Econometric Theory, **40**, 2024, 827–858.
doi:10.1017/S0266466622000615

TESTING A CLASS OF SEMI- OR NONPARAMETRIC CONDITIONAL MOMENT RESTRICTION MODELS USING SERIES METHODS

JESPER RIIS-VESTERGAARD SØRENSEN
University of Copenhagen

This paper proposes a new test for a class of conditional moment restrictions (CMRs) whose parameterization involves unknown, unrestricted conditional expectation functions. Motivating examples of such CMRs arise from models of discrete choice under uncertainty including certain static games of incomplete information. The proposed test may be viewed as a semi-/nonparametric extension of the Bierens (1982, *Journal of Econometrics* 20, 105–134) goodness-of-fit test of a parametric model for the conditional mean. Estimating conditional expectations using series methods and employing a Gaussian multiplier bootstrap to obtain critical values, the test is shown to be asymptotically correctly sized and consistent. Simulation studies indicate good finite-sample properties. In an empirical application, the test is used to study the validity of a game-theoretical model for discount store market entry, treating equilibrium beliefs as nonparametric conditional expectations. The test indicates that Walmart and Kmart entry decisions do not result from a static discrete game of incomplete information with linearly specified profits.

1. INTRODUCTION

Econometric models are often stated in terms of conditional moment restrictions (CMRs) in which a known function of observables as well as both finite- and infinite-dimensional model parameters is said to have zero mean given known conditioning variables. In many instances, the infinite-dimensional part of the model parameterization corresponds to one or more conditional expectation functions (CEFs). Examples are random-effects dynamic binary-response models for panel data (Honoré and Kyriazidou, 2000; Honoré and Tamer, 2006), single-agent discrete choice under uncertainty (Manski, 1991; Ahn and Manski, 1993), and static games of incomplete information (Bajari et al., 2010; Ellickson and Misra, 2011). Specifically, in random-effects dynamic binary-response models for

Parts of this paper derive from my doctoral dissertation, completed under the guidance and encouragement of Denis Chetverikov and Jinyong Hahn. I thank Andres Santos, Rasmus Søndergaard Pedersen, the Editor, and two anonymous referees for highly constructive comments that helped improve this paper. All remaining errors are my own. Address correspondence to Jesper Riis-Vestergaard Sørensen, Department of Economics, University of Copenhagen, Copenhagen, Denmark; e-mail: jrvs@econ.ku.dk.

panel data, CEFs arise from the initial conditions problem. In a model of discrete choice under uncertainty, CEFs are introduced via the model assumption that the agents' beliefs are correct in the aggregate—a rational expectations hypothesis. Moreover, in static games of incomplete information, CEFs appear under the model assumption that beliefs are correct in a Bayesian Nash equilibrium (BNE).

In this paper, we propose a general method for constructing omnibus specification tests for a wide class of semi-/nonparametric CMR models pararameterized in part by CEFs. More precisely, the paper aims to test the model assertion that there exists a finite-dimensional parameter β such that

$$\mathrm{E}[\rho_\ell(Z,\beta,\mathrm{E}[Y_\ell|W_\ell])|X_\ell] = 0 \text{ almost surely } X_\ell \text{ for all } \ell \in \{1,\ldots,L\}, \qquad \textbf{(1.1)}$$

where the ρ_ℓ's are known (residual) functions, each $\mathrm{E}[Y_\ell|W_\ell]$ is an unrestricted, possibly vector, conditional expectation, each W_ℓ is a subvector of the conditioning variables X_ℓ, and Z denotes all model observables (i.e., the union of distinct elements of the X_ℓ's and Y_ℓ's). The alternative hypothesis is that (1.1) is violated.

The proposed test is an extension of the Bierens (1982) test given in the context of parametric mean regression. The idea of Bierens's method is to recast a CMR as a collection of testable *un*conditional moment restrictions (UMRs), which are then suitably integrated (or otherwise aggregated). Within the context of parametric mean regression, a test of correct specification may be obtained by checking whether the least-squares residuals correlate with any member of a suitably rich family of transformations of the regressors. Bierens's idea carries over to any setting where one may speak of model residuals, including (1.1).[1]

The suggested test statistic is a Cramér–von Mises-type (CM-type) measure of distance between the collection of residual-to-transformation correlations and zero. One rejects the null hypothesis that the semi-/nonparametric model in (1.1) is correctly specified whenever said distance is "unreasonably" large. Under the null hypothesis, the proposed test statistic has a nonpivotal limiting distribution and therefore cannot be tabulated. We propose and formally justify the use of a multiplier bootstrap procedure for obtaining critical values. Calculation of the test statistic and critical values requires estimation of CEFs. These are here estimated using series methods and therefore boil down to linear regressions.

The resulting test is shown to have attractive theoretical properties in that it is both asymptotically of correct size and consistent against any fixed alternative. To illustrate these properties, we implement our procedure in a comprehensive simulation study testing the specification of a binary-action simultaneous-move game of incomplete information. The simulations by and large reproduce the asymptotic guarantees in relatively small samples. We also apply the specification test

[1] An alternative approach not pursued in this paper estimates the model under both the null and alternative and contrasts the estimates according to some notion of distance (see, e.g., Härdle and Mammen, 1993; Zheng, 1996; Kristensen, 2011). The Bierens approach is convenient in that it only requires estimation of the (potentially substantially) simpler null model. However, the two approaches cannot be ranked in terms of their local power properties and should be viewed as complementary (Fan and Li, 2000).

developed in this paper to an entry game between Walmart and Kmart discount stores using the Jia (2008) dataset. All implementations of the test considered result in bootstrap p-values less than 1%, which indicates that Walmart and Kmart entry decisions do not result from a simple static discrete game with linearly specified profits.

Originating from Bierens (1982), there exists a vast literature on omnibus specification testing (i.e., tests consistent against any violation of the null hypothesis) of both parametric, semiparametric and nonparametric models of features of i.i.d. data. We here focus on Li, Hsiao, and Zinn (2003), Song (2010), Bravo (2012), and the recent working paper Lapenta (2021), all of whom develop test statistics for a class of semiparametric CMRs similar to (1.1), and to which this paper is closest in content.

Lapenta (2021) provides a specification test for semiparametric models where some variables are nonparametrically generated and studies a moment condition involving a residual from a semiparametric regression with nonparametrically generated variables. While the model structure is here different, in the present paper, the unobserved but identified conditional expectations $\mathrm{E}[Y_{\ell i}|W_{\ell i}]$ are similarly replaced with nonparametric estimates, and can therefore be considered as nonparametrically generated variables. Loosely speaking, in Lapenta's paper, the residual function is itself nonparametric, whereas in the present work, each ρ_ℓ has a known form. The bootstrap provided in Lapenta (2021) requires estimating the finite-dimensional parameter at every bootstrap iteration, which might be computationally demanding. In contrast, the bootstrap provided here is computationally convenient, in that it only involves a single round of estimation.

Li et al. (2003) provide a specification test for a semiparametric additive partially linear model using a series approach. The main departures from their paper here lie in the presence of generated variables and different methods of proofs. In Li et al. (2003), the nonparametric components are identified within the moment condition that is tested and, thus, cannot be considered as nonparametrically generated variables. In comparison, in the present work, the nonparametric CEFs are identified using auxiliary moment conditions, and may be viewed as generated variables. Due to these differences, the Bahadur expansions of the empirical processes at the basis of the test statistic are different from the one obtained in Li et al. (2003). Another difference stands in the method of proofs, as the present work is based on empirical process theory and employs maximal inequalities based on bracketing entropy.

Song (2010) confines interest to the case where the nonparametric part of the parameterization takes a composite-index form. His treatment of the nonparametric part rules out unrestricted CEFs, but does allow for single-index models not nested in (1.1). Song's framework is thus neither more nor less general. Unlike the nonpivotal test statistic proposed in this paper, Song (2010) uses a conditional martingale transform to obtain an asymptotically distribution-free test statistic, thus allowing for tabulation of critical values. However, since the martingale transform is generally unknown, pivotality comes at the cost of additional steps of

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

nonparametric estimation. In addition, as indicated by Song's simulation studies (and remarked by Song), the martingale transform approach appears more sensitive to the choice of tuning parameters than the bootstrap—the latter approach being the one taken in this paper.

Bravo (2012) uses a generalized empirical likelihood approach to obtain specification tests similar in spirit to classical Kolmogorov–Smirnov (KS) and Cramér–von Mises goodness-of-fit statistics. As in this paper, Bravo's test statistic has a nonpivotal limit distribution and a multiplier bootstrap procedure is used to obtain critical values. His framework is broader than (1.1) in that the residual function may depend on arbitrary nonparametric components and in a functional manner. Naturally, Bravo's greater generality comes at the cost of relatively abstract conditions. Specifically, Bravo's treatment implicitly presumes that the adjustments terms necessary to account for parametric and nonparametric estimation have already been derived. (See also Remark 1.) These adjustments must be consistently estimated in order to obtain valid critical values and therefore constitute crucial elements of the implementation of his test. In contrast, by restricting attention to a simpler setting with nonparametric CEFs (a type of mean-square projection), we obtain the necessary adjustments for parametric and nonparametric estimation in closed form (in (3.6)–(3.8)) under relatively primitive conditions, such as (ordinary) differentiability, using arguments familiar from the two-step generalized method of moments (GMM) literature with a nonparametric first step (see, e.g., Newey, 1994). We also provide explicit estimators thereof (in (3.20)–(3.22)) and establish the validity of the resulting bootstrap procedure (Lemma 3). The added CEF structure also allows us to tailor our assumptions to the nonparametric method of estimation, here chosen to be series estimation.

The remainder of this paper is organized as follows. We define the testing problem and test statistic in Section 2. In Section 3, we analyze the limiting behavior of the test statistic, construct critical values, and derive the limiting properties of the resulting test, in turn. We conduct simulation studies in Section 4 and give an empirical illustration in Section 5. Section 6 concludes. Proofs can be found in the Appendix with further discussion, proof details, and supporting lemmas provided in the Supplementary Material.

2. TESTING SEMI-/NONPARAMETRIC CMRS

Let $\{Z_i\}_1^n$ be n independent copies of Z, such that Z_i is a random element of $\mathbf{R}^{d_z}$ composed of the distinct elements of $X_{\ell i}$, thus subsuming $W_{\ell i}$,[2] and $Y_{\ell i}, \ell \in \{1,\ldots,L\}$. The support of Z is denoted by $\mathcal{Z}$ and that of X_ℓ by $\mathcal{X}_\ell \subseteq \mathbf{R}^{d_{x,\ell}}$. Let $\mathcal{B} \subseteq \mathbf{R}^{d_\beta}$ be a finite-dimensional parameter space.

[2]Here, W_ℓ need not be a literal subvector of X_ℓ; only X_ℓ-measurability is required.

2.1. Testing Problem

The *null hypothesis* (H_0) we wish to test is

$$H_0 : \text{For some } \beta \in \mathcal{B},\ \mathrm{E}\left[\rho_\ell\left(Z,\beta,h_\ell^*(W_\ell)\right)\middle|X_\ell\right]=0 \text{ a.s. } X_\ell \text{ for all } \ell \in \{1,\ldots,L\}, \tag{2.1}$$

where "a.s." connotes "almost surely," and we have abbreviated $h_\ell^*(W_\ell) :=\mathrm{E}[Y_\ell|W_\ell]$ for convenience. The null is pitted against the general *alternative hypothesis* (H_1),

$$H_1 : \text{For all } \beta \in \mathcal{B},\ \mathrm{P}\left(\mathrm{E}\left[\rho_\ell\left(Z,\beta,h_\ell^*(W_\ell)\right)\middle|X_\ell\right]=0\right) < 1 \text{ for some } \ell \in \{1,\ldots,L\}, \tag{2.2}$$

under regularity conditions presented below. In this paper, we propose a procedure for testing (2.1) versus (2.2) assuming the existence of some $\beta_0 \in \mathcal{B}$ such that (a) β_0 is regularly estimable (in the sense of Assumption 1) and (b) equation (2.1) is satisfied at β_0 under H_0. Due to property (b), β_0 will be referred to as *pseudo-true*.[3]

2.2. Recasting the Problem

The presence of a pseudo-truth implies that H_0 may be equivalently stated as

$$H_0 : \mathrm{E}\left[\rho_\ell\left(Z,\beta_0,h_\ell^*(W_\ell)\right)\middle|X_\ell\right]=0 \text{ a.s. } X_\ell \text{ for all } \ell \in \{1,\ldots,L\}. \tag{2.3}$$

Suppose for now that $L=1$, abbreviate $U := \rho_1\left(Z,\beta_0,h_1^*(W_1)\right)$, and drop the ℓ subscripts. Then H_0 means

$$\mathrm{E}[U|X]=0 \text{ a.s.}, \tag{2.4}$$

which holds if and only if $\mathrm{E}[Ug(X)]=0$ for all "test functions" $g \in L^\infty(\mathcal{X}), L^\infty(\mathcal{X})$ denoting the space of bounded functions $g : \mathcal{X} \to \mathbf{R}$. Following Bierens and Ploberger (1997), Stute (1997), and Stinchcombe and White (1998), among others, we construct a test of the CMR in (2.3) by testing the UMRs

$$\mathrm{E}[U\omega(t,X)]=0 \quad \text{for } F_X\text{-a.e. } t \in \mathcal{X}, \tag{2.5}$$

where "a.e." connotes "almost every," and $\{g=\omega(t,\cdot);t\in\mathcal{X}\}$ denotes a weight function family chosen so as to make (2.4) and (2.5) equivalent.[4] Examples include the exponential $\omega(t,x)=\exp(t'x)$, logistic $\omega(t,x)=1/[1+\exp(c-t'x)]$ with $c\neq 0$, and cosine–sine $\omega(t,x)=\cos(t'x)+\sin(t'x)$ families.[5]

[3] In the context of the two-player, binary-action static games of incomplete information of Sections 4 and 5, expected (pseudo) log-likelihood is presumed to have a unique maximizer (β_0), and the parameters are estimated using two-step (pseudo) maximum likelihood.

[4] We use F_V to denote both V's distribution and cumulative distribution function (CDF).

[5] See Bierens and Ploberger (1997) (with its Bierens (2017) addendum) and Stinchcombe and White (1998) for detailed guidance on this choice. Strictly speaking, exponential weighting requires X bounded in order to ensure $\omega(t,\cdot)$ bounded. However, for unbounded X, one may replace X with any bounded, one-to-one transformation thereof. See also the discussion following Assumption 2.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

In general, one chooses a weight function ω_ℓ for each ℓ. Then H_0 can be written as

$$\mathrm{H}_0 : M_\ell(t_\ell) = 0 \quad \text{for } F_{X_\ell}\text{-a.e. } t_\ell \in \mathcal{X}_\ell \text{ and all } \ell \in \{1,\ldots,L\}, \tag{2.6}$$

with $M_\ell : \mathcal{X}_\ell \to \mathbf{R}$ defined by $M_\ell(t_\ell) := \mathrm{E}[\rho_\ell(Z,\beta_0,h_\ell^*(W_\ell))\omega_\ell(t_\ell,X_\ell)]$. Aggregating the UMRs involved, we may express H_0 and H_1 as

$$\mathrm{H}_0 : \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} M_\ell(t_\ell)^2\,\mathrm{d}F_{X_\ell}(t_\ell) = 0, \tag{2.7}$$

$$\mathrm{H}_1 : \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} M_\ell(t_\ell)^2\,\mathrm{d}F_{X_\ell}(t_\ell) > 0. \tag{2.8}$$

This representation suggests the CM-type of statistic[6]

$$T_n := n\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} \widehat{M}_\ell(t_\ell)^2\,\mathrm{d}\widehat{F}_{X_\ell}(t_\ell) = \sum_{\ell=1}^{L}\sum_{i=1}^{n}\widehat{M}_\ell(X_{\ell i})^2, \tag{2.9}$$

where $\widehat{F}_{X_\ell}$ is the empirical distribution,

$$\widehat{F}_{X_\ell}(t_\ell) := \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_{\ell i} \leqslant t_\ell), \quad \ell \in \{1,\ldots,L\}, \tag{2.10}$$

and we estimate each M_ℓ by the plug-in method

$$\widehat{M}_\ell(t_\ell) := \frac{1}{n}\sum_{i=1}^{n}\rho_\ell(Z_i,\widehat{\beta},\widehat{h}_\ell(W_{\ell i}))\omega_\ell(t_\ell,X_{\ell i}), \quad \ell \in \{1,\ldots,L\}. \tag{2.11}$$

Formal requirements of the estimate $\widehat{\beta}$ of β_0 are stated in Assumption 1. Each coordinate m of $\widehat{h}_\ell$ is a series estimator

$$\widehat{h}_{\ell m}(w_\ell) := \widehat{h}_{\ell m,k_{\ell m,n}}(w_\ell) := p_\ell^{k_{\ell m,n}}(w_\ell)'\,\widehat{\pi}_{\ell m,k_{\ell m,n}},$$
$$\widehat{\pi}_{\ell m,k} := \left(\frac{1}{n}\sum_{i=1}^{n}p_\ell^k(W_{\ell i})p_\ell^k(W_{\ell i})'\right)^{-}\left(\frac{1}{n}\sum_{i=1}^{n}p^k(W_{\ell i})\,Y_{\ell i m}\right),$$

where $p_\ell^k(w_\ell) := (p_{\ell 1}(w_\ell),\ldots,p_{\ell k}(w_\ell))'$ is based on a known dictionary $\{p_{\ell j}\}_1^k$ of approximating functions, each $k_{\ell m,n}$ is a sequence of positive integers growing without bound as $n \to \infty$, and $(\cdot)^-$ indicates Moore–Penrose inversion.[7]

[6]One could alternatively use a KS-type statistic or base a test statistic on empirical likelihood as in Bravo (2012) without affecting the limiting properties of the test. See Section S.1 of the Supplementary Material for details.

[7]The approximating functions could depend on m or even k, neither of which is reflected in our notation. Under assumptions invoked in this paper, the matrix $n^{-1}\sum_{i=1}^{n}p_\ell^{k_{\ell m,n}}(W_{\ell i})p_\ell^{k_{\ell m,n}}(W_{\ell i})'$ is nonsingular with probability approaching one and the particular generalized inverse therefore asymptotically irrelevant. Detailed accounts of the properties of least-squares series estimators can be found in Newey (1995, 1997), Chen (2007), and Belloni et al. (2015).

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

Under general conditions presented in Section 3, the stochastic processes $\{\widehat{M}_\ell\}_1^L$ all converge to the zero function in probability under H_0, whereas at least one of them converges to a nonzero probability limit under H_1. A "large" realization of T_n thus telegraphs a violation of H_0.

3. THEORETICAL PROPERTIES

In this section, we first establish the limiting behavior of the test statistic T_n (Section 3.1). Its limiting null distribution is generally nonpivotal and its dependence on the data-generating process (DGP) involved. To tackle this problem, we next introduce a multiplier bootstrap procedure and establish its asymptotic validity (Section 3.2). These findings translate into limiting behavior of the resulting test (Section 3.3).

3.1. Limiting Behavior of Test Statistic

Some regularity is required to derive the limiting behavior of the test statistic. To control the influence of estimation of β_0 on this limiting behavior, we invoke the following assumption.

Assumption 1 (Parametric estimation). For each $n \in \mathbf{N}, \widehat{\beta}$ is a random element of $\mathcal{B}$, a subset of $\mathbf{R}^{d_\beta}$ with nonempty interior, and, under both H_0 and $\mathrm{H}_1, \widehat{\beta} \to_{\mathrm{P}} \beta_0$ for some β_0 interior to $\mathcal{B}$. There exists a function $s : \mathcal{Z} \to \mathbf{R}^{d_\beta}$ such that $\|s(Z)\|^2$ is integrable and, if H_0 is true, $s(Z)$ is both centered and

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(Z_i) + o_{\mathrm{P}}(1). \tag{3.1}$$

Assumption 1 is common in the specification testing literature involving a finite-dimensional parameter estimated by nonlinear criteria (see, e.g., Horowitz, 2006, Assumption 3). Note that root-n asymptotic linearity (3.1) is only required under H_0.[8] Example 1 shows how asymptotic linearity can arise from more primitive assumptions.

Example 1 (Asymptotic linearity in two-step GMM). Let β_0 be the unique solution to $\mathrm{E}[m(Z,\beta,h^*(W))] = 0,$ where $h^*(W)$ gathers the unique elements of $\{h_\ell^*(W_\ell)\}_1^L$, and $m(Z,\beta,h^*(W))$ is a $d_m(\geqslant d_\beta)$-vector arising from interacting one or more $\rho_\ell\left(Z,\beta,h_\ell^*(W_\ell)\right)$ with (transformations of) the corresponding X_ℓ and then stacking. Let $\widehat{\beta}$ minimize $\widehat{m}(\beta)'\widehat{W}\widehat{m}(\beta)$, where $\widehat{m}(\beta) := n^{-1}\sum_{i=1}^{n} m(Z_i,\beta,\widehat{h}(W_i))$, $\widehat{h}$ is some nonparametric estimator of h^*, and $\widehat{W}$ is a $d_m \times d_m$ weight matrix consistent for W positive definite and nonstochastic. Newey (1994, Lem. 5.3) provides conditions under which such a two-step GMM estimator based on a nonparametric first step is $\sqrt{n}$-asymptotically normal as well as tools for calculating its asymptotic variance. Inspection of his proof reveals that the same set of conditions actually

[8] Under $\mathrm{H}_1,$ the pseudo-truth β_0 need not even be root-n estimable (cf. Hall and Inoue, 2003; Chen and Pouzo, 2015; Hong and Li, 2022).

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

yields (the slightly stronger) asymptotic linearity

$$\sqrt{n}(\widehat{\beta}-\beta_0) = -\left(M'WM\right)^{-1} M'W \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{m\left(Z_i, \beta_0, h^*(W_i)\right) + \alpha(Z_i)\right\} + o_{\mathrm{P}}(1), \tag{3.2}$$

where $M := \mathrm{E}[(\partial/\partial\beta')m(Z,\beta_0,h^*(W))]$ is a Jacobian term, and α is an adjustment due to estimation of h^*. Because h^* involves CEFs, Newey (1994, Proposition 4 and p. 1357) shows that, irrespective of the choice of nonparametric estimator,

$$\alpha(z) = \sum_{\ell=1}^{L} \delta_\ell(w_\ell)\left\{y_\ell - h_\ell^*(w_\ell)\right\}, \quad \delta_\ell(W_\ell) := \mathrm{E}\left[\frac{\partial}{\partial h_\ell'} m\left(Z,\beta_0,h^*(W)\right)\middle| W_\ell\right], \tag{3.3}$$

where $\partial/\partial h_\ell'$ denotes (ordinary) differentiation with respect to the $h_\ell^*(W_\ell)$ arguments.

Assumption 1 leaves freedom in choice beyond the two-step GMM estimation outlined in Example 1. For example, (3.1) allows for other or more general two- or multistep estimation procedures, such as two-step extremum estimation. Such procedures typically estimate the nonparametric components in a first step, use their estimates to construct a criterion function, and maximize or minimize over β in order to produce a second-step estimator $\widehat{\beta}$. Specifically, one may let $\widehat{\beta}$ be a sieve minimum distance estimator (Ai and Chen, 2003) or a penalized sieve minimum distance estimator (Chen and Pouzo, 2009, 2012, 2015).

We impose the following conditions on the choice of weight functions.

Assumption 2 (Weight function). Each $\mathcal{X}_\ell \subset \mathbf{R}^{d_{x,\ell}}$ is compact. Each weight function $\omega_\ell : \mathcal{X}_\ell \times \mathcal{X}_\ell \to \mathbf{R}$ is continuous, has the property that (2.4) if and only if (2.5), and satisfies the Lipschitz condition: for all $t_1, t_2, x_\ell \in \mathcal{X}_\ell$ and some finite constant C_ℓ, $|\omega_\ell(t_1,x_\ell) - \omega_\ell(t_2,x_\ell)| \leqslant C_\ell \|t_1 - t_2\|$.

Weight functions examples satisfying Assumption 2 and references giving detailed discussion of the equivalence between (2.4) and (2.5) were provided in Section 2.2.

For the moment, drop the ℓ subscript. If X is not bounded, one may replace it with $\widetilde{X} := \Phi(X)$ using a known bounded transformation Φ. Provided Φ is also one-to-one, such a transformation entails no loss in generality in the sense that $\mathrm{E}[U|X] = \mathrm{E}[U|\Phi(X)]$ a.s. The compactness "assumption" thus only acts as a reminder to conduct such a preliminary transformation, if necessary. In the simulations (Section 4) and empirical application (Section 5), we use an element-wise arctan transform to reduce otherwise unbounded conditioning variables to a bounded set prior to calculating weights.

To impose conditions on the residual functions, let d_ℓ be the number of elements in Y_ℓ (hence $h_\ell^*(W_\ell)$), let $\mathcal{W}_\ell$ be the support of W_ℓ, and write $\|f\|_{\mathcal{D}} := \sup_{x\in\mathcal{D}} |f(x)|$.

Assumption 3 (Residual). For each $\ell \in \{1,\ldots,L\}$, the following holds:

1. For each $z \in \mathcal{Z}, v_\ell \in \mathbf{R}^{d_\ell}, \beta \mapsto \rho_\ell(z,\beta,v)$ is continuous on $\mathcal{B}$ and continuously differentiable on an open neighborhood $\mathcal{N}_\ell$ of β_0. Moreover, there exist $c_\ell \in (0,1]$ and $a_\ell : \mathcal{Z} \to \mathbf{R}_+$ integrable such that for each $z \in \mathcal{Z}, \beta \in \mathcal{N}_\ell, v_\ell \in \mathbf{R}^{d_\ell}$,
$$\left\| \frac{\partial}{\partial\beta}\rho_\ell(z,\beta,v_\ell) - \frac{\partial}{\partial\beta}\rho_\ell\left(z,\beta,h_\ell^*(w_\ell)\right)\right\| \leqslant a_\ell(z)\left\|v_\ell - h_\ell^*(w_\ell)\right\|^{c_\ell}.$$
2. For each $z \in \mathcal{Z}, v_\ell \mapsto \rho_\ell(z,\beta_0,v_\ell)$ is continuously differentiable on $\mathbf{R}^{d_\ell}$. Moreover, there exists $\gamma_\ell \in (0,1]$ and $R_\ell : \mathcal{Z} \to \mathbf{R}_+$, such that for each $z \in \mathcal{Z}, v_\ell \in \mathbf{R}^{d_\ell}$,
$$\left\| \frac{\partial}{\partial h_\ell}\rho_\ell(z,\beta_0,v_\ell) - \frac{\partial}{\partial h_\ell}\rho_\ell\left(z,\beta_0,h_\ell^*(w_\ell)\right)\right\| \leqslant R_\ell(z)\left\|v_\ell - h_\ell^*(w_\ell)\right\|^{\gamma_\ell}, \tag{3.4}$$
and $\mathrm{E}[R_\ell(Z)]\sqrt{n}\max_{1\leqslant m\leqslant d_\ell}\|\widehat{h}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^{1+\gamma_\ell} \to_{\mathrm{P}} 0$.
3. The following are integrable: $\left|\rho_\ell\left(Z,\beta_0,h_\ell^*(W_\ell)\right)\right|$, $\|(\partial/\partial h)\rho_\ell(Z,\beta_0,h_\ell^*(W_\ell))\|^2$, and $\sup_{\beta\in\mathcal{N}_\ell}\|(\partial/\partial\beta)\rho_\ell(Z,\beta,h_\ell^*(W_\ell))\|$.

Assumptions 3.1 and 3.2 are smoothness conditions facilitating a linearization around (β_0,h^*) in order to extract the dominant component of the processes $\{\widehat{M}_\ell\}_1^L$ used in constructing the test statistic (2.9).[9] Assumption 3.2 generally requires each element of $\widehat{h}_\ell$ to converge to the corresponding element of h_ℓ^* uniformly over $\mathcal{W}_\ell$ at a sufficiently fast rate.[10] Such a rate requirement often boils down to assuming that the estimand is sufficiently smooth relative to its number of arguments.

While the previous assumptions allow for general nonparametric estimation methods, the following three are tailored to series estimation.

Assumption 4 (Variance). $\mathrm{var}(Y_{\ell m}|W_\ell)$ is bounded for all $m \in \{1,\ldots,d_\ell\}$, $\ell \in \{1,\ldots,L\}$.

Assumption 5 (Eigenvalues). The eigenvalues of $\mathrm{E}[p_\ell^k(W_\ell)p_\ell^k(W_\ell)']$ are bounded from above and away from zero uniformly over $k \in \mathbf{N}$ for all $\ell \in \{1,\ldots,L\}$.

Assumption 6 (Approximation). Each $h_{\ell m}^*$ is bounded. Moreover, for each $\ell \in \{1,\ldots,L\}, m \in \{1,\ldots,d_\ell\}$ and each $k \in \mathbf{N}$, there exist constants $\alpha_{\ell m} \in (0,1), C_{\ell m} \in (0,\infty)$, and $\widetilde{\pi}_{\ell m} \in \mathbf{R}^k$ such that $\|p_\ell^{k\prime}\widetilde{\pi}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell} \leqslant C_{\ell m}k^{-\alpha_{\ell m}}$.

Assumption 4 is prevalent in the series estimation literature (see, e.g., Stone, 1985; Newey 1994, 1997; Belloni et al., 2015). Assumption 5 imposes regularity conditions on the approximating functions and implies that, loosely speaking, the technical regressors $p_\ell^k(W_\ell)$ cannot be too co-linear. See Belloni et al. (2015, Proposition 2.1) for more primitive sufficient conditions. While Assumptions 4

[9] Differentiability may likely be relaxed at the cost of longer proofs. We leave such an extension to future research.

[10] A notable exception occurs when the residual is linear in $h_\ell^*(w)$. In this case, R_ℓ may be taken as the zero function, and $\mathrm{E}[R_\ell(Z)]\sqrt{n}\max_{1\leqslant m\leqslant d_\ell}\|\widehat{h}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^{1+\gamma_\ell} \to_{\mathrm{P}} 0$ becomes vacuous.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

and 5 are used to control the variance of the series estimators, Assumption 6 concerns the approximation error relative to the supremum metric. While the latter assumption is high level, it is satisfied in many cases. The exponent $\alpha_{\ell m}$ usually depends on the smoothness of the estimand $h^*_{\ell m}$ and its number of arguments d_ℓ. When the estimand can be viewed as a member of some smooth class of functions, this exponent is typically available from the approximation theory literature.[11]

Assumption 5 is a normalization that restricts the magnitude of the series terms. The theory to follow will also require that the size of each p^k_ℓ does not grow too fast relative to the sample size, where "size" is quantified by

$$\zeta_{\ell,k} := \sup_{w_\ell \in \mathcal{W}_\ell} \left\| p^k_\ell (w_\ell) \right\|. \tag{3.5}$$

For specific choices of approximating functions p^k_ℓ, bounds on the corresponding $\zeta_{\ell,k}$ are readily available. For example, under suitable conditions, $\zeta_{\ell,k} \leqslant C_\ell k$ for power series, and $\zeta_{\ell,k} \leqslant C_\ell \sqrt{k}$ for regression splines (cf. Newey, 1997). See Belloni et al. (2015, Sect. 3) for a comprehensive list.

The probabilistic behavior of the test statistic T_n in (2.9) depends crucially on that of the stochastic processes $\{\sqrt{n}\widehat{M}_\ell\}_1^L$ in (2.11). A linearization argument (cf. Lemma 1) shows that, under H_0, each $\sqrt{n}\widehat{M}_\ell$ is asymptotically equivalent to a stochastic process $\mathcal{X}_\ell \ni t_\ell \mapsto n^{-1/2} \sum_{i=1}^n g_\ell (t_\ell, Z_i)$, where

$$g_\ell (t_\ell, z) := \rho_\ell \left(z, \beta_0, h^*_\ell (w_\ell)\right) \omega_\ell (t_\ell, x_\ell) + b_\ell (t_\ell)' s(z) + \delta_\ell (t_\ell, w_\ell)' \left\{ y_\ell - h^*_\ell (w_\ell) \right\}, \tag{3.6}$$

$$b_\ell (t_\ell) := \mathrm{E}\left[\omega_\ell (t_\ell, X_\ell) \frac{\partial}{\partial \beta} \rho_\ell \left(Z, \beta_0, h^*_\ell (W_\ell)\right) \right], \tag{3.7}$$

$$\delta_\ell (t_\ell, W_\ell) := \mathrm{E}\left[\omega_\ell (t_\ell, X_\ell) \frac{\partial}{\partial h_\ell} \rho_\ell \left(Z, \beta_0, h^*_\ell (W_\ell)\right) \middle| W_\ell \right], \tag{3.8}$$

with s provided by Assumption 1. Here, $b_\ell (t_\ell)' s(z)$ and $\delta_\ell (t_\ell, w_\ell)' \left\{ y_\ell - h^*_\ell (w_\ell) \right\}$ are adjustments to the moment function $z \mapsto \rho_\ell \left(z, \beta_0, h^*_\ell (w_\ell)\right) \omega_\ell (t_\ell, x_\ell)$ due to estimation of β_0 and h^*_ℓ, respectively. The form of the β-adjustment follows from a mean-value expansion with $b_\ell (t_\ell)$ being a Jacobian term. The form of the h-adjustment is akin to the adjustment (3.3) to the influence function in two-step GMM estimation with a nonparametric first step as summarized in Example 1. The main difference is that, while two-step semiparametric GMM estimation requires adjustment of a finite number of moments used in defining the GMM criterion function, we here need to adjust a possibly infinite collection of moment functions $\left\{ z \mapsto \rho_\ell \left(z, \beta_0, h^*_\ell (w_\ell)\right) \omega_\ell (t_\ell, x_\ell) ; t_\ell \in \mathcal{X}_\ell \right\}$ for estimation of h^*_ℓ.

[11] For example, if $h^*_{\ell m}$ belongs to a Hölder ball with Hölder exponent $s_{\ell m}$ (often referred to as $h^*_{\ell m}$ being "$s_{\ell m}$-smooth"), then Assumption 6 holds with $\alpha_{\ell m} = s_{\ell m}/d_\ell$, provided p^k_ℓ is constructed using either power series (see, e.g., Timan, 1963, Sect. 5.3.2; Lorentz, 1966, Thm. 8) or splines (see, e.g., DeVore and Lorentz, 1993; Schumaker, 2007).

To state the next assumption, define the mean-square projection coefficients

$$\pi_{h_{\ell m},k} := \underset{\pi \in \mathbf{R}^k}{\operatorname{argmin}} \mathrm{E}\left[\left\{p_\ell^k(W_\ell)'\pi - h_{\ell m}^*(W_\ell)\right\}^2\right], \tag{3.9}$$

$$\pi_{\delta_{\ell m},k}(t_\ell) := \underset{\pi \in \mathbf{R}^k}{\operatorname{argmin}} \mathrm{E}\left[\left\{p_\ell^k(W_\ell)'\pi - \delta_{\ell m}(t_\ell, W_\ell)\right\}^2\right], \tag{3.10}$$

and their induced mean-square errors

$$r_{h_{\ell m},k}^2 := \min_{\pi \in \mathbf{R}^k} \mathrm{E}\left[\left\{p_\ell^k(W_\ell)'\pi - h_{\ell m}^*(W_\ell)\right\}^2\right], \tag{3.11}$$

$$r_{\delta_{\ell m},k}^2(t_\ell) := \min_{\pi \in \mathbf{R}^k} \mathrm{E}\left[\left\{p_\ell^k(W_\ell)'\pi - \delta_{\ell m}(t_\ell, W_\ell)\right\}^2\right], \tag{3.12}$$

$$R_{\delta_{\ell m},k}^2 := \mathrm{E}\left[\sup_{t_\ell \in \mathcal{X}_\ell}\left\{p_\ell^k(W_\ell)'\pi_{\delta_{\ell m},k}(t_\ell) - \delta_{\ell m}(t_\ell, W_\ell)\right\}^2\right], \tag{3.13}$$

where $\ell \in \{1,\ldots,L\}, t_\ell \in \mathcal{X}_\ell$, and $m \in \{1,\ldots,d_\ell\}$. Assumption 7 contains rate conditions sufficient to show that the difference between $\sqrt{n}\widehat{M}_\ell$ and $n^{-1/2}\sum_{i=1}^n g_\ell(\cdot, Z_i)$ is asymptotically negligible under H_0, $\ell \in \{1,\ldots,L\}$.

Assumption 7 (Rate conditions). For all $\ell \in \{1,\ldots,L\}$ and $m \in \{1,\ldots,d_\ell\}$,

$$\zeta_{\ell,k_{\ell m,n}} r_{h_{\ell m},k_{\ell m,n}} \to 0, \qquad \frac{\zeta_{\ell,k_{\ell m,n}}^2 k_{\ell m,n}\ln\left(k_{\ell m,n}\right)}{n} \to 0, \qquad n r_{h_{\ell m},k_{\ell m,n}}^2 \|r_{\delta_{\ell m},k_{\ell m,n}}\|_{\mathcal{X}_\ell}^2 \to 0,$$

$$R_{\delta_{\ell m},k_{\ell m,n}} \to 0, \qquad R_{\delta_{\ell m},k_{\ell m,n}}\sqrt{\ln\left(\frac{k_{\ell m,n}}{R_{\delta_{\ell m},k_{\ell m,n}}}\right)} \to 0,$$

and for $\alpha_{\ell m}$ provided by Assumption 6,

$$\left(\sum_{j=1}^{k_{\ell m,n}} \|p_{\ell j}\|_{\mathcal{W}_\ell}^2\right)^{1/2}\left(\sqrt{\frac{k_{\ell m,n}}{n}} + k_{\ell m,n}^{-\alpha_{\ell m}}\right) \to 0.$$

In discussing the rate conditions, consider the scalar case and drop the ℓ and m subscripts. Given that $\zeta_k \leqslant (\sum_{j=1}^k \|p_{jk}\|_{\mathcal{W}}^2)^{1/2}$, the last rate condition ensures that $\zeta_{k_n}(\sqrt{k_n/n} + k_n^{-\alpha}) \to 0$, which we use to argue uniform consistency of the series estimators. While the presence of ζ_k in the rate conditions formally requires one to use bounded approximating functions, the simulations in Section 4—where we construct approximating functions based on power series with unbounded conditioning variables—suggest that this requirement can be relaxed.

Rate conditions involving mean-square projection errors such as a finite number of $r_{\delta,k_n}(t)$'s appear in, e.g., Newey (1994, Assumption 6.6). The new part of Assumption 7 lies in the (stronger) assumptions placed on the *uniform* error arising from such projections, as captured by R_{δ,k_n}.

Observe that the mean-square projection error r_{h,k_n} resulting from approximating h^* by linear forms is not required to go to zero at a rate faster than $n^{-1/2}$. Such a condition would otherwise require choosing k_n larger than what would

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

optimize the rate of convergence, sometimes referred to as "undersmoothing." Instead, Assumption 7 requires the *product* of r_{h,k_n} and the maximal approximation mean-square error $\|r_{\delta,k_n}\|_{\mathcal{X}}$ to be $o(n^{-1/2})$. This (weaker) requirement arises from the orthogonality property of mean-square projections. Specifically, for the projections $h_k(\cdot) = p^k(\cdot)'\pi_{h,k}$ and $\delta_k(t,\cdot) = p^k(\cdot)'\pi_{\delta,k}(t)$ of h^* and $\delta(t,\cdot)$, respectively, the bias term $\mathrm{E}[\delta(t,W)\{h_k(W) - h^*(W)\}]$ equals $\mathrm{E}[\{\delta_k(t,W) - \delta(t,W)\}\{h_k(W) - h^*(W)\}]$ for each $t \in \mathcal{X}$. Consequently, if the family $\{\delta(t,\cdot); t \in \mathcal{X}\}$ can be sufficiently well approximated by linear forms, there is no need to undersmooth.[12] Newey (1994) shows that a similar feature arises in the context of two-step GMM estimation with a first step based on series estimation of projection functionals, such as CEFs.

The previous assumptions suffice for the asymptotic equivalence posited above.

LEMMA 1 (Asymptotic equivalence). *If Assumptions 1–7 hold and* H_0 *is true, then for* $\{\widehat{M}_\ell\}_1^L$ *in (2.11) and* $\{g_\ell\}_1^L$ *in (3.6), we have*

$$\max_{1\leqslant \ell\leqslant L}\left\|\sqrt{n}\widehat{M}_\ell(\cdot) - \frac{1}{\sqrt{n}}\sum_{i=1}^n g_\ell(\cdot, Z_i)\right\|_{\mathcal{X}_\ell} \xrightarrow{\mathrm{P}} 0.$$

Lemma 1 implies that the probabilistic behavior of $\|\sqrt{n}\widehat{M}_\ell\|$ under H_0 may be approximated by that of $\|n^{-1/2}\sum_i g(\cdot, Z_i)\|$ for any norm $\|\cdot\|$ weaker than the supremum norm, such as the empirical L^2-norms implicit in the definition of T_n.

Remark 1 (Comparing with Bravo, 2012). The asymptotic equivalence established in Lemma 1 is directly assumed in Bravo (2012, Assumption 2.1(e)), which in his paper is invoked under the null hypothesis. Implicit in Bravo's assumption is the presumption that the user is both willing and able to derive the dominant part (his $\sum_{i=1}^n l_i(\theta_0, h_0, \cdot)/n^{1/2}$) of a stochastic process (his $\sum_{i=1}^n v_i(\widehat{\theta}, \widehat{h}, \cdot)/n^{1/2}$) as well as suitably estimate it. Knowledge of the dominant process is crucial for implementation as estimates thereof (his $\widehat{l_i}$) figure in his bootstrap procedure. The derivations needed to arrive at the dominant process as well as construct suitable estimators thereof are elegantly illustrated by Bravo's examples. However, these derivations are case specific and may be quite involved (as illustrated by the same examples). In contrast, in (3.6)–(3.8) and (3.20)–(3.22), we give explicit formulas for the $g_\ell(t_\ell, z)$'s and estimators thereof, respectively, and we establish the validity of these estimates in Lemma 3.

Recall that a class $\mathcal{F}$ of real-valued functions is called *Donsker* (van der Vaart and Wellner, 1996, pp. 81–82), if the sequence of empirical processes $\{n^{-1/2}\sum_{i=1}^n\{f(Z_i) - \mathrm{E}[f(Z)]\}; f \in \mathcal{F}\}$ induced by $\mathcal{F}$—viewed as random elements of $L^\infty(\mathcal{F})$—converges weakly to a centered Gaussian process $\{\mathbb{G}(f); f \in \mathcal{F}\}$

[12] While undersmoothing may not be necessary to achieve the claimed asymptotic approximation, it may be optimal in the sense of minimizing the remainder resulting from this approximation as remarked by Donald and Newey (1994) in the context of partially linear regression.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

with covariance kernel $(f_1,f_2) \mapsto \mathrm{E}[\mathbb{G}(f_1)\mathbb{G}(f_2)] = \mathrm{E}[f_1(Z)f_2(Z)] - \mathrm{E}[f_1(Z)]\mathrm{E}[f_2(Z)]$.[13] Define function classes $\mathcal{G}_\ell := \{g_\ell(t_\ell,\cdot) : \mathcal{Z} \to \mathbf{R}; t_\ell \in \mathcal{X}_\ell\}, \ell \in \{1,\ldots,L\}$, and $\mathcal{G} := \times_{\ell=1}^{L}\mathcal{G}_\ell$. The same set of assumptions then also establishes the following result.

LEMMA 2 (Donsker class). *If Assumptions 1–7 hold, then $\mathcal{G}$ is Donsker.*

Under H_0, each $t_\ell \mapsto \mathrm{E}[g_\ell(t_\ell,Z)]$ vanishes. Since we may identify each $\mathcal{G}_\ell$ with the corresponding $\mathcal{X}_\ell$, Lemma 2 then means that the L-variate stochastic process

$$G_n(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(t,Z_i), \quad t \in \mathcal{T}, \quad \mathcal{T} := \times_{\ell=1}^{L}\mathcal{X}_\ell,$$

converges weakly in $\times_{\ell=1}^{L}L^\infty(\mathcal{X}_\ell)$ to an L-variate zero-mean Gaussian process G_0 indexed by $\mathcal{T}$ with (matrix) covariance kernel

$$\mathbb{C}_0\left(t,t'\right) := \mathrm{E}[g(t,Z)\,g\left(t',Z\right)'], \quad t,t' \in \mathcal{T}. \tag{3.14}$$

Under H_1, each $\widehat{M}_\ell$ converges uniformly in probability to M_ℓ, and at least one M_ℓ is nonvanishing. The behavior of the test statistic follows.

THEOREM 1 (Asymptotic behavior of test statistic). *If Assumptions 1–7 hold, then*

$$T_n \xrightarrow{d} \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} G_{0\ell}(t_\ell)^2\,\mathrm{d}F_{X_\ell}(t_\ell) =: T_0 \text{ under } \mathrm{H}_0,$$

$$\frac{T_n}{n} \xrightarrow{\mathrm{P}} \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} M_\ell(t_\ell)^2\,\mathrm{d}F_{X_\ell}(t_\ell) > 0 \text{ under } \mathrm{H}_1.$$

The proof of Theorem 1 invokes a (second-order) functional delta method argument to show that the limiting null behavior is unaffected by the use of empirical distributions in place of their (unknown) population counterparts. The first claim then follows from Lemmas 1 and 2 via the continuous mapping theorem. The second claim of Theorem 1 implies that $T_n \to_{\mathrm{P}} \infty$ at the rate n under the alternative, which is key to establishing consistency (Theorem 2).

The limit results in Theorem 1 cannot be operationalized without appropriate critical values. For this purpose, we rely on a multiplier bootstrap procedure.

3.2. Bootstrap Critical Values

The limiting law of T_n under H_0 is given by that of $T_0 = \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell} G_{0\ell}(t_\ell)^2\,\mathrm{d}F_{X_\ell}(t_\ell)$ (cf. Theorem 1). To obtain a consistent bootstrap, it is therefore necessary to estimate the law of the Gaussian process G_0. To this end, let $\{\xi_i\}_1^\infty$ be i.i.d.

[13] A sequence X_n of stochastic processes taking values in a metric space $\mathbb{D}$ are said to *converge weakly* to X if $\mathrm{E}[h(X_n)] \to \mathrm{E}[h(X)]$ for all $h : \mathbb{D} \to \mathbf{R}$ continuous and bounded.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

standard normal random variables independent of the stream of data $\{Z_i\}_1^\infty$ and let $\overline{\xi} := n^{-1}\sum_{i=1}^n \xi_i$.[14]

To fix ideas, consider first the multiplier process $G_n^* := \left(G_{1n}^*, \ldots, G_{Ln}^*\right)'$ defined by

$$G_n^*(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\xi_i - \overline{\xi}\right) g(t, Z_i), \quad t \in \mathcal{T}. \tag{3.15}$$

By independence, the summands of G_n^* are centered even if one or more of the $g_\ell(t_\ell, Z)$'s are not, i.e., even when H_0 is false. The demeaning aims for less conservative critical values in finite sample by correctly accounting for sample variation.

The multiplier process G_n^* is said to converge weakly in probability to G^* in $\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)$, written $G_n^* \rightsquigarrow_{\mathrm{P},\xi} G^*$, if G_n^* converges weakly to G^* conditional on the data, in probability.[15] Given that $\mathcal{G}$ is Donsker (Lemma 2), Kosorok (2008, Thm. 10.4) shows that $G_n^* \rightsquigarrow_{\mathrm{P},\xi} G$ in $\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)$, where G is a centered Gaussian process with covariance kernel

$$\mathbb{C}\left(t, t'\right) := \mathrm{E}\left[\{g(t,Z) - \mathrm{E}[g(t,Z)]\}\{g(t,Z) - \mathrm{E}[g(t,Z)]\}'\right], \quad t, t' \in \mathcal{T}. \tag{3.16}$$

Under $\mathrm{H}_0, \mathrm{E}[g(\cdot, Z)] \equiv 0$ on $\mathcal{T}$, and the covariance kernels $\mathbb{C}$ and $\mathbb{C}_0$ coincide. Since both G and G_0 are Gaussian, the two must then be identically distributed. This observation suggests the $(1-\alpha)$-quantile of $\sum_{\ell=1}^L \int_{\mathcal{X}_\ell} G_{\ell n}^*(t_\ell)^2 \, \mathrm{d}F_{X_\ell}(t_\ell)$, calculated conditional on $\{Z_i\}_1^n$, as an approximation to

$$c_T(\alpha) := (1-\alpha)\text{-quantile of } T, \text{ where } T := \sum_{\ell=1}^L \int_{\mathcal{X}_\ell} G_\ell(t_\ell)^2 \, \mathrm{d}F_{X_\ell}(t_\ell). \tag{3.17}$$

While the g_ℓ's and F_{X_ℓ}'s are generally unknown, endowed with an estimator $\widehat{s}$ of the (null influence) function s from Assumption 1, one may estimate g and define the *bootstrap process* $\widehat{G} := (\widehat{G}_1, \ldots, \widehat{G}_L)'$ as the feasible version of G_n^*. Specifically, we let

$$\widehat{G}(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\xi_i - \overline{\xi}\right)\widehat{g}(t, Z_i), \quad t \in \mathcal{T}, \tag{3.18}$$

$$\widehat{g}(t,z) := \left(\widehat{g}_1(t_1,z), \ldots, \widehat{g}_L(t_L,z)\right)', \tag{3.19}$$

$$\widehat{g}_\ell(t_\ell, z) := \rho_\ell(z, \widehat{\beta}, \widehat{h}_\ell(w_\ell))\omega_\ell(t_\ell, x_\ell) + \widehat{b}_\ell(t_\ell)'\widehat{s}(z) + \widehat{\delta}_\ell(t_\ell, w_\ell)'\{y_\ell - \widehat{h}_\ell(w_\ell)\}, \tag{3.20}$$

---

[14] Standard normality of the multipliers is chosen mainly for the sake of concreteness. The limiting behavior of the test introduced below remains the same if one instead uses any zero mean and unit variance distribution, provided it has a finite weak second moment. The latter requirement is implied by $\mathrm{E}[|\xi|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$. See, e.g., van der Vaart and Wellner (1996, Sect. 2.9) for details.

[15] That is, $\sup_{h \in \mathrm{BL}_1\left(\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)\right)} \left|\mathrm{E}\left[h\left(G_n^*\right) \middle| \{Z_i\}_1^n\right] - \mathrm{E}\left[h(G^*)\right]\right| \to_{\mathrm{P}} 0$, where $\mathrm{BL}_1(\mathbb{D})$ denotes the space of functionals $h : \mathbb{D} \to \mathbf{R}$ defined on the metric space $(\mathbb{D}, d)$ whose Lipschitz norm is bounded by one, i.e., $\|h\|_{\mathbb{D}} \leqslant 1$ and $|h(f) - h(g)| \leqslant d(f,g)$ for all $f, g \in \mathbb{D}$.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

where

$$\widehat{b}_\ell(t) := \frac{1}{n}\sum_{i=1}^{n}\omega_\ell(t_\ell, X_{\ell i})\frac{\partial}{\partial\beta}\rho_\ell(Z_i,\widehat{\beta},\widehat{h}_\ell(W_{\ell i})), \tag{3.21}$$

$$\widehat{\delta}_{\ell m}(t_\ell, w_\ell) := p_\ell^{k_{\ell m,n}}(w_\ell)'\left(\frac{1}{n}\sum_{i=1}^{n}p_\ell^{k_{\ell m,n}}(W_{\ell i})p_\ell^{k_{\ell m,n}}(W_{\ell i})'\right)^{-} \times\frac{1}{n}\sum_{i=1}^{n}p_\ell^{k_{\ell m,n}}(W_{\ell i})\,\omega_\ell(t_\ell, X_{\ell i})\frac{\partial}{\partial h_{\ell m}}\rho_\ell(Z_i,\widehat{\beta},\widehat{h}_\ell(W_{\ell i})). \tag{3.22}$$

Based on the above process, we get the feasible bootstrap test statistic,

$$\widehat{T} := \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}\widehat{G}_\ell(t_\ell)^2\,\mathrm{d}\widehat{F}_{X_\ell}(t_\ell) = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{i=1}^{n}\widehat{G}_\ell(X_{\ell i})^2, \tag{3.23}$$

and a feasible critical value as follows:

$$c_{\widehat{T}}(\alpha) := (1-\alpha)\text{-quantile of } \widehat{T} \text{ conditional on } \{Z_i\}_1^n. \tag{3.24}$$

The test rejects H_0 in favor of H_1 if and only if $T_n > c_{\widehat{T}}(\alpha)$ for some prespecified significance level $\alpha \in (0,1)$, where the test statistic is defined in (2.9) and the critical value in (3.24). For any α, $c_{\widehat{T}}(\alpha)$ can be obtained through simulation holding the data constant. In practice, this simulation is terminated after a finite but large number of draws. In the empirical illustration of Section 5, we use 250,000.

Remark 2 (Additively separable residuals). If a residual function is *additively separable* in the conditioning variables, in the sense that $\rho_\ell(z,\beta_0,h_\ell^*(w_\ell)) = \phi_\ell(y_\ell,\beta_0)+\varphi_\ell(x_\ell,\beta_0,h_\ell^*(w_\ell))$, then $(\partial/\partial h_\ell)\rho_\ell(z,\beta_0,h_\ell^*(w_\ell)) = (\partial/\partial h_\ell)\varphi_\ell(x_\ell,\beta_0, h_\ell^*(w_\ell))$ depends on z through x_ℓ alone. If, in addition, X_ℓ and W_ℓ (measurably) coincide, then the term $\omega_\ell(t_\ell,X_\ell)(\partial/\partial h_\ell)\rho_\ell(Z,\beta_0,h_\ell^*(W_\ell))$ is conditionally known given W_ℓ (up to β_0 and h_ℓ^*) and hence equal to $\delta_\ell(t_\ell,W_\ell)$. For such models, one may therefore drop the projection element of the estimator in (3.22) and replace it by the simpler $\widehat{\delta}_\ell(t_\ell,W_{\ell i}) = \omega_\ell(t_\ell,X_{\ell i})(\partial/\partial h_\ell)\varphi_\ell(X_{\ell i},\widehat{\beta},\widehat{h}_\ell(W_{\ell i}))$. This simplification is utilized in both the simulations (Sections 4) and the empirical illustration (Section 5).

Remark 3 (Alternative nonparametric estimators). The test proposed in this paper is based on series estimation of the CEFs h_ℓ^* and $\delta_\ell(t_\ell,\cdot), t_\ell \in \mathcal{W}_\ell, \ell \in \{1,\dots,L\}$. From a theoretical point of view, series estimation has the advantage that no undersmoothing is required, which follows from the built-in orthogonality of mean-square projections (cf. the discussion following Assumption 7). While the theory in this paper is tailored to series estimation (in particular, Assumptions 4–7), in practice, one may replace the series estimators with alternative nonparametric estimation techniques, such as kernel regression. Note, however, that the orthogonality property of mean-square projections is not shared by kernel estimators, which typically require undersmoothing to eliminate the relevant bias terms.

A potentially difficult step in this bootstrap procedure is constructing $\widehat{s}$. One strategy to estimation involves first obtaining an analytic formula for s and then replacing any unknowns with consistent estimates. This approach can be taken in the case of two-step GMM estimators (Example 1). However, at the level of generality for estimation of the parametric component considered in this paper, it does not appear possible to give primitive conditions under which $\widehat{s}$ is consistent for s. Letting $\|f\|_{n,2} := \{n^{-1}\sum_{i=1}^{n} f(Z_i)^2\}^{1/2}$, we invoke the following assumption.

Assumption 8 (Bootstrap conditions).

1. $n^{-1}\sum_{i=1}^{n}\|\widehat{s}(Z_i) - s(Z_i)\|^2 \to_{\mathrm{P}} 0.$

   Moreover, for each $\ell \in \{1,\dots,L\}$, the following holds.
2. For each $z \in \mathcal{Z}, v_\ell \in \mathbf{R}^{d_\ell}, \beta \mapsto \rho_\ell(z,\beta,v_\ell)$ is continuously differentiable on $\mathcal{N}_\ell$. Moreover, there exists $a' : \mathcal{Z} \to \mathbf{R}_+$ such that for each $z \in \mathcal{Z}, \beta \in \mathcal{N}_\ell, v_\ell \in \mathbf{R}^{d_\ell}$,

$$\left\| \frac{\partial}{\partial\beta}\rho_\ell(z,\beta,v_\ell) - \frac{\partial}{\partial\beta}\rho_\ell\left(z,\beta_0,h_\ell^*(w_\ell)\right)\right\| \leqslant a_\ell'(z)\left(\|\beta-\beta_0\| + \|v_\ell - h_\ell^*(w_\ell)\|\right),$$

   where $\mathrm{E}\left[a_\ell'(Z)\right]\sqrt{n}\|\widehat{\beta}-\beta_0\|\max\{\|\widehat{\beta}-\beta_0\|, \max_{1\leqslant m\leqslant d_\ell}\|\widehat{h}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}\} \to_{\mathrm{P}} 0.$
3. For each $z \in \mathcal{Z}, \beta \in \mathcal{N}_\ell, v_\ell \mapsto \rho_\ell(z,\beta,v_\ell)$ is continuously differentiable on $\mathbf{R}^{d_\ell}$. Moreover, there exists $R_\ell' : \mathcal{Z} \to \mathbf{R}_+$ such that for each $z \in \mathcal{Z}, \beta \in \mathcal{N}_\ell, v_\ell \in \mathbf{R}^{d_\ell}$,

$$\left\| \frac{\partial}{\partial h_\ell}\rho_\ell(z,\beta,v_\ell) - \frac{\partial}{\partial h_\ell}\rho_\ell\left(z,\beta_0,h_\ell^*(w_\ell)\right)\right\| \leqslant R_\ell'(z)\left(\|\beta-\beta_0\| + \|v_\ell - h_\ell^*(w_\ell)\|\right),$$

   where $\mathrm{E}\left[R_\ell'(Z)\right]\max_{1\leqslant m\leqslant d_\ell}\|\widehat{h}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^2 \to_{\mathrm{P}} 0.$
4. For all $m \in \{1,\dots,d_\ell\}$ and $\alpha_{\ell m}$'s provided by Assumption 6,

$$\left\{\mathrm{E}\left[R_\ell'(Z)^2\right]\right\}^{1/2}\zeta_{\ell,k_{\ell m,n}}\sqrt{k_{\ell m,n}}\max\left\{\|\widehat{\beta}-\beta_0\|, \max_{1\leqslant m'\leqslant d_\ell}\|\widehat{h}_{\ell m'} - h_{\ell m'}^*\|_{n,2}\right\} \xrightarrow{\mathrm{P}} 0,$$

$$\left(\sum_{j=1}^{k_{\ell m,n}}\|p_{\ell j}\|_{\mathcal{W}_\ell}^2\right)^{1/2}\max_{1\leqslant m'\leqslant d_\ell}\left(\sqrt{k_{\ell m',n}/n} + k_{\ell m',n}^{-\alpha_{\ell m'}}\right) \to 0.$$

With the addition of Assumption 8, we obtain the following result.

LEMMA 3 (Bootstrap equivalence). *If Assumptions 1–8 hold, then* $\max_{1\leqslant\ell\leqslant L}\|\widehat{G}_\ell - G_{\ell n}^*\|_{\mathcal{X}_\ell} \to_{\mathrm{P}} 0.$

Lemma 3 establishes that the unknown character of g is asymptotically irrelevant. Given that G_n^* converges weakly in probability to G, by the lemma, so must the feasible analog $\widehat{G}$.

The limit T in (3.17) is a nonnegative random variable arising from applying a convex functional (the sum of squares of L^2-type norms) to a Gaussian process G. Since G is centered and indexed by $\mathcal{T}$, the function $t \mapsto 0_{L\times 1}$ identically zero on $\mathcal{T}$ is in the support of G. It follows from Davydov et al. (1998, Thm. 11.1 and Prob. 11.3) that its CDF F_T is everywhere continuous, except possibly at the point of separation—here zero. We explicitly rule out mass at separation by invoking the following assumption.

Assumption 9 (Continuity). $F_T(0) = 0$.

More primitive conditions may be used to satisfy Assumption 9. For example, using the continuity of sample paths of G, $F_T(0) = 0$ may be obtained under the "nondegeneracy" assumption that $\text{var}[g_\ell(t_\ell, Z)] > 0$ for some $t_\ell \in \mathcal{X}_\ell$ and some $\ell \in \{1, \dots, L\}$, when combined with an assumption that the corresponding distribution F_{X_ℓ} is absolutely continuous with density bounded away from zero.

From the (now) continuous nature of the weak in-probability limit T of $\widehat{T}$, convergence of quantiles essentially follows.

LEMMA 4 (Quantile consistency). *If Assumptions 1–9 hold, and F_T is increasing at its $(1-\alpha)$-quantile $c_T(\alpha)$, then $c_{\widehat{T}}(\alpha) \to_{\mathrm{P}} c_T(\alpha) \in (0, \infty)$.*

3.3. Limiting Behavior of Test

Theorem 2 contains the main results of this paper, namely that the test which rejects H_0 if and only if $T_n > c_{\widehat{T}}(\alpha)$ is: (1) asymptotically correctly sized and (2) consistent against any fixed alternative.

THEOREM 2 (Asymptotic properties of test). *If Assumptions 1–9 hold, and F_T is increasing at its $(1-\alpha)$-quantile, then*

$$\mathrm{P}\left(T_n > c_{\widehat{T}}(\alpha)\right) \to \begin{cases} \alpha, & \textit{under } \mathrm{H}_0, \\ 1, & \textit{under } \mathrm{H}_1. \end{cases}$$

As pointed out by a referee, Assumptions 1–9 alone actually suffice for the (weaker) size control $\limsup_{n\to\infty} \mathrm{P}(T_n > c_{\widehat{T}}(\alpha)) \leqslant \alpha$ under H_0. This result essentially follows from the (uniform) validity of the bootstrap (under H_0), which does not hinge on F_T being increasing at its $(1-\alpha)$-quantile.

4. SIMULATIONS

To demonstrate the usefulness of the proposed testing procedure and assess its finite-sample properties, we carry out a simulation experiment in a game-theoretical econometric model (Sections 4.1–4.3). We also numerically compare with the existing literature using a partially linear model (Section 4.4).

4.1. Setup: A Two-by-Two Game of Incomplete Information

A potential application of the test lies in testing for correct specification of static binary choice models with social and/or strategic interactions.[16] Such models may

[16]Static discrete-choice models with social and/or strategic interactions have been applied in numerous contexts including firm entry (Seim, 2006), the timing of radio commercials (Sweeting, 2009), labor force participation (Bjorn and Vuong, 1984), and teen sex (Card and Giuliano, 2013).

be conveniently estimated in two steps. In the first step, the conditional choice probabilities (CCPs) are estimated in a nonparametric manner. The estimated CCPs are then employed in the second step to estimate the structural parameters of the model (see, e.g., Bajari et al., 2010). Construction of the test follows along the same lines.[17]

We use a two-player, binary-action game of incomplete information as the DGP.[18] Two players, indexed $j \in \{1,2\}$, simultaneously choose one out of two alternatives $y_j \in \{0,1\}$. Utility of the players is parameterized as

$$u\left(y_j, y_{-j}, x_j, \varepsilon_j(0), \varepsilon_j(1); \theta\right) = \begin{cases} Ax_j + Cx_j^2 + \gamma_0 y_{-j} + \varepsilon_j(1), & y_j = 1, \\ Bx_j + Dx_j^2 + \gamma_0\left(1 - y_{-j}\right) + \varepsilon_j(0), & y_j = 0, \end{cases}$$

where y_{-j} denotes the action of the other player, x_j is a player-specific *public* payoff shock, and $\left(\varepsilon_j(0), \varepsilon_j(1)\right)$ is a vector of i.i.d. (over both players and alternatives) payoff shocks private to player j drawn from a commonly known distribution. In a BNE, both players maximize their expected utility given their beliefs, and their beliefs turn out correct, thus leading to the decision rule

$$Y_j = \mathbf{1}\left(\alpha_0 X_j + \delta_0 X_j^2 + \gamma_0\left(2\mathrm{E}\left[Y_{-j}\middle|X\right] - 1\right) \geqslant \varepsilon_j\right), \tag{4.1}$$

where we abbreviate $\alpha_0 := A - B, \delta_0 := C - D, X := (X_1, X_2)$, and $\varepsilon_j := \varepsilon_j(0) - \varepsilon_j(1)$. The $\varepsilon_j\left(y_j\right)$'s are here taken to be Type 1 Extreme Value distributed independent of the X_j's. Correctness of beliefs therefore leads to the CCPs

$$\mathrm{E}\left[Y_j\middle|X\right] = \Lambda\left(\alpha_0 X_j + \delta X_j^2 + \gamma_0\left(2\mathrm{E}\left[Y_{-j}\middle|X\right] - 1\right)\right), \quad j \in \{1,2\}, \tag{4.2}$$

with $\Lambda(t) = 1/(1 + \mathrm{e}^{-t})$ being the logistic CDF.

Let $\{\{(Y_{ij}, X_{ij})\}_{j=1}^2\}_{i=1}^n$ denote data from n independent games. We wish to test the hypothesis

$$\exists \beta := (\alpha, \gamma) \text{ s.t. } \mathrm{E}\left[Y_j - \Lambda\left(\alpha X_j + \gamma\left(2\mathrm{E}\left[Y_{-j}\middle|X\right] - 1\right)\right)\middle|X\right] = 0 \text{ a.s. for both } j \in \{1,2\}.$$

To generate data from the model, we first draw conditioning variables $X = (X_1, X_2)$, which are taken to be bivariate normal with unit variances and correlation ρ (specified below). We then solve the two equations

$$\sigma_j = \Lambda\left(\alpha_0 X_j + \delta_0 X_j^2 + \gamma_0\left(2\sigma_{-j} - 1\right)\right), \quad j \in \{1,2\}, \tag{4.3}$$

---

[17] Implicit in this two-step estimation approach is an assumption of equilibrium uniqueness. See Hahn, Moon, and Snider (2017) for a test, which can be used to test for equilibrium multiplicity.

[18] The DGP considered here is a slight modification of the one in Hahn et al. (2017) with the addition of continuous conditioning variables. When conditioning variables are discrete, the CEFs may be represented by a finite set of values, and the estimation problem becomes parametric.

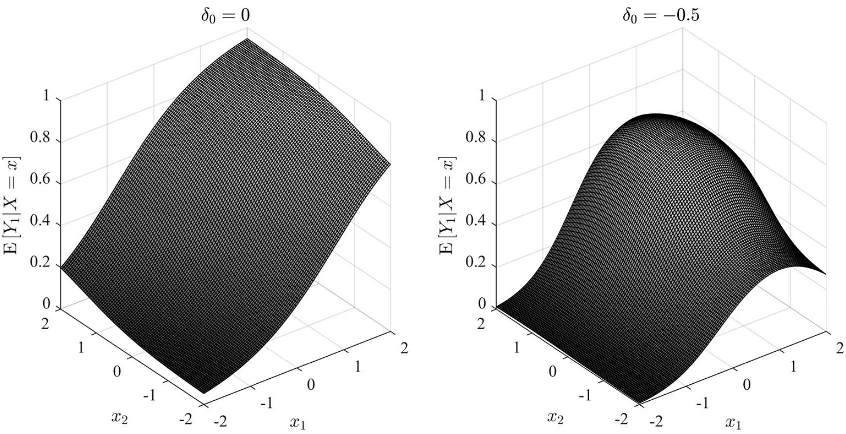

FIGURE 1. Equilibrium beliefs of Player 1 as a function of public information.

in the unknowns (σ_1, σ_2) to obtain beliefs consistent with a BNE. Outcomes are subsequently generated using the decision rules in (4.1).[19] Throughout, we set $\alpha_0 = \gamma_0 = 1$. To generate data consistent with H_0, we set $\delta_0 = 0$. To generate data under H_1, we set $\delta_0 = -0.5$. The resulting equilibrium beliefs from the perspective of Player 1 as a function of the public signals are depicted in Figure 1. (Player 2's beliefs follow from symmetry.) In both cases, equilibrium beliefs are analytic functions in public information (cf. the analytic implicit function theorem).

To allow for different parts of the equilibrium belief surface to be likely to be explored, we allow for different levels of public information correlation, $\rho \in \{0, .1, \ldots, .5\}$.

4.2. Construction of Test

To construct the test statistic, we first take a series approach to estimating the equilibrium beliefs $h_j^*(\cdot) := \mathrm{E}\left[Y_j \middle| X = \cdot\right]$ of both players. For both estimands, we employ the power series approximating functions $p_\ell^k := p^k$ defined as the tensor product

$$p^k(x)' := (1, x_1, \ldots, x_1^{\sqrt{k}-1}) \otimes (1, x_2, \ldots, x_2^{\sqrt{k}-1})$$

of the monomials in each argument up to the same order. The formal results of this paper are developed under the assumption that the series length $k = k_n$ grows

[19]Depending on the value of X, the nonlinear system (4.3) may in principle have multiple solutions resulting in different equilibria. The notation employed in (4.1) and (4.2) implicitly assumes uniqueness of equilibrium beliefs. The parameter values are here selected to guarantee a unique solution to this nonlinear system of equations no matter the realization of X, thus ensuring equilibrium uniqueness.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

with n at a suitable rate. However, for a given sample size, one must settle on a particular k. In order to investigate the sensitivity of the test with respect to this (user) choice, we carry out our procedure for each series length $k \in \{4, 9, 16\}$.

Next, based on the logit CCPs,

$$f\left(y_j \middle| x, (\alpha, \gamma), h\right) = \Lambda\left(\alpha x_j + \gamma(2h_{-j} - 1)\right)^{y_j}\left[1 - \Lambda\left(\alpha x_j + \gamma(2h_{-j} - 1)\right)\right]^{1-y_j},$$

we formulate a (pseudo) maximum likelihood estimator,

$$\widehat{\beta} := \operatorname*{argmax}_{\beta \in \mathbf{R}^2} \sum_{i=1}^{n} \sum_{j=1}^{2} \ln f\left(Y_{ij} \middle| X_i, \beta, \widehat{h}(X_i)\right).$$

Following Bierens (1990), we use exponential weighting $\omega(\tilde{t}, \tilde{x}) = \exp(\tilde{t}'\tilde{x})$ combined with a preliminary arctan transformation $\widetilde{X}_j := \tan^{-1}\left(X_j\right)$ of each (otherwise unbounded) conditioning variable. We use the same weights for both residuals.[20] The test statistic then follows from (2.9)–(2.11) using

$$\rho_\ell(z, \beta, h) := y_\ell - \Lambda(\alpha x_\ell + \gamma(2h_{-\ell} - 1)), \quad \ell \in \{1, 2\},$$

as residual functions,[21] and integration is understood to be against the empirical distribution of the *transformed* conditioning variables.

We obtain a critical value using (3.18)–(3.24). Given that the argument of the CEFs coincides with the conditioning variables and that $(\partial/\partial h_{-\ell})\rho_\ell(z, \beta_0, h^*(x)) = -2\gamma_0 \Lambda'(\alpha_0 x_\ell + \gamma_0(2h^*_{-\ell}(x) - 1))$ depends on z only through x, we construct the $\widehat{\delta}_{\ell,-\ell}$'s defined in (3.22) without projections, i.e.,

$$\widehat{\delta}_{\ell,-\ell}(\tilde{t}, X_i) := -2\widehat{\gamma}\omega(\tilde{t}, \widetilde{X}_i)\Lambda'(\widehat{\alpha}X_{i\ell} + \widehat{\gamma}[2\widehat{h}_{-\ell}(X_i) - 1]).$$

(See also Remark 2.) The $\widehat{\delta}_{\ell,\ell}$'s are zero. To obtain the $\widehat{s}(Z_i)$ estimates needed in (3.20) to adjust for estimation of β_0, we first derive the influence function of $\sqrt{n}(\widehat{\beta} - \beta_0)$ as outlined in Example 1 for general two-step GMM estimation with a nonparametric first step, using the (pseudo) scores $\sum_{j=1}^{2}(\partial/\partial\beta)\ln f(y_j|x, \beta, h)$ as moment functions $m(z, \beta, h)$. We then replace unknowns (including the moment Jacobian) with sample analogs.

We consider sample sizes $n \in \{100, 200, 400\}$ and 10,000 Monte Carlo (MC) replications. We use a 5% nominal level ($\alpha = 5\%$) and approximate the critical value $c_{\widehat{T}}(.05)$ in (3.24) using 1,000 Gaussian multiplier draws within each replication.

[20] Given that the arctan function is close to constant for arguments far from zero, prior to applying the transformation, Bierens (1990) advocates centering and scaling the conditioning variables by their sample means and variances, respectively. This centering and scaling is strictly speaking not allowed for in our notation, which treats the weight function as known.

[21] For simplicity, the test only makes use of the CMRs arising from the marginal distributions of the Y_j's (given X). Three additional CMRs may be deduced from their joint distribution (given X).

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

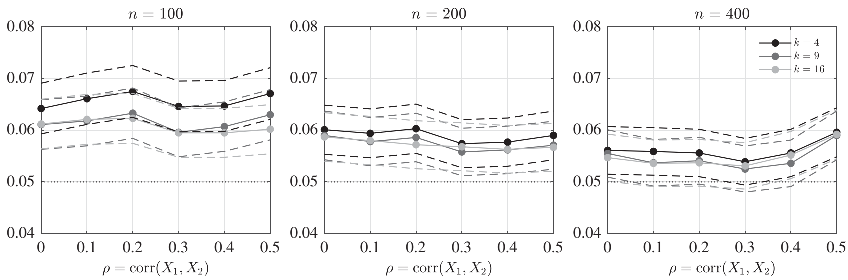

FIGURE 2. Size estimates (±2 MC standard errors).

When using a tensor-product basis of monomials, the (mean-square) optimal series length k_n^* grows at the rate $n^{\alpha/(2\alpha+1)}$, where α denotes the number of continuous derivatives of the target relative to its number of arguments (see, e.g., Newey, 1997, p. 151). Since each target is here analytic and bivariate, k_n^* is of order $n^{1/2}$. The series length set $\{4, 9, 16\}$ allows such sequences. Consider, for example, the formula $k = \lfloor \sqrt{Cn^{1/2}} \rfloor^2$, where $\lfloor a \rfloor$ is the largest integer less than or equal to a. Then $C = \frac{1}{2}$ produces $k = 4, 4$, and 9 for $n = 100, 200$, and 400, respectively, whereas $C = 1$ leads to $k = 9, 9$, and 16.

4.3. Results

Figure 2 shows the size estimates of the test for each sample size and series length as a function of the public information correlation level. The test is oversized by 1–1.5 percentage points for $n = 100$. For this (limited) sample size, the amount of overrejection may depend on the choice of series length by about half a percentage point. However, as the sample size increases, the size estimates appear to converge toward the nominal level across all series lengths and all correlation levels, except perhaps $\rho = 0.5$.

Figure 3 plots the power of the test when $\delta_0 = -0.5$. The power may depend on the choice of series length by upward of 10 percentage points. As the sample size increases, the power appears to converge to one for all series lengths and all correlations. This convergence is expected, since the test is consistent against all deviations from the null.

Lastly, we briefly explore the local power of the test developed in this paper. While the proposed test is not formally proved to exhibit nontrivial local power, in Figure 4, we depict estimates of its local power for the sequence of alternatives resulting from the now n-dependent $\delta_0 = -5/\sqrt{n}$. The test does have nontrivial local power (of 20%–40%), at least against this particular sequence of alternatives. Moreover, its local power appears stable across correlations as well as series lengths, at least for $n = 400$.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

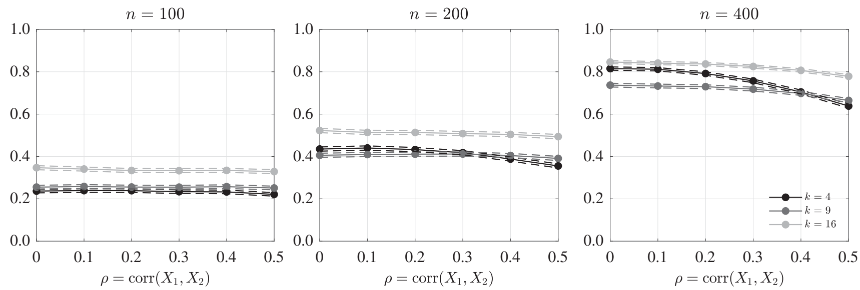

FIGURE 3. Global power estimates (±2 MC standard errors).

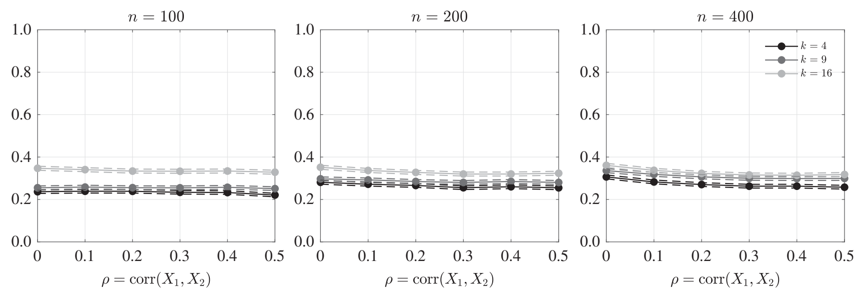

FIGURE 4. Local power estimates (±2 MC standard errors).

4.4. Partially Linear Model

To facilitate numerical comparison with existing tests, we also consider the model in Song (2010) as reported in Bravo (2012, Sect. 4.3). The DGP is here

$$Y = r(X_1, \gamma) + \tau(X_2) + \varepsilon_i,$$

where $\tau(x_2) = 2\Phi(x_2) - 1$, Φ being the standard normal CDF, $X_j = \widetilde{X}_j + .8\zeta_j + \zeta_0$, $j \in \{1,2\}$, $\widetilde{X}_j \sim \mathrm{U}(-1,1)$, $j \in \{1,2\}$, $\zeta_j \sim \mathrm{N}(0,1)$, $j \in \{0,1,2\}$, $\varepsilon \sim \mathrm{N}(0,.25)$, and the random variables $(\widetilde{X}_1, \widetilde{X}_2, \zeta_0, \zeta_1, \zeta_2, \varepsilon)$ are mutually independent. The function $r(x_1, \gamma) = \gamma a(x_1) + (1-\gamma)x_1$ with a being the nonlinear function $a(x_1) = 4\phi(x_1) - 2$, where ϕ is the N(0,.25) probability density function (PDF) (Song, 2010, p. 80). The null hypothesis of partial linearity (in X_1) holds when $\gamma = 0$. When $\gamma \neq 0$, the data are generated under the alternative. As in Bravo (2012, p. 16), we use $\gamma \in \{.05, .15\}$.

Partial linearity $\mathrm{E}[Y|X] = \beta X_1 + \tau(X_2)$ implies the CMR $\mathrm{E}[\rho(Z, \beta, h^*(X_2)|X] = 0$ for $\rho(z, \beta, h(x_2)) = y - h_Y(x_2) - \beta\{x_1 - h_{X_1}(x_2)\}$, where $h_Y^*(X_2) := \mathrm{E}[Y|X_2]$ and $h_{X_1}^*(X_2) := \mathrm{E}[X_1|X_2]$. The pseudo-truth, here available in closed

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

form, is

$$\beta_0 = \frac{\mathrm{E}\left[\left\{X_1 - h^*_{X_1}(X_2)\right\}\left\{Y - h^*_Y(X_2)\right\}\right]}{\mathrm{E}\left[\left\{X_1 - h^*_{X_1}(X_2)\right\}^2\right]},$$

which we estimate using Robinson's (1988) two-step procedure, except that we use series (instead of kernel) estimation for the nonparametric first step. As in Section 4.2, we use exponential weighting combined with a preliminary arctan transformation of each (otherwise unbounded) conditioning variable. Differentiation shows that

$$\begin{aligned} g(t,Z_i) &= \{\omega(t,X_i) - \mathrm{E}[\omega(t,X)|X_{2i}]\}\rho\left(Z_i,\beta_0,h^*(X_{2i})\right) \\ &\quad - \mathrm{E}\left[\omega(t,X)\left\{X_1 - h^*_{X_1}(X_2)\right\}\right]\frac{\left\{X_{1i} - h^*_{X_1}(X_{2i})\right\}\rho\left(Z_i,\beta_0,h^*(X_{2i})\right)}{\mathrm{E}\left[\left\{X_1 - h^*_{X_1}(X_2)\right\}^2\right]}, \end{aligned}$$

each of which is plug-in estimated and then fed the bootstrap. For series estimation, we here use the monomial basis, i.e., $p^k(x_2) = (1, x_2, x_2^2, \ldots, x_2^{k-1})'$. To compare with Bravo (2012) and Song (2010), we consider sample sizes $n \in \{100, 300\}$. We again use $\alpha = 5\%$, 10,000 MC replications, and 1,000 Gaussian multiplier draws.

Table 1 shows the finite-sample size and size-adjusted power of the test proposed in this paper resulting from series lengths $k \in \{2, 4, \ldots, 10\}$, which contains the range of k-values in Song (2010). Here, D^{KS}_{GEL} and D^{CM}_{GEL}, $GEL \in \{EL, EU, ET\}$ are Bravo's (2012, eqn. (4.1)) generalized empirical likelihood statistics of the KS and CM forms, respectively, and the KS- and CM-type statistics KS_Ω and CM_Ω are defined in his equation (2.8). Finally, KS^{ADF} and CM^{ADF} are the asymptotically distribution-free KS and CM statistics based on Song's (2010) conditional martingale transform.

The table shows that, unlike the existing tests, the test proposed in this paper is here correctly sized at $n = 300$, at least for large enough k. (For $k = 2$, the resulting linear approximations to CEFs are insufficient.) However, no test stands out in terms of (size-corrected) power. Specifically, the test proposed here has slightly lower power for $n = 100$ but generally (and in some cases much) greater power for $n = 300$.

5. EMPIRICAL ILLUSTRATION: ENTRY OF DISCOUNT STORES

As a proof of concept, we apply the specification test developed in this paper to an entry game between Walmart (W) and Kmart (K) discount stores using the Jia (2008) dataset. The setup is similar to Ellickson and Misra (2011, Sect. 4), who focus on the nuts and bolts of estimation. As in their paper, we consider a (much) simplified version of Jia's model, in which the two chains $j \in \{W, K\}$ make independent entry decisions $y_j = 1$ ("enter") or $= 0$ ("do not") across a collection

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

TABLE 1. Partially linear model: Size and size-adjusted power (in percent)

n	100			300		
γ	0	.05	.15	0	.05	.15
D_{EL}^{KS}	7.2	13.0	48.4	6.2	18.0	64.0
D_{EU}^{KS}	7.2	13.8	49.1	6.4	18.0	62.8
D_{ET}^{KS}	7.2	13.4	48.7	6.3	17.7	61.9
KS_{Ω}	7.4	11.2	42.8	6.7	16.4	56.0
KS^{ADF}	7.2	9.2	40.1	6.2	14.0	53.0
D_{EL}^{CM}	7.0	15.2	50.9	6.2	19.2	64.1
D_{EU}^{CM}	7.0	15.2	51.2	6.2	18.6	64.3
D_{ET}^{CM}	7.1	14.8	50.6	6.2	18.7	62.6
CM_{Ω}	7.0	13.4	49.4	6.3	18.3	62.2
CM^{ADF}	6.9	11.0	45.0	6.1	20.1	59.7
$k=2$	7.6	8.1	31.9	9.0	13.7	73.4
$k=4$	6.9	9.5	46.2	6.1	20.0	91.0
$k=6$	6.7	10.0	48.1	5.4	21.6	93.4
$k=8$	7.2	9.8	47.5	5.3	22.0	93.8
$k=10$	7.2	10.0	47.7	5.4	21.8	93.6

Notes: Columns correspond to finite-sample size ($\gamma = 0$) and size-adjusted power ($\gamma \neq 0$). Rows D_{EL}^{KS} to CM^{ADF} are imported from Bravo (2012, Table 9) (PDF alternative) and rounded to nearest decimal. The nominal level is 5%.

of counties (markets). Like Jia, we consider the $n = 2,065$ counties in which both Walmart and Kmart operate at most one store. Akin to the Section 4 simulations, player profits are parameterized as

$$u_j\left(y_j, y_{-j}, x_c, z_j, \varepsilon_j; \beta\right) = y_j\left(x_c'\beta_c + z_j'\beta_j + \gamma y_{-j} + \varepsilon_j\right), \quad j \in \{W, K\},$$

where market characteristics common to both firms enter through x_c, firm characteristics through z_W/z_K, and $(\beta_c, \beta_W, \beta_K, \gamma) =: \beta$ are model parameters. Firm-specific profit functions accommodate asymmetric players, such as the here more dominant Walmart and relatively weak Kmart (present in 47 and 17 pct. of the counties, respectively). We include three market characteristics: population (`pop`), retail sales per capita (`spc`), and fraction of urban population (`urban`). On top of firm-specific intercepts, we include the distance to Benton County, Arkansas (`dBenton`; the location of Walmart headquarters) and a dummy for the southern region (`south`) as Walmart characteristics, and a dummy for the Mid-West (`midwest`) as Kmart characteristics.[22] Here (`pop`, `spc`, `urban`, `dBenton`, `south`, `midwest`) is treated as public

[22] See Jia (2008) for a detailed discussion of the industry, market definition, and covariate relevance.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

TABLE 2. Bootstrap p-values (in percent) for correct specification of discount store market entry decisions

Transform	Description	rank P	p-value (%)
`Linear`	Constant, levels	7	.007
`Interaction`	Constant, levels, pairwise interactions	21	.022
`Pure Quadratic`	Constant, levels, squares	11	.004
`Quadratic`	At most second-order terms	25	.012
`Cubic`	At most third-order terms	65	.140
`Quartic`	At most fourth-order terms	140	.054

information (X), whereas ε_j is treated as private to Player j and unobserved by the econometrician. The ε_j's are here taken independent standard Logistic, which in a BNE leads to a decision rule akin to (4.1) and CCPs similar to (4.2) with some player-specific coefficients allowed.

The construction of the test runs parallel to Section 4.2.[23] To show the impact of the choice of approximating functions, we consider various transformations of the original conditioning variables as described in Table 2. Since some powers and/or interactions of elements of X are redundant, for series estimation, we use Moore–Penrose inversion, and report the rank of the matrix $[p^k(X_i)^\top] =: P$ of approximating functions, which one may think of as the "effective k." Table 2 shows the p-values resulting from 250,000 bootstrap repetitions and each of the options for approximating functions considered. The null of correct specification is rejected at a 1% significance level across the board, which indicates that entry decisions do not result from the simple static discrete game with linearly specified profits considered here.[24] These model rejections may also stem from chain stores not making market-independent decisions, an observation which motivated the network structure in Jia (2008).

6. CONCLUSION AND DISCUSSION

In this paper, we develop an omnibus specification test for a class of semi- or nonparametric CMRs in part parameterized by CEFs. The test is a suitable semi-/nonparametric extension of the Bierens (1982) goodness-of-fit test of a parametric model for the conditional mean. Estimating conditional expectations using series methods, we construct a bootstrap-based test which is proved both asymptotically correctly sized and consistent against any fixed alternative. We implement our

[23] In calculating weights of the exponential form, prior to conducting an arctan transform, we center and scale the conditioning variables by their sample means and standard deviations so as to put them on the same scale, as advocated by Bierens (1990).

[24] Using $\omega(t,x) = \cos(t'x) + \sin(t'x)$ instead of weights of the exponential form yields essentially all-zero p-values (not reported), when studentizing (but not arctan transforming) the conditioning variables. All-zero p-values follow from using weights $\mathbf{1}(t \leqslant x)$, so this finding appears weight robust.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

procedure in a comprehensive simulation study testing the specification of a static game of incomplete information. The simulations by and large reproduce the asymptotic properties in small samples.

In an empirical application using data from Jia (2008), we study the validity of a game-theoretical model for discount store market entry, treating equilibrium beliefs of Walmart and Kmart as nonparametric conditional expectations. Our findings indicate that Walmart and Kmart entry decisions do not result from a simple static discrete game of incomplete information with linearly specified profits.

Our simulations also indicate that the test has nontrivial power versus root-n-local alternatives, although further effort is needed to investigate the local power properties of the test in a formal manner. Future research might also consider relaxing the assumption of root-n estimability of the parametric component under the null hypothesis, relaxing the requirement of differentiability to allow for non-smooth residual functions, and developing formally justified data-driven methods for choosing the series length(s).

The test proposed in this paper involves estimation of the influence function representation of the estimator of the parametric part of the model, $\sqrt{n}(\widehat{\beta}-\beta_0)$. One way to avoid influence function estimation is to use a correction of the weighting function ω as proposed in Sant'Anna and Song (2019). For example, such a correction can be applied when (i) ρ_ℓ is a residual from a regression and (ii) the explanatory variables entering the regression function are a transformation of X_ℓ. Both these conditions are satisfied for the game-theoretical model considered in the simulations and empirical application. The present paper could be extended so as to consider such correction of ω. However, a drawback of the Sant'Anna and Song (2019) approach is that the resulting test would not have power against some fixed specific alternatives, i.e., it would not be consistent.

A. APPENDIX

Appendix Abbreviations and Notation

To conserve space, we use the abbreviations CS, H, J, M, and T for the Cauchy–Schwarz, Hölder, Jensen, Markov, and triangle inequalities, respectively. CMT, LOIE, MVE, and MVT are short for the "continuous mapping theorem," "law of iterated expectations," "mean-value expansion," and "mean-value theorem," respectively. We also abbreviate "with probability approaching one" by wp $\to$ 1. We employ empirical process notation and write $\mathbb{E}_n[f] := \mathbb{E}_n[f(Z_i)]$ for the average $n^{-1}\sum_{i=1}^n f(Z_i)$ and $\mathbb{G}_n(f)$ for the centered and scaled average $\mathbb{G}_n(f) := \mathbb{G}_n[f(Z_i)] = \sqrt{n}(\mathbb{E}_n - \mathrm{E})[f]$. If need be, we subscript the expectation operator E to highlight which variables are integrated out (e.g., E_Z). $\|f\|_{n,2}$ is short for the empirical L^2-norm (i.e., $\|f\|_{n,2}^2 = \mathbb{E}_n[f^2]$). The bracketing number $N_{[\,]}(\varepsilon,\mathcal{F},L^2(P))$ is defined as in van der Vaart and Wellner (1996, p. 83). Whenever $\mathcal{F}$ has a square-integrable envelope F, we denote by $J_{[\,]}(\delta,\mathcal{F}_j,L^2(P))$ the bracketing integral $\int_0^\delta\{1+N_{[\,]}(\varepsilon\|F\|_{P,2},\mathcal{F},L^2(P))\}^{1/2}\mathrm{d}\varepsilon$. $\|A\|_{\mathrm{op}}$ is the operator norm of a matrix A induced by the

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

ℓ^2-norm for vectors. $C, C_1, C_2, \dots$ denote positive and finite constants, the meaning of which may change between appearances. $a \vee b$ means $\max\{a,b\}$, and $a \wedge b$ means $\min\{a,b\}$. For nonrandom sequences, the notation $a_n \lesssim b_n$ means that $|a_n| \leqslant Cb_n$ for C not depending on n. For potentially random sequences, the relation $X_n \lesssim_{\mathrm{P}} b_n$ means $X_n/b_n = O_{\mathrm{P}}(1)$, where $O_{\mathrm{P}}(1)$ denotes stochastic boundedness.

A.1. Proofs for Section 3.1

Lemma A.1. *If Assumption 3 holds, then for any $z \in \mathcal{Z}$ and any $h_\ell : \mathcal{W}_\ell \to \mathbf{R}^{d_\ell}$,*

$$\left|\rho_\ell(z, h_\ell(w_\ell)) - \rho_\ell\left(z, h_\ell^*(w_\ell)\right) - \frac{\partial}{\partial h_\ell'}\rho_\ell\left(z, h_\ell^*(w_\ell)\right)\left[h_\ell(w_\ell) - h_\ell^*(w_\ell)\right]\right|$$
$$\leqslant R_\ell(z)\, d_\ell^{(1+\gamma_\ell)/2} \max_{1\leqslant m\leqslant d_\ell} \|h_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^{1+\gamma_\ell},$$

where $\rho_\ell(z, v_\ell) := \rho_\ell(z, \beta_0, v_\ell)$.

Proof. Let $z \in \mathcal{Z}, h_\ell : \mathcal{W}_\ell \to \mathbf{R}^{d_\ell}$ be arbitrary. Then $h_\ell(w) \in \mathbf{R}^{d_\ell}$, so Assumption 3 and an MVE of $v_\ell \mapsto \rho_\ell(z, v_\ell)$ at $h_\ell(w)$ around $h_\ell^*(w_\ell)$ yield

$$\left|\rho_\ell(z, h_\ell(w_\ell)) - \rho_\ell\left(z, h_\ell^*(w_\ell)\right) - \left(\partial/\partial h_\ell'\right)\rho_\ell\left(z, h_\ell^*(w_\ell)\right)\left[h_\ell(w_\ell) - h_\ell^*(w_\ell)\right]\right|$$
$$= \left|\left[\frac{\partial}{\partial h_\ell'}\rho_\ell(z, \widetilde{h}_\ell(w_\ell)) - \frac{\partial}{\partial h_\ell'}\rho_\ell\left(z, h_\ell^*(w_\ell)\right)\right]\left[h_\ell(w_\ell) - h_\ell^*(w_\ell)\right]\right|$$
$$\leqslant R_\ell(z)\, \|\widetilde{h}_\ell(w_\ell) - h_\ell^*(w_\ell)\|^{\gamma_\ell}\, \|h_\ell(w_\ell) - h_\ell^*(w_\ell)\| \leqslant R_\ell(z)\, \|h_\ell(w_\ell) - h_\ell^*(w_\ell)\|^{1+\gamma_\ell}$$
$$\leqslant R_\ell(z)\, d_\ell^{(1+\gamma_\ell)/2} \max_{1\leqslant m\leqslant d_\ell} \|h_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^{1+\gamma_\ell},$$

where $\widetilde{h}_\ell(w_\ell)$ lies on the line segment connecting $h_\ell(w_\ell)$ and $h_\ell^*(w_\ell)$, thus satisfying $\|\widetilde{h}_\ell(w_\ell) - h_\ell^*(w_\ell)\| \leqslant \|h_\ell(w_\ell) - h_\ell^*(w_\ell)\|$. □

The following result is crucial to proving Lemma 1 (Asymptotic equivalence).

Lemma A.2. *If Assumptions 1–7 hold and H_0 is true, then for each $\ell \in \{1, \dots, L\}$,*

$$\|\sqrt{n}\widehat{M}_\ell - \sqrt{n}\mathbb{E}_n[g_\ell(\cdot, Z_i)]\|_{\mathcal{X}_\ell}$$
$$\lesssim_{\mathrm{P}} \max_{1\leqslant m\leqslant d_\ell}\Bigg\{\mathrm{E}[R_\ell(Z)]\sqrt{n}\|\widehat{h}_{\ell m} - h_{\ell m}^*\|_{\mathcal{W}_\ell}^{1+\gamma_\ell}$$
$$+ \left(\sum_{j=1}^{k_{\ell m,n}} \|p_{\ell,j}\|_{\mathcal{W}_\ell}^2\right)^{1/2}\left(\sqrt{k_{\ell m,n}/n} + k_{\ell m,n}^{-\alpha_{\ell m}}\right)$$
$$+ \sqrt{n} r_{h_{\ell m}, k_{\ell m,n}} \sup_{t_\ell \in \mathcal{X}_\ell} r_{\delta_{\ell m}, k_{\ell m,n}}(t_\ell) + \sqrt{\zeta_{\ell, k_{\ell m,n}}^2 k_{\ell m,n} \ln\left(k_{\ell m,n}\right)/n}$$
$$+ R_{\delta_{\ell m}, k_{\ell m,n}}\sqrt{\ln\left(k_{\ell m,n}/R_{\delta_{\ell m}, k_{\ell m,n}}\right)} + \zeta_{\ell, k_{\ell m,n}} r_{h_{\ell m}, k_{\ell m,n}}\Bigg\} + o_{\mathrm{P}}(1).$$

The proof of Lemma A.2 is long and technical in nature and has therefore been relegated to the Supplementary Material.

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

Proof of Lemma 1. The claim follows from Lemma A.2 and Assumption 7. □

Proof of Lemma 2. A multivariate central limit theorem (CLT) shows joint convergence of all marginals of the sequences of processes $\{\mathbb{G}_n[g_\ell(t_\ell, Z_i)]; t_\ell \in \mathcal{X}_\ell\}, n \in \mathbf{N}$, $\ell \in \{1,\dots,L\}$. To show $\mathcal{G}$ is Donsker, it therefore suffices to show that each

$$\mathcal{G}_\ell := \{z \mapsto g_\ell(t_\ell, z); t_\ell \in \mathcal{X}_\ell\}, \quad \ell \in \{1,\dots,L\},$$

is Donsker (cf. van der Vaart and Wellner, 1996, Prob. 1.5.3). In what follows, we therefore omit the subscript ℓ. Moreover, given that β_0 and h^* are held fixed throughout the argument, we write $\rho(z) := \rho(z, \beta_0, h^*(w))$, $(\partial/\partial\beta)\rho(z) := (\partial/\partial\beta)\rho(z, \beta_0, h^*(w))$, and $(\partial/\partial h)\rho(z) := (\partial/\partial h)\rho\left(z, \beta_0, h^*(w)\right)$. Let $g_1 = g(t_1, \cdot), g_2 = g(t_2, \cdot) \in \mathcal{G}$ be arbitrary. Then, by T and CS, followed by J, CS, and Assumption 2,

$$\begin{aligned}|g(t_1,z) - g(t_2,z)| &\leqslant |\omega(t_1,x) - \omega(t_2,x)|\,|\rho(z)| + \|b(t_1) - b(t_2)\|\,\|s(z)\| \\ &\quad + \|y - h^*(w)\|\,\|\delta(t_1,w) - \delta(t_2,w)\| \\ &\leqslant C\Big(|\rho(z)| + \mathrm{E}[\|(\partial/\partial\beta)\rho(Z)\|]\,\|s(z)\| \\ &\quad + \|y - h^*(w)\|\mathrm{E}[\|(\partial/\partial h)\rho(Z)\|\,|W = w]\Big)\|t_1 - t_2\|,\end{aligned}$$

with s stemming from Assumption 1. Write the right-hand side as $G(z)\|t_1 - t_2\|$. Then, taking the expectation and using the inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$ repeatedly alongside the integrability and boundedness parts of Assumptions 1 and 3, we see that $G(Z)^2$ is integrable. Given that $\mathcal{T}$ is compact (Assumption 2), hence bounded, the Donsker property of $\mathcal{G}$ now follows from van der Vaart (2000, Example 19.7). □

Proof of Theorem 1. Under H_0, $\mathrm{E}[g(\cdot, Z)] \equiv 0$, so Lemma 2 means that $G_n = \sqrt{n}\mathbb{E}_n[g(\cdot, Z_i)] \rightsquigarrow G_0$ in $\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)$. Donsker's theorem shows that $\sqrt{n}(\widehat{F}_X - F_X) \rightsquigarrow G_{F_X}$ in the Skorokhod space $D([-\infty,\infty]^{d_x})$, where d_x is the number of distinct elements of the X_ℓ's and G_{F_X} is a centered Gaussian process indexed by $\mathbf{R}^{d_x}$ with covariance kernel $\left(x, x'\right) \mapsto \mathrm{P}\left(X \leqslant x \cap X \leqslant x'\right) - F_X(x)F_X\left(x'\right)$. A multivariate CLT establishes joint convergence of the marginals of the above processes from which we may deduce joint convergence in their product space (cf. van der Vaart and Wellner, 1996, Prob. 1.5.3). A CMT therefore shows weak convergence of $(\sqrt{n}\mathbb{E}_n[g(\cdot, Z_i)], \{\sqrt{n}(\widehat{F}_{X_\ell} - F_{X_\ell})\}_1^L)$ in $[\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)] \times [\times_{\ell=1}^L D([-\infty,\infty]^{d_{x,\ell}})]$ to a $2L$-variate centered Gaussian process. Lemma 1 implies that $(\sqrt{n}\widehat{M}, \{\sqrt{n}(\widehat{F}_{X_\ell} - F_{X_\ell})\}_1^L)$ has the same distributional limit. Let $BV_K(\mathcal{A})$ be the set of real-valued functions on $\mathcal{A}$ which are nondecreasing in each argument (holding the other arguments fixed) and of variation no more than $K \in (0,\infty)$. Then the functional $\phi : [\times_{\ell=1}^L C(\mathcal{X}_\ell)] \times [\times_{\ell=1}^L BV_1(\mathcal{X}_\ell)] \subseteq [\times_{\ell=1}^L L^\infty(\mathcal{X}_\ell)] \times [\times_{\ell=1}^L D([-\infty,\infty]^{d_{x,\ell}})] \to \mathbf{R}$ defined by $\phi(\{m_\ell\}_1^L, \{f_\ell\}_1^L) := \sum_{\ell=1}^L \int_{\mathcal{X}_\ell} m_\ell^2 \mathrm{d}f_\ell$ is second-order Hadamard differentiable at $(0, \{F_{X_\ell}\}_1^L) \in [\times_{\ell=1}^L C(\mathcal{X}_\ell)] \times [\times_{\ell=1}^L BV_1(\mathcal{X}_\ell)]$ with vanishing first-order Hadamard derivative and second-order Hadamard derivative $\phi''_{(0,\{F_{X_\ell}\}_1^L)} : [\times_{\ell=1}^L C(\mathcal{X}_\ell)] \times [\times_{\ell=1}^L BV_1(\mathcal{X}_\ell)] \to \mathbf{R}$ given by $\phi''_{(0,\{F_{X_\ell}\}_1^L)}(h_1, h_2) := 2\sum_{\ell=1}^L \int_{\mathcal{X}_\ell} h_{1\ell}^2 \mathrm{d}F_{X_\ell}$. The (second-order) functional delta method therefore produces

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

$$
\begin{aligned}
T_n &= n\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}\widehat{M}_\ell^2 \mathrm{d}\widehat{F}_{X_\ell} = n[\phi(\widehat{M},\{\widehat{F}_{X_\ell}\}_1^L)-\phi(0,\{F_{X_\ell}\}_1^L)]\\
&= \tfrac{1}{2}\phi''_{(0,\{F_{X_\ell}\}_1^L)}(\sqrt{n}\widehat{M},\{\sqrt{n}(\widehat{F}_{X_\ell}-F_{X_\ell})\}_1^L))+o_{\mathrm{P}}(1)\\
&= \sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}(\sqrt{n}\widehat{M}_\ell)^2\mathrm{d}F_{X_\ell}+o_{\mathrm{P}}(1).
\end{aligned}
$$

The claimed null distribution now follows from the CMT.

Arguments paralleling the proof of Lemma A.2 show that even under H_1 (where $\widehat{\beta}$ is consistent for β_0 but not necessarily asymptotically linear), both $\max_{1\leqslant \ell\leqslant L}\|M_\ell\|_{\mathcal{X}_\ell}<\infty$ and $\max_{1\leqslant \ell\leqslant L}\|\widehat{M}_\ell - M_\ell\|_{\mathcal{X}_\ell}\to_{\mathrm{P}} 0$. Decompose T_n/n as

$$
\frac{T_n}{n}=\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell^2\mathrm{d}\widehat{F}_{X_\ell}+\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}(\widehat{M}_\ell-M_\ell)^2\mathrm{d}\widehat{F}_{X_\ell}+2\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell(\widehat{M}_\ell-M_\ell)\mathrm{d}\widehat{F}_{X_\ell}.
$$

The calculations

$$
\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}(\widehat{M}_\ell-M_\ell)^2\mathrm{d}\widehat{F}_{X_\ell}\leqslant L\max_{1\leqslant\ell\leqslant L}\|\widehat{M}_\ell-M_\ell\|_{\mathcal{X}_\ell}^2\xrightarrow{\mathrm{P}}0,
$$

$$
\left|\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell(\widehat{M}_\ell-M_\ell)\mathrm{d}\widehat{F}_{X_\ell}\right|\leqslant L\max_{1\leqslant\ell\leqslant L}\|M_\ell\|_{\mathcal{X}_\ell}\max_{1\leqslant\ell\leqslant L}\|\widehat{M}_\ell-M_\ell\|_{\mathcal{X}_\ell}^2\xrightarrow{\mathrm{P}}0,
$$

show that $T_n/n=\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell^2\mathrm{d}\widehat{F}_{X_\ell}+o_{\mathrm{P}}(1)$. The law of large numbers (LLN) then shows that

$$
\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell^2\mathrm{d}\widehat{F}_{X_\ell}=\sum_{\ell=1}^{L}\mathbb{E}_n\left[M_\ell(X_{\ell i})^2\right]\xrightarrow{\mathrm{P}}\sum_{\ell=1}^{L}\mathrm{E}\left[M_\ell(X_{\ell i})^2\right]=\sum_{\ell=1}^{L}\int_{\mathcal{X}_\ell}M_\ell^2\mathrm{d}F_{X_\ell},
$$

which is positive under H_1 by the choice of weights (Assumption 2). □

A.2. Proofs for Section 3.2

The proof of Lemma 3 parallels that of Lemma A.2 with some added complexity due to the parametric estimator not necessarily being asymptotically linear, the presence of multipliers, and the error introduced from estimating the g_ℓ's and recentering at the sample values. Like the proof of Lemma A.2, we relegate the long and technical argument to the Supplementary Material.

Proof of Lemma 4. Given that $\mathcal{G}$ is Donsker (Lemma 2), Kosorok (2008, Thm. 10.4(iv)) implies that $\mathbb{G}_n''\rightsquigarrow_{\mathrm{P},\xi}\mathbb{G}$ in $\times_{\ell=1}^{L}L^\infty(\mathcal{G}_\ell)$, where $\mathbb{G}_n''(g):=n^{-1/2}\sum_{i=1}^{n}\xi_i\{g(Z_i)-\mathrm{E}[g(Z)]\}$ and $\mathbb{G}$ is an L-variate zero-mean Gaussian process with covariance kernel $\mathrm{E}[g(Z)g'(Z)']-\mathrm{E}[g(Z)]\mathrm{E}[g'(Z)'], g,g'\in\times_{\ell=1}^{L}\mathcal{G}_\ell$. Since we may identify each $\mathcal{G}_\ell$ with $\mathcal{X}_\ell$, this result is equivalent to $G_n^*\rightsquigarrow_{\mathrm{P},\xi}G$ in $\times_{\ell=1}^{L}L^\infty(\mathcal{X}_\ell)$, where G is a centered Gaussian process with covariance kernel $\mathbb{C}$ given in (3.16). Lemmas 3 and S.8 now imply $\widehat{G}_n\rightsquigarrow_{\mathrm{P},\xi}G$ in $\times_{\ell=1}^{L}L^\infty(\mathcal{X}_\ell)$. An application of the (second-order) delta method for the bootstrap now establishes that $\widehat{T}$ converges weakly in probability to T. Invoking continuity (Assumption 9),

https://doi.org/10.1017/S0266466622000615 Published online by Cambridge University Press

Kosorok (2008, Lem. 10.11) therefore shows that the $F_{\widehat{T}}$ converges in probability to F_T pointwise on $[0,\infty)$. Fix $\varepsilon > 0$ and $\alpha \in (0,1)$. Since F_T is continuous, by the hypothesis that it is also increasing at $c_T(\alpha)$, there exists $r_1 \in \mathbf{R}$ such that $c_T(\alpha) - \varepsilon < r_1 < c_T(\alpha)$ and $F_T(r_1) < 1-\alpha$. Then $F_{\widehat{T}}(r_1) < 1-\alpha$ wp $\to 1$, which implies $c_T(\alpha) - \varepsilon < r_1 \leqslant c_{\widehat{T}}(\alpha)$ wp $\to 1$. In particular, $\mathrm{P}(c_{\widehat{T}}(\alpha) \geqslant c_T(\alpha) - \varepsilon) \to 1$. Similarly, there exists $r_2 \in \mathbf{R}$ such that $c_T(\alpha) < r_2 < c_T(\alpha) + \varepsilon$ and $1-\alpha < F_T(r_2)$. Then $1-\alpha < F_{\widehat{T}}(r_2)$ wp $\to 1$, which implies $c_{\widehat{T}}(\alpha) \leqslant r_2 < c_T(\alpha) + \varepsilon$ wp $\to 1$. In particular, $\mathrm{P}(c_{\widehat{T}}(\alpha) < c_T(\alpha) + \varepsilon) \to 1$. It follows that

$$\begin{aligned}
&\limsup_{n\to\infty} \mathrm{P}(|c_{\widehat{T}}(\alpha) - c_T(\alpha)| \geqslant \varepsilon) \\
&\quad \leqslant \limsup_{n\to\infty} \mathrm{P}(c_{\widehat{T}}(\alpha) \geqslant c_T(\alpha) + \varepsilon) + \limsup_{n\to\infty} \mathrm{P}(c_{\widehat{T}}(\alpha) \leqslant c_T(\alpha) - \varepsilon) = 0.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, the lemma follows. □

A.3. Proofs for Section 3.3

Proof of Theorem 2. Fix $\alpha \in (0,1)$. Under H_0, $T_n \to_d T_0$ (Theorem 1). Letting F_{T_0} denote the CDF of T_0, by F_{T_0} being continuous on $\mathbf{R}$ (using Assumption 9) and increasing at $c_{T_0}(\alpha)$ (by hypothesis), it follows from Lemma 4 that $c_{\widehat{T}}(\alpha) \to_{\mathrm{P}} c_{T_0}(\alpha) \in (0,\infty)$. Slutsky's theorem then shows $T_n - c_{\widehat{T}}(\alpha) \to_d T_0 - c_{T_0}(\alpha)$, which establishes the first claim.

Under H_1, $T_n/n \to_{\mathrm{P}} \sum_{\ell=1}^{L} \int_{\mathcal{X}_\ell} M_\ell^2 \mathrm{d}F_{X_\ell} \in (0,\infty)$ (Theorem 1). Since F_T is increasing at $c_T(\alpha)$, Lemma 4 yields $c_{\widehat{T}}(\alpha) \to_{\mathrm{P}} c_T(\alpha) \in (0,\infty)$. In particular, $c_{\widehat{T}}(\alpha) \lesssim_{\mathrm{P}} 1$, so for any $\varepsilon \in (0,1)$, there exists $K_\varepsilon \in (0,\infty)$ such that $\limsup_{n\to\infty} \mathrm{P}(c_{\widehat{T}}(\alpha) > K_\varepsilon) \leqslant \varepsilon$. Letting $\varepsilon \in (0,1)$ be arbitrary, we see that

$$\begin{aligned}
\mathrm{P}\left(T_n \leqslant c_{\widehat{T}}(\alpha)\right) &= \mathrm{P}\left(T_n \leqslant c_{\widehat{T}}(\alpha) \cap c_{\widehat{T}}(\alpha) \leqslant K_\varepsilon\right) + \mathrm{P}\left(T_n \leqslant c_{\widehat{T}}(\alpha) \cap c_{\widehat{T}}(\alpha) > K_\varepsilon\right) \\
&\leqslant \mathrm{P}(T_n \leqslant K_\varepsilon) + \mathrm{P}\left(c_{\widehat{T}}(\alpha) > K_\varepsilon\right) \\
&= \mathrm{P}(T_n/n \leqslant K_\varepsilon/n) + \mathrm{P}\left(c_{\widehat{T}}(\alpha) > K_\varepsilon\right),
\end{aligned}$$

which—by the preceding remarks—implies $\limsup_{n\to\infty} \mathrm{P}(T_n \leqslant c_{\widehat{T}}(\alpha)) \leqslant \varepsilon$. The second claim now follows from taking $\varepsilon \to 0_+$. □

SUPPLEMENTARY MATERIAL

To view supplementary material for this paper, please visit: https://doi.org/10.1017/S0266466622000615.

REFERENCES

Ahn, H. & C.F. Manski (1993) Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of Econometrics* 56, 291–321.

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

Bajari, P., H. Hong, J. Krainer, & D. Nekipelov (2010) Estimating static models of strategic interactions. *Journal of Business & Economic Statistics* 28, 469–482.

Belloni, A., V. Chernozhukov, D. Chetverikov, & K. Kato (2015) Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics* 186, 345–366.

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

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

Bierens, H.J. (2017) *Econometric Model Specification: Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference*. World Scientific Publishing Co. Pte. Ltd.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.

Bjorn, P.A. & Q.H. Vuong (1984) Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation. Social Science Working paper 537, Caltech.

Bravo, F. (2012) Generalized empirical likelihood testing in semiparametric conditional moment restrictions models. *The Econometrics Journal* 15, 1–31.

Card, D. & L. Giuliano (2013) Peer effects and multiple equilibria in the risky behavior of friends. *Review of Economics and Statistics* 95, 1130–1149.

Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 5549–5632.

Chen, X. & D. Pouzo (2009) Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics* 152, 46–60.

Chen, X. & D. Pouzo (2012) Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* 80, 277–321.

Chen, X. & D. Pouzo (2015) Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* 83, 1013–1079.

Davydov, Y.A., M.A. Lifshits, & N.V. Smorodina (1998) *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society.

DeVore, R.A. & G.G. Lorentz (1993) *Constructive Approximation*. Grundlehren der mathematischen Wissenschafte, vol. 303. Springer Science & Business Media.

Donald, S.G. & W.K. Newey (1994) Series estimation of semilinear models. *Journal of Multivariate Analysis* 50, 30–40.

Ellickson, P.B. & S. Misra (2011) *Structural workshop paper—estimating discrete games*. *Marketing Science* 30, 997–1010.

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

Hahn, J., H.R. Moon, & C. Snider (2017) LM test of neglected correlated random effects and its application. *Journal of Business & Economic Statistics* 35, 359–370.

Hall, A.R. & A. Inoue (2003) The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114, 361–394.

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

Hong, H. & J. Li (2022) Rate-Adaptive Bootstrap for Possibly Misspecified GMM. Working paper, UC Santa Cruz.

Honoré, B.E. & E. Kyriazidou (2000) Panel data discrete choice models with lagged dependent variables. *Econometrica* 68, 839–874.

Honoré, B.E. & E. Tamer (2006) Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74, 611–629.

Horowitz, J.L. (2006) Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica* 74, 521–538.

Jia, P. (2008) What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica* 76, 1263–1316.

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

Kristensen, D. (2011) Semi-nonparametric estimation and misspecification testing of diffusion models. *Journal of Econometrics* 164, 382–403.

Lapenta, E. (2021) A Bootstrap Specification Test for Semiparametric Models with Generated Regressors. Working paper, CREST-ENSAE.

Li, Q., C. Hsiao, & J. Zinn (2003) Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.

Lorentz, G.G. (1966) *Approximation of Functions*. Holt, Rinehart and Winston.

Manski, C.F. (1991) Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, pp. 259–275.

Newey, W.K. (1994) The asymptotic variance of semiparametric estimators. *Econometrica* 62, 1349–1382.

Newey, W.K. (1995) Convergence rates for series estimators. In G. Maddala, P.C.B. Phillips, and T. Srinivasan (eds), *Advances in Econometrics and Quantitative Economics*. Blackwell, pp. 254–275.

Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.

Robinson, P.M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.

Sant'Anna, P.H. & X. Song (2019) Specification tests for the propensity score. *Journal of Econometrics* 210, 379–404.

Schumaker, L. (2007) *Spline Functions: Basic Theory*. Cambridge University Press.

Seim, K. (2006) An empirical model of firm entry with endogenous product-type choices. *The Rand Journal of Economics* 37, 619–640.

Song, K. (2010) Testing semiparametric conditional moment restrictions using conditional martingale transforms. *Journal of Econometrics* 154, 74–84.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

Stone, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics*, 689–705.

Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25, 613–641.

Sweeting, A. (2009) The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *The Rand Journal of Economics* 40, 710–742.

Timan, A.F. (1963) *Theory of Approximation of Functions of a Real Variable*. International Series of Monographs in Pure and Applied Mathematics, vol. 34. Courier Corporation.

van der Vaart, A.W. (2000) *Asymptotic Statistics*, vol. 3. Cambridge University Press.

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

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