Testing Conditional Independence using Conditional Martingale Transforms

Kyungchul Song[1]
Department of Economics, University of Pennsylvania

November 28, 2006

Abstract

This paper investigates the problem of testing conditional independence between Y and Z given $\lambda_0(X)$ where λ_0 is an unknown parametric or nonparametric real-valued function and a consistent estimator $\hat{\lambda}$ is available. First, the paper analyzes the asymptotic power properties of the tests via approximate Bahadur asymptotic relative efficiency and demonstrates that there is improvement in power when we use nonparametric estimators of conditional distribution functions in place of true ones. Second, the paper proposes a method of conditional martingale transforms under which tests are asymptotically distribution free and asymptotically unbiased against $\sqrt{n}$-converging Pitman local alternatives. We find that when Y and Z are continuous, the power improvement from the nonparametric estimation disappears after the conditional martingale transform, but when either of Y or Z is binary, this phenomenon of power improvement appears even after the conditional martingale transform. We perform Monte Carlo simulation studies to compare the martingale transform approach and the bootstrap approach, and also compare small sample powers of tests with nonparametrically estimated distribution functions and true ones.

Key words and Phrases: Conditional independence, Asymptotic pivotal tests, conditional martingale transforms, approximate Bahadur ARE.

JEL Classifications: C12, C14, C52.

1 Introduction

This paper proposes a new class of tests of conditional independence between Y and Z given an index $\lambda_0(X)$ of X, where λ_0 is either parametrically or nonparametrically specified. Conditional

[1] I am grateful to Yoon-Jae Whang and Guido Kuersteiner who gave me valuable advice at the initial stage of this research. I also thank Hyungsik Roger Moon, Qi Li, Joon Park, Norm Swanson and seminar participants at a KAEA meeting and a workshop at Rutgers University and Texas A&M for their valuable comments. All errors are mine. Address correspondence to: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297.

independence restrictions are used in a variety of contexts of econometric modeling, and often serve as a testable implication of identifying restrictions. See the next section for relevant examples.

The literature of testing conditional independence is recent and, apparently, involves relatively few researches as compared to that of other nonparametric or semiparametric tests. Linton and Gozalo (1999) proposed tests that are slightly stronger than tests of conditional independence. Recently, Su and White (2003a, 2003b, 2003c) studied several methods of testing conditional independence. Angrist and Kuersteiner (2004) suggested distribution-free tests of conditional independence.

This paper's framework of hypothesis testing is based on an unconditional-moment formulation of the null hypothesis using indicator functions that run through an index space. Tests in our framework have the usual local power properties that are shared by other empirical-process based approaches such as Linton and Gozalo (1999) and Angrist and Kuersteiner (2004). In particular, the tests are asymptotically unbiased against $\sqrt{n}$-converging Pitman local alternatives. In contrast with Linton and Gozalo (1999), our tests are asymptotically pivotal (or asymptotically distribution-free), which means that asymptotic critical values do not depend on unknowns, so that one does not need to resort to a simulation-based method like bootstrap to obtain approximate critical values.

This latter property of asymptotic pivotalness is shared, in particular, by Angrist and Kuersteiner (2004) whose approach is close to ours in two aspects.[2] First, their test has local power properties that are similar to ours. Second, both their approach and ours employ a method of martingale-transform to obtain asymptotically distribution free tests. However, they assume that Z is binary and the conditional probability $\mathbf{P}\{Z_i = 1|X_i\}$ is parametrically specified, whereas we do not require such conditions. This latter contrast primarily stems from completely different approaches of martingale transforms employed by the two papers. In particular, Angrist and Kuersteiner (2004) employ martingale transforms that apply to tests involving estimators of finite-dimensional parameters, whereas this paper uses conditional martingale transforms that are amenable to testing semiparametric restrictions that involve nonparametric estimators (Song (2006b)).

The martingale transform approach was pioneered by Khmaladze (1988,1993) and has been used both in the statistics and economics literature. For example, Stute, Thies, and Zhu (1998) and Khmaladze and Koul (1994) studied testing nonlinear parametric specification of regressions. Koul and Stute (1999) proposed specification tests of nonlinear autoregressions. Recently in the econometrics literature, Koenker and Xiao (2002) and Bai (2003) employed the martingale transform approach for testing parametric models, and Angrist and Kuersteiner (2004) investigated testing conditional independence as mentioned before. Song (2006b) developed a method of conditional martingale transform that applies to tests of semiparametric conditional moment restrictions. Al-

[2] Several tests suggested by Su and White are also asymptotically pivotal, and consistent, but have different asymptotic power properties. Their tests are asymptotically unbiased against Pitman local alternatives that converge to the null at a rate slower than $\sqrt{n}$. However, when we extend the space of local alternatives beyond Pitman local alternatives, the rate of convergence of local alternatives that is optimal is typically slower than $\sqrt{n}$.(e.g. Horowitz and Spokoiny (2002)).

though the conditional independence restrictions can be viewed as a special case of semiparametric conditional moment restrictions, the restrictions are not encompassed by the framework of Song (2006b).

As by-product, we find that the nonparametric estimation error of $\hat{F}(y|u)$ improves the asymptotic power of the test. Here $\hat{F}(y|u)$ denotes a nonparametric estimator of $F(y|u)$, the conditional distribution function of Y given $U \triangleq F_{\lambda_0}(\lambda_0(X))$, where $F_{\lambda_0}(\cdot)$ denotes the distribution function of $\lambda_0(X)$. This gives rise to a counter-intuitive consequence that even when the true nonparametric function $F(y|u)$ is known, the estimation of the function improves the performance of the test. We formally show this in terms of approximate Bahadur asymptotic relative efficiency. We provide intuitive explanation for this, by identifying the restrictions that are utilized more effectively in nonparametric estimation than when we use true ones.

When Y and Z are continuous, we apply conditional martingale transforms to the indicator functions involving Y and Z. Since the conditional martingale transforms remove the additional local shift caused by the nonparametric estimation error in $\hat{F}(y|u)$, the phenomenon of improved asymptotic power by nonparametric estimation errors does not arise after we apply the conditional martingale transforms. However, when Z is binary, this phenomenon of improved local asymptotic power reappears, because in this case it suffices to take conditional martingale transform only on the continuous variable Y.

It is interesting to analyze the effect of the conditional martingale transform on the asymptotic power properties of tests. To this end, we compute the approximate Bahadur asymptotic relative efficiency of a Kolmogorov-Smirnov type test that uses the original empirical process relative to one that uses the transformed empirical process. From the result, we characterize a certain set of alternatives under which the conditional martingale transform improves the asymptotic power of the test.

We perform a small scale Monte Carlo simulations to compare the approach of bootstrap and that of martingale transforms. Although the results are at a preliminary stage, the martingale transform appears to perform as well as bootstrap.

In the next section, we define the null hypothesis of conditional independence and present the asymptotic representation of the semiparametric empirical process. In Section 3, we construct conditional martingale transforms and show how they can be used to generate a class of asymptotically pivotal tests. In Section 4, we present feasible transforms using series-based methods. In Section 5, we consider the case where Z_i is binary as in Angrist and Kuersteiner (2004) and discuss how this case can be accommodated into our framework. The implementation of the transform approach becomes even simpler in this case. In Section 5, we discuss results from simulation studies that compare the bootstrap approach and the martingale transform approach. In Section 6, we conclude.

2 Testing Conditional Independence

2.1 The Null and Alternative Hypotheses

Suppose that we are given with a random vector (Y, Z, X) distributed by $\mathbf{P}$ and an unknown real valued function $\lambda_0(\cdot)$ on $\mathbf{R}^{d_X}$ that satisfy the following conditions:

Assumption 1 : (i) Y and Z are continuous random variables with support in $[0, 1]$
(ii) $\lambda_0(X)$ is a continuous random variable with a distribution function F_0.

When either Y or Z is binary, the development of this paper's thesis becomes simpler. We will discuss the case when Z is binary in a later section. The thesis of this paper can be extended to the case of Y and Z being random vectors, only to cause more complicated notations. Our restricting the support of (Y, Z) into a unit square is innocuous, since we can transform its marginals using a strictly increasing function taking values in $[0, 1]$. Condition (ii) is mostly satisfied, unless the function $\lambda_0(\cdot)$ is equal to a constant over an area in which X has a positive support. Condition (ii) is convenient, enabling us to employ a quantile transform of the conditioning variable $\lambda_0(X)$. The quantile transform is useful because as it turns out later, the estimation error from the estimator $\hat{\lambda}$ of λ_0 does not alter the asymptotic representation of the test statistic.

We are interested in testing the null hypothesis of conditional independence of Y and Z given $\lambda_0(X)$. Since the distribution function F_0 of $\lambda_0(X)$ is strictly increasing on the support of $\lambda_0(X)$, the σ-field generated by $\lambda_0(X)$ is equal to that generated by its quantile transform $U = F_0(\lambda_0(X))$. By using notations for indicator functions:

$$\gamma_z(Z) \triangleq 1\{Z \leq z\},\ \gamma_y(Y) \triangleq 1\{Y \leq y\},\ \text{and } \gamma_u(U) \triangleq 1\{U \leq u\},$$

we can write the null hypothesis of conditional independence as (e.g. see Theorem 9.2.1. in Chung (2001), p.322):

$$H_0 : \mathbf{E}(\gamma_y(Y)|U, Z) = \mathbf{E}(\gamma_y(Y)|U),\ \mathbf{P}\text{-a.s. } \forall y \in [0, 1]. \tag{1}$$

Then the alternative hypothesis is given by the negation of the null:

$$H_1 : \mathbf{P}\{\mathbf{E}(\gamma_y(Y)|U, Z) = \mathbf{E}(\gamma_y(Y)|U) \text{ for some } y \in [0, 1]\} > 0.$$

We assume throughout the paper that to each potential data generating process $\mathbf{P}$ corresponds a unique function $\lambda_0(\cdot)$. This means that a unique function $\lambda_0(\cdot)$ is identified from the data generating process $\mathbf{P}$ regardless of whether it belongs to the null hypothesis or to the alternatives. In the next subsection, we discuss two examples of econometric specifications that use conditional independence.

2.2 Examples

2.2.1 Nonseparable Models

Conditional independence restriction is often used in nonseparable models. The first example is a model of weakly separable endogenous dummy variables. Suppose Y and Z are generated by

$$\begin{aligned} Y &= g(\nu_1(X,Z),\varepsilon) \text{ and} \\ Z &= 1\{\nu_2(X) \geq u\}. \end{aligned}$$

For example, this model can be useful when a researcher attempts to analyze the effect of job training of a worker upon her earnings prospects. Here Y denotes the earnings outcome of a worker, and Z denotes the indicator for the job training. It will be of interest to a researcher whether the indicator for the job training remains still endogenous even after controlling for covariates X, because in the presence of endogeneity, the researcher might need to include some additional covariate into the specification of the endogenous job training that has a sufficient degree of variation independent from X. (e.g. Vytlacil and Yildiz (2006)). The exogeneity of Z for the outcome Y given the covariate X will be represented by the conditional independence restriction:

$$Y \perp Z|\nu_1(X,Z),\nu_2(X).$$

A second example is from Altonji and Matzkin (2005) who considered the following nonseparable regression model.

$$Y = m(Z,\varepsilon).$$

One of the assumption used to identify the local average response was that ε is independent of Z given an instrumental variable X. This gives the following testable implication of conditional independence restriction (e.g. Dawid (1980)),

$$Y \perp Z|X.$$

2.2.2 Heterogeneous Treatment Effects

In the literature of heterogeneous treatment effects, often the restriction of conditional independence

$$(Y_0,Y_1) \perp Z \mid X$$

is considered.(e.g. Rubin (1978), Rosenbaum and Rubin (1983), and Hirano, Imbens, and Ridder (2003) to name but a few.)[3] This conditions is also called the condition of unconfounded treatment assignment. Here Y_0 represents the outcome when an agent is not treated and Y_1, the outcome when treated. The conditional probability of being treated $p(X) = \mathbf{P}(Z=1|X)$ given the covariate X

[3] It is also worth noting that Heckman, Ichimura, and Todd (1998) point out that a weaker condition of mean independence suffices for the identication of the average treatment effect.

is called the propensity score. When $p(X) \in (0,1)$, it follows that (Rosenbaum and Rubin (1983))

$$(Y_0, Y_1) \perp Z \mid p(X).$$

This restriction is not directly testable because we do not observe Y_0 and Y_1 simultaneously for each individual. We can at least test its implications:

$$(1-Z)Y \perp Z \mid p(X) \text{ or } ZY \perp Z \mid p(X).$$

While we may specify $p(\cdot)$ parametrically, the procedure of this paper also allows for the case when $p(\cdot)$ is specified as a nonparametric function.

3 An Asymptotic Representation of a Semiparametric Empirical Process

3.1 Asymptotic Representation

The null hypothesis of (1) is a conditional moment restriction. We can obtain an equivalent formulation in terms of unconditional moment restrictions. Let us define

$$g_y(u) \triangleq \mathbf{E}(\gamma_y(Y)|U=u) \text{ and } g_z(u) \triangleq \mathbf{E}(\gamma_z(Y)|U=u)$$

Then the null hypothesis can be written as:

$$H_0 : \mathbf{E}[\gamma_u(U)\gamma_z(Z)(\gamma_y(Y) - g_y(U))] = 0,\ \forall (y,z,u) \in [0,1]^3.$$

Test statistics we analyze in the paper are based on the sample analogue of the above moment restrictions. Suppose that we are given a random sample $\{S_i\}_{i=1}^n = \{(Y_i, Z_i, U_i)\}_{i=1}^n$ of $S = (Y, Z, U)$. If the function $g_y(\cdot)$ were known and the quantile transformed data U_i were observed, a test statistic could be constructed as a functional of the following stochastic process:

$$\nu_n(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_u(U_i)\gamma_z(Z_i)(\gamma_y(Y_i) - g_y(U_i)),\ r = (y,z,u). \tag{2}$$

The process cannot be used to construct a test statistic because it involves unknown g_y. In this section, we introduce a series estimator of g_y and investigate the asymptotic properties of the resulting empirical process that involves the estimator.

We approximate $g_y(u)$ by $p^K(u)'\pi_y$ which is constituted by a basis function vector $p^K(u)$ and a coefficient vector π_y. Given an appropriate estimator $\hat{\lambda}(\cdot)$ of $\lambda(\cdot)$, define

$$\hat{U}_i \triangleq \frac{1}{n} \sum_{j=1, j\neq i}^n 1\left\{\hat{\lambda}(X_j) \leq \hat{\lambda}(X_i)\right\}. \tag{3}$$

The series-based estimator is defined as $\hat{g}_y(u) \triangleq p^K(u)'\hat{\pi}_y$, where $\hat{\pi}_y \triangleq [\hat{P}'\hat{P}]^{-1}\hat{P}'a_{y,n}$ with $a_{y,n}$ and $\hat{P}$ being defined by

$$a_{y,n} \triangleq \begin{bmatrix} \gamma_y(Y_1) \\ \vdots \\ \gamma_y(Y_n) \end{bmatrix} \text{ and } \hat{P} \triangleq \begin{bmatrix} p^K(\hat{U}_1)' \\ \vdots \\ p^K(\hat{U}_n)' \end{bmatrix}.$$

Using the estimated conditional mean function $\hat{g}_y$, we can construct a feasible version of the process $\nu_n(r)$:

$$\hat{\nu}_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i)\gamma_z(Z_i)(\gamma_y(Y_i) - \hat{g}_y(\hat{U}_i)).$$

We call the process $\hat{\nu}_n(r)$ a semiparametric empirical process. A test statistic can be constructed as a known functional of the process $\hat{\nu}_n(r)$. Our main result in this section provides a uniform asymptotic representation of the process $\hat{\nu}_n(r)$ that holds both under the null hypothesis and under the alternatives. To this end, we introduce notations and assumptions.

We define the $L_p(\mathbf{P})$ norm $\|\cdot\|_p$ by $||g||_p \triangleq (\mathbf{E}\,\|g(S_i)\|^p)^{1/p}$ and define a uniform norm $||\cdot||_\infty$ by $||g||_\infty \triangleq \sup_s |g(s)|$. We also introduce the notation

$$\zeta_{\nu,K} \triangleq \sup_{\mu \leq \nu} \sup_u |D^\mu p^K(u)|, \tag{4}$$

where D^μ is the derivative operator defined by

$$D^\mu g(u) \triangleq (\partial^{|\mu|}/\partial z_1^{\mu_1} \cdots \partial z_{d_Z}^{\mu_{d_Z}})(g(u)). \tag{5}$$

The notation $\rightsquigarrow$ denotes weak-convergence in the sense of Hoffman-Jorgensen (e.g. van der Vaart and Wellner (1996)). Let $F_{Z|U}(z|u)$ and $f_{Z|U}(z|u)$ denote the conditional distribution function and the conditional density of Z_i given U_i. We define similarly $F_{Y|U}(y|u)$ and $f_{Y|U}(y|u)$. Let $N_{[]}(\varepsilon, \Lambda, ||\cdot||_p)$ denote the L_p-bracketing number, i.e., the smallest number r of pairs of functions $(\lambda_j, \Delta_j)_{j=1}^r$ in $L_p(\mathbf{P})$ such that $||\Delta_j||_p < \varepsilon$, and for all $\lambda \in \Lambda$, there exists a pair (λ_j, Δ_j), $j \in \{1, \cdot\cdot, r\}$ satisfying the inequality: $|\lambda_j(x) - \lambda(x)| < \Delta_j(x)$.

Assumption 2 : (i) $(Y_i, Z_i, X_i)_{i=1}^n$ is a random sample from $\mathbf{P}$ where $\mathbf{E}||Y||^p < \infty$, $\mathbf{E}||Z||^p < \infty$ and $\mathbf{E}||X||^p < \infty$ for some $p > 4$. (ii) There exist a class Λ of functions that contain λ_0 and satisfy that:(a) for a constant $b_1 \in [0, 2)$, $\log N_{[]}(\varepsilon, \Lambda, ||\cdot||_\infty) < C\varepsilon^{-b_1}$, (b) $||\hat{\lambda} - \lambda_0||_\infty = o_P(n^{-1/4})$, and (c) $\mathbf{P}\{\hat{\lambda} \in \Lambda\} \to 1$ as $n \to \infty$, where F is the distribution function of X.

When Λ is a parametric class such that $\Lambda \triangleq \left\{\lambda_\theta : \theta \in \Theta \subset \mathbf{R}^{d_\theta}\right\}$ with Θ being compact and λ_θ is locally uniformly L_p-continuous in $\theta \in \Theta$ in the sense of Chen, Linton, and van Keilegom (2003), the condition (ii)(a) holds for arbitrarily small $b > 0$. The condition of L_p-uniform continuity is much weaker than the requirement that λ_θ be Lipschitz in θ. When λ_0 is a nonparametric function, we can take Λ to be a function space containing λ_0. The uniform consistency of $\hat{\lambda}$ for λ_0 is required to hold both under the null hypothesis and under the alternatives. We may view λ_0 as a uniform probability limit of $\hat{\lambda}$. Condition (ii)(c) can be fulfilled, for instance, by taking a smooth basis functions.

Let us define a shrinking neighborhood of λ_0 : $\Lambda_n \triangleq \{\lambda \in \Lambda : ||\lambda - \lambda_0||_\infty \leq Cn^{-b}\}$, for $b \in (1/4, 1/2]$ and for some positive constant $C > 0$. The following are a set of assumptions for basis functions.

Assumption 3 : (i) $\lambda_{\min}\left(\int p^K(u)p^K(u)du\right) > 0$.
(ii) There exist d_1 and d_2 such that (a) for some $d > 0$, there exist sets of vectors $\{\pi_y : y \in [0,1]\}$ and $\{\pi_z : z \in [0,1]\}$ such that

$$\begin{aligned}\sup_{(y,u)\in[0,1]^2} \left|p^K(u)'\pi_y - g_y(u)\right| &= O(K^{-d_1}) \text{ and} \\ \sup_{(z,u)\in[0,1]^2} \left|p^K(u)'\pi_z - g_z(u)\right| &= O(K^{-d_2}),\end{aligned}$$

(b) for each $\bar{u} \in [0,1]$ there exist sets of vectors in $\mathbf{R}^K$, $\{\pi_{y,\bar{u}} : y \in [0,1]\}$, such that

$$\sup_{(y,\bar{u})\in[0,1]\times[0,1]} \left(\int_0^1 \left|P^K(u)'\pi_{y,\bar{u}} - g_y(u)1\{\bar{u} \leq u\}\right|^p du\right)^{1/p} = O(K^{-d_1}).$$

(iii) For d_1 and d_2 in (ii), $\sqrt{n}\zeta_{0,K}^2 K^{-d} = o(1)$ and $\sqrt{n}\zeta_{1,K}K^{-d_2} = o(1)$
(iv) For b in the definition of Λ_n, we have $n^{1/2-2b}\zeta_{0,K}^3\zeta_{2,K} = o(1)$, $n^{-1/2+1/p}K^{1-1/p}\zeta_{0,K}^2 = o(1)$ and $n^{-b}\zeta_{0,K}\{\sqrt{\zeta_{0,K}\zeta_{2,K}} + \zeta_{1,K}\} = o(1)$.

Condition (i) is used by Andrews (1991) and Newey (1997), and can be satisfied by choosing an orthonormal basis function that is supported in $[0,1]$. Condition (ii) requires the rate of convergence for uniform approximation errors to be of a polynomial order in K.

Assumption 4 : (i) The density $f_\lambda(\bar{\lambda})$ of $\lambda(X)$ satisfies $\sup_{\lambda\in\Lambda_n}\sup_{\bar{\lambda}\in R} f_\lambda(\bar{\lambda}) < \infty$.
(ii) For some $C > 0$, $\sup_{\lambda\in\Lambda_n}\sup_{\bar{\lambda}\in\mathcal{S}_\lambda}\left|F_\lambda(\bar{\lambda}+\delta) - F_\lambda(\bar{\lambda}-\delta)\right| < C\delta$, for all $\delta > 0$.
(iii) There exists $C > 0$ such that for each $u, u_1 \in \mathcal{U} \triangleq \{F_\lambda \circ \lambda : \lambda \in \Lambda_n\}$ and for each $\bar{u}_2 \in [0,1]$,

$$\begin{aligned}&\sup_{(\bar{u}_1,\bar{u})\in[0,1]^2:|\bar{u}_2-\bar{u}_1|<\delta} |P(Y \leq y, u_1(X) \leq \bar{u}|u(X) = \bar{u}_1) - P(Y \leq y, u_1(X) \leq \bar{u}|u(X) = \bar{u}_2)| \\ &\leq \varphi_{u,u_1}(y,x)\delta,\end{aligned}$$

where $\varphi_{u,u_1}(\cdot,\cdot)$ is a real function that satisfies $\sup_{x\in\mathcal{S}_X}\int|\tilde{\alpha}(y)|\varphi_{u,u_1}(y,x)dy < C$ and $\int\varphi_{u,u_1}(y,x)dx < Cf_Y(y)$ with $f_Y(\cdot)$ denoting the density of Y.
(iv) The conditional distribution functions $F_{Y|U}(y|u)$ and $F_{Z|U}(z|u)$ are second-order continuously differentiable with uniformly bounded derivatives.

Conditions in Assumption 4 are concerned about the densities and conditional densities of Y, Z, and $\lambda(X)$. The conditions are primarily used to resort to a lemma in Escanciano and Song (2006) on which the result of Theorem 1 below relies. Condition (i) requires that the density of $\lambda(X)$ for $\lambda \in \Lambda_n$ is uniformly bounded. A similar condition was used by Stute and Zhu (2005). Condition (ii) says that the distribution function of $\lambda(X)$ should be Lipschitz continuous with Lipschitz coefficient uniformly bounded over Λ_n. Condition (iii) requires Lipschitz continuity of conditional distribution of $(Y, u_1(X)),\ u \in \mathcal{U}$, given $u(X)$ in the conditioning variable.

Theorem 1 : Suppose Assumptions 1-4 hold. Then the following holds:

$$\sup_{r\in[0,1]^3} \left| \hat{\nu}_n(r) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left\{ \gamma_z(Z_i) - g_z(U_i) \right\} \left\{ \gamma_y(Z_i) - g_y(U_i) \right\} \right| = o_P(1). \tag{6}$$

The result of Theorem 1 provides a uniform asymptotic representation of the semiparametric empirical process $\hat{\nu}_n(r)$. This asymptotic representation motivates the construction of proper conditional martingale transforms that we develop in a later section. The representation also constitutes the basis upon which we derive the power properties of the test. Indeed, observe that the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left\{ \gamma_z(Z_i) - g_z(U_i) \right\} \left\{ \gamma_y(Z_i) - g_y(U_i) \right\}$$

is a zero mean process only when the null hypothesis of conditional independence is true. From the asymptotic representation, we can derive the asymptotic properties of the process $\hat{\nu}_n(r)$ as follows.

Corollary 1 : (i) (a) Under the null hypothesis, $\nu_n(r) \rightsquigarrow \nu(r)$, where ν is a Gaussian process whose covariance kernel is given by

$$c(r_1; r_2) = \int_{-\infty}^{u_1 \wedge u_2} C(w_1, w_2; u) du$$

with $C(w_1, w_2; u) = F_{Z|U}(z_1 \wedge z_2|u) \times \{F_{Y|U}(y_1 \wedge y_2|u) - F_{Y|U}(y_1|u)F_{Y|U}(y_2|u)\}$.
(b) Under fixed alternatives, $n^{-1/2}\nu_n(r) \rightsquigarrow \langle \gamma_u, \gamma_z(\gamma_y - g_y) \rangle$.
(ii) (a) Under the null hypothesis, $\hat{\nu}_n(r) \rightsquigarrow \nu_1(r)$, where ν_1 is a Gaussian process whose covariance kernel is given by

$$c_1(r_1; r_2) = \int_{-\infty}^{u_1 \wedge u_2} C_1(w_1, w_2; u) dF(u)$$

with

$$C_1(w_1, w_2; u) = \left[F_{Z|U}(z_1 \wedge z_2|u) - F_{Z|U}(z_1|u)F_{Z|U}(z_2|u)\right] \times \left[F_{Y|U}(y_1 \wedge y_2|u) - F_{Y|U}(y_1|u)F_{Y|U}(y_2|u)\right].$$

(b) Under fixed alternatives, $n^{-1/2}\hat{\nu}_n(r) \rightsquigarrow \langle \gamma_u, \gamma_z(\gamma_y - g_y) \rangle$.

The results of Corollary 1 lead to the asymptotic properties of tests based on the processes $\nu_n(r)$ and $\hat{\nu}_n(r)$. The limiting processes $\nu(r)$ and $\nu_1(r)$ under the null hypothesis are Gaussian processes with different covariance kernels. Under the fixed alternatives, the two empirical processes $\nu_n(r)$

and $\hat{\nu}_n(r)$ weakly converge to the same shift term after the normalization by $\sqrt{n}$. This has two interesting consequences. First, the local power properties under $\sqrt{n}$-converging Pitman local alternatives naturally follow from this result. Second, the result in Corollary 1 gives rise to a surprising consequence that using the nonparametrically estimated distribution function $\hat{F}(y|u)$ instead of the true one $F(y|u)$ can improve the asymptotic power of the tests. In particular, the covariance kernel of the limit process under the null hypothesis "shrinks" due to the use of $\hat{F}(y|u)$ instead of $F(y|u)$ whereas the limit shift $\langle \gamma_u, \gamma_z(\gamma_y - g_y)\rangle$ under the alternatives remains intact. In a subsequent subsection, we formalize this observation using approximate Bahadur asymptotic relative efficiency.

It is worth noting that the estimation error in $\hat{\lambda}$ does not play a role in determining the limit behavior of the process $\hat{\nu}_n(r)$. In other words, the results of Theorem 1 and Corollary 1 do not change when we replace $\hat{\lambda}$ by λ_0. A similar phenomenon is discovered and analyzed in Stute and Zhu (2005) in the context of testing single-index restrictions using a kernel method. This convenient phenomenon is due to the use of the empirical quantile transform of $\hat{\lambda}(X_i)$.

We construct the following test statistics:

$$T_{KS} = \sup_{r\in[0,1]^3} |\hat{\nu}_n(r)| \text{ and } T_{CM} = \int_{[0,1]^3} \hat{\nu}_n(r)^2 dr. \tag{7}$$

The test statistic T_{KS} is of Kolmogorov-Smirnov type and the test statistic T_{CM} is of Cramér-von Mises type. The asymptotic properties of the tests based on T_{KS} and T_{CM} follow from Theorem 1. Indeed, under the null hypothesis,

$$T_{KS} \rightsquigarrow \sup_{r\in[0,1]^3} |\nu(r)| \text{ and } T_{CM} \rightsquigarrow \int_{[0,1]^3} \nu_1(r)^2 dr,$$

and under the Pitman local alternatives $\mathbf{P}_n$ such that for some $a_{y,z}(\cdot)$, $\mathbf{E}_n(\gamma_z(Z)(\gamma_y(Y)-g_y(U))|U = u) = a_{y,z}(u)/\sqrt{n} + o(n^{-1/2})$,

$$\hat{\nu}_n(r) \rightsquigarrow \nu_1(r) + \langle \gamma_u, a_{y,z}\rangle.$$

The presence of the non-zero shift term $\langle \gamma_u, a_{y,z}\rangle$ renders the test to have nontrivial asymptotic power against such local alternatives.

The asymptotic critical values cannot be computed directly from the limiting distributions of the test statistics, because the test statistics depend on unknown components of the data generating process and in consequence the tests are not asymptotically pivotal. This stems from the fact that the limiting Gaussian processes $\nu(r)$ and $\nu_1(r)$ have complicated kernels which, in particular, contain Brownian bridge type components

$$F_{Y|U}(y_1 \wedge y_2|u) - F_{Y|U}(y_1|u)F_{Y|U}(y_2|u) \text{ and}$$
$$F_{Z|U}(z_1 \wedge z_2|u) - F_{Z|U}(z_1|u)F_{Z|U}(z_2|u).$$

This source of complication is different in nature from the one that is simply due to the estimation error of parameters in the conditional moment restriction (e.g. Stute (1997), Stute, Thies and Zhu (1998), Khmaladze and Koul (2005), Song (2006b)). Rather, the situation is close to the fundamental problem of goodness-of-fit tests of multivariate distributions (e.g. Khmaladze (1993)) in which the limiting distribution of a test statistic is equal to a functional of a multiparameter Brownian bridge.

There are primarily two approaches to deal with this problem: the bootstrap approach and the martingale transform approach. The bootstrap approach uses critical values from the distribution of bootstrap samples. The martingale transform approach transforms the test so that critical values can be tabulated from the first order asymptotic theory. This paper focuses on developing the latter approach.

3.2 The Effect of Nonparametric Estimation Error

We analyze how our use of a nonparametric estimator for $F(y|u)$ in the process $\hat{\nu}_n(r)$ affects the power of the test. Surprisingly, we find that the power of the test improves by using the estimator instead of a true one. In other words, the test becomes better when we use estimator $\hat{F}(y|u)$ even if we know the true conditional distribution function $F(y|u)$. This is because the use of nonparametric estimator has an effect of "shrinking" the covariance of the limiting process $\nu(r)$ under the null hypothesis while preserves the local shift under the Pitman local alternatives.

We formalize this observation by employing the notion of approximate Bahadur asymptotic relative efficiency (ARE). Approximate Bahadur ARE was introduced by Bahadur (1960) and has been widely used in the literature (see Nikitin (1995) and references therein). While it is more desirable to use exact Bahadur efficiency, approximate Bahadur efficiency is very easy to compute and can provide a quick idea in comparison of tests (Nikitin (1995)). We relegate a further analysis using exact Bahadur efficiency to a future research.

Suppose Γ is a known continuous functional that is homogeneous of degree one, is applied to stochastic processes $\nu_n(r)$ and $\hat{\nu}_n(r)$ and used to construct a test statistic. For example, the Cramér-von Mises type test statistics can be defined as $\Gamma\nu_n$ and $\Gamma\hat{\nu}_n$ with Γ being a functional defined by $\Gamma g = \int g^2(r)dr$. Then denote by $H_0(v)$ and $H_1(v)$ the limits of the rejection probability of tests based on $\Gamma\nu_n$ and $\Gamma\hat{\nu}_n$ under the null hypothesis. That is, under the null hypothesis,

$$\mathbf{P}\left\{\Gamma\nu_n \geq v\right\} \to h_0(\kappa) \text{ and } \mathbf{P}\left\{\Gamma\hat{\nu}_n \geq \kappa\right\} \to h_1(\kappa)$$

as $n \to \infty$. This often boils down to the computation of tail probabilities of a functional of Gaussian processes. Suppose the functional Γ is taken to be homogeneous of degree one and there exist constants b_0 and b_1 satisfying

$$\ln h_0(\kappa) = -\frac{1}{2}b_0\kappa^2 + o(1) \text{ and } \ln h_1(\kappa) = -\frac{1}{2}b_1\kappa^2 + o(1) \text{ as } \kappa \to \infty.$$

On the other hand, under the fixed alternative $\mathbf{P}_1$, suppose we have functions φ_0 and φ_1 on

$[0,1]^3$ such that

$$\begin{aligned} \frac{1}{\sqrt{n}}\Gamma\nu_n &= \Gamma\left(\frac{1}{\sqrt{n}}\nu_n\right) \to_p \Gamma\varphi_0, \text{ and} \\ \frac{1}{\sqrt{n}}\Gamma\hat{\nu}_n &= \Gamma\left(\frac{1}{\sqrt{n}}\hat{\nu}_n\right) \to_p \Gamma\varphi_1. \end{aligned} \quad (8)$$

Then the approximate Bahadur ARE (denoted by $e_{0,1}(\mathbf{P}_1)$) of a test based on the test statistic $\Gamma\nu_n$ relative to one that is based on $\Gamma\hat{\nu}_n$ is computed from (Nikitin (1995)),

$$\sqrt{e_{0,1}(\mathbf{P}_1)} = \sqrt{\frac{b_0}{b_1}} \times \frac{\Gamma\varphi_0}{\Gamma\varphi_1}.$$

The approximate Bahadur ARE is composed of two ratios. The first ratio b_1/b_0 measures the relative tail behavior of the limiting distribution of the test statistic under the null hypothesis. The second ratio $\Gamma\varphi_0/\Gamma\varphi_1$ measures how sensitively the test statistics deviate from the null limiting distribution in large samples under the alternatives. When $e_{0,1}(\mathbf{P}_1) < 1$, it is likely that the power of the test based on $\Gamma\hat{\nu}_n$ is better than one based on $\Gamma\nu_n$. The following result shows that this is indeed the case.

Corollary 2 : Suppose the conditions of Theorem 1 hold and choose the Kolmogorov-Smirnov functional $\Gamma\nu(r) = \sup_{r\in[0,1]^3} |\nu(r)|$. Then

$$e_{1,0}(\mathbf{P}_1) = \frac{1}{4}.$$

The result in Corollary 2 is well expected from the result of Theorem 1. The main reason is because using the nonparametric estimator $\hat{F}(y|u)$ in place of $F(y|u)$ has an effect of "shrinking" the kernel of the limiting Gaussian process from

$$c_0(r_1, r_2) = (u_1 \wedge u_2)(z_1 \wedge z_2)(y_1 \wedge y_2 - y_1 y_2)$$

to $c_1(r_1, r_2) = (u_1 \wedge u_2)(z_1 \wedge z_2 - z_1 z_2)(y_1 \wedge y_2 - y_1 y_2)$. Hence with the same fixed level α, the critical value based on the second Gaussian process is smaller than the first Gaussian process. On the other hand the local limit shifts of the test statistics under the alternatives are identical to $\langle \gamma_u, \rho_w \rangle$ regardless of whether one uses $F(y|u)$ or $\hat{F}(y|u)$. This yields the counter-intuitive result that even when we know the conditional distribution function $F(y|u)$, it is better to use its estimator $F(y|u)$ in the test statistic.[4]

[4] Khmaladze and Koul (2005) also discuss the case when the power is increased by using estimators instead of true ones in the context of parametric tests.

3.3 Discussion

In the literature of estimation, there are findings that report that under certain circumstances, the estimation of nonparametric functions improves the efficiency of the estimators upon that when true functions are used. (e.g. Hahn (1998), and Hirano, Imbens, and Ridder (2003) and see references therein.) This efficiency gain arises due to the fact that the nonparametric estimation utilizes the additional restrictions in a manner that is not utilized when we use the true ones. We give an explanation of our findings in the preceding section in a similar vein.

The primary source of the improvement of the asymptotic power stems from the following asymptotic representation (for a general result, see Lemma 1U of Escanciano and Song (2006))

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_u(U_i)\gamma_z(Z_i)\{\gamma_y(Y_i)-\hat{g}_y(U_i)\} \\ = \; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_u(U_i)\{\gamma_z(Y_i)-g_z(U_i)\}\{\gamma_y(Y_i)-g_y(U_i)\}+o_P(1). \end{aligned}$$

As noted in the remarks after Theorem 1, the variance of the leading sum on the right-hand side is smaller than that of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_u(U_i)g_z(Z_i)\{\gamma_y(Y_i)-g_y(U_i)\}$ under the null hypothesis. When we replace $\gamma_z(Z)$ by $g_z(U)$ on the left-hand side, we obtain the following result

$$\frac{1}{n}\sum_{i=1}^{n}\gamma_u(U_i)g_z(U_i)\{\gamma_y(Y_i)-\hat{g}_y(U_i)\}=o_P(n^{-1/2}). \tag{9}$$

This fact provides a contrast to the result obtained using the true nonparametric function g_y :

$$\frac{1}{n}\sum_{i=1}^{n}\gamma_u(U_i)g_z(U_i)\{\gamma_y(Y_i)-g_y(U_i)\}=O_P(n^{-1/2}). \tag{10}$$

The rate of convergence for the sample version of $\mathbf{E}\left[\gamma_u(U)g_z(U)\{\gamma_y(Y)-g_y(U)\}\right]$ is faster with the nonparametric estimator $\hat{g}_y$ than the true one. Actually we can obtain the same contrast between (9) and (10) when we replace $\gamma_u(U)g_z(U)$ by any function $\varphi(U)$. In other words, by using the nonparametric estimation in the test statistic, we utilize the conditional moment restriction $\mathbf{E}\left[\gamma_y(Y)-g_y(U)|U\right]=0$ more effectively than by using the true function g_y. The conditional moment restriction holds both under the null hypothesis and under the alternatives. Under the alternatives, this effective utilization of the conditional moment restriction by nonparametric estimation has an effect of order that is dominated by the $\sqrt{n}$-diverging shift term which remains the same regardless of whether we use g_y or $\hat{g}_y$. Combined with this, the first order effect of nonparametric estimation under the null hypothesis results in improvement in power over the test using the true one g_y.

4 Conditional Martingale Transform

A recent work by the author (Song (2006b)) demonstrates that the method of conditional martingale transform can be useful to obtain asymptotically pivotal semiparametric tests. The method of conditional martingale transform proposed in the work does not apply to the testing problem of conditional independence for two reasons. First, the generalized residual function $\rho_{y,z} \triangleq \gamma_z(\gamma_y - g_y)$ is indexed by $(y,z) \in [0,1]^2$ whereas the generalized residual function $\rho_{\tau,\theta}$ in Song (2006b) does not depend on the index of the empirical process. Second, while Song (2006b) requires that the conditional distribution of instrumental variables conditional on the variable inside the nonparametric function should not be degenerate, this condition does not hold here. In the current situation of testing conditional independence, the instrumental variable in the conditional moment restriction and the variable inside the nonparametric function are identically X. Therefore, the nondegeneracy of the conditional distribution fails.

The main idea of this paper is that first, we take into account the null restriction of conditional independence when we search for a proper transformation of the semiparametric empirical process $\hat{\nu}_n(r)$ using conditional martingale transforms, and then analyze the asymptotic behavior of the semiparametric empirical process after the transform.[5]

4.1 Preliminary Heuristics

In this section, we provide some heuristics of the mechanics by which conditional martingale transforms work. Let us define the conditional inner product $\langle \cdot, \cdot \rangle_u$ by

$$\langle f, g \rangle_u \triangleq \int (fg)(s) dF(s|u)$$

where $F(s|u)$ is the conditional distribution function of S given $U = u$. Suppose that we are given with an operator $\mathcal{K}$ on the space of indicator functions such that the transformed indicator functions $\gamma_z^{\mathcal{K}} \triangleq \mathcal{K}\gamma_z$ and $\gamma_y^{\mathcal{K}} \triangleq \mathcal{K}\gamma_y$ satisfy the orthogonality condition and the isometry condition with respect to the conditional inner product almost surely:

$$\begin{aligned} \text{Conditional Orthogonality :} \quad \langle \gamma_z^{\mathcal{K}}, 1 \rangle_U &= 0 \text{ and } \langle \gamma_y^{\mathcal{K}}, 1 \rangle_U = 0, \text{and} \\ \text{Conditional Isometry :} \quad \langle \gamma_{z_1}^{\mathcal{K}}, \gamma_{z_2}^{\mathcal{K}} \rangle_U &= \langle \gamma_{z_1}, \gamma_{z_2} \rangle_U \text{ and } \langle \gamma_{y_1}^{\mathcal{K}}, \gamma_{y_2}^{\mathcal{K}} \rangle_U = \langle \gamma_{y_1}, \gamma_{y_2} \rangle_U. \end{aligned}$$

An operator $\mathcal{K}$ that satisfies these two properties can be used to construct a transformed empirical process from which asymptotically distribution free tests can be generated. Recall that by Theorem 1, the process $\hat{\nu}_n(r)$ has the following asymptotic representation (using our conditional inner product

[5] By noting that the limiting processes ν and ν_1 are variants of a 4-sided tied down Kiefer process, we may consider a martingale transform for a 4-sided tied down Kiefer process that was suggested by McKeague and Sun (1996). The martingale transform for this case is represented as two sequential martingale transforms. However, the sequential martingale transforms will be very complicated in practice because it involves taking repreated martingale transforms of a function.

notation: $\langle \gamma_y, 1\rangle_{U_i} \triangleq \mathbf{E}\left[\gamma_y(Y_i)|U_i\right]$)

$$\hat{\nu}_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left\{\gamma_z(Y_i) - \langle \gamma_z, 1\rangle_{U_i}\right\} \left\{\gamma_y(Y_i) - \langle \gamma_y, 1\rangle_{U_i}\right\} + o_P(1).$$

When we replace the indicator functions γ_z and γ_y by the transformed ones $\gamma_z^{\mathcal{K}}$ and $\gamma_y^{\mathcal{K}}$, the right-hand side asymptotic representation turns into the following (by the conditional orthogonality of the transform):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left\{\gamma_z^{\mathcal{K}}(Y_i) - \langle \gamma_z^{\mathcal{K}}, 1\rangle_{U_i}\right\} \left\{\gamma_y^{\mathcal{K}}(Y_i) - \langle \gamma_y^{\mathcal{K}}, 1\rangle_{U_i}\right\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i)\gamma_z^{\mathcal{K}}(Z_i)\gamma_y^{\mathcal{K}}(Y_i). \quad (11)$$

We can show that the last term is equal to

$$\sqrt{n}\mathbf{E}\left[\gamma_u(U)\langle \gamma_z^{\mathcal{K}}, \gamma_y^{\mathcal{K}}\rangle_U\right] + \nu^{\mathcal{K}}(r) + o_P(1) \quad (12)$$

where $\nu^{\mathcal{K}}(r)$ is a certain Gaussian process. The asymptotic behavior of the leading two terms determine the asymptotic properties of tests that are based on the transformed empirical process. Suppose that we are under the null hypothesis. Then the first term in (12) becomes zero by the conditional independence restriction. As shown in the appendix, the Gaussian process $\nu^{\mathcal{K}}(r)$ has a covariance function of the form

$$c(r_1, r_2) \triangleq \mathbf{E}[\gamma_{u_1 \wedge u_2}(U)\gamma_{z_1 \wedge z_2}(Z)\gamma_{y_1 \wedge y_2}(Y)]. \quad (13)$$

This form is obtained by using the conditional isometry of the operator $\mathcal{K}$ and the null hypothesis of conditional independence. Hence the resulting Gaussian process is a time-transformed Brownian sheet and we can rescale it into a standard Brownian sheet by using an appropriate normalization. Against alternatives $\mathbf{P}_1$ such that $\mathbf{P}_1\{|\langle \gamma_z^{\mathcal{K}}, \gamma_y^{\mathcal{K}}\rangle_U| > 0\} > 0$, a test properly based on the transformed process will be consistent.

The interesting phenomenon that the nonparametric estimation error improves the power disappears for tests based on transformed processes of the kind in (11). Consider the asymptotic representation of the empirical process $\nu_n(r)$ that uses the true conditional distribution function $F_y(y|u)$:

$$\nu_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i)\gamma_z(Y_i) \left\{\gamma_y(Y_i) - \langle \gamma_y, 1\rangle_{U_i}\right\} + o_P(1).$$

When we apply the transform $\mathcal{K}$ to the indicator functions of γ_y and γ_z, the right-hand side is equal to

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \gamma_z^{\mathcal{K}}(Y_i) \left\{ \gamma_y^{\mathcal{K}}(Y_i) - \langle \gamma_y^{\mathcal{K}}, 1 \rangle_{U_i} \right\} + o_P(1) \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \gamma_z^{\mathcal{K}}(Y_i) \gamma_y^{\mathcal{K}}(Y_i) + o_P(1).
\end{aligned}
$$

Hence the asymptotic representation of the process after the transform $\mathcal{K}$ coincides with that in (11). The asymptotic power properties remain the same regardless of whether we use $F_y(y|u)$ or $\hat{F}_y(y|u)$ once the transform $\mathcal{K}$ is applied.

4.2 Conditional Martingale Transforms

In this section, we discuss a method to obtain the operator $\mathcal{K}$ that satisfies the conditional orthogonality and the conditional isometry properties. Khmaladze (1993) proposes using a class of isometric projection operators on L_2-space. This class of isometric projection operators are designed to be an isometry in the usual inner-product in L_2-space. Song (2006b) extends this isometric projection to conditional isometric projection by using conditional inner-product in place of the usual inner-product and finds that this class of projection operators can be useful to generate semiparametric tests that are asymptotically distribution free. Just as the isometric projection is called a martingale transform, Song (2006b) calls the conditional isometric projection a *conditional martingale transform*. See Song (2006b) for the details.

By applying the conditional martingale transform in Song (2006b) to our indicator functions, we can obtain the transform $\mathcal{K}$ as follows:

$$
\begin{aligned}
\gamma_y^{\mathcal{K}}(U,Y) &\triangleq \gamma_y(Y) - \mathbf{E}\left[\gamma_y(Y) h_{\bar{y}}(U,Y) | U = u\right]_{\bar{y}=Y} \text{ and} \\
\gamma_z^{\mathcal{K}}(U,Z) &\triangleq \gamma_z(Z) - \mathbf{E}\left[\gamma_z(Z) h_{\bar{z}}(U,Z) | U = u\right]_{\bar{z}=Z},
\end{aligned}
\tag{14}
$$

where

$$
h_{\bar{y}}(U,Y) = \frac{1\{Y \leq \bar{y}\}}{1 - F_{Y|U}(Y|U)} \text{ and } h_{\bar{z}}(U,Z) = \frac{1\{Z \leq \bar{z}\}}{1 - F_{Z|U}(Z|U)}. \tag{15}
$$

Observe that if $h_{\bar{y}}(U,Y)$ and $h_{\bar{z}}(U,Z)$ were taken to be a constant 1, the operator $\mathcal{K}$ is just an orthogonal projection onto the orthogonal complement of a constant function in $L_2(\mathbf{P})$. The factors $h_{\bar{y}}(U,Y)$ and $h_{\bar{z}}(U,Z)$ as defined in (15) render the transforms isometric.

As compared to the conditional martingale transform in Song (2006b), the transform in this paper corresponds to the case of function q_0 in Song (2006b) being equal to 1 and hence is simpler than typical conditional martingale transforms used to test semiparametric restrictions. This is primarily because the estimation error of $\hat{\lambda}$ is no longer needed to be taken care of in the transform, as it does not play a role in determining the asymptotic representation in Theorem 1 due to the sample quantile transform of the conditioning variable. In order to substantiate the properties of the conditional martingale transforms defined in (14), we introduce the following assumptions.

Assumption 5 : The distribution functions $F_{Y|U}(\cdot)$ and $F_{Z|U}(\cdot)$ of Y and of Z satisfy the following: (a) for some constants $C > 0$, and $\eta > 0$,

$$\begin{aligned}\sup_{u\in[0,1]} |F_{Y|U}(y_1|u) - F_{Y|U}(y_2|u)| &\leq C|y_1 - y_2|^{\eta} \text{ and} \\ \sup_{u\in[0,1]} |F_{Z|U}(z_1|u) - F_{Z|U}(z_2|u)| &\leq C|z_1 - z_2|^{\eta}.\end{aligned}$$

(b) For any $\eta > 0$, there exists $0 < c_0 < \eta$ such that for some $\varepsilon > 0$,

$$\begin{aligned}\inf_{(y,u)\in([0,1]\cap B_{Y,0})\times[0,1]} |1 - F_{Y|U}(y|u)| &> \varepsilon \text{ and} \\ \inf_{(z,u)\in([0,1]\cap B_{Z,0})\times[0,1]} |1 - F_{Z|U}(z|u)| &> \varepsilon,\end{aligned}$$

where $B_{Y,0} = \{y : y < 1 - c_0\}$ and $B_{Z,0} = \{z : z < 1 - c_0\}$.

Condition (i) is used to deal with the indicator functions $1\{Y \leq y\}$ and $1\{Z \leq z\}$ in the definition of conditional martingale transforms. The conditional martingale transform involves the inverse of $1 - F_{Y|U}(Z|U)$ and $1 - F_{Z|U}(Z|U)$. We restrict the index space for y and z to be the set $[0,1]\cap B_{Y,0}$ and $[0,1]\cap B_{Z,0}$ respectively. For notational brevity, we define S_0 to be the corresponding restricted index space for r :

$$S_0 \triangleq [0,1] \times ([0,1] \cap B_{Y,0}) \times ([0,1] \cap B_{Z,0}).$$

The following result shows that the transform $\mathcal{K}$ defined in (14) is a martingale transform that we have sought for. This result leads to the asymptotic properties of the process $\nu_n^{\mathcal{K}}(r)$ defined by

$$\nu_n^{\mathcal{K}}(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i)\gamma_z^{\mathcal{K}}(Z_i)(\gamma_y(Y_i) - \hat{g}_y^{\mathcal{K}}(\hat{U}_i)),$$

where $\hat{g}_y^{\mathcal{K}}(u)$ denotes the nonparametric estimator $\hat{g}_y(u)$ that uses $\gamma_y^{\mathcal{K}}$ in place of γ_y.

Theorem 2 : (i) For all y, y_1 and y_2 in $[0,1]$,

$$\begin{aligned}\langle \gamma_y^{\mathcal{K}}, 1\rangle_U &= 0 \text{ and } \langle \gamma_{y_1}^{\mathcal{K}}, \gamma_{y_2}^{\mathcal{K}}\rangle_U = \langle \gamma_{y_1}, \gamma_{y_2}\rangle_U \text{ almost everywhere and} \\ \langle \gamma_z^{\mathcal{K}}, 1\rangle_U &= 0 \text{ and } \langle \gamma_{z_1}^{\mathcal{K}}, \gamma_{z_2}^{\mathcal{K}}\rangle_U = \langle \gamma_{z_1}, \gamma_{z_2}\rangle_U \text{ almost everywhere.}\end{aligned}$$

(ii) Suppose Assumptions 1-5 hold.
(a) Under the null hypothesis,

$$\nu_n^{\mathcal{K}}(r) \rightsquigarrow \nu^{\mathcal{K}}(r), \tag{16}$$

where $\nu^{\mathcal{K}}$ is a Gaussian process whose covariance kernel is given by

$$c_3(r_1; r_2) \triangleq \int_{-\infty}^{u_1\wedge u_2} C_3(w_1, w_2; u)dF(u),$$

and the function $C_3(w_1, w_2; u)$ is defined by $C_3(w_1, w_2; u) \triangleq F_{Z|U}(z_1 \wedge z_2|u) \times F_{Y|U}(y_1 \wedge y_2|u)$.
(b) Under the fixed alternatives,

$$n^{-1/2}\nu_n^{\mathcal{K}}(r) \rightsquigarrow \mathbf{E}[\gamma_u(U)\langle\gamma_z^{\mathcal{K}}, \gamma_y^{\mathcal{K}}\rangle_U],$$

where $\gamma_z^{\mathcal{K}}$ and $\gamma_y^{\mathcal{K}}$ are defined in (14).

The first statement of the theorem is immediately adapted from Theorem 6.1 of Khmaladze and Koul (2004) to the "conditional inner product". This result tells us that the transform satisfies the orthogonality condition and the isometry condition. From the results of (i) and (ii), we can derive the asymptotic properties of tests that are constructed based on the transformed process $\nu_n^{\mathcal{K}}(r)$. The limiting Gaussian process $\nu^{\mathcal{K}}(r)$ is a time-transformed Brownian sheet whose kernel still depends on the unknown nonparametric functions $F_{Z|U}$ and $F_{Y|U}$. Following a similar idea in Khmaladze (1993), we can turn the process into a standard Brownian sheet by replacing $\gamma_z^{\mathcal{K}}(u, \bar{z})$ and $\gamma_y^{\mathcal{K}}(u, \bar{y})$ by

$$\tilde{\gamma}_z^{\mathcal{K}}(u, \bar{z}) \triangleq \frac{\gamma_z^{\mathcal{K}}(u, \bar{z})}{\sqrt{f_{Z|U}(\bar{z}|u)}} \text{ and } \gamma_y^{\mathcal{K}}(u, \bar{y}) \triangleq \frac{\gamma_y^{\mathcal{K}}(u, \bar{y})}{\sqrt{f_{Y|U}(\bar{y}|u)}},$$

where $f_{Z|U}$ and $f_{Y|U}$ denote conditional density functions. Then the weak limit $\nu^{\mathcal{K}}(r)$ in (16) has a covariance function

$$c(r_1, r_2) \triangleq (u_1 \wedge u_2)\,(z_1 \wedge z_2)\,(y_1 \wedge y_2),\ r_1, r_2 \in S_0,$$

which is that of a standard Brownian sheet.

Our test statistics are constructed based on the transformed empirical process $\tilde{v}_n^{\mathcal{K}}(r)$ defined by

$$\tilde{\nu}_n^{\mathcal{K}}(r) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i)\tilde{\gamma}_z^{\mathcal{K}}(\hat{U}_i, Z_i)\{\tilde{\gamma}_y^{\mathcal{K}}(\hat{U}_i, Y_i) - (\hat{\mathbf{P}}^U\tilde{\gamma}_y^{\mathcal{K}})(\hat{U}_i, Y_i)\}.$$

For instance, we can construct the transformed test statistics in the following way:

$$T_{KS}^{\mathcal{K}} = \sup_{r \in S_0} \left|\tilde{\nu}_n^{\mathcal{K}}(r)\right| \text{ and } T_{CM}^{\mathcal{K}} = \int_{S_0} \tilde{\nu}_n^{\mathcal{K}}(r)^2 dr. \tag{17}$$

From Theorem 2, we obtain the following null limiting distribution of the test statistics:

$$T_{KS}^{\mathcal{K}} \to_d \sup_{r \in S_0} |W(r)| \text{ and } T_{CM}^{\mathcal{K}} \to_d \int_{S_0} ||W(r)||^2 dr,$$

where W is a standard Brownian sheet on S_0. The resulting tests based on the statistics are asymptotically distribution free.

4.3 The Effect of Conditional Martingale Transforms on Asymptotic Power of Tests

Based on the result of Theorem 2, we can perform an analysis on the effect of conditional martingale transforms on the asymptotic power using approximate Bahadur ARE. In this subsection, we compare two tests based on the processes $\nu_{1n}(r)$ and $\nu_n^{\mathcal{K}}(r)$. At first glance, the comparison is not unambiguous because the martingale transform dilates the variance of the empirical process under the null hypothesis, but at the same time moves the shift term under the local alternatives farther from zero. The eventual comparison should take the trade-off between these two changes into account. To facilitate the analysis, we confine ourselves to those alternatives that make such comparison straightforward. More specifically, we consider the following specification of Y:

$$Y = \xi(Z, \varepsilon),$$

where $\xi(z, \bar{\varepsilon}) : [0,1]^2 \to [0,1]$ and ε is a random variable taking values in $[0,1]$ and is conditionally independent of Z given U. In the following we give an approximate Bahadur ARE of Kolmogorov-Smirnov tests based on the processes $\nu_{1n}(r)$ and $\nu_n^{\mathcal{K}}(r)$ and characterize a set of alternatives under which the conditional martingale transform improves the asymptotic power of the test.

Let us denote the conditional copula of Z and Y given U by

$$C_{Z,Y|U}(z, y|U) \triangleq P\left\{F_{Z|U}(Z|U) \le z, F_{Y|U}(Z|U) \le y|U\right\}.$$

The conditional copula $C_{Z,Y|U}(z, y|U)$ is the conditional joint distribution function of Z and Y after normalized by the conditional quantile transforms. It summarizes the conditional joint dependence of Z and Y given U. For introductory details about copulas, see Nelsen (1999) and for conditional copulas, see Patton (2006).

Corollary 3 : (i) *Suppose the conditions of* Theorems 1 *and* 2 *hold. Furthermore assume that* $\xi(z, \bar{\varepsilon})$ *is either strictly increasing in* z *for all* $\bar{\varepsilon}$ *in the support of* ε *or strictly decreasing in* z *for all* $\bar{\varepsilon}$ *in the support of* ε*. Then we have*

$$e_{1,\mathcal{K}}(\mathbf{P}_1) = 4 \sup_{z,y} \left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U), F_{Y|U}(y|U)|U) - F_{Z|U}(z|U) F_{Y|U}(y|U)\right]\right|$$

where $e_{1,\mathcal{K}}$ *represents the approximate Bahadur efficiency of the test based on* $\Gamma\hat{\nu}_n$ *relative to that based on* $\Gamma\nu_n^{\mathcal{K}}$ *and* Γ *is the Kolmogorov-Smirnov functional in* Corollary 1.
(ii) *Suppose further that* $F_{Z|U}(z|u)$ *is either increasing in* u *for all* $z \in [0,1]$ *or decreasing in* u *for all* $z \in [0,1]$*. Then we have*

$$e_{1,\mathcal{K}}(\mathbf{P}_1) \le 1. \tag{18}$$

When $F_{Z|U}(z|U)$ *and* $F_{Y|U}(y|U)$ *are uncorrelated for all* $(z, y) \in [0,1]^2$*, the inequality in* (18) *becomes equality if and only if in the case of* $\xi(z, \varepsilon)$ *strictly increasing in* z*,* $C_{Z,Y|U}(z, y|u) = z \wedge y$ *and in the case of* $\xi(z, \varepsilon)$ *strictly decreasing in* z*,* $C_{Z,Y|U}(z, y|u) = \max\{z + y - 1, 0\}$.

Corollary 3(i) provides a representation of the approximate Bahadur ARE in terms of the conditional copula $C_{Z,Y|U}$. The improvement of asymptotic power arises when

$$\left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)-F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\right|\leq\frac{1}{4}.$$

When the alternative is close to the null of conditional independence, the left-hand side term becomes close to zero, making it more likely that $e_{1,\mathcal{K}}(\mathbf{P}_1)$ is less than 1, in other words, the conditional martingale transform improves the asymptotic power of the test. Corollary 3(ii) tells us that when $F_{Z|U}(z|u)$ is monotone in u for all $z\in[0,1]$, the conditional martingale transform improves the asymptotic power in terms of approximate Bahadur ARE. In particular, this implies that improvement in power by the conditional martingale transform arises when Z and U are actually independent. When $F_{Z|U}(z|U)$ and $F_{Y|U}\left(y|U\right)$ are uncorrelated, the approximate Bahadur ARE becomes one if and only if the conditional copula achieves the Frechet-Hoeffding upper bound or lower bound depending on whether $\xi(z,\varepsilon)$ is strictly increasing in z or strictly decreasing in z. It is worth noting that either of these cases represents an alternative farthest from the null in the sense that under this alternative Z and Y are perfectly positively dependent given U.

4.4 Feasible Conditional Martingale Transforms

The conditional martingale transform proposed in the previous section is not feasible. It has unknown components to be estimated: conditional density functions $f_{Y|U},f_{Z|U}$ and conditional mean operators. We propose a feasible version of the martingale transform where the conditional mean functions are replaced by series estimators.

Let us define $\mathbf{P}^U$ to be conditional mean operator: $(\mathbf{P}^U\gamma)(u)\triangleq\mathbf{E}\left[\gamma(Z)|U=u\right]$ and denote $\hat{\mathbf{P}}^U$ to be its estimator. We define a feasible transform as follows. Let us define $p^K(u)\triangleq[p_0(u),\cdots,p_K(u)]'$,

$$\hat{P}\triangleq\begin{bmatrix}p^K(\hat{U}_1)'\\ \vdots\\ p^K(\hat{U}_n)'\end{bmatrix},\text{ and }\gamma_{y,n}\triangleq\begin{bmatrix}1\{c_0\leq Y_1\leq y\}\\ \vdots\\ 1\{c_0\leq Y_n\leq y\}\end{bmatrix}. \qquad (19)$$

Using $\hat{F}_{Y|U}(y|u)\triangleq p^K(u)'\hat{\pi}(y)$, where $\hat{\pi}(y)\triangleq[\hat{P}'\hat{P}]^{-1}\hat{P}'\gamma_{y,n}$, we define

$$\beta(y)\triangleq\begin{bmatrix}\hat{\gamma}_y(Y_1,\hat{U}_1)1\{Y_1\leq y\}[1-\hat{F}_{Y|U}(Y_1|\hat{U}_1)]^{-1}\\ \vdots\\ \hat{\gamma}_y(Y_n,\hat{U}_n)1\{Y_n\leq y\}[1-\hat{F}_{Y|U}(Y_n|\hat{U}_n)]^{-1}\end{bmatrix},$$

with $\hat{\gamma}_y(\bar{y},u)$ and $\hat{\gamma}_z(\bar{z},u)$ denoting the estimated versions of $\tilde{\gamma}_z(\bar{z},u)$ and $\tilde{\gamma}_z(\bar{z},u)$:

$$\hat{\gamma}_y(\bar{y},u)\triangleq\frac{\gamma_y(\bar{y})}{\sqrt{\hat{f}_{Y|U}(\bar{y}|u)}}\text{ and }\hat{\gamma}_z(\bar{z},u)\triangleq\frac{\gamma_z(\bar{z})}{\sqrt{\hat{f}_{Z|U}(\bar{z}|u)}}.$$

and $\hat{f}_{Y|U}(y|u)$ and $\hat{f}_{Z|U}(z|u)$ are conditional density estimators for $f_{Y|U}(y|u)$ and $f_{Z|U}(z|u)$.[6] Thus the feasible transformation is obtained by

$$(\hat{\gamma}_y^{\mathcal{K}})(u,\bar{y}) = \gamma_y(\bar{y}) - p^K(u)'\hat{\pi}^{\beta}(\bar{y}).$$

where $\hat{\pi}_K^{\beta}(y) \triangleq [\hat{P}'\hat{P}]^{-1}\hat{P}'\beta(y)$. We can similarly obtain $(\hat{\gamma}_z^{\mathcal{K}})(u,\bar{z})$ by exchanging the roles of Y_i and Z_i in the above construction.

An estimated version of the transformed process is defined by

$$\hat{\nu}_n^{\mathcal{K}}(r) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_u(\hat{U}_i)\hat{\rho}_w^{\mathcal{K}}(S_i) \tag{20}$$

where $\hat{\rho}_w^{\mathcal{K}} \triangleq [(\hat{\gamma}_z^{\mathcal{K}})(\hat{U}_i, Z_i)]\{(\hat{\gamma}_y^{\mathcal{K}})(\hat{U}_i, Y_i) - \hat{\mathbf{P}}^U(\hat{\gamma}_y^{\mathcal{K}})(\hat{U}_i, Y_i)\}$. Using the process $\hat{\nu}_n^{\mathcal{K}}(r)$, we can construct the transformed test statistics in the following way:

$$T_{KS}^{\mathcal{K}} \triangleq \sup_{r\in S_0}\left|\hat{\nu}_n^{\mathcal{K}}(r)\right| \text{ and } T_{CM}^{\mathcal{K}} \triangleq \int_{S_0}\hat{\nu}_n^{\mathcal{K}}(r)^2 dr.$$

In the context of testing semiparametric conditional moment restrictions, Song (2006b) delineated conditions under which the conditional martingale transforms are asymptotically valid. A similar study for the feasible transform in this section is in progress.

5 When Z_i is a Binary Variable

5.1 Semiparametric Empirical Process

In some applications, either Y or Z is binary. Specifically, we consider the case where Z is a binary variable taking zero or one, and the variables Y and X are continuous. This setting is similar to Angrist and Kuersteiner (2004), but the difference is that we do not assume a parametric specification of the conditional probability $\mathbf{P}\{Z = 1|X_i\}$. The application of conditional martingale transform in this case becomes even simpler. First, we do not need to transform the indicator function of Z_i. Second, the dimension of the index space for the empirical process in the test statistic reduces from 3 to 2.

First, observe that the null hypothesis becomes

$$\mathbf{E}[\gamma_u(U)1\{Z = z\}\{\gamma_y(Y) - \mathbf{E}[\gamma_y(Y)|U]\}] = 0$$

for all $(u, y, z) \in [0,1]^2 \times \{0,1\}$. From the binary character of Z, we observe that the null hypothesis

[6] Note that since U is uniformly distributed, we have $f_{Z|U}(z|u) = f_{Z,U}(z,u)$ where $f_{Z,U}(z,u)$ is a joint-density of Z and U. Hence we can use simply the estimator of $f_{Z,U}(z,u)$ instead of $f_{Z|U}(z|u)$

is equivalent to[7]

$$\mathbf{E}[\gamma_u(U)Z\{\gamma_y(Y) - \mathbf{E}[\gamma_y(Y)|U]\}] = 0.$$

This latter formulation of the null hypothesis is convenient because the index for the indicator of Z_i is fixed to be one and hence the corresponding empirical process has an index running over $[0,1]$ instead of $[0,1]^2 \times \{0,1\}$. The corresponding feasible and infeasible empirical processes in the test statistic before the martingale transform are given by

$$\begin{aligned}
\nu_n(u,y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_u(U_i)\left[Z_i\{\gamma_y(Y_i) - (\mathbf{P}^U\gamma_y)(U_i)\}\right] \text{ and} \\
\hat{\nu}_n(u,y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_u(\hat{U}_i)\left[Z_i\{\gamma_y(Y_i) - (\hat{\mathbf{P}}^U\gamma_y)(\hat{U}_i)\}\right], \ (u,y) \in [0,1]^2.
\end{aligned}$$

In this case of binary Z, we have only to transform the indicator function of γ_y. Using similar arguments in the decomposition of the processes $\hat{\nu}_{2n}(u,y)$ in the proof of Theorem 1, we obtain that

$$\hat{\nu}_n(u,y) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_u(U_i)(Z_i - \mathbf{E}\left[Z_i|U_i\right])\{\gamma_y(Y_i) - \mathbf{E}(\gamma_y(Y_i)|U_i)\} + o_P(1).$$

Therefore, we consider the following three martingale transformed processes:

$$\begin{aligned}
\tilde{\nu}^{\mathcal{K}}_{1na}(u,y) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_i\gamma_u(\hat{U}_i)(\tilde{\gamma}^{\mathcal{K}}_y)(\hat{U}_i, Y_i)}{\sqrt{\sigma^2_{2a}(\hat{U}_i)}}, \\
\tilde{\nu}^{\mathcal{K}}_{1nb}(u,y) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(1-Z_i)\gamma_u(\hat{U}_i)(\tilde{\gamma}^{\mathcal{K}}_y)(\hat{U}_i, Y_i)}{\sqrt{\sigma^2_{2b}(\hat{U}_i)}}, \text{ and} \\
\tilde{\nu}^{\mathcal{K}}_{2n}(u,y) &\triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_i\gamma_u(\hat{U}_i)\{(\tilde{\gamma}^{\mathcal{K}}_y)(\hat{U}_i, Y_i) - (\hat{\mathbf{P}}^U\tilde{\gamma}^{\mathcal{K}}_y)(\hat{U}_i, Y_i)\}}{\sqrt{\sigma^2_2(\hat{U}_i) - \sigma^4_2(\hat{U}_i)}},
\end{aligned}$$

where $\sigma^2_{2a}(U_i) \triangleq \mathbf{E}[Z_i|U_i]$ and $\sigma^2_{2b}(U_i) \triangleq \mathbf{E}[1 - Z_i|U_i]$. Then in view of the previous results, this process will weakly converge to a standard Brownian sheet on $[0,1]^2$ under the null hypothesis, i.e., a Gaussian process with covariance function:

$$c(u_1, y_1; u_2, y_2) \triangleq (u_1 \wedge u_2)(y_1 \wedge y_2).$$

Note that in the case of Z_i being binary, the Gaussian process has index over $[0,1]^2$, not over $[0,1]^3$.

[7] First note that

$$\begin{aligned}
\mathbf{E}[\gamma_u(U_i)1\{Z_i &= 1\}\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|U_i]\}] = 0 \text{ implies} \\
\mathbf{E}[\gamma_u(U_i)1\{Z_i &= 0\}\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|U_i]\}] = \mathbf{E}[\gamma_u(U_i)(1 - 1\{Z_i = 1\})\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|U_i]\}] \\
&= \mathbf{E}[\gamma_u(U_i)\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|U_i]\}] \\
&= \mathbf{E}[\gamma_u(U_i)\mathbf{E}\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|U_i]|U_i\}] = 0.
\end{aligned}$$

Now we obtain the equivalence by interchanging 1 and 0 in the indicator function of Z_i.

Also we have only to take the conditional martingale transform of γ_y. This is computationally convenient when we construct a test statistic using a functional of the transformed process.

In the case of a Kolmogorov-Smirnov type test, the test statistics we consider are the following.

$$\begin{aligned} T_{1A}^{\mathcal{K}} &\triangleq \sup_{(u,y)\in[0,1]^2} \left|\tilde{\nu}_{1n}^{\mathcal{K}}(u,y)\right|, \\ T_{1B}^{\mathcal{K}} &\triangleq \sup_{(u,y)\in[0,1]^2} \frac{\left|\tilde{\nu}_{1na}^{\mathcal{K}}(u,y)\right| + \left|\tilde{\nu}_{1nb}^{\mathcal{K}}(u,y)\right|}{\sqrt{2}}, \text{ and} \\ T_{2}^{\mathcal{K}} &\triangleq \sup_{(u,y)\in[0,1]^2} \left|\tilde{\nu}_{2n}^{\mathcal{K}}(u,y)\right|. \end{aligned}$$

We can construct similarly Cramér-von Mises type test statistics. The first and second test statistics $T_{1A}^{\mathcal{K}}$ and $T_{1B}^{\mathcal{K}}$ are computationally simpler than $T_2^{\mathcal{K}}$ because they do not involve the term $(\hat{\mathbf{P}}^U\tilde{\gamma}_y^{\mathcal{K}})(\hat{U}_i, Y_i)$ that is to be nonparametrically estimated. The three test statistics have the same limiting distribution under the null hypothesis and so we can use the identical critical values for these test statistics. As we shall see in the following subsection, the test statistic $T_2^{\mathcal{K}}$ has a better local asymptotic power property than the test statistics $T_{1A}^{\mathcal{K}}$ and $T_{1B}^{\mathcal{K}}$. The motivation for the second test statistic $T_{1B}^{\mathcal{K}}$ stems from our finding from simulation studies that the power properties of the test based on $T_{1A}^{\mathcal{K}}$ are disproportionate depending on whether alternatives represent positive or negative conditional dependence of Y and Z. This inconveniency is removed in the test statistic $T_{1B}^{\mathcal{K}}$. For more details, see a subsequent section devoted to simulation studies.

5.2 Analysis of Asymptotic Power Properties

We noted in a preceding section that the nonparametric estimation error does not affect the asymptotic properties of the martingale-transformed processes. However, in the case when Z is binary, the nonparametric estimation error can improve the power of the test even after the martingale transform. This is because we are taking a conditional martingale transform only on γ_y. Let us analyze the behavior of the two processes $\tilde{\nu}_{1n}^{\mathcal{K}}(u,y)$ and $\tilde{\nu}_{2n}^{\mathcal{K}}(u,y)$ under a fixed alternative. Based on Theorems 1 and 2, we can easily show that

$$\begin{aligned} n^{-1/2}\tilde{\nu}_{1n}^{\mathcal{K}}(u,y) &\to_p \mathbf{E}\left[\frac{Z_i\gamma_u(U_i)(\tilde{\gamma}_y^{\mathcal{K}})(U_i,Y_i)}{\sqrt{\sigma_2^2(U_i)}}\right] \text{ and} \\ n^{-1/2}\tilde{\nu}_{2n}^{\mathcal{K}}(u,y) &\to_p \mathbf{E}\left[\frac{\left(Z_i - \mathbf{E}\left[Z_i|U_i\right]\right)\gamma_u(U_i)(\tilde{\gamma}_y^{\mathcal{K}})(U_i,Y_i)}{\sqrt{\sigma_2^2(U_i) - \sigma_2^4(U_i)}}\right]. \end{aligned}$$

However, observe that the second local shift term is equal to

$$\mathbf{E}\left[\frac{Z_i\gamma_u(U_i)(\tilde{\gamma}_y^{\mathcal{K}})(U_i,Y_i)}{\sqrt{\sigma_2(U_i) - \sigma_2^2(U_i)}}\right]$$

because

$$\mathbf{E}\left[\frac{\mathbf{E}\left[Z_i|U_i\right]\gamma_u(U_i)(\tilde{\gamma}_y^{\mathcal{K}})(U_i,Y_i)}{\sqrt{\sigma_2(U_i)-\sigma_2^2(U_i)}}\right]=\mathbf{E}\left[\frac{\mathbf{E}\left[Z_i|U_i\right]\gamma_u(U_i)\mathbf{E}\left[(\tilde{\gamma}_y^{\mathcal{K}})(U_i,Y_i)|U_i\right]}{\sqrt{\sigma_2(U_i)-\sigma_2^2(U_i)}}\right]=0.$$

by the conditional orthogonality condition for $\tilde{\gamma}_y^{\mathcal{K}}$. Since the denominator of the shift term for $\tilde{\nu}_{1n}^{\mathcal{K}}(u,y)$ dominates that of the shift term for $\tilde{\nu}_{2n}^{\mathcal{K}}(u,y)$, the test based on the process $\tilde{\nu}_{2n}^{\mathcal{K}}(u,y)$ has a better local asymptotic power than that based on $n^{-1/2}\tilde{\nu}_{1n}^{\mathcal{K}}(u,y)$. This also shows the interesting phenomena that the test has a better power when we use the estimator $(\hat{\mathbf{P}}^U\tilde{\gamma}_y^{\mathcal{K}})(U_i)$ for $(\mathbf{P}^U\tilde{\gamma}_y^{\mathcal{K}})(U_i)$ instead of the true one. As before, we can formalize this phenomenon in terms of approximate Bahadur ARE. Let $e_{1,0}(\mathbf{P}_1)$ be the approximate Bahadur ARE of the Kolmogorov-Smirnov test base on $\tilde{\nu}_{1n}^{\mathcal{K}}(u,y)$ and that based on $\tilde{\nu}_{2n}^{\mathcal{K}}(u,y)$.

Corollary 4 : Suppose that the conditions of Theorem 1 hold except for that of Z. Then

$$e_{1,0}(\mathbf{P}_1)=\frac{1}{4}.$$

This result is in contrast to that in the case with continuous Z and Y where no improvement of the power was introduced by the nonparametric estimation error. The result of the improved power property in Corollary 4 is demonstrated in the simulation studies.

5.3 Testing Procedure in Summary

In this subsection, we summarize the testing procedure for the case where Z is binary. First we quantile-transform the data $(Y_i,\hat{\lambda}(X_i))_{i=1}^n$ to obtain $(Y_i,\hat{U}_i)_{i=1}^n$. (We use the same notation Y_i here.) Then construct the test statistic in the following way. First estimate the conditional density $f_{Y|U}(y|u)$ using the data $(Y_i,\hat{U}_i)_{i=1}^n$ and the conditional distribution function $F_{Y|U}(y|u)$ of Y given U. Let us denote the estimated joint density by $\hat{f}_{Y,U}(y,u)$ and the conditional distribution function by $\hat{F}_{Y|U}(y|u)$. Then we compute

$$\alpha_i(y,\bar{y})=\frac{1\{Y_i\leq\min(y,\bar{y})\}}{\sqrt{\hat{f}_{Y,U}(Y_i,\hat{U}_i)}\left(1-\hat{F}_{Y|U}(Y_i'|\hat{U}_i)\right)}.$$

Note that our use of $\hat{f}_{Y,U}(Y_i,\hat{U}_i)$ instead of $\hat{f}_{Y|U}(Y_i|\hat{U}_i)$ is based on the fact that U_i is uniformly distributed. Choose a basis function $p^K=(p_1,p_2,\cdots,p_K)'$. Let us define $p^K(u)=[p_0(u),\cdots,p_K(u)]'$ and define $\alpha(y,\bar{y})=[\alpha_1(y,\bar{y}),\cdots,\alpha_n(y,\bar{y})]'$. Then we define a $K\times 1$ vector $\pi_n(y,\bar{y})$ by

$$\pi_n(y,\bar{y})=[\hat{P}'\hat{P}]^{-1}\hat{P}'\alpha(y,\bar{y}).$$

The choice of K can be done using the procedure of cross-validations. We investigate the sensitivity of the performance of tests to the choice of K in detail in the next section. Let $\hat{\mathbf{P}}(Z_i=1|\hat{U}_i)$ denote a nonparametric estimator of $\mathbf{P}\{Z_i = 1|U_i\}$ and let

$$B_i(u,y) = 1\{\hat{U}_i \le u\} \left(\frac{1\{Y_i \le y\}}{\sqrt{\hat{f}_{Y,U}(Y_i, \hat{U}_i)}} - p^K(\hat{U}_i)'\pi_n(y, Y_i) \right).$$

Then define

$$\nu^{\mathcal{K}}_{1na}(u,y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i B_i(u,y)}{\sqrt{\hat{\mathbf{P}}(Z_i = 1|\hat{U}_i)}} \text{ and } \nu^{\mathcal{K}}_{1nb}(u,y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(1 - Z_i) B_i(u,y)}{\sqrt{1 - \hat{\mathbf{P}}(Z_i = 1|\hat{U}_i)}}.$$

A Kolmogorov-Smirnov type test statistic is given by

$$T^{\mathcal{K}}_{1B} = \sup_{(u,y)\in[0,1]^2} \frac{\left|\nu^{\mathcal{K}}_{1na}(u,y)\right| + \left|\nu^{\mathcal{K}}_{1nb}(u,y)\right|}{\sqrt{2}}$$

In the actual computation of the test statistic, we use grid points. We have found from our simulation studies that we discuss in the next section, the equally spaced grid points of 10 by 10 in a unit square appear to work well. The limiting distribution under the null is the Kolmogorov-Smirnov functional of a two-parameter Brownian sheet. Its critical values are replicated from Table 3 of Khmaladze and Koul (2004) in the following table.

significance level	0.5	0.25	0.20	0.10	0.05	0.025	0.01
critical values	1.46	1.81	1.91	2.21	2.46	2.70	3.03

The critical values are based on the computation of Brownrigg and can be found at his website: http://www.mcs.vuw.ac.nz/~ray/Brownian. A Matlab program to compute p-values based on the interpolation of the critical values is available from the author. Therefore, one reject the null hypothesis of conditional independence at the level of 0.05 if $T_n > 2.46$.[8]

6 Simulation Studies

6.1 Conditional Martingale Transform and Wild Bootstrap

In this section, we present and discuss the results from simulation studies that compare the approach of conditional martingale transform and the approach of bootstrap. We consider testing conditional

[8] In the case of continuous variable Z_i, we obtain $\pi_n(z, \bar{z})$ and $\hat{f}_{Z|U}(y|u)$ in a similar manner and construct the test statistic:

$$\sup_{(u,y,z)\in[0,1]^3} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1\{\hat{U}_i \le u\} \left(\frac{1\{Z_i \le y\}}{\sqrt{\hat{f}_{Z|U}(Z_i|\hat{U}_i)}} - p^K(\hat{U}_i)'\pi_n(z, Z_i) \right) \left(\frac{1\{Y_i \le y\}}{\sqrt{\hat{f}_{Y|U}(Y_i|\hat{U}_i)}} - p^K(\hat{U}_i)'\pi_n(y, Y_i) \right) \right|.$$

In this case, we need to use different asymptotic critical values that are based on the Kolmogorov-Smirnov functional of a three-parameter standard Brownian sheet.

independence of Y_i and Z_i given X_i. Each variable is set to be real-valued. The variable Z_i is binary taking zero or one depending on the following rule:

$$Z_i = 1\{X_i + \eta_i > 0\}.$$

The variable Y_i is determined by the data generating process

$$Y_i = \beta X_i + \kappa Z_i + \varepsilon_i.$$

The parameter β is set to be 0.5. The variable X_i is drawn from the uniform distribution on $[-1, 1]$ and the errors η_i and ε_i are independently drawn from $N(0, 1)$.

Note that $\kappa = 0$ corresponds to the null hypothesis of conditional independence:

$$\mathbf{E}[\gamma_u(X_i)\rho_y(S_i)] = 0, \text{ for all } (x, y) \in [0, 1]^{1+d_Y}$$

where

$$\rho_y(S_i) = 1\{Z_i = 1\}\{\gamma_y(Y_i) - \mathbf{E}[\gamma_y(Y_i)|X_i]\}.$$

The magnitude of β controls the strength of the relation between Y_i and X_i.

We consider two tests, one based on the conditional martingale transform approach and the other based on the bootstrap approach. The bootstrap procedure is as follows:

(1) Take $(\hat{\varepsilon}_{y,i})_{i=1}^n$ defined by $\hat{\varepsilon}_{y,i} = \gamma_y(Y_i) - \hat{\mathbf{E}}(\gamma_y(Y_i)|\hat{U}_i)$.

(2) Then we take a wild bootstrap sample $(\varepsilon^*_{y,i})_{i=1}^n = (\omega_i \hat{\varepsilon}_{y,i})_{i=1}^n$ where $\{\omega_i\}$ is an i.i.d. sequence of random variables that are independent of $\{Y_i, Z_i, \hat{U}_i\}$ such that $\mathbf{E}(\omega_i) = 0$, $\mathbf{E}(\omega_i^2) = 1$.

(3) Construct $\gamma^*_{y,i} = \hat{\mathbf{E}}(\gamma_y(Y_i)|\hat{U}_i) + \varepsilon^*_{y,i}$.

(4) Using the bootstrap data $(\gamma^*_{y,i}, \hat{U}_i)_{i=1}^n$, we estimate the bootstrap conditional mean function $\hat{\mathbf{E}}(\gamma^*_{y,i}|\hat{U}_i)$.

(5) Based on this, we define

$$\rho^*_{y,i} = Z_i \left[\gamma^*_{y,i} - \hat{\mathbf{E}}(\gamma^*_{y,i}|\hat{U}_i)\right].$$

(6) We obtain the bootstrap empirical process

$$\nu^*_n(u, w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_u(\hat{U}_i)\rho^*_{y,i}.$$

The bootstrap procedure is the same as that considered in Delgado and Gonzáles Manteiga (2001), p.1490, except that they weight the process by $f(U_i)$ where f is the density function of U_i. Based upon the bootstrap process, we can construct the bootstrap test statistics:

$$T_{KS}^B = \sup_{s \in [0,1]^{d_Y+1}} |\nu_n^*(u,w)| \text{ and } T_{CM}^B = \int \int \nu_n^*(u,w)^2 dudw.$$

Nonparametric estimations are done using series-estimation with Legendre polynomials. We rescale the variables X_i and Y_i so that they have a unit interval support. The rescaling was done by taking their (empirical) quantile transformations. We also take into account unknown conditional heteroskedasticity by normalizing the semiparametric empirical process by $\hat{\sigma}(\hat{U}_i)^{-1}$. The test statistics are formed by using a Kolmogorov-Smirnov type. The number of Monte Carlo iterations and the number of bootstrap Monte Carlo iterations are set to be 1000. The sample size is 300.

First, consider the size of the tests whose results are in Table 1. The nominal size is set at 0.05. Order 1 represents the number of the terms included in the nonparametric estimation of the bivariate density function $f_{Y,U}(y,u)$ and Order 2 represents the number of the terms included in the nonparametric estimation of the conditional expectations given U_i. The martingale-transform based tests appear more sensitive to the choice of orders in the series estimation than the bootstrap-based test. This is not surprising because the martingale transform approach involves more incidence of nonparametric estimations than the bootstrap approach. The empirical size seems reasonable overall.

[INSERT TABLE 1]

Tables 2 and 3 contain the results for power of the tests. The deviation from the null hypothesis is controlled by the magnitude of κ. Results in Table 2 are under the alternatives of positive dependence between Y_i and Z_i and those in Table 3 are under the alternatives of negative dependence. First, the power of the test based on $T_{1A}^{\mathcal{K}}$ is shown to be disproportionate between alternatives of positive dependence (with κ positive) and negative dependence (with κ negative). This disproportionate behavior disappears in the case of its "symmetrized" version $T_{1B}^{\mathcal{K}}$. Second, the power of tests based on $T_2^{\mathcal{K}}$ is better than that based on $T_{1A}^{\mathcal{K}}$ or $T_{1B}^{\mathcal{K}}$. Recall that the test statistic $T_2^{\mathcal{K}}$ involves the nonparametric estimator for $(\mathbf{P}^U \tilde{\gamma}_y^{\mathcal{K}})(U_i, Y_i) = 0$ whereas $T_{1A}^{\mathcal{K}}$ and $T_{1B}^{\mathcal{K}}$ simply uses zero in its place. The result shows the interesting phenomenon that the use of the nonparametric estimator improves the power of the test even when we know that it is equal to zero in population. The results from the bootstrap works well and the performance is stable over the choice of order in the series and over the alternatives under which the test is performed.

[INSERT TABLES 2 AND 3]

6.2 The Effect of Nonparametric Estimation of $F(y|u)$

We consider testing the conditional independence of Y and Z given U where Z is a binary variable and (Y,U) has uniform marginals over $[0,1]^2$. We generate

$$Y = \alpha((1-\kappa)U + \kappa(Z\varepsilon + (1-Z)U)) + (1-\alpha)\varepsilon,$$

where ε follows a standard normal distribution and independent of U and Z, and Z is generated by

$$Z = 1\{U - \eta > 0\}$$

where η is a uniform variate $[0,1]$. Hence $P\{Z=1\} = 1/2$. When $\kappa = 0$, the null hypothesis of conditional independence of Y and Z given U holds. As κ gets bigger the role of Z in directly contributing to Y without through U becomes more significant. In this case, we can explicitly compute the conditional distribution function $F_{Y|U}(y|u)$ as follows

$$\begin{aligned} F_{Y|U}(y|u) &= \mathbf{P}\{Y \le y|U=u, Z=1\}\mathbf{P}\{Z=1\} + \mathbf{P}\{Y \le y|U=u, Z=0\}\mathbf{P}\{Z=0\} \\ &= \frac{1}{2}\left(\Phi\left(\frac{y-\alpha(1-\kappa)u}{1-\alpha(1-\kappa)}\right) + \Phi\left(\frac{y-\alpha u}{1-\alpha}\right)\right). \end{aligned}$$

Note that $\mathbf{E}[Z|U] = U$. Hence $\sigma_2(U) = \sqrt{U - U^2}$.

We noted before that the estimation error in $\hat{F}_{Y|U}(y|U_i)$ has an effect of improving the power of the test. In order to investigate this in finite samples, we consider the two test statistics

$$\begin{aligned} T_n^{IINF} &= \sup_{(u,y)\in[0,1]^2}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1\{U_i \le u\}Z_i}{\sqrt{\sigma_2^2(U_i)}}\left(1\{Y_i \le y\} - F_{Y|U}(y|U_i)\right)\right| \text{ and} \\ T_n^{FEAS} &= \sup_{(u,y)\in[0,1]^2}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1\{U_i \le u\}Z_i}{\sqrt{\sigma_2^2(\hat{U}_i) - \sigma_2^4(\hat{U}_i)}}\left(1\{Y_i \le y\} - \hat{F}_{Y|U}(y|U_i)\right)\right|. \end{aligned}$$

Note that $F_{Y|U}(y|u)$ and $\sigma_2(U)$ are completely known in this case.

We take the sample size equal to $n = 3000$ and set the Monte-Carlo simulation number to equal to 1000. We consider the estimated density of the test statistics both under the null ($\kappa = 0$) and under the alternative ($\kappa = 0.5$). Here we use the quantile transform of the data, use the series estimation with Legendre polynomials with the number of terms equal to 5 when univariate and 5^2 when bivariate. The result is shown in Figure 1. Under the null, the behavior of the two test statistics is more or less similar as consistent with the asymptotic theory. However, under the alternative, the behavior becomes dramatically different.[9] The shift is more prominent in the case of T_n^{FEAS} than in the case of T_n^{IINF}. This was expected from our analysis of power from the previous section, and reflects our theoretical observation that the use of $\hat{F}_{Y|U}(y|U_i)$ instead

[9] The different dispersions in the distribution of the test statistics between the null and the alternative appear to mainly stem from the inclusion of $\hat{\sigma}_2(U_i)$ in the test statistics. More specifically, under the null hypothesis, the

of $\hat{F}_{Y|U}(y|U_i)$ improves the local power properties of the tests. This is also observed in terms of empirical processes inside T_n^{IINF} and T_n^{FEAS} as shown in Figure 2.

[INSERT FIGURES 1 AND 2]

7 Conclusion

A test of conditional independence has been proposed. The test is asymptotically unbiased against Pitman local alternatives and yet has critical values that can be tabulated from the limiting distribution theory. As in Song (2006b), the main idea is to employ conditional martingale transforms, but the manner that we apply the transforms in testing conditional independence is different. We apply the transform to the indicator functions of Z_i and Y_i with the conditioning on $\hat{U}_i$.

There are two possible extensions: first, the analysis of the asymptotic power function, and second, the accommodation of weakly dependent data. The analysis of the asymptotic power function will illuminate the effect of the martingale transform on the power of the tests. The i.i.d. assumption in testing conditional independence is restrictive considering that testing the presence of causal relations often presupposes the context of time series. Extending the results in the paper to time series appears nontrivial because the methods of this paper heavily draws on the results of empirical process theory that are based on independent series. However, there have been numerous efforts in the literature to extend the results of empirical processes to time series (e.g., Andrews (1991), Arcones and Yu (1994), and Nishyama (2000).) It seems that the results in this paper can be extended to time series context based on those results.

---

covariance kernel of the process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1\{U_i\leq u\}Z_i}{\sigma_2(U_i)}\left(1\{Y_i\leq y\}-F_{Y|U}(y|U_i)\right)$$

is equal to

$$\int_0^{u_1\wedge u_2}\left\{F_{Y|U}(y_1\wedge y_2|u)-F_{Y|U}(y_1|u)F_{Y|U}(y_2|u)\right\}du$$

whereas under the Pitman local alternatives, the process is asymptotically equivalent to a shifted version of the following process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1\{U_i\leq u\}}{\sigma_2(U_i)}\left(Z_i1\{Y_i\leq y\}-\mathbf{E}\left[Z_i1\{Y_i\leq y\}|U_i\right]\right).$$

The covariance kernel of this process is equal to

$$\begin{aligned}
&\int_0^{u_1\wedge u_2}\Phi(u)^{-2}\left\{\mathbf{E}\left[Z_i1\{Y_i\leq y_1\wedge y_2\}|U_i=u\right]-\mathbf{E}\left[Z_i1\{Y_i\leq y_1\}|U_i=u\right]\mathbf{E}\left[Z_i1\{Y_i\leq y_2\}|U_i=u\right]\right\}du\\
=&\int_0^{u_1\wedge u_2}\Phi(u)^{-2}\left\{\Phi(u)F_{Y|U}(y_1\wedge y_2|u)-\Phi(u)^2F_{Y|U}(y_1|u)F_{Y|U}(y_2|u)\right\}du\\
=&\int_0^{u_1\wedge u_2}\left\{\Phi^{-1}(u)F_{Y|U}(y_1\wedge y_2|u)-F_{Y|U}(y_1|u)F_{Y|U}(y_2|u)\right\}
\end{aligned}$$

since $\sigma_2(U)=P(Z=1|U)=\Phi(U)$. Since $\Phi(U)\in[0,1]$, the covariance kernel above dominates that of the process under the null hypothesis.

8 Appendix: Mathematical Proofs

8.1 Main Results

We first prove Theorem 1. The following lemma is useful for this end. Define

$$\gamma_{u,n,\lambda}(x) \triangleq 1\left\{\frac{1}{n}\sum_{i=1}^{n} 1\left\{\lambda(X_i) \leq \lambda(x)\right\} \leq u\right\}.$$

The function $\gamma_{u,n,\lambda}(\cdot)$ is indexed by $(u,\lambda) \in [0,1] \times \Lambda_n$. The indicator function contains the nonparametric function which is an empirical distribution function. Note that we cannot directly apply the framework of Chen, Linton and van Keilegom (2003) to analyze the class of functions that the realizations of $\gamma_{u,n,\lambda}(x)$ falls into. This is because when the indicator function contains a nonparametric function, the procedure of Chen, Linton, and van Keilegom (2003) requires the entropy condition for the nonparametric functions with respect to the sup norm. This entropy condition with respect to the sup norm does not help because L_p uniform continuity fails with respect to the sup norm. Hence we establish the following result for our purpose.

Lemma A1 : *There exists a sequence of classes $\mathcal{G}_n$ of measurable functions such that*

$$\mathbf{P}\left\{\gamma_{u,n,\lambda} \in \mathcal{G}_n \text{ for all } (u,\lambda) \in [0,1] \times \Lambda_n\right\} = 1,$$

and for any $r \geq 1$,

$$\log N_{[]}(C_2\varepsilon, \mathcal{G}_n, ||\cdot||_r) \leq \log N(\varepsilon^r/2, \Lambda_n, ||\cdot||_\infty) + \frac{C_1}{\varepsilon},$$

where C_1 and C_2 are positive constants depending only on r.

Proof of Lemma A1 : For a fixed sequence $\{x_i\}_{i=1}^n$ with $\{x_i\} \in \mathbf{R}^{nd_X}$, we define a function:

$$g_{n,\lambda,u}(\bar{\lambda}) \triangleq 1\left\{\frac{1}{n}\sum_{i=1}^{n} 1\left\{\lambda(x_i) \leq \bar{\lambda}\right\} \leq u\right\},$$

where $g_{n,\lambda,u}(\cdot)$ depends on the chosen sequence $\{x_i\}_{i=1}^n$. We introduce classes of functions as follows:

$$\begin{aligned} \mathcal{G}_n' &\triangleq \left\{g_{n,\lambda,u} : (\lambda, \{x_i\}_{i=1}^n, u) \in \Lambda_n \times \mathbf{R}^{nd_X} \times [0,1]\right\} \text{ and} \\ \mathcal{G}_n &\triangleq \left\{g \circ \lambda : (g,\lambda) \in \mathcal{G}_n' \times \Lambda_n\right\}. \end{aligned}$$

Then we have $\gamma_{u,n,\lambda} \in \mathcal{G}_n$. Note that $g_{n,\lambda,u}(\bar{\lambda})$ is decreasing in $\bar{\lambda}$. Hence by Theorem 2.7.5 in van der Vaart and Wellner (1996), for any $r \geq 1$

$$\log N_{[]}(\varepsilon, \mathcal{G}_n', ||\cdot||_{Q,r}) \leq \frac{C_1}{\varepsilon}, \tag{21}$$

for any probability measure Q and for some constant $C_1 > 0$.

Without loss of generality, assume that $\lambda(X) \in [0,1]$. We choose $\{\lambda_1, \cdots, \lambda_{N_2}\}$ such that for any $\lambda \in \Lambda_n$, there exists j with $\|\lambda_j - \lambda\|_\infty < \varepsilon^r/2$. For each j, we define $\tilde{\lambda}_j(x)$ as follows.

$$\tilde{\lambda}_j(x) = m\varepsilon^r/2 \text{ when } \lambda_j(x) \in [m\varepsilon^r/2, (m+1)\varepsilon^r/2) \text{ for some } m \in \{0, 1, 2, \cdots, \lfloor 2/\varepsilon^r \rfloor\},$$

where $\lfloor z \rfloor$ denotes the greatest integer that does not exceed z. Note that the range of $\tilde{\lambda}_j$ is finite and

$$\begin{aligned}
||\lambda - \tilde{\lambda}_j||_\infty &= \sup_x \sum_{k=0}^{\lfloor 2/\varepsilon^r \rfloor} |\lambda(x) - k\varepsilon^r/2| \, 1\{\lambda_j(x) \in [k\varepsilon^r/2, (k+1)\varepsilon^r/2)\} \\
&\leq \sup_x \sum_{k=0}^{\lfloor 2/\varepsilon \rfloor} (|\lambda(x) - \lambda_j(x)| + |\lambda_j(x) - k\varepsilon^r/2|) \, 1\{\lambda_j(x) \in [k\varepsilon^r/2, (k+1)\varepsilon^r/2)\} \leq \varepsilon^r.
\end{aligned}$$

From (21), we have $\log N(\varepsilon, \mathcal{G}_n', ||\cdot||_{Q,r}) \leq C_1/\varepsilon$. We choose $\{g_1, \cdots, g_{N_1}\}$ such that for any $g \in \mathcal{G}_n'$, there exists j such that $\int |g(u) - g_j(u)|^r du < \varepsilon^r$. Then we have for any $\lambda \in \Lambda_n$,

$$\mathbf{E}\,|g(\lambda(X)) - g_j(\lambda(X))|^r = \int \left|g(\bar{\lambda}) - g_j(\bar{\lambda})\right|^r f_\lambda(\bar{\lambda}) du \leq \sup_\lambda \sup_{\bar{\lambda}} f(\bar{\lambda})\varepsilon^r,$$

where $f_\lambda(\cdot)$ denotes the density of $\lambda(X)$.

Now, take any $h \in \mathcal{G}_n$ such that $h \triangleq g \circ \lambda$ and choose (g_{j_1}, λ_{j_2}) from $\{g_1, \cdots, g_{N_1}\}$ and $\{\lambda_1, \cdots, \lambda_{N_2}\}$ such that

$$\int |g(u) - g_{j_1}(u)|^r du < \varepsilon^r \text{ and } \|\lambda_{j_2} - \lambda\|_\infty < \varepsilon^r/2.$$

Then consider

$$\begin{aligned}
\left\|h - g_{j_1} \circ \tilde{\lambda}_{j_2}\right\|_r &\leq \|g \circ \lambda - g_{j_1} \circ \lambda\|_r + \left\|g_{j_1} \circ \lambda - g_{j_1} \circ \tilde{\lambda}_{j_2}\right\|_r \\
&\leq \left(\sup_\lambda \sup_{\bar{\lambda}} f(\bar{\lambda})\right)^{1/r} \varepsilon + \left(\mathbf{E}\left|(g_{j_1} \circ \lambda)(X) - (g_{j_1} \circ \tilde{\lambda}_{j_2})(X)\right|^r\right)^{1/r}.
\end{aligned}$$

Note that the absolute value in the expectation of the second term is bounded by

$$1\left\{\frac{1}{n}\sum_{i=1}^n 1\left\{\lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) - \varepsilon^r\right\} \leq u_{j_1}\right\} - 1\left\{\frac{1}{n}\sum_{i=1}^n 1\left\{\lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) + \varepsilon^r\right\} \leq u_{j_1}\right\}$$

or by

$$
\begin{aligned}
& 1\left\{\frac{1}{n}\sum_{i=1}^{n}1\left\{\lambda_{j_1}(x_i)\leq\tilde{\lambda}_{j_2}(X)-\varepsilon^r\right\}\leq u_{j_1}<\frac{1}{n}\sum_{i=1}^{n}1\left\{\lambda_{j_1}(x_i)\leq\tilde{\lambda}_{j_2}(X)+\varepsilon^r\right\}\right\} \\
= & \sum_{m=0}^{\lfloor 2/\varepsilon\rfloor}1\left\{\frac{1}{n}\sum_{i=1}^{n}1\left\{\lambda_{j_1}(x_i)\leq\frac{m\varepsilon^r}{2}-\varepsilon^r\right\}\leq u_{j_1}<\frac{1}{n}\sum_{i=1}^{n}1\left\{\lambda_{j_1}(x_i)\leq\frac{m\varepsilon^r}{2}+\varepsilon^r\right\}\right\} \\
& \times 1\left\{\frac{m\varepsilon^r}{2}\leq\lambda_{j_2}(X)<\frac{(m+1)\varepsilon^r}{2}\right\}.
\end{aligned}
$$

Each summand in the last sum, as a range of a fixed value u_{j_1}, overlaps with precisely three other adjacent ones, and hence is bounded by

$$1\left\{\lambda_{j_2}(X)\in[k\varepsilon^r/2,(k+3)\varepsilon^r/2)\right\}$$

for some $k\in\{0,1,2,\cdots,\lfloor 2/\varepsilon^r\rfloor\}$. By taking expectation, this last indicator function is bounded by

$$
\begin{aligned}
\mathbf{P}\left\{\lambda_{j_2}(X)\in[m\varepsilon^r/2,(m+2)\varepsilon^r/2)\right\} &= F_{\lambda_{j_2}}((m+2)\varepsilon^r/2)-F_{\lambda_{j_2}}(m\varepsilon^r/2) \\
&\leq \sup_{\bar{\lambda}} f_{\lambda_{j_2}}(\bar{\lambda})\varepsilon^r,
\end{aligned}
$$

where $F_{\lambda_{j_2}}$ is the distribution function of $\lambda_{j_2}(X)$. Therefore,

$$\left\|h-g_{j_1}\circ\tilde{\lambda}_{j_2}\right\|_r\leq 2\left(\sup_{\lambda}\sup_{\bar{\lambda}}f_{\lambda}(\bar{\lambda})\right)^{1/r}\varepsilon.$$

This leads to the result that

$$
\begin{aligned}
\log N_{[]}\left(2\left(\sup_{\lambda}\sup_{\bar{\lambda}}f_{\lambda}(\bar{\lambda})\right)^{1/r}\varepsilon,\mathcal{G}_n,||\cdot||_r\right) &\leq \log N(\varepsilon^r/2,\Lambda_n,||\cdot||_\infty)+\log N_{[]}(\varepsilon,\mathcal{G}_n',||\cdot||_r) \\
&\leq \log N(\varepsilon^r/2,\Lambda_n,||\cdot||_\infty)+\frac{C}{\varepsilon}.
\end{aligned}
$$

■

We define two notations that we use throughout the remaining proofs. First, $\mathbb{P}$ denotes the operation of sample mean: $\mathbb{P}f\triangleq\frac{1}{n}\sum_{i=1}^{n}f(S_i)$ and second, $^{\perp}$ denotes the conditional mean deviation: $\gamma_y^{\perp}(Y,U)\triangleq\gamma_y(Y)-\mathbf{E}\left[\gamma_y(Y)|U=u\right]$.

Proof of Theorem 1 : Recall that

$$\nu_n(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma_u(U_i)\gamma_z(Z_i)\left\{\gamma_y(Z_i)-\mathbf{E}(\gamma_y(Z_i)|U_i)\right\}=\sqrt{n}\mathbb{P}\gamma_u\gamma_z(\gamma_y^{\perp}).$$

Then we analyze the asymptotic behavior of the empirical process above under the null hypothesis and under the alternatives. Choose a sequence $\tilde{\lambda}$ within the shrinking neighborhood $B(\lambda_0,\delta_n)\subset\Lambda$

of λ_0 with diameter $\delta_n = o(1)$ in terms of $||\cdot||_2$. Consider

$$(\hat{\mathbf{P}}^U_{\tilde{\lambda}}\gamma)(u) \triangleq p^K(u)'\hat{\pi}_{\gamma,\tilde{\lambda}},$$

where $\hat{\pi}_{\gamma,\tilde{\lambda}} \triangleq [P'_{\tilde{\lambda}}P_{\tilde{\lambda}}]^{-1}P'_{\tilde{\lambda}}y_{\gamma,n}$ with $y_{\gamma,n}$ and $P_{\tilde{\lambda}}$ being defined by

$$y_{\gamma,n} \triangleq \begin{bmatrix} \gamma(Y_1) \\ \vdots \\ \gamma(Y_n) \end{bmatrix} \text{ and } P_{\tilde{\lambda}} \triangleq \begin{bmatrix} p^K(U_{n,\tilde{\lambda}}(X_1))' \\ \vdots \\ p^K(U_{n,\tilde{\lambda}}(X_n))' \end{bmatrix}.$$

Note that $\hat{\mathbf{P}}^U_{\tilde{\lambda}}$ is the same as $\hat{\mathbf{P}}^U$ except that $\hat{\lambda}$ is replaced by $\tilde{\lambda}$.

We define a process $\hat{\nu}_n(g, r, \lambda)$ indexed by $(g, r, \lambda) \in \mathcal{G}_n \times [0,1]^{d_r} \times \Lambda_n$ as

$$\hat{\nu}_n(g, r, \lambda) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i)\gamma_z(Z_i)\left[\gamma_y(Y_i) - p^K(U_{n,\tilde{\lambda}}(X_i))'\hat{\pi}_{\gamma,\tilde{\lambda}}\right].$$

Then by Lemma 1U of Escanciano and Song (2006), we obtain the result that $\hat{\nu}_n(g, r, \lambda)$ is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [g(X_i)\gamma_z(Z_i) - \mathbf{E}\left[g(X_i)\gamma_z(Z_i)|U_i\right]\left[\gamma_y(Y_i) - \mathbf{P}\left\{Y_i \leq y|U_i\right\}\right] + o_P(1)$$

where $o_P(1)$ is uniform over $(g, r, \lambda) \in \mathcal{G}_n \times [0,1]^{d_r} \times \Lambda_n$. Hence the above holds for any arbitrary sequence $g_n \in \mathcal{G}_n$ such that $g_n \to \gamma_u$. Since $\gamma_u(U_i)$ is a function of U_i, this yields the result. ■

Proof of Corollary 1 : (i) Assume the null hypothesis. We first consider $\nu_{1n}(r)$. The convergence of finite dimensional distributions of the process $\sqrt{n}\mathbb{P}_n\left[\gamma_u\gamma_z^{\perp}\gamma_y^{\perp}\right]$ follows by the usual CLT. Note that the finite second moment condition trivially follows from the uniform boundedness of the indicator functions. And it is trivial to compute the covariance function of the process $\sqrt{n}\mathbb{P}_n[\gamma_u\gamma_z{}^{\perp}\gamma_y{}^{\perp}]$ to verify the covariance function of the limit Gaussian process in the theorem. Since the rectangle $[0,1]^3$ is obviously totally bounded with respect to the Euclidean norm, it remains to show that the process$\sqrt{n}\mathbb{P}_n a$ is stochastically equicontinuous in $a \in \mathcal{A}$ where $\mathcal{A}$ is defined by $\{\gamma_u(\gamma_z)^{\perp}(\gamma_y)^{\perp} : (u, z, y) \in [0,1]^3\}$. However, it is well-known that the class of univariate indicator functions is VC and products of a finite number of the indicator functions are also VC by the stability properties of VC classes (see van der Vaart and Wellner(1996)). Hence the required stochastic equicontinuity follows. Therefore we obtain the wanted result. It is a trivial matter to check the covariance kernel. We can likewise analyze the limit behavior of the process $\nu_n(r) = \sqrt{n}\mathbb{P}_n\left[\gamma_u\gamma_z\gamma_y^{\perp}\right]$ to obtain the wanted result.

(ii) First consider $\nu_n(r)$ which we write

$$\begin{aligned} \nu_n(r) &= \mathbb{P}_n\left[\gamma_u\gamma_z\gamma_y^{\perp}\right] \\ &= \mathbb{P}_n\gamma_u\left[\gamma_z\gamma_y^{\perp}\right]^{\perp} + \mathbb{P}_n[\gamma_u[\mathbf{P}^U\rho_w] - \mathbf{P}\{\gamma_u[\mathbf{P}^U\rho_w]\}] + \mathbf{P}[\gamma_u\rho_w], \end{aligned}$$

where $\rho_w \triangleq \gamma_z(\gamma_y - \mathbf{P}^U\gamma_y)$. Now, assume that the null hypothesis does not necessarily hold. Using the preliminary result in (6), we write

$$\begin{aligned}\hat{\nu}_n(r) &= \mathbb{P}_n\left[\gamma_u\gamma_z^\perp\gamma_y^\perp\right] + o_P(1) \\ &= \mathbb{P}_n\gamma_u\left[\gamma_z^\perp\gamma_y^\perp\right]^\perp + \mathbb{P}_n[\gamma_u[\mathbf{P}^U\rho_w] - \mathbf{P}\{\gamma_u[\mathbf{P}^U\rho_w]\}] + \mathbf{P}[\gamma_u\rho_w] + o_P(1)\end{aligned}$$

by adding and subtracting terms. The first and second process is a mean-zero empirical process, and using the similar arguments in (i), it is not hard to show that the processes are indexed by a Glivenko-Cantelli class. Hence, the result follows. ■

Proof of Theorem 2 (i) The proof is similar to the proof of Proposition 6.1 of Khmaladze and Koul (2004).

(ii) (a) We apply Theorem 1 to the process

$$\nu_n^{\mathcal{K}}(r) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_u(U_i)\gamma_z^{\mathcal{K}}(Z_i)\{\gamma_y^{\mathcal{K}}(Y_i) - (\hat{\mathbf{P}}^U\gamma_y^{\mathcal{K}})(Y_i, U_i)\}$$

by replacing γ_z and γ_y by $\gamma_z^{\mathcal{K}}$ and $\gamma_y^{\mathcal{K}}$. In view of the proof of Theorem 1, the main step is applying Lemma 1U of Escanciano and Song (2006) to the above process. To this end, it suffices to show that the classes $\{\gamma_y^{\mathcal{K}} : y \in [0,1]\}$ and $\{\gamma_z^{\mathcal{K}} : z \in [0,1]\}$ satisfy the required bracketing entropy conditions. This is shown in Lemma A2 below. Hence by following the same steps in the proof of Theorem 1, we obtain that under the null hypothesis,

$$\nu_n^{\mathcal{K}}(r) = \sqrt{n}\mathbb{P}_n[\gamma_u(\gamma_z^{\mathcal{K}})^\perp(\gamma_y^{\mathcal{K}})^\perp] + o_P(1). \tag{22}$$

Furthermore, by observing from (i) of this theorem,

$$\mathbf{P}^U(\gamma_y^{\mathcal{K}}) = 0 \text{ and } \mathbf{P}^U(\gamma_z^{\mathcal{K}}) = 0,$$

Note that under the null hypothesis of conditional independence, $(\gamma_z^{\mathcal{K}})(U_i, Z_i)$ and $(\gamma_y^{\mathcal{K}})(U_i, Y_i)$ are conditionally independent given U_i. Therefore

$$\mathbf{P}^U(\gamma_z^{\mathcal{K}}\gamma_y^{\mathcal{K}}) = \mathbf{P}^U(\gamma_z^{\mathcal{K}})\mathbf{P}^U(\gamma_y^{\mathcal{K}}) = 0.$$

Therefore, under the null hypothesis, we have

$$\nu_n^{\mathcal{K}}(r) = \sqrt{n}\mathbb{P}_n[\gamma_u(\gamma_z^{\mathcal{K}})(\gamma_y^{\mathcal{K}})] + o_P(1).$$

We investigate the process $\sqrt{n}\mathbb{P}_n[\gamma_u(\gamma_z^{\mathcal{K}})(\gamma_y^{\mathcal{K}})]$ and establish its weak convergence. The convergence of finite dimensional distributions follows by the usual central limit theorem. Note that the functions $\gamma_y^{\mathcal{K}}$ and $\gamma_z^{\mathcal{K}}$ are uniformly bounded. By Lemma A2, the process $\sqrt{n}\mathbb{P}_n[\gamma_u(\gamma_z^{\mathcal{K}})(\gamma_y^{\mathcal{K}})]$ is stochastically equicontinuous in S_0 and the index space S_0 is totally bounded with respect to the Euclidean norm in $[0,1]^3$, the weak convergence of the process $\sqrt{n}\mathbb{P}_n[\gamma_u(\gamma_z^{\mathcal{K}})(\gamma_y^{\mathcal{K}})]$ follows by Proposition in Andrews (1994), p. 2251.

The covariance function of the process is equal to

$$\mathbf{P}\left[\gamma_{u_1\wedge u_2}(\gamma_{z_1}^{\mathcal{K}})(\gamma_{z_2}^{\mathcal{K}})(\gamma_{y_1}^{\mathcal{K}})(\gamma_{y_2}^{\mathcal{K}})\right]. \tag{23}$$

It is easy to see that under the null hypothesis of conditional independence, the covariance function becomes

$$\begin{aligned}\mathbf{P}\left[\gamma_{u_1\wedge u_2}\mathbf{P}^U\left[(\gamma_{z_1}^{\mathcal{K}})(\gamma_{z_2}^{\mathcal{K}})(\gamma_{y_1}^{\mathcal{K}})(\gamma_{y_2}^{\mathcal{K}})\right]\right] &= \mathbf{P}\left[\gamma_{u_1\wedge u_2}\mathbf{P}^U\left[\gamma_{z_1}^{\mathcal{K}}\gamma_{z_2}^{\mathcal{K}}\right]\mathbf{P}^U\left[\gamma_{y_1}^{\mathcal{K}}\gamma_{y_2}^{\mathcal{K}}\right]\right] \\ &= \mathbf{P}\left[\gamma_{u_1\wedge u_2}\mathbf{P}^U\left[\gamma_{z_1}\gamma_{z_2}\right]\mathbf{P}^U\left[\gamma_{y_1}\gamma_{y_2}\right]\right] \\ &= \mathbf{P}\left[\gamma_{u_1\wedge u_2}\gamma_{z_1\wedge z_2}\gamma_{y_1\wedge y_2}\right].\end{aligned}$$

By the null hypothesis of conditional independence, the above is equal to

$$\int^{u_1\wedge u_2} F_{Z|U}(z_1\wedge z_2|u)F_{Y|U}(y_1\wedge y_2|u)dF(u).$$

Now, when we replace $\gamma_u^{\mathcal{K}}$, $\gamma_z^{\mathcal{K}}$, and $\gamma_y^{\mathcal{K}}$ by $\tilde{\gamma}_u^{\mathcal{K}}$, $\tilde{\gamma}_z^{\mathcal{K}}$, and $\tilde{\gamma}_y^{\mathcal{K}}$, the covariance function becomes

$$\begin{aligned}&\int^{u_1\wedge u_2} f^{-1}(u)\left[\int^{z_1\wedge z_2} f_{Z|U}^{-1}(z|u)f_{Z|U}(z|u)dz\right]\left[\int^{y_1\wedge y_2} f_{Y|U}^{-1}(y|u)f_{Y|U}(y|u)dy\right]f(u)du \\ =\ &(u_1\wedge u_2)(z_1\wedge z_2)(y_1\wedge y_2).\end{aligned}$$

(b) The result is immediate from (22). ■

We introduce notations for classes of functions:

$$\mathcal{G}_Y^{\mathcal{K}} \triangleq \{\gamma_y^{\mathcal{K}} : y\in[0,1]\} \text{ and } \mathcal{G}_{u,y}^{\mathcal{K}} \triangleq \{\mathbf{P}^U\gamma_z^{\mathcal{K}} : z\in[0,1]\}.$$

The following result provides the bracketing entropy bound for these classes. The result implies that these classes are P-Donsker.

Lemma A2 : *Suppose the conditions of* Theorem 2 *holds. Then we have for* $p\geq 2$,

$$N_{[]}(\varepsilon,\mathcal{G}_Y^{\mathcal{K}},||\cdot||_p)\leq C_1\varepsilon^{-2p} \text{ and } N_{[]}(\varepsilon,\mathcal{G}_{u,y}^{\mathcal{K}},||\cdot||_p)\leq C_2\varepsilon^{-2p}.$$

Proof : Define $\mathcal{G}_Y \triangleq \{\gamma_y : y\in[0,1]\}$. Then we have (e.g. the proof of Lemma A1(d) in Song (2006b)), $N_{[]}(\varepsilon,\mathcal{G}_Y,||\cdot||_p)\leq 2\varepsilon^{-p}$. Now, recall that

$$(\mathcal{M}_u\gamma_y)(u,\bar{y}) \triangleq (\mathbf{P}^U\gamma_{y\wedge\bar{y}}H)(u), \text{ where } H(\tilde{y},u)\triangleq\frac{1}{1-F_{Y|U}(\tilde{y};u)}.$$

From the previous observation, for the class $\mathcal{G}_Y' \triangleq \{\gamma_{y\wedge\bar{y}}(\cdot)/(1-F_{Y|U}(\cdot;\cdot)) : (y,\bar{y})\in[0,1]^2\}$, we have $N_{[]}(\varepsilon, \mathcal{G}'_Y, ||\cdot||_p) \leq C\varepsilon^{-p}$ and hence for the class $\mathcal{G}''_Y \triangleq \{\mathbf{P}^U g : g \in \mathcal{G}'_Y\}$, we also have $N_{[]}(\varepsilon, \mathcal{G}''_Y, ||\cdot||_p) \leq C\varepsilon^{-p}$. Note that the denominator of $(1 - F_{Y|U}(\cdot;\cdot))$ does not depend on the index $(y, \bar{y})^2 \in [0,1]^2$. Hence

$$N_{[]}(C\varepsilon, \mathcal{G}^{\mathcal{K}}_Y, ||\cdot||_p) \leq N_{[]}(\varepsilon, \mathcal{G}_Y, ||\cdot||_p) \times N_{[]}(\varepsilon, \mathcal{G}''_Y, ||\cdot||_p) \leq C\varepsilon^{-2p}.$$

From this, the second statement $N_{[]}(\varepsilon, \mathcal{G}^{\mathcal{K}}_{u,y}, ||\cdot||_p) \leq C_2\varepsilon^{-2p}$ immediately follows. ■

Proof of Corollary 2 : The covariance kernels of the limit Gaussian processes of $\nu(r)$ and $\nu_1(r)$ in Theorem 1 are respectively given by

$$\begin{aligned} c(r_1, r_2) &= (u_1 \wedge u_2)(z_1 \wedge z_2)(y_1 \wedge y_2 - y_1 y_2), \\ c_1(r_1, r_2) &= (u_1 \wedge u_2)(z_1 \wedge z_2 - z_1 z_2)(y_1 \wedge y_2 - y_1 y_2). \end{aligned}$$

Therefore, (e.g. Lemma 4.8.1 in Nikitin (1995)) we have

$$\begin{aligned} \lim_{\kappa\to\infty} \kappa^{-2} \log \mathbf{P}\left\{ \sup_{r\in[0,1]^{d_r}} |\nu(r)| \geq \kappa \right\} &= -\frac{1}{2} b_0 \\ \lim_{\kappa\to\infty} \kappa^{-2} \log \mathbf{P}\left\{ \sup_{r\in[0,1]^{d_r}} |\nu_1(r)| \geq \kappa \right\} &= -\frac{1}{2} b_1, \end{aligned}$$

where the constants b_0 and b_1 are given by

$$b_0^{-1} = \sup_{r\in[0,1]^3} c(r,r) = \frac{1}{4} \text{ and } b_1^{-1} = \sup_{r\in[0,1]^3} c_1(r,r) = \frac{1}{16}.$$

Now let us consider the case of alternatives. The results of Theorem 1 yields that the function φ_0 and φ_1 in (8) are equal, given by

$$\varphi_0(r) = \mathbf{P}_1 \left[\gamma_u (\langle \gamma_z, \gamma_y \rangle_U - \langle \gamma_z, 1 \rangle_U \langle \gamma_y, 1 \rangle_U) \right].$$

Therefore, we obtain the result of the approximate Bahadur efficiency $e_{0,1}(\mathbf{P}_1)$. ■

Proof of Corollary 3 : (i) The covariance kernels of the limit Gaussian processes of $\nu^{\mathcal{K}}(r)$ is given by

$$c_{\mathcal{K}}(r_1, r_2) = (u_1 \wedge u_2)(z_1 \wedge z_2)(y_1 \wedge y_2),$$

yielding the result that (e.g. Lemma 4.8.1 in Nikitin (1995)) we have

$$\lim_{\kappa\to\infty} \kappa^{-2} \log \mathbf{P}\left\{ \sup_{r\in[0,1]^{d_r}} |\nu_{\mathcal{K}}(r)| \geq \kappa \right\} = -\frac{1}{2} b_{\mathcal{K}},$$

where the constant $b_{\mathcal{K}}$ is given by $b_{\mathcal{K}}^{-1} = \sup_{r\in[0,1]^{d_r}} c_{\mathcal{K}}(r,r) = 1$. On the other hand, under $\mathbf{P}_1$, we have

$$\begin{aligned}\varphi_0(r) &= \mathbf{E}_1\left[\gamma_u\{\langle\gamma_z,\gamma_y\rangle_U - \langle\gamma_z,1\rangle_U\langle\gamma_y,1\rangle_U\}\right] \\ &= \mathbf{E}_1\left[\gamma_u(U)\left\{C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U) - F_{Z|U}(z|U)F_{Y|U}(y|U)\right\}\right].\end{aligned}$$

Now, we turn to $\varphi_{\mathcal{K}}(r)$. Observe that

$$\begin{aligned}\varphi_{\mathcal{K}}(r) &= \mathbf{P}_1\left[\gamma_u(\langle\gamma_z^{\mathcal{K}},\gamma_y^{\mathcal{K}}\rangle_U - \langle\gamma_z^{\mathcal{K}},1\rangle_U\langle\gamma_y^{\mathcal{K}},1\rangle_U)\right] \\ &= \mathbf{P}_1\left[\gamma_u\langle\gamma_z^{\mathcal{K}},\gamma_y^{\mathcal{K}}\rangle_U\right] = \mathbf{E}_1\left[(\mathcal{K}\gamma_z)(U;Z)(\mathcal{K}\gamma_y)(U;Y)|U\right].\end{aligned} \tag{24}$$

The second equality follows from the conditional orthogonality property of $\mathcal{K}$.

First, assume without loss of generality that $\xi(z,\varepsilon)$ is strictly increasing in z. We observe that the last term in (24) is equal to

$$\mathbf{E}_1\left[(\mathcal{K}\gamma_z)(U;Z)(\mathcal{K}\gamma_{\xi^{-1}(y,\varepsilon)})(U;Z)|U\right] \tag{25}$$

in this case. (When $\xi(z,\varepsilon)$ is strictly decreasing in z, the last term in (24) is equal to

$$\begin{aligned}&\mathbf{E}_1\left[(\mathcal{K}\gamma_z)(U;Z)(\mathcal{K}(1-\gamma_{\xi^{-1}(y,\varepsilon)}))(U;Z)|U\right] \\ = \; & -\mathbf{E}_1\left[(\mathcal{K}\gamma_z)(U;Z)(\mathcal{K}\gamma_{\xi^{-1}(y,\varepsilon)})(U;Z)|U\right],\end{aligned}$$

because $\mathcal{K}$ is a linear operator and $\mathcal{K}a = 0$ for any constant a. Then we can follow the subsequent steps in a similar manner.) By the conditional isometry condition, the term in (25) is equal to

$$\begin{aligned}\mathbf{E}_1\left[\gamma_z(Z)\gamma_{\xi^{-1}(y,\varepsilon)}(Z)|U\right] &= \mathbf{E}_1\left[F_{Z|U}\left(z\wedge\xi^{-1}(y,\varepsilon)|U\right)|U\right] \\ &= \mathbf{P}\left\{Z\le z, Z\le\xi^{-1}(y,\varepsilon)|U\right\} \\ &= C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U).\end{aligned}$$

Therefore,

$$\begin{aligned}\Gamma\varphi_0 &= \sup_{z,y,u}\left|\mathbf{E}\left[\gamma_u(U)\left\{C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U) - F_{Z|U}(z|U)F_{Y|U}(y|U)\right\}\right]\right| \\ &= \sup_{z,y}\left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U) - F_{Z|U}(z|U)F_{Y|U}(y|U)\right]\right| \\ \Gamma\varphi_{\mathcal{K}} &= \sup_{z,y,u}\mathbf{E}\left[\gamma_u(U)C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)\right] \\ &= \sup_{z,y}\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)\right]\end{aligned}$$

Combining the results, we obtain the first statement.

$$\begin{aligned} e_{1,\mathcal{K}}(\mathbf{P}_1) &= \frac{4\sup_{z,y}\left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)-F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\right|}{\sup_{z,y}\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)\right]} \\ &= 4\sup_{z,y}\left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)-F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\right|. \end{aligned}$$

The last equality follows as the supremum in the denominator is achieved when $z = y = 1$.

(ii) Write the approximate Bahadur ARE as

$$\begin{aligned} e_{1,\mathcal{K}}(\mathbf{P}_1) &= 4\sup_{z,y}\left|\mathbf{E}\left[C_{Z,Y|U}(F_{Z|U}(z|U),F_{Y|U}(y|U)|U)\right]-\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\right| \\ &= 4\sup_{z,y}\left|C_{Z,Y}(z,y)-\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\right|. \end{aligned}$$

First, we assume that $\xi(z,\varepsilon)$ is strictly increasing in z. Then

$$\begin{aligned} &\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]-zy \\ =\; &\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]-\mathbf{E}\left[F_{Z|U}(z|U)\right]\mathbf{E}\left[F_{Y|U}\left(y|U\right)\right]\geq 0 \end{aligned}$$

because $F_{Z|U}(z|U)$ and $F_{Y|U}(y|U)$ are nonnegatively correlated. Hence take $h(z,y) \triangleq z\wedge y - C_{Z,Y}(z,y)\geq 0$ and observe

$$\begin{aligned} e_{1,\mathcal{K}}(\mathbf{P}_1) &\leq 4\sup_{z,y}\left\{C_{Z,Y}(z,y)-zy\right\}\leq 4\sup_{z,y}\left\{C_{Z,Y}(z,y)+h(z,y)-zy\right\} \\ &= 4\sup_{z,y}\left\{z\wedge y-zy\right\}=1. \end{aligned}$$

The last equality follows as the supremum is achieved by $z = y = 1/2$.

Now, assume that $\xi(z,\varepsilon)$ is strictly decreasing in z. Then $F_{Z|U}(z|U)$ and $F_{Y|U}(y|U)$ are non-positively correlated and we have

$$\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]-zy\leq 0.$$

We take $h(z,y) = C_{Z,Y}(z,y) - \max\{z+y-1,0\} \geq 0$. Since Z and Y are negatively dependent conditional on U, $C_{Z,Y}(z,y)-\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]\leq 0$ and hence

$$\begin{aligned} e_{1,\mathcal{K}}(\mathbf{P}_1) &= 4\sup_{z,y}\left\{\mathbf{E}\left[F_{Z|U}(z|U)F_{Y|U}\left(y|U\right)\right]-C_{Z,Y}(z,y)\right\} \\ &\leq 4\sup_{z,y}\left\{zy-C_{Z,Y}(z,y)\right\}\leq 4\sup_{z,y}\left\{zy+h(z,y)-C_{Z,Y}(z,y)\right\} \\ &= 4\sup_{z,y}\left\{zy-\max\{z+y-1,0\}\right\}=1. \end{aligned}$$

■

Proof of Corollary 4 : The covariance kernels of the limit Gaussian processes of $\nu^{\mathcal{K}}(r)$ and $\nu_1^{\mathcal{K}}(r)$

in Theorem 1 are respectively given by

$$\begin{aligned} c_{\mathcal{K}}(r_1, r_2) &= C(u_1 \wedge u_2)(y_1 \wedge y_2), \\ c_{1,\mathcal{K}}(r_1, r_2) &= C(u_1 \wedge u_2)(y_1 \wedge y_2 - y_1 y_2). \end{aligned}$$

Therefore, (e.g. Lemma 4.8.1 in Nikitin (1995)) we have

$$\begin{aligned} \lim_{\kappa \to \infty} \kappa^{-2} \log \mathbf{P} \left\{ \sup_{r \in [0,1]^3} \left| \nu^{\mathcal{K}}(r) \right| \geq \kappa \right\} &= -\frac{1}{2} b_0 \\ \lim_{\kappa \to \infty} \kappa^{-2} \log \mathbf{P} \left\{ \sup_{r \in [0,1]^3} \left| \nu_1^{\mathcal{K}}(r) \right| \geq \kappa \right\} &= -\frac{1}{2} b_1, \end{aligned}$$

where the constants b_0 and b_1 are given by

$$b_0^{-1} = \sup_{r \in [0,1]^3} c_{\mathcal{K}}(r, r) = 1 \text{ and } b_1^{-1} = \sup_{r \in [0,1]^3} c_{1,\mathcal{K}}(r, r) = \frac{1}{4}.$$

Now let us consider the case of alternatives. The results of Theorem 1 yields that the function φ_0 and φ_1 in (8) are equal, given by

$$\varphi_0(r) = \mathbf{E} \left[Z \gamma_u(U) (\tilde{\gamma}_y^{\mathcal{K}})(U, Y) \right].$$

Therefore, we obtain the result of the approximate Bahadur efficiency $e_{0,1}(\mathbf{P}_1)$. ■

References

[1] Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica* 71, 1795-1843.

[2] Altonji, J. and R. Matzkin (2005), "Cross-section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica* 73, 1053-1102.

[3] Andrews, D. W. K (1991), "An empirical process central limit theorem for dependent non-identically distributed random variables," *Journal of Multivariate Analysis*, 38, 187-203.

[4] Andrews, D. W. K (1994), "Empirical process method in econometrics," in *The Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden, Amsterdam: North-Holland.

[5] Andrews, D. W. K (1995), "Nonparametric kernel estimator for semiparametric models," *Econometric Theory*, 11, 560-595.

[6] Andrews, D. W. K (1997), "A conditional Kolmogorov test," *Econometrica*, 65, 1097-1128.

[7] Angrist, J. D. and G. M. Kuersteiner (2004), "Semiparametric causality tests using the policy propensity score," Working paper.

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

[9] Bahadur, R. R. (1960), "Stochastic comparison of tests," *Annals of Mathematical Statistics*, 31, 276-295.

[10] Bierens, H. J. (1990), "A consistent conditional moment test of functional form," *Econometrica* 58, 1443-1458.

[11] Bierens, H. J. and W. Ploberger, (1997), "Asymptotic theory of integrated conditional moment tests," *Econometrica* 65, 1129-1151.

[12] Chen, X., O. Linton, and I. van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica* 71, 1591-1608.

[13] Chen, X., and X. Shen (1998), "Sieve extremum estimates for weakly dependent data," *Econometrica*, 66, 289-314.

[14] Chung, K. L. (2001), *A Course in Probability Theory*, Academic Press, 3rd Ed.

[15] Dawid, P. (1980), "Conditional independence for statistical operations," *Annals of Statistics*, 8, 598-617.

[16] Delgado, M. A., W. Gonzáles Manteiga (2001), "Significance testing in nonparametric regression based on the bootstrap," *Annals of Statistics* 29, 1469-1507.

[17] Escanciano, J. C. and K. Song (2006), "Asymptotically optimal tests for single index restrictions with a focus on average partial effects," Working paper.

[18] Fan, Y. and Q. Li (1996), "Consistent model specification tests: omitted variables, parametric and semiparametric functional forms," *Econometrica*, 64, 865-890.

[19] Härdle, W. and E. Mammen (1993), "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21, 1926-1947.

[20] Hirano, K., G. W. Imbens, and G. Ridder (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161-1189.

[21] Khmaladze, E. V. (1993), "Goodness of fit problem and scanning innovation martingales," *Annals of Statistics*, 21, 798-829.

[22] Khmaladze, E. V. and H. Koul (2004), "Martingale transforms goodness-of-fit tests in regression models," *Annals of Statistics*, 32, 995-1034.

[23] Linton, O. and P. Gozalo (1997), "Conditional independence restrictions: testing and estimation", Cowles Foundation Discussion Paper 1140.

[24] McKeague, I. W. and Y. Sun (1996), "Transformations of gaussian random fields to Brownian sheet and nonparametric change-point tests," *Statistics and Probability Letters*, 28, 311-319.

[25] Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer-Verlag, New York

[26] Newey, W. K. (1997), "Convergence rates and asymptotic normality of series estimators," *Journal of Econometrics*, 79, 147-168.

[27] Nikitin, Y. (1995), *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York.

[28] Nishiyama, Y. (2000), *Entropy methods for martingales*, CWI Tract, 128.

[29] Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41-55.

[30] Patton, A. (2006), "Modeling asymmetric exchange rate dependence," *International Economic Review*, 47, 527-556.

[31] Rosenblatt, M. (1952), "Remarks on a multivariate transform," *Annals of Mathematical Statistics*, 23, 470-472.

[32] Rubin, D. "Bayesian inference for causal effects: the role of randomization," *Annals of Statistics*, 6, 34-58.

[33] Song, K. (2006a), "Uniform convergence of series estimator over functional spaces," Working paper.

[34] Song, K. (2006b), "Testing semiparametric conditional moment restrictions using conditional martingale transforms," Working Paper.

[35] Stinchcombe, M. B. and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory*, 14, 295-325.

[36] Stute, W. and L. Zhu (2005), "Nonparametric checks for single-index models," *Annals of Statistics*, 33, 1048-1083.

[37] Su, L. and H. White (2003a), "Testing conditional independence via empirical likelihood," Discussion Paper, University of California San Diego.

[38] Su, L. and H. White (2003b), "A nonparametric Hellinger metric test for conditional independence," Discussion Paper, University of California San Diego.

[39] Su, L. and H. White (2003c), "A characteristic-function-based test for conditional independence," Discussion Paper, University of California San Diego.

[40] van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

Table 1 : Size of the Tests : Nominal size is set at 5 Percent [10]

$n=300$ Order 1	Order 2	MG trns ($T_{1A}^{\mathcal{K}}$)	MG trns ($T_{1B}^{\mathcal{K}}$)	MG trns ($T_{2}^{\mathcal{K}}$)	Bootstrap
	4	0.034	0.054	0.061	0.060
4^2	5	0.050	0.077	0.076	0.055
	6	0.058	0.072	0.073	0.056
	7	0.076	0.118	0.088	0.055
	4	0.050	0.044	0.070	0.039
5^2	5	0.037	0.064	0.059	0.056
	6	0.059	0.099	0.078	0.046
	7	0.090	0.120	0.115	0.042
	4	0.035	0.040	0.055	0.050
6^2	5	0.032	0.068	0.057	0.055
	6	0.054	0.092	0.085	0.056
	7	0.070	0.106	0.096	0.049
	4	0.041	0.048	0.067	0.061
7^2	5	0.047	0.073	0.066	0.051
	6	0.062	0.084	0.082	0.048
	7	0.080	0.121	0.089	0.055

[10] Order 1 represents the number of terms included in the bivariate density function estimation using Legendre polynomials. Order 2 represents the number of terms included in the estimation of conditional expectation given U_i using Legendre polynomials.

MG trns ($T_{1A}^{\mathcal{K}}$) and MG trns ($T_{2}^{\mathcal{K}}$) indicate tests based on martingale transforms, the former using the test statistic $T_{1A}^{\mathcal{K}}$ which is based on the true conditional expectation of $\gamma_y^{\mathcal{K}}(Y)$ given U and the latter using the test statistic $T_2^{\mathcal{K}}$ which is based on the estimated conditional expectation of $\gamma_y^{\mathcal{K}}(Y)$ given U. MG trns ($T_{1B}^{\mathcal{K}}$) is the symmetrized version of MG trns ($T_{1A}^{\mathcal{K}}$).

Table 2 : Power of the Tests : Conditional Positive Dependence

$n = 300$	Order 1	Order 2	MG trns ($T_{1A}^{\mathcal{K}}$)	MG trns ($T_{1B}^{\mathcal{K}}$)	MG trns ($T_{2}^{\mathcal{K}}$)	Bootstrap
$\kappa = 0.2$	5^2	5	0.193	0.191	0.229	0.216
		6	0.214	0.190	0.225	0.203
	6^2	5	0.187	0.196	0.226	0.230
		6	0.245	0.213	0.269	0.243
$\kappa = 0.4$	5^2	5	0.575	0.581	0.678	0.676
		6	0.604	0.605	0.667	0.664
	6^2	5	0.591	0.585	0.679	0.673
		6	0.626	0.612	0.694	0.701
$\kappa = 0.6$	5^2	5	0.909	0.938	0.953	0.950
		6	0.930	0.921	0.957	0.956
	6^2	5	0.898	0.927	0.949	0.949
		6	0.919	0.912	0.951	0.948

Table 3 : Power of the Tests : Conditional Negative Dependence

$n = 300$	Order 1	Order 2	MG trns ($T_{1A}^{\mathcal{K}}$)	MG trns ($T_{1B}^{\mathcal{K}}$)	MG trns ($T_{2}^{\mathcal{K}}$)	Bootstrap
$\kappa = -0.2$	5^2	5	0.024	0.176	0.208	0.235
		6	0.039	0.217	0.204	0.208
	6^2	5	0.029	0.197	0.191	0.238
		6	0.036	0.229	0.201	0.201
$\kappa = -0.4$	5^2	5	0.070	0.614	0.652	0.636
		6	0.068	0.647	0.634	0.640
	6^2	5	0.072	0.644	0.644	0.649
		6	0.062	0.656	0.628	0.650
$\kappa = -0.6$	5^2	5	0.367	0.944	0.956	0.935
		6	0.319	0.965	0.957	0.928
	6^2	5	0.359	0.943	0.941	0.925
		6	0.306	0.949	0.950	0.933

Figure 1: The Empirical Density of Test Statistics both under the Null and under the Alternative: n=3000

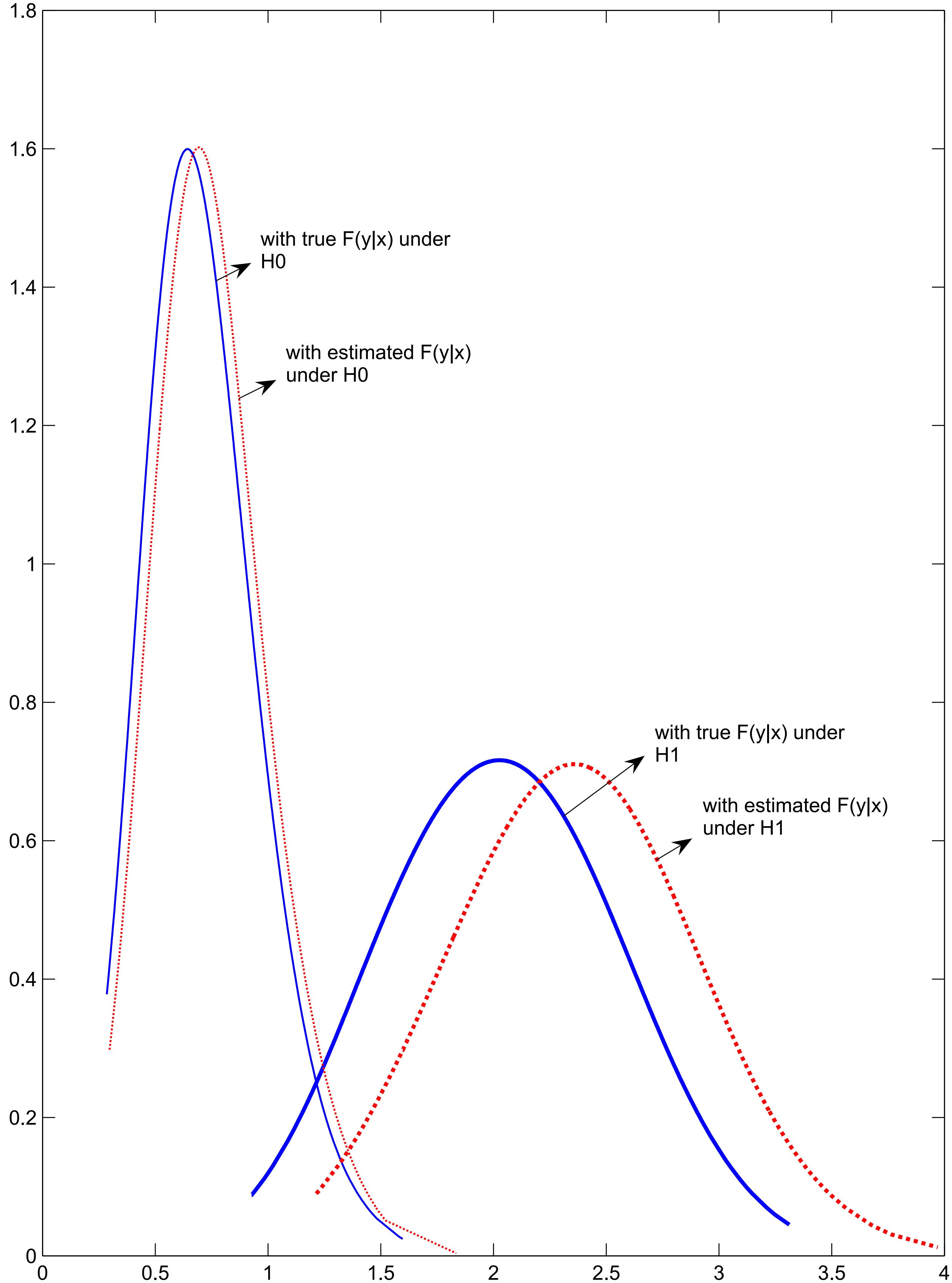

Figure 2: Empirical Density of Empirical Processes at (x,y)=(0.5,0.5) under the Null and under the Alternative: n=3000