard (1989, p. 1033)), there exist constants $(\overline{A}, \overline{V})$ such that, for each $F \in \mathcal{F}$, the following classes of indicator functions are Euclidean $(\overline{A}, \overline{V})$ for the constant envelope 1,

$$\left\{m : \mathcal{X} \times \mathcal{T} \longrightarrow \mathbb{R} : m(x,t) = \mathbb{1}\left\{-b \leq \mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) < 0\right\} \text{ for some } 0 < b \leq b_0\right\},$$
$$\left\{m : \mathcal{X} \longrightarrow \mathbb{R} : m(x) = \mathbb{1}\left\{-b \leq \mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) < 0\right\} \text{ for some } 0 < b \leq b_0 \text{ and } t \in \mathcal{T}\right\},$$
$$\left\{m : \mathcal{X} \longrightarrow \mathbb{R} : m(x) = \mathbb{1}\left\{\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) \geq 0\right\} \text{ for some } t \in \mathcal{T}\right\}.$$

This will yield manageability conditions for empirical processes involved in the asymptotic properties of $\widehat{T}_1$ and $\widehat{T}_2$.

A linear representation result for $\widehat{T_1} - T_{1,F}$

The last building block towards obtaining our main asymptotic result for $\widehat{T_1}$ is the following assumption, which involves smoothness conditions for some additional functionals, along with a regularity condition for the density of $\mathcal{B}(Q_F(X,t,\theta_F^*))$.

Assumption 7A ***(Smoothness and a regularity condition relevant for $\widehat{T_1}$)*** *As described in* (14)*, let $\mathcal{S}_{g,F}$ denote the support of $g(X,\theta_F^*)$. Recall that $\mathcal{G} \subset int\left(\mathcal{S}_{g,F}\right)$ for each $F \in \mathcal{F}$ by assumption. There exists a finite constant $\overline{\mu}_{\nabla_{\mathcal{B}}}$ such that, for each $p = 1,\ldots,P$,*

$$E_F\left[\left|\frac{\partial \mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\right|\right] \leq \overline{\mu}_{\nabla_{\mathcal{B}}} \quad \forall\, F \in \mathcal{F},$$

where the above is taken with respect to (X,t). For a given $g \equiv (g_1,\ldots,g_D)$ and $y \in \mathcal{S}_Y$, and for each $p = 1,\ldots,P$, let

$$\Omega_{T_1}^p(y,g) = E_F\left[\left(S_p(y,t) - \Gamma_{p,F}(X,t,\theta_F^*)\right)\frac{\partial \mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\phi(X,t)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right) \geq 0\right\}\middle| g(X,\theta_F^*) = g\right] \tag{17}$$

where the above expectation is taken with respect to (X,t), conditional on $g(X,\theta_F^*) = g$. There exists a set $\mathcal{G}'$ such that $\mathcal{G} \subset \mathcal{G}'$ such that,

(i) $f_g(g) \geq \underline{f}_g$ for all $g \in \mathcal{G}'$ and each $F \in \mathcal{F}$.

(ii) Let M be the integer described in Assumption 2. There exists a constant $\overline{C}_3 < \infty$ such that, for each p, and for all $0 \leq m \leq M+1$ and $(j_1,\ldots,j_D)$ such that $\sum_{d=1}^D j_d = m$,

$$\sup_{\substack{(g_1,\ldots,g_D)\in\mathcal{G}' \\ y \in \mathcal{S}_Y}} \left|\frac{\partial^m \Omega_{T_1}^p(y,g_1,\ldots,g_D)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right| \leq \overline{C}_3 \quad \forall\, F \in \mathcal{F}$$

(iii) ***(Behavior of $\mathcal{B}(Q_F(X,t,\theta_F^*))$ at zero from below):*** *There exist finite constants $\underline{b}_1 > 0$ and $\overline{C}_{\mathcal{B},1} > 0$ such that, for all $0 < b \leq \underline{b}_1$,*

$$E_F\left[\mathbb{1}\left\{-b \leq \mathcal{B}\left(Q_F(X,t,\theta_F^*)\right) < 0\right\}\cdot \mathbb{1}\left\{(X,t) \in \mathcal{X}\times\mathcal{T}\right\}\right] \leq \overline{C}_{\mathcal{B},1}\cdot b \quad \forall\, F \in \mathcal{F},$$

where the above expectation is taken with respect to (X,t).

The functional $\Omega_{T_1}^p$ described in Assumption 7A appears in a projection term of the Hoeffding decomposition of a U-process that comprises $\widehat{T_1}$. Combined with our bias-reducing properties for kernels and bandwidths, the bound in part (i) and smoothness conditions in part (ii) result in some bias terms vanishing asymptotically at the appropriate rate. Part (iii) essentially presupposes that, conditional on $(X,t) \in \mathcal{X} \times \mathcal{T}$, the functional $\mathcal{B}(Q_F(X,t,\theta_F^*))$ has a **density that is bounded** in a neighborhood $[-\underline{b}_1, 0)$ to the left of, but excluding, zero. This condition allows for $\mathcal{B}(Q_F(X,t,\theta_F^*))$ to have a point-mass at zero. We are now ready to present the main asymptotic result for $\widehat{T}_1$.

Proposition 1A *Let $\Delta \equiv \epsilon \wedge (\delta_0/2)$, where ϵ and δ_0 are as described in Assumption 4. If Assumptions 1-6 and 7A hold,*

$$\widehat{T}_1 = T_{1,F} + \frac{1}{n}\sum_{i=1}^{n} \psi_F^{T_1}(V_i) + \varepsilon_n^{T_1}, \quad \textit{where} \quad |\varepsilon_n^{T_1}| = o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \quad \textit{uniformly over } \mathcal{F}. \tag{18}$$

The influence function $\psi_F^{T_1}(V)$ has two key features,

$$\begin{aligned} &(i) \quad E_F\left[\psi_F^{T_1}(V)\right] = 0 \quad \forall\, F \in \mathcal{F}, \\ &(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*)) < 0 \,\middle|\, (X,t) \in \mathcal{X}_F^* \times \mathcal{T}\right) = 1 \implies P_F\left(\psi_F^{T_1}(V) = 0\right) = 1. \end{aligned} \tag{19}$$

Thus, if the functional inequality $\mathcal{B}\left(\Gamma_F(X,t,\theta_F^*)\right) \leq 0$ holds strictly for $F-a.e$ (x,t), the influence function $\psi_F^{T_1}(V)$ is equal to zero w.p.1.

The steps of the proof of Proposition 1A are described in Appendix A1, along with the precise description of the influence function $\psi_F^{T_1}(V)$, in Appendix A1.2. The step-by-step details are included in the online Econometric Supplement (Section S3). Let us outline the main steps. Let,

$$\widetilde{T}_{1,F} \equiv \frac{1}{n}\sum_{i=1}^{n} \mathcal{B}\left(\widehat{Q}(X_i,t_i,\widehat{\theta})\right) \mathbb{1}\{\mathcal{B}(Q_F(X_i,t_i,\theta_F^*)) \geq 0\}\, \phi(X_i,t_i)$$

$\widetilde{T}_{1,F}$ replaces the indicator function $\mathbb{1}\{\mathcal{B}\left(\widehat{Q}(X_i,t_i,\widehat{\theta})\right) \geq -b_n\}$ with $\mathbb{1}\{\mathcal{B}\left(Q_F(X_i,t_i,\theta_F^*)\right) \geq 0\}$. The **first step of the proof** is to show that, under Assumptions 1-6 and 7A, we have

$$\begin{aligned} \widehat{T}_1 =& T_{1,F} + \frac{1}{n}\sum_{i=1}^{n}\Big((\mathcal{B}(Q_F(X_i,t_i,\theta_F^*)))_+ \phi(X_i,t_i) - T_{1,F} \Big) \\ &+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \psi_F^{\mathcal{B}}(V_j, X_i, t_i, \theta_F^*, h_n) \mathbb{1}\{\mathcal{B}(Q_F(X_i,t_i,\theta_F^*)) \geq 0\}\, \phi(X_i,t_i) + \xi_{T_1,n}, \end{aligned} \tag{20}$$

where $\psi_F^{\mathcal{B}}$ is as described in equation (A-9) of the appendix and $\xi_{T_1,n} = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right)$ uniformly over $\mathcal{F}$, with $\epsilon > 0$ being the constant described in Assumption 4. From here, the **concluding step** consist of computing the Hoeffding decomposition (see Serfling (1980, pages 177-178) or Sherman (1994, equations (6)-(7))) of the U-process on the right-hand side of (20). The result in Proposition 1A follows from here using the maximal inequality results in Sherman (1994, Corollary 4A).

A linear representation result for $\widehat{T}_2 - T_{2,F}$

We obtain an analogous result to Proposition 1A for $\widehat{T}_2 - T_{2,F}$. The following restriction plays the same role as Assumption 7A.

Assumption 7B ***(Smoothness and a regularity condition relevant for $\widehat{T}_2$)*** *As described in* (14)*, let $\mathcal{S}_{g,F}$ denote the support of $g(X,\theta_F^*)$. Then $\mathcal{G} \subset int\left(\mathcal{S}_{g,F}\right)$ for each $F \in \mathcal{F}$. There exists a finite constant $\overline{\mu}_{\nabla_{\mathcal{B}}}$ such that, for each $p = 1,\ldots,P$,*

$$\sup_{t\in\mathcal{T}} E_F\left[\left|\frac{\partial\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\right|\right] \leq \overline{\mu}_{\nabla_{\mathcal{B}}} \quad \forall\ F \in \mathcal{F},$$

where the above is taken with respect to X. For each $F \in \mathcal{F}$, and a given $g \equiv (g_1,\ldots,g_D)$, $(y,t) \in \mathcal{S}_Y \times \mathcal{T}$, and for each $p = 1,\ldots,P$, let

$$\Omega_{T_0}^p(y,t,g) = E_F\left[\left(S_p(y,t) - \Gamma_{p,F}(X,t,\theta_F^*)\right)\frac{\partial\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\phi(X)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right) \geq 0\right\}\middle| g(X,\theta_F^*) = g\right] \tag{21}$$

where the above expectation is taken with respect to X, conditional on $g(X,\theta_F^*) = g$. There exists a set $\mathcal{G}'$ such that $\mathcal{G} \subset \mathcal{G}'$ (recall that $\mathcal{G}$ is our testing range for $g_F(x,\theta_F^*)$) such that,

(i) $f_g(g) \geq \underline{f}_g$ for all $g \in \mathcal{G}'$ and each $F \in \mathcal{F}$.

(ii) Let M be the integer described in Assumption 2. There exists a constant $\overline{C}_4 < \infty$ such that, for each p, and for all $0 \leq m \leq M+1$ and $(j_1,\ldots,j_D)$ such that $\sum_{d=1}^D j_d = m$,

$$\sup_{\substack{(g_1,\ldots,g_D)\in\mathcal{G}'\\(y,t)\in\mathcal{S}_Y\times\mathcal{T}}}\left|\frac{\partial^m\Omega_{T_0}^p(y,t,g_1,\ldots,g_D)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right| \leq \overline{C}_4 \quad \forall\ F \in \mathcal{F}$$

(iii) ***(Behavior of $\mathcal{B}(Q_F(X,t,\theta_F^*))$ at zero from below):*** *There exist finite constants $\underline{b}_2 > 0$ and $\overline{C}_{\mathcal{B},2} > 0$ such that, for all $0 < b \leq \underline{b}_2$,*

$$\sup_{t\in\mathcal{T}} E_F\left[\mathbb{1}\left\{-b \leq \mathcal{B}\left(Q_F(X,t,\theta_F^*)\right) < 0\right\}\cdot\mathbb{1}\left\{X \in \mathcal{X}\right\}\right] \leq \overline{C}_{\mathcal{B},2}\cdot b \quad \forall\ F \in \mathcal{F},$$

where the above expectation is taken with respect to X.

Our assumptions yield a result analogous to Proposition 1A, but for $\widehat{T}_2$.

Proposition 1B *Let $\Delta \equiv \epsilon \wedge (\delta_0/2)$, where ϵ and δ_0 are the positive constants described in Assumption 4. If Assumptions 1-6 and 7B hold, then*

$$\widehat{T}_2 = T_{2,F} + \frac{1}{n}\sum_{i=1}^{n} \psi_F^{T_2}(V_i) + \varepsilon_n^{T_2}, \quad \textit{where} \quad |\varepsilon_n^{T_2}| = o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \quad \textit{uniformly over } \mathcal{F}. \tag{22}$$

The influence function $\psi_F^{T_2}(V)$ has two key features,

$$\begin{aligned} &(i) \quad E_F\left[\psi_F^{T_2}(V)\right] = 0 \quad \forall\ F \in \mathcal{F}, \\ &(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*)) < 0 \,\middle|\, X \in \mathcal{X}_F^*\right) = 1 \ \textit{ for } \mathcal{W}\textit{–a.e } t \in \mathcal{T} \quad \Longrightarrow \ P_F\left(\psi_F^{T_2}(V) = 0\right) = 1. \end{aligned} \tag{23}$$

The influence function $\psi_F^{T_0}(V,t)$ described in Proposition 1B has two key features,

$$\begin{aligned} &(i) \quad E_F\left[\psi_F^{T_0}(V,t)\right] = 0 \quad \forall\ t \in \mathcal{T},\ \forall\ F \in \mathcal{F}, \\ &(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*)) < 0 \,\middle|\, X \in \mathcal{X}_F^*\right) = 1 \ \Longrightarrow \ P_F\left(\psi_F^{T_0}(V,t) = 0\right) = 1. \end{aligned} \tag{24}$$

From (24), we have the following properties for the influence function $\psi_F^{T_2}(V)$,

$$\begin{aligned} &(i) \quad E_F\left[\psi_F^{T_2}(V)\right] = 0 \quad \forall\ F \in \mathcal{F}, \\ &(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*)) < 0 \,\middle|\, X \in \mathcal{X}_F^*\right) = 1 \ \text{ for } \mathcal{W}\text{–a.e } t \in \mathcal{T} \quad \Longrightarrow \ P_F\left(\psi_F^{T_2}(V) = 0\right) = 1. \end{aligned} \tag{25}$$

The steps of the proof are analogous to those of Proposition 1A, and the details are included in the online Econometric Supplement (Section S3). In Appendix A1.3 we describe precisely the structure of the influence function $\psi_F^{T_2}(V)$.

3.5 Constructing test-statistics

We use the results in Propositions 1A and 1B to construct our tests and we present the details here. The first step is to construct estimators for $\sigma_{1,F}^2 \equiv E_F[\psi_F^{T_1}(V)^2]$ and $\sigma_{2,F}^2 \equiv E_F[\psi_F^{T_2}(V)^2]$.

3.5.1 Estimation of $\sigma_{1,F}^2 \equiv E_F[\psi_F^{T_1}(V)^2]$ and $\sigma_{2,F}^2 \equiv E_F[\psi_F^{T_2}(V)^2]$

Under the conditions of Propositions 1A and 1B, there exists a finite $\overline{\mu}_T > 0$ such that $E_F\left[\psi_F^{T_1}(V)^2\right] \leq \overline{\mu}_T$ and $E_F\left[\psi_F^{T_2}(V)^2\right] \leq \overline{\mu}_T$ for all $F \in \mathcal{F}$. From here, a Chebyshev inequality yields,

$$\left|\frac{1}{n}\sum_{i=1}^{n} \psi_F^{T_1}(V_i)^2 - \sigma_{1,F}^2\right| = o_p(1), \quad \text{and} \quad \left|\frac{1}{n}\sum_{i=1}^{n} \psi_F^{T_2}(V_i)^2 - \sigma_{2,F}^2\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{26}$$

We estimate $\sigma^2_{1,F}$ and $\sigma^2_{2,F}$ by constructing estimators for the influence functions $\psi_F^{T_1}(V)$ and $\psi_F^{T_2}(V)$. Appendix A2 describes the construction of the estimated influence functions under the following assumption.

Assumption 8 ***(An estimator for the influence function of $\widehat{\boldsymbol{\theta}}$)*** *We have an estimator $\widehat{\psi}^{\theta}(Z)$ for the influence function $\psi_F^{\theta}(Z)$ that satisfies,*

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\psi}^{\theta}(Z_i)-\psi_F^{\theta}(Z_i)\right\|^2 = o_p(1) \quad \textit{uniformly over } \mathcal{F}. \quad \blacksquare$$

Our estimators for $\sigma^2_{1,F}$ and $\sigma^2_{2,F}$ are $\widehat{\sigma}^2_1 \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}^{T_1}(V_i)^2$, and $\widehat{\sigma}^2_2 \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\psi}^{T_2}(V_i)^2$. Under the conditions of Propositions 1A and 1B, combined with Assumption 8, uniformly over $\mathcal{F}$ we have $\frac{1}{n}\sum_{i=1}^{n}|\widehat{\psi}^{T_1}(V_i)-\psi_F^{T_1}(V_i)|^2 = o_p(1)$, and $\frac{1}{n}\sum_{i=1}^{n}|\widehat{\psi}^{T_2}(V_i)-\psi_F^{T_2}(V_i)|^2 = o_p(1)$. From here, (26) yields

$$\left|\widehat{\sigma}^2_1-\sigma^2_{1,F}\right| = o_p(1), \quad \text{and} \quad \left|\widehat{\sigma}^2_2-\sigma^2_{2,F}\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{27}$$

3.5.2 Null hypotheses

Our population statistics are designed to test two different versions of our null hypothesis, depending on how we integrate out the index parameter t. These are

$$\begin{aligned}
&H_0^1 : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_{X,t}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.\\
&H_0^2 : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_X\otimes\mathcal{W}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.
\end{aligned} \tag{28}$$

$T_{1,F}$ and $T_{2,F}$ are designed to test H_0^1 and H_0^2, respectively.

3.5.3 Our proposed tests

It will be useful to introduce notation to distinguish the distributions that satisfy the functional inequalities in (28) and, within that class, those that satisfy them as strict inequalities. Let

$$\begin{aligned}
&\mathcal{F}_1^0 \equiv \{F\in\mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_{X,t}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.\}\\
&\mathcal{F}_2^0 \equiv \{F\in\mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_X\otimes\mathcal{W}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.\}
\end{aligned}$$

$\mathcal{F}_1^0$ and $\mathcal{F}_2^0$ denote, respectively, the sets of distributions for which the inequalities in H_0^1 and in H_0^2 are satisfied. Now let,

$$\begin{aligned}
&\overline{\mathcal{F}}_1 \equiv \{F\in\mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) < 0 \text{ for } F_{X,t}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.\}\\
&\overline{\mathcal{F}}_2 \equiv \{F\in\mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) < 0 \text{ for } F_X\otimes\mathcal{W}\text{–a.e } (x,t)\in\mathcal{X}_F^*\times\mathcal{T}.\}
\end{aligned}$$

$\overline{\mathcal{F}}_1 \subseteq \mathcal{F}_1^0$ and $\overline{\mathcal{F}}_2 \subseteq \mathcal{F}_2^0$ denote, respectively, the subfamilies of distributions for which the inequalities in H_0^1 and in H_0^2 are satisfied as *strict inequalities*. From part (ii) of (19), $\sigma_{1,F} = 0$ for all $F \in \overline{\mathcal{F}}_1$. Similarly, from part (ii) of (25), $\sigma_{2,F} = 0$ for all $F \in \overline{\mathcal{F}}_2$. Thus, in order to studentize $\widehat{T_1}$ and $\widehat{T_2}$ we need to regularize their estimated standard deviations. Let $\kappa_1 > 0$ and $\kappa_2 > 0$ denote pre-specified, arbitrarily small but strictly positive constants. We consider the test-statistics,

$$\widehat{t_1} = \frac{\sqrt{n}\widehat{T_1}}{(\widehat{\sigma_1} \vee \kappa_1)}, \quad \text{and} \quad \widehat{t_2} = \frac{\sqrt{n}\widehat{T_2}}{(\widehat{\sigma_2} \vee \kappa_2)}. \tag{29}$$

We regularize $\widehat{\sigma_1}$ by taking $(\widehat{\sigma_1} \vee \kappa_1)$ since, as we described in Proposition 1A, we have $\sigma_{1,F} = 0$ for all $F \in \overline{\mathcal{F}}_1$. Similarly, we use $(\widehat{\sigma_2} \vee \kappa_2)$ instead of $\widehat{\sigma_2}$ because Proposition 1B implies $\sigma_{2,F} = 0$ for all $F \in \overline{\mathcal{F}}_2$. Under the conditions of Proposition 1A we have

$$\widehat{t_1} = \begin{cases} \frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1}\vee\kappa_1)} & \forall\ F \in \overline{\mathcal{F}}_1, \\ \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_1}(V_i)}{(\widehat{\sigma_1}\vee\kappa_1)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1}\vee\kappa_1)} & \forall\ F \in \mathcal{F}_1^0 \setminus \overline{\mathcal{F}}_1, \\ \frac{\sqrt{n}\cdot T_{1,F}}{(\widehat{\sigma_1}\vee\kappa_1)} + \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_1}(V_i)}{(\widehat{\sigma_1}\vee\kappa_1)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1}\vee\kappa_1)} & \forall\ F \in \mathcal{F} \setminus \mathcal{F}_1^0 \end{cases} \tag{30A}$$

Note that $\left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1}\vee\kappa_1)}\right| \leq \left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{\kappa_1}\right|$. Therefore, from Proposition 1A,

$$\left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1} \vee \kappa_1)}\right| \leq \left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{\kappa_1}\right| = o_p\left(\frac{n^{1/2}}{n^{1/2+\Delta}}\right) = o_p\left(\frac{1}{n^{\Delta}}\right) = o_p(1) \quad \text{uniformly over } \mathcal{F}$$

with $\Delta > 0$ being the constant described in Proposition 1A. Thus, the first implication of (30A) is that $\widehat{t_1} = o_p(1)$ uniformly over $F \in \overline{\mathcal{F}}_1$. Thus, for any $c > 0$,

$$\lim_{n\to 0} \sup_{F \in \overline{\mathcal{F}}_1} P_F\left(\widehat{t_1} \geq c\right) = 0. \tag{31A}$$

Similarly, if the conditions of Proposition 1B hold,

$$\widehat{t_2} = \begin{cases} \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma_2}\vee\kappa_2)} & \forall\ F \in \overline{\mathcal{F}}_2, \\ \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_2}(V_i)}{(\widehat{\sigma_2}\vee\kappa_2)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma_2}\vee\kappa_2)} & \forall\ F \in \mathcal{F}_2^0 \setminus \overline{\mathcal{F}}_2, \\ \frac{\sqrt{n}\cdot T_{2,F}}{(\widehat{\sigma_2}\vee\kappa_2)} + \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_2}(V_i)}{(\widehat{\sigma_2}\vee\kappa_2)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma_2}\vee\kappa_2)} & \forall\ F \in \mathcal{F} \setminus \mathcal{F}_2^0 \end{cases} \tag{30B}$$

where, as in (30A), $|\varepsilon_n^{T_2}| = o_p\left(\frac{1}{n^{1/2+\Delta}}\right)$ uniformly over $\mathcal{F}$. Therefore, (30B) implies $\widehat{t_2} = o_p(1)$ uniformly over $F \in \overline{\mathcal{F}}_2$. Thus, for any $c > 0$,

$$\lim_{n\to 0} \sup_{F\in\overline{\mathcal{F}}_2} P_F\left(\widehat{t_2} \geq c\right) = 0. \tag{31B}$$

Let $\alpha \in (0,1)$ denote our target asymptotic significance level and let $z_{1-\alpha}$ denote the $(1-\alpha)^{th}$ quantile for the $\mathcal{N}(0,1)$ distribution. Based on the asymptotic properties summarized in (30A)-(31A) and (30B)-(31B), we consider the following rejection rules for H_0^1 and for H_0^2 in (28),

$$\text{Reject } H_0^1 \text{ iff } \widehat{t_1} \geq z_{1-\alpha} \qquad \text{Reject } H_0^2 \text{ iff } \widehat{t_2} \geq z_{1-\alpha} \tag{32}$$

From (30A) and (30B), a uniform Berry-Esseen condition would suffice for our proposed tests to be uniformly asymptotically level α. We add such an assumption next.

Assumption 9 ***(A sufficient condition for a uniform Berry-Esseen bound)*** *There exists a $B < \infty$ such that*

$$\frac{E_F\left[\left|\psi_F^{T_1}(V)\right|^3\right]}{\sigma_{1,F}^3} < B \quad \forall\, F \in \mathcal{F}\setminus\overline{\mathcal{F}}_1, \quad \textit{and} \quad \frac{E_F\left[\left|\psi_F^{T_2}(V)\right|^3\right]}{\sigma_{2,F}^3} < B \quad \forall\, F \in \mathcal{F}\setminus\overline{\mathcal{F}}_2 \quad \blacksquare$$

Assumption 9 allows for $\sigma_{1,F}$ and $\sigma_{2,F}$ to be arbitrarily close to zero over $\mathcal{F}\setminus\overline{\mathcal{F}}_1$ and $\mathcal{F}\setminus\overline{\mathcal{F}}_2$, respectively. By the Berry-Esseen Theorem (Lehmann and Romano (2005, Theorem 11.2.7)), the condition in Assumption 9 is sufficient to ensure that there exists $C > 0$ such that

$$\begin{aligned}
&\sup_{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_1} \sup_d \left| P_F\left(\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n \psi_F^{T_1}(V_i)}{\sigma_{1,F}} \leq d\right) - \Phi(d)\right| \leq \frac{C}{n^{1/2}}, \\
&\sup_{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_2} \sup_d \left| P_F\left(\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n \psi_F^{T_2}(V_i)}{\sigma_{2,F}} \leq d\right) - \Phi(d)\right| \leq \frac{C}{n^{1/2}}.
\end{aligned} \tag{33}$$

where $\Phi(\cdot)$ denotes the standard normal c.d.f. Take a given target asymptotic level $\alpha \in (0,1)$ and let $z_{1-\alpha}$ denote the $(1-\alpha)^{th}$ quantile of the standard normal distribution. Assumption 9 yields,

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_1:\\ \sigma_{1,F}\geq\kappa_1}} \left| P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi_F^{T_1}(V_i)}{(\sigma_{1,F}\vee\kappa_1)} \geq z_{1-\alpha}\right) - \alpha\right| = 0, \quad \limsup_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_1:\\ \sigma_{1,F}<\kappa_1}} P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi_F^{T_1}(V_i)}{(\sigma_{1,F}\vee\kappa_1)} \geq z_{1-\alpha}\right) \leq \alpha, \\
&\lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_2:\\ \sigma_{2,F}\geq\kappa_2}} \left| P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi_F^{T_2}(V_i)}{(\sigma_{2,F}\vee\kappa_2)} \geq z_{1-\alpha}\right) - \alpha\right| = 0, \quad \limsup_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}_2:\\ \sigma_{2,F}<\kappa_2}} P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\psi_F^{T_2}(V_i)}{(\sigma_{2,F}\vee\kappa_2)} \geq z_{1-\alpha}\right) \leq \alpha.
\end{aligned} \tag{34}$$

From (27), we have $\left|\frac{1}{(\widehat{\sigma_1}\vee\kappa_1)}-\frac{1}{(\sigma_{1,F}\vee\kappa_1)}\right|=o_p(1)$ and $\left|\frac{1}{(\widehat{\sigma_2}\vee\kappa_2)}-\frac{1}{(\sigma_{2,F}\vee\kappa_2)}\right|=o_p(1)$ uniformly over $\mathcal{F}\setminus\overline{\mathcal{F}}_1$ and $\mathcal{F}\setminus\overline{\mathcal{F}}_2$, respectively. Thus, from (34),

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_1^0\setminus\overline{\mathcal{F}}_1}P_F\left(\widehat{t_1}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{\substack{F\in\mathcal{F}_1^0\setminus\overline{\mathcal{F}}_1:\\ \sigma_{1,F}\geq\kappa_1}}\left|P_F\left(\widehat{t_1}\geq z_{1-\alpha}\right)-\alpha\right|=0,$$
$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_2^0\setminus\overline{\mathcal{F}}_2}P_F\left(\widehat{t_2}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{\substack{F\in\mathcal{F}_2^0\setminus\overline{\mathcal{F}}_2:\\ \sigma_{2,F}\geq\kappa_2}}\left|P_F\left(\widehat{t_2}\geq z_{1-\alpha}\right)-\alpha\right|=0.$$

Combining (31A)-(31B) with the above result, we have

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_1^0}P_F\left(\widehat{t_1}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{\substack{F\in\mathcal{F}_1^0\setminus\overline{\mathcal{F}}_1:\\ \sigma_{1,F}\geq\kappa_1}}\left|P_F\left(\widehat{t_1}\geq z_{1-\alpha}\right)-\alpha\right|=0,\\ &\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_2^0}P_F\left(\widehat{t_2}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{\substack{F\in\mathcal{F}_2^0\setminus\overline{\mathcal{F}}_2:\\ \sigma_{2,F}\geq\kappa_2}}\left|P_F\left(\widehat{t_2}\geq z_{1-\alpha}\right)-\alpha\right|=0.\end{aligned}\tag{35}$$

Thus, our proposed tests have *uniformly asymptotically level* α (Lehmann and Romano (2005, Definition 11.1.2)). Next, let us focus on $\mathcal{F}\setminus\mathcal{F}_1^0$ (the subset of distributions that violate H_0^1) and $\mathcal{F}\setminus\mathcal{F}_2^0$ (the subset of distributions that violate H_0^2). Take any sequence of distributions $F_n\in\mathcal{F}\setminus\mathcal{F}_1^0$ such that $\sqrt{n}\cdot T_{1,F_n}\geq\delta_nD$ for some fixed $D>0$ and a sequence of positive constants $\delta_n\to\infty$. From Assumption 9 and equation (33), we have that for any $c>0$,

$$\lim_{n\to\infty}P_{F_n}\left(\frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n\psi_{F_n}^{T_1}(V_i)}{(\sigma_{1,F_n}\vee\kappa_1)}+\frac{\sqrt{n}\cdot T_{1,F_n}}{(\widehat{\sigma_1}\vee\kappa_1)}\geq c\right)=1,\quad\text{and therefore}\quad \lim_{n\to\infty}P_{F_n}\left(\widehat{t_1}\geq c\right)=1.$$

And we have the analogous result for our proposed test for H_0^2. More generally, consider a sequence of distributions F_n such that $\left(\frac{\sigma_{1,F_n}\vee\kappa_1}{\sigma_{1,F_n}}\right)\to s_1^a$ and $\frac{\sqrt{n}\cdot T_{1,F_n}}{(\sigma_{1,F_n}\vee\kappa_1)}\to s_1^b$ (with $s_1^a=\infty$ and $s_1^b=\infty$ as special cases). For any such sequence we have

$$\lim_{n\to\infty}\left|P_{F_n}(\widehat{t_1}\geq z_{1-\alpha})-\left[1-\Phi\left(s_1^a\cdot(z_{1-\alpha}-s_1^b)\right)\right]\right|=0,$$

Following the terminology of Lee, Song, and Whang (2018, Definition 3), we say that our rejection rule for H_0^1 has *nontrivial asymptotic power* for a sequence $F_n\in\mathcal{F}\setminus\mathcal{F}_1^0$ if $\lim_{n\to\infty}P_{F_n}\left(\widehat{t_1}\geq z_{1-\alpha}\right)>\alpha$. It follows that our test for H_0^1 will have nontrivial power for F_n iff $s_1^a\cdot(z_{1-\alpha}-s_1^b)<z_{1-\alpha}$. An analogous result holds our proposed test for H_0^2. The following theorem summarizes the previous results and the properties of our proposed tests.

Theorem 1 *Consider the tests for H_0^1 and H_0^2 described by the rejection rules given in (32) for a target significance level α. If Assumptions 1-6, 7A-7B, 8 and 9 hold, these tests have the following properties.*

(i) Uniformly asymptotically level α: Our proposed tests satisfy,

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}_1^0} P_F\left(\widehat{t_1} \geq z_{1-\alpha}\right) \leq \alpha, \quad \lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}_1^0\setminus\overline{\mathcal{F}}_1: \\ \sigma_{1,F}\geq\kappa_1}} \left|P_F\left(\widehat{t_1} \geq z_{1-\alpha}\right) - \alpha\right| = 0,$$
$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}_2^0} P_F\left(\widehat{t_2} \geq z_{1-\alpha}\right) \leq \alpha, \quad \lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}_2^0\setminus\overline{\mathcal{F}}_2: \\ \sigma_{2,F}\geq\kappa_2}} \left|P_F\left(\widehat{t_2} \geq z_{1-\alpha}\right) - \alpha\right| = 0.$$

(ii) Consistency: Take any sequence of distributions $F_n \in \mathcal{F}\setminus\mathcal{F}_1^0$ such that $\sqrt{n}\cdot T_{1,F_n} \geq \delta_n D$ for some fixed $D > 0$ and a sequence of positive constants $\delta_n \to \infty$. We have $\lim_{n\to\infty} P_{F_n}\left(\widehat{t_1} \geq z_{1-\alpha}\right) = 1$. Similarly, take any sequence of distributions $F_n \in \mathcal{F}\setminus\mathcal{F}_2^0$ such that $\sqrt{n}\cdot T_{2,F_n} \geq \delta_n D$ for some fixed $D > 0$ and a sequence of positive constants $\delta_n \to \infty$. For any such sequence, we have $\lim_{n\to\infty} P_{F_n}\left(\widehat{t_2} \geq z_{1-\alpha}\right) = 1$.

(iii) Nontrivial asymptotic power against local alternatives: Consider a sequence of distributions F_n such that $\left(\frac{\sigma_{1,F_n}\vee\kappa_1}{\sigma_{1,F_n}}\right) \to s_1^a$ and $\frac{\sqrt{n}\cdot T_{1,F_n}}{(\sigma_{1,F_n}\vee\kappa_1)} \to s_1^b$ (with $s_1^a = \infty$ and $s_1^b = \infty$ as special cases). Then, $\lim_{n\to\infty} P_{F_n}\left(\widehat{t_1} \geq z_{1-\alpha}\right) > \alpha$ iff $s_1^a\cdot(z_{1-\alpha} - s_1^b) < z_{1-\alpha}$. Similarly, take any sequence of distributions F_n such that $\left(\frac{\sigma_{2,F_n}\vee\kappa_2}{\sigma_{2,F_n}}\right) \to s_2^a$ and $\frac{\sqrt{n}\cdot T_{2,F_n}}{(\sigma_{2,F_n}\vee\kappa_2)} \to s_2^b$ (with $s_2^a = \infty$ and $s_2^b = \infty$ as special cases). Then, $\lim_{n\to\infty} P_{F_n}\left(\widehat{t_2} \geq z_{1-\alpha}\right) > \alpha$ iff $s_2^a\cdot(z_{1-\alpha} - s_2^b) < z_{1-\alpha}$.

3.5.4 A stronger version of Theorem 1 when $\sigma_{1,F}$ and $\sigma_{2,F}$ are assumed to be bounded away from zero over $\mathcal{F}\setminus\overline{\mathcal{F}}_1$ and $\mathcal{F}\setminus\overline{\mathcal{F}}_2$

Assumption 9 does not preclude $\sigma_{1,F}$ and $\sigma_{2,F}$ from becoming arbitrarily close to zero over $\mathcal{F}\setminus\overline{\mathcal{F}}_1$ and $\mathcal{F}\setminus\overline{\mathcal{F}}_2$, respectively. Consider the following stronger version of Assumption 9.

Assumption 9' ***(A strengthening of Assumption 9)*** *There exists $\underline{\sigma} > 0$ and $B < \infty$ such that $\sigma_{1,F} \geq \underline{\sigma}$ and $E_F\left[\left|\psi_F^{T_1}(V)\right|^3\right] \leq B$ for all $F \in \mathcal{F}\setminus\overline{\mathcal{F}}_1$, and $\sigma_{2,F} \geq \underline{\sigma}$ and $E_F\left[\left|\psi_F^{T_2}(V)\right|^3\right] \leq B$ for all $F \in \mathcal{F}\setminus\overline{\mathcal{F}}_2$.*

Let $\kappa_{1,n}$ and $\kappa_{2,n}$ be any pair of positive sequences such that $\kappa_{1,n} \to 0$, $n^\Delta\cdot\kappa_{1,n} \to \infty$, and $\kappa_{2,n} \to 0$, $n^\Delta\cdot\kappa_{2,n} \to \infty$, where $\Delta > 0$ is the constant described in Proposition 1A (and 1A). Under Assumption 9' we can replace the constants κ_1 and κ_2 with the sequences sequences $\kappa_{1,n} \to 0$ and $\kappa_{2,n} \to 0$. Our modified test-statistics would then be

$$\widetilde{t_1} = \frac{\sqrt{n}\widehat{T_1}}{(\widehat{\sigma_1}\vee\kappa_{1,n})}, \quad \text{and} \quad \widetilde{t_2} = \frac{\sqrt{n}\widehat{T_2}}{(\widehat{\sigma_2}\vee\kappa_{2,n})}.$$

Going back to (30A)-(30B), note that $\left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{(\widehat{\sigma_1}\vee\kappa_{1,n})}\right| \leq \left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_1}}{\kappa_{1,n}}\right| = o_p\left(\frac{n^{1/2}}{n^{1/2+\Delta}\cdot\kappa_{1,n}}\right) = o_p\left(\frac{1}{n^\Delta\cdot\kappa_{1,n}}\right) = o_p(1)$ and

$\left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma_2}\vee\kappa_{2,n})}\right| \leq \left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{\kappa_{2,n}}\right| = o_p\left(\frac{n^{1/2}}{n^{1/2+\Delta}\cdot\kappa_{2,n}}\right) = o_p\left(\frac{1}{n^{\Delta}\cdot\kappa_{2,n}}\right) = o_p(1)$, uniformly over $\mathcal{F}$. Therefore, the results in (31A) and (31B) are preserved for $\widetilde{t_1}$ and $\widetilde{t_2}$, respectively. The stronger conditions in Assumption 9' imply that

$$\sup_{F\in\mathcal{F}\backslash\overline{\mathcal{F}}_1}\left|\frac{\sigma_{1,F}\vee\kappa_{1,n}}{\sigma_{1,F}}-1\right|\to 0 \quad\text{and}\quad \sup_{F\in\mathcal{F}\backslash\overline{\mathcal{F}}_2}\left|\frac{\sigma_{2,F}\vee\kappa_{2,n}}{\sigma_{2,F}}-1\right|\to 0,$$

and thus, by the Berry-Esseen Theorem and the conditions in Assumption 9',

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}\backslash\overline{\mathcal{F}}_1}\left|P_F\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi_F^{T_1}(V_i)}{(\sigma_{1,F}\vee\kappa_{1,n})}\geq z_{1-\alpha}\right)-\alpha\right|=0,\ \lim_{n\to\infty}\sup_{F\in\mathcal{F}\backslash\overline{\mathcal{F}}_2}\left|P_F\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi_F^{T_2}(V_i)}{(\sigma_{2,F}\vee\kappa_{2,n})}\geq z_{1-\alpha}\right)-\alpha\right|=0.$$

This strengthens the result in (34). These results combined yield,

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_1^0}P_F\left(\widetilde{t_1}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{F\in\mathcal{F}_1^0\backslash\overline{\mathcal{F}}_1}\left|P_F\left(\widetilde{t_1}\geq z_{1-\alpha}\right)-\alpha\right|=0,$$
$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_2^0}P_F\left(\widetilde{t_2}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{F\in\mathcal{F}_2^0\backslash\overline{\mathcal{F}}_2}\left|P_F\left(\widetilde{t_2}\geq z_{1-\alpha}\right)-\alpha\right|=0.$$

This is a stronger version of the result in (35) for $\widehat{t_1}$ and $\widehat{t_2}$. Under Assumption 9', replacing $\widehat{t_1}$ and $\widehat{t_2}$ with $\widetilde{t_1}$ and $\widetilde{t_2}$, the result in part (i) of Theorem 1 (uniformly asymptotically level α) strengthens to the following,

$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_1^0}P_F\left(\widetilde{t_1}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{F\in\mathcal{F}_1^0\backslash\overline{\mathcal{F}}_1}\left|P_F\left(\widetilde{t_1}\geq z_{1-\alpha}\right)-\alpha\right|=0,$$
$$\limsup_{n\to\infty}\sup_{F\in\mathcal{F}_2^0}P_F\left(\widetilde{t_2}\geq z_{1-\alpha}\right)\leq\alpha,\quad \lim_{n\to\infty}\sup_{F\in\mathcal{F}_2^0\backslash\overline{\mathcal{F}}_2}\left|P_F\left(\widetilde{t_2}\geq z_{1-\alpha}\right)-\alpha\right|=0.$$

Part (ii) of Theorem 1 (consistency) remains the same and part (iii) (nontrivial asymptotic power against local alternatives) is preserved after we replace κ_1 and κ_2 with $\kappa_{1,n}$ and $\kappa_{2,n}$, respectively and we redefine the limits s_1^a, s_1^b, s_2^a and s_2^b accordingly.

3.6 On the choice of tuning parameters

Our procedure uses three tuning parameters: the bandwidth sequences h_n and b_n, along with the regularization constants κ_1 (for $\widehat{t_1}$) or κ_2 (for $\widehat{t_2}$), which can be replaced with sequences $\kappa_{1,n}\to 0$ or $\kappa_{2,n}\to 0$ under the conditions of Assumption 9'. While we leave the development of a general theory of how to choose these tuning parameters for future work, we can provide recommendations based on the results of our Monte Carlo experiments in Section 5. First, we consider a bandwidth sequence of the form $h_n = c_h\cdot\|\widehat{\sigma}(g(X,\widehat{\theta}))\|\cdot n^{-\alpha_h}$, where $\alpha_h>0$ denotes the rate of convergence of h_n. Next, we set b_n and the regularization tuning parameters κ_1 and κ_2 proportional to a measure of the scale of the transformation $\mathcal{B}(\cdot)$. If $\mathcal{B}(\cdot)$ is bounded[6], let $\overline{\mathcal{B}} \equiv \sup|\mathcal{B}(\cdot)|$. In this case we can set $\kappa_\ell = c_k \cdot \overline{\mathcal{B}}$ for $\ell = 1, 2$ and $b_n = c_b \cdot \overline{\mathcal{B}} \cdot n^{-\alpha_b}$. In Appendix A3, we show that if we set $\alpha_h = \frac{1}{4(D+1)} - \frac{\epsilon' + \delta'}{2(D+1)}$ and $\alpha_b = \frac{1}{4} + \Delta_b$, where $\frac{\epsilon'}{2} < \Delta_b < \frac{\epsilon' + \delta'}{2}$, with $\epsilon' > 0$ and $\delta' > 0$ small enough (this is made precise in Appendix A3), then the bandwidth convergence restrictions in Assumption 4 are satisfied with $M \geq 2D + 3$. Thus, we can use a bias-reducing kernel of order $M = 2D + 3$.

Regarding the choice of the constants c_h, c_κ and c_b, in our Monte Carlo experiments we found the following. First, Assumption 4 requires a relatively slow rate of convergence α_h. We find that choosing values of c_h slightly smaller than some of the usual choices in the literature yields good results. In our experiments we set $c_h = 0.7$, which is about 1/3 below the "rule of thumb" value of 1.06 (see Silverman (1986, Equation 3.28)). Next, to concentrate on the effect of b_n, we set the regularization parameter κ_1 to $c_\kappa = 10^{-10}$ and we focused on b_n. We find that, in order to reduce the risk of over-rejection in finite samples when the inequalities hold but are binding with positive probability, it is best to set a value of b_n that is nontrivially bounded away from zero. In our experiments we focus on values of c_b between 0.5 and 1 and we find that the finite-sample size and power properties of our procedure resemble closely the asymptotic results in Theorem 1. We find that even conservative values of c_b (e.g, $c_b = 1$ in our experiments) are capable of generating nontrivial power to detect violations that occur with probability as small as 1%.

4 Testing multiple inequalities

Our approach can be readily extended to testing multiple functional inequalities of the type described in (1). Suppose we have a model that predicts,

$$\mathcal{B}_r(\Gamma_F(x, t, \theta_F^*)) \leq 0 \;\text{ for } F\text{–a.e } x \in \mathcal{S}_X \text{ and } \forall\, t \in \mathcal{T}, \text{ for each } r = 1, \ldots, \mathcal{R}.$$

Suppose each transformation $\mathcal{B}_r$ satisfies the condition in (6) for weight functions $(\omega_{p,r})_{p=1}^{P}$, so

$$\begin{aligned}
&\mathcal{B}_r(\Gamma_1 \cdot \omega_{1,r}, \ldots, \Gamma_P \cdot \omega_{P,r}) = \mathcal{B}_r(\Gamma_1, \ldots, \Gamma_P) \cdot \mathcal{H}_r(\omega_{1,r}, \ldots, \omega_{P,r}),\\
&\text{where} \quad \begin{cases} \mathcal{H}_r(\cdot) \geq 0, \\ \mathcal{H}_r(\omega_{1,r}, \ldots, \omega_{P,r}) > 0 \iff \omega_{p,r} > 0 \;\forall\, p. \end{cases}
\end{aligned}$$

[6]If $|\mathcal{B}(\cdot)|$ is not bounded by a constant, we can use $\overline{\mathcal{B}} \equiv \sup_{(x,t) \in \mathcal{X} \times \mathcal{T}} \left|\mathcal{B}(\widehat{Q}(x, t, \widehat{\theta}))\right|$.

for each $r = 1,\ldots,\mathcal{R}$. Assume for simplicity that we use the same target testing range $\mathcal{X}_F^* = \left\{x \in \mathcal{S}_X : x \in \mathcal{X} \text{ and } g(x,\theta_F^*) \in \mathcal{G}\right\}$ for each one of the $\mathcal{R}$ inequalities and, as in (5), denote

$$\begin{aligned}
Q_{p,F}^r(x,t,\theta_F^*) &\equiv \Gamma_{p,F}(x,t,\theta_F^*) \cdot \omega_{p,r}(g(x,\theta_F^*)), \\
Q_F^r(x,t,\theta_F^*) &\equiv \left(Q_{1,F}^r(x,t,\theta_F^*),\ldots,Q_{P,F}^r(x,t,\theta_F^*)\right)', \\
\omega_r(g(x,\theta_F^*) &\equiv (\omega_{1,r}(g(x,\theta_F^*)),\ldots,\omega_{P,r}(g(x,\theta_F^*)))'.
\end{aligned}$$

Then (7) holds for each $r = 1,\ldots,\mathcal{R}$,

$$\begin{aligned}
&\mathcal{B}_r\left(Q_F^r(x,t,\theta_F^*)\right) = \mathcal{B}_r\left(\Gamma_F(x,t,\theta_F^*)\right) \cdot \mathcal{H}_r\left(\omega^r(g(x,\theta_F^*))\right), \\
&\text{where} \begin{cases} \mathcal{H}_r(\cdot) \geq 0, \\ \mathcal{H}_r(\omega_r(g(x,\theta_F^*)) > 0 \iff g^r(x,\theta_F^*) \in \mathcal{G}, \end{cases}
\end{aligned}$$

Let $\phi_r(x,t)$ be a pre-specified function satisfying $\phi_r(x,t) \geq 0$ and $\phi_r(x,t) > 0$ if and only if $(x,t) \in \mathcal{X} \times \mathcal{T}$. For brevity, let us focus on the population statistic $T_{1,F}$ since our treatment of $T_{2,F}$ would follow analogous steps. One way to generalize $T_{1,F}$ is by aggregating the $\mathcal{R}$ restrictions as[7],

$$\begin{aligned}
T_{1,F} &\equiv \sum_{r=1}^{\mathcal{R}} E_{F_{X,t}}\left[\left(\mathcal{B}_r\left(Q_F^r(X,t,\theta_F^*)\right)\right)_+ \phi_r(X,t)\right] = \sum_{r=1}^{\mathcal{R}}\left[\int_{x,t}\left(\mathcal{B}_r\left(Q_F^r(x,t,\theta_F^*)\right)\right)_+ \phi_r(x,t)dF_{X,t}(x,t)\right] \\
&= \sum_{r=1}^{\mathcal{R}}\left[\int_{x,t}\left(\mathcal{B}_r\left(\Gamma_F(x,t,\theta_F^*)\right)\right)_+ \mathcal{H}_r(\omega^r(g(x,\theta_F^*))\phi_r(x,t)dF_{X,t}(x,t)\right]
\end{aligned}$$

As in the case of a single functional inequality, $T_{1,F}$ has the following key properties,

- $T_{1,F} \geq 0$.
- $T_{1,F} = 0$ if and only if $\mathcal{B}_r(\Gamma_F(x,t,\theta_F^*)) \leq 0$ for $F_{X,t}$–a.e $(x,t) \in \mathcal{X}_F^* \times \mathcal{T}$, for each $r = 1,\ldots,\mathcal{R}$.

To construct an estimator $\widehat{T}_1$, we would first construct estimators $\widehat{Q}_p^r(x,t,\widehat{\theta})$ for each r in the manner described in equation (12). Our estimator for $T_{1,F}$ is a generalization of (13),

$$\widehat{T}_1 \equiv \sum_{r=1}^{\mathcal{R}}\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{B}_r\left(\widehat{Q}^r(X_i,t_i,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}_r\left(\widehat{Q}^r(X_i,t_i,\widehat{\theta})\right) \geq -b_n\right\}\phi_r(X_i,t_i)\right] \tag{13'}$$

We can use a different sequence b_n^r for each restriction, as long as each one of them satisfies the bandwidth convergence conditions described in Assumption 4. Maintaining that Assumptions 5, 6 and 7A are satisfied for each of the $\mathcal{R}$ restrictions, then the results in (A-8) and (A-9) are

[7] We can use specific weights for each of the inequalities. For simplicity we show the case where they are each weighted equally.

satisfied for each $\mathcal{B}(\widehat{Q}^r(X_i,t_i,\widehat{\theta})) - \mathcal{B}(Q^r(X_i,t_i,\widehat{\theta}))$ and each of the summands in (13') satisfies the linear representation result in Proposition 1A. From here, we have

$$\widehat{T}_1 = T_{1,F} + \frac{1}{n}\sum_{i=1}^{n}\psi_F^{T_1}(V_i) + \varepsilon_n^{T_1}, \quad \text{where} \quad |\varepsilon_n^{T_1}| = o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \quad \text{uniformly over } \mathcal{F}.$$

$\Delta > 0$ is the constant described in Proposition 1A. The influence function $\psi_F^{T_1}(V_i)$ corresponds to the sum of the influence functions for each of the $\mathcal{R}$ restrictions, each one with the structure described in Appendix A1.2. From (19), $\psi_F^{T_1}(V)$ has two key features,

$$\begin{aligned}
&(i) \quad E_F\left[\psi_F^{T_1}(V)\right] = 0 \quad \forall\, F \in \mathcal{F}, \\
&(ii) \quad P_F\left(\mathcal{B}_r\left(\Gamma_F(X,t,\theta_F^*)\right) < 0 \;\middle|\; (X,t) \in \mathcal{X}_F^* \times \mathcal{T}\right) = 1 \text{ for each } r = 1,\ldots,\mathcal{R} \implies P_F\left(\psi_F^{T_1}(V) = 0\right) = 1.
\end{aligned}$$

Thus, if each of the $\mathcal{R}$ inequalities holds strictly F–a.s, then the influence function $\psi_F^{T_1}(V)$ is F–a.s equal to zero. From here we estimate $\sigma_{1,F}^2 \equiv E_F[\psi_F^{T_1}(V)^2]$ by constructing an estimator for $\psi_F^{T_1}(V)$ in the manner described in Appendix A2. Under Assumption 8, the resulting estimator $\widehat{\sigma}_1^2$ will have the consistency features described in (27). Our null hypothesis is now,

$$H_0^1 : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_{X,t}\text{–a.e } (x,t) \in \mathcal{X}_F^* \times \mathcal{T}, \text{ for each } r = 1,\ldots,\mathcal{R}.$$

Our test-statistic would once again be of the form $\widehat{t}_1 = \frac{\sqrt{n}\cdot\widehat{T}_1}{(\widehat{\sigma}_1 \vee \kappa_1)}$, as in (29). To extend the results from Section 3.5.3, we first re-define,

$$\begin{aligned}
&\mathcal{F}_1^0 \equiv \{F \in \mathcal{F} : \mathcal{B}_r(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F_{X,t}\text{–a.e } (x,t) \in \mathcal{X}_F^* \times \mathcal{T}, \text{ for each } r = 1,\ldots,\mathcal{R}.\}, \\
&\overline{\mathcal{F}}_1 \equiv \{F \in \mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) < 0 \text{ for } F_{X,t}\text{–a.e } (x,t) \in \mathcal{X}_F^* \times \mathcal{T}, \text{ for each } r = 1,\ldots,\mathcal{R}.\}
\end{aligned}$$

From here, if we maintain the integrability condition in Assumption 9, the test that rejects H_0^1 iff $\widehat{t}_1 \geq z_{1-\alpha}$ has the asymptotic properties described in Theorem 1.

4.1 Comparison to existing methods

To our knowledge, this is the first paper devoted to testing functional inequalities that are conditioned on functions that are estimated in a first step. As such, there are no existing procedures that our results can be directly compared with. However, we can do a general comparison to existing conditional moment inequalities (CMI) methods. We note first that methods based on "instrument functions" (e.g, Andrews and Shi (2013)) cannot be used to test inequalities that involve nonlinear transformations of conditional moments (such as our covariance or affiliation examples), while our approach can be applied in such cases. Our method is an extension of Aradillas-López, Gandhi, and Quint (2016) to the case of estimated conditioning functions. It shares conceptual similarities to other CMI criterion function based approaches such as those of Andrews and Shi (2013), Lee, Song, and Whang (2013), Lee, Song, and Whang (2018), Armstrong (2015), Armstrong (2014), Chetverikov (2017), Armstrong and Chan (2016) and Armstrong (2018), but with key differences. Regarding the scaling of the CMI violations, while the aforementioned methods use test-statistics that measure violations scaled by their standard errors, ours first aggregates these violations and then scales the aggregate violation. Our tuning parameter b_n is similar to that used by Armstrong (2014) when scaling individual moment inequality violations. Armstrong (2014) shows that for a test based on a Kolmogorov-Smirnov statistic this can lead to improvements in estimation rates and local asymptotic power relative to using bounded weights. For the statistic considered here, which is based on aggregate moment inequality violations, truncation through the decreasing sequence b_n ensures that the violation is asymptotically weighted by its inverse standard error which, combined with our regularization, is used to establish the asymptotic validity of fixed standard normal critical values.

The use of the decreasing sequence b_n to measure violations of the inequalities allows our procedure to adapt asymptotically to the measure of the contact sets[8]. This avoids the need to estimate the contact sets in a first step, as is done in the method of Lee, Song, and Whang (2018). It also helps us avoid conservative methods based on least-favorable configurations where the standard errors are computed assuming that the inequalities are binding everywhere, as is done, e.g, in Lee, Song, and Whang (2013). Regularizing the estimator for the asymptotic variance of our statistic allows us to standardize it in a way that produces asymptotically pivotal properties. Adapting asymptotically to the contact sets and the pivotal features of our test are novel features of our approach, shared by the conditional functional inequalities test proposed in Aradillas-López, Gandhi, and Quint (2016) and, as we showed above, these features are readily extended to the case of testing multiple inequalities. In Aradillas-López, Gandhi, and Quint (2016), the authors show that the type of one-sided Cramér–von Mises (CvM) test-statistic we employ can perform as well or better than other non-CvM tests, including methods based on sup-norm statistics, particularly in cases where the nonparametric functions are flat near the contact sets. On the other hand, sup-norm statistics would out-perform procedures like ours when violations to the inequalities take the form of localized spikes. Finally we note that all existing methods require the choice of either tuning parameters or instrument functions and that, like our case, developing a general theory of how to choose these tuning parameters has been left to future work. Characterizing the asymptotic properties of existing methods when moments are conditioned on estimated functions is outside the scope of this paper.

[8]We refer to the contact sets as the set of values of the conditioning variables and the index parameter where the inequalities are binding

5 Monte Carlo experiments

We apply our method to test for conditional, first-order stochastic dominance, $F_{Y_1|X}(t|X) \leq F_{Y_2|X}(t|X)$ $F-$a.e X, $\forall\ t$. Let $Y \equiv (Y_1, Y_2)$ and $S(Y,t) \equiv \mathbb{1}\{Y_1 \leq t\} - \mathbb{1}\{Y_2 \leq t\}$. Our FOSD conjecture is

$$E_F[S(Y,t)|X] \leq 0\ F\text{−a.e } X,\ \forall\ t \tag{34}$$

5.1 Designs

Our random vector X will include eight independent, continuously distributed covariates, with $X_1, X_5 \sim \mathcal{N}(0,1)$, $X_2, X_6 \sim$ logistic, $X_3, X_7 \sim$ log-normal, and $X_4, X_8 \sim U[-1,1]$. In addition, we have two i.i.d, unobservable shocks, $\varepsilon_1, \varepsilon_2 \sim \mathcal{N}(0,1)$, independent of X. Let $m_I(X) \equiv -X_1 + X_2 + X_3 + X_4$ and $m_{II}(X) \equiv -X_5 + X_6 + X_7 + X_8$. We produced several data generating processes (DGPs), described in the following table.

Table 1: Monte Carlo designs

Design	Description	Is the FOSD inequality (34) satisfied?
DGP 1	$Y_1 = m_I(X) \vee m_{II}(X) + \varepsilon_1$ $Y_2 = m_I(X) \wedge m_{II}(X) + \varepsilon_2$	**Yes**, and it is satisfied as a **strict inequality** w.p.1 for each $t \in \mathbb{R}$.
DGP 2	$Y_1 = m_I(X) + \varepsilon_1$ $Y_2 = m_I(X) + \varepsilon_2$	**Yes**, and it holds as an **equality** w.p.1 for each $t \in \mathbb{R}$.
DGP 3	$Y_1 = m_I(X) + \varepsilon_1$ $Y_2 = m_I(X) \vee m_{II}(X) + \varepsilon_2$	**No**. It is violated with probability **50%** for each $t \in \mathbb{R}$.
DGP 4	$Y_1 = m_I(X) + 1.5 + \varepsilon_1$ $Y_2 = (m_I(X)+1.5) \vee (m_{II}(X) - 1.5) + \varepsilon_2$	**No**. It is violated with probability **≈ 20%** for each $t \in \mathbb{R}$.
DGP 5	$Y_1 = m_I(X) + 4.5 + \varepsilon_1$ $Y_2 = (m_I(X) + 4.5) \vee (m_{II}(X) - 2) + \varepsilon_2$	**No**. It is violated with probability **≈ 5%** for each $t \in \mathbb{R}$.
DGP 6	$Y_1 = m_I(X) + 5 + \varepsilon_1$ $Y_2 = (m_I(X) + 5) \vee (m_{II}(X) - 5) + \varepsilon_2$	**No**. It is violated with probability **≈ 1%** for each $t \in \mathbb{R}$.

First, we want to evaluate our procedure in the two extreme scenarios under the null hypothesis: when the inequalities are satisfied as strict inequalities w.p.1. (DPG 1), and when they are binding w.p.1 (DGP 2). Second, we want to study the power properties of our method as the probability that the inequalities are violated diminishes. In our designs (DGPs 3-6) the probability of a violation goes from 50% to only 1%.

5.2 Conditioning function employed

X includes eight continuously distributed covariates, representing the type of scenario where researchers would want to aggregate X into a lower-dimensional function. We apply our methodology using a single index $g(X,\theta) = X'\theta$, estimated as follows. Let $\Delta Y \equiv Y_1 - Y_2$. We run an OLS regression of ΔY against X to compute $\widehat{\theta}$ and we use $g(X,\widehat{\theta}) = X'\widehat{\theta}$. This index is not designed to be "optimal" in any way, but instead is meant to represent a situation where a researcher wants to aggregate X into an arbitrarily chosen function. Group $Z_i \equiv (\Delta Y_i, X_i')'$, and define $\theta_F^* \equiv (E_F[XX'])^{-1} \cdot E_F[X\Delta Y]$, $\nu_i \equiv \left(\Delta Y_i - X_i'\theta_F^*\right)$, and $\psi_F^\theta(Z_i) \equiv (E_F[XX'])^{-1} \cdot X_i \nu_i$. Note that $E_F[\psi_F^\theta(Z_i)] = 0$. As shown in Appendix A4.1, under conditions that are satisfied by our designs, the OLS estimator satisfies the conditions in Assumption 1, with $\widehat{\theta} = \theta_F^* + \frac{1}{n}\sum_{i=1}^n \psi_F^\theta(Z_i) + \varepsilon_n^\theta$. Appendix A4.1 also shows that the remaining conditions in Assumption 1 are satisfied with $r_n = n^{1/2}$, any τ and $\overline{\delta}$ such that $0 < \tau < 1/2$, and $0 < \overline{\delta} < 1/2$, and for any $q \geq 2$. Thus, our conditioning function is $g(X, \theta_F^*) \equiv x'\theta_F^*$ and our goal is to test

$$E_F[S(Y,t)|X'\theta_F^*] \leq 0 \;\; F\text{–a.s}, \forall\, t \tag{34'}$$

Note that (34') follows from (34) by iterated expectations since $X'\theta_F^*$ is a deterministic function of X.

5.3 Test statistic used

We applied our test based on the statistic $\widehat{t_1}$. Going back to (13), we have in this case

$$\widehat{T}_1 \equiv \frac{1}{n}\sum_{i=1}^n \widehat{Q}(X_i, t_i, \widehat{\theta}) \mathbb{1}\left\{\widehat{Q}(X_i, t_i, \widehat{\theta}) \geq -b_n\right\} \phi(X_i, t_i).$$

We chose $t_i = Y_{1i}$ as the index-parameter variable. Our estimator $\widehat{Q}(x,t,\widehat{\theta}) \equiv \frac{\widehat{R}(x,t,\widehat{\theta})}{\widehat{f_g}(g(x,\widehat{\theta}))}$, is as described in (12), which simplifies to the following here,

$$\widehat{R}_p(x,t,\theta) = \frac{1}{n \cdot h_n}\sum_{j=1}^n \Big(\mathbb{1}\{Y_{1j} \leq t\} - \mathbb{1}\{Y_{2j} \leq t\}\Big)\omega(X_j'\widehat{\theta})K\left(\frac{(X_j - x)'\widehat{\theta}}{h_n}\right), \quad \widehat{f_g}(g(x,\theta)) \quad = \frac{1}{n \cdot h_n}\sum_{j=1}^n K\left(\frac{(X_j - x)'\widehat{\theta}}{h_n}\right)$$

Our tuning parameters and testing range are described next.

5.3.1 Tuning parameters, kernel used and target testing range

The bandwidth sequences h_n and b_n and the tuning parameter κ_1 had the general form described in Section 3.6. Accordingly, let $\widehat{\sigma}(X'\widehat{\theta})$ denote the sample standard deviation of $X'\widehat{\theta}$. Note that $\sup|S(\cdot)| \equiv \overline{\mathcal{B}} = 1$. Our tuning parameters are of the form,

$$h_n = c_h \cdot \widehat{\sigma}(X'\widehat{\theta}) \cdot n^{-\alpha_h}, \quad b_n = c_b \cdot \overline{\mathcal{B}} \cdot n^{-\alpha_b} = c_b \cdot n^{-\alpha_b}, \quad \kappa_1 = c_\kappa \cdot \overline{\mathcal{B}} = c_\kappa.$$

The bandwidth convergence rates α_h and α_b were set as described in Section 3.6. We used $\alpha_h = \frac{1}{8} - \frac{\epsilon'+\delta'}{2}$ and $\alpha_b = \frac{1}{4} + \Delta_b$, with $\epsilon' = \delta' = 10^{-5}$ and $\Delta_b = \frac{3}{4} \cdot 10^{-5}$. The convergence rate α_h is relatively slow. To compensate for the resulting oversmoothing, we choose a value of c_h smaller than some of the usual choices in the literature. In our experiments we set $c_h = 0.7$, which is about 1/3 below the "rule of thumb" value of 1.06 (see Silverman (1986, Equation 3.28)). Next, to concentrate on the effect of b_n, we set the regularization parameter κ_1 to $c_\kappa = 10^{-10}$ and we focused on the choice of c_b. To reduce the risk of over-rejection in finite samples when the inequalities hold but are binding with positive probability (DGP 2), we selected values of c_b such that b_n is nontrivially bounded away from zero. In our experiments we used $c_b \in \{0.5, 0.75, 1\}$. For each of these choices, we found that the finite-sample size and power properties resemble closely the asymptotic predictions of Theorem 1. Notably, even conservative values of c_b ($c_b = 1$ in our experiments) produce nontrivial power to detect violations that occur with probability as small as 1%. Violations that occur with larger probability in our DGPs are rejected with large probabilty even when $n = 100$.

As discussed in Section 3.6, we can use a kernel of order $M = 2D + 3 = 5$. We employed $K(\psi) = \Big(c_1 \cdot (S^2 - \psi^2)^2 + c_2 \cdot (S^2 - \psi^2)^4 + c_3 \cdot (S^2 - \psi^2)^6\Big) \cdot \mathbb{1}\{|\psi| \le S\}$, where c_1, c_2 and c_3 are chosen to satisfy[9] $\int_{-S}^{S} K(\psi)d\psi = 1$, $\int_{-S}^{S} \psi^2 K(\psi)d\psi = 0$ and $\int_{-S}^{S} \psi^4 K(\psi)d\psi = 0$. The support of the kernel is $[-S, S]$ and, as in the case of the Epanechnikov kernel (Pagan and Ullah (1999, Equation 2.61)), we chose $S = 1$. Let $(\xi)_{(\tau)}$ denote the τ^{th} sample quantile of ξ, and let $\overline{X}_i \equiv \frac{1}{8}\sum_{\ell=1}^{8} X_{\ell i}$. We used $\omega(X_i'\widehat{\theta}) = \mathbb{1}\Big\{(X'\widehat{\theta})_{0.005} \le X_i'\widehat{\theta} \le (X'\widehat{\theta})_{0.995}\Big\}$, and $\phi(X_i, t_i) = \mathbb{1}\Big\{\Big(\overline{X}\Big)_{0.005} \le \overline{X}_i \le \Big(\overline{X}\Big)_{0.995}, (t)_{0.005} \le t_i \le (t)_{0.995}\Big\}$ as our weight functions and testing range. We chose a wide testing range to detect violations to (34') that occur with small probability.

5.4 Results

The results of our experiments are summarized in Table 2 for a target significance level of 5%. Overall, our rejection frequencies are in line with the asymptotic predictions of Theorem 1 in terms of size and power. Our test has remarkable power when the inequalities are violated with moderate probability (20% or more in our experiments). Even for samples of size $n = 100$, our test rejected the null hypothesis with frequency at least 60% when the probability of a violation was 20%. The frequency of rejection was greater than 97% when the inequalities are violated with probability 50%. Our test also has nontrivial power when the probability of a violation is small. For example, when the inequalities are violated with probability 5%, our test rejects with a frequency greater than 14% when $n = 100$. This figure jumps to 85% for samples of size $n = 1,000$. Nontrivial power is present even when the inequalities are violated with probability 1%. In this case, our test rejects FOSD with probability at least 6% in a sample of size $n = 100$. This figure

[9]Due to the symmetry of $K(\cdot)$ around zero, it satisfies $\int_{-S}^{S} \psi^j K(\psi)d\psi = 0$ for all odd j. In particular, it also satisfies $\int_{-S}^{S} \psi^5 K(\psi)d\psi = 0$, so ours is technically a bias-reducing kernel of order 6.

jumps to 37% when $n = 1,000$ and 62% when $n = 2,000$. When $n = 4,000$, our test will reject with a frequency of at least 79% when the probability of a violation is just 1%. As the results show, the power properties of our test were only moderately sensitive to the choice of $c_b \in [0.5, 1]$.

Next, let us summarize the size properties of our test. The values we chose for c_b result in values of b_n that are nontrivially bounded away from zero, helping us ward off against the possibility of over-rejecting the null hypothesis when the inequalities are binding w.p.1 (DGP 2). With our choices of b_n, the size properties of our test are very much aligned with the results in Theorem 1 which predict that, asymptotically, our test should reject our conjecture in DGP1 with probability zero, and DGP2 with probability 5% (our target significance level). In fact, the test never rejected the null hypothesis for any sample size in the case of DGP1 (when the FOSD hold strictly w.p.1) and the probability of rejection in DGP2 approaches the target significance level of 5% as n increased. Overall, we find in our experiments that choosing values of b_n that are nontrivially bounded away from zero lead to finite-sample properties in line with our asymptotic predictions.

6 Concluding remarks

Many economic models produce testable implications in the form of functional inequalities. Even though such tests can be conducted fully nonparametrically, in many instances, the data may consist of a rich collection of conditioning variables, and researchers may want to aggregate them into a lower-dimensional collection of conditioning functions or indices. Motivated by this problem, we introduced a test for functional inequalities conditional on estimated functions. We focused on the case where the conditioning functions have a parametric functional form and are indexed by a finite-dimensional parameter vector which is estimated in a first step. Our proposed tests are based on one-sided CvM statistics which adapt to the properties of the contact sets (the set of values of conditioning variables where the inequalities are binding) and have asymptotically pivotal properties. Their construction can be straightforwardly applied to single or multiple functional inequalities. Furthermore, the conditions we studied for the conditioning functions include a wide variety of $\sqrt{n}$–consitent extremum estimators as special cases. In Monte Carlo experiments, our test displayed good power properties, capable of detecting violations to the inequalities that occur with very low probability.

Table 2: Monte Carlo results. Rejection frequencies with 2,000 simulations in each case.

	$c_b = 0.50$					
Sample size	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
$n = 100$	0.000	0.026	0.973	0.637	0.171	0.076
$n = 250$	0.000	0.032	0.992	0.947	0.397	0.171
$n = 500$	0.000	0.039	0.998	0.986	0.662	0.269
$n = 1,000$	0.000	0.045	0.998	0.995	0.904	0.459
$n = 2,000$	0.000	0.049	0.999	0.995	0.976	0.718
$n = 3,000$	0.000	0.052	0.999	0.998	0.989	0.827
$n = 4,000$	0.000	0.051	0.999	0.998	0.992	0.887

	$c_b = 0.75$					
Sample size	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
$n = 100$	0.000	0.022	0.973	0.610	0.150	0.066
$n = 250$	0.000	0.026	0.992	0.937	0.347	0.137
$n = 500$	0.000	0.030	0.994	0.986	0.618	0.224
$n = 1,000$	0.000	0.032	0.996	0.994	0.869	0.400
$n = 2,000$	0.000	0.034	0.998	0.996	0.967	0.646
$n = 3,000$	0.000	0.038	0.999	0.997	0.988	0.762
$n = 4,000$	0.000	0.039	0.999	0.998	0.992	0.829

	$c_b = 1.00$					
Sample size	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
$n = 100$	0.000	0.021	0.972	0.596	0.142	0.060
$n = 250$	0.000	0.023	0.990	0.932	0.326	0.130
$n = 500$	0.000	0.024	0.992	0.984	0.599	0.202
$n = 1,000$	0.000	0.025	0.994	0.992	0.851	0.372
$n = 2,000$	0.000	0.030	0.994	0.995	0.963	0.618
$n = 3,000$	0.000	0.032	0.998	0.996	0.986	0.722
$n = 4,000$	0.000	0.034	0.999	0.998	0.991	0.793

References

Andrews, D. (1994). Empirical process methods in econometrics. In R. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. North-Holland.

Andrews, D. W. K. and X. Shi (2013). Inference for parameters defined by conditional moment inequalities. *Econometrica 81*(2), 609–666.

Aradillas-López, A. and A. Gandhi (2016). Estimation of games with ordered actions: An application to chain-store entry. *Quantitative Economics 7*(1), 727–780.

Aradillas-López, A., A. Gandhi, and D. Quint (2016). A simple test for moment inequality models with an application to english auctions. *Journal of Econometrics 194*, 96–115.

Armstrong, T. B. (2014). Weighted ks statistics for inference on conditional moment inequalities. *Journal of Econometrics 181*(2), 92–116.

Armstrong, T. B. (2015). Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics 186*(1), 51–65.

Armstrong, T. B. (2018). On the choice of test statistic for conditional moment inequalities. *Journal of Econometrics 203*(1), 241–255.

Armstrong, T. B. and H. P. Chan (2016). Multiscale adaptive inference on conditional moment inequalities. *Journal of Econometrics 194*(1), 24–43.

Chetverikov, D. (2017). Adaptive test of conditional moment inequalities. *Econometric Theory 34*(1), 186–227.

Donkers, B. and M. Schafgans (2008). Specification and estimation of semiparametric multiple-index models. *Econometric Theory 24*, 1584–1606.

Han, A. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics 58*, 71–120.

Ichimura, H. and L.-F. Lee (1991). Semiparametric least squares estimation of multiple index models: Single equation estimation. In W. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and semiparametric methods in econometrics and statistics*, pp. 3–51. Cambridge University Press.

Lee, S., K. Song, and Y.-J. Whang (2013). Testing functional inequalities. *Journal of Econometrics 172*(1), 14–32.

Lee, S., K. Song, and Y.-J. Whang (2018). Testing for a general class of functional inequalities. *Econometric Theory* (34), 1018–1064.

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

Milgrom, P. and R. Weber (1982). A theory of auctions and competitive bidding. *Econometrica 50*(5), 1089–1122.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*, 780–799.

Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Picone, G. and J. Butler (2000). Semiparametric estimation of multiple equation models. *Econometric Theory 16*, 551–575.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics.

Powell, J., J. Stock, and T. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*, 1403–1430.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley. New York, NY.

Sherman, R. (1994). Maximal inequalities for degenerate u-processes with applications to optimization estimators. *Annals of Statistics 22*, 439–459.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC.