Econometric Theory, **0**, 2026, 1–28.
doi:10.1017/S0266466626100528

UNIFIED INFERENCE ON MOMENT RESTRICTIONS WITH NUISANCE PARAMETERS

XINGYU LI
Zhejiang University

XIAOJUN SONG
Peking University

ZHENTING SUN
University of Melbourne

This article proposes a simple unified inference approach on moment restrictions in the presence of nuisance parameters. The proposed test is constructed based on a new characterization that avoids the estimation of nuisance parameters and can be broadly applied across diverse settings. Under suitable conditions, the test is shown to be asymptotically size controlled and consistent for both independent and dependent samples. Monte Carlo simulations show that the test performs well in finite samples. Numerical results from the application to conditional moment restriction models with weak instruments demonstrate that the proposed method may improve upon existing approaches in the literature.

1. INTRODUCTION

The analysis of moment restriction models plays a central role in econometric theory and applications. Considerable efforts have been devoted to estimating unknown key parameters and to testing hypotheses related to these parameters in the moment restrictions (see, e.g., Chamberlain, 1987, Newey, 1993, Domínguez and Lobato, 2004, Kitamura, Tripathi, and Ahn, 2004, Smith, 2007, and Lavergne and Patilea, 2013, among many others; see also Kunitomo, McAleer, and Nishiyama, 2011, for an overview of the moment restriction-based econometric methods).

Valid statistical inference on these parameters relies crucially on the correct specification of the postulated moment restriction models. Assessing the suitability of the moment restrictions has therefore generated an extensive literature (see, e.g., Bierens, 1982; Newey, 1985; Tauchen, 1985; Donald, Imbens, and Newey, 2003).

This article is a revised version of Chapters 2 and 3 of the first author's doctoral dissertation at Peking University. The authors are grateful to the Editors, two anonymous referees, Zheng Fang and Yixiao Sun, and all seminar and conference participants for their insightful and constructive comments. Address correspondence to Zhenting Sun, Department of Economics, University of Melbourne, Australia, e-mail: zhenting.sun@unimelb.edu.au

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

In testing the moment restrictions, the unknown parameters may not be of primary interest under the null hypothesis and can be regarded as nuisance parameters. Handling nuisance parameters in the considered testing procedures is an important theoretical issue. Existing specification tests for moment restrictions typically employ procedures that first estimate the nuisance parameters and then test the moment restrictions using the estimators (see, e.g., Tripathi and Kitamura, 2003; Delgado, Domínguez, and Lavergne, 2006; Muandet, Jitkrittum, and Kübler, 2020). As a result, classical approaches are generally model- or estimator-dependent, requiring different theories and implementation procedures for different cases. In addition, these approaches may encounter theoretical difficulties due to the estimation of nuisance parameters. For example, obtaining reliable estimates of nuisance parameters may be nontrivial in conditional moment restriction models when instruments are weak.

In this article, we propose a unified testing framework for moment restrictions with nuisance parameters that is broadly applicable to various settings. The critical values of our test are constructed using the numerical delta methods developed by Hong and Li (2018) and Chen and Fang (2019b) who provide novel methodologies for addressing nonstandard testing issues.[1] The proposed method in the article effectively circumvents the estimation of nuisance parameters, thus providing a general and robust inferential tool for different settings where nuisance parameters are present. A comparison between the proposed test and existing approaches in conditional moment restriction models with weak instruments demonstrates that the test can achieve performance improvement.

We summarize the main features of the proposed test as follows: (i) it is case-independent; (ii) it is free of the estimation of nuisance parameters and is particularly appealing in cases where desirable estimation is challenging; (iii) it is asymptotically size controlled and consistent against a broad class of alternatives to the null; (iv) it works for both independent and dependent samples; and (v) the bootstrap test procedure is simple.

Now we introduce our testing framework. Let $d_\theta \in \mathbb{Z}_+$ and $d_z \in \mathbb{Z}_+$. Let $\Theta \subset \mathbb{R}^{d_\theta}$ be a parameter space. Let $\Psi = \{\psi_{x,\theta} : \mathbb{R}^{d_z} \to \mathbb{R} : x \in \mathbb{R}, \theta \in \Theta\}$ be a class of functions indexed by $(x,\theta) \in \mathbb{R} \times \Theta$ such that $\psi_{x,\theta}$ is measurable for all $(x,\theta) \in \mathbb{R} \times \Theta$. Throughout the article, all random elements are defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let P be an unknown probability distribution on $(\mathbb{R}^{d_z}, \mathscr{B}(\mathbb{R}^{d_z}))$ and $Z \sim P$ be a random vector such that for every Borel set $B \subset \mathbb{R}^{d_z}$, $P(B) = \mathbb{P}(Z \in B)$. We are interested in the null hypothesis

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, \mathbb{E}_P[\psi_{x,\theta}(Z)] = 0 \text{ for all } x \in \mathbb{R}. \tag{1}$$

[1] More discussions on this topic can be found in Dümbgen (1993), Andrews (2000), Hirano and Porter (2012), Hansen (2017), and Fang and Santos (2019). Other discussions and applications of related bootstrap methods can be found in Beare and Moon (2015), Beare and Fang (2017), Seo (2018), Beare and Shi (2019), Chen and Fang (2019a), Hong and Li (2020), Sun and Beare (2021), and Sun (2023).

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

This can be viewed as a specification test on a set of moment restrictions. The parameter θ in (1) is the nuisance parameter we need to take into account. Let $\phi_P : \mathbb{R} \times \Theta \to \mathbb{R}$ be a function depending on P such that $\phi_P(x,\theta) = P(\psi_{x,\theta}) = \mathbb{E}_P[\psi_{x,\theta}(Z)]$ for every $(x,\theta) \in \mathbb{R} \times \Theta$. Clearly, the null hypothesis in (1) is equivalent to

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, \phi_P(x,\theta) = 0 \text{ for all } x \in \mathbb{R}. \tag{2}$$

The above formulation can easily be extended to cases where $x \in \mathbb{R}^k$ for some $k > 1$. To simplify exposition, we present the results for scalar x in the main text.

The testing approach provided in the article can be readily applied in a wide range of empirical studies. In the following, we present several important examples where the hypothesis of interest can be formulated into (2).

1.1. Examples

Example 1.1 (Conditional Moment Restrictions). Let $Z = (X, Y)$ be a d_z-dimensional random vector with scalar X and d_y-dimensional vector Y, where $d_z = d_y + 1 \geq 2$. Let $g : \mathbb{R}^{d_y} \times \Theta \to \mathbb{R}$ be a known function. The null hypothesis of interest is

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, \mathbb{E}_P[g(Y,\theta)|X] = 0 \text{ almost surely.}$$

This null hypothesis is equivalent to

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, \mathbb{E}_P[g(Y,\theta)\mathbb{1}\{X \leq x\}] = 0 \text{ for all } x \in \mathbb{R}.$$

In this case, $\psi_{x,\theta}(z) = g(y,\theta)\mathbb{1}\{w \leq x\}$ for every $z = (w,y) \in \mathbb{R} \times \mathbb{R}^{d_y}$ and every $(x,\theta) \in \mathbb{R} \times \Theta$, and $\phi_P(x,\theta) = \mathbb{E}_P[g(Y,\theta)\mathbb{1}\{X \leq x\}]$ for every $(x,\theta) \in \mathbb{R} \times \Theta$. Tripathi and Kitamura (2003) construct a smoothed empirical likelihood-based test for the conditional moment restrictions, Escanciano and Goh (2014) use a projected empirical process to eliminate the estimation effect of nuisance parameters, Domínguez and Lobato (2015) introduce an omnibus test statistic as the minimized value of the objective function considered in Domínguez and Lobato (2004), and Berger (2022) proposes a new empirical likelihood test for parameters of conditional moment restriction models.

Jun and Pinkse (2009) propose semi-parametric tests of conditional moment restrictions with weak instruments. The null rejection probabilities of their tests are shown to be asymptotically no greater than the nominal significance level, suggesting possible conservativeness. Under suitable conditions, the test proposed in this article has an exact asymptotic size, which allows for dependent data as well. The performance improvement of our method over existing approaches is illustrated through simulation studies in Section 4.1, where the data-generating processes (DGPs) are tailored to conditional moment restriction models with weak instruments.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Example 1.2 (Symmetry). Let G be the cumulative distribution function of the random variable Z. The null hypothesis of symmetry about center θ is

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, G(x) = 1 - G(2\theta - x) \text{ for all } x \in \mathbb{R}.$$

In this case, $\psi_{x,\theta}(z) = \mathbb{1}\{z \le x\} + \mathbb{1}\{z \le 2\theta - x\} - 1$ for every $z \in \mathbb{R}$ and every $(x, \theta) \in \mathbb{R} \times \Theta$, and $\phi_P(x, \theta) = G(x) + G(2\theta - x) - 1$ for every $(x, \theta) \in \mathbb{R} \times \Theta$. Psaradakis and Vávra (2015) use a quantile-based measure of skewness to test symmetry about an unspecified center, Psaradakis (2016) considers the autoregressive sieve bootstrap to obtain critical values for tests of symmetry, and Psaradakis and Vávra (2022) employ a U-statistic involving triples of observations to assess symmetry. Psaradakis and Vávra (2019) provide an overview of symmetry tests.

Example 1.3 (Goodness of Fit). Let G be the cumulative distribution function of the random variable Z. Suppose there is a given class of distribution functions $\{G_0(\cdot, \theta) : \theta \in \Theta\}$ so that $x \mapsto G_0(x, \theta)$ is a distribution function on $\mathbb{R}$ for every $\theta \in \Theta$. We assume the identifiability of θ in the sense that for all $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, there exists $x_0 \in \mathbb{R}$ such that $G_0(x_0, \theta_1) \neq G_0(x_0, \theta_2)$. The null hypothesis of correct specification is

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, G(x) = G_0(x, \theta) \text{ for all } x \in \mathbb{R}.$$

In this case, $\psi_{x,\theta}(z) = \mathbb{1}\{z \le x\} - G_0(x, \theta)$ for every $z \in \mathbb{R}$ and every $(x, \theta) \in \mathbb{R} \times \Theta$, and $\phi_P(x, \theta) = G(x) - G_0(x, \theta)$. Goodness-of-fit tests based on parametric empirical processes have been extensively studied since Durbin (1973). For example, the martingale approach proposed by Khmaladze (1982) is applied to the problem of testing goodness of fit with estimated parameters, and Genest and Rémillard (2008) consider goodness-of-fit tests using a parametric bootstrap approach. A more recent work is Parker (2013), which recommends conducting sup-norm inference for tests based on Durbin's (1985) approximations.

Example 1.4 (Location-Scale Transformation). We wish to test the null hypothesis of equal distributions up to some location-scale transformation. This is a generalization of the classical two-sample problem. Let $Z = (X, Y)$ be a two-dimensional random vector and H be the joint cumulative distribution function of Z with marginal distribution functions F (for X) and G (for Y). The null hypothesis is

$$\mathrm{H}_0 : \text{For some } \theta = (\theta_1, \theta_2) \in \Theta, F(x) = G\left(\frac{x - \theta_1}{\theta_2}\right) \text{ for all } x \in \mathbb{R}. \tag{3}$$

In this case, $\psi_{x,\theta}(z) = \mathbb{1}\{z_1 \le x\} - \mathbb{1}\{z_2 \le (x - \theta_1)/\theta_2\}$ for every $z = (z_1, z_2) \in \mathbb{R}^2$ and every $(x, \theta) \in \mathbb{R} \times \Theta$, and $\phi_P(x, \theta) = F(x) - G((x - \theta_1)/\theta_2)$. A substantial number of tests exist for comparing two or multiple distributions. See, for example, Lehmann and Romano (2005) and Chen and Pokojovy (2018) for extensive reviews. Hall, Lombard, and Potgieter (2013) propose an extension of the Cramér–von Mises-type test based on empirical characteristic functions to examine whether the two samples come from the same location-scale family of distributions. Henze,

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Klar, and Zhu (2005) and Jiménez-Gamero et al. (2017) deal with the two-sample problem using similar test statistics.

An important special case of Example 1.4 is testing for heterogeneous treatment effects. We follow Ding, Feller, and Miratrix (2016) and Chung and Olivares (2021) and consider a randomized experiment model. Let Y denote the observable outcome of interest, and D denote the binary treatment variable. If an individual is randomly assigned to the treatment group and receives treatment, then $D = 1$; otherwise, the individual is randomly assigned to the control group and does not receive treatment, with $D = 0$. Suppose that $Y(1)$ is the potential outcome of an individual if treated, and $Y(0)$ is the potential outcome if not treated. The treatment effect is constant if $Y(1) - Y(0) = \theta$ almost surely for some fixed constant θ; otherwise, the treatment effect is said to be heterogeneous. The null hypothesis of constant treatment effect is

$$\mathrm{H}_0^s : \text{For some } \theta \in \Theta, Y(1) - Y(0) = \theta \text{ almost surely.} \tag{4}$$

Hypothesis (4) is a more restrictive sharp null and is usually not directly testable. A necessary and weaker condition of this sharp null hypothesis, which is considered by Ding et al. (2016) and Chung and Olivares (2021), is

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta, F(x) = G(x-\theta) \text{ for all } x \in \mathbb{R},$$

where F and G are the CDFs of $Y(1)$ and $Y(0)$, respectively. Clearly, this condition can be incorporated into (3).

Organization of the Article: Section 2 introduces the framework and develops theoretical results for testing general moment restrictions in the presence of nuisance parameters. Section 3 extends the results to dependent data. Section 4 provides Monte Carlo simulation evidence to show the performance of the test in finite samples. Section 5 concludes the article. Auxiliary lemmas, analyses and extensions of examples, all mathematical proofs, and additional simulation results are collected in the Supplementary Material.

Notation. Notation follows common conventions (e.g., van der Vaart and Wellner, 1996; Kosorok, 2008). We use M^{T} to denote the transpose of a matrix M. For $a, b \in \mathbb{R}$, we define $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. We use two forms of indicator functions: $\mathbb{1}\{S\} = 1$ if the statement S is true, and $\mathbb{1}\{S\} = 0$ otherwise; $\mathbb{1}_A(x) = 1$ if $x \in A$, and $\mathbb{1}_A(x) = 0$ if $x \notin A$. For an arbitrary set A, let $\ell^\infty(A)$ be the set of bounded real-valued functions on A. Equip $\ell^\infty(A)$ with the supremum norm $\|\cdot\|_\infty$ such that $\|f\|_\infty = \sup_{x\in A} |f(x)|$ for every $f \in \ell^\infty(A)$. For a subset B of a metric space, let $\mathcal{C}(B)$ be the set of continuous real-valued functions on B, and $\mathcal{C}_{\mathrm{b}}(B)$ be the set of bounded continuous functions on B, that is, $\mathcal{C}_{\mathrm{b}}(B) = \mathcal{C}(B) \cap \ell^\infty(B)$. Following the notation of van der Vaart and Wellner (1996), for every normed space $\mathbb{B}$ with a norm $\|\cdot\|_{\mathbb{B}}$, we define

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

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

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

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

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

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

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

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

for every $f \in L^p(\nu)$.

Let μ be the Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. For an arbitrary space $\mathcal{F}$, we say that $\mathbb{W}$ is a P-Brownian bridge in $\ell^\infty(\mathcal{F})$ if and only if $\mathbb{W}$ is a tight Borel measurable Gaussian process with $\mathbb{E}_P[\mathbb{W}(f_1)] = 0$ and $\mathbb{E}_P[\mathbb{W}(f_1)\mathbb{W}(f_2)] = P(f_1 f_2) - P(f_1)P(f_2)$ for all $f_1, f_2 \in \mathcal{F}$. Let $\rightsquigarrow$ denote the weak convergence defined in van der Vaart and Wellner (1996, p. 4). Let $\overset{\mathbb{P}}{\rightsquigarrow}$ and $\overset{\text{a.s.}}{\rightsquigarrow}$ denote the weak convergence in probability conditional on the sample and almost sure weak convergence conditional on the sample, respectively, as defined in Kosorok (2008, pp. 19–20).

2. TEST FORMULATION

2.1. Setup

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

Assumption 2.1. For every $\theta \in \Theta$, the function $x \mapsto \phi_P(x, \theta)$ is continuous.

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

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

Assumption 2.4. For every $\theta_0 \in \Theta$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{x \in \mathbb{R}} P\left[(\psi_{x,\theta} - \psi_{x,\theta_0})^2 \right] < \varepsilon$$

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

Assumption 2.1 shows that we focus on moment restrictions that are continuous in x for every $\theta \in \Theta$. Assumption 2.2 requires the absolute continuity of the Lebesgue measure μ with respect to the probability measure ν. For example, ν could be set as the probability measure corresponding to a normal distribution

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

with a large variance.[2] Assumption 2.3 is a common condition on the compactness of Θ. Assumption 2.4 can be understood as the continuity of $\psi_{x,\theta}$ with respect to θ under a certain metric.

Define a function space

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

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

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

for all $\theta \in \Theta$ with $\|\theta - \theta_0\|_2 < \delta$. Note that for every $x \in \mathbb{R}$ and all $\theta, \theta_0 \in \Theta$, by Jensen's inequality,

$$[\phi_P(x, \theta) - \phi_P(x, \theta_0)]^2 \leq P\left[(\psi_{x,\theta} - \psi_{x,\theta_0})^2\right].$$

Since ν is a probability measure, Assumption 2.4 implies that $\phi_P \in \mathbb{D}_{\mathcal{L}0}$.

The proposition below provides an equivalent characterization of the null hypothesis in (2). We construct the test based on this equivalent characterization to avoid estimating the nuisance parameter θ under the null.

PROPOSITION 2.1. *If Assumptions 2.1–2.4 hold, then the null hypothesis in (2) is equivalent to*

$$\mathrm{H}_0 : \inf_{\theta \in \Theta} \int_{\mathbb{R}} [\phi_P(x, \theta)]^2 \, \mathrm{d}\nu(x) = 0. \tag{5}$$

It is worth noting that different measures ν may deliver different power properties of the test. However, searching for the optimal ν to maximize power is challenging, as it may depend in a complicated manner on the DGP.

The measure $\nu(\mathbb{R})$ is assumed to be finite (Assumption 2.2) to obtain the theoretical results in the article. In practice, we suggest setting ν to a normal probability measure with a large variance so that it does not heavily concentrate on some region of the real line, given no prior information about the DGP. Other probability measures satisfying Assumption 2.2 also work asymptotically for the proposed method. For finite samples, the simulation results in Section 4 show that normal probability measures with different variances ($\mathcal{N}(0, 1)$, $\mathcal{N}(0, 5^2)$, $\mathcal{N}(0, 10^2)$) perform well.

2.2. Test Statistic

We first restrict our attention to independent and identically distributed (i.i.d.) samples and will extend the results to dependent data in Section 3. Let $\widehat{P}_n$ be the empirical probability measure of the sample $\mathbf{Z}_n$, which assigns weight $1/n$ to each

[2] See the discussion and simulation results in Section 4.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

observation Z_i with $i \in \{1, \ldots, n\}$. Then the sample analog of ϕ_P is defined as

$$\widehat{\phi}_n(x,\theta) = \widehat{P}_n(\psi_{x,\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_{x,\theta}(Z_i)$$

for every $(x,\theta) \in \mathbb{R} \times \Theta$. We present the exact function form of $\widehat{\phi}_n$ in every example.

Example 1.1 (Cont). With the known function g, it follows by definition that

$$\widehat{\phi}_n(x,\theta) = \widehat{P}_n(\psi_{x,\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_{x,\theta}(Z_i) = \frac{1}{n}\sum_{i=1}^{n} g(Y_i,\theta)\mathbb{1}\{X_i \le x\}$$

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

Example 1.2 (Cont). The cumulative distribution function G can be estimated by the empirical distribution function $\widehat{G}_n$ such that for every $x \in \mathbb{R}$,

$$\widehat{G}_n(x) = \widehat{P}_n(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(Z_i).$$

Then

$$\widehat{\phi}_n(x,\theta) = \widehat{P}_n(\psi_{x,\theta}) = \widehat{P}_n(\mathbb{1}_{(-\infty,x]}) + \widehat{P}_n(\mathbb{1}_{(-\infty,2\theta-x]}) - 1 = \widehat{G}_n(x) + \widehat{G}_n(2\theta - x) - 1$$

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

Example 1.3 (Cont). The cumulative distribution function G can be estimated by the empirical distribution function $\widehat{G}_n$ such that for every $x \in \mathbb{R}$,

$$\widehat{G}_n(x) = \widehat{P}_n(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(Z_i).$$

Then $\widehat{\phi}_n(x,\theta) = \widehat{P}_n(\psi_{x,\theta}) = \widehat{P}_n(\mathbb{1}_{(-\infty,x]}) - G_0(x,\theta) = \widehat{G}_n(x) - G_0(x,\theta)$ for every $(x,\theta) \in \mathbb{R} \times \Theta$.

Example 1.4 (Cont). For every $i \in \{1, \ldots, n\}$, the observation $Z_i = (X_i, Y_i)$. Let $\widehat{P}_n$ be the empirical distribution of $\{Z_i\}_{i=1}^n$, and $\widehat{H}_n$ be its empirical distribution function so that

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

for all $(x,y) \in \mathbb{R}^2$. Let $\widehat{P}_{X,n}$ and $\widehat{P}_{Y,n}$ be the marginal distributions of $\widehat{P}_n$, that is, the (marginal) empirical distributions of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, respectively. It follows that

$$\widehat{\phi}_n(x,\theta) = \widehat{P}_n(\psi_{x,\theta}) = \widehat{P}_{X,n}(\mathbb{1}_{(-\infty,x]}) - \widehat{P}_{Y,n}(\mathbb{1}_{(-\infty,(x-\theta_1)/\theta_2]})$$

for every $(x,\theta) \in \mathbb{R} \times \Theta$. The marginal distribution functions F and G can be estimated by the empirical distribution functions $\widehat{F}_n$ and $\widehat{G}_n$, respectively, where

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

for every $x \in \mathbb{R}$,

$$\widehat{F}_n(x) = \widehat{P}_{X,n}(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i) \text{ and}$$

$$\widehat{G}_n(x) = \widehat{P}_{Y,n}(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(Y_i).$$

This implies that $\widehat{\phi}_n(x,\theta) = \widehat{F}_n(x) - \widehat{G}_n[(x-\theta_1)/\theta_2]$ for every $(x,\theta) \in \mathbb{R} \times \Theta$.

We may also test this null hypothesis with two independent samples of different sizes. This case will be discussed in Section C of the Supplementary Material, where we present the results for comparing multiple samples.

To obtain the asymptotic law of the stochastic process $\widehat{\phi}_n$, we need the following assumption on the function class Ψ.

Assumption 2.5. The function class $\Psi = \{\psi_{x,\theta} : (x,\theta) \in \mathbb{R} \times \Theta\}$ satisfies that

$$\sup_{f\in\Psi} |P(f)| < \infty \text{ and } \sup_{f\in\Psi} |f(z) - P(f)| < \infty \tag{6}$$

for all $z \in \mathbb{R}^{d_z}$ and is P-Donsker in the sense that

$$\sqrt{n}(\widehat{P}_n - P) \rightsquigarrow \mathbb{W} \text{ in } \ell^\infty(\Psi) \tag{7}$$

as $n \to \infty$, where $\mathbb{W}$ is a P-Brownian bridge in $\ell^\infty(\Psi)$.

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

LEMMA 2.1. *If Assumptions 2.4 and 2.5 hold, then* $(\widehat{\phi}_n - \phi_P) \in \ell^\infty(\mathbb{R} \times \Theta)$ *for all* $n \in \mathbb{Z}_+$. *In addition,*

$$\sup_{(x,\theta)\in\mathbb{R}\times\Theta} \left|\widehat{\phi}_n(x,\theta) - \phi_P(x,\theta)\right| \xrightarrow{\mathbb{P}} 0 \text{ and } \sqrt{n}(\widehat{\phi}_n - \phi_P) \rightsquigarrow \mathbb{G}_0 \text{ in } \ell^\infty(\mathbb{R} \times \Theta)$$

as $n \to \infty$, *where* $\mathbb{G}_0$ *is some tight random element which almost surely takes values in* $\mathbb{D}_{\mathcal{L}0}$.

Define a function space

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

Define a map $\mathcal{L}$ on $\mathbb{D}_{\mathcal{L}}$ such that $\mathcal{L}(\varphi) = \inf_{\theta\in\Theta} \int_{\mathbb{R}} [\varphi(x,\theta)]^2 \, \mathrm{d}\nu(x)$ for every $\varphi \in \mathbb{D}_{\mathcal{L}}$. Then under Assumptions 2.1–2.5, the null and the alternative hypotheses can be expressed as

$$\mathrm{H}_0 : \mathcal{L}(\phi_P) = 0 \text{ and } \mathrm{H}_1 : \mathcal{L}(\phi_P) > 0. \tag{8}$$

To test the null hypothesis in (8), we set the test statistic to $n\mathcal{L}(\widehat{\phi}_n)$.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Next, we show that the map $\mathcal{L}$ is Hadamard directionally differentiable, but its Hadamard directional derivative is degenerate under H_0.[3] Define

$$\mathbb{D}_0 = \{\varphi \in \mathbb{D}_{\mathcal{L}0} : \mathcal{L}(\varphi) = 0\}.$$

The following lemma provides the Hadamard directional derivative of $\mathcal{L}$ and its first-order degeneracy under H_0.

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

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

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

The first-order degeneracy of $\mathcal{L}$ under H_0 implies that we may need to find the second-order Hadamard directional derivative of $\mathcal{L}$.[4] We assume the following conditions to guarantee the existence of the second-order Hadamard directional derivative of $\mathcal{L}$.

Assumption 2.6. The function ϕ_P is twice differentiable with respect to θ, and the second partial derivative satisfies

$$\int_{\mathbb{R}} \sup_{\theta\in\Theta} \left\| \frac{\partial^2 \phi_P(z,\vartheta)}{\partial\vartheta\,\partial\vartheta^{\mathrm{T}}}\bigg|_{(z,\vartheta)=(x,\theta)} \right\|_2^2 \mathrm{d}\nu(x) < \infty, \tag{9}$$

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

Assumption 2.7. The set $\Theta_0 \equiv \{\theta \in \Theta : \int_{\mathbb{R}} [\phi_P(x,\theta)]^2\, \mathrm{d}\nu(x) = 0\} \subset \mathrm{int}(\Theta)$, and there exist $\kappa \in (0,1]$, $\bar{\varepsilon} > 0$, and $C > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$,

$$\inf_{\theta\in\Theta\setminus\Theta_0^\varepsilon} \left\{ \int_{\mathbb{R}} [\phi_P(x,\theta)]^2\, \mathrm{d}\nu(x) \right\}^{1/2} \geq C\varepsilon^\kappa. \tag{10}$$

We provide Assumptions 2.6 and 2.7 following the basic idea of Chen and Fang (2019b). Assumption 2.6 requires the boundedness of the second partial derivative of ϕ_P in the sense of (9). Assumption 2.7 requires that the set Θ_0 is in the interior of Θ and it is well separated. The condition in (10) is similar to the partial identification assumption used in Chernozhukov et al. (2007, p. 1265). It is worth noting that these conditions are sufficient but not necessary for our results, as also mentioned by Chen and Fang (2019b). We impose such high-level

[3] See Definition D.1 in the Supplementary Material for Hadamard directional differentiability.

[4] See Definition D.2 in the Supplementary Material for second-order Hadamard directional differentiability.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

conditions for theoretical completeness. In Section 4, we verify these assumptions for a conditional moment restriction model.

LEMMA 2.3. *If Assumptions 2.3, 2.4, 2.6, and 2.7 hold, and $\phi_P \in \mathbb{D}_0$, then the function $\mathcal{L}$ is second-order Hadamard directionally differentiable at ϕ_P tangentially to $\mathbb{D}_{\mathcal{L}0}$ with the second-order Hadamard directional derivative*

$$\mathcal{L}''_{\phi_P}(h) = \inf_{\theta\in\Theta_0(\phi_P)} \inf_{v\in\mathbb{R}^{d_\theta}} \left\| \left[\Phi'(\theta)\right]^T v + \mathscr{H}(\theta) \right\|^2_{L^2(\nu)} \quad \text{for all } h \in \mathbb{D}_{\mathcal{L}0},$$

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

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

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

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

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

PROPOSITION 2.2. *If Assumptions 2.1–2.7 hold and* H_0 *is true ($\phi_P \in \mathbb{D}_0$), then*

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

2.3. Bootstrap Procedure

The distribution of $\mathcal{L}''_{\phi_P}(\mathbb{G}_0)$ in Proposition 2.2 is unknown because both the function $\mathcal{L}''_{\phi_P}$ and the stochastic process $\mathbb{G}_0$ depend on the unknown underlying distribution P. Motivated by Hong and Li (2018) and Chen and Fang (2019b), we propose to approximate $\mathcal{L}''_{\phi_P}$ by a consistent estimator and approximate the distribution of $\mathbb{G}_0$ by bootstrap.[5] We use the numerical second-order Hadamard directional derivative $\widehat{\mathcal{L}}''_n$ to approximate $\mathcal{L}''_{\phi_P}$, which is defined as

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

for all $h \in \ell^\infty(\mathbb{R}\times\Theta)$, where $\{\tau_n\}$ is a sequence of tuning parameters satisfying the assumption below.[6]

[5] Bootstrap may not be the only method to approximate the distribution of $\mathbb{G}_0$ in our framework. Other consistent estimators of $\mathbb{G}_0$ might also suffice for the proposed approach.

[6] As discussed in Chen and Fang (2019b), the modified bootstrap in Babu (1984) (Babu correction) is inappropriate when $\mathcal{L}$ is only second-order Hadamard directionally differentiable but $\mathcal{L}''_{\phi_P}$ is not "continuous" in ϕ_P. To ensure that our method can accommodate more general cases, we employ the bootstrap method of Hong and Li (2018) and Chen and Fang (2019b).

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Assumption 2.8. $\{\tau_n\} \subset \mathbb{R}_+$ is a sequence of scalars such that $\tau_n \downarrow 0$ and $\tau_n\sqrt{n} \to \infty$ as $n \to \infty$.

Assumption 2.8 provides the rate at which $\tau_n \downarrow 0$. Under this condition, we show that $\widehat{\mathcal{L}}''_n$ approximates $\mathcal{L}''_{\phi_P}$ well in the following lemma.

LEMMA 2.4. *If Assumptions 2.1–2.8 hold and* H_0 *is true* ($\phi_P \in \mathbb{D}_0$), *then for every sequence* $\{h_n\} \subset \ell^\infty(\mathbb{R} \times \Theta)$ *and every* $h \in \mathbb{D}_{\mathcal{L}0}$ *such that* $h_n \to h$ *in* $\ell^\infty(\mathbb{R} \times \Theta)$ *as* $n \to \infty$, *we have*

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

We next approximate the distribution of $\mathbb{G}_0$ via bootstrap. The bootstrap sample $\mathbf{Z}^*_n = \{Z^*_i\}_{i=1}^n$ is i.i.d. drawn from the empirical distribution $\widehat{P}_n$ of the original sample $\mathbf{Z}_n$. Equivalently, $\mathbf{Z}^*_n$ is a random sample of size n, drawn from the set $\mathbf{Z}_n$ with replacement. Let $\widehat{P}^*_n$ be the empirical distribution of $\mathbf{Z}^*_n$. Then the bootstrap version of $\widehat{\phi}_n$ is $\widehat{\phi}^*_n$ such that

$$\widehat{\phi}^*_n(x,\theta) = \widehat{P}^*_n(\psi_{x,\theta}) = \frac{1}{n}\sum_{i=1}^n \psi_{x,\theta}(Z^*_i)$$

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

Example 1.1 (Cont). It follows by definition that

$$\widehat{\phi}^*_n(x,\theta) = \widehat{P}^*_n(\psi_{x,\theta}) = \frac{1}{n}\sum_{i=1}^n \psi_{x,\theta}(Z^*_i) = \frac{1}{n}\sum_{i=1}^n g(Y^*_i,\theta)\mathbb{1}\{X^*_i \le x\}$$

for every $(x,\theta) \in \mathbb{R} \times \Theta$, where $Z^*_i = (X^*_i, Y^*_i)$.

Example 1.2 (Cont). Define

$$\widehat{G}^*_n(x) = \widehat{P}^*_n(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}\left(Z^*_i\right)$$

for every $x \in \mathbb{R}$. Then

$$\widehat{\phi}^*_n(x,\theta) = \widehat{P}^*_n(\psi_{x,\theta}) = \widehat{P}^*_n(\mathbb{1}_{(-\infty,x]}) + \widehat{P}^*_n(\mathbb{1}_{(-\infty,2\theta-x]}) - 1 = \widehat{G}^*_n(x) + \widehat{G}^*_n(2\theta - x) - 1$$

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

Example 1.3 (Cont). Define

$$\widehat{G}^*_n(x) = \widehat{P}^*_n(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}\left(Z^*_i\right)$$

for every $x \in \mathbb{R}$. Then

$$\widehat{\phi}^*_n(x,\theta) = \widehat{P}^*_n(\psi_{x,\theta}) = \widehat{P}^*_n(\mathbb{1}_{(-\infty,x]}) - G_0(x,\theta) = \widehat{G}^*_n(x) - G_0(x,\theta)$$

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

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Example 1.4 (Cont). Define $\widehat{P}_n^*$ as the empirical distribution of $\{Z_i^*\}_{i=1}^n$ with $Z_i^* = (X_i^*, Y_i^*)$. Let $\widehat{P}_{X,n}^*$ and $\widehat{P}_{Y,n}^*$ be the marginal distributions of $\widehat{P}_n^*$, that is, the (marginal) empirical distributions of $\{X_i^*\}_{i=1}^n$ and $\{Y_i^*\}_{i=1}^n$, respectively. It follows that

$$\widehat{\phi}_n^*(x,\theta) = \widehat{P}_n^*(\psi_{x,\theta}) = \widehat{P}_{X,n}^*(\mathbb{1}_{(-\infty,x]}) - \widehat{P}_{Y,n}^*(\mathbb{1}_{(-\infty,(x-\theta_1)/\theta_2]})$$

for every $(x,\theta) \in \mathbb{R} \times \Theta$. Define $\widehat{F}_n^*$ and $\widehat{G}_n^*$ to be the (marginal) empirical distribution functions of $\{X_i^*\}_{i=1}^n$ and $\{Y_i^*\}_{i=1}^n$, respectively, such that for every $x \in \mathbb{R}$,

$$\widehat{F}_n^*(x) = \widehat{P}_{X,n}^*(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}\left(X_i^*\right) \text{ and}$$

$$\widehat{G}_n^*(x) = \widehat{P}_{Y,n}^*(\mathbb{1}_{(-\infty,x]}) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}\left(Y_i^*\right).$$

This implies that $\widehat{\phi}_n^*(x,\theta) = \widehat{F}_n^*(x) - \widehat{G}_n^*((x-\theta_1)/\theta_2)$ for every $(x,\theta) \in \mathbb{R} \times \Theta$.

The following lemma establishes the conditional weak convergence of $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ in probability as $n \to \infty$.

LEMMA 2.5. *If Assumption 2.5 holds, then as* $n \to \infty$,

$$\sup_{\Gamma \in \mathrm{BL}_1(\ell^\infty(\mathbb{R}\times\Theta))} \left|\mathbb{E}\left[\Gamma\left(\sqrt{n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right)\middle|\mathbf{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0\right)\right]\right| \xrightarrow{\mathbb{P}} 0,$$

and $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ *is asymptotically measurable, where* $\mathbb{G}_0$ *is defined as in Lemma 2.1.*

With the numerical estimator $\widehat{\mathcal{L}}_n''$ for $\mathcal{L}_{\phi_P}''$ and a suitable bootstrap approximation $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ for $\mathbb{G}_0$ at hand, we can naturally approximate the distribution of $\mathcal{L}_{\phi_P}''(\mathbb{G}_0)$ by the conditional distribution of the bootstrap test statistic $\widehat{\mathcal{L}}_n''\{\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)\}$ given the original sample. This is justified by the following proposition.

PROPOSITION 2.3. *If Assumptions 2.1–2.8 hold and* H_0 *is true* ($\phi_P \in \mathbb{D}_0$), *then*

$$\sup_{\Gamma \in \mathrm{BL}_1(\mathbb{R})} \left|\mathbb{E}\left[\Gamma\left(\widehat{\mathcal{L}}_n''\left[\sqrt{n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n\right)\right]\right)\middle|\mathbf{Z}_n\right] - \mathbb{E}\left[\Gamma\left(\mathcal{L}_{\phi_P}''\left(\mathbb{G}_0\right)\right)\right]\right| \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$.

2.4. Asymptotic Properties

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

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

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

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

THEOREM 2.1. *Suppose that Assumptions 2.1–2.8 hold.*

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

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

(ii) *If* H_0 *is false, then*

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

2.5. Local Power

In this section, we consider the local power of the test following the discussion in Chen and Fang (2019b). For each $n \in \mathbb{Z}_+$, let the sample $\mathbf{Z}_n = \{Z_i\}_{i=1}^n$ be distributed according to the joint law $P_n^n = \prod_{i=1}^n P_n$, where P_n is a probability distribution on $(\mathbb{R}^{d_z}, \mathscr{B}(\mathbb{R}^{d_z}))$ with $P_n(B) = \mathbb{P}(Z_i \in B)$ for every Borel set B. That is, for each $n \in \mathbb{Z}_+$, the observations $Z_1, \ldots, Z_n$ are i.i.d. with distribution P_n. We suppose that the null hypothesis H_0 is false for each P_n, that is, for all $\theta \in \Theta$, $P_n(\psi_{x,\theta}) \neq 0$ for some $x \in \mathbb{R}$. Suppose that P_n converges (in a way as described in the following assumption) to the probability measure P, and that P satisfies H_0, that is, for some $\theta \in \Theta$, $P(\psi_{x,\theta}) = 0$ for all $x \in \mathbb{R}$.

Assumption 2.9. The probability distributions P_n and P satisfy that

$$\lim_{n\to\infty} \int \left[\sqrt{n}\left(\mathrm{d}P_n^{1/2} - \mathrm{d}P^{1/2}\right) - \frac{1}{2} v_0 \, \mathrm{d}P^{1/2}\right]^2 = 0 \tag{11}$$

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

Our local power results rely on Assumption 2.9, which is similar to (3.10.10) of van der Vaart and Wellner (1996). The following proposition states formally the local power property of the test.

PROPOSITION 2.4. *Suppose that Assumptions 2.1–2.9 hold and* $\sup_{f\in\Psi} |P_n(f^2)| = O(1)$. *Then* $\sqrt{n}(\widehat{\phi}_n - \phi_P) \rightsquigarrow \mathbb{G}_0 + \zeta_P$, *where* $\mathbb{G}_0$ *is some tight random element, and* $\zeta_P(x,\theta) = P(\psi_{x,\theta} v_0)$ *for every* $(x,\theta) \in \mathbb{R} \times \Theta$. *Furthermore, if the CDF of*

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

$\mathcal{L}''_{\phi_P}(\mathbb{G}_0)$ is strictly increasing and continuous at its $1-\alpha$ quantile $c_{1-\alpha}$, then it follows that

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

Proposition 2.4 follows from Lemma C.1 of Chen and Fang (2019b) and provides lower bounds for the power of the test under local perturbations to the null.

3. DEPENDENT DATA

This section considers cases where the observations $\{Z_i\}_{i=1}^n$ may be dependent. For results established in Section 2, it is worth noting that Lemmas 2.2–2.4, Propositions 2.1–2.3, and Theorem 2.1 do not directly rely on the i.i.d. nature of the data observations, possibly given the consistency and weak convergence of $\widehat{\phi}_n$ (Lemma 2.1) and the conditional weak convergence of $\widehat{\phi}_n^*$ in probability (Lemma 2.5). Thus, to obtain the asymptotic properties of the proposed test in dependent samples, it suffices to establish the consistency and weak convergence of $\widehat{\phi}_n$ and the conditional weak convergence of $\widehat{\phi}_n^*$ in probability under dependency.

A sequence of d_z-dimensional random vectors, $\{Z_i : i \in \mathbb{Z}\}$, is said to be strictly stationary, if for all $\{i_1, \ldots, i_n\} \subset \mathbb{Z}$ and all $n \in \mathbb{Z}_+$, the joint distribution of $(Z_{i_1+k}, \ldots, Z_{i_n+k})$ does not depend on k. For $-\infty \leq s \leq t \leq \infty$, let $\mathscr{S}_s^t$ be the σ-field generated by $\{Z_s, \ldots, Z_t\}$. Following Equation (II) of Volkonskii and Rozanov (1959) and (1.1) of Arcones and Yu (1994), the β-mixing coefficient β_k of the sequence $\{Z_i : i \in \mathbb{Z}\}$ is defined as

$$\beta_k = \sup_{t\in\mathbb{Z}} \mathbb{E}\left[\sup_{A\in\mathscr{S}_{t+k}^{\infty}} \left|\mathbb{P}\left(A \,\middle|\, \mathscr{S}_{-\infty}^{t}\right) - \mathbb{P}(A)\right|\right],$$

and $\{Z_i : i \in \mathbb{Z}\}$ is said to be β-mixing if and only if $\beta_k \to 0$ as $k \to \infty$.

Throughout our discussion of cases with dependent data, we assume that the sample $\mathbf{Z}_n = \{Z_i : i = 1, \ldots, n\}$ is a finite segment of the strictly stationary sequence $\{Z_i : i \in \mathbb{Z}\}$ in which the common marginal distribution of Z_i is P. We impose the following assumptions.

Assumption 3.1. The class $\Psi = \{\psi_{x,\theta} : (x, \theta) \in \mathbb{R} \times \Theta\}$ is a VC-subgraph class of functions satisfying (6) with $P(\overline{\psi}^p) < \infty$ for some $p \in (2, \infty)$, where $\overline{\psi}(z) = \sup_{f\in\Psi} |f(z)|$ for every $z \in \mathbb{R}^{d_z}$, and Ψ is totally bounded under $\|\cdot\|_{L^2(P)}$.[7]

With p specified as in Assumption 3.1, we introduce the following condition for β_k.

[7] See the definition of VC-subgraph class of functions in Section 2.6 of van der Vaart and Wellner (1996, p. 141).

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Assumption 3.2. The sequence $\{Z_i : i \in \mathbb{Z}\}$ is β-mixing with coefficient $\beta_k = O(k^{-q})$ as $k \to \infty$ for some $q > p/(p-2)$.

Assumption 3.1 emerges as one of the conditions in Theorem 2.1 of Arcones and Yu (1994) and Theorem 1 of Radulović (1996). Assumption 3.2 corresponds to one of the conditions in Theorem 1 of Radulović (1996).

Let $\widehat{\phi}_n$ and ϕ_P be defined as in Section 2. The lemma below establishes the consistency and weak convergence of $\widehat{\phi}_n$ as $n \to \infty$.

LEMMA 3.1. *If Assumptions 2.3, 2.4, 3.1, and 3.2 hold, then* $(\widehat{\phi}_n - \phi_P) \in \ell^\infty(\mathbb{R} \times \Theta)$ *for all* $n \in \mathbb{Z}_+$. *In addition,*

$$\sup_{(x,\theta)\in\mathbb{R}\times\Theta} \left|\widehat{\phi}_n(x,\theta) - \phi_P(x,\theta)\right| \xrightarrow{\mathbb{P}} 0 \text{ and } \sqrt{n}(\widehat{\phi}_n - \phi_P) \rightsquigarrow \mathbb{G}_0 \text{ in } \ell^\infty(\mathbb{R} \times \Theta)$$

as $n \to \infty$, *where* $\mathbb{G}_0$ *is tight and almost surely takes values in* $\mathbb{D}_{\mathcal{L}0}$.

To construct the bootstrap sample $\mathbf{Z}_n^* = \{Z_i^*\}_{i=1}^n$, we follow Radulović (1996) and use the moving blocks bootstrap (MBB) procedure. Recall that the original sample is $\{Z_i\}_{i=1}^n$. Let $b \in \mathbb{Z}_+$ be the block size satisfying $b \to \infty$ and $b/n \to 0$, and $k \in \mathbb{Z}_+$ be the number of blocks. Without loss of generality, we may assume that k and b satisfy $kb = n$.[8] For $i \in \{1, \ldots, b-1\}$, we set $Z_{n+i} = Z_i$. Let the random variables $I_1, \ldots, I_k$ be i.i.d. from Unif$\{1, \ldots, n\}$ and independent of the original sample. For all $\ell \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, b\}$, set the bootstrap observation $Z^*_{(\ell-1)b+j} = Z_{I_\ell + j - 1}$. That is, the bootstrap sample is

$$\mathbf{Z}_n^* = \{Z_{I_1}, Z_{I_1+1}, \ldots, Z_{I_1+b-1}, Z_{I_2}, Z_{I_2+1}, \ldots, Z_{I_2+b-1}, \ldots, Z_{I_k}, Z_{I_k+1}, \ldots, Z_{I_k+b-1}\}.$$

Let $\widehat{P}_n^*$ be the empirical distribution of $\mathbf{Z}_n^*$. The bootstrap version of $\widehat{\phi}_n$ is defined as

$$\widehat{\phi}_n^*(x,\theta) = \widehat{P}_n^*(\psi_{x,\theta}) = \frac{1}{n}\sum_{i=1}^n \psi_{x,\theta}(Z_i^*)$$

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

We impose the assumption below on the block size b, which treats b as a function of n, that is, $b = b(n)$. This assumption corresponds to one of the conditions in Theorem 1 of Radulović (1996).

Assumption 3.3. The block size b is a function of the sample size n such that $b = b(n) = O(n^r)$ as $n \to \infty$ for some $0 < r < (p-2)/(2p-2)$.

The following lemma establishes the conditional weak convergence of $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ in probability.

[8] In practice, n/b may not always be an integer. In this case, we set $k = \lceil n/b \rceil$ and generate $kb > n$ bootstrap observations according to the algorithm described in the main text, and then keep the first n observations as the bootstrap sample.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

LEMMA 3.2. *If Assumptions 3.1–3.3 hold, then as* $n \to \infty$,

$$\sup_{\Gamma \in \mathrm{BL}_1(\ell^\infty(\mathbb{R}\times\Theta))} \left| \mathbb{E}\left[\Gamma\left(\sqrt{n}\left(\widehat{\phi}_n^* - \widehat{\phi}_n \right) \right) \middle| \mathbf{Z}_n \right] - \mathbb{E}\left[\Gamma\left(\mathbb{G}_0 \right) \right] \right| \xrightarrow{\mathbb{P}} 0,$$

where $\mathbb{G}_0$ *is defined as in Lemma 3.1.*

Given the modification to the construction of the bootstrap sample, the remaining steps of the test follow the procedure in Section 2. For dependent data, the test is also asymptotically size controlled and consistent, as shown in Theorem 3.1.

THEOREM 3.1. *Suppose that Assumptions 2.1–2.4, 2.6–2.8, and 3.1–3.3 hold, and that* $\sqrt{n}(\widehat{\phi}_n^* - \widehat{\phi}_n)$ *is asymptotically measurable.*

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

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

(ii) *If* H_0 *is false, then*

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

4. MONTE CARLO EXPERIMENTS

In this section, we construct the Monte Carlo experiments based on the conditional moment restriction models with weak instrumental variables (IVs) in Jun and Pinkse (2009, Example II). Let y_i be a scalar outcome variable, Y_i be a scalar endogenous variable, and z_i be a scalar IV. The model of interest is

$$\mathbb{E}_P[y_i - Y_i\theta_0 | z_i] = 0 \text{ almost surely} \tag{12}$$

for a true structural parameter $\theta_0 \in \Theta \subset \mathbb{R}$. We consider the null hypothesis

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta,\ \mathbb{E}_P[y_i - Y_i\theta | z_i] = 0 \text{ almost surely},$$

which is equivalent to

$$\mathrm{H}_0 : \text{For some } \theta \in \Theta,\ \mathbb{E}_P\left[(y_i - Y_i\theta)\mathbb{1}\{z_i \le x\} \right] = 0 \text{ for all } x \in \mathbb{R}.$$

As noted by Jun and Pinkse (2009), there are several specification tests for (12) under strong point identification and the assumption that θ_0 can be $\sqrt{n}$-consistently estimated under the null (e.g., Bierens, 1990; Fan and Li, 1996, 2000; Zheng, 1996). Since $\mathbb{E}_P[y_i - Y_i\theta_0 | z_i] = 0$ almost surely for some θ_0 under the null, typical estimators of θ_0 include two-stage least squares (2SLS) and semiparametric methods. However, when instruments are weak, these estimators of θ_0 may be undesirable (e.g., Staiger and Stock, 1997; Stock and Wright, 2000; Jun and Pinkse, 2012), and thus two-step tests plugging in preliminary estimators of θ_0 may not perform well.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Jun and Pinkse (2009) propose semi-parametric specification tests of conditional moment restrictions with weak instruments, which do not require a consistent first-step estimator. As shown in Theorems 1 and 2 of Jun and Pinkse (2009), their tests yield limiting rejection probabilities no greater than the nominal significance level under the null. Based on their Example II, Jun and Pinkse (2009) study the finite sample performance of their tests with weak instruments via Monte Carlo experiments. We first follow Jun and Pinkse (2009) and consider two cases under the null. In the first case (Case 1), $\mathbb{E}_P[z_i Y_i] = 0$, that is, the rank condition fails and z_i is not a valid instrument for Y_i when estimating θ_0 by 2SLS in two-step tests, which may be seen as an extreme case of weak instruments. Thus, the 2SLS estimator of θ_0 that uses z_i as the instrument for Y_i is unreliable. In the second case (Case 2), $\mathbb{E}_P[Y_i|z_i] \to 0$ almost surely as $n \to \infty$, that is, all measurable functions f of z_i with $\mathbb{E}_P[|f(z_i)Y_i|] < \infty$ may be weak instruments for Y_i when estimating θ_0 by 2SLS in two-step tests because $\mathbb{E}_P[f(z_i)Y_i]$ may converge to 0. As discussed in Jun and Pinkse (2012), semi-parametric estimators of θ_0 may also break down when $\mathbb{E}_P[Y_i|z_i]$ decays too fast in n. In addition, we consider a third case (Case 3), which is an extreme case of Case 2: $\mathbb{E}_P[Y_i|z_i] = 0$ almost surely.

As demonstrated in Tables 1 and 2 of Jun and Pinkse (2009), their tests improve greatly upon two-step plug-in methods in the presence of weak instruments, while they are often conservative, which is in line with their theoretical results. The proposed test in this article is asymptotically exactly size controlled and consistent under certain conditions, regardless of the strength of instruments. We numerically present these properties through Monte Carlo experiments, where the DGPs are designed for conditional moment restriction models with weak instruments as in the above cases.

Now we introduce the designs of our simulations. For i.i.d. samples, we follow the design of Jun and Pinkse (2009):

$$y_i = Y_i + \delta \ln\left(Y_i^2 + 1\right) + u_i,$$
$$Y_i = \lambda g(z_i) + v_i,$$

where $\{(u_i, v_i, z_i) : i = 1, \ldots, n\}$ are i.i.d. with

$$\begin{bmatrix} u_i \\ v_i \\ z_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).$$

The aforementioned three cases are realized in the following manner:

- *Case 1*: $\rho = 0.5$, $\lambda = 1$, and $g(z) = z^2 - 1$. The moment $\mathbb{E}_P[z_i Y_i] = 0$.
- *Case 2*: $\rho = -0.99$, $\lambda = 0.07\sqrt{200/n}$, and $g(z) = z$. The conditional moment $\mathbb{E}_P[Y_i|z_i] \to 0$ almost surely as $n \to \infty$.
- *Case 3*: $\rho = -0.5$ and $\lambda = 0$. The conditional moment $\mathbb{E}_P[Y_i|z_i] = 0$ almost surely.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

For each case, we consider four DGPs characterized by the values of δ:

- *DGP (0)*: $\delta = 0$. The null is true.
- *DGP (1)*: $\delta = 0.2$. The null is false.
- *DGP (2)*: $\delta = 0.6$. The null is false.
- *DGP (3)*: $\delta = 1$. The null is false.

We also consider dependent data. For every DGP introduced above, we construct the dependent-data counterpart by generating $\{z_i : i = 1, \ldots, n\}$ as

$$z_0 = 0, z_i = 0.5 z_{i-1} + \varepsilon_i,$$

where $\{\varepsilon_i : i = 1, \ldots, n\}$ are i.i.d. $\mathcal{N}(0,1)$ and independent of $\{(u_i, v_i) : i = 1, \ldots, n\}$.

Remark 4.1. For illustration of the high-level assumptions in Section 2.2, we consider Case 1 with $\delta = 0$. With $\theta_0 = 1$, H_0 is true and thus

$$\mathbb{E}_P[(y_i - Y_i\theta_0)\,\mathbb{1}\{z_i \le x\}] = 0$$

for all x. We have that for all θ,

$$\begin{aligned}\phi_P(x,\theta) &= \mathbb{E}_P[(y_i - Y_i\theta)\,\mathbb{1}\{z_i \le x\}] \\ &= \mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}] - \mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}]\theta.\end{aligned}$$

It follows that for all θ,

$$\begin{aligned}&\int_{\mathbb{R}} \phi_P(x,\theta)^2\,\mathrm{d}\nu(x) \\ &= \int_{\mathbb{R}} (\mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}])^2 \mathrm{d}\nu(x) - 2\int_{\mathbb{R}} \mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}]\,\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}]\,\mathrm{d}\nu(x)\,\theta \\ &\quad + \int_{\mathbb{R}} (\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}])^2 \mathrm{d}\nu(x)\,\theta^2 \ge 0.\end{aligned}$$

The value $\theta_0 = 1$ satisfies $\int_{\mathbb{R}} \phi_P(x,\theta_0)^2\,\mathrm{d}\nu(x) = 0$, so we have

$$\Theta_0 = \{\theta_0\}, \theta_0 = \frac{\int_{\mathbb{R}} \mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}]\,\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}]\,\mathrm{d}\nu(x)}{\int_{\mathbb{R}} (\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}])^2 \mathrm{d}\nu(x)} = 1.$$

For every $\varepsilon > 0$,

$$\begin{aligned}&\int_{\mathbb{R}} \phi_P(x,\theta_0-\varepsilon)^2\,\mathrm{d}\nu(x) \\ &= \int_{\mathbb{R}} (\mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}])^2 \mathrm{d}\nu(x) \\ &\quad - 2\int_{\mathbb{R}} \mathbb{E}_P[y_i\mathbb{1}\{z_i \le x\}]\,\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}]\,\mathrm{d}\nu(x)\,(\theta_0 - \varepsilon) \\ &\quad + \int_{\mathbb{R}} (\mathbb{E}_P[Y_i\mathbb{1}\{z_i \le x\}])^2 \mathrm{d}\nu(x)\,(\theta_0 - \varepsilon)^2\end{aligned}$$

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

$$
\begin{aligned}
&= 2\int_{\mathbb{R}} \mathbb{E}_P\left[y_i \mathbb{1}\left\{z_i \le x\right\}\right] \mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right] \mathrm{d}\nu\left(x\right) \varepsilon \\
&\quad - 2\int_{\mathbb{R}} \left(\mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right]\right)^2 \mathrm{d}\nu\left(x\right) \theta_0 \varepsilon \\
&\quad + \int_{\mathbb{R}} \left(\mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right]\right)^2 \mathrm{d}\nu\left(x\right) \varepsilon^2 \\
&= \int_{\mathbb{R}} \left(\mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right]\right)^2 \mathrm{d}\nu\left(x\right) \varepsilon^2.
\end{aligned}
$$

This implies that

$$
\left\{\int_{\mathbb{R}} \phi_P\left(x, \theta_0 - \varepsilon\right)^2 \mathrm{d}\nu\left(x\right)\right\}^{1/2} = \left\{\int_{\mathbb{R}} \left(\mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right]\right)^2 \mathrm{d}\nu\left(x\right)\right\}^{1/2} \varepsilon.
$$

In this case, Assumptions 2.6 and 2.7 hold. The asymptotic limit of the test statistic is

$$
\mathcal{L}''_{\phi_P}\left(\mathbb{G}_0\right) = \inf_{v \in \mathbb{R}} \int_{\mathbb{R}} \left(\mathbb{G}_0\left(x, \theta_0\right) - \mathbb{E}_P\left[Y_i \mathbb{1}\left\{z_i \le x\right\}\right] v\right)^2 \mathrm{d}\nu\left(x\right). \tag{13}
$$

Theorem 2.1(i) requires that the CDF of $\mathcal{L}''_{\phi_P}\left(\mathbb{G}_0\right)$ in (13) is strictly increasing and continuous at its $1-\alpha$ quantile.

The sample size is set to $n \in \{100, 200, 400, 800\}$. We set the tuning parameter τ_n as $\tau_n = \sqrt{\ln(n)/n}$, $n^{-2/5}$, $n^{-1/3}$, $n^{-1/4}$, $n^{-1/5}$, and $n^{-1/6}$, which all satisfy Assumption 2.8. For dependent data, the MBB involves an additional tuning parameter $b(n)$. We set $b(n) = n^{1/6}$, $n^{1/5}$, $n^{1/4}$, and $n^{1/3}$. Recall that the test statistic involves an integration with respect to a measure ν and an infimum. The integration is approximated by an equally weighted average on the grid $\{-3, -2.998, -2.996, \ldots, 3\}$ of x, and the infimum is achieved by a search on the grid $\{0.7, 0.702, 0.704, \ldots, 1.3\}$ of θ. Furthermore, we apply the warp-speed method (Giacomini, Politis, and White, 2013) to implement all the Monte Carlo experiments. Specifically, for each DGP and sample size, we generate 1,000 samples and compute one original statistic $n\mathcal{L}(\widehat{\phi}_n)$ and one bootstrap statistic $\widehat{\mathcal{L}}''_n[\sqrt{n}(\widehat{\phi}^*_n - \widehat{\phi}_n)]$ for each sample. The critical value $\widehat{c}_{1-\alpha,n}$ is approximated by the $(1-\alpha)$-empirical quantile of the 1,000 bootstrap statistics, and the rejection rate is computed by comparing the 1,000 original statistics with the critical value $\widehat{c}_{1-\alpha,n}$.

We present some main simulation results in the following and leave the remaining results to Section E of the Supplementary Material. Tables 1–6 and Tables E.12–E.22 in the Supplementary Material show the rejection rates for different DGPs, tuning parameters, and nominal significance levels with measure ν being the probability measure of $\mathcal{N}(0, 10^2)$. Tables E.1–E.10 in the Supplementary Material display the rejection rates for Case 1 with the measure ν being the probability measure of $\mathcal{N}(0,1)$ or $\mathcal{N}(0, 5^2)$. The results are stable for different choices of τ_n, $b(n)$, and ν. Most of the rejection rates under the null are close to the nominal significance levels. The rejection rates under the alternatives increase to

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

TABLE 1. Size for Case 1 with i.i.d. data

α	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
0.01	100	0.011	0.007	0.011	0.011	0.011	0.011
	200	0.004	0.003	0.004	0.006	0.010	0.011
	400	0.003	0.003	0.003	0.005	0.006	0.006
	800	0.008	0.008	0.008	0.014	0.014	0.014
0.025	100	0.026	0.017	0.026	0.027	0.026	0.027
	200	0.020	0.015	0.020	0.022	0.025	0.026
	400	0.022	0.021	0.022	0.022	0.023	0.023
	800	0.019	0.016	0.023	0.026	0.026	0.026
0.05	100	0.043	0.038	0.043	0.051	0.054	0.054
	200	0.040	0.035	0.041	0.046	0.050	0.051
	400	0.058	0.044	0.067	0.069	0.062	0.067
	800	0.052	0.046	0.052	0.069	0.074	0.076
0.1	100	0.101	0.090	0.101	0.111	0.111	0.111
	200	0.098	0.091	0.103	0.110	0.110	0.111
	400	0.109	0.101	0.114	0.124	0.130	0.131
	800	0.128	0.112	0.136	0.127	0.133	0.137
0.2	100	0.219	0.209	0.219	0.241	0.244	0.244
	200	0.213	0.198	0.213	0.228	0.238	0.247
	400	0.219	0.206	0.229	0.235	0.238	0.241
	800	0.238	0.215	0.240	0.248	0.255	0.255

one as the sample size n increases. For dependent samples, the rejection rates under the null may exceed the significance level α for some τ_n, $b(n)$, and ν as shown, for example, in Table 3. As we increase the sample sizes, the results become closer to α, as shown in Table E.11 in the Supplementary Material.

4.1. Performance Improvement in Conditional Moment Restriction Models with Weak Instruments

Note that Cases 1 and 2 ($n = 200$) with $\delta = 0$ (under the null hypothesis) are identical to the DGPs in Tables 1 and 2 of Jun and Pinkse (2009), respectively. Thus, the results in Table 1 with $n = 100$ and Table E.12 in the Supplementary Material with $n = 200$ can be compared with those in Tables 1 and 2 of Jun and Pinkse (2009), respectively. We present the comparisons in Tables 7 and 8, where $\widehat{T}_2(\widehat{\theta}_*)$ and $\widehat{T}_{\rm k}(\widehat{\theta}_*)$ with $\widehat{\theta}_* \in \{\widehat{\theta}_{\rm 2SLS}, \widehat{\theta}_{\rm SP}\}$ are two-step plug-in test statistics computed by using either a 2SLS or a semi-parametric estimator of θ_0 as described in Jun and Pinkse (2009), and $\widehat{T}_1(\widehat{\theta}_{\rm CUE1})$, $\widehat{T}_2(\widehat{\theta}_{\rm CUE2})$, and $\widehat{T}_{\rm k}(\widehat{\theta}_{\rm CUEk})$ are the test

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

TABLE 2. Power for Case 1 with i.i.d. data ($\alpha = 0.05$)

DGP	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
DGP (1)	100	0.245	0.184	0.246	0.313	0.345	0.373
	200	0.440	0.362	0.460	0.573	0.623	0.638
	400	0.679	0.583	0.709	0.820	0.850	0.860
	800	0.888	0.822	0.924	0.976	0.990	0.992
DGP (2)	100	0.865	0.797	0.866	0.926	0.949	0.956
	200	0.997	0.986	0.997	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
DGP (3)	100	0.992	0.983	0.992	0.999	0.999	0.999
	200	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 3. Size for Case 1 with dependent data ($\alpha = 0.05$)

$b(n)$	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
$n^{1/6}$	100	0.037	0.028	0.037	0.046	0.050	0.052
	200	0.050	0.037	0.054	0.058	0.067	0.069
	400	0.071	0.065	0.072	0.075	0.080	0.079
	800	0.060	0.050	0.068	0.077	0.082	0.083
$n^{1/5}$	100	0.037	0.029	0.037	0.045	0.046	0.047
	200	0.035	0.029	0.037	0.040	0.045	0.049
	400	0.071	0.065	0.072	0.075	0.080	0.079
	800	0.047	0.045	0.064	0.078	0.081	0.085
$n^{1/4}$	100	0.037	0.029	0.037	0.045	0.046	0.047
	200	0.039	0.035	0.044	0.054	0.058	0.061
	400	0.067	0.058	0.068	0.072	0.072	0.072
	800	0.083	0.072	0.088	0.097	0.097	0.100
$n^{1/3}$	100	0.056	0.046	0.057	0.065	0.070	0.072
	200	0.047	0.037	0.049	0.055	0.059	0.064
	400	0.067	0.058	0.067	0.069	0.074	0.075
	800	0.057	0.036	0.071	0.078	0.085	0.084

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

TABLE 4. Power for DGP (1) of Case 1 with dependent data ($\alpha = 0.05$)

$b(n)$	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
$n^{1/6}$	100	0.317	0.249	0.318	0.387	0.410	0.433
	200	0.520	0.393	0.547	0.655	0.683	0.697
	400	0.759	0.671	0.804	0.895	0.915	0.924
	800	0.988	0.964	0.992	1.000	1.000	1.000
$n^{1/5}$	100	0.257	0.206	0.258	0.333	0.356	0.380
	200	0.482	0.368	0.509	0.617	0.673	0.686
	400	0.759	0.671	0.804	0.895	0.915	0.924
	800	0.990	0.969	0.992	1.000	1.000	1.000
$n^{1/4}$	100	0.257	0.206	0.258	0.333	0.356	0.380
	200	0.547	0.421	0.569	0.680	0.688	0.703
	400	0.757	0.651	0.797	0.892	0.919	0.927
	800	0.987	0.963	0.992	1.000	1.000	1.000
$n^{1/3}$	100	0.263	0.176	0.264	0.331	0.364	0.370
	200	0.486	0.381	0.507	0.632	0.672	0.688
	400	0.749	0.645	0.775	0.883	0.916	0.922
	800	0.978	0.950	0.988	1.000	1.000	1.000

TABLE 5. Power for DGP (2) of Case 1 with dependent data ($\alpha = 0.05$)

$b(n)$	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
$n^{1/6}$	100	0.976	0.923	0.976	0.989	0.992	0.993
	200	1.000	0.998	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/5}$	100	0.966	0.920	0.966	0.989	0.992	0.993
	200	1.000	0.999	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/4}$	100	0.966	0.920	0.966	0.989	0.992	0.993
	200	1.000	0.999	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/3}$	100	0.961	0.913	0.961	0.986	0.991	0.992
	200	1.000	0.999	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

TABLE 6. Power for DGP (3) of Case 1 with dependent data ($\alpha = 0.05$)

$b(n)$	n	τ_n					
		$\sqrt{\ln(n)/n}$	$n^{-2/5}$	$n^{-1/3}$	$n^{-1/4}$	$n^{-1/5}$	$n^{-1/6}$
$n^{1/6}$	100	0.999	0.995	0.999	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/5}$	100	0.999	0.994	0.999	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/4}$	100	0.999	0.994	0.999	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000
$n^{1/3}$	100	0.999	0.995	0.999	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000	1.000	1.000
	400	1.000	1.000	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 7. Comparison with Table 1 of Jun and Pinkse (2009)

α	Plug-in		Jun and Pinkse (2009)			Proposed test
	$\widehat{T}_2(\widehat{\theta}_{\text{2SLS}})$	$\widehat{T}_{\text{k}}(\widehat{\theta}_{\text{2SLS}})$	$\widehat{T}_1(\widehat{\theta}_{\text{CUE1}})$	$\widehat{T}_2(\widehat{\theta}_{\text{CUE2}})$	$\widehat{T}_{\text{k}}(\widehat{\theta}_{\text{CUEk}})$	$\tau_n = n^{-1/4}$
0.01	0.511	0.480	0.004	0.012	0.012	0.011
0.025	0.533	0.509	0.007	0.016	0.022	0.027
0.05	0.551	0.531	0.015	0.025	0.030	0.051
0.1	0.584	0.559	0.027	0.049	0.045	0.111
0.2	0.626	0.602	0.053	0.078	0.075	0.241

statistics proposed by Jun and Pinkse (2009).[9] Our test uses $\tau_n = n^{-1/4}$ for illustration. The plug-in method suffers from substantial size distortion. The tests of Jun and Pinkse (2009) improve upon the plug-in approach, but could be conservative as shown in their theoretical results. The proposed method achieves rejection rates closer to the nominal significance levels compared to the results of Jun and Pinkse (2009). These numerical observations provide supporting evidence for the theoretical results in the article.

[9]The function $\widehat{T}_{\text{k}}(\cdot)$ is proposed by Zheng (1996), and the test statistic $\widehat{T}_{\text{k}}(\widehat{\theta}_{\text{CUEk}})$ based on minimization of $\widehat{T}_{\text{k}}(\cdot)$ follows the idea of Jun and Pinkse (2009).

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

TABLE 8. Comparison with Table 2 of Jun and Pinkse (2009)

α	Plug-in		Jun and Pinkse (2009)			Proposed test
	$\widehat{T}_2(\widehat{\theta}_{\mathrm{SP}})$	$\widehat{T}_{\mathrm{k}}(\widehat{\theta}_{\mathrm{SP}})$	$\widehat{T}_1(\widehat{\theta}_{\mathrm{CUE1}})$	$\widehat{T}_2(\widehat{\theta}_{\mathrm{CUE2}})$	$\widehat{T}_{\mathrm{k}}(\widehat{\theta}_{\mathrm{CUEk}})$	$\tau_n = n^{-1/4}$
0.01	0.341	0.362	0.018	0.018	0.024	0.009
0.025	0.360	0.395	0.027	0.027	0.030	0.023
0.05	0.382	0.419	0.036	0.036	0.046	0.049
0.1	0.422	0.455	0.052	0.056	0.067	0.109
0.2	0.487	0.519	0.077	0.093	0.106	0.233

5. CONCLUSION

This article provides a unified framework for inference on moment restriction models with nuisance parameters. We employ a new characterization that does not require the estimation of nuisance parameters, along with a numerical delta method to construct the test. The test is asymptotically size controlled and consistent. We conduct extensive Monte Carlo simulations to illustrate the finite sample properties of the proposed test. The numerical results show that the proposed method may achieve improvement in testing conditional moment restriction models with weak instruments. While the present article focuses primarily on models with finite-dimensional nuisance parameters, the proposed approach might be extended to moment restrictions involving infinite-dimensional nuisance parameters that have been widely investigated in the literature. This framework may also be applied to other inference problems in the presence of nuisance parameters.

SUPPLEMENTARY MATERIAL

Li, X., Song, X., & Sun, Z. (2026). Supplement to "Unified inference on moment restrictions with nuisance parameters," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466626100528.

FUNDING STATEMENT

The work of X.S. was supported by the National Natural Science Foundation of China (Grant Nos. 72373007 and 72333001). The author also gratefully acknowledges the research support from the Center for Statistical Science of Peking University, China, and the Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University) of the Ministry of Education, China. The work of Z.S. was supported by the National Natural Science Foundation of China (Grant No. 72103004).

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

COMPETING INTERESTS

The authors declare that no competing interests exist.

REFERENCES

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

Arcones, M. A., & Yu, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability*, 7(1), 47–71.

Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā: The Indian Journal of Statistics, Series A*, 46(1), 85–93.

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

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

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

Berger, Y. G. (2022). Testing conditional moment restriction models using empirical likelihood. *The Econometrics Journal*, 25(2), 384–403.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1), 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58(6), 1443–1458.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305–334.

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

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

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

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

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

Delgado, M. A., Domínguez, M. A., & Lavergne, P. (2006). Consistent tests of conditional moment restrictions. *Annales d'Économie et de Statistique*, 81, 33–67.

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

Domínguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5), 1601–1615.

Domínguez, M. A., & Lobato, I. N. (2015). A simple omnibus overidentification specification test for time series econometric models. *Econometric Theory*, 31(4), 891–910.

Donald, S. G., Imbens, G. W., & Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117(1), 55–93.

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

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics*, 1(2), 279–290.

Durbin, J. (1985). The first-passage density of a continuous Gaussian process to a general boundary. *Journal of Applied Probability*, 22(1), 99–122.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Escanciano, J. C., & Goh, S.-C. (2014). Specification analysis of linear quantile models. *Journal of Econometrics*, 178(3), 495–507.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, 64(4), 865–890.

Fan, Y., & Li, Q. (2000). Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory*, 16(6), 1016–1041.

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

Genest, C., & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'I.H.P. Probabilités et statistiques*, 44(6), 1096–1127.

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

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

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

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

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

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

Hong, H., & Li, J. (2020). The numerical bootstrap. *Annals of Statistics*, 48(1), 397–412.

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

Jun, S. J., & Pinkse, J. (2009). Semiparametric tests of conditional moment restrictions under weak or partial identification. *Journal of Econometrics*, 152(1), 3–18.

Jun, S. J., & Pinkse, J. (2012). Testing under weak identification with conditional moment restrictions. *Econometric Theory*, 28(6), 1229–1282.

Khmaladze, E. V. (1982). Martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications*, 26(2), 240–257.

Kitamura, Y., Tripathi, G., & Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica*, 72(6), 1667–1714.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.

Kunitomo, N., McAleer, M., & Nishiyama, Y. (2011). Moment restriction-based econometric methods: An overview. *Journal of Econometrics*, 165(1), 1–4.

Lavergne, P., & Patilea, V. (2013). Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory. *Journal of Econometrics*, 177(1), 47–59.

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer.

Muandet, K., Jitkrittum, W., & Kübler, J. (2020). Kernel conditional moment test via maximum moment restriction. In J. Peters, & D. Sontag (Eds.), *Proceedings of the 36th conference on uncertainty in artificial intelligence (UAI)*, volume 124 of Proceedings of Machine Learning Research. (pp. 41–50) PMLR.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics*, 29(3), 229–256.

Newey, W. K. (1993). 16 Efficient estimation of models with conditional moment restrictions. In *Econometrics*, volume 11 of Handbook of Statistics (pp. 419–454). Elsevier.

Parker, T. (2013). A comparison of alternative approaches to supremum-norm goodness-of-fit tests with estimated parameters. *Econometric Theory*, 29(5), 969–1008.

Psaradakis, Z. (2016). Using the bootstrap to test for symmetry under unknown dependence. *Journal of Business & Economic Statistics*, 34(3), 406–415.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press

Psaradakis, Z., & Vávra, M. (2015). A quantile-based test for symmetry of weakly dependent processes. *Journal of Time Series Analysis*, 36(4), 587–598.
Psaradakis, Z., & Vávra, M. (2019). Bootstrap-assisted tests of symmetry for dependent data. *Journal of Statistical Computation and Simulation*, 89(7), 1203–1226.
Psaradakis, Z., & Vávra, M. (2022). Using triples to assess symmetry under weak dependence. *Journal of Business & Economic Statistics*, 40(4), 1538–1551.
Radulović, D. (1996). The bootstrap for empirical processes based on stationary observations. *Stochastic Processes and their Applications*, 65(2), 259–279.
Seo, J. (2018). Tests of stochastic monotonicity with improved power. *Journal of Econometrics*, 207(1), 53–70.
Smith, R. J. (2007). Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics*, 138(2), 430–460.
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.
Stock, J. H., & Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5), 1055–1096.
Sun, Z. (2023). Instrument validity for heterogeneous causal effects. *Journal of Econometrics*, 237(2), 105523.
Sun, Z., & Beare, B. K. (2021). Improved nonparametric bootstrap tests of Lorenz dominance. *Journal of Business & Economic Statistics*, 39(1), 189–199.
Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 30(1–2), 415–443.
Tripathi, G., & Kitamura, Y. (2003). Testing conditional moment restrictions. *Annals of Statistics*, 31(6), 2059–2095.
van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer.
Volkonskii, V. A., & Rozanov, Y. A. (1959). Some limit theorems for random functions. I. *Theory of Probability & Its Applications*, 4(2), 178–197.
Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2), 263–289.

https://doi.org/10.1017/S0266466626100528 Published online by Cambridge University Press