Dimension reduction and testing of functional inequalities conditional on estimated functions

Andres Aradillas-Lopez*

June 9, 2025

Abstract

A number of economic models produce testable implications in the form of inequalities involving conditional functionals of the distribution of an outcome variable Y conditional on X, a vector of observable covariates. In applications where X includes a large collection of predictors, researchers may wish to pursue dimension reduction and aggregate X into a lower-dimensional, parameterized function $g(X,\theta)$, indexed by a finite-dimensional parameter θ, and proceed to test the functional inequalities conditional on $g(X,\widehat{\theta})$ instead of X, where $\widehat{\theta}$ is a first-step estimator. Motivated by this, we introduce tests for functional inequalities conditional on estimated, aggregate functions of X. Our 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. Our proposed test-statistics 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. In Monte Carlo experiments, our procedure displays good power properties, capable of detecting violations to the inequalities that occur with very small probability.
Keywords: Functional inequalities, nonparametric tests, conditional moments, dimension reduction.
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 nonparametrically. In a number of applications, X may contain a rich number of continuous variables,

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

making nonparametric estimators susceptible to the curse of dimensionality. In such cases, researchers may pursue dimension reduction and aggregate X into a lower-dimensional index. We focus on the case where this is done through parametric functions $g(X,\theta)$ indexed by a finite-dimensional parameter θ, and proceed to test the functional inequalities conditional on $g(X,\widehat{\theta})$ instead of X, where $\widehat{\theta}$ is a first-step estimator. The choice of the conditioning functions can be justified by the *testable* assumption that $Y|X \sim Y|g(X,\theta^*)$, where θ^* is the probability-limit of $\widehat{\theta}$. Alternatively, conditioning on $g(X,\widehat{\theta})$ instead of X can be justified by iterated-expectation arguments when the inequalities are linear in the conditional functionals involved. Motivated by the general problem described above, we will introduce tests for functional inequalities conditional on estimated functions, and characterize their asymptotic properties. To our knowledge, this appears to be the first paper explicitly devoted to the study functional inequalities and dimension reduction where tests are conditioned on estimated functions of the conditioning predictors X.

Our approach is a generalized version 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 estimated in a first step (i.e, a "generated regressor"). Conditioning directly on X, without relying on estimated conditioning functions, is a special case of our tests. Therefore, our methodology also contributes to the existing toolbox for the usual problem of testing conditional functional inequalities (without estimated conditioning functions).

The paper proceeds as follows. Section 2 provides the setup and the type of functional inequalities we study. Based on this setup, Section 3 describes the population statistic that our test will focus on. Section 4 describes our proposed econometric test and studies its asymptotic properties. Section 5 includes results from Monte Carlo experiments. Section 6 concludes. Appendix A describes the proofs of our main econometric results, along with extensions and other details mentioned throughout the paper. Appendix B describes examples of estimators that satisfy a key condition in our setup. An accompanying Econometric Supplement includes all the step-by-step derivations and the full details of our proofs. Appendix B can be downloaded at `https://aaradill.github.io/condit-ineq-functions-appendix-B.pdf` . The Econometric Supplement can be downloaded at `https://aaradill.github.io/condit-ineq-functions-supplement.pdf`

2 Setup and description of the inequalities to be tested

2.1 Some preliminaries

Our analysis includes a triple of *observable* random variables (Y, X, Z). Our inequalities will involve functionals of the distribution of $Y|X$, while Z will denote covariates used in the construction of a first-step estimator $\widehat{\theta}$ (to be described below). Z can include elements from (Y, X), with $Z = (Y, X)$ as a special case. We observe a sample $(Y_i, X_i, Z_i)_{i=1}^n$ of independent observations of a distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is a space of distributions. Our goal will be to describe conditions that yield asymptotic properties that hold uniformly over $\mathcal{F}$. Let $\mathcal{S}_\xi$ denote the support of a r.v ξ. We will group $V \equiv (Y, X) \cup Z$, and we will maintain that $\mathcal{S}_V$ is the same for all $F \in \mathcal{F}$. We will indicate functionals of the distribution F by including the subscript F except when this produces cumbersome notation. In any case, the exposition and definitions will clarify which objects are functionals of F. We will maintain that the space of distributions $\mathcal{F}$ satisfies the following compactness feature. For any measurable set S, $\sup\limits_{F \in \mathcal{F}} P_F(S) = p \Longrightarrow \exists\, F^* \in \mathcal{F} : P_{F^*}(S) = p$. 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, $\sup\limits_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) \longrightarrow 0 \quad \forall\, c > 0$.

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

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 $n \geq n_0$. We say $\sup\limits_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) = O(n^\gamma)$ for a given $c > 0$, if $\exists\, n_0$ and $\Delta > 0$ such that $n^{-\gamma} \sup\limits_{F \in \mathcal{F}} P_F\left(n^{-\lambda}\|\xi_n\| > c\right) < \Delta \quad \forall\, n \geq n_0$.

2.2 Components of the models studied here

2.2.1 Functional inequalities

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 $F \in \mathcal{F}$, define

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

Models without index parameters t will be a a special case of our general setup. Next, we have a known, real-valued transformation $\mathcal{B} : \mathbb{R}^P \longrightarrow \mathbb{R}$ of the functionals $\Gamma_F(\cdot)$, and a model that predicts,

$$\mathcal{B}(\Gamma_F(x,t)) \leq 0 \;\; F\text{–a.e } x \in \mathcal{S}_X, \; \forall \; t \in \mathcal{T}. \tag{2}$$

Equation (2) describes a functional inequality produced by some underlying economic or statistical model, and our ultimate goal will be to construct a test for it. We present examples next.

2.2.2 Examples

Example 1: First order stochastic dominance

Suppose Y_1, Y_2 are two scalar random variables, and consider the first-order stochastic dominance restriction $F_{Y_1|X}(\cdot|X) \gtrsim_{FOSD} F_{Y_2|X}(\cdot|X)$ 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, $\forall\; t$, a special case of (2) where $\mathcal{B}(\Gamma) = \Gamma$.

Example 2: Second order stochastic dominance

Consider a second-order stochastic dominance restriction $F_{Y_1|X}(\cdot|X) \gtrsim_{SOSD} F_{Y_2|X}(\cdot|X)$ F–a.s. That is, $\int_{-\infty}^{t} F_{Y_1|X}(v|x)dv \leq \int_{-\infty}^{t} F_{Y_2|X}(v|x)dv$ F–a.e $x \in \mathcal{S}_X$, $\forall\; t$. Note that $\int_{-\infty}^{t} \mathbb{1}\{\xi \leq v\}\, dv = \max\{t - \xi,\, 0\}$. Thus, $E_F[\max\{t - Y_\ell,\, 0\}|X = x] = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{t} \mathbb{1}\{y \leq v\}\, dv\right) f_{Y_\ell|X}(y|x)dy = \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} \mathbb{1}\{y \leq v\} f_{Y_\ell|X}(y|x)dy\right) dv$ $= \int_{-\infty}^{t} F_{Y_\ell|X}(v|x)dv$ for $\ell = 1,2$. 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, $\forall\; t$. This is a a special case of (2) where $\mathcal{B}(\Gamma) = \Gamma$.

Example 3: Covariance inequalities

There exist economic models that yield restrictions of the form $Cov(\eta_1(Y,t), \eta_2(Y,t)\big|X) \leq 0$ F–a.s, $\forall\; t \in \mathcal{T}$, where η_1 and η_2 are parametric functions. As shown in Aradillas-López and Gandhi (2016), restrictions of this form arise in incomplete information games with ordinal action spaces when we conjecture that some parametric "aggregate index" $\varphi(Y_{-p})$ (e.g, $\varphi(Y_{-p}) = \sum_{q \neq p} Y_q$) of the actions of player p's opponents is a *strategic substitute* for Y_p. Under payoff-shape restrictions described by the authors, we must have $Cov(\mathbb{1}\{Y_p \geq t\}, \varphi(Y_{-p})|X)$ F–a.s, $\forall\; t \in \mathcal{T}$. Here, t denotes a generic element in the action space of player p, and $\mathcal{T}$ is p's 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, $E_F[S_1(Y,t)|X] - E_F[S_2(Y,t)|X] \cdot E_F[S_3(Y,t)|X] \leq 0$ F–a.s, $\forall\; t \in \mathcal{T}$. This is a special case of (2), with $\mathcal{B}(\Gamma_1, \Gamma_2, \Gamma_3) = \Gamma_1 - \Gamma_2 \cdot \Gamma_3$.

Example 4: Affiliation

Let $Y \equiv (Y_1, \ldots, Y_L) \in \mathbb{R}^L$. 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 $G_F(u, \delta|X) = P_F\left(u - \delta \leq Y \leq u + \delta \big| X\right)$. 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 $S_1(Y,t) \equiv \mathbb{1}\{t_1 - t_3 \leq Y \leq t_1 + t_3\}$, $S_2(Y,t) \equiv \mathbb{1}\{t_2 - t_3 \leq Y \leq t_2 + t_3\}$, $S_3(Y,t) \equiv \mathbb{1}\{t_1 \vee t_2 - t_3 \leq Y \leq t_1 \vee t_2 + t_3\}$, and $S_4(Y,t) \equiv \mathbb{1}\{t_1 \wedge t_2 - t_3 \leq Y \leq t_1 \wedge t_2 + t_3\}$. From here, the functional inequality in (2) 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 $E[S(Y)|X] \leq 0$ $F-$a.s (without an index parameter t) as special cases. These models are a special case of (2), with $\mathcal{B}(\Gamma) = \Gamma$.

2.3 Conditioning functions

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

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

Our setup is motivated by the case where $g(X,\theta)$ is lower-dimensional than X, and the researcher aggregates X through this parametric index and proceeds to test the functional inequalities in (2) conditional on $g(X,\widehat{\theta})$ (instead of the full vector X), where $\widehat{\theta}$ is estimated in a first-step. We will assume that the estimator $\widehat{\theta}$ is obtained from the same sample[2] that produced the observations for (Y,X). In our setup, we can express $\widehat{\theta}$ as a statistic $\widehat{\theta}(Z_1,\ldots,Z_n)$, where $(Z_i)_{i=1}^n$ denote the observable covariates in our sample used in the construction of $\widehat{\theta}$. As we mentioned previously, Z_i can have elements in common with (Y_i,X_i), with $Z_i = (Y_i,X_i)$ as a special case. We will take the choice of the estimator $\widehat{\theta}$ as a given, pre-specified modeling choice made by the researcher, and we leave the problem of how to estimate $\widehat{\theta}$ "optimally" for future work. We will maintain that the estimator $\widehat{\theta}$ used by the researcher satisfies the following asymptotic properties.

Assumption 1 *The probability-limit of $\widehat{\theta}$ is denoted as θ^*_F. We let $\Theta \subseteq \mathbb{R}^k$ denote the parameter space, assumed to be a bounded and convex subset of $\mathbb{R}^k$. The estimator $\widehat{\theta}$ satisfies the linear representation,*

$$\widehat{\theta} = \theta^*_F + \frac{1}{n}\sum_{i=1}^n \psi^\theta_F(Z_i) + \varepsilon^\theta_n, \tag{3}$$

[1] The conditioning functions can also potentially depend on the index parameter t. Our results and derivations will illustrate how this extension could be handled.

[2] Our results can be readily extended to the case where $\widehat{\theta}$ is obtained from an auxiliary sample.

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

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

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

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

θ_F^* will depend on the assumptions of the econometric model used to estimate $\widehat{\theta}$. For example, suppose the researcher maintains the exclusion restriction $Y|X \sim Y|g(X,\theta_0)$ (a testable assumption). Existing semiparametric methods (see Powell, Stock, and Stoker (1989), Ichimura and Lee (1991), Ichimura (1993), Picone and Butler (2000), Donkers and Schafgans (2008)) can be used to construct an estimator $\widehat{\theta}$ for θ_0 (modulo scale and location normalizations). In this case, $\theta_F^* = \theta_0$.

Examples of estimators that satisfy Assumption 1

Appendix B presents various examples of estimators that satisfy the restrictions in Assumption 1. The examples include OLS, GMM, density-weighted average derivatives and a semiparametric multiple-index model. As shown there, these estimators can satisfy Assumption 1 under commonly used regularity and integrability conditions. Appendix B can be downloaded at `https://aaradill.github.io/condit-ineq-functions-appendix-B.pdf`

2.4 Conditioning on $g(X,\theta_F^*)$

Our setup is one where the researcher chooses to aggregate X through $g(X,\widehat{\theta})$, and proceeds to test the functional inequalities described above, *conditional on* $g(X,\widehat{\theta})$. For a given θ, let us begin by defining the counterparts to the functionals in (1) when we condition on $g(X,\theta)$ instead of X. For a given $x \in \mathcal{S}_X$, $t \in \mathcal{T}$, and $\theta \in \Theta$, we will denote

$$\Gamma_{p,F}(x,t,\theta) = E_F\left[S_p(Y,t)\;\middle|\; g(X,\theta) = g(x,\theta)\right], \quad \text{and} \quad \underbrace{\Gamma_F(x,t,\theta)}_{P\times 1} \equiv \left(\Gamma_{1,F}(x,t,\theta),\ldots,\Gamma_{P,F}(x,t,\theta)\right)'. \tag{4}$$

By Assumption 1, if we condition on $g(X,\widehat{\theta})$, then asymptotically the relevant functionals will be $\Gamma_F(x,t,\theta_F^*)$, and the functional inequalities we will test are,

$$\mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \;\; F\text{–a.e } x \in \mathcal{S}_X,\; \forall\; t \in \mathcal{T}. \tag{5}$$

Remark 1 ***(Conditioning directly on X)*** *Our setup and results will always include $g(X,\theta) = X$ (no conditioning functions, no estimator $\widehat{\theta}$) as a special case. In this instance, all the results bellow will follow with $\psi_F(Z_i) \equiv 0$ and $\varepsilon_n^\theta \equiv 0$.*

When aggregating X into the index $g(X,\theta_F^*)$, two questions arise. First, under what conditions will (5) follow from (2)? Second, and more generally, what is the difference between testing (5) and (2)? We discuss these questions next.

2.4.1 Testing (5) vs. testing (2)

Before proceeding, there are two key questions to address. When do the inequalities in (5) follow from those in (2), and what is the difference between testing (5) and (2)? To address these questions, we can focus on two general cases.

(A) The case where $Y|X \sim Y|g(X,\theta_F^*)$

Suppose we maintain that the conditional distribution of $Y|X$ depends on X only through the parametric function $g(X,\theta_F^*)$ (i.e, $Y|X \sim Y|g(X,\theta_F^*)$). In this case, all the functionals of $Y|X$ can be written as functionals of $Y|g(X,\theta_F^*)$, and therefore (2) can be re-written as (5). In this case, *testing* (5) *is equivalent to testing* (2). Note that the exclusion restriction $Y|X \sim Y|g(X,\theta_F^*)$ is *testable* (see, e.g, Hardle and Mammen (1993), Fan (1995), Fan and Li (1996), Zheng (1996), and Fan (1998)), and an estimator $\widehat{\theta}$ can be constructed through existing semiparametric methods (see Powell, Stock, and Stoker (1989), Ichimura and Lee (1991), Ichimura (1993), Picone and Butler (2000), Donkers and Schafgans (2008)). Appendix B provides an example of such an estimator and it shows conditions under which it satisfies the restrictions in Assumption 1.

(B) The case where the transformation $\mathcal{B}$ is linear

Suppose $\mathcal{B}(\Gamma) = \alpha'\Gamma$, where α is a known vector of constants. Then, (2) is simply "$\alpha' E_F[S(Y,t)|X] \leq 0$ a.e X, $\forall\ t \in \mathcal{T}$". Let $g(X)$ be a measurable function of X. By iterated expectations,

$$\alpha' E_F[S(Y,t)|g(X)] = E_F[\alpha' E_F[S(Y,t)|g(X),X]|g(X)] = E_F[\underbrace{\alpha' E_F[S(Y,t)|X]}_{\leq 0 \text{ a.e } X,\ \forall\ t \in \mathcal{T}}|g(X)] \leq 0 \quad \text{a.e } X,\ \forall\ t \in \mathcal{T}$$

As long as $g(X,\theta_F^*)$ is a measurable function of X, the functional inequality in (2) implies that (5) is valid. This does not require that $Y|X \sim Y|g(X,\theta_F^*)$. Without this exclusion restriction, testing (5) is no longer equivalent to testing (2), but rejecting (5) would immediately reject (2).

Revisiting the examples in Section 2.2.2

Examples 1, 2 and 5 (FOSD, SOSD and CMI) are instances where $\mathcal{B}$ is linear. Thus, the inequality (5) will be valid (modulo measurability of $g(X,\theta_F^*)$) without having to assume that $Y|X \sim Y|g(X,\theta_F^*)$. Examples 3 and 4 (covariance restrictions and affiliation) involve nonlibear transformations $\mathcal{B}$. In these cases, (5) will be valid if we assume that $Y|X \sim Y|g(X,\theta_F^*)$ (a testable assumption).

3 A population test-statistic for (5)

Our goal moving forward is to construct a test for the functional inequality (5) over a prespecified testing range of values of x. We describe next the population test-statistics we will focus on.

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 a prespecified, compact subset of $\mathbb{R}^D$. Let $\mathcal{X}_F^* = \left\{x \in \mathcal{S}_X : x \in \mathcal{X} \ \text{ and } \ g(x,\theta_F^*) \in \mathcal{G}\right\}$. The set $\mathcal{X}_F^*$ constitutes our *target testing range* for the functional inequalities (5). $\mathcal{X}_F^*$ will be assumed to satisfy conditions that yield uniform asymptotic properties for our nonparametric estimators.

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

$$\begin{aligned} Q_{p,F}(x,t,\theta_F^*) &\equiv \Gamma_{p,F}(x,t,\theta_F^*) \cdot \omega_p(g(x,\theta_F^*)), \\ Q_F(x,t,\theta_F^*) &\equiv (Q_{1,F}(x,t,\theta_F^*),\ldots,Q_{P,F}(x,t,\theta_F^*))', \\ \omega(g(x,\theta_F^*)) &\equiv (\omega_1(g(x,\theta_F^*)),\ldots,\omega_P(g(x,\theta_F^*)))'. \end{aligned} \tag{6}$$

3.3 A separability assumption for $\mathcal{B}$

The rest of the paper will focus on cases where the transformation $\mathcal{B}$ satisfies a separability condition. Specifically, we assume that $\mathcal{B}$ is such that we can construct a collection of nonnegative "weights" $(\omega_p)_{p=1}^P$ such that,

$$\begin{aligned} &\mathcal{B}(\Gamma_1 \cdot \omega_1, \Gamma_2 \cdot \omega_2, \ldots, \Gamma_P \cdot \omega_P) = \mathcal{B}(\Gamma_1, \Gamma_2, \ldots, \Gamma_P) \cdot \mathcal{H}(\omega_1, \omega_2, \ldots, \omega_P), \\ &\text{where} \begin{cases} \mathcal{H}(\omega_1, \omega_2 \ldots, \omega_P) \geq 0 \\ \mathcal{H}(\omega_1, \omega_2, \ldots, \omega_P) > 0 \iff \omega_p \neq 0 \ \forall \ p. \end{cases} \end{aligned} \tag{7}$$

Note that we do not assume that (7) holds for any collection of nonnegative weights, but that there exists a particular collection of weights for which (7) is satisfied. From (7) and the properties of the weight functions ω described in (6), it follows that,

$$\begin{aligned} &\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) = \mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right) \cdot \mathcal{H}\left(\omega(g(x,\theta_F^*))\right), \\ &\text{where} \begin{cases} \mathcal{H}(\omega(g(x,\theta_F^*)) \geq 0 \ \ \forall \ x, \\ \mathcal{H}(\omega(g(x,\theta_F^*)) > 0 \iff g(x,\theta_F^*) \in \mathcal{G}. \end{cases} \end{aligned} \tag{8}$$

The condition in (6) is trivially satisfied by any linear transformation $\mathcal{B}$, but it can also be satisfied by nonlinear transformations. In particular, it holds all the examples in Section 2.2.2.

Examples 1, 2 and 5 (FOSD, SOSD and CMI): Here, we have $\mathcal{B}(\Gamma) = \Gamma$ and the condition in (7) 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 have $P = 3$ and $\mathcal{B}(\Gamma_1, \Gamma_2, \Gamma_3) = \Gamma_1 - \Gamma_2\Gamma_3$. Take any $\omega_1 \geq 0$ and let $\omega_2 = \omega_3 = \omega_1^{1/2}$. Then, $\mathcal{B}(\Gamma_1 \cdot \omega_1, \Gamma_2 \cdot \omega_2, \Gamma_3 \cdot \omega_3) = \Gamma_1\omega_1 - \Gamma_2\omega_1^{1/2}\Gamma_3\omega_1^{1/2} = (\Gamma_1 - \Gamma_2 \cdot \Gamma_3) \cdot \omega_1 \equiv \mathcal{B}(\Gamma_1, \Gamma_2, \Gamma_3) \cdot \mathcal{H}(\omega)$. Thus, the condition in (7) 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 $\omega_1 \geq 0$ and set $\omega_2 = \omega_3 = \omega_4 = \omega_1$. Then, $\mathcal{B}(\Gamma_1 \cdot \omega_1, \Gamma_2 \cdot \omega_2, \Gamma_3 \cdot \omega_3, \Gamma_4 \cdot \omega_4) = \Gamma_1\omega_1\Gamma_2\omega_1 - \Gamma_3\omega_1\Gamma_4\omega_1 = (\Gamma_1\Gamma_2 - \Gamma_3\Gamma_4) \cdot \omega_1^2 \equiv \mathcal{B}(\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4) \cdot \mathcal{H}(\omega)$, and condition in (7) holds with $\mathcal{H}(\omega) \equiv \omega_1^2$. ■

3.4 A population statistic

We will construct a test for the inequality

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

From (8), testing (9) is equivalent to testing the restriction,

$$\mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right) \leq 0 \ \text{ for } F\text{-a.e } x \in \mathcal{X}_F^* \text{ and } \forall \ t \in \mathcal{T} \tag{10}$$

The goal if this paper is to test (10) by constructing a test for (9). Let $(a)_+ \equiv a \vee 0$. Note from (8),

$$\left(\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right)\right)_+ = \left(\mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right)\right)_+ \mathcal{H}(\omega(g(x,\theta_F^*)) \tag{11}$$

We will focus on one-sided Cramér–von Mises (CvM) population statistics that integrate out x in (11) with respect to dF_X (the distribution of X), and integrate t with respect to a prespecified weight function $d\mathcal{W} \geq 0$ which satisfies $d\mathcal{W}(t) > 0 \Leftrightarrow t \in \mathcal{T}$. For simplicity, we normalize

$\int_{t\in\mathcal{T}} d\mathcal{W}(t) = 1$. Let $\phi \geq 0$ be a weight function satisfying $\phi(x) > 0 \Leftrightarrow x \in \mathcal{X}$. For a given $t \in \mathcal{T}$, let

$$T_{0,F}(t) \equiv E_F\Big[\left(\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)\right)_+ \phi(X)\Big] = \int_x \left(\mathcal{B}\left(\Gamma_F(x,t,\theta_F^*)\right)\right)_+ \mathcal{H}\left(\omega(g(x,\theta_F^*))\right)\phi(x)dF_X(x).$$

Where the last equality follows from (11). We focus on the following population statistic,

$$T_{2,F} \equiv \int_t T_{0,F}(t)d\mathcal{W}(t) = \int_t \Big(E_F\Big[\left(\mathcal{B}\left(\Gamma_F(X,t,\theta_F^*)\right)\right)_+ \mathcal{H}\left(\omega(g(X,\theta_F^*))\right)\phi(X)\Big]\Big)d\mathcal{W}(t) \tag{12}$$

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

Remark 2 *In models without index parameters t, we simply have*

$$T_{2,F} = E_F\Big[\left(\mathcal{B}\left(Q_F(X,\theta_F^*)\right)\right)_+ \phi(X)\Big] = E_F\Big[\left(\mathcal{B}\left(\Gamma_F(X,\theta_F^*)\right)\right)_+ \mathcal{H}\left(\omega(g(X,\theta_F^*))\right)\phi(X)\Big].$$

Again, the last equality follows from (11)*. Models without index parameters t will be a special case of our framework throughout.* ■

4 Proposed econometric test and its asymptotic properties

Our proposal is to construct a test-statistic based on $T_{2,F}$. First, we present some preliminary definitions and estimators.

4.1 Notational definitions of some key functionals

We will focus on the case where the conditioning functions $g(X,\theta_F^*)$ are jointly continuously distributed[3], 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 $R_{p,F}(x,t,\theta_F^*) \equiv \Gamma_{p,F}(x,t,\theta_F^*) \cdot \omega_p(g(x,\theta_F^*)) \cdot f_g(g(x,\theta_F^*))$. Note that $Q_{p,F}(x,t,\theta_F^*) = \frac{R_{p,F}(x,t,\theta_F^*)}{f_g(g(x,\theta_F^*))}$.

4.2 Estimators of the functionals involved

We will rely on kernel-based estimators. Let $K : \mathbb{R}^D \rightarrow \mathbb{R}$ be a kernel function and let $h_n \longrightarrow 0$ be a bandwidth sequence (their properties will be described below). For any pair (x_1, x_2) and θ, denote $\Delta g(x_1,x_2,\theta) \equiv g(x_1,\theta) - g(x_2,\theta)$. For a given (x,θ), let $\widehat{f_g}(g(x,\theta)) \equiv \frac{1}{n\cdot h_n^D}\sum_{i=1}^n K\left(\frac{\Delta g(X_i,x,\theta)}{h_n}\right)$ be our estimator of $\widehat{f_g}(g(x,\widehat{\theta}))$. 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

[3]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.

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

$$\widehat{Q}_p(x,t,\widehat{\theta}) \equiv \frac{\widehat{R}_p(x,t,\widehat{\theta})}{\widehat{f}_g(g(x,\widehat{\theta}))}, \quad \text{with} \quad \widehat{Q}(x,t,\widehat{\theta}) \equiv \left(\widehat{Q}_1(x,t,\widehat{\theta}),\ldots,\widehat{Q}_P(x,t,\widehat{\theta})\right)' \tag{13}$$

4.3 Our estimator for $T_{2,F}$

Letting $\widehat{Q}(x,t,\widehat{\theta})$ be as described in (13), our estimator for $T_{2,F}$ is,

$$\begin{aligned} \widehat{T}_2 &\equiv \int_t \left(\frac{1}{n}\sum_{i=1}^n \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\} \phi(X_i) \right) d\mathcal{W}(t) \equiv \int_t \widehat{T}_0(t) d\mathcal{W}(t), \\ \text{where} \quad \widehat{T}_0(t) &\equiv \frac{1}{n}\sum_{i=1}^n \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\} \phi(X_i), \end{aligned} \tag{14}$$

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

4.4 Asymptotic properties of $\widehat{T}_2$

In this section we will describe assumptions that will yield an asymptotic linear representation result for $\widehat{T}_2$. These assumptions will be technical in nature, but we will summarize intuitively how each one contributes to our asymptotic results. Let $\mathcal{S}_{g,F}$ denote the support of $g(X,\theta_F^*)$ for a given F. That is,

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

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

$$\mathcal{S}_{g,F}^{\mathcal{X}} = \left\{ g \equiv (g_1,\ldots,g_D) \in \mathbb{R}^D : g = g(x,\theta_F^*) \text{ for some } x \in \mathcal{X} \right\}. \tag{16}$$

We will maintain 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 restriction will involve smoothness of the conditioning functions g, and of some key functionals in our model. These conditions are similar to smoothness restrictions assumed in general 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 $F \in \mathcal{F}$,*

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

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 (15) and (16). Then, $\mathcal{S}_{g,F}^{\mathcal{X}} \subset int\left(\mathcal{S}_{g,F}\right)$ for each $F \in \mathcal{F}$. Furthermore, there exists $c > 0$ such that if, we define

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

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

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

There exists an integer M such that, for each ℓ, d, and every $1 \leq m \leq M+1$ and $(j_d)_{d=1}^D$ such that $\sum_{d=1}^D j_d = m$, both $\frac{\partial^m f_g(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}$ and $\frac{\partial^m \Omega_{f_g}^{d,\ell}(g_1, \ldots, g_D)}{\partial g_1^{j_1} \cdots \partial g_D^{j_D}}$ are well-defined for any $(g_1, \ldots, g_D) \in \overline{\mathcal{S}}_{g,F}^{\mathcal{X}}$, $F \in \mathcal{F}$, and

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

for some $\overline{C}_1 < \infty$. Next let,

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

(note that $\Omega_{R_p,0}\left(g(x, \theta_F^*), t\right) = \Gamma_{p,F}(x, t, \theta_F^*)$). 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 $\sum_{d=1}^{D} j_d = m$,

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

For every $1 \leq m \leq M+1$ and $(j_d)_{d=1}^{D}$ such that $\sum_{d=1}^{D} j_d = m$, the weight function $\omega_p(\cdot)$ satisfies, $\sup_{(g_1,\ldots,g_D)\in\mathbb{R}^D} \left| \frac{\partial^m \omega_p(g_1,\ldots,g_D)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}} \right| \leq \overline{C}_\omega < \infty$ *and* $\omega_p(\cdot) \leq \overline{\omega} < \infty$. ■

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, uniformly over our testing range. Our next set of assumptions will lead to *manageability* properties (see Pollard (1990, Definition 7.9) or Andrews (1994, Assumption A, Theorem 1)) for some empirical processes that will be relevant in our setting. We will focus on *Euclidean* classes of functions, whose manageability properties have been 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 $i \neq j$. Let $\mathscr{G}$ be a class of functions on $\mathbb{R}$. We say that G is an* envelope *for $\mathscr{G}$ if $\sup_{\mathscr{G}} |g(\cdot)| \leq G(\cdot)$. Let μ be a measure on $\mathcal{S}_Z^k$ and denote $\mu h \equiv \int h(z_1, \ldots, z_k) d\mu(z_1, \ldots, z_k)$. The class $\mathscr{G}$ is Euclidean (A, V) for the envelope G if, for any measure μ s.t $\mu G^2 < \infty$, we have $D(x, d_\mu, \mathscr{G}) \leq A x^{-V}$ $\forall\ 0 < x \leq 1$, where $d_\mu(g_1, g_2) = (\mu |g_1 - g_2|^2 / \mu G^2)^{1/2}$ $\forall\ g_1, g_2 \in \mathscr{G}$. The constants (A, 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 many examples found in econometric models. A partial list includes 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 $|g(x,t) -$

$g(x,t')| \leq \phi(x) \cdot \|t-t'\|^{\alpha}$ $\forall$ $x \in \mathscr{X}$ 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 $x \rightarrow \lambda(\alpha' x + \beta)$, with $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ 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)) 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)) The Type I, II and III classes of functions described in Andrews (1994) are 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* $F \in \mathcal{F}$. ■

The conditions in Assumption 3 will lead to manageability of various empirical processes relevant to our problem. Next we describe restrictions for our kernel functions and bandwidths.

Assumption 4 ***(Kernels and bandwidths)*** *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*, symmetric around zero and with support* $[-S,S]$*. That is,* $\kappa(v) = \kappa(-v)\ \forall\ v$, $\kappa(v) = 0\ \forall\ v \notin (-S,S)$, $\int_{-S}^{S}\kappa(v)dv = 1$, $\int_{-S}^{S} v^j\kappa(v)dv = 0$ *for* $j = 1,\ldots,M-1$, *and* $\int_{-S}^{S}|v|^M\kappa(v)dv < \infty$. $\kappa(\cdot)$ *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* $|\kappa(\cdot)| \leq \overline{\kappa}$, $\left|\kappa^{(1)}(\cdot)\right| \leq \overline{\kappa}$ *and* $\left|\kappa^{(2)}(\cdot)\right| \leq \overline{\kappa}$ *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* $\overline{\kappa}$,

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

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

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

(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* $\widehat{T}_2$ *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* $\delta_0 > 0$.

Existing results for Euclidean classes can be invoked to verify the restrictions described above. For instance, if $g(x,\theta) = x'\theta$, the Euclidean property in part (ii) of Assumption 4 follows from Nolan and Pollard (1987, Lemma 22). Next, we impose smoothness restrictions on the transformation $\mathcal{B}$.

Assumption 5 ***(Smoothness II)*** *We impose the following restrictions on the transformation* $\mathcal{B}$.

(i) $\exists\ M_1 > 0, M_2 > 0$ *such that,* $\forall\ Q, Q' \in \mathbb{R}^P$, $\|Q - Q'\| \le M_2 \Longrightarrow \|\mathcal{B}(Q) - \mathcal{B}(Q')\| \le M_1 \cdot \|Q - Q'\|$.

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

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

(iii) *Let* $\overline{Q} \equiv \overline{\Gamma} \cdot \overline{\omega}$, *where* $\overline{\Gamma}$ *and* $\overline{\omega}$ *are the constants described in Assumption 2.* $\exists\ C_Q > 0, \overline{H}_Q > 0$ *such that,* $\|Q\| \le \overline{Q} + C_Q \Longrightarrow \|\nabla_Q \mathcal{B}(Q)\| \le \overline{H}_Q$, *and* $\|\nabla_{QQ'} \mathcal{B}(Q)\| \le \overline{H}_Q$. ■

It is easy to verify that all the examples in Section 2.2.2 satisfy the restrictions in Assumption 5. Combined, the restrictions in Assumptions 1-5 ultimately yield the following result,

$$\sup_{F\in\mathcal{F}} P_F\left(\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) - \mathcal{B}(Q_F(x,t,\theta_F^*))\right| \ge b_n\right) \longrightarrow 0, \tag{17}$$

This is the first key step towards the proof of our main result, as shown in Appendix A and in the Econometric Supplement. The next step focuses on $\mathbb{1}\{\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) \ge -b_n\} - \mathbb{1}\{\mathcal{B}(Q_F(x,t,\theta_F^*)) \ge 0\}$. 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,* $\mathcal{C}_F(b) = \left\{(x,t) \in \mathcal{X}\times\mathcal{T} : \mathcal{B}\left(Q_F(x,t,\theta_F^*)\right) \ge b\right\}$. *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* $\overline{D}_1^{VC}$.

Classes of indicator functions over VC classes of sets are Euclidean (see Pakes and Pollard (1989, p. 1033)). This fact, combined with Assumption 6, will lead manageability properties of empirical processes relevant to the study of $\mathbb{1}\{\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) \ge -b_n\} - \mathbb{1}\{\mathcal{B}(Q_F(x,t,\theta_F^*)) \ge 0\}$. The last building block towards our main asymptotic result for $\widehat{T}_2$ 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 7 ***(Smoothness III, and a regularity condition)*** *There exists a finite constant* $\overline{\mu}_{\nabla_{\mathcal{B}}}$ *such that, for each* p, $\sup_{t\in\mathcal{T}} E_F\left[\left|\frac{\partial\mathcal{B}(Q_F(X,t,\theta_F^*))}{\partial Q_p}\right|\right] \le \overline{\mu}_{\nabla_{\mathcal{B}}}\ \forall\ F \in \mathcal{F}$. *For a given* (g,y,t) *and* p, *let*

$$\Omega_{T_0}^p(y,t,g) = E_F\left[\left(S_p(y,t) - \Gamma_{p,F}(X,t,\theta_F^*)\right)\frac{\partial\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\phi(X)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right) \ge 0\right\}\middle| g(X,\theta_F^*) = g\right] \tag{18}$$

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. $\exists\, \overline{C}_4 < \infty$ *such that, for each p, and* $0 \leq m \leq M+1$, $(j_d)_{d=1}^D$ *such that* $\sum_{d=1}^D j_d = m$, *we have* $\sup_{\substack{(g_1,\ldots,g_D)\in\mathcal{G}' \\ (y,t)\in\mathcal{S}_Y\times\mathcal{T}}} \left|\frac{\partial^m \Omega_{T_0}^p(y,t,g_1,\ldots,g_D)}{\partial g_1^{j_1}\cdots\partial g_D^{j_D}}\right| \leq \overline{C}_4 \;\forall\, F \in \mathcal{F}$.

(iii) There exist finite constants $\underline{b}_2 > 0$ *and* $\overline{C}_{\mathcal{B},2} > 0$ *such that, for all* $0 < b \leq \underline{b}_2$,

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

The functional $\Omega_{T_0}^p$ described in Assumption 7 appears in the decomposition of a U-process that is relevant for $\widehat{T}_2$. Combined with our bias-reducing properties for kernels and bandwidths, parts (i) and (ii) of Assumption 7 ensure that relevant bias terms vanish asymptotically at the appropriate rate, uniformly over our testing range. **Part (iii) essentially presupposes that, conditional on $X \in \mathcal{X}$, the functional $\mathcal{B}(Q_F(X,t,\theta_F^*))$ has a density that is bounded, uniformly over $t \in \mathcal{T}$ and $F \in \mathcal{F}$, in a neighborhood of the form $[-\underline{b}_2, 0)$ (to the left of, but excluding, zero). Note that this condition allows for $\mathcal{B}(Q_F(X,t,\theta_F^*))$ to have a point-mass at zero.** A point-mass at zero would occur if the inequalities we are testing are binding with positive probability. We are now ready to present the main asymptotic result for $\widehat{T}_2$.

Proposition 1 *Let $\Delta \equiv \epsilon \wedge (\delta_0/2)$, where ϵ and δ_0 are the constants described in Assumption 4. Group $V_i \equiv (Y_i, X_i, Z_i)$. If Assumptions 1-7 hold, we have*

$$\widehat{T}_2 = T_{2,F} + \frac{1}{n}\sum_{i=1}^n \psi_F^{T_2}(V_i) + \varepsilon_n^{T_2}, \text{ where } \varepsilon_n^{T_2}| = o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \text{ uniformly over } \mathcal{F}, \text{ and} \tag{19}$$

$$\begin{aligned} &(i) \quad E_F\left[\psi_F^{T_2}(V)\right] = 0 \quad \forall\, F \in \mathcal{F}, \\ &(ii) \quad P_F\left(\mathcal{B}\left(\Gamma_F(X,t,\theta_F^*)\right) < 0 \;\middle|\; X \in \mathcal{X}_F^*\right) = 1 \;\text{ for } \mathcal{W}\text{-a.e } t \in \mathcal{T} \implies P_F\left(\psi_F^{T_2}(V) = 0\right) = 1. \end{aligned} \tag{20}$$

Proof: The steps of the proof of Proposition 1 are described in Appendix A, and the exact expression for the influence function $\psi_F^{T_2}(V)$ can be found in equations (A17)-(A20) there. Step-by-step details of the proof are included in the Econometric Supplement. ■

4.5 Constructing a test-statistic

4.5.1 Null hypotheses

Our population statistic is designed to test the null hypothesis,

$$H_0 : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F\text{–a.e } x \in \mathcal{X}_F^*, \text{ and } \mathcal{W}\text{–a.e } t \in \mathcal{T}. \tag{21}$$

Let $\mathcal{F}^0 \equiv \left\{F \in \mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \leq 0 \text{ for } F\text{–a.e } x \in \mathcal{X}_F^*, \text{ and } \mathcal{W}\text{–a.e } t \in \mathcal{T}.\right\}$. This is the subspace of distributions that satisfy H_0. Therefore, $\mathcal{F} \setminus \mathcal{F}^0$ is the subspace of distributions that violate H_0. Now let, $\overline{\mathcal{F}} \equiv \left\{F \in \mathcal{F} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) < 0 \text{ for } F\text{–a.e } x \in \mathcal{X}_F^*, \text{ and } \mathcal{W}\text{–a.e } t \in \mathcal{T}.\right\}$. This is the subset of distributions in $\mathcal{F}^0$ for which the inequalities in H_0 are satisfied as *strict inequalities* almost surely.

Remark 3 ***(Contact sets and $\sigma_{2,F}^2 \equiv E_F[\psi_F^{T_2}(V)^2]$)***
For each $F \in \mathcal{F}$, we define the contact sets as $\{(x,t) \in \mathcal{X}_F^* \times \mathcal{T} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) \geq 0\}$. Thus, for each $F \in \mathcal{F}^0$, the contact sets are $\{(x,t) \in \mathcal{X}_F^* \times \mathcal{T} : \mathcal{B}(\Gamma_F(x,t,\theta_F^*)) = 0\}$ (all (x,t) in our target testing range for which the inequalities in H_0 are binding). Denote $\sigma_{2,F}^2 \equiv E_F[\psi_F^{T_2}(V)^2]$. From the properties of the influence function $\psi_F^{T_2}(V)$ summarized in (20), ***$\sigma_{2,F}$ will be the relevant measure of the contact sets for our test.*** *For any $F \in \mathcal{F}^0$, the contact sets have measure zero if and only if $\sigma_{2,F} = 0$. From part (ii) of equation (20), we have $\sigma_{2,F} = 0 \; \forall \; F \in \overline{\mathcal{F}}$.*

We will use the results in Proposition 1 to construct a test for the null hypothesis H_0 described in (21). The first step is to construct an estimator for $\sigma_{2,F}^2$.

4.5.2 Estimation of $\sigma_{2,F}^2$

Under the conditions of Proposition 1, $\exists \; \overline{\sigma}_2 < \infty$ such that $E_F\left[\psi_F^{T_2}(V)^2\right] \equiv \sigma_{2,F}^2 \leq \overline{\sigma}_2$ for all $F \in \mathcal{F}$. From here, a Chebyshev inequality yields,

$$\left| \frac{1}{n}\sum_{i=1}^{n} \psi_F^{T_2}(V_i)^2 - \sigma_{2,F}^2 \right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{22}$$

Based on the expression of the influence function $\psi_F^{T_2}(V)$, described in Appendix A (see equations A17-A20), we estimate $\sigma_{2,F}^2$ by constructing an estimator $\widehat{\psi}^{T_2}(V)$ for $\psi_F^{T_2}(V)$. Appendix A describes its construction under the following assumption.

Assumption 8 ***(An estimator for the influence function of $\widehat{\theta}$)*** *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 estimator for $\sigma^2_{2,F}$ is $\widehat{\sigma}_2^2 \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}^{T_2}(V_i)^2$. The exact expression for $\widehat{\psi}^{T_2}(V)$ is included in Appendix A (Section A2, equation A22). Under the conditions of Proposition 1, combined with Assumption 8, we show that

$$\left|\widehat{\sigma}_2^2 - \sigma^2_{2,F}\right| = o_p(1) \quad \text{uniformly over } \mathcal{F}. \tag{23}$$

4.5.3 Our proposed test

Recall that $\mathcal{F}^0$ denotes the class of distributions that satisfy H_0, and $\overline{\mathcal{F}} \subseteq \mathcal{F}^0$ denotes the subclass for which the inequalities in H_0 are satisfied as *strict inequalities*. As we pointed out above, $\sigma_{2,F} = 0$ $\forall\ F \in \overline{\mathcal{F}}$. Thus, to studentize $\widehat{T_2}$ we need to regularize $\widehat{\sigma}_2$. Let $\kappa_2 > 0$ denote a pre-specified, small but strictly positive constant. We consider the test-statistic,

$$\widehat{t_2} = \frac{\sqrt{n}\widehat{T_2}}{(\widehat{\sigma}_2 \vee \kappa_2)}. \tag{24}$$

We use $(\widehat{\sigma}_2 \vee \kappa_2)$ instead of $\widehat{\sigma}_2$ in our studentization because $\sigma_{2,F} = 0\ \forall\ F \in \overline{\mathcal{F}}$. Under the conditions of Proposition 1, we have

$$\widehat{t_2} = \begin{cases} \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma}_2 \vee \kappa_2)} & \forall\ F \in \overline{\mathcal{F}}, \\ \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_2}(V_i)}{(\widehat{\sigma}_2 \vee \kappa_2)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma}_2 \vee \kappa_2)} & \forall\ F \in \mathcal{F}^0 \setminus \overline{\mathcal{F}}, \\ \frac{\sqrt{n}\cdot T_{2,F}}{(\widehat{\sigma}_2 \vee \kappa_2)} + \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n \psi_F^{T_2}(V_i)}{(\widehat{\sigma}_2 \vee \kappa_2)} + \frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma}_2 \vee \kappa_2)} & \forall\ F \in \mathcal{F} \setminus \mathcal{F}^0 \end{cases} \tag{25}$$

From Proposition 1, $\left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{(\widehat{\sigma}_2 \vee \kappa_2)}\right| \leq \left|\frac{\sqrt{n}\cdot\varepsilon_n^{T_2}}{\kappa_2}\right| = o_p\left(\frac{n^{1/2}}{n^{1/2+\Delta}}\right) = o_p\left(\frac{1}{n^\Delta}\right) = o_p(1)$ uniformly over $\mathcal{F}$, with $\Delta > 0$ being the constant described in Proposition 1. Thus, the first implication of (25) is that $\widehat{t_2} = o_p(1)$ uniformly over $F \in \overline{\mathcal{F}}$. Thus, for any $c > 0$,

$$\lim_{n\to 0} \sup_{F \in \overline{\mathcal{F}}} P_F\left(\widehat{t_2} \geq c\right) = 0. \tag{26}$$

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 (25)-(26), we propose the following rejection rule for H_0 in (21),

$$\text{Reject } H_0 \text{ iff } \widehat{t_2} \geq z_{1-\alpha} \tag{27}$$

Our alternative hypothesis is simply that H_0 is violated. From (25), a uniform Berry-Esseen condition would suffice for our proposed tests to be uniformly asymptotically level α. We describe that condition next.

Assumption 9 ***(A sufficient condition for a uniform Berry-Esseen bound)*** *For some $B < \infty$, we have* $\frac{E_F\left[\left|\psi_F^{T_2}(V)\right|^3\right]}{\sigma_{2,F}^3} < B \;\forall\; F \in \mathcal{F} \setminus \overline{\mathcal{F}}$. ■

Assumption 9 allows for $\sigma_{2,F}$ (the relevant measure of the contact sets in our test) to be arbitrarily close to zero over $\mathcal{F} \setminus \overline{\mathcal{F}}$. 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

$$\sup_{F\in\mathcal{F}\setminus\overline{\mathcal{F}}} \sup_{d} \left| P_F\left(\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} \psi_F^{T_2}(V_i)}{\sigma_{2,F}} \leq d \right) - \Phi(d) \right| \leq \frac{C}{n^{1/2}}. \tag{28}$$

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,

$$\lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}: \\ \sigma_{2,F}\geq\kappa_2}} \left| P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi_F^{T_2}(V_i)}{(\sigma_{2,F} \vee \kappa_2)} \geq z_{1-\alpha} \right) - \alpha \right| = 0, \quad \limsup_{n\to\infty} \sup_{\substack{F\in\mathcal{F}\setminus\overline{\mathcal{F}}: \\ \sigma_{2,F}<\kappa_2}} P_F\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\psi_F^{T_2}(V_i)}{(\sigma_{2,F} \vee \kappa_2)} \geq z_{1-\alpha} \right) \leq \alpha. \tag{29}$$

From (23), we have $\left|\frac{1}{(\widehat{\sigma_2}\vee\kappa_2)} - \frac{1}{(\sigma_{2,F}\vee\kappa_2)}\right| = o_p(1)$ uniformly over $\mathcal{F} \setminus \overline{\mathcal{F}}$. Thus, from (29),

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

Combining (26) with the above result, we have

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{F}^0} P_F\left(\widehat{t_2} \geq z_{1-\alpha}\right) \leq \alpha, \quad \text{with} \quad \lim_{n\to\infty} \sup_{\substack{F\in\mathcal{F}^0\setminus\overline{\mathcal{F}}: \\ \sigma_{2,F}\geq\kappa_2}} \left| P_F\left(\widehat{t_2} \geq z_{1-\alpha}\right) - \alpha \right| = 0. \tag{30}$$

Thus, our proposed test has *uniformly asymptotically level α* (Lehmann and Romano (2005, Definition 11.1.2)). Next, consider $\mathcal{F}\setminus\mathcal{F}^0$ (the subset of distributions that violate H_0). Take any sequence $F_n \in \mathcal{F} \setminus \mathcal{F}^0$ such that $\sqrt{n} \cdot T_{2,F_n} \geq \delta_n D$ for some fixed $D > 0$, and a sequence of positive constants $\delta_n \to \infty$. From Assumption 9 and equation (28), we have that for any $c > 0$,

$$\lim_{n\to\infty} P_{F_n}\left(\frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} \psi_{F_n}^{T_2}(V_i)}{(\sigma_{2,F_n} \vee \kappa_2)} + \frac{\sqrt{n} \cdot T_{2,F_n}}{(\widehat{\sigma_2} \vee \kappa_2)} \geq c \right) = 1, \quad \text{and therefore} \quad \lim_{n\to\infty} P_{F_n}\left(\widehat{t_2} \geq c\right) = 1.$$

More generally, consider a 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). For any such sequence we have

$$\lim_{n\to\infty} \left| P_{F_n}(\widehat{t_2} \geq z_{1-\alpha}) - \left[1 - \Phi\left(s_2^a \cdot (z_{1-\alpha} - s_2^b)\right)\right] \right| = 0. \tag{31}$$

We say that our rejection rule for H_0 has *nontrivial asymptotic power* for a sequence $F_n \in \mathcal{F} \setminus \mathcal{F}^0$ if $\lim_{n\to\infty} P_{F_n}\left(\widehat{t_2} \geq z_{1-\alpha}\right) > \alpha$ (see Lee, Song, and Whang (2018, Definition 3)). It follows that our test for H_0 will have nontrivial power for F_n iff $s_2^a \cdot (z_{1-\alpha} - s_2^b) < z_{1-\alpha}$. Combining the results in equations (25)-(31), the following theorem summarizes the properties of our proposed test.

Theorem 1 *Consider the test for H_0 described by the rejection rule given in* (27) *for a target significance level α. If Assumptions 1-9 hold, our test has the following properties.*

(i) Uniformly asymptotically level α: Our proposed test satisfies,

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

(ii) Consistency: Take any sequence of distributions distributions $F_n \in \mathcal{F}\setminus\mathcal{F}^0$ 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) Local alternatives against which our test has nontrivial asymptotic power: Consider a sequence of distributions F_n such that $\left(\frac{\sigma_{2,F_n}\vee\kappa_2}{\sigma_{2,F_n}}\right) \longrightarrow s_2^a$ and $\frac{\sqrt{n}\cdot T_{2,F_n}}{(\sigma_{2,F_n}\vee\kappa_2)} \longrightarrow s_2^b$ (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* $s_2^a \cdot (z_{1-\alpha} - s_2^b) < z_{1-\alpha}$.

Proof: Theorem 1 is a summary of the results in equations (25)-(31). ■

4.6 On the choice of tuning parameters

While we leave a fully developed theory of bandwidth selection for future work, we present reasonably detailed recommendations in Section 5, where we perform a series of Monte Carlo experiments. Our proposed choice of h_n follows a "rule of thumb" approach (see Silverman (1986, Section 3.4)), while our tuning parameters b_n and κ_2 are chosen to be proportional to $\overline{\mathcal{B}} \equiv \sup_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta}))\right|$. Our tuning parameter recommendations perform well across our range of experiments, both in terms of power and size.

4.7 Extensions included in Appendix A

Appendix A (Section A4) presents two extensions of our results. The first extension describes conditions under which we can let the regularization parameter κ_2 vanish asymptotically (i.e, $\kappa_{2,n} \longrightarrow 0$). This can be done if we assume that $\sigma_{2,F}$ (the relevant measure of the contact sets) is bounded away from zero over $\mathcal{F}\setminus\overline{\mathcal{F}}$ (our current assumptions allow for $\sigma_{2,F}$ to be arbitrarily close to zero over $\mathcal{F}\setminus\overline{\mathcal{F}}$). In a second extension, we show that our approach can be readily extended to testing multiple functional inequalities of the form (5), and we describe a test-statistic that generalizes the results we obtained for the single inequality case.

4.8 Comparison to existing methods

To our knowledge, this is the first paper devoted to testing functional inequalities that are conditioned on estimated functions or *generated regressors*. For this reason, 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 $\mathcal{B}$ 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. This avoids the need to estimate the contact sets in a first step, as is done, e.g, 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). 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 in our case, a general theory of how to choose these tuning parameters has been left to future work. A formal analysis of how to adapt the aforementioned existing methods to the case of estimated conditioning functions is outside the scope of this paper.

5 Monte Carlo experiments

We apply our method to test for conditional, first-order stochastic dominance relationship between two scalar random variables, (Y_1, Y_2). Our conjecture is $F_{Y_1|X}(t|X) \leq F_{Y_2|X}(t|X)$ $F-$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 rewritten as,

$$E_F[S(Y,t)|X] \leq 0 \; F\text{–a.e } X, \; \forall \; t \tag{29}$$

5.1 Designs

X includes 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 seven data generating processes (DGPs), described in Table 1.

Table 1: Monte Carlo designs

Design	Description	Is the FOSD inequality (29) 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	$Y_1 = m_I(X) + 3.2 + \varepsilon_1$ $Y_2 = (m_I(X) + 3.2) \vee (m_{II}(X) - 3.2) + \varepsilon_2$	**No**. It is violated with probability **≈ 5%** for each $t \in \mathbb{R}$.
DGP 6	$Y_1 = m_I(X) + 4.1 + \varepsilon_1$ $Y_2 = (m_I(X) + 4.1) \vee (m_{II}(X) - 4.1) + \varepsilon_2$	**No**. It is violated with probability **≈ 2.5%** for each $t \in \mathbb{R}$.
DGP 7	$Y_1 = m_I(X) + 5.4 + \varepsilon_1$ $Y_2 = (m_I(X) + 5.4) \vee (m_{II}(X) - 5.4) + \varepsilon_2$	**No**. It is violated with probability **≈ 1%** for each $t \in \mathbb{R}$.

Our first goal is to evaluate our procedure in two extreme scenarios where the null hypothesis is satisfied: when the inequalities are satisfied as strict inequalities w.p.1. (DPG 1), and when they are binding w.p.1 (DGP 2). Our second goal is to study its power properties as the probability that the inequalities are violated diminishes. In DGPs 3-7, the probability of a violation goes from 50% (DGP 3) to only 1% (DGP 7).

5.2 Conditioning function employed

We consider a case where X is aggregated into a single linear index $g(X,\widehat{\theta}) = X'\widehat{\theta}$, where $\widehat{\theta}$ is estimated through an OLS regression of $\Delta Y \equiv Y_1 - Y_2$ on X. This index is not designed to be "optimal" in any way, but instead is meant to depict a plausible way in which an applied researcher may want to aggregate X. 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 $\psi_F^{\theta}(Z_i) \equiv (E_F[XX'])^{-1} \cdot X_i\nu_i$. As shown formally in Appendix B, under conditions that are satisfied by our designs, the OLS estimator satisfies the restrictions in Assumption 1, with[4] $\widehat{\theta} = \theta_F^* + \frac{1}{n}\sum_{i=1}^n \psi_F^{\theta}(Z_i) + \varepsilon_n^{\theta}$. Thus, our population conditioning function is $g(x,\theta_F^*) \equiv x'\theta_F^*$, and our goal is to test

$$E_F[S(Y,t)|X'\theta_F^*] \le 0 \;\; F\text{–a.s}, \forall\, t \tag{29'}$$

Note that (29') follows from (29) by iterated expectations. Thus, a rejection of (29') would lead to a rejection of the FOSD null hypothesis in (29).

5.3 Test statistic

Our test is conducted using the test-statistic described in Section 4. First, $\widehat{T}_2$ is constructed as described in equations (13)-(14),

$$\widehat{T}_2 \equiv \int_t \left(\frac{1}{n}\sum_{i=1}^n \widehat{Q}(X_i,t,\widehat{\theta})\mathbb{1}\left\{\widehat{Q}(X_i,t,\widehat{\theta}) \ge -b_n\right\}\phi(X_i)\right) d\mathcal{W}(t), \quad \text{with} \quad \widehat{Q}(x,t,\widehat{\theta}) \equiv \frac{\widehat{R}(x,t,\widehat{\theta})}{\widehat{f}_g(g(x,\widehat{\theta}))},$$

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

5.3.1 Testing range and tuning parameters

For $\tau \in (0,1)$, let $\psi_{(\tau)}$ denote the $\tau-$ quantile of the r.v ψ. Let $\underline{Y}_{(\tau)} \equiv Y_{1(\tau)} \vee Y_{2(\tau)}$ and $\overline{Y}_{(1-\tau)} \equiv Y_{1(1-\tau)} \wedge Y_{2(1-\tau)}$. Our testing range $\mathcal{T}$ for the index variable t was set to $\mathcal{T} = [\underline{Y}_{(\tau)}, \overline{Y}_{(1-\tau)}]$, with $\tau = 10^{-3}$, with the weight function $d\mathcal{W}$ set to be the uniform measure over $\mathcal{T}$. Our testing range $\mathcal{X}$ was set to, $\mathcal{X} = \left\{x \in \mathbb{R}^8\colon X_{(\tau)} \le x \le X_{(1-\tau)}\,,\, (X'\widehat{\theta})_{(\tau)} \le x'\widehat{\theta} \le (X'\widehat{\theta})_{(1-\tau)}\right\}$, where, once again, $\tau = 10^{-3}$

[4]Appendix B 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 $q \ge 2$.

and where $X_{(\tau)} \leq x \leq X_{(1-\tau)}$ denotes element-wise inequalities[5]. As weight function, we simply used $\phi(X_i) = \omega(X_i'\widehat{\theta}) = \mathbb{1}\{X_i \in \mathcal{X}\}$.

We follow a conventional approach to choose h_n, letting $h_n = c_h \cdot n^{-\alpha_h}$, where $\alpha_h > 0$ denotes the rate of convergence of h_n. To satisfy the bandwidth convergence restrictions in Assumption 4, we set $\alpha_h = 1/8 - 10^{-5}$. We chose c_h according to "rule of thumb" recommendations (see Silverman (1986, Section 3.4)), setting $c_h = 0.9 \cdot \min\{\widehat{\sigma}(X'\widehat{\theta})\ , [(X'\widehat{\theta})_{(0.75)} - (X'\widehat{\theta})_{(0.25)}]/1.34\}$. We chose our remaining tuning parameters, b_n and κ_2, to be proportional to a measure of the scale of $\widehat{Q}(\cdot)$. Let $\overline{\mathcal{B}} \equiv \sup\limits_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left|\widehat{Q}(x,t,\widehat{\theta})\right|$. We set $\kappa_2 = c_k \cdot \overline{\mathcal{B}}$ and $b_n = c_b \cdot \overline{\mathcal{B}} \cdot n^{-\alpha_b}$, with $\alpha_b = 1/4 + 3/4 \cdot 10^{-5}$, which satisfies the bandwidth convergence restrictions in Assumption 4. By construction, small values of c_b and c_k enhance the power properties of our test, but could lead to over-rejection of the null hypothesis when the latter is binding w.p.1 (DGP 2). In our experiments we found that an aggressive (i.e, small) choice for c_b coupled with a relatively more conservative choice for c_k achieved a good compromise between power and size. We obtained good results in all our experiments by choosing $c_b = 10^{-4}$ and $c_k = 0.01\cdot(1/\log(n))$. Note that our choice of c_k implies that the regularization parameter κ_2 vanishes (slowly) asymptotically. In Appendix A (Section A4.1), we show conditions under which this is valid. These conditions are satisfied in our experiments.

Regarding our choice of kernel, the restrictions in Assumption 4 are satisfied[6] if we use a bias-reducing kernel of order $M = 5$. We employed a symmetric kernel with support $[-S, S]$ of the form, $K(\psi) = \left(c_1 \cdot (S^2 - \psi^2)^2 + c_2 \cdot (S^2 - \psi^2)^4 + c_3 \cdot (S^2 - \psi^2)^6\right) \cdot \mathbb{1}\{|\psi| \leq S\}$, where c_1, c_2 and c_3 are chosen to satisfy[7] $\int_{-S}^{S} K(\psi)d\psi = 1$, $\int_{-S}^{S} \psi^2 K(\psi)d\psi = 0$ and $\int_{-S}^{S} \psi^4 K(\psi)d\psi = 0$. In our experiments we chose $S = 5$, so the support of the kernel is $[-5, 5]$.

5.4 Results

The results of our experiments are summarized in Table 2 for a range of sample sizes n between 250 and 4,000, and a target significance level of 5%. Overall, with our tuning parameter choices, our rejection frequencies are in line with the asymptotic predictions of Theorem 1 in terms of size and power. Our first finding is that our test has remarkable power even when the probability of violation is low. When this probability is 5% (DGP 5), our test rejects the null hypothesis with probability greater than 50% in samples as small as $n = 500$. This figure jumps to 78% when $n = 1,000$. If violations occur with probability 50% (DGP 3), our test will reject the null

[5]Our testing range was obtained for each sample generated in our experiments by using the corresponding sample quantiles.

[6]In Appendix A 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 (as in our experiments), 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 $M = 2D + 3$. Since $D = 1$ in our experiments, a bias-reducing kernel of order $M = 5$ satisfies our restrictions.

[7]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.

with frequency greater than 98% when $n = 1,000$. Our test also has nontrivial power when the probability of a violation is very small. For example, when the inequalities are violated with probability 1% (DGP 7), our test rejects with a frequency greater than 13% when $n = 100$. This figure grows to 20% when $n = 1,000$ and 40% when $n = 2,000$. For samples of size $n = 4,000$, our test will reject the null hypothesis with frequency 74% when FOSD is violated with just 1% probability.

The power performance of our test with our tuning parameter choices did not come at the expense of distortions in size. As our results show, the rejection frequencies were close to 5% in both DGPs 1 and 2 for all sample sizes. This was particularly welcome news for DGP 2, where the FOSD inequalities are binding w.p.1. The tuning parameter κ_2 plays an important role in DGP 1, where the FOSD inequalities hold strictly w.p.1, and regularization is crucial to achieve our target size asymptotically. Our choice for this tuning parameter (whose intuition was described above) brought our finite sample rejection rates in line with our asymptotic predictions.

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

Sample size	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
$n = 250$	0.0%	6.3%	96.3%	82.0%	33.8%	21.5%	13.0%
$n = 500$	0.0%	4.3%	98.5%	94.3%	53.1%	29.6%	15.7%
$n = 1,000$	0.3%	4.1%	99.0%	98.0%	77.7%	47.9%	20.1%
$n = 2,000$	2.1%	6.1%	99.3%	99.2%	97.6%	81.3%	39.8%
$n = 3,000$	4.1%	5.2%	99.4%	99.4%	99.0%	94.1%	56.9%
$n = 4,000$	4.7%	4.7%	99.7%	99.7%	99.5%	97.4%	74.2%

• Target significance level 5% in all cases.

6 Concluding remarks

Many economic models produce testable implications in the form of functional inequalities, which can be tested nonparametrically. In many instances the data may consist of a rich collection of conditioning variables, and researchers may want to pursue dimension reduction and aggregate them into lower-dimensional conditioning functions or "indices" in an effort to mitigate curse of dimensionality issues. Motivated by this problem, we introduced a test for functional inequalities conditional on estimated functions. We focused on the case where the researcher chooses a parametric form for these conditioning functions and estimates its parameters in a first step. The researcher then proceeds to construct a test of the functional inequalities in question, conditional on the estimated functions. Taking the choice of the conditioning functions and their estimators as given, we proposed tests 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 applied to single or multiple functional inequalities. Furthermore, our general conditions for the first-step estimators encompass a wide variety of extremum estimators as special cases. In Monte Carlo experiments, our test displayed good size control and power properties, capable of detecting violations to the inequalities that occur with very small probability.

Appendix A

A1 Proposition 1

This section outlines the steps of the proof of Proposition 1. The step-by-step details are included in the **Econometric Supplement**, which can be downloaded at `https://aaradill.github.io/condit-ineq-functions-supplement.pdf`. As we defined in Section 4.2, 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, t, θ, our estimators are $\widehat{f_g}(g(x,\theta)) \equiv \frac{1}{n \cdot h_n^D} \sum_{i=1}^n K\left(\frac{\Delta g(X_i, x, \theta)}{h_n}\right)$, $\widehat{R}_p(x,t,\theta) \equiv \frac{1}{n \cdot h_n^D} \sum_{i=1}^n S_p(Y_i, t)\omega_p(g(X_i,\theta))K\left(\frac{\Delta g(X_i, x, \theta)}{h_n}\right)$, and $\widehat{Q}_p(x,t,\widehat{\theta}) \equiv \frac{\widehat{R}_p(x,t,\widehat{\theta})}{\widehat{f_g}(g(x,\widehat{\theta}))}$, with $\widehat{Q}(x,t,\widehat{\theta}) \equiv \left(\widehat{Q}_1(x,t,\widehat{\theta}), \ldots, \widehat{Q}_P(x,t,\widehat{\theta})\right)'$.

A1.1 Two key preliminary results for $\widehat{Q}(x,t,\widehat{\theta}) - Q_F(x,t,\theta_F^*)$

Let $\Omega_{f_g}^{d,\ell}$, $\Omega_{R_p,1}^{d,\ell}$, $\Omega_{R_p,2}$ and $\Omega_{R_p,3}^{\ell}$ be as defined in Assumption 2. For a given $(x,t) \in \mathcal{X} \times \mathcal{T}$, let

$$
\begin{aligned}
\Xi_{\ell,f_g}(x,\theta_F^*) &\equiv \sum_{d=1}^{D} \left(\frac{\partial g_d(x,\theta_F^*)}{\partial \theta_\ell} \cdot \frac{\partial f_g(g(x,\theta_F^*))}{\partial g_d} - \frac{\partial \left[\Omega_{f_g}^{d,\ell}(g(x,\theta_F^*)) \cdot f_g(g(x,\theta_F^*))\right]}{\partial g_d} \right), \\
\underbrace{\Xi_{f_g}(x,\theta_F^*)}_{1\times k} &\equiv \left(\Xi_{1,f_g}(x,\theta_F^*), \ldots, \Xi_{k,f_g}(x,\theta_F^*)\right), \\
\Xi_{\ell,R_p}(x,t,\theta_F^*) &\equiv \sum_{d=1}^{D} \left(\frac{\partial \left[\Omega_{R_p,2}(g(x,\theta_F^*),t) f_g(g(x,\theta_F^*))\right]}{\partial g_d} \cdot \frac{\partial g_d(x,\theta_F^*)}{\partial \theta_\ell} - \frac{\partial \left[\Omega_{R_p,1}^{d,\ell}(g(x,\theta_F^*),t) f_g(g(x,\theta_F^*))\right]}{\partial g_d} \right. \\
&\qquad \left. + \Omega_{R_p,3}^{\ell}(g(x,\theta_F^*),t) \cdot f_g(g(x,\theta_F^*)) \right), \\
\underbrace{\Xi_{R_p}(x,t,\theta_F^*)}_{1\times k} &\equiv \left(\Xi_{1,R_p}(x,t,\theta_F^*), \ldots, \Xi_{k,R_p}(x,t,\theta_F^*)\right).
\end{aligned}
\tag{A1}
$$

Note that, for each p, d, ℓ, $\Omega^{d,\ell}_{R_p,1}(g,t) = \Omega_{R_p,2}(g,t) = \Omega^{\ell}_{R_p,3}(g,t) = 0 \;\forall\; g \notin \mathcal{G}$, and therefore,

$$\Xi_{R_p}(x,t,\theta^*_F) = 0 \quad \forall\; (x,t) : g(x,\theta^*_F) \notin \mathcal{G}. \tag{A2}$$

Let $\Xi_{f_g}(x,\theta^*_F)$ and $\Xi_{R_p}(x,t,\theta^*_F)$ be as described in (A1) and, for each p, define

$$\underbrace{\Xi_{Q_p}(x,t,\theta^*_F)}_{1\times k} \equiv \frac{\Xi_{R_p}(x,t,\theta^*_F) - Q_{p,F}(x,t,\theta^*_F)\cdot \Xi_{f_g}(x,\theta^*_F)}{f_g(g(x,\theta^*_F))} \tag{A3}$$

Note from (A2) and the definition of $Q_{p,F}$ that

$$\Xi_{Q_p}(x,t,\theta^*_F) = 0 \quad \forall\; (x,t) : g(x,\theta^*_F) \notin \mathcal{G}. \tag{A4}$$

Let,

$$\begin{aligned}\psi^{Q_p}_F(V_i,x,t,\theta^*_F,h_n) \equiv & \frac{1}{h_n^D}\Bigg\{\left(\frac{S_p(Y_i,t)-\Gamma_{p,F}(x,t,\theta^*_F)}{f_g(g(x,\theta^*_F))}\right)\cdot\omega_p(g(X_i,\theta^*_F))\cdot K\left(\frac{\Delta g(X_i,x,\theta^*_F)}{h_n}\right)\\ & - E_F\left[\left(\frac{S_p(Y_i,t)-\Gamma_{p,F}(x,t,\theta^*_F)}{f_g(g(x,\theta^*_F))}\right)\cdot\omega_p(g(X_i,\theta^*_F))\cdot K\left(\frac{\Delta g(X_i,x,\theta^*_F)}{h_n}\right)\right]\Bigg\}\\ & + \Xi_{Q_p}(x,t,\theta^*_F)\psi^{\theta}_F(Z_i), \quad \text{with}\\ \psi^{Q}_F(V_i,x,t,\theta^*_F,h_n) \equiv & \left(\psi^{Q_1}_F(V_i,x,t,\theta^*_F,h_n),\ldots,\psi^{Q_P}_F(V_i,x,t,\theta^*_F,h_n)\right)'.\end{aligned} \tag{A5}$$

Proposition S1 in the Econometric Supplement shows that, under Assumptions 1-4,

$$\widehat{Q}(x,t,\widehat{\theta}) = Q_F(x,t,\theta^*_F) + \frac{1}{n}\sum_{i=1}^{n}\psi^{Q}_F(V_i,x,t,\theta^*_F,h_n) + \zeta^{Q}_n(x,t), \quad \text{where}$$

$$\left.\begin{aligned}&\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left\|\zeta^{Q}_n(x,t)\right\| = o_p\left(\tfrac{1}{n^{1/2+\epsilon}}\right),\\ &\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left\|\widehat{Q}(x,t,\widehat{\theta}) - Q_F(x,t,\theta^*_F)\right\| = o_p\left(\tfrac{1}{n^{1/4+\epsilon/2}}\right)\end{aligned}\right\}\text{ uniformly over } \mathcal{F},$$

where $\epsilon > 0$ is the constant described in Assumption 4. In addition, we also show that

$$\sup_{F\in\mathcal{F}} P_F\left(\sup_{(x,t)\in\mathcal{X}}\left|\widehat{Q}(x,t,\widehat{\theta}) - Q_F(x,t,\theta^*_F)\right| \geq b_n\right) \longrightarrow 0. \tag{A6}$$

Combined with Assumptions 1-5, we show in the Econometric Supplement that (A6) yields,

$$\sup_{F\in\mathcal{F}} P_F\left(\sup_{(x,t)\in\mathcal{X}}\left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) - \mathcal{B}(Q_F(x,t,\theta^*_F))\right| \geq b_n\right) \longrightarrow 0, \tag{A7}$$

and we have the following linear representation result,

$$\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) = \mathcal{B}(Q_F(x,t,\theta_F^*)) + \frac{1}{n}\sum_{i=1}^{n}\psi_F^{\mathcal{B}}(V_i,x,t,\theta_F^*,h_n) + \zeta_n^{\mathcal{B}}(x,t), \quad \text{where,}$$

$$\psi_F^{\mathcal{B}}(V_i,x,t,\theta_F^*,h_n) \equiv \nabla_Q\mathcal{B}(Q_F(x,t,\theta_F^*))\psi_F^{Q}(V_i,x,t,\theta_F^*,h_n) = \sum_{p=1}^{P}\frac{\partial\mathcal{B}\left(Q_F(x,t,\theta_F^*)\right)}{\partial Q_p}\psi_F^{Q_p}(V_i,x,t,\theta_F^*,h_n),$$

$$\left.\begin{array}{l} \sup\limits_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left\|\zeta_n^{\mathcal{B}}(x,t)\right\| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right), \text{ and} \\ \sup\limits_{(x,t)\in\mathcal{X}\times\mathcal{T}} \left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta})) - \mathcal{B}(Q_F(x,t,\theta_F^*))\right| = o_p\left(\frac{1}{n^{1/4+\epsilon/2}}\right) \end{array}\right\} \text{uniformly over } \mathcal{F}, \tag{A8}$$

where $\epsilon > 0$ is the constant described in Assumption 4 and ψ_F^{Q}, $\psi_F^{Q_p}$ are as described in (A5).

A1.2 Steps of the proof of Proposition 1

The results in (A7)-(A8) are the building blocks of the proof of Proposition 1. We outline the main steps of the proof here, with all the details included in the Econometric Supplement. Recall that $T_{2,F} \equiv \int_t T_{0,F}(t)d\mathcal{W}(t)$, where $T_{0,F}(t) \equiv E_F\left[\left(\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)\right)_+\phi(X,t)\right]$, and our estimators are $\widehat{T_2} \equiv \int_t \widehat{T_0}(t)d\mathcal{W}(t)$, where $\widehat{T_0}(t) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}\phi(X_i)$. Let $\widetilde{T}_{0,F}(t) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) \geq 0\right\}\phi(X_i)$. Note that $\widetilde{T}_{0,F}(t)$ takes $\widehat{T_0}(t)$ and replaces the indicator function $\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}$ with $\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) \geq 0\right\}$. Let $\widehat{T_0}(t) - T_{0,F}(t) \equiv \xi_{T_0,n}^{a}(t)$. Note that,

$$\begin{aligned}
&\left|\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\} - \mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) \geq 0\right\}\right| \\
&= \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\,,\ -2b_n \leq \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) < 0\right\} + \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\,,\ \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) < -2b_n\right\} \\
&+ \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) < -b_n\,,\ \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) \geq 0\right\} \\
&\leq \mathbb{1}\left\{-2b_n \leq \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) < 0\right\} + \mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) - \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\right| \geq b_n\right\}
\end{aligned}$$

From here, we obtain,

$$\begin{aligned}
&\left|\xi_{T_0,n}^{a}(t)\right| \\
&\leq \left(2b_n + \sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\mathcal{B}\left(\widehat{Q}(x,t,\widehat{\theta})\right) - \mathcal{B}\left(Q_F(x,t,\theta_F^*)\right)\right|\right) \times \frac{1}{n}\sum_{i=1}^{n}\phi(X_i,t)\mathbb{1}\left\{-2b_n \leq \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right) < 0\right\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\right|\phi(X_i,t)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) - \mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\right| \geq b_n\right\}
\end{aligned}$$

In the Econometric Supplement we show, $\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\mathcal{B}\left(\widehat{Q}(x,t,\widehat{\theta})\right)-\mathcal{B}\left(Q(x,t,\theta_F^*)\right)\right| = o_p\left(\frac{1}{n^{1/4+\epsilon/2}}\right)$ uniformly over $\mathcal{F}$, where $\epsilon>0$ is as described in Assumption 4. Thus, uniformly over $\mathcal{F}$,

$$\begin{aligned}\left|\xi^a_{T_0,n}(t)\right| \leq &\left(2b_n + o_p\left(\frac{1}{n^{1/4+\epsilon/2}}\right)\right)\times\frac{1}{n}\sum_{i=1}^n \phi(X_i,t)\mathbb{1}\left\{-2b_n \leq \mathcal{B}(Q(X_i,t,\theta_F^*))<0\right\}\\ &+\frac{1}{n}\sum_{i=1}^n\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\right|\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}(Q(X_i,t,\theta_F^*))\right|\geq b_n\right\}.\end{aligned} \tag{A9}$$

In the Econometric Supplement we analyze the two terms in (A9). For a given $b>0$, $t\in\mathcal{T}$, let $m^a_{T_0,n}(b,t) \equiv \frac{1}{n}\sum_{i=1}^n\left(\phi(X_i)\mathbb{1}\left\{-b\leq\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)<0\right\}-E_F\left[\phi(X_i)\mathbb{1}\left\{-b\leq\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)<0\right\}\right]\right)$. From Assumption 6, we show that, $\sup_{\substack{0<b<b_0\\ t\in\mathcal{T}}}\left|m^a_{T_0,n}(b,t)\right| = O_p\left(\frac{1}{n^{1/2}}\right)$ uniformly over $\mathcal{F}$. For n large enough we have $0<2b_n\leq b_0$. Therefore, for n large enough,

$$\left|m^a_{T_0,n}(2b_n,t)\right| \leq \sup_{\substack{0<b<b_0\\ t\in\mathcal{T}}}\left|m^a_{T_0,n}(b,t)\right| = O_p\left(\frac{1}{n^{1/2}}\right)\quad\text{uniformly over }\mathcal{F}. \tag{A10}$$

We have $\frac{1}{n}\sum_{i=1}^n\phi(X_i)\mathbb{1}\left\{-2b_n\leq\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)<0\right\} = m^a_{T_0,n}(2b_n,t)+E_F\left[\phi(X)\mathbb{1}\left\{-2b_n\leq\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)<0\right\}\right]$. From Assumption 7 and the bounded properties of $\phi(\cdot)$, there exist finite constants $\underline{b}_2>0$ and $\overline{C}_{\mathcal{B},2}>0$ such that, for all $0<b\leq\underline{b}_2$, $\sup_{t\in\mathcal{T}}E_F\left[\phi(X)\mathbb{1}\left\{-b\leq\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)<0\right\}\right]\leq\overline{\phi C}_{\mathcal{B},2}\cdot b\quad\forall\, F\in\mathcal{F}$. For n large enough, $0<2b_n\leq\underline{b}_2\wedge b_0$, and from Assumption 4, we have $n^{1/2}\cdot b_n\longrightarrow\infty$. This, combined with equation (A10) and Assumption 7 yields,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n\phi(X_i)\mathbb{1}\left\{-2b_n\leq\mathcal{B}(Q_F(X_i,t,\theta_F^*))<0\right\}&\leq O_p\left(\frac{1}{n^{1/2}}\right)+O(b_n) = b_n\cdot\left(O_p\left(\frac{1}{b_n\cdot n^{1/2}}\right)+O(1)\right)\\ &= b_n\cdot\left(o_p(1)+O(1)\right) = O_p(b_n)\quad\text{uniformly over }\mathcal{F}.\end{aligned} \tag{A11}$$

Next, under our assumptions, $\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\mathcal{B}\left(\widehat{Q}(x,t,\widehat{\theta})\right)\right| = O_p(1)$ uniformly over $\mathcal{F}$. Therefore,

$$\begin{aligned}&\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\right|\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}(Q(X_i,t,\theta_F^*))\right|\geq b_n\right\}\\ &\leq \sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\mathcal{B}\left(\widehat{Q}(x,t,\widehat{\theta})\right)\right|\times\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}(Q(X_i,t,\theta_F^*))\right|\geq b_n\right\}\\ &=O_p(1)\times\left(\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}(Q(X_i,t,\theta_F^*))\right|\geq b_n\right\}\right),\quad\text{uniformly over }\mathcal{F}.\end{aligned}$$

Next, note that

$$\begin{aligned}
&\sup_{F\in\mathcal{F}} P_F\left(\sup_{t\in\mathcal{T}} \frac{1}{n}\sum_{i=1}^{n}\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}\left(Q(X_i,t,\theta_F^*)\right)\right|\geq b_n\right\}\neq 0\right)\\
\leq&\sup_{F\in\mathcal{F}} P_F\left(\sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\mathcal{B}(\widehat{Q}(x,t,\widehat{\theta}))-\mathcal{B}(Q_F(x,t,\theta_F^*))\right|\geq b_n\right)\longrightarrow 0
\end{aligned}$$

where the last result follows from (A6). It follows from here that, for any $\delta>0$ and $\Delta>0$, $\sup_{F\in\mathcal{F}} P_F\left(\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}\left(Q(X_i,t,\theta_F^*)\right)\right|\geq b_n\right\}\geq\frac{\delta}{n^{1/2+\Delta}}\right)\longrightarrow 0$. Immediately, this implies that, $\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}\left(Q(X_i,t,\theta_F^*)\right)\right|\geq b_n\right\}=o_p\left(\frac{1}{n^{1/2+\Delta}}\right)\quad\forall\,\Delta>0$, uniformly over $\mathcal{F}$. Therefore,

$$\sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^{n}\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\right|\phi(X_i)\mathbb{1}\left\{\left|\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)-\mathcal{B}\left(Q(X_i,t,\theta_F^*)\right)\right|\geq b_n\right\}=o_p\left(\frac{1}{n^{1/2+\Delta}}\right)\quad\forall\,\Delta>0, \tag{A12}$$

uniformly over $\mathcal{F}$. Plugging the results in (A11) and (A12) into (A9), for any $\Delta>0$ we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}}\left|\xi^a_{T_0,n}(t)\right|&\leq\left(2b_n+o_p\left(\frac{1}{n^{1/4+\epsilon/2}}\right)\right)\times O_p(b_n)+o_p\left(\frac{1}{n^{1/2+\Delta}}\right)\\
&=O_p\left(b_n^2\right)+o_p\left(\frac{b_n}{n^{1/4+\epsilon/2}}\right)+o_p\left(\frac{1}{n^{1/2+\Delta}}\right)\quad\text{uniformly over }\mathcal{F}.
\end{aligned}$$

Take any $\Delta>0$ and note that $\left(\frac{b_n}{n^{1/4+\epsilon/2}}\right)\cdot n^{1/2+\Delta}=\left(n^{1/2+2\Delta-\epsilon}\cdot b_n^2\right)^{1/2}$. In Assumption 4 we stated that there exists $\delta_0>0$ such that $n^{1/2+\delta_0}\cdot b_n^2\longrightarrow 0$. Therefore, $\frac{b_n}{n^{1/4+\epsilon/2}}=o\left(\frac{1}{n^{1/2+\Delta}}\right)\quad\forall\,0<\Delta\leq\frac{\delta_0}{2}$. From here, we obtain $\sup_{t\in\mathcal{T}}\left|\xi^a_{T_0,n}(t)\right|=o_p\left(\frac{1}{n^{1/2+\delta_0/2}}\right)$ uniformly over $\mathcal{F}$. Therefore, using the linear representation result in (A8),

$$\begin{aligned}
&\widehat{T}_0(t)=\widetilde{T}_{0,F}(t)+\xi^a_{T_0,n}(t)\\
&=\frac{1}{n}\sum_{i=1}^{n}\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\geq 0\right\}\phi(X_i)+\xi^a_{T_0,n}(t)\\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{B}(Q_F(X_i,t,\theta_F^*))+\frac{1}{n}\sum_{j=1}^{n}\psi_F^{\mathcal{B}}(V_j,X_i,t,\theta_F^*,h_n)+\zeta_n^{\mathcal{B}}(X_i,t)\right)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\geq 0\right\}\phi(X_i)+\xi^a_{T_0,n}(t)\\
&=\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\right)_+\phi(X_i)+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_F^{\mathcal{B}}(V_j,X_i,t,\theta_F^*,h_n)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\geq 0\right\}\phi(X_i,t)\\
&+\frac{1}{n}\sum_{i=1}^{n}\zeta_n^{\mathcal{B}}(X_i,t)\mathbb{1}\left\{\mathcal{B}\left(Q_F(X_i,t,\theta_F^*)\right)\geq 0\right\}\phi(X_i,t)+\xi^a_{T_0,n}(t).
\end{aligned}$$

Recall that $T_{0,F}(t) \equiv E_F\left[\left(\mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)\right)_+ \phi(X)\right]$. Thus, from the above expression we have,

$$\begin{aligned}\widehat{T_0}(t) =& T_{0,F}(t) + \frac{1}{n}\sum_{i=1}^{n}\Big((\mathcal{B}(Q_F(X_i,t,\theta_F^*)))_+ \phi(X_i) - T_{0,F}(t)\Big)\\ &+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_F^{\mathcal{B}}(V_j,X_i,t,\theta_F^*,h_n)\mathbb{1}\{\mathcal{B}(Q_F(X_i,t,\theta_F^*))\geq 0\}\phi(X_i)\\ &\underbrace{+ \frac{1}{n}\sum_{i=1}^{n}\zeta_n^{\mathcal{B}}(X_i,t)\mathbb{1}\{\mathcal{B}(Q_F(X_i,t,\theta_F^*))\geq 0\}\phi(X_i) + \xi_{T_0,n}^{a}(t).}_{\equiv \xi_{T_0,n}^{b}(t)}\end{aligned} \tag{A13}$$

From the results in (A8), we have,

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\xi_{T_0,n}^{b}(t)\right| \equiv \left|\frac{1}{n}\sum_{i=1}^{n}\zeta_n^{\mathcal{B}}(X_i,t)\mathbb{1}\{\mathcal{B}(Q_F(X_i,t,\theta_F^*))\geq 0\}\phi(X_i)\right| &\leq \overline{\phi}\cdot \sup_{(x,t)\in\mathcal{X}\times\mathcal{T}}\left|\zeta_n^{\mathcal{B}}(x,t)\right|\\ &= o_p\left(\frac{1}{n^{1/2+\epsilon}}\right) \quad \text{uniformly over } \mathcal{F}\end{aligned} \tag{A14}$$

The proof proceeds from here by 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 (A13). Let $\Xi_{Q_p}(x,t,\theta_F^*)$ be as defined in equation (A3) and define,

$$\Xi_{T_0,F}^{p}(t) \equiv E_F\left[\frac{\partial \mathcal{B}\left(Q_F(X,t,\theta_F^*)\right)}{\partial Q_p}\mathbb{1}\{\mathcal{B}(Q_F(X,t,\theta_F^*))\geq 0\}\phi(X)\Xi_{Q_p}(X,t,\theta_F^*)\right], \quad \Xi_{T_0,F}(t) \equiv \sum_{p=1}^{P}\Xi_{T_0,F}^{p}(t) \tag{A15}$$

Let $\Omega_{T_0}^{p}(y,t,g)$ be as defined in Assumption 7 (eq. (18)). In the Econometric Supplement we show,

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_F^{\mathcal{B}}(V_j,X_i,t,\theta_F^*,h_n)\mathbb{1}\{\mathcal{B}(Q_F(X_i,t,\theta_F^*))\geq 0\}\phi(X_i)\\ &=\frac{1}{n}\sum_{i=1}^{n}\left(\left(\sum_{p=1}^{P}\Omega_{T_0}^{p}(Y_i,t,g(X_i,\theta_F^*))\cdot\omega_p(g(X_i,\theta_F^*))\right) + \Xi_{T_0,F}(t)\psi_F^{\theta}(Z_i)\right) + \xi_{T_0,n}^{c}(t),\end{aligned} \tag{A16}$$

where $\sup_{t\in\mathcal{T}}\left|\xi_{T_0,n}^{c}(t)\right| = o_p\left(\frac{1}{n^{1/2+\epsilon}}\right)$ uniformly over $\mathcal{F}$, with $\epsilon > 0$ as described in Assumption 4. Let

$$\psi_F^{T_0}(V_i,t) \equiv \Big((\mathcal{B}(Q_F(X_i,t,\theta_F^*)))_+ \phi(X_i) - T_{0,F}(t)\Big) + \sum_{p=1}^{P}\Omega_{T_0}^{p}(Y_i,t,g(X_i,\theta_F^*))\cdot\omega_p(g(X_i,\theta_F^*)) + \Xi_{T_0,F}(t)\psi_F^{\theta}(Z_i). \tag{A17}$$

By inspection, we can verify that $\psi_F^{T_0}(V,t)$ has two key features,

$$\begin{aligned}&(i) \quad E_F\left[\psi_F^{T_0}(V,t)\right]=0 \quad \forall\, t\in\mathcal{T},\ \forall\, F\in\mathcal{F},\\&(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*))<0 \,\middle|\, X\in\mathcal{X}_F^*\right)=1 \implies P_F\left(\psi_F^{T_0}(V,t)=0\right)=1.\end{aligned} \tag{A18}$$

Let $\Delta\equiv\epsilon\wedge(\delta_0/2)$. Plugging (A16) and (A14) into (A13), we have

$$\widehat{T_0}(t)=T_{0,F}(t)+\frac{1}{n}\sum_{i=1}^{n}\psi_F^{T_0}(V_i,t)+\varepsilon_n^{T_0}(t), \quad \text{where} \quad \sup_{t\in\mathcal{T}}\left|\varepsilon_n^{T_0}(t)\right|=o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \quad \text{uniformly over } \mathcal{F}. \tag{A19}$$

Since $\widehat{T_2}\equiv\int_t\widehat{T_0}(t)d\mathcal{W}(t)$ and $T_{2,F}\equiv\int_t T_{0,F}(t)d\mathcal{W}(t)$ (with the normalization $\int_{t\in\mathcal{T}}d\mathcal{W}(t)=1$), the result in (A19) yields,

$$\begin{aligned}&\widehat{T_2}=T_{2,F}+\frac{1}{n}\sum_{i=1}^{n}\psi_F^{T_2}(V_i)+\varepsilon_n^{T_2}, \quad \text{where} \quad \psi_F^{T_2}(V_i)\equiv\int_{t\in\mathcal{T}}\psi_F^{T_0}(V_i,t)d\mathcal{W}(t), \quad \varepsilon_n^{T_2}\equiv\int_{t\in\mathcal{T}}\varepsilon_n^{T_0}(t)d\mathcal{W}(t),\\&\text{and} \quad \left|\varepsilon_n^{T_2}\right|\le\sup_{t\in\mathcal{T}}\left|\varepsilon_n^{T_0}(t)\right|=o_p\left(\frac{1}{n^{1/2+\Delta}}\right) \quad \text{uniformly over } \mathcal{F}.\end{aligned} \tag{A20}$$

This is the result in Proposition 1. ■

A1.3 Properties of the influence function $\psi_F^{T_2}(V)$

From (A18) and (A20), we obtain the following properties for the influence function $\psi_F^{T_2}(V)$,

$$\begin{aligned}&(i) \quad E_F\left[\psi_F^{T_2}(V)\right]=0 \quad \forall\, F\in\mathcal{F},\\&(ii) \quad P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*))<0 \,\middle|\, X\in\mathcal{X}_F^*\right)=1 \ \text{ for } \mathcal{W}\text{–a.e } t\in\mathcal{T} \implies P_F\left(\psi_F^{T_2}(V)=0\right)=1.\end{aligned} \tag{A21}$$

Part (i) of (A21) follows directly from part (i) of (A18) since,

$$E_F\left[\psi_F^{T_2}(V)\right]=E_F\left[\int_t\psi_F^{T_0}(V,t)d\mathcal{W}(t)\right]=\int_t\underbrace{\left(E_F\left[\psi_F^{T_0}(V,t)\right]\right)}_{=0\ \forall\, t\in\mathcal{T}}d\mathcal{W}(t)=0.$$

Similarly, part (ii) of (A21) follows directly from part (ii) of (A18) since,

$$P_F\left(\mathcal{B}(\Gamma_F(X,t,\theta_F^*))<0 \,\middle|\, X\in\mathcal{X}_F^*\right)=1 \ \text{ for } \mathcal{W}\text{–a.e } t\in\mathcal{T} \implies P_F\left(\psi_F^{T_0}(V,t)=0 \ \text{ for } \mathcal{W}\text{–a.e } t\in\mathcal{T}\right)=1,$$

$$\text{and} \quad P_F\left(\psi_F^{T_0}(V,t)=0 \ \text{ for } \mathcal{W}\text{–a.e } t\in\mathcal{T}\right)=1 \implies P_F\left(\psi_F^{T_2}(V)=0\right)=P_F\left(\int_t\psi_F^{T_0}(V,t)d\mathcal{W}(t)=0\right)=1.$$

The two properties in (A21) **are those described in the statement of Proposition 1.** ■

A2 An estimator for the influence function $\psi_F^{T_2}(V)$

Our test relies on an estimator for $\sigma^2_{2,F} \equiv \text{Var}[\psi_F^{T_2}(V)]$. To this end, we construct an estimator for $\psi_F^{T_2}(V)$ based on its characterization in equations (A17)- (A20). As before, let $\widehat{f_g}(g) \equiv \frac{1}{nh_n^D}\sum_{i=1}^n K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right)$. Our estimators for the functionals described in Assumption 2 are,

$$\widehat{\Omega}^{d,\ell}_{f_g}(g) \equiv \frac{\widehat{A}^{d,\ell}_{f_g}(g)}{\widehat{f_g}(g)}, \quad \text{where} \quad \widehat{A}^{d,\ell}_{f_g}(g) \equiv \frac{1}{nh_n^D}\sum_{i=1}^n \frac{\partial g_d(X_i,\widehat{\theta})}{\partial\theta_\ell} K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right),$$
$$\widehat{\Omega}^{d,\ell}_{R_p,1}(g,t) \equiv \frac{\widehat{A}^{d,\ell}_{R_p,1}(g,t)}{\widehat{f_g}(g)}, \quad \text{where} \quad \widehat{A}^{d,\ell}_{R_p,1}(g,t) \equiv \frac{1}{nh_n^D}\sum_{i=1}^n S_p(Y_i,t)\omega_p(g(X_i,\widehat{\theta}))\frac{\partial g_d(X_i,\widehat{\theta})}{\partial\theta_\ell} K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right),$$
$$\widehat{\Omega}_{R_p,2}(g,t) \equiv \frac{\widehat{A}_{R_p,2}(g,t)}{\widehat{f_g}(g)}, \quad \text{where} \quad \widehat{A}_{R_p,2}(g,t) \equiv \frac{1}{nh_n^D}\sum_{i=1}^n S_p(Y_i,t)\omega_p(g(X_i,\widehat{\theta})) K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right),$$
$$\widehat{\Omega}^{\ell}_{R_p,3}(g,t) \equiv \frac{\widehat{A}^{\ell}_{R_p,3}(g,t)}{\widehat{f_g}(g)}, \quad \text{where} \quad \widehat{A}^{\ell}_{R_p,3}(g,t) \equiv \frac{1}{nh_n^D}\sum_{i=1}^n S_p(Y_i,t)\frac{\partial\omega_p(g(X_i,\widehat{\theta}))}{\partial\theta_\ell} K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right),$$

From the above definitions, we have $\frac{\partial\widehat{A}^{d,\ell}_{f_g}(g)}{\partial g_d} = -\frac{1}{nh_n^{D+1}}\sum_{i=1}^n \frac{\partial g_d(X_i,\widehat{\theta})}{\partial\theta_\ell}\frac{\partial K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right)}{\partial g_d}$,

$$\frac{\partial\widehat{A}^{d,\ell}_{R_p,1}(g,t)}{\partial g_d} = -\frac{1}{nh_n^{D+1}}\sum_{i=1}^n S_p(Y_i,t)\omega_p(g(X_i,\widehat{\theta}))\frac{\partial g_d(X_i,\widehat{\theta})}{\partial\theta_\ell}\frac{\partial K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right)}{\partial g_d}, \quad \text{and}$$
$$\frac{\partial\widehat{A}_{R_p,2}(g,t)}{\partial g_d} = -\frac{1}{nh_n^{D+1}}\sum_{i=1}^n S_p(Y_i,t)\omega_p(g(X_i,\widehat{\theta}))\frac{\partial K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right)}{\partial g_d}.$$

Let $\Xi_{f_g}(x,\theta^*_F)$, $\Xi_{R_p}(x,t,\theta^*_F)$ and $\Xi_{Q_p}(x,t,\theta^*_F)$ be the functionals defined in equations (A1) and (A3). Our estimators for these functionals are,

$$\widehat{\Xi}_{\ell,f_g}(x,\widehat{\theta}) \equiv \sum_{d=1}^D\left(\frac{\partial g_d(x,\widehat{\theta})}{\partial\theta_\ell}\cdot\frac{\partial\widehat{f_g}(g(x,\widehat{\theta}))}{\partial g_d} - \frac{\partial\widehat{A}^{d,\ell}_{f_g}(g(x,\widehat{\theta})}{\partial g_d}\right),$$
$$\widehat{\Xi}_{f_g}(x,\widehat{\theta}) \equiv \left(\widehat{\Xi}_{1,f_g}(x,\widehat{\theta}),\ldots,\widehat{\Xi}_{k,f_g}(x,\widehat{\theta})\right),$$
$$\widehat{\Xi}_{\ell,R_p}(x,t,\widehat{\theta}) \equiv \sum_{d=1}^D\left(\frac{\partial\widehat{A}_{R_p,2}(g(x,\widehat{\theta}),t)}{\partial g_d}\cdot\frac{\partial g_d(x,\widehat{\theta})}{\partial\theta_\ell} - \frac{\partial\widehat{A}^{d,\ell}_{R_p,1}(g(x,\widehat{\theta}),t)}{\partial g_d} + \widehat{A}^{\ell}_{R_p,3}(g(x,\widehat{\theta}),t)\right),$$
$$\widehat{\Xi}_{R_p}(x,t,\widehat{\theta}) \equiv \left(\widehat{\Xi}_{1,R_p}(x,t,\widehat{\theta}),\ldots,\widehat{\Xi}_{k,R_p}(x,t,\widehat{\theta})\right),$$
$$\widehat{\Xi}_{Q_p}(x,t,\widehat{\theta}) \equiv \frac{\widehat{\Xi}_{R_p}(x,t,\widehat{\theta}) - \widehat{Q}_p(x,t,\widehat{\theta})\cdot\widehat{\Xi}_{f_g}(x,\widehat{\theta})}{\widehat{f_g}(g(x,\widehat{\theta}))},$$

For a given t we estimate the functional $\Xi_{T_0,F}(t)$ described in Proposition 1, equation (A15) with

$$\widehat{\Xi}^p_{T_0}(t) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)}{\partial \widehat{Q}_p} \mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}\phi(X_i,t)\widehat{\Xi}_{Q_p}(X_i,t,\widehat{\theta}), \quad \text{and} \quad \widehat{\Xi}_{T_0}(t) \equiv \sum_{p=1}^{P} \widehat{\Xi}^p_{T_0}(t).$$

Next, for a given x, t, we estimate $\widehat{\Gamma}_p(x,t,\widehat{\theta}) = \frac{\frac{1}{n\cdot h_n^D}\sum_{i=1}^n S_p(Y_i,t)K\left(\frac{\Delta g(X_i,x,\widehat{\theta})}{h_n}\right)}{\widehat{f}_g(g(x,\widehat{\theta}))}$. For a given (y,g,t), we estimate the functional $\Omega^p_{T_0}(y,t,g)$ described in Assumption 7, equation (18) with

$$\widehat{\Omega}^p_{T_0}(y,t,g) = \frac{\frac{1}{n\cdot h_n^D}\sum_{i=1}^n \left(S_p(y,t)-\widehat{\Gamma}_p(X_i,t,\widehat{\theta})\right)\frac{\partial \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)}{\partial Q_p}\phi(X_i)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\}K\left(\frac{g(X_i,\widehat{\theta})-g}{h_n}\right)}{\widehat{f}_g(g)}$$

Based on (A17)- (A20), our estimator for the influence function $\psi_F^{T_2}(V)$ is,

$$\begin{aligned}\widehat{\psi}^{T_2}(V_i) \equiv & \left(\int_t \mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_n\right\} d\mathcal{W}(t)\phi(X_i) - \widehat{T}_2\right) \\ & + \sum_{p=1}^{P}\int_t \widehat{\Omega}^p_{T_0}(Y_i,t,g(X_i,\widehat{\theta}))d\mathcal{W}(t)\cdot\omega(g(X_i,\widehat{\theta})) + \int_t \widehat{\Xi}_{T_0}(t)d\mathcal{W}(t)\widehat{\psi}^{\theta}(Z_i)\end{aligned} \tag{A22}$$

A2.1 Estimator for $\sigma^2_{2,F}$

Our estimator for $\sigma^2_{2,F}$ is $\widehat{\sigma}^2_2 \equiv \frac{1}{n}\sum_{i=1}^n \widehat{\psi}^{T_2}(V_i)^2$. As we show in the Econometric Supplement, under the conditions in Assumptions 1-6, 7 and 8,

$$\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{\psi}^{T_2}(V_i) - \psi_F^{T_2}(V_i)\right|^2 = o_p(1) \text{ uniformly over } \mathcal{F}. \tag{A23}$$

Next, under the conditions of Proposition 1, there exists a finite $\overline{\sigma}_2 > 0$ such that $E_F\left[\psi_F^{T_2}(V)^2\right] \leq \overline{\sigma}_2$ for all $F \in \mathcal{F}$. From here, a Chebyshev inequality yields,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\psi_F^{T_2}(V_i)^2 - \sigma^2_{2,F}\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{A24}$$

(A23) and (A24) yield,

$$\left|\widehat{\sigma}^2_2 - \sigma^2_{2,F}\right| = o_p(1), \quad \text{uniformly over } \mathcal{F}. \tag{A25}$$

This is the result stated in equation (23) in the paper.

A3 Bandwidths that satisfy Assumption 4

We focus on bandwidth sequences of the form $h_n \propto n^{-\alpha_h}$ and $b_n \propto n^{-\alpha_b}$. Since b_n plays an important role in our problem, we focus on the fastest rates of convergence for it that is compatible with Assumption 4. Let $\alpha_b = \frac{1}{4} + \Delta_b$. To satisfy the restrictions $n^{1/2+\epsilon'} \cdot b_n^2 \longrightarrow 0$ and $n^{1/2} \cdot h_n^{D+1} \cdot b_n \longrightarrow \infty$, we must have $\frac{\epsilon'}{2} < \Delta_b < \frac{1}{4} - (D+1) \cdot \alpha_h$. Thus, we must have $\alpha_h < \frac{1}{4(D+1)} - \frac{\epsilon'}{2(D+1)}$. Accordingly, let $\alpha_h = \frac{1}{4(D+1)} - \Delta_h$, where $\Delta_h > \frac{\epsilon'}{2(D+1)}$. Thus, we can express $\Delta_h = \frac{\epsilon'+\delta'}{2(D+1)}$ with $\delta' > 0$. Set[8] $\epsilon' < \left(\frac{1}{2D(D+1)} \wedge \frac{1}{8}\right)$. Then, having $\alpha_h = \frac{1}{4(D+1)} - \Delta_h$ automatically satisfies the condition $n^{1/2-\epsilon'} \cdot (h_n^{2D} \wedge h_n^{D+2}) \longrightarrow \infty$. The last bandwidth convergence restriction we need to satisfy is $n^{1/2+\epsilon'} \cdot h_n^M \longrightarrow 0$. We are interested in the smallest integer M that can satisfy this condition given the restrictions on α_h. The condition will be satisfied if and only if $M \cdot \left(\frac{1-4(D+1)\Delta_h}{4(D+1)}\right) > \frac{1}{2} + \epsilon'$. Since $\Delta_h = \frac{\epsilon'+\delta'}{2(D+1)}$, this becomes $M > \left(\frac{4(D+1)}{1-2(\epsilon'+\delta')}\right) \cdot \left(\frac{1}{2} + \epsilon'\right)$. Note that $\left(\frac{4(D+1)}{1-2(\epsilon'+\delta')}\right) \cdot \left(\frac{1}{2} + \epsilon'\right) > 2(D+1) = 2D+2$, so the smallest integer that can be greater than the right-hand side of the previous expression is $2D+3$. Accordingly, choose $\epsilon' > 0$ and $\delta' > 0$ small enough that $\left(\frac{4(D+1)}{1-2(\epsilon'+\delta')}\right) \cdot \left(\frac{1}{2} + \epsilon'\right) < 2D+3$. Then, $M = 2D+3$ satisfies our restriction and this is the smallest integer that can do so. In summary, the bandwidth convergence rates would be of the form $\alpha_h = \frac{1}{4(D+1)} - \frac{\epsilon'+\delta'}{2(D+1)}$ 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 such that $\epsilon' < \left(\frac{1}{2D(D+1)} \wedge \frac{1}{8}\right)$ and $\left(\frac{4(D+1)}{1-2(\epsilon'+\delta')}\right) \cdot \left(\frac{1}{2} + \epsilon'\right) < 2D+3$. 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 $M = 2D+3$.

A4 Extensions

A4.1 Allowing $\kappa_{2,n} \longrightarrow 0$

Assumption 9 allows for $\sigma_{2,F}$ to be arbitrarily close to zero over $\mathcal{F} \setminus \overline{\mathcal{F}}$. Suppose we rule this out and we maintain the following (stronger) version of Assumption 9.

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

The above states that, for any $F \in \mathcal{F}$, either $\sigma_{2,F} = 0$, or $\sigma_{2,F}$ is bounded away from zero by some $\underline{\sigma} > 0$. Let $\kappa_{2,n}$ be any positive sequence such that $\kappa_{2,n} \to 0$, $n^{\Delta} \cdot \kappa_{2,n} \to \infty$, where $\Delta > 0$ is the constant described in Proposition 1. Under Assumption 9′ we can replace the constant κ_2 with the sequence $\kappa_{2,n} \to 0$. Our modified test-statistic would then be $\widetilde{t}_2 = \frac{\sqrt{n}\widehat{T}_2}{(\widehat{\sigma}_2 \vee \kappa_{2,n})}$. Going back to (25), note that $\left|\frac{\sqrt{n} \cdot \varepsilon_n^{T_2}}{(\widehat{\sigma}_2 \vee \kappa_{2,n})}\right| \leq \left|\frac{\sqrt{n} \cdot \varepsilon_n^{T_2}}{\kappa_{2,n}}\right| = o_p\left(\frac{n^{1/2}}{n^{1/2+\Delta} \cdot \kappa_{2,n}}\right) = o_p\left(\frac{1}{n^{\Delta} \cdot \kappa_{2,n}}\right) = o_p(1)$, uniformly over $\mathcal{F}$. Therefore, the results in (26) are preserved for $\widetilde{t}_2$. The stronger conditions in Assumption 9′ imply that

[8] We will have $\frac{1}{2D(D+1)} \wedge \frac{1}{8} = \frac{1}{2D(D+1)}$ and $h_n^{2D} \wedge h_n^{D+2} = h_n^{2D}$ for all $D \geq 2$.

$\sup_{F\in\mathcal{F}\backslash\overline{\mathcal{F}}}\left|\frac{\sigma_{2,F}\vee\kappa_{2,n}}{\sigma_{2,F}}-1\right|\to 0$. Thus, by the Berry-Esseen Theorem and the conditions in Assumption 9',

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

This strengthens the result in (29). These results combined yield,

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

This is a stronger version of the result we showed in (30). Under Assumption 9', replacing $\widehat{t_2}$ with $\widetilde{t_2}$, the result in part (i) of Theorem 1 (uniformly asymptotically level α) strengthens to,

$$\lim_{n\to\infty}\sup_{F\in\mathcal{F}^0\backslash\overline{\mathcal{F}}}\left|P_F\left(\widetilde{t}_2\geq z_{1-\alpha}\right)-\alpha\right|=0.$$

Part (ii) of Theorem 1 (consistency) remains unchanged, and part (iii) (nontrivial asymptotic power) is preserved after we replace κ_2 with $\kappa_{2,n}$, redefining the limits s_2^a and s_2^b accordingly.

A4.2 Testing multiple inequalities

Our approach can be readily extended to testing multiple functional inequalities of the form (5). Suppose we have a model that predicts,

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

Suppose each $\mathcal{B}_r$ satisfies the condition in (7) for weight functions $(\omega_{p,r})_{p=1}^P$, so, for each r,

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

Assume for simplicity 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 r. As in (6), let $Q_{p,F}^r(x,t,\theta_F^*)\equiv\Gamma_{p,F}(x,t,\theta_F^*)\cdot\omega_{p,r}(g(x,\theta_F^*))$, $Q_F^r(x,t,\theta_F^*)\equiv\left(Q_{1,F}^r(x,t,\theta_F^*),\ldots,Q_{P,F}^r(x,t,\theta_F^*)\right)'$, and $\omega_r(g(x,\theta_F^*)\equiv\left(\omega_{1,r}(g(x,\theta_F^*)),\ldots,\omega_{P,r}(g(x,\theta_F^*))\right)'$. Then (8) holds for each $r=1,\ldots,\mathcal{R}$, and we have, $\mathcal{B}_r\left(Q_F^r(x,t,\theta_F^*)\right)=\mathcal{B}_r\left(\Gamma_F(x,t,\theta_F^*)\right)\cdot\mathcal{H}_r\left(\omega_r(g(x,\theta_F^*))\right)$, where $\mathcal{H}_r(\cdot)\geq 0$, and $\mathcal{H}_r(\omega_r(g(x,\theta_F^*))>0$ iff $g^r(x,\theta_F^*)\in\mathcal{G}$. For each r, let $\phi_r(x)$ be a function satisfying $\phi_r(x)\geq 0$ for all x and $\phi_r(x)>0$ if and only if $x\in\mathcal{X}$. For a given $t\in\mathcal{T}$ let $T_{0,F}^r(t)\equiv E_{F_X}\left[\left(\mathcal{B}_r\left(Q_F^r(X,t,\theta_F^*)\right)\right)_+\phi_r(X)\right]$. Let $d\mathcal{W}^r$ be a pre-specified weight function for the index parameter t for the r^{th} restriction. One way to generalize

$T_{2,F}$ is by aggregating the $\mathcal{R}$ restrictions as[9],

$$T_{2,F} \equiv \sum_{r=1}^{R}\left[\int_t T_{0,F}^r(t)d\mathcal{W}^r(t)\right] \tag{A26}$$

As in the single functional-inequality case, $T_{2,F}$ has the following key properties,

- $T_{2,F} \geq 0$.
- $T_{2,F} = 0 \Longleftrightarrow \mathcal{B}_r(\Gamma_F(x,t,\theta_F^*)) \leq 0$ for $F-$a.e $x \in \mathcal{X}_F^*$, and $\mathcal{W}-$a.e $t \in \mathcal{T}$ for each r.

For each r, construct the estimator $\widehat{Q}_p^r(x,t,\widehat{\theta})$ in the manner described in equation (13), and group $\widehat{Q}^r(x,t,\widehat{\theta}) \equiv \left(\widehat{Q}_1^r(x,t,\widehat{\theta}),\ldots,\widehat{Q}_P^r(x,t,\widehat{\theta})\right)'$. We can estimate $T_{2,F}$ with,

$$\widehat{T}_2 \equiv \sum_{r=1}^{R}\left[\int_t\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{B}_r\left(\widehat{Q}^r(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}_r\left(\widehat{Q}^r(X_i,t,\widehat{\theta})\right) \geq -b_{r,n}\right\}\phi(X_i)\right)d\mathcal{W}^r(t)\right] \equiv \sum_{r=1}^{R}\left[\int_t \widehat{T}_0^r(t)d\mathcal{W}^r(t)\right],$$
$$\text{where} \quad \widehat{T}_0^r(t) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathcal{B}_r\left(\widehat{Q}^r(X_i,t,\widehat{\theta})\right)\mathbb{1}\left\{\mathcal{B}\left(\widehat{Q}(X_i,t,\widehat{\theta})\right) \geq -b_{r,n}\right\}\phi(X_i). \tag{A27}$$

Each $b_{r,n}$ satisfies the restrictions in Assumption 4. If the assumptions leading to Proposition 1 are satisfied by each of the $\mathcal{R}$ restrictions, then each of the summands in (A27) will satisfy the linear representation result in Proposition 1. From here, we would have

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

$\Delta > 0$ is a constant with the features described in Proposition 1. $\psi_F^{T_2}(V_i)$ corresponds to the sum of the influence functions for each of the $\mathcal{R}$ restrictions, each one with the structure described in equations (A17)-(A20). From here, the results in Proposition 1 would yield,

$$\begin{aligned}
&(i) \quad E_F\left[\psi_F^{T_2}(V)\right] = 0 \quad \forall\, F \in \mathcal{F},\\
&(ii) \quad P_F\left(\mathcal{B}_r\left(\Gamma_F(X,t,\theta_F^*)\right) < 0 \;\middle|\; X \in \mathcal{X}_F^*\right) = 1 \text{ for } \mathcal{W}\text{−a.e } t \in \mathcal{T}, \forall\, r = 1,\ldots,\mathcal{R} \implies P_F\left(\psi_F^{T_2}(V) = 0\right) = 1.
\end{aligned} \tag{A28}$$

[9]We can aggregate using different weights for each of the $r = 1,\ldots,R$ restrictions,

$$T_{2,F} = \sum_{r=1}^{\mathcal{R}}\left[\int_t T_{0,F}^r(t)d\mathcal{W}^r(t)\right]\cdot\gamma_r,$$

where γ_r is the weight given to the r^{th} restriction. For simplicity, (A26) considers the uniform-weight case.

Our test-statistic would once again be of the form $\widehat{t_2} = \frac{\sqrt{n}\cdot\widehat{T}_2}{(\widehat{\sigma_2} \vee \kappa_2)}$, as in (24). From here, if we maintain the integrability condition in Assumption 9, the test that rejects H_0 iff $\widehat{t_2} \geq z_{1-\alpha}$ will have the asymptotic properties described in Theorem 1.

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.

Fan, W. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test. *Econometric Reviews 14*, 367–382.

Fan, W. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory 14*, 604–621.

Fan, Y. and Q. Li (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* (4), 865–890.

Hardle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics 21*, 1926–1947.

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.

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.

Zheng, J. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics 75*, 263–289.