Journal of Econometrics 154 (2010) 74–84

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing semiparametric conditional moment restrictions using conditional martingale transforms

Kyungchul Song*

Department of Economics, University of Pennsylvania, United States

ARTICLE INFO

Article history:
Received 1 April 2008
Received in revised form
25 June 2009
Accepted 14 July 2009
Available online 18 July 2009

JEL classification:
C12
C14
C52

Keywords:
Semiparametric models
Conditional moment restrictions
Martingale transform
Asymptotically distribution-free tests

ABSTRACT

This paper studies conditional moment restrictions that contain unknown nonparametric functions, and proposes a general method of obtaining asymptotically distribution-free tests via martingale transforms. Examples of such conditional moment restrictions are single index restrictions, partially parametric regressions, and partially parametric quantile regressions. This paper introduces a conditional martingale transform that is conditioned on the variable in the nonparametric function, and shows that we can generate distribution-free tests of various semiparametric conditional moment restrictions using this martingale transform. The paper proposes feasible martingale transforms using series estimation and establishes their asymptotic validity. Some results from a Monte Carlo simulation study are presented and discussed.

1. Introduction

This paper proposes a general method to produce tests of semiparametric conditional moment restrictions that are asymptotically distribution-free. The method is obtained by extending the approach of martingale transforms to accommodate nonparametric components in the moment restrictions. This extension involves introduction of a conditional martingale transform that is conditioned on the random variable in the nonparametric function.

To be more specific, let $\{S_i\}_{i=1}^n$ be a random sample from a distribution P of a random vector S taking values in $\mathbf{R}^{d_S}$. Let S be constituted by (possibly overlapping) subvectors X and W which take values in $\mathbf{R}^{d_X}$ and $\mathbf{R}^{d_W}$. This paper focuses on the following type of null hypotheses:

$$H_0 : P\{\mathbf{E}[\rho_\alpha(S, h(X))|W] = 0\} = 1 \quad \text{for some } (\alpha, h) \in \mathcal{A} \times \mathcal{H}, \tag{1}$$

where $\rho_\alpha(\cdot,\cdot) : \mathbf{R}^{d_S+1} \to \mathbf{R}^{d_\rho}$ is a vector-valued function known up to a finite-dimensional parameter $\alpha \in \mathcal{A} \subset \mathbf{R}^{d_\alpha}$ and h belongs to an infinite-dimensional class $\mathcal{H}$ of real-valued functions $h(\cdot) : \mathbf{R}^{d_X} \to \mathbf{R}$. The function ρ_α is often called a *generalized residual function,* reminiscent of a residual in the regression. The function $\rho_\alpha(\cdot,\cdot)$ can be nonsmooth, even discontinuous in $(\alpha, \bar{h}) \in \mathcal{A} \times \mathbf{R}$, and hence the framework includes such models as semiparametric quantile regression models.

In particular, throughout this paper, the interest is confined to the case where h takes the following form:

$$h_{\beta,\tau}(x) \triangleq \tau(\lambda_\beta(x)), \tag{2}$$

where τ is an unknown function on $[0, 1]$ in an infinite dimensional space $\mathcal{T}$, and $\lambda_\beta : \mathbf{R}^{d_X} \to [0, 1]$ is a real valued function known up to $\beta \in \mathcal{B} \subset \mathbf{R}^{d_\beta}$. (Throughout the paper, the notation $\triangleq$ denotes a definitional relation.) This paper also assumes that for each $\beta \in \mathcal{B}$, $\lambda_\beta(X)$ is measurable with respect to the σ-field of W. Hence, the set-up of the null hypothesis is such that the nonparametric component h depends on X only through $\lambda_\beta(X)$ that is measurable with respect to the σ-field of W. It is assumed that the support of W lies in $[0, 1]^{d_W}$. By doing so, we lose no generality, for we can always take a strictly monotone transformation of W's marginals.

The alternative hypothesis is the negation of the null:

$$H_1 : P\{\mathbf{E}[\rho_\alpha(S, h(X))|W] = 0\} < 1 \quad \text{for any } (\alpha, h) \in \mathcal{A} \times \mathcal{H}, \tag{3}$$

where $\mathcal{H} \triangleq \{h_{\beta,\tau} : (\beta, \tau) \in \mathcal{B} \times \mathcal{T}\}$, with $h_{\beta,\tau}$ as defined in (2). The main focus of this paper rests on omnibus tests that are consistent against any models that violate the null restriction. However, our

* Corresponding address: Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297, United States. Tel.: +1 215 898 8206.
E-mail address: kysong@sas.upenn.edu.

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2009.07.002

framework can also be utilized for directional tests that focus on a specific subset of alternatives. (For the directional approach, see Stute (1997) and Stute et al. (1998) for example.)

In general, the testing problem of (1) is dealt with by employing a certain metric that measures the "distance" between the function $m(\cdot) \triangleq \mathbf{E}[\rho_\alpha(S, h(X))|W = \cdot]$ and zero. In particular, many studies have focused on the following type of metrics

$$\sup_{\bar{w}} \left\| \int m(w)\gamma_{\bar{w}}(w)\mathrm{d}F(w) \right\| \quad \text{or} \quad \int \left\| \int m(w)\gamma_{\bar{w}}(w)\mathrm{d}F(w) \right\|^2 \mathrm{d}F(\bar{w}), \tag{4}$$

using appropriate test functions $\gamma_{\bar{w}}(\cdot)$ indexed by $\bar{w}$ in the support of W. Here $\|\cdot\|$ denotes the Euclidean norm: $\|a\| = \sqrt{tr(aa^\top)}$ and F denotes the distribution function of W. The first choice leads to a Kolmogorov–Smirnov type test and the second, a Cramér–von Mises type test. In either case, the metric is conveniently represented as a functional of an infinite number of unconditional expectations and their sample version normalized by $\sqrt{n}$ forms an empirical process under the null hypothesis.

The testing problem of this type has very long been studied in the literature, if not with the same generality assumed in this paper. Two approaches have been mainly employed to test conditional moment restrictions in the literature: the smoothing approach and the empirical process-based approach. The smoothing approach focuses on the local behavior of the conditional mean function by using its smoothed estimator. Tests using this approach are known to have good power against local alternatives of high frequency (Horowitz and Spokoiny, 2001; Guerre and Lavergne, 2002; Fan and Li, 2000) and are often asymptotically distribution-free. Studies adopting this approach include Härdle and Mammen (1993), Horowitz and Spokoiny (2001). Tripathi and Kitamura (2003) and Kitamura et al. (2004) applied the empirical likelihood approach to test conditional moment restrictions.

The empirical process-based approach formulates a conditional moment restriction as unconditional moment restrictions indexed by a set of functions. For earlier literature on goodness-of-fit tests, see, for example, Durbin (1973), Durbin et al. (1975), and Neuhaus (1976). Bierens (1990), Stute (1997) and Bierens and Ploberger (1997) considered testing nonlinear regression models, and Andrews (1997) proposed conditional goodness-of-fit tests. Also see Stinchcombe and White (1998) and Whang (2000). Recently Horowitz (2006) proposed a test of a conditional moment restriction when the regression function is identified through instrumental variables. As for semiparametric tests, Chen and Fan (1999) proposed tests of semiparametric and nonparametric conditional moments using the CLT for Hilbert-valued dependent random arrays established in Chen and White (1998). Stute and Zhu (2005) studied tests of single-index restrictions. See also bootstrap tests by Li et al. (2003) and Delgado and González Manteiga (2001).

It is well-known that many tests based on the metrics (4) are not asymptotically distribution-free. To see this, let

$$V_n^1(\gamma_w) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_w(W_i)\rho_{\hat{\alpha}}(S_i, \hat{h}(X_i)), \tag{5}$$

where $(\hat{\alpha}, \hat{h})$ is a consistent estimator for (α, h). Then the test statistics based on the metrics in (4) take the following form

$$T_{KS} \triangleq \max_w \|V_n^1(\gamma_w)\| \quad \text{or} \quad T_{CM} \triangleq \int \|V_n^1(\gamma_w)\|^2 \mathrm{d}F(w). \tag{6}$$

Under certain mild assumptions for the estimators $\hat{\alpha}$ and $\hat{h}$, the weak limit of $V_n^1(\gamma_w)$ (in the sense of Hoffman–Jorgensen, e.g. see van der Vaart and Wellner (1996)) is a Gaussian process but with a covariance kernel that depends on the particulars of the data generating process. Generally speaking, there are primarily two situations in which this type of complexity arises. First, the multivariate nature of the tests may cause complexity. It is well known that goodness-of-fit tests for a composite null hypothesis of multivariate distribution functions are not asymptotically distribution-free (Simpson, 1951; Rosenblatt, 1952). Second, complexity may arise due to the presence of estimation errors that leave their mark on the limiting distribution. Koenker and Xiao (2002) call this type of complexity the Durbin problem, observing that Durbin (1973) analyzed this phenomenon in the context of goodness-of-fit tests. In the case of testing conditional moment restrictions, the complexity is mainly due to the second cause, i.e., due to the estimation errors in $(\hat{\alpha}, \hat{h})$.

Since the pioneering papers by Khmaladze (1981, 1993), there have been numerous studies that employ a transform of the empirical process into an innovation martingale to generate asymptotically distribution-free tests.[1] For example, the martingale transform approach was applied to nonlinear regressions by Stute et al. (1998) and Khmaladze and Koul (2004), and to specifications of autoregressive processes and linear processes by Koul and Stute (1999) and Delgado et al. (2005). In the econometrics literature, an early working paper version of Bai (2003) was the first to apply this approach to specification tests, where he developed a testing procedure of parametric dynamic models. Bai and Ng (2001) proposed a test of conditional symmetry and Koenker and Xiao (2002) considered specification tests of quantile regression processes. See also Angrist and Kuersteiner (2004), Delgado et al. (2005) and Delgado and Stute (2007).[2]

The aim of this paper is to generalize this approach of innovation martingales and introduce asymptotically distribution-free tests of the null hypothesis of the form in (1) with h satisfying (2). The main contribution of this paper's proposal lies in its unprecedented generality, examples including single index restrictions, partially parametric regressions, and partially parametric quantile regressions. This paper demonstrates that the existing method of innovation martingales cannot be directly applied to this general framework. This paper proposes a conditional (nonorthogonal) projection of the empirical processes, which is called, here, *a conditional martingale transform*, and shows that conditional martingale transforms can be used to generate distribution-free tests of semiparametric conditional moment restrictions. For a feasible version of the transform, we consider series estimation of the conditional mean function and show that the feasible transform is asymptotically valid under general conditions.

In principle, the martingale transform approach and the method of bootstrap are not substitutes because we can always bootstrap the martingale-transformed tests. However, it is interesting to compare their small sample properties, for a martingale transform involves more estimation than bootstrap, and the asymptotic power functions are very different. Using a simple partial linear model, we performed Monte Carlo simulations to compare their size and power. We found that neither of the two approaches dominated the other: along a certain direction of alternatives, the martingale transform showed a conspicuously greater power than

1 See Cabaña and Cabaña (1997) for a general class of isometric nonorthogonal projections that encompasses the martingale transform of Khmaladze.

2 Phillips and Sun (2002), in a brief note, considered two detrending methods of semimartingale processes and showed that an efficient detrending is a nonorthogonal projection with respect to the inner product $\langle Z_1, Z_2\rangle = \int Z_1(t)Z_2(t)\mathrm{d}t$. This efficient detrending coincides with the martingale transform considered by Khmaladze (1981) when the detrending is performed over the future period $(t, 1]$. Bai (2003) offers a nice interpretation of the martingale transform in Khmaladze (1981) as a detrending operation in continuous time regressions.

the bootstrap approach, but this order of performance was easily reversed when another direction of alternatives was considered. This empirical result emphasizes the fact that the two approaches generate tests with very different power properties and should be viewed as complementary, not as substitutes.

The next section defines the scope of the testing framework, introduces conditional martingale transforms, and presents their properties and examples. In Section 3, some results from simulation studies are discussed. Section 4 concludes. All the technical proofs are relegated to the Appendix.

2. Conditional martingale transforms

2.1. The asymptotic representation

The null hypothesis in (1) is composite, demanding that every element (α, h) in $\mathcal{A} \times \mathcal{H}$ be checked against the conditional moment restriction. Under weak regularity conditions, we can reformulate the null hypothesis so that the moment condition is required to be satisfied by only one element in $\mathcal{A} \times \mathcal{H}$. Let $\mathcal{P}$ denote the whole class of probabilities to which the data generating process P for S belongs. Let $\mathcal{P}_0 \subset \mathcal{P}$ be the class of probabilities under the null hypothesis in (1). Let $\|\cdot\|_\infty$ denote the sup norm: $\|f\|_\infty = \sup_s \|f(s)\|$, where s runs in the domain of f.

Assumption 1. For each $P \in \mathcal{P}$ and the corresponding i.i.d. data $\{S_i(P)\}_{i=1}^n$, there exists an estimator $(\hat{\alpha}, \hat{h})(P)$, based on the data such that for each $P \in \mathcal{P}$, $\hat{\alpha} \to_P \alpha_P$ and $\|\hat{h} - h_P\|_\infty \to_P 0$ for some pseudo value $(\alpha_P, h_P) \in \mathcal{A} \times \mathcal{H}$, and for each $P \in \mathcal{P}_0$,

$$P\left\{\mathbf{E}\left[\rho_{\alpha_P}(S, h_P(X))|W\right] = 0\right\} = 1.$$

Assumption 1 requires that a single parameter $(\alpha_P, h_P) \in \mathcal{A} \times \mathcal{H}$ satisfying (1) be identified for each $P \in \mathcal{P}_0$, and that a pseudo-value $(\alpha_P, h_P) \in \mathcal{A} \times \mathcal{H}$ to which the estimator $(\hat{\alpha}, \hat{h})$ converges in probability exist for each $P \in \mathcal{P}/\mathcal{P}_0$. Under Assumption 1, (1) can be reformulated as

$$H_0 : P\left\{\mathbf{E}\left[\rho_{\alpha_P}(S, h_P(X))|W\right] = 0\right\} = 1. \quad (7)$$

Using appropriate test functions $\{\gamma_w(\cdot); w \in [0,1]^{d_W}\}$, we can rewrite (1) as

$$H_0 : \mathbf{E}[\gamma_w(W)\rho_{\alpha_P}(S, h_P(X))] = 0 \quad \text{for all } w \in [0,1]^{d_W}. \quad (8)$$

This formulation is useful because its sample version leads to an empirical process in (5). For example, Bierens (1990) and Bierens and Ploberger (1997) considered an exponential type test functions, and Andrews (1997) and Stute (1997) used indicator functions. See Escanciano (2006) for other interesting choices of test functions. For a general discussion, see Stinchcombe and White (1998). In this paper, we choose

$$\gamma_w(\cdot) = 1\{\cdot \le w\}.$$

We first establish an asymptotic representation of the semiparametric empirical process V_n^1 defined in (5). Let $\|\cdot\|_p$ denote the L_p-norm: $\|f\|_p \triangleq (\int |f|^p dP)^{1/p}$. For a class $\mathcal{F}$ of functions, the number $N_{[]}(\varepsilon, \mathcal{F}, L_p(P))$ denotes the L_p-bracketing number, i.e., the smallest number N of pairs of functions $(f_j, \Delta_j)_{j=1}^N$ in $L_p(P)$ such that $\|\Delta_j\|_p < \varepsilon$, and for all $f \in \mathcal{F}$, there exists a pair $(f_j, \Delta_j), j \in \{1, \ldots, N\}$ satisfying the inequality: $|f_j(x) - f(x)| < \Delta_j(x)$. The notation $N(\varepsilon, \mathcal{F}, \|\cdot\|_\infty)$ indicates the covering number. i.e. the smallest number of ε-balls with respect to $\|\cdot\|_\infty$ needed to cover $\mathcal{F}$.

Definition 1. (i) For a function $\varphi(\cdot) : \mathbf{R}^{d_S} \to \mathbf{R}$ and its estimator $\hat{\varphi}(\cdot)$, we say that $\hat{\varphi}$ satisfies RC$(r, p, \mathcal{J})$ (regularity condition) if (a) $\|\hat{\varphi} - \varphi\|_\infty = o_P(n^{-1/4})$ and (b) for some class $\mathcal{J}$, $\varphi \in \mathcal{J}$, $P\{\hat{\varphi} \in \mathcal{J}\} \to 1$ as $n \to \infty$, for some $C > 0$,

$$\log N_{[]}(\varepsilon, \mathcal{J}, L_p(P)) \le C\varepsilon^{-r}, \quad \text{for all } \varepsilon > 0, \quad (9)$$

and there exists an envelope $\bar{\varphi}$ for $\mathcal{J}$ such that $\|\bar{\varphi}\|_{p+\delta} < \infty$ for some $\delta > 0$.

(ii) For any estimator $\hat{\varphi}(s)$ of a matrix-valued function $\varphi(s)$, we say $\hat{\varphi}(s)$ satisfies RC$(r, p, \mathcal{J})$ if each entry $\hat{\varphi}_{jk}$ in $\hat{\varphi}$ as an estimator of φ_{jk} in φ satisfies RC$(r, p, \mathcal{J}_{jk})$, where $\mathcal{J} = \Pi_{j,k}\mathcal{J}_{jk}$.

For brevity, let $\lambda_P(X_i) \triangleq \lambda_{\beta_P}(X_i)$, and introduce the following notations

$$\rho_P(S) \triangleq \rho_{\alpha_P}(S; (\tau_P \circ \lambda_P)(X)),\ \hat{\rho}(S) \triangleq \rho_{\hat{\alpha}}(S; (\hat{\tau} \circ \lambda_{\hat{\beta}})(X)),$$
$$m_{\alpha,\beta,\tau}(w) \triangleq \mathbf{E}[\rho_\alpha(S; (\tau \circ \lambda_\beta)(X))|W = w],$$
$$\text{and} \quad m_P(w) \triangleq \mathbf{E}[\rho_P(S)|W = w].$$

Let $\hat{\tau}$ and $\hat{\beta}$ be estimators of τ_P and β_P. In semiparametric restrictions, there are many pieces of literatures that proposed estimators of τ_P and β_P. The approach of this paper does not hinge upon a specific method of estimation. The following assumption is all we need for these estimators.

Assumption 2. (i) $\hat{\tau} \circ \lambda_{\hat{\beta}}$ satisfies RC$(b_1, p, \mathcal{H})$ where $\mathcal{H} \triangleq \{\tau \circ \lambda_\beta : (\tau, \beta) \in \mathcal{T} \times \mathcal{B}\}$ for some $b_1 \in [0, 2)$ and $p > 4$ and for some class of functions $\mathcal{T}$.
(ii) $\Lambda \triangleq \{\lambda_\beta : \beta \in \mathcal{B}\}$ is uniformly bounded and $\|\lambda_{\hat{\beta}} - \lambda_P\|_\infty = o_P(n^{-1/4})$.

Assumption 2(i) is satisfied in many situations. The uniform consistency condition (a) in Definition 1 can be established by using the standard procedure as in the nonparametric estimation literatures. As for the condition (b) there for $\mathcal{H}$, observe that when $\mathcal{B}$ is bounded and $\lambda_\beta(\cdot)$ is Lipschitz in β with a bounded coefficient, the bracketing entropy of Λ is bounded by $-C \log \varepsilon$ for some constant C. When $\mathcal{T}$ is uniformly bounded and of bounded variation, we can immediately use Lemma A1 of Song (2009) to establish the bracketing entropy condition in (9). In many situations, one can directly obtain a bracketing entropy bound for $\mathcal{H}$ in terms of entropy bounds for $\mathcal{T}$ and for Λ. When the functions in $\mathcal{T}$ are differentiable up to a certain order, Andrews (1994) delineates sufficient conditions that ensure bracketing entropy conditions for $\mathcal{T}$ and the convergence $P\{\hat{\tau} \in \mathcal{T}\} \to 1$. See also van der Vaart and Wellner (1996). Uniform boundedness of Λ in Condition (ii) is innocuous because we can take a strictly increasing bounded map G to redefine $\tau' = \tau \circ G^{-1}$ and $\lambda'_\beta = G \circ \lambda_\beta$.

Assumption 3. (i) $\|\hat{\alpha} - \alpha_P\| = o_P(n^{-1/4})$ and $\mathcal{A}$ is compact.
(ii) There exist $s \in (b_1/2, 1]$ and $C > 0$ such that for each $(\alpha, t) \in \mathcal{A} \times [0, 1]$, for each $\delta > 0$, for some $p > 4$,

$$\mathbf{E}\left[\sup_{(\alpha,t') \in \mathcal{A}\times[0,1]:\|\alpha-\alpha'\|+\|t-t'\|<\delta} \left\|\rho_\alpha(S, t) - \rho_{\alpha'}(S, t')\right\|^p |X\right] \le C\delta^{sp}, \text{ a.s.}$$

Condition (i) is satisfied when $\hat{\alpha}$ is $\sqrt{n}$-consistent. Condition (ii) admits the case where $\rho_\alpha(\cdot, t)$ is discontinuous in (α, t) when the conditional distribution of S given X is absolutely continuous with a bounded density. This condition is closely related to the local uniform L_p-continuity condition used by Chen et al. (2003).

Let $\hat{h} \triangleq \hat{\tau} \circ \lambda_{\hat{\beta}}$ and $h_P \triangleq \tau_P \circ \lambda_P$. Given consistent estimators $(\hat{\alpha}, \hat{h})$ that converge in probability to (α_P, h_P) at a certain rate, it suffices to consider a shrinking neighborhood $N_P(\delta_n)$ of (α_P, h_P) in $\mathcal{A} \times \mathcal{H}$:

$$N_P(\delta_n) \triangleq B_1(\alpha_P, \delta_n) \times B_2(h_P, \delta_n), \quad \delta_n \to 0,$$

where $B_1(\alpha_P, \delta_n)$ and $B_2(h_P, \delta_n)$ are δ_n-neighborhoods of α_P and h_P with respect to $\|\cdot\|$ and $\|\cdot\|_\infty$.

Assumption 4. (i) $\lambda_P(X)$ is measurable with respect to the σ-field of W.

(ii) For any $\delta_n \to 0$ as $n \to \infty$, there exists a $(d_\alpha+1)\times d_\rho$ matrix-valued function $q(w)$ such that $\mathbf{E}\|q(W)\|^p < \infty$, $p > 4$, and

$$\mathbf{E}\left[\sup_{(\alpha,h)\in N_P(\delta_n)}\left\|\Delta m_{\alpha,h}(W) - q(W)^\top \begin{bmatrix}\alpha-\alpha_P\\ h(X)-h_P(X)\end{bmatrix}\right\|^2\right] = O(\delta_n^4),$$

where $\Delta m_{\alpha,h}(W) \triangleq m_{\alpha,h}(W) - m_P(W)$.

(iii) For all $\beta \in \{\beta \in \mathcal{B} : \|\lambda_\beta - \lambda_P\|_\infty < \delta_0\}$, the density function $f_\beta(s|\bar\lambda_1,\bar\lambda_2)$ of S conditional on $(W,\lambda_\beta(X),\lambda_P(X)) = (w,\bar\lambda_1,\bar\lambda_2)$ with respect to a σ-finite measure ω_β satisfies that for all $s\in\mathbf{R}^{d_S}$, for all $\bar\lambda_1,\bar\lambda_2\in\Lambda$, and for all $\delta>0$,

$$\sup_{\bar\lambda\in\bar\Lambda:|\bar\lambda-\bar\lambda_2|<\delta}\left|f_\beta(s|w,\bar\lambda_1,\bar\lambda) - f_\beta(s|w,\bar\lambda_1,\bar\lambda_2)\right| \le \varphi_\beta(w,\bar\lambda_1,\bar\lambda_2)\delta,$$

where φ_β is a real function that satisfies $\varphi_\beta(w,\bar\lambda_1,\bar\lambda_2) < Cf(w)$ with f denoting the density of W.

(iv) For some $\delta > 0$ and $p > 4$, $\mathbf{E}\left[\sup_{(\alpha,h)\in N_P(\delta)}\|\rho_\alpha(S,h(X))\|^p\right] < \infty$.

Condition (i) makes Condition (ii) plausible. This condition excludes certain models that allow X to be endogenous (e.g. Ai and Chen (2003) and Newey and Powell (2003)). Despite the restrictiveness, Assumption 4(i) is weaker than the usual exogeneity assumption because it requires only that $\lambda_P(X)$ (instead of the whole vector X) is measurable with respect to the σ-field of W.

Condition (ii) is satisfied by many models under Condition (i). For example, suppose that $m(w;\alpha,t) \triangleq \mathbf{E}[\rho_\alpha(S,t)|W=w]$, viewed as a function of (α,t), is twice continuously differentiable in (α,t). Let the first order derivatives of $m(w;\alpha,h_P(x))$ in α at $\alpha=\alpha_P$ be $\dot m_\alpha(w)$, where $\dot m_\alpha$ is a $(d_\rho\times d_\alpha)$ matrix, and let the first order derivative of $m(w;\alpha_P,t)$ in t at $t=h_P(x)$ be $\dot m_h(w)$, a $d_\rho\times 1$ vector. Then we define q by

$$q(w) = [\dot m_\alpha(w),\dot m_h(w)]^\top.$$

This q serves as the function q in Condition (ii).

Condition (iii) is employed to ensure a nice behavior of conditional expectations given $\lambda_\beta(X)$. However, it is worth noting that Condition (iii) alone does not guarantee a smooth behavior of conditional expectations given $\lambda_\beta(X)$ with respect to a perturbation of β, because this conditional expectation as a measurable function of $\lambda_\beta(X)$ also depends on β. In fact, that this condition leads to a nice behavior of conditional expectations in conditioning variables is due to Lemma A2 of Song (2008).

Let $\langle\cdot,\cdot\rangle_\lambda$ denote the conditional "inner product": for conformable matrix-valued functions f and g,

$$\langle f,g\rangle_\lambda \triangleq \int f(w)g^\top(w)\mathrm{d}F(w|\lambda), \tag{10}$$

where $F(\cdot|\cdot)$ denotes the conditional distribution of W given $\lambda_P(X)$. As before, when γ is a real valued function, we simply write $\langle\gamma,g\rangle_\lambda \triangleq \langle\gamma I,g\rangle_\lambda$, where I denotes the conformable identity matrix. The following theorem presents an asymptotic representation of the marked empirical process $V_n^1(\gamma_w)$.

Theorem 1. *Suppose that Assumptions 1–4 hold and let $V_n^1(\gamma_w)$ be as in (5). Then*

$$V_n^1(\gamma_w) = \frac{1}{\sqrt n}\sum_{i=1}^n \gamma_w(W_i)\rho_P(S_i) + \frac{1}{\sqrt n}\sum_{i=1}^n\langle\gamma_w,q\rangle_{\lambda_P(X_i)}\begin{bmatrix}\hat\alpha-\alpha_P\\ \hat h(X_i)-h_P(X_i)\end{bmatrix} + o_P(1) \tag{11}$$

uniformly over $w\in[0,1]^{d_W}$, where $\langle\cdot,\cdot\rangle_\lambda$ is as defined in (10).

The representation in Theorem 1 is incomplete because it still depends on estimators $\hat\alpha$ and $\hat h$. However, this representation suffices for our purpose of introducing a proper martingale transform. This representation reveals that we cannot directly apply the isometric projection of Khmaladze (1993) to this situation in general. In other words, there does not exist an isometric projection that can be used to eliminate the second sum in (11). In general, this fact remains unchanged even when we further employ an asymptotic linear representation of the estimators, $\hat\alpha$, $\hat\beta$, and $\hat\tau$. The main idea of this paper is to construct a linear operator $\mathcal{K}$ on $L_2([0,1]^{d_W})$ that eliminates the second process. This can be accomplished by modifying the martingale transform of Khmaladze (1993).

2.2. Conditional martingale transforms

In this section, we search for an appropriate transform $\mathcal{K}$, guided by the decomposition in Theorem 1, to eliminate the second process in the decomposition. We normalize $V_n^1(\gamma_w)$ and $V_n(\gamma_w)$ by $\Omega^{-1/2}$ to deal with conditional heteroskedasticity:

$$V_{n,\Omega}^1(\gamma_w) \triangleq \frac{1}{\sqrt n}\sum_{i=1}^n\gamma_w(W_i)\Omega^{-1/2}(W_i)\hat\rho(S_i)\quad\text{and}$$

$$V_{n,\Omega}(\gamma_w) \triangleq \frac{1}{\sqrt n}\sum_{i=1}^n\gamma_w(W_i)\Omega^{-1/2}(W_i)\rho_P(S_i),$$

where $\Omega(W) \triangleq \mathbf{E}[(\rho_P(S)-\mathbf{E}(\rho_P(S)|W))(\rho_P(S)-\mathbf{E}(\rho_P(S)|W))^\top|W]$.

Suppose that there exists a transform $\mathcal{K}$ that satisfies the following: for any γ_1 and γ_2 in $\{\gamma_w : w\in[0,1]^{d_W}\}$,

$$\langle\mathcal{K}\gamma_1,\mathcal{K}\gamma_2\rangle_{\lambda_P(X)} = \langle\gamma_1,\gamma_2\rangle_{\lambda_P(X)}\quad\text{and}$$
$$\langle\mathcal{K}\gamma_1,\tilde q\rangle_{\lambda_P(X)} = 0,\ \text{a.s.}, \tag{12}$$

where $\tilde q(w) \triangleq q(w)\Omega^{-1/2}(w)$. By applying $\mathcal{K}$ to the decomposition in Theorem 1, we deduce (by adding and subtracting terms) that under H_0,

$$V_{n,\Omega}^1(\mathcal{K}\gamma_w) = \frac{1}{\sqrt n}\sum_{i=1}^n(\mathcal{K}\gamma_w)(W_i)\Omega^{-1/2}(W_i)\times\{\rho_P(S_i)-\mathbf{E}[\rho_P(S_i)|W_i]\} + o_P(1).$$

Due to the (conditional) isometry condition in (12), the covariance function of the right-hand side process at $(\gamma_{w_1},\gamma_{w_2})$ is equal to $\langle\gamma_{w_1},\gamma_{w_2}\rangle$, i.e., the covariance function of the first process on the right-hand side in Theorem 1.

Similarly as in Khmaladze (1993) and Khmaladze and Koul (2004), we define $\mathcal{K}$ as follows: for an arbitrary small number $c>0$,

$$(\mathcal{K}\gamma_{\bar w})(w,\bar\lambda) \triangleq \gamma_{\bar w}(w) - \langle A_w\gamma_{\bar w},C_P^{-1}\tilde q\rangle_{\lambda_P(X)=\bar\lambda}\times\tilde q(w), \tag{13}$$

where $C_P(w,\bar\lambda) \triangleq \langle(1-A_w)\tilde q,(1-A_w)\tilde q\rangle_{\lambda_P(X)=\bar\lambda}$ and $A_{\bar w}(w) \triangleq 1\{0\le w^\top\mathbf{1}\le(\bar w\wedge(1-c))^\top\mathbf{1}\}$.This choice of A_w is a particular example following the general prescriptions given by Khmaladze (1993). The introduction of $c>0$ in $A_{\bar w}(w)$ is made to ensure the existence of $C_P^{-1}(w,\bar\lambda)$. We assume that c is set to be such that for some $\varepsilon>0$, $\inf_{(w,\bar\lambda)\in([0,1]^{d_W}\cap B_c)\times\mathrm{supp}(\lambda_P(X))}\lambda_{\min}(C_P(w,\bar\lambda)) > \varepsilon > 0$, where $\lambda_{\min}(A)$ for a symmetric matrix A denotes its smallest eigenvalue and $B_c \triangleq \{w\in[0,1]^{d_W} : w^\top\mathbf{1} < d_W(1-c)\}$. Due to the introduction of c in (13), the index w is restricted to the set $[0,1]^{d_W}\cap B_c$.

Assumption 5. (i) The conditional distribution of W given $\lambda_P(X)$ is nondegenerate.

(ii) $\|\tilde q\|_p < \infty$, $p > 8$.

Assumption 5(i) is needed for the conditioning in (13) to be meaningful. While all the examples in Section 2.4 below satisfy this condition, there exist interesting examples excluded by this condition. For example, the hypothesis of conditional independence that compares conditional distribution functions and semiparametric models of the kind in Honoré and Lewbel (2002) do not satisfy this condition. The result (i) below is due to Proposition 6.1 of Khmaladze and Koul (2004).

Lemma 1. (i) *Suppose that Assumption 5 holds. Then for any w, w_1 and w_2 in $[0,1]^{d_W}$*

$$\langle \mathcal{K}\gamma_w, \tilde{q}\rangle_{\lambda_P(X)} = 0 \quad \text{and}$$
$$\langle \mathcal{K}\gamma_{w_1}, \mathcal{K}\gamma_{w_2}\rangle_{\lambda_P(X)} = \langle \gamma_{w_1}, \gamma_{w_2}\rangle_{\lambda_P(X)}, \text{ a.s.}$$

(ii) *Suppose that Assumptions 1–5 hold. Then we have*

$$\sup_{w\in[0,1]^{d_W}\cap B_c} \|V^1_{n,\Omega}(\mathcal{K}\gamma_w) - V_{n,\Omega}(\mathcal{K}\gamma_w)\| = o_P(1),$$

where $V_{n,\Omega}(\mathcal{K}\gamma_w) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n (\mathcal{K}\gamma_w)(W_i)\Omega^{-1/2}(W_i)\rho_P(S_i)$.

(iii) *Suppose that Assumptions 1–5 hold. Then under H_0, the process $V_{n,\Omega}(\mathcal{K}\gamma_w)$ weakly converges to a Gaussian process with a covariance kernel, $\langle \gamma_{w_1}, \gamma_{w_2}\rangle$.*

Applying the orthogonality condition in (i) to the decomposition of Theorem 1 leads to (ii) of Lemma 1. The limit process in (iii) is a time-transformed Brownian sheet with a covariance kernel

$$\langle \gamma_{w_1}, \gamma_{w_2}\rangle = F(w_1 \wedge w_2) \triangleq F(w_{11}\wedge w_{21}, \ldots, w_{1d}\wedge w_{2d}).$$

Using the density f of W, let us introduce a normalized indicator function $\tilde{\gamma}_w$ defined by $\tilde{\gamma}_w = \gamma_w \times f^{-1/2}$. Then the limit process of $V_{n,\Omega}(\mathcal{K}\tilde{\gamma}_w)$ becomes a standard Brownian sheet (see Khmaladze (1993)).

2.3. Martingale transforms and the power of tests

It is interesting to investigate the effect of $\mathcal{K}$ upon the test, as $\mathcal{K}$ alters its asymptotic power property. Against certain alternatives, the transform may strengthen the power, whereas, against others, it may weaken the power or even eliminate it completely. However, in most practical cases, the transform maintains consistency of the test, as we argue heuristically below. This argument runs parallel to the argument of Khmaladze (1981) in the context of maximum likelihood estimation. For some results on asymptotic power properties in regression models, see Khmaladze and Koul (2004).

For expositional brevity, we assume $d_\rho = 1$ so that the generalized residual $\rho_\alpha(\cdot,\cdot)$ is a real-valued function. The alternative hypotheses that are eliminated by $\mathcal{K}$ will be such that $\tilde{m}(W) \triangleq \Omega^{-1/2}(W)\mathbf{E}(\rho_P(S)|W)$ satisfies

$$\tilde{m}(w) = v(\lambda_P(x))^\top \tilde{q}(w), \tag{14}$$

for some $(d_\alpha+1)\times 1$ vector-valued function $v(\cdot)$ on $\mathbf{R}$ such that $\mathbf{E}\,\|1_B v(\lambda_P(X))\|^2 > 0$ for some P_X-measurable set B with $P_X B > 0$. The function $\tilde{m}(w)$ lies (conditionally on x) in the subspace spanned by $\tilde{q}(w)$. Hence when $\inf_{\bar{\lambda}\in\mathbf{R}} \mathbf{E}\left[\|\tilde{q}(W)\|^2|\lambda_P(X)=\bar{\lambda}\right] > 0$,

$$\left\|\langle \tilde{m}, \tilde{q}\rangle\right\| > 0. \tag{15}$$

Against alternatives such that (14) holds, $\mathcal{K}$ eliminates the asymptotic power of the test.

These alternatives are excluded, however, when the probability limits α_P, β_P and τ_P of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\tau}$ are well-defined under the alternatives. Suppose the probability limits α_P and h_P of $\hat{\alpha}$ and $\hat{h}$ are solutions to the following m-estimation problem (in population):

$$(\alpha_P, h_P) = \arg\min_{(\alpha,h)\in\mathcal{A}\times\mathcal{H}} \mathbf{E}\left\|\mathbf{E}[\rho_\alpha(S,h(X))|W]\Omega^{-1/2}(W)\right\|^2. \tag{16}$$

This identification strategy is equivalent to the integrated regression function approach in Domínguez and Lobato (2004). Assuming the differentiability of the conditional moment in parameters, and the interchangeability of the integral and the derivative, the first order condition implies the following moment condition:

$$0 = \mathbf{E}[\mathbf{E}[\rho_P(S)|W]\Omega^{-1}(W)q^\top(W)] = \langle \tilde{m}, \tilde{q}\rangle,$$

resulting in the violation of (15). Thus, the identification of (α_P, h_P) in (16) removes the models satisfying (14) from the space of alternatives, and the consistency of the test is left intact after the transform.

2.4. Examples

Example 1: Single Index Restrictions

Single index restrictions can be written as:

$$\mathbf{E}(Y-\tau(X^\top\beta)|X) = 0 \quad \text{for some } (\beta,\tau)\in\mathcal{B}\times\mathcal{T},$$

where $\tau(\bar{\lambda}) = \mathbf{E}\left[Y|X^\top\beta=\bar{\lambda}\right]$. The restriction is mapped to (1) by writing $W = X$, $\lambda_\beta(x) = x^\top\beta$ and $\rho_\alpha(s,\tau) = y-\tau$. Here, S consists of Y and X.

We can obtain a $\sqrt{n}$-consistent estimator $\hat{\beta}$ for $\beta_P\in\mathcal{B}$ as in Powell et al. (1989). Using $\hat{\beta}$ we can construct $\hat{\tau}(\bar{\lambda}) \triangleq \hat{\mathbf{E}}[Y|X^\top\hat{\beta}=\bar{\lambda}]$ for $\tau(\bar{\lambda})$, where $\hat{\mathbf{E}}[Y|X^\top\hat{\beta}=\bar{\lambda}]$ denotes a nonparametric estimator of $\mathbf{E}[Y|X^\top\beta_P=\bar{\lambda}]$.

To find q in Assumption 4(ii), note that

$$m(x;\tau(\lambda_\beta(x))) = \mathbf{E}[Y|X=x] - \tau(\lambda_\beta(x)).$$

From $m(x;\tau(\lambda_\beta(x))) - m(x;\tau_P(\lambda_P(x))) = -\{\tau(\lambda_\beta(x)) - \tau_P(\lambda_P(x))\}$, we take $q(x) = -1$. The conditional martingale transform is obtained by using this q in (13) and by conditioning the transform on $X^\top\beta_P$.

Example 2: Conditional Mean Independence

Given random variables X, Y, and Z, conditional mean independence takes a form: $\mathbf{E}(Y|X) = \mathbf{E}(Y|X,Z)$ or

$$\mathbf{E}(Y-\mathbf{E}(Y|X)|X,Z) = 0.$$

This restriction can be used to test omitted variables in regressions (e.g. Fan and Li (1996)). We define $\tau(x) = \mathbf{E}(Y|X=x)$, $W = (X,Z)$, $\rho_\alpha(s,\tau) = y-\tau$, and $\lambda_\beta(x) = x$. Then the conditional mean independence is a special case of (1). Assumption 4(ii) is satisfied by taking $q=-1$. We condition the martingale transform on X.

We may alternatively consider the restriction : $\mathbf{E}(Y|X^\top\beta) = \mathbf{E}(Y|X,Z)$. This is the joint restriction of conditional mean independence and a single-index restriction. In this case, we can take $\tau(\bar{\lambda}) = \mathbf{E}(Y|X^\top\beta=\bar{\lambda})$ and $\lambda_\beta(x) = X^\top\beta$. For $\mathcal{K}$, we take $q=-1$ and condition the martingale transform on $X^\top\beta$.

Example 3: Partially Parametric Regression with Endogenous Regressors

Consider the partial linear regression model,

$$Y = \mu(Z;\alpha) + \tau(X) + \varepsilon.$$

Suppose that there exists W satisfying $\mathbf{E}(\varepsilon|W) = 0$ and that X is exogenous, contained in W. We focus on the restriction:

$$\mathbf{E}[Y-\mu(Z;\alpha)-\tau(X)|W] = 0.$$

By putting $\rho_\alpha(s,\tau) = y-\mu(z;\alpha)-\tau$, and $\lambda(x) = x$, we can map this model to (1). Robinson (1988) proposes estimating the model based on the identification of $\tau(X) = \mathbf{E}[Y-\mu(Z;\alpha_P)|X]$.

To find q in Assumption 4(ii), note that

$$m_{\alpha,\tau}(w) = \mathbf{E}[Y-\mu(Z;\alpha)|W=w] - \tau(x)$$

and that $m_\alpha(w;\tau(x)) - m_{\alpha_P}(w;\tau_P(x))$ is equal to

$$= -\mathbf{E}[\mu(Z;\alpha_P)-\mu(Z;\alpha)|W=w] - \{\tau(x)-\tau_P(x)\}$$
$$\approx -\mathbf{E}[\dot{\mu}_\alpha(Z;\alpha_P)|W=w](\alpha-\alpha_P) - \{\tau(x)-\tau_P(x)\},$$

by assuming the interchangeability of the conditional expectation and differentiation, where $\dot{\mu}_\alpha(\cdot;\alpha_P)=(\partial/\partial\alpha)\mu(\cdot;\alpha_P)$. Hence we find

$$q(W)=-\begin{bmatrix}\mathbf{E}\left[\dot{\mu}_\alpha(Z;\alpha_P)|W\right]\\ 1\end{bmatrix}.$$

When $\mu(z;\alpha)=z^\top\alpha$, $q(W)=[\mathbf{E}[Z|W],1]^\top$. We condition the transform on X.

Example 4: Partially Parametric Quantile Regressions

Consider the conditional moment restriction:

$$\mathbf{E}\left[\pi-1\{Y-\mu(Z,\alpha_P)-\tau(X)<0\}|W\right]=0,\quad \pi\in[0,1],$$

for a nonparametric function τ. We assume that X is included in W. Let the distribution of $Y-\mu(Z,\alpha_P)$ be continuous and let its conditional distribution function given W be $F_e(\cdot|W)$. Then $\tau(X)$ is identified by

$$\tau(X)=\mathbf{E}\left[F_e^{-1}(\pi|W)|X\right].$$

We let $\rho_\alpha(s,\tau)=\pi-1\{y-\mu(z,\alpha)-\tau<0\}$.

For $m_\alpha(W,\tau)\triangleq -P\{Y-\mu(Z,\alpha)\le\tau|W\}$ and $m_P(W,\tau_P)\triangleq -P\{Y-\mu(Z,\alpha_P)\le\tau_P|W\}$,

$$\begin{aligned}&m_\alpha(W,\tau(X))-m_P(W,\tau_P(X))\\ &\quad=-\left(P\{Y-\mu(Z,\alpha)\le\tau(X)|W\}\right.\\ &\qquad\left.-P\{Y-\mu(Z,\alpha_P)\le\tau(X)|W\}\right)\\ &\qquad-\left(P\{Y-\mu(Z,\alpha_P)\le\tau(X)|W\}\right.\\ &\qquad\left.-P\{Y-\mu(Z,\alpha_P)\le\tau_P(X)|W\}\right).\end{aligned}\tag{17}$$

The first term becomes (under regularity conditions that allow the interchange of differentiation and integration)

$$\begin{aligned}&-\mathbf{E}\left[P\{Y-\mu(Z,\alpha)\le\tau(X)|W,Z\}\right.\\ &\qquad\left.-P\{Y-\mu(Z,\alpha_P)\le\tau(X)|W,Z\}|W\right]\\ &\quad\approx-\mathbf{E}\left[f_{Y|W,Z}(\mu(Z,\alpha_P)+\tau(X)|W,Z)\dot{\mu}_\alpha(Z,\alpha_P)^\top|W\right](\alpha-\alpha_P)\\ &\quad\approx-\mathbf{E}\left[f_{Y|W,Z}(\mu(Z,\alpha_P)+\tau_P(X)|W,Z)\dot{\mu}_\alpha(Z,\alpha_P)^\top|W\right](\alpha-\alpha_P),\end{aligned}$$

where $\dot{\mu}_\alpha(z,\alpha_P)$ denotes the derivative of $\mu(z,\alpha)$ in α at $\alpha=\alpha_P$, and $f_{Y|W,Z}$ denotes the conditional density of Y given W and Z. Similarly, the second term in (17) becomes approximately

$$-\mathbf{E}\left[f_{Y|W,Z}(\mu(Z,\alpha_P)+\tau_P(X)|W,Z)|W\right](\tau(X)-\tau_P(X)).$$

Therefore, we choose

$$q(W)=-\mathbf{E}\left[f_{Y|W,Z}(\mu(Z,\alpha_P)+\tau(X)|W,Z)\begin{bmatrix}\dot{\mu}_\alpha(Z,\alpha_P)\\ 1\end{bmatrix}\middle|W\right].$$

When Z is exogenous, i.e., it is measurable with respect to the σ-field generated by W,

$$q(W)=-f_{Y|W}(\mu(Z,\alpha_P)+\tau(X)|W)\begin{bmatrix}\dot{\mu}_\alpha(Z,\alpha_P)\\ 1\end{bmatrix}$$

where $f_{Y|W}(\cdot|W)$ denotes the conditional density of Y given W. The conditional martingale transform is obtained by conditioning the martingale transform on X.

2.5. Feasible conditional martingale transforms

The transform $\mathcal{K}$ is not feasible because it involves unknown components. In this section, we suggest a feasible transform and establish its asymptotic validity. To ease the exposition, we assume $d_\rho=1$ so that the generalized residual function, $\rho_\alpha(\cdot,\cdot)$, is a real-valued function.

Let us define $\hat{u}(X)\triangleq F_{n,\hat{\beta}}(\lambda_{\hat{\beta}}(X))$, where $F_{n,\hat{\beta}}(\bar{\lambda})=n^{-1}\sum_{i=1}^n 1\{\lambda_{\hat{\beta}}(X_i)\le\bar{\lambda}\}$, and define $\hat{\gamma}_{\bar{w}}(w)=\gamma_w(\bar{w})\hat{f}^{-1/2}(w)$, where $\hat{f}(\cdot)$ is a nonparametric estimator of the density function $f(\cdot)$ of W. Also let $\hat{q}(w)$ be a consistent estimator of $\tilde{q}(w)$. Define a series estimator

$$\hat{C}_{k,m}(w,u)\triangleq p^K(u)^\top\hat{\pi}_{k,m}(w),$$

where $p^K(u)$ denotes a $K\times 1$ basis function vector and $\hat{\pi}_{k,m}(w)=[P_n^\top P_n]^{-1}P_n^\top\tilde{a}_{n,k,m}(w)$ with

$$\tilde{a}_{n,k,m}(w)\triangleq\begin{bmatrix}\tilde{a}_{k,m}(W_1;w)\\ \vdots\\ \tilde{a}_{k,m}(W_n;w)\end{bmatrix},\qquad P_n\triangleq\begin{bmatrix}p^K(\hat{u}(X_1))^\top\\ \vdots\\ p^K(\hat{u}(X_n))^\top\end{bmatrix},$$

and $\tilde{a}_{k,m}(W_i;w)\triangleq(1-A_w(W_i))\hat{Q}_{k,m}(W_i)$, $\hat{Q}_{k,m}(W_i)$ denoting the (k,m)-th element of the matrix $\hat{q}(W_i)\hat{q}(W_i)^\top$. Hence for each k and m, $\tilde{a}_{n,k,m}(w)$ is an $n\times 1$ vector. Let $\hat{C}(w;u)$ be a $(d_\alpha+1)\times(d_\alpha+1)$ matrix whose (k,m)-th element is $\hat{C}_{k,m}(w,u)$. Using this, construct an $n\times(d_\alpha+1)$ matrix:

$$a_{n,\bar{w}}(w)\triangleq\begin{bmatrix}(A_w\hat{\gamma}_{\bar{w}}\hat{q}^\top)(W_1)\hat{C}^{-1}(W_1;\hat{u}(X_1))\\ \vdots\\ (A_w\hat{\gamma}_{\bar{w}}\hat{q}^\top)(W_n)\hat{C}^{-1}(W_n;\hat{u}(X_n))\end{bmatrix}$$

and a $K\times(d_\alpha+1)$ matrix, $\hat{\pi}_{\bar{w}}(w)\triangleq[P_n^\top P_n]^{-1}P_n^\top a_{n,\bar{w}}(w)$. We define a feasible conditional martingale transform as

$$(\hat{\mathcal{K}}\hat{\gamma}_{\bar{w}})(w,u)\triangleq\hat{\gamma}_{\bar{w}}(w)-p^K(u)^\top\hat{\pi}_{\bar{w}}(w)\hat{q}(w).\tag{18}$$

To simplify the exposition, we resort to high-level assumptions for certain components of the estimator. Let

$$\eta_P(w,u)\triangleq\left(f^{-1/2}\tilde{q}^\top C_P^{-1}\right)(w,u)\quad\text{and}$$

$$\hat{\eta}(w,u)\triangleq\left(\hat{f}^{-1/2}\hat{q}^\top\hat{C}^{-1}\right)(w,u).\tag{19}$$

Then we assume the following.

Assumption 6 (*Regularity Conditions for Estimated Components*). Estimators $(\hat{\mathcal{K}}\hat{\gamma}_w)(\cdot,\hat{u}(\cdot))$ and $\hat{\Omega}^{-1/2}$ of $(\mathcal{K}\tilde{\gamma}_w)(\cdot,u_P(\cdot))$ and $\Omega_P^{-1/2}$ satisfy RC$(b,4,\mathcal{J}_\mathcal{K})$ and RC$(b,4,\mathcal{W})$, for $b\in[0,2)$ and for some classes of functions $\mathcal{J}_\mathcal{K}$ and $\mathcal{W}$, respectively.

The conditions for the basis function vector p^K and the number of terms K are subsumed into the regularity condition for $\hat{\mathcal{K}}\hat{\gamma}_w$. (See Newey (1997) and Song (2008) for uniform convergence of series estimators.) The following theorem shows that the feasible martingale transform is asymptotically valid.

Theorem 2. *Suppose that Assumptions 1–6 hold. Then under H_0 or under local alternatives P_n such that $\sup_{w\in[0,1]^{d_W}\cap B_c}|m_{P_n}(w)|=O(n^{-1/4})$,*

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\{(\hat{\mathcal{K}}\hat{\gamma}_w)(W_i,\hat{u}(X_i))\hat{\Omega}^{-1/2}(W_i)-(\mathcal{K}\tilde{\gamma}_w)(W_i,u_P(X_i))\Omega_P^{-1/2}(W_i)\}\hat{\rho}(S_i)\right|=o_P(1),$$

uniformly over $w\in[0,1]^{d_W}\cap B_c$.

Under local alternatives that converge to the null hypothesis at least at the rate of $n^{-1/4}$, the difference between the feasible and infeasible martingale-transformed processes vanishes in probability. Obviously, such local alternatives include the usual Pitman local alternatives that converge to the null at the rate of $\sqrt{n}$. Theorem 2, combined with previous results, leads to feasible test statistics:

$$T_{KS}^{\hat{\mathcal{K}}}=\sup_{w\in[0,1]^{d_W}\cap B_c}\left|\left|V_{n,\Omega}^1(\hat{\mathcal{K}}\hat{\gamma}_w)\right|\right|\quad\text{and}$$

$$T_{CM}^{\hat{\mathcal{K}}}=\int\left|\left|V_{n,\Omega}^1(\hat{\mathcal{K}}\hat{\gamma}_w)\right|\right|^2\mathrm{d}w.\tag{20}$$

The limit process of $V_{n,\Omega}^1(\hat{\mathcal{K}}\hat{\gamma}_w)$ under the null hypothesis is the same as that of $V_{n,\Omega}^1(\mathcal{K}\tilde{\gamma}_w)$.

Table 1
Empirical size (The nominal size is set at 0.05.).

$n = 500$		$K = 5$	$K = 6$	$K = 7$	$K = 8$	$K = 9$	$K = 10$
$b = 0.2$	Mart. trns	0.050	0.045	0.060	0.071	0.081	0.082
	Bootstrap	0.089	0.048	0.044	0.050	0.050	0.049
$b = 0.6$	Mart. trns	0.074	0.068	0.071	0.093	0.104	0.149
	Bootstrap	0.056	0.044	0.042	0.040	0.035	0.038

3. Simulation studies

This section presents and discusses results from Monte Carlo simulation studies. We consider a partially linear model and investigate two testing procedures: the martingale transform and the wild bootstrap (Härdle and Mammen, 1993). More specifically, consider the following:

$$H_0 : \mathbf{E}[Y - Z\beta - \tau(X)|X, Z] = 0 \quad \text{for some } (\tau, \beta) \in \mathcal{T} \times \mathcal{B}$$

for a certain class $\mathcal{T}$ and an interval $\mathcal{B}$. The data generating process is as follows:

$$\tilde{Z}_i \sim U[-1, 1], \tilde{X}_i \sim U[-1, 1],$$
$$Z_i = \tilde{Z}_i + (1 - b)\varepsilon_{1i} + \varepsilon_i, \quad \text{and} \quad X_i = \tilde{X}_i + (1 - b)\varepsilon_{2i} + \varepsilon_i$$
$$\varepsilon_i, \varepsilon_{1i}, \varepsilon_{2i} \sim N(0, 1) \quad \text{with } \varepsilon_i, \varepsilon_{1i}, \text{ and } \varepsilon_{2i} \text{ being independent}$$
$$u_i \sim 0.5 \times N(0, 1),$$
$$Y_i = r(Z_i, a) + \tau(X_i) + u_i, \quad \text{where } \tau(X_i) = 2\Phi(X_i) - 1$$

with Φ being a standard normal cdf. When $a = 0, r(Z_i, a) = Z_i$, and becomes nonlinear as a moves away from zero. The data generating processes corresponding to a's different from zero represent alternative hypotheses. See Fig. 1.

The number of Monte Carlo iterations and the number of bootstrap Monte Carlo iterations are set at 1000. To estimate the model, we use Robinson (1988)'s two step procedure, but for the nonparametric estimation, we employ series estimation. For this, we use normalized Hermite polynomials for estimating the model and Legendre polynomials for estimating the conditional martingale transform. We took the sample quantile transform of the instrumental variables using the empirical distribution functions and rescaled the variables to have support $[-1, 1]$ for the series-estimation using the Legendre basis functions. The K number of the basis functions in the series estimation were used for univariate conditional mean functions, and (integer of $K^{0.7})^2$ number of basis functions for a bivariate function. The test statistic we used is of Kolmogorov–Smirnov type, i.e., $T_{KS}^{\tilde{\mathcal{K}}}$ defined in (20).

First, we consider the small sample size and powers of the martingale transform-based test and the bootstrap-based test. The nominal size is set at 0.05 and the sample size equal to 500. Table 1 contains the results of empirical sizes of tests based on the martingale transform and the bootstrap approaches.

The empirical sizes for the martingale transform approach are more sensitive to the choice of K than those from the bootstrap approach. This is expected because the martingale transform approach involves additional steps of nonparametric estimation.

In the investigation of empirical power properties, we considered four types of alternatives in the specification of $r(z, a)$. The four specifications for $r(z, a)$ under the alternatives are depicted in Fig. 1. The solid lines represent $r(Z_i, 0)$, and dotted lines, $r(Z_i, a)$ with $a \neq 0$.[3]

[3] The specification of the functions is as follows: $r(Z_i, a) = 2h(Z_i, a)/\max\{|h(Z_i, a)|\}_{i=1}^n - 1$, where Type-QRTIC: $h(z, a) = a(2z^4 - 1) + (1 - a)z$; Type-PDF: $h(z, a) = a(4\phi(z) - 2) + (1 - a)z$, ϕ being the density of $N(0, 0.5)$; Type-COS: $h(z, a) = a(2\cos(\pi z/5) - 1) + (1 - a)z$; Type-CUBIC: $h(z, a) = a((z - 1/2)^3 - 15(x - 1/2)) + (1 - a)z$.

Table 2
Empirical power: $n = 300$.

		K	QRTIC	PDF	COS	CUBIC
Small a	Mart. trns.	5	0.132	0.103	0.122	0.131
		8	0.201	0.166	0.201	0.199
	Bootstrap	5	0.119	0.083	0.076	0.122
		8	0.071	0.090	0.080	0.166
Medim a	Mart. trns.	5	0.225	0.140	0.180	0.468
		8	0.277	0.193	0.261	0.562
	Bootstrap	5	0.169	0.112	0.123	0.720
		8	0.124	0.122	0.132	0.770
Large a	Mart. trns.	5	0.466	0.446	0.531	0.904
		8	0.458	0.507	0.594	0.929
	Bootstrap	5	0.287	0.378	0.311	0.961
		8	0.217	0.426	0.377	0.973

Table 3
Empirical power: $n = 500$.

		K	QRTIC	PDF	COS	CUBIC
Small a	Mart. trns.	5	0.128	0.065	0.100	0.105
		8	0.144	0.104	0.129	0.132
	Bootstrap	5	0.168	0.101	0.101	0.162
		8	0.091	0.096	0.102	0.239
Medim a	Mart. trns.	5	0.263	0.139	0.215	0.491
		8	0.270	0.152	0.271	0.556
	Bootstrap	5	0.238	0.169	0.165	0.783
		8	0.138	0.200	0.183	0.857
Large a	Mart. trns.	5	0.524	0.562	0.705	0.923
		8	0.548	0.575	0.751	0.944
	Bootstrap	5	0.448	0.551	0.517	0.977
		8	0.314	0.612	0.601	0.985

Table 4
Empirical Power: $n = 700$.

		K	QRTIC	PDF	COS	CUBIC
Small a	Mart. trns.	5	0.100	0.065	0.111	0.105
		8	0.125	0.076	0.137	0.133
	Bootstrap	5	0.186	0.130	0.128	0.217
		8	0.086	0.119	0.106	0.308
Medim a	Mart. trns.	5	0.283	0.174	0.277	0.619
		8	0.299	0.182	0.307	0.658
	Bootstrap	5	0.343	0.220	0.182	0.871
		8	0.172	0.221	0.195	0.926
Large a	Mart. trns.	5	0.575	0.719	0.817	0.950
		8	0.604	0.719	0.825	0.960
	Bootstrap	5	0.540	0.710	0.630	0.979
		8	0.373	0.762	0.698	0.987

The empirical powers of the tests are reported in Tables 2–4. To gauge the sensitivity of the powers, we chose K from $\{5, 8\}$. Here b is set to be 0.2. Note that $K = 5$ for martingale transform and $K = 8$ for bootstrap yields the size of the tests approximately equal to 5%.

First, the relative performances of both approaches dramatically change depending on alternatives. For example, under the alternatives of COS, the empirical power of the martingale transform-based test is stronger than that of the bootstrap-based test. This order of performances is reversed when we take the alternatives of CUBIC. This finding reinforces the fact that the boot-

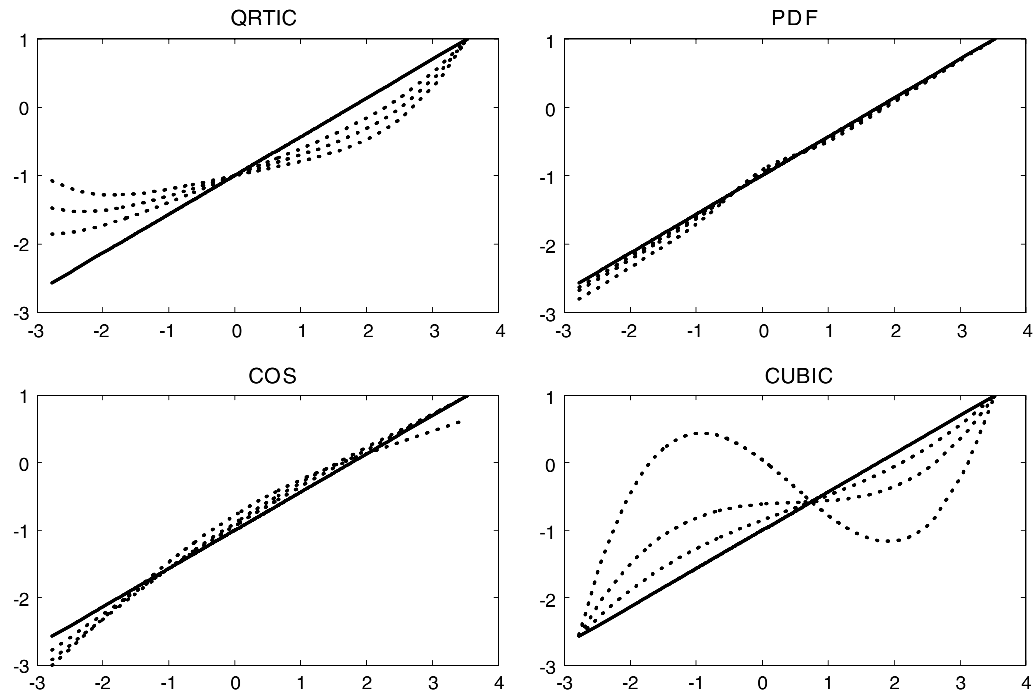

Fig. 1. Four types of alternatives: $r(z, a)$.

strap test and the martingale transform-based test are far from perfect substitutes to each other. The two tests have very different power properties, and hence should be viewed as complementary. Second, the power depends on the choice of K, with varied degrees of sensitivity. In the case of QRTIC, the power of the bootstrap-based test is sensitive to the choice of K relative to that of the martingale transform-based test. Third, the increase in power as the sample size becomes larger is more prominent for the bootstrap method than the martingale method when the alternative hypothesis is very close to the null hypothesis (with small a).

4. Conclusion

This paper develops new tests for semiparametric conditional moment restrictions by generalizing the martingale transform approach to accommodate semiparametric restrictions. This approach provides a unifying testing framework in which asymptotic pivotal tests for many semiparametric models can be obtained.

The martingale transform approach alters the asymptotic power function of test statistics. It would be interesting to investigate the specific manner in which the martingale transform affects power properties of the original tests. For example, Stute et al. (1998) and Escanciano (2006) investigated the asymptotic power function of the tests after the martingale transform. Future research might consider applying a similar idea to the tests of semiparametric conditional moment restrictions, based on the results of this paper.

Acknowledgments

My thanks go to Peter Phillips, for his guidance, advice, and continual encouragement. I am grateful to Don Andrews and Yoon-Jae Whang for their continual advice and comments throughout this research. I thank Jushan Bai, Xiaohong Chen, Juan Carlos Escanciano, Yuichi Kitamura, Guido Kuersteiner, Vadim Marmer, Taisuke Otsu, and seminar participants at various places. I thank the Associate Editor and a referee for valuable comments which led to an improvement of the paper. Finally, I am grateful to the Cowles Foundation for research support under the Cowles Foundation Prize and the Carl Arvid Anderson Prize. All errors are mine.

Appendix. Mathematical proofs

Throughout the proofs, C indicates a constant that takes different values in different places. In the proofs in the following we say a class of functions is $\mathrm{BEP}_p(r)$ (or BEP_∞) briefly if the class satisfies a bracketing entropy bound of order r with respect to $\|\cdot\|_p$, (or $\|\cdot\|_\infty$), as in (9), and when the class of $\mathrm{BEP}_p(r)$ for any $r > 0$ and $p \geq 1$, we say simply it is BEP.

A.1. Proof of the main results

Proof of Theorem 1. Consider $N_P(\delta_n) \subset \mathcal{A} \times \mathcal{H}$ of (α_P, h_P) defined prior to Assumption 4 with $\delta_n \to 0$ such that $\delta_n = o(n^{-1/4})$ and $\lim_{n\to\infty} P\{(\hat{\alpha}, \hat{h}) \in N_P(\delta_n)\} = 1$, and define $N_n \triangleq [0, 1]^{d_W} \times N_P(\delta_n)$. Consider the processes

$$V_n^1(\gamma_w; \alpha, h) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_w(W_i)\rho_\alpha(S_i, h(X_i)) \quad \text{and}$$

$$\tilde{V}_n(\gamma_w; \hat{\alpha}, \hat{h}) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_w(W_i)\rho_P(S_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \langle \gamma_w, q\rangle_{\lambda_P(X_i)} \begin{bmatrix} \hat{\alpha} - \alpha_P \\ \hat{h}(X_i) - h_P(X_i) \end{bmatrix}.$$

Since $P\left\{(\hat{\alpha}, \hat{h}) \notin N_P(\delta_n)\right\} = o(1)$, it suffices to show that

$$\sup_{(w,\alpha,h)\in N_n} \left|V_n^1(\gamma_w; \alpha, h) - \tilde{V}_n(\gamma_w; \alpha, h)\right| = o_P(1).$$

Lemma A.1 below shows that

$$\zeta_n(w, \alpha, h) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_w(W_i)\left\{\rho_\alpha(S_i, h(X_i)) - \mathbf{E}\left[\rho_\alpha(S_i, h(X_i))|W_i\right]\right\}$$

is stochastically equicontinuous in $(w, \alpha, h) \in N_n$, yielding the fact that

$$\sup_{(w,\alpha,h)\in N_n} \|\zeta_n(w, \alpha, h) - \zeta_n(w, \alpha_P, h_P)\| = o_P(1).$$

From this, we deduce

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)(\rho_\alpha(S_i,h(X_i))-\rho_P(S_i))$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)\mathbf{E}(\rho_\alpha(S_i,h(X_i))-\rho_P(S_i)|W_i)+o_P(1). \quad (21)$$

Now, we can write the first term in the second line as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\gamma_w(W_i)\Delta_{\alpha,h}(W_i,X_i)-\mathbf{E}[\gamma_w(W_i)\Delta_{\alpha,h}(W_i,X_i)|]\right\}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)$$
$$+\sqrt{n}\mathbf{E}[\gamma_w(W_i)\Delta_{\alpha,h}(W_i,X_i)|] \quad (22)$$

where $\Delta_{\alpha,h}(W_i,X_i)\triangleq m_{\alpha,h}(W_i)-m_P(W_i)-\Delta^0_{\alpha,h}(W_i,X_i)$ and

$$\Delta^0_{\alpha,h}(W_i,X_i)\triangleq q(W_i)^\top\begin{bmatrix}\alpha-\alpha_P\\ h(X_i)-h_P(X_i)\end{bmatrix}.$$

The stochastic equicontinuity of the first process can be established similarly as in Lemma A.1 and hence it is $O_P(\delta_n)=o_P(1)$. By Assumption 4(ii), the last term in (22) is $O_P(n^{1/2}\delta_n^2)=o_P(1)$. Therefore,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)\mathbf{E}(\rho_\alpha(S_i,h(X_i))-\rho_P(S_i)|W_i)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)+o_P(1). \quad (23)$$

By Assumption 4(iii) and by the choice of δ_n,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\mathbf{E}\left[\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)|\lambda_P(W_i),\lambda_\beta(X_i)\right]\right.$$
$$\left.-\mathbf{E}\left[\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)|\lambda_P(W_i)\right]\right\}=o_P(1).$$

Furthermore, by Lemma A.1 below,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)-\mathbf{E}[\gamma_w(W_i)\Delta^0_{\alpha,h}(W_i,X_i)|\lambda_P(X_i)]\}$$

is stochastically equicontinuous in $(w,\alpha,h)\in N_n$ and hence is $o_P(1)$. This leads to the conclusion that the last sum in (23) is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[\gamma_w q(W_i)^\top|\lambda_\beta(X_i),\lambda_P(X_i)\right]\begin{bmatrix}\alpha-\alpha_P\\ \tau(\lambda_\beta(X_i))-\tau_P(\lambda_P(X_i))\end{bmatrix}.$$

The application of Assumption 4(iii) to Lemma A2(ii) of Song (2008) renders the sum above asymptotically equivalent to $\tilde{V}_n(\gamma_w;\alpha,h)$ uniformly over $(w,\alpha,h)\in N_n$. ■

Lemma A.1. *Suppose that the conditions of Theorem 1 hold. The processes*

$$\nu_{1n}(w,\alpha,h)\triangleq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_w(W_i)(\rho_\alpha(S_i,h(X_i))-\mathbf{E}[\rho_\alpha(S_i,h(X_i))|W_i]),$$
$$\nu_{2n}(w,\alpha,h)\triangleq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\gamma_w(W_i)q^\top(W_i)-\mathbf{E}(\gamma_w(W_i)q^\top(W_i)|\lambda_P(W_i))\}\begin{bmatrix}\alpha\\ h(X_i)\end{bmatrix},$$

are stochastically equicontinuous in $(w,\alpha,h)\in[0,1]^{d_W}\times\mathcal{A}\times\mathcal{H}$.

Proof of Lemma A.1. Define $\mathcal{J}_{j,\gamma}\triangleq\{\gamma_{w_j}(\cdot):w_j\in[0,1]\}, j=1,\ldots,d_W$ and

$$\mathcal{J}_{j,\rho}\triangleq\{\rho_{j\alpha}(\cdot,h(\cdot)):(\alpha,h)\in N_n(\delta_n)\},\ j=1,\ldots,d_\rho, \quad (24)$$

where $\rho_{j\alpha}$ is the j-th entry of ρ_α. We first show that $\mathcal{J}_{j,\gamma}$ and $\mathcal{J}_{j,\rho}$ are BEP and $\text{BEP}_p(b_1/s)$ for b_1, s and p in Assumption 2. We know that (Example 2.5.4. in van der Vaart and Wellner (1996), p.129) for any $p'>0$, $N_{[]}(\varepsilon^{p'},\mathcal{J}_{j,\gamma},\|\cdot\|_1)\le 2/\varepsilon^{p'}$. With F_k denoting the distribution function of W_k, the k-th element of W,

$$\left(\int\left|1\{w\le w_k\}-1\{w\le w_k'\}\right|^{p'}dF_k(w)\right)^{1/p'}$$
$$=\left(\int\left|1\{w\le w_k\}-1\{w\le w_k'\}\right|dF_k(w)\right)^{1/p'}.$$

Hence $N_{[]}(\varepsilon,\mathcal{J}_{j,\gamma},\|\cdot\|_{p'})\le N_{[]}(\varepsilon^{p'},\mathcal{J}_{j,\gamma},\|\cdot\|_1)\le 2/\varepsilon^{p'}$, i.e, $\mathcal{J}_{j,\gamma}$ is BEP. From the straightforward arguments, it follows that $\mathcal{J}_\gamma\triangleq\otimes_{j=1}^{d_W}\mathcal{J}_{j,\gamma}$ is BEP.

Let us turn to $\mathcal{J}_{j,\rho}$. Following the proof of Theorem 3 in Chen et al. (2003) along with Assumption 2(i) and Assumption 3(ii), we deduce that

$$N_{[]}(\varepsilon,\mathcal{J}_{j,\rho},\|\cdot\|_p)\le N((\varepsilon/2C)^{1/s},\mathcal{A}\times\mathcal{H},\|\cdot\|+\|\cdot\|_p).$$

Therefore $\mathcal{J}_{j,\rho}$ is $\text{BEP}_p(b_1/s)$.

Let $\mathcal{J}''_{j,\rho}\triangleq\{\mathbf{E}[\rho(S)|W=\cdot]:\rho\in\mathcal{J}_{j,\rho}\}$. That $\mathcal{J}'_{j,\rho}\triangleq\{\rho_1-\rho_2:\rho_1\in\mathcal{J}_{j,\rho},\rho_2\in\mathcal{J}''_{j,\rho}\}$ is $\text{BEP}_p(b_1/s)$ follows from applying Theorem 6 in Andrews (1994) (see (A.23), p.2291) to $\mathcal{J}_{j,\rho}$ and $\mathcal{J}''_{j,\rho}$. That $\mathcal{J}''_{j,\rho}$ is $\text{BEP}_p(b_1/s)$ immediately follows from the fact that $\|\mathbf{E}[\rho_1(S)|W]-\mathbf{E}[\rho_2(S)|W]\|_p\le\|\rho_1-\rho_2\|_p$. Let $\mathcal{J}_\rho\triangleq\otimes_{j=1}^{d_\rho}\mathcal{J}_{j,\rho}$. Now, for $\mathcal{J}_{\gamma\rho}\triangleq\{\gamma\rho:\rho\in\mathcal{J}_\rho,\gamma\in\mathcal{J}_\gamma\}$, we observe that for a small $\varepsilon>0$,

$$N_{[]}(C\varepsilon,\mathcal{J}_{\gamma\rho},\|\cdot\|_2)$$
$$\le N_{[]}(\varepsilon,\mathcal{J}_\rho,\|\cdot\|_{2(1+\varepsilon)})N_{[]}(\varepsilon,\mathcal{J}_\gamma,\|\cdot\|_{2(1+\varepsilon)/\varepsilon}),$$

which implies that $\mathcal{J}_{\gamma\rho}$ is $\text{BEP}_2(b_1/s)$. Therefore, by combining this bracketing entropy condition with the moment condition for the envelope $\tilde{\rho}$ in Assumption 5, the stochastic equicontinuity of the first process in the lemma follows by Theorem 4 of Andrews (1994), p. 2277.

For the second process in this lemma, recall that $\mathcal{H}$ is $\text{BEP}_p(b_1)$ by Assumption 2(i). Let q_{ij} be the (i,j)-th element of q. From the result that $\mathcal{J}_\gamma$ is BEP and the assumption in Assumption 4(ii) about the moment conditions for q, $\mathcal{J}_{\gamma q}\triangleq\{\gamma q_{ij}:\gamma\in\mathcal{J}_\gamma\}$ is $\text{BEP}_4(b)$ for any $b>0$. Obviously the bracketing entropy for $\{\mathbf{E}[(\gamma_w q_{ij})(W)|\lambda_P(X)=\cdot]:w\in[0,1]^{d_W}\}$ is always bounded by the bracketing entropy of $\mathcal{J}_{\gamma q}$. Hence $\mathcal{J}_{\gamma q}$ is $\text{BEP}_4(b)$ for any $b>0$ and the class

$$\left\{\{(\gamma q_j^\top)(\cdot)-\mathbf{E}((\gamma q_j^\top)(W)|\lambda_P(X)=\cdot)\}\times\begin{bmatrix}\alpha\\ h(\cdot)\end{bmatrix}:\gamma\in\mathcal{J}_\gamma,(\alpha,h)\in\mathcal{A}\times\mathcal{H}\right\} \quad (25)$$

is $\text{BEP}_2(b_1)$ by the Hölder inequality. Hence the second process in the lemma is stochastically equicontinuous in (w,α,h), again by Theorem 4 of Andrews (1994). Note that the required moment conditions are satisfied by Assumptions 2 and 4(ii). ■

Proof of Lemma 1. (i) The proof follows by modifying slightly the proof of Proposition 6.1 in Khmaladze and Koul (2004). The modification simply involves replacing the inner product by the conditional inner product.

(ii) Suppose we have proven Theorem 1 with $\mathcal{K}\gamma_w$ in place of γ_w so that

$$V^1_{n,\Omega}(\mathcal{K}\gamma_w) - V_{n,\Omega}(\mathcal{K}\gamma_w) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \langle \mathcal{K}\gamma_w, \tilde{q}\rangle_{\lambda_P(X_i)} \times \begin{bmatrix} \hat{\alpha} - \alpha_P \\ \hat{\tau}(\lambda_{\hat{\beta}}(X_i)) - \tau_P(\lambda_P(X_i)) \end{bmatrix} + o_P(1).$$

Then the wanted result follows by the fact that $\langle \mathcal{K}\gamma_w, \tilde{q}\rangle_{\lambda_P(X_i)} = 0$ a.s. due to Lemma 1(i).

The proof of Theorem 1 depends on stochastic equicontinuity of certain processes in (w, α, h). Stochastic equicontinuity of these processes are established in Lemma A.1. The same procedure can be taken when we replace γ_w by $\mathcal{K}\gamma_w$ as stochastic equicontinuity of these processes with γ_w replaced by $\mathcal{K}\gamma_w$ are established in Lemma A.2. Hence the wanted result follows.

(iii) Weak convergence of the process $V_{n,\Omega}(\mathcal{K}\gamma_w)$ follows once we prove that (1) the process is stochastically equicontinuous in $w \in [0, 1]^{d_W} \cap B_c$ and (2) the index space of w is totally bounded, and (3) the finite dimensional distributions of the process converge in distribution to a jointly normal random vector (e.g., Proposition in Andrews (1994), p.2251). Condition (2) is trivially satisfied because $[0, 1]^{d_W}$ is compact. Condition (1) is demonstrated in the proof of Lemma A.2. Condition (3) follows by the usual CLT.

The covariance function of the process $V_{n,\Omega}(\mathcal{K}\gamma_w)$ is written as

$$\mathbf{E}\left[(\mathcal{K}\gamma_{w_1})\Omega_P^{-1/2}[\rho_P - \mathbf{E}(\rho_P|W)][\rho_P - \mathbf{E}(\rho_P|W)]^\top \Omega_P^{-1/2}(\mathcal{K}\gamma_{w_2})\right] = \mathbf{E}\left[\langle \mathcal{K}\gamma_{w_1}, \mathcal{K}\gamma_{w_2}\rangle_{\lambda_P(X)}\right].$$

Note that the above equality uses the measurability condition in Assumption 4(i). By (i) of this lemma, $\langle \mathcal{K}\gamma_{w_1}, \mathcal{K}\gamma_{w_2}\rangle_{\lambda_P(X_i)} = \langle \gamma_{w_1}, \gamma_{w_2}\rangle_{\lambda_P(X_i)}$. Hence the covariance function is equal to $\langle \gamma_{w_1}, \gamma_{w_2}\rangle$. ■

Lemma A.2. *Suppose that the conditions of Theorem 1 and Lemma 1 hold.*

(a) *The class $\mathcal{J}_G \triangleq \{\mathcal{K}\gamma_w : w \in [0, 1]^{d_W}\}$ is BEP$_p(b)$ for any $b > 0$, for some $p > 4$.*
(b) *Lemma A.1 above holds when we replace γ_w by $\mathcal{K}\gamma_w$.*

Proof of Lemma A.2. (a) We saw in the proof of Lemma A.1 that the class $\mathcal{J}_\gamma$ is BEP. Let

$$(\mathcal{M}_c\gamma_{\bar{w}})(w, \bar{\lambda}) \triangleq \langle A_w\gamma_{\bar{w}}, C_P^{-1}\tilde{q}\rangle_{\lambda_P(X)=\bar{\lambda}} \times \tilde{q}(w). \quad (26)$$

Let v_m be the m-th column vector of $(\tilde{q}^\top C_P^{-1})^\top$. For the (m, k)-th element $[\mathcal{M}_c\gamma_w]_{mk}$ of $\mathcal{M}_c\gamma_w$,

$$\|[\mathcal{M}_c\gamma_w]_{mk}\|_p^p = \mathbf{E}\left|\mathbf{E}\left[(A_{\bar{w}}\gamma_w v_m^\top)(W)|\lambda_P(X)\right]_{\bar{w}=W}\tilde{q}_k(W)\right|^p.$$

Since $\|v_m\| \le \sqrt{tr\left[\tilde{q}\tilde{q}^\top C_P^{-2}\right]} \le \lambda_{\min}(C_P)^{-1}\left\|\tilde{q}\right\|$, we deduce that

$$\|\mathcal{M}_c(\gamma_{w_1} - \gamma_{w_2})\|_p \le C\|\gamma_{w_1} - \gamma_{w_2}\|_{2p}.$$

Therefore $N_{[]}(\varepsilon, \mathcal{J}_G, \|\cdot\|_p) \le N_{[]}(C\varepsilon, \mathcal{J}_\gamma, \|\cdot\|_{2p})$, yielding the wanted conclusion that $\mathcal{J}_G$ is BEP$_p(b)$.

(b) We can follow similarly the steps in the proof of Lemma A.1 using (a). ■

Proof of Theorem 2. First, write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[(\hat{\mathcal{K}}\hat{\gamma}_w)(W_i, \hat{u}(X_i))\hat{\Omega}^{-1/2}(W_i) - (\mathcal{K}\tilde{\gamma}_w)(W_i, u_P(X_i))\Omega_P^{-1/2}(W_i)\right]\hat{\rho}(S_i) = A_{1n} + A_{2n},$$

where

$$A_{1n} \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)\{\hat{\rho}(S_i) - \rho_P(S_i)\},$$

$$A_{2n} \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)\rho_P(S_i).$$

We show A_{1n} and A_{2n} are $o_P(1)$ uniformly over $w \in [0, 1]^{d_w} \cap B_c$.

We first deal with A_{1n}. Following the proof of Theorem 1,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)\{\hat{\rho}(S_i) - \rho_P(S_i)\} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)\Delta_{\hat{\alpha},\hat{h}}(W_i, X_i) + o_P(1)\end{aligned}$$

for $\Delta_{\alpha,h}(W_i, X_i)$ defined in the proof of Theorem 1. Since

$\sup_{w,(\bar{w},x)}\left\|\left[\hat{\mathcal{K}}\hat{\gamma}_w(\bar{w}, \hat{u}(x)) - \mathcal{K}\tilde{\gamma}_w(\bar{w}, u_P(x))\right]\right\| = o_P(n^{-1/4})$ and $\sup_{w,x}\left\|\Delta_{\hat{\alpha},\hat{h}}(w, x)\right\| = o_P(n^{-1/4})$,

we conclude that $A_{1n} = o_P(1)$.

We deal with A_{2n} which we write as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)\{\rho_P(S_i) - m_P(W_i)\} \\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \times \hat{\Omega}^{-1/2}(W_i)m_P(W_i). \quad (27)\end{aligned}$$

Let $\mathcal{J}_{\mathcal{K},n}$ be a subset of $\mathcal{J}_{\mathcal{K}}$ such that for each $\phi \in \mathcal{J}_{\mathcal{K},n}$, $\|\phi(\cdot,\cdot) - \mathcal{K}\tilde{\gamma}_w(\cdot, u_P(\cdot))\|_\infty < \varepsilon n^{-1/4}$ for a small $\varepsilon > 0$. The first sum is bounded by

$$\sup_{(\phi,\omega)\in\mathcal{J}_{\mathcal{K},n}\times\mathcal{W}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\phi(W_i, X_i) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))\right] \omega(W_i)\{\rho_P(S_i) - m_P(W_i)\}\right|.$$

The above supremum is $o_P(1)$ using the fact that the process is mean zero and stochastically equicontinuous in $(\phi, \omega) \in \mathcal{J}_{\mathcal{K}} \times \mathcal{W}$. As for the second sum in (27),

$$\begin{aligned}&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n[\hat{\mathcal{K}}\hat{\gamma}_w(W_i, \hat{u}(X_i)) - \mathcal{K}\tilde{\gamma}_w(W_i, u_P(X_i))]\hat{\Omega}^{-1/2}(W_i)m_P(W_i)\right| \\ &\le C\sqrt{n}\sup_{w,(\bar{w},x)}\left\|\left[\hat{\mathcal{K}}\hat{\gamma}_w(\bar{w}, \hat{u}(x)) - \mathcal{K}\tilde{\gamma}_w(\bar{w}, u_P(x))\right]\right\| \times \sup_w|m_P(w)| \times O_P(1) \\ &= \sqrt{n} \times o_P(n^{-1/4}) \times \sup_w|m_P(w)|,\end{aligned}$$

where the supremum over w is over the support of W. Since $\sup_w|m_P(w)| = 0$ under the null hypothesis and $\sup_w|m_P(w)| = O(n^{-1/4})$ under the local alternatives, we conclude that $A_{2n} = o_P(1)$. ■

References

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

Andrews, D.W.K., 1994. Empirical process method in econometrics. In: Engle, R.F., McFadden, D.L. (Eds.), The Handbook of Econometrics, Vol. IV. North-Holland, Amsterdam.
Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.
Angrist, J.D., Kuersteiner, G.M., 2004. Semiparametric causality tests using the policy propensity score. NBER Working Paper 10975.
Bai, J., 2003. Testing parametric conditional distributions of dynamic models. Review of Economics and Statistics 85, 531–549.
Bai, J., Ng, S., 2001. A consistent test for conditional symmetry in time series models. Journal of Econometrics 103, 225–258.
Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.
Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1151.
Cabaña, A., Cabaña, E.M., 1997. Transformed empirical processes and modified Kolmogorov–Smirnov tests for multivariate distributions. Annals of Statistics 25, 2388–2409.
Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics 91, 373–401.
Chen, X., Linton, O., van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591–1608.
Chen, X., White, H., 1998. Central limit and functional central limit theorems for Hilbert-valued dependent random arrays. Econometric Theory 14, 260–284.
Delgado, M.A., González Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics 29, 1469–1507.
Delgado, M.A., Hidalgo, J., Velasco, C., 2005. Distribution free goodness-of-fit tests for linear processes. Annals of Statistics 33, 2568–2609.
Delgado, M.A., Stute, W., 2007. Distribution-free specification tests of conditional models. Journal of Econometrics 143, 37–55.
Domínguez, M.A., Lobato, I.N., 2004. Consistent estimation of models defined by conditional moment restrictions. Econometrica 72, 1601–1615.
Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. Annals of Statistics 1, 279–290.
Durbin, J., Knott, M., Taylor, C.C., 1975. Components of Cramér–von Mises statistics. II. Journal of the Royal Statistical Society, Series B 37, 216–237.
Escanciano, J.C., 2006. Consistent diagnostic test for regression models using projections. Econometric Theory 22, 1030–1051.
Fan, Y., Li, Q., 1996. Consistent model specification tests: Omitted variables, parametric and semiparametric functional forms. Econometrica 64, 865–890.
Fan, Y., Li, Q., 2000. Consistent model specification tests: Kernel based tests vs. Bierens' ICM tests. Econometric Theory 16, 1016–1041.
Guerre, E., Lavergne, P., 2002. Optimal minimax rates for nonparametric specification testing in regression models. Econometric Theory 18, 1139–1171.
Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.
Honoré, B., Lewbel, A., 2002. Semiparametric binary choice panel data models without strictly exogenous regressors. Econometrica 70, 2053–2063.
Horowitz, J., 2006. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. Econometrica 74, 521–538.
Horowitz, J., Spokoiny, V., 2001. An adaptive, rate optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69, 599–631.
Khmaladze, E.V., 1981. Martingale approach to the goodness of fit tests. Theory of Probability and Its Applications 26, 246–265.
Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovation martingales. Annals of Statistics 21, 798–829.
Khmaladze, E.V., Koul, H., 2004. Martingale transforms goodness-of-fit tests in regression models. Annals of Statistics 32, 995–1034.
Kitamura, Y., Tripathi, G., Ahn, H., 2004. Empirical likelihood-based inference in conditional moment restriction models. Econometrica 72, 1667–1714.
Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70, 1583–1612.
Koul, H., Stute, W., 1999. Nonparametric model checks for time series. Annals of Statistics 27, 204–236.
Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. Journal of Econometrics 112, 295–325.
Neuhaus, G., 1976. Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives. Journal of Multivariate Analysis 6, 95–110.
Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.
Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.
Phillips, P.C.B., Sun, Y., 2002. Non-orthogonal Hilbert projections in trend regression. Econometric Theory 18, 1011–1015.
Powell, J., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.
Robinson, P., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.
Rosenblatt, M., 1952. Remark on multivariate transformation. Annals of Mathematical Statistics 23, 470–472.
Simpson, P.B., 1951. Note on the estimation of a bivariate distribution function. Annals of Mathematical Statistics 22, 476–478.
Song, K., 2008. Uniform convergence of series estimators over function spaces. Econometric Theory 24, 1463–1499.
Song, K., 2009. Testing conditional independence via Rosenblatt transforms. Annals of Statistics (forthcoming).
Stinchcombe, M.B., White, H., 1998. Consistent specification testing when the nuisance parameters present only under the alternative. Econometric Theory 14, 295–325.
Stute, W., 1997. Nonparametric model checks for regression. Annals of Statistics 25, 613–641.
Stute, W., Thies, S., Zhu, L., 1998. Model checks for regression: An innovation process approach. Annals of Statistics 26, 1916–1934.
Stute, W., Zhu, L., 2005. Nonparametric checks for single-index models. Annals of Statistics 33, 1048–1083.
Tripathi, G., Kitamura, Y., 2003. Testing conditional moment restrictions. Annals of Statistics 31, 2059–2095.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag.
Whang, Y., 2000. Consistent bootstrap tests of parametric regression functions. Journal of Econometrics 98, 27–46.