Testing Conditional Independence using Conditional Martingale Transforms

Kyungchul Song

Department of Economics, University of Pennsylvania

March 13, 2007

Abstract

This paper investigates testing conditional independence between $Y$ and $Z$ given $\lambda_\theta(X)$ for some $\theta \in \Theta \subset \mathbb{R}^d$, for a function $\lambda_\theta(\cdot)$ known up to a parameter $\theta \in \Theta$. First, the paper proposes a new method of conditional martingale transforms under which tests are asymptotically pivotal and asymptotically unbiased against $\sqrt{n}$-converging Pitman local alternatives. Second, the paper performs an analysis of asymptotic power comparison using the limiting Pitman efficiency. From this analysis, it is found that there is improvement in power when we use nonparametric estimators of conditional distribution functions in place of true ones. When both $Y$ and $Z$ are continuous, the improvement in power due to the nonparametric estimation disappears after the conditional martingale transform, but when either $Y$ or $Z$ is binary, this phenomenon of power improvement reappears even after the conditional martingale transform. We perform Monte Carlo simulation studies to compare the martingale transform approach and the bootstrap approach.

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

JEL Classifications: C12, C14, C52.

1I 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, Bernard Salanié, Norm Swanson and seminar participants at a KAED meeting and workshops at Greater New York Metropolitan Econometrics Colloquium, Rutgers University, Texas A&M, UC Riverside, UCLA 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.
1 Introduction

Suppose that $Y$ and $Z$ are random variables, and let $\lambda_\theta(X)$ be a real valued function of a random vector $X$ indexed by a parameter $\theta \in \Theta$. This paper investigates the problem of testing the conditional independence of $Y$ and $Z$ given $\lambda_\theta(X)$ for some $\theta \in \Theta$, i.e., (using the notation of Dawid (1980))

$$Y \perp Z | \lambda_\theta(X) \text{ for some } \theta \in \Theta.$$  \hspace{1cm} (1)

The function $\lambda_\theta(\cdot)$ is known up to a finite dimensional parameter $\theta \in \Theta$.

The conditional independence restriction is often used as part of the specification of an econometric model such as an identifying restriction. It is crucial for identification of parameters of interest in nonseparable models. (e.g. Altonji and Matzkin (2005) and Vytlacil and Yildiz (2006).) The restriction is also frequently used in the literature of program evaluations. (e.g. Rubin (1978), Rosenbaum and Rubin (1983), and Hirano, Imbens, and Ridder (2003)). The restriction is, sometimes, a direct implication of economic theory. For example, in the literature of insurance, presence of positive conditional dependence between coverage and risk is known to be a direct consequence of adverse selection. (e.g. Chiappori and Salanié (2000)).

The literature of testing conditional independence for continuous variables appears rather recent and, apparently, involves relatively few researches as compared to that of other non-parametric or semiparametric tests. Linton and Gozalo (1999) proposed tests that are slightly stronger than tests of conditional independence. Delgado and Gonzáles Manteiga (2001) also proposed a bootstrap-based test of conditional independence. 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), Delgado and Gonzáles Manteiga (2001), and Angrist and Kuersteiner (2004). In contrast with Linton and Gozalo (1999) and Delgado and Gonzáles Manteiga (2001), 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. 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 pivotal tests. However, they assume that $Z$ is binary and the conditional probability $P\{Z_i = 1|X_i\}$ 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 (2007)).

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 (2004) 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. See also Bai and Ng (2003) for testing conditional symmetry in a similar vein. Song (2007) developed a method of conditional martingale transform that applies to tests of semiparametric conditional moment restrictions. Although the conditional independence restriction can be viewed as a special case of semiparametric conditional moment restrictions, the restriction lies beyond the scope of Song (2007) and the manner the conditional martingale transforms are employed is quite different.

One of the most distinct aspects of the hypothesis testing framework of focus in this paper is that we allow the conditioning variable $\lambda_{\theta}(X)$ to depend on unknown parameter $\theta$. This is particularly important when the dimension of the random vector $X$ is large. For example, the number of variables in $X$ used in Chiappori and Salanié (2000) is 55. In such a situation, instead of testing

$$Y \perp Z|X,$$

it is reasonable to impose a single-index restriction upon $X$ and test the hypothesis of the form in (1). One can check the robustness of the result by varying the parameterization used for the single-index $\lambda_{\theta}(X).

---

\(^2\)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)).
This paper contains also an extensive analysis of asymptotic power properties of tests by using the limiting Pitman efficiency. First, we find that using the nonparametric estimator of the conditional distribution function of $Y$ given $\lambda_\theta(X)$ in the test statistic improves the asymptotic power of the test as compared to the case where one uses a true one. This may strike one as counter-intuitive because even when the true nonparametric function is known, one does better by estimating the function. We affirm this phenomenon in terms of the limiting Pitman efficiency, and provide intuitive explanation. In fact, this phenomenon can be viewed precisely as a hypothesis testing version of what Hirano, Imbens, and Ridder (2003) have found in estimation.

When $Y$ and $Z$ are continuous, our approach suggests applying 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 of the conditional distribution function, 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 $Y$.

A conditional martingale transform alters the asymptotic power properties of the original test. Under the setup of a nonseparable model: $Y = \xi(Z, \varepsilon)$, we compute the limiting Pitman efficiency of two Kolmogorov-Smirnov type tests, one using the original empirical process and the other using the transformed empirical process. We characterize the set of alternatives under which the conditional martingale transform weakly improves the asymptotic power of the original test.

We perform Monte Carlo simulations to compare the approach of bootstrap and that of martingale transforms. We find that the martingale transform-based tests overall appear to perform as well as the bootstrap-based test.

In the next section, we define the null hypothesis of conditional independence and discuss examples. In Section 3, we present the asymptotic representation of the semiparametric empirical process. In Section 4, we introduce conditional martingale transforms and in Section 5, we focus on the case where $Z$ is binary. Section 6 is devoted to the asymptotic power analysis of a variety of tests. In Section 7, we present and discuss results from simulation studies and in Section 8, we conclude.
2 Testing Conditional Independence

2.1 The Null and the Alternative Hypotheses

Suppose that we are given with a random vector \((Y, Z, X)\) distributed by \(P\) and an unknown real valued function \(\lambda_\theta(\cdot)\) on \(\mathbb{R}^d\times, \theta \in \Theta \subset \mathbb{R}^d\). We assume the following for \((Y, Z)\).

Assumption 1: (i) \(Y\) and \(Z\) are continuous random variables. (ii) \(\Theta\) is compact in \(\mathbb{R}^d\).

When either \(Y\) or \(Z\) is binary, the development of this paper’s thesis becomes simpler, as we will see when \(Z\) is binary in a later section. We introduce notations for indicator functions: for a random variable \(S\) and a real number \(s\),

\[ \gamma_s(S) = 1\{S \leq s\}. \]

Hence we write \(\gamma_z(Z) = 1\{Z \leq z\}\), \(\gamma_y(Y) = 1\{Y \leq y\}\), and \(\gamma_\lambda(\lambda_\theta(X)) = 1\{\lambda_\theta(X) \leq \lambda\}\).

The null hypothesis in (1) requires to check the conditional independence of \(Y\) and \(Z\) given \(\lambda_\theta(X)\) for all \(\theta \in \Theta\) until we find one \(\theta\) that satisfies the conditional independence restriction. The following assumption simplifies this problem by making it suffice to focus on a specific \(\theta_0\) in \(\Theta\).

Assumption 2: There exists a unique parameter \(\theta_0 \in \Theta\) such that (i) \(Y \perp Z|\lambda_{\theta_0}(X)\), whenever the null hypothesis of (1) holds, and (ii) for some \(\delta > 0\), \(\lambda_\theta(X)\) is continuous for all \(\theta \in B(\theta_0, \delta) \triangleq \{\theta \in \Theta : ||\theta - \theta_0|| < \delta\}\).

For example, the unique parameter \(\theta_0\) can be defined as a unique solution to the following problem

\[
\theta_0 = \arg\min_{\theta \in \Theta} \sup_{(y,z,\lambda) \in \mathbb{R}^3} \left| \mathbb{E}[\gamma_\lambda(\lambda_\theta(X))\gamma_z(Z)(\gamma_y(Y) - \mathbb{E} [\gamma_y(Y)|\lambda_\theta(X)])] \right|, \tag{2}
\]

when a unique solution exists. Under Assumption 2, we can write the null hypothesis equivalently as (see Theorem 9.2.1 in Chung (2001), p.322, and Lemma 1 in Song (2007))

\[ H_0 : P \left\{ \mathbb{E}(\gamma_y(Y)|U, Z) = \mathbb{E}(\gamma_y(Y)|U), \text{ for all } y \in [0, 1] \right\} = 1, \]

where \(U \triangleq F_{\theta_0}(\lambda_{\theta_0}(X))\) and \(F_\theta\) denotes the distribution function of \(\lambda_\theta(X)\). The alternative hypothesis is given by the negation of the null:

\[ H_1 : P \left\{ \mathbb{E}(\gamma_y(Y)|U, Z) = \mathbb{E}(\gamma_y(Y)|U), \text{ for all } y \in \mathbb{R} \right\} < 1. \]
The next subsection provides examples in which this testing framework is relevant.

2.2 Examples

2.2.1 Nonseparable Models

The 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

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

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. Vytlačil 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 $\varepsilon$ is independent of $Z$ given an instrumental variable $X$. This gives the following testable implication of conditional independence restriction, $Y \perp Z|X$.

2.2.2 Heterogeneous Treatment Effects

In the literature of program evaluations, 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.)

Here $Y_0$ represents the outcome when the agent is not treated and $Y_1$, the outcome when treated. The variable $X$ represents the agent’s covariates.

---

3It is also worth noting that Heckman, Ichimura, and Todd (1998) point out that a weaker condition of mean independence suffices for the identification of the average treatment effect.
It is well-known (Rosenbaum and Rubin (1983)) that when \( p(X) \triangleq P(Z = 1|X) \in (0, 1) \),

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

This restriction is not directly testable because we do not observe \( Y_0 \) and \( Y_1 \) simultaneously for each individual. However, as pointed out by Heckman, Smith, and Clemens (1997), we have the testable implications: \((1 - Z)Y \perp Z \mid p(X)\) or \(ZY \perp Z \mid p(X)\). When we specify \( p(\cdot) \) parametrically, the testing framework belongs to (1).

### 2.2.3 Testing No Conditional Positive Dependence

In the literature of contract theory, there has been an interest in testing relevance of asymmetric information. Under the asymmetric information, it is known that the risk is positively related to the coverage of the contract conditional on all publicly observed variables. (See Cawley and Phillipson (1999), Chiappori and Salanié (2000), Chiappori, Jullien, Salanié and Salanié (2002), and references therein.) In this case, the null hypothesis is conditional independence of the risk and the coverage, but the alternative hypothesis is restricted to positive dependence of the risk (or probability of claims in the insurance data) and the coverage of the insurance contract conditional on all observable variables. This paper’s framework can be used to deal with this case of one-sided testing. At the end of the paper, we provide a table of critical values for one-sided Kolmogorov-Smirnov statistics.

### 3 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

\[g^Y_y(u) \triangleq E(\gamma_y(Y)|U = u) \quad \text{and} \quad g^Z_z(u) \triangleq E(\gamma_z(Z)|U = u)\]

and \( \mathcal{S} \triangleq \mathbb{R}^2 \times [0, 1] \). Then the null hypothesis can be written as:

\[H_0 : E[\gamma_y(U)\gamma_z(Z)(\gamma_y(Y) - g^Y_y(U))] = 0, \forall (y, z, u) \in \mathcal{S}.\]
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 \( S = (Y, Z, U) \). Let us define

\[ \hat{U}_{\theta,i} = F_{n,\theta,i}(\lambda_{\theta}(X_i)), \]

where \( F_{n,\theta,i}(\cdot) \) denotes the empirical distribution function of \( \{\lambda_{\theta}(X_i)\}_{i=1}^n \) with the omission of the \( i \)-th data point \( \lambda_{\theta}(X_i) \). We assume the following.

**Assumption 3:**

(i) \( (Y_i, Z_i, X_i)_{i=1}^n \) is a random sample from \( P \) where \( E\|Y\|^p < \infty \) and \( E\|Z\|^p < \infty \) for some \( p > 4 \).

(ii) There exists an estimator \( \hat{\theta} \) of \( \theta_0 \) in Assumption 2 such that \( \|\hat{\theta} - \theta_0\| = o_P(n^{-1/4}) \).

(iii) \( \Lambda \triangleq \{\lambda_{\theta}(\cdot) : \theta \in \Theta\} \) is uniformly bounded and there exists \( \delta > 0 \), such that for any \( \theta_1, \theta_2 \in B(\theta_0, \delta) \),

\[ |\lambda_{\theta_1}(x) - \lambda_{\theta_2}(x)| \leq C\|\theta_1 - \theta_2\| \]

for some constant \( C > 0 \).

The estimator \( \hat{\theta} \) in (ii) can be obtained from a sample version of the problem in (2). The consistency can be obtained using an approach of Chen, Linton, and van Keilegom (2003). The uniform boundedness condition in (iii) for \( \Lambda \) is innocuous. For any strictly increasing function \( \Phi \) on \([0, 1]\), we can redefine \( \lambda'_{\theta} = \Phi \circ \lambda_{\theta} \) so that \( \lambda'_{\theta} \) is now uniformly bounded. The null hypothesis and the alternative hypothesis, and the random variable \( \hat{U}_{\theta,i} \) remain invariant to this redefinition of \( \lambda_{\theta} \).

If the function \( g^Y_y(\cdot) \) were known and the quantile transformed data \( (U_i)_{i=1}^n \) for \( U \) were observed, a test statistic could be constructed as a functional of the following stochastic process:

\[ \nu_n(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_u(U_i)\gamma_z(Z_i)\gamma_y(Y_i) - g^Y_y(U_i), \quad r = (y, z, u) \in \mathcal{S}. \]  

(3)

In this section, we introduce a series estimator of \( g^Y_y \) and investigate the asymptotic properties of the resulting empirical process that involves the estimator. We approximate \( g^Y_y(u) \) by \( p^K(u)'\pi_y \) which is constituted by a basis function vector \( p^K(u) \) and a coefficient vector \( \pi_y \). Given an appropriate estimator \( \hat{\lambda}(\cdot) \) of \( \lambda(\cdot) \), define

\[ \hat{U}_i \triangleq \frac{1}{n} \sum_{j=1,j\neq i}^n 1\{\lambda_{\theta}(X_j) \leq \lambda_{\theta}(X_i)\} \cdot \]  

(4)

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

$$a_y \triangleq \begin{bmatrix} \gamma_y(Y_1) \\ \vdots \\ \gamma_y(Y_n) \end{bmatrix} \text{ and } \hat{P} \triangleq \begin{bmatrix} p^K(\hat{U}_1) \\ \vdots \\ p^K(\hat{U}_n) \end{bmatrix}. \quad (5)$$

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

$$\hat{\nu}_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i) \gamma_z(Z_i)(\gamma_y(Y_i) - \hat{g}_y^Y(\hat{U}_i)). \quad (6)$$

A test statistic can be constructed as a known functional of the process $\hat{\nu}_n(r)$. We assume the following for the distribution of $\lambda_\theta(X)$, $\theta \in B(\theta_0, \delta)$, for some $\delta > 0$. In the following $S_\theta$ denotes the support of $\lambda_\theta(X)$.

**Assumption 4**: (i) There exists $\delta > 0$ such that (a) the density $f_\theta(\lambda)$ of $\lambda_\theta(X)$ satisfies that $\sup_{\theta \in B(\theta_0, \delta)} \sup_{\lambda \in S_\theta} f_\theta(\lambda) < \infty$, (b) for some $C > 0$, $\sup_{\theta \in B(\theta_0, \delta)} \sup_{\lambda \in S_\theta} |F_\theta(\lambda + \delta') - F_\theta(\lambda - \delta')| < C\delta'$, for all $\delta' > 0$, and (c) there exists $C > 0$ such that for each $u \in U \triangleq \{F_\theta \circ \lambda_\theta : \theta \in B(\theta_0, \delta)\}$, and each $\delta' > 0$,

$$\sup_{(\bar{u}_1, \bar{u}) \in [0,1]^2 : |\bar{u}_2 - \bar{u}_1| < \delta'} |f_{Y,X|u(X)}(y, x|\bar{u}_1) - f_{Y,X|u(X)}(y, x|\bar{u}_2)| \leq \varphi_u(y, x) \delta',$$

where $f_{Y,X|u(X)}$ denotes the conditional density function of $(Y, X)$ given $u(X)$, and $\varphi_u(\cdot, \cdot)$ is a real function that satisfies $\sup_{x \in \mathbb{R}^d} \int \varphi_u(y, x) dy < C$ and $\int \varphi_u(y, x) dx < Cf_Y(y)$ with $f_Y(\cdot)$ denoting the density of $Y$.

(ii) The conditional distribution functions $g_y^Y(u)$ and $g_z^Z(u)$ are second-order continuously differentiable with uniformly bounded derivatives.

The conditions in Assumption 4 are primarily used to resort to a general result of Escanciano and Song (2007) which Theorem 1 below relies on. A condition similar to Condition (i)(a) was used by Stute and Zhu (2005). Condition (i)(b) says that the distribution function of $\lambda_\theta(X)$ should be Lipschitz continuous with a uniformly bounded Lipschitz coefficient. Condition (i)(c) requires Lipschitz continuity of conditional density of $(Y, X)$, given $F_\theta(\lambda_\theta(X))$ in the evaluating point of the conditioning variable.

**Theorem 1**: Suppose Assumptions 1-4 hold. Furthermore, assume that the basis function
Corollary 1: (i) (a) Let us define $p^K$ in (5) satisfies Assumption A1 in the Appendix. Then the following holds:

$$
\sup_{(y,z,u) \in S} \left| \hat{\nu}_n(y, z, u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left\{ \gamma_z(Z_i) - g_z^Z(U_i) \right\} \left\{ \gamma_y(Y_i) - g_y^Y(U_i) \right\} \right| = o_P(1). \tag{7}
$$

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 asymptotic power properties of a test constructed using $\hat{\nu}_n(r)$. Indeed, observe that the process

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

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. Let $D(S)$ denote the space of bounded functions on $S$ that is right-continuous and have left-limits, and we endow $D(S)$ with a uniform norm $|| \cdot ||_\infty$. The notation $\rightsquigarrow$ denotes weak-convergence in $D(S)$ in the sense of Hoffman-Jorgensen (e.g. van der Vaart and Wellner (1996)). 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 $U = u$.

Corollary 1: (i) (a) Under the null hypothesis, $\nu_n(r) \rightsquigarrow \nu(r)$ in $D(S)$, where $\nu$ 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 $C(w_1, w_2; u) = g_{z_1 \wedge z_2}(U) \{g_{y_1 \wedge y_2}(U) - g_{y_1}(U)g_{y_2}(U)\}$, $w_j = (z_j, y_j)$, $j = 1, 2$.

(b) Under fixed alternatives, $n^{-1/2} \nu_n(r) \rightsquigarrow \langle \gamma_u, \gamma_z(\gamma_y - g_y^Y) \rangle$ in $D(S)$.

(ii) (a) Under the null hypothesis, $\hat{\nu}_n(r) \rightsquigarrow \nu_1(r)$ in $D(S)$ where $\nu_1$ is a Gaussian process whose covariance kernel is given by

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

with $C_1(w_1, w_2; u) = \{g_{z_1 \wedge z_2}(U) - g_{z_1}(U)g_{z_2}(U)\} \{g_{y_1 \wedge y_2}(U) - g_{y_1}(U)g_{y_2}(U)\}$.
(b) Under fixed alternatives, \( n^{-1/2} \hat{\nu}_n(r) \rightsquigarrow \langle \gamma_u, \gamma_z(\gamma_y - g_Y^') \rangle \) in \( D(S) \).

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 fixed alternatives, the two empirical processes \( \nu_n(r) \) and \( \hat{\nu}_n(r) \) weakly converge to the same shift term after normalization by \( \sqrt{n} \), providing the basis for local power properties against \( \sqrt{n} \)-converging Pitman local alternatives. The result in Corollary 1 gives rise to a surprising consequence that using the nonparametrically estimated distribution function \( \hat{g}_Y^' (u) \) instead of the true one \( g_Y^' (u) \) 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 \( \hat{g}_Y^' (u) \) instead of \( g_Y^' (u) \) whereas the limit shift \( \langle \gamma_u, \gamma_z(\gamma_y - g_Y^') \rangle \) under the alternatives remains intact. In a later section, we formalize this observation.

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 \( \lambda_{\hat{\theta}} \) by \( \lambda_{\theta_0} \). A similar phenomenon is discovered and analyzed by 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 \( \{\lambda_{\hat{\theta}}(X_i)\}_{i=1}^n \).

We construct the following two sided test statistics:

\[
T_{KS} = \sup_{r \in S} |\hat{\nu}_n(r)| \quad \text{and} \quad T_{CM} = \int_S \hat{\nu}_n(r)^2 dr.
\]  

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,

\[
T_{KS} \rightsquigarrow \sup_{r \in S} |\nu(r)| \quad \text{and} \quad T_{CM} \rightsquigarrow \int_S \nu_1(r)^2 dr, \quad \text{in} \quad D(S)
\]  

and under the Pitman local alternatives \( P_n \) such that for some \( a_{y,z}(\cdot) \), \( E_n(\gamma_z(Z)(\gamma_y(Y) - g_Y^U)|U = u) = a_{y,z}(u)/\sqrt{n} + o(n^{-1/2}) \), we have

\[
\hat{\nu}_n(r) \rightsquigarrow \nu_1(r) + \langle \gamma_u, a_{y,z} \rangle, \quad \text{in} \quad D(S).
\]

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 testing procedure based on (9) is not feasible. In particular, we cannot tabulate
critical values from the limiting distribution, because the test statistics depend on unknown components of the data generating process and in consequence the tests are not asymptotically pivotal.

4 Conditional Martingale Transform

A recent work by the author (Song (2007)) proposes the method of conditional martingale transform which 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 \( \rho_{y,z} \triangleq \gamma_z(\gamma_y - g_y^z) \) here is indexed by \((y,z) \in \mathbb{R}^2\) whereas the generalized residual function \( \rho_{\tau,\theta} \) in Song (2007) does not depend on the index of the empirical process. Second, while Song (2007) 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 \( \lambda_{\theta_0}(X) \). Therefore, the nondegeneracy of the conditional distribution fails. Third, the instrumental variable \( \lambda_{\theta_0}(X) \) here depends on the unknown parameter \( \theta_0 \).

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.\(^4\)

4.1 Preliminary Heuristics

In this section, we provide some heuristics of the mechanics by which conditional martingale transforms work. 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

\footnote{By noting that the limiting processes \( \nu \) and \( \nu_1 \) 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 repeated martingale transforms of a function.}
product almost surely:

Conditional Orthogonality: \( \langle \gamma^K, 1 \rangle_U = 0 \) and \( \langle \gamma^K, 1 \rangle_U = 0 \), and

Conditional Isometry: \( \langle \gamma^K_1, \gamma^K_2 \rangle_U = \langle \gamma_1, \gamma_2 \rangle_U \) and \( \langle \gamma^K_1, \gamma^K_2 \rangle_U = \langle \gamma_1, \gamma_2 \rangle_U \).

An operator \( K \) that satisfies these two properties can be used to construct a transformed empirical process from which asymptotically pivotal 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 notation:

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

When we replace the indicator functions \( \gamma_z \) and \( \gamma_y \) by the transformed ones \( \gamma^K_z \) and \( \gamma^K_y \), the right-hand side asymptotic representation turns into the following (by the conditional orthogonality of the transform):

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \{ \gamma^K_z(Y_i) - \langle \gamma^K_z, 1 \rangle_U \} \{ \gamma^K_y(Y_i) - \langle \gamma^K_y, 1 \rangle_U \} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \gamma^K_z(Z_i) \gamma^K_y(Y_i).
\]

We can show that the last term is equal to

\[
\sqrt{n} \mathbb{E} [\gamma_u(U) \langle \gamma^K_z, \gamma^K_y \rangle_U] + \nu^K(r) + o_P(1)
\]

where \( \nu^K(r) \) 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 (11) becomes zero by the conditional independence restriction. As shown in the appendix, the Gaussian process \( \nu^K(r) \) has a covariance function of the form

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

This form is obtained by using the conditional isometry of the operator \( K \) 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 \( P_1 \) such that \( P_1\{|\langle \gamma^K_z, \gamma^K_y \rangle_U| > 0\} > 0 \), 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 (10). Consider the asymptotic representation of the empirical process $\nu_n(r)$ that uses the true $\gamma_y(U_i)\gamma_z(Y_i)$:

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

When we apply the transform $K$ to the indicator functions of $\gamma_y$ and $\gamma_z$, the right-hand side is equal to

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \gamma_z^K(Y_i) \gamma_y^K(Y_i) + o_P(1).$$

Hence the asymptotic representation of the process after the transform $K$ coincides with that in (10). The asymptotic power properties remain the same regardless of whether we use $g^Y(u)$ or $\hat{g}_Y(u)$ once the transform $K$ is applied.

4.2 Conditional Martingale Transforms

In this section, we discuss a method to obtain the operator $K$ 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 (2007) 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 pivotal. Just as the isometric projection is called a martingale transform, Song (2007) calls the conditional isometric projection a conditional martingale transform. See Song (2007) for the details.

By applying the conditional martingale transform in Song (2007) to our indicator functions, we can obtain the transform $K$ as follows:

$$\gamma_y^K(U, Y) \triangleq \gamma_y(Y) - \mathbb{E} [\gamma_y(Y) h_y(U, Y) | U = u]_{\hat{g} = Y} \quad \text{and}$$

$$\gamma_z^K(U, Z) \triangleq \gamma_z(Z) - \mathbb{E} [\gamma_z(Z) h_z(U, Z) | U = u]_{\hat{z} = Z}. \quad (13)$$
where

\[ h_y(U, Y) \triangleq \frac{1\{Y \leq \bar{y}\}}{1 - F_{Y|U}(Y|U)} \quad \text{and} \quad h_z(U, Z) \triangleq \frac{1\{Z \leq \bar{z}\}}{1 - F_{Z|U}(Z|U)}. \quad (14) \]

Observe that if \( h_y(U, Y) \) and \( h_z(U, Z) \) were taken to be a constant 1, the operator \( K \) is just an orthogonal projection onto the orthogonal complement of a constant function in \( L_2(P) \). The factors \( h_y(U, Y) \) and \( h_z(U, Z) \) as defined in (14) render the transforms isometric.

As compared to the conditional martingale transform in Song (2007), the transform in this paper corresponds to the case of function \( q_0 \) in Song (2007) being equal to 1 and hence is simpler. 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.

The following result shows that the transform \( K \) defined in (13) is a martingale transform that we have sought for. Eventually, we focus on test statistics based on the following two processes \( \nu_n^K(r) \) and \( \nu_{n,E}^K(r) \) defined by

\[
\nu_n^K(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i) \gamma_z^K(Z_i) \gamma_y^K(Y_i) \quad \text{and} \\
\nu_{n,E}^K(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i) \gamma_z^K(Z_i) (\gamma_y^K(Y_i) - \hat{g}_y^K(\hat{U}_i)),
\]

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

**Theorem 2 :** (i) For all \( y_1, y_2, z_1 \) and \( z_2 \) in \( \mathbf{R} \),

\[
\langle \gamma_{y_1}^K, 1 \rangle_U = 0 \quad \text{and} \quad \langle \gamma_{y_1}^K, \gamma_{y_2}^K \rangle_U = \langle \gamma_{y_1}, \gamma_{y_2} \rangle_U \quad \text{a.s. and} \\
\langle \gamma_{z_1}^K, 1 \rangle_U = 0 \quad \text{and} \quad \langle \gamma_{z_1}^K, \gamma_{z_2}^K \rangle_U = \langle \gamma_{z_1}, \gamma_{z_2} \rangle_U \quad \text{a.s.}
\]

(ii) **Suppose Assumptions 1-4 hold.**

(a) Under \( H_0 \), the processes \( \nu_n^K(r) \) and \( \nu_{n,E}^K(r) \) defined in (15) satisfy that

\[
\nu_n^K(r) \rightsquigarrow \nu^K(r) \quad \text{and} \quad \nu_{n,E}^K(r) \rightsquigarrow \nu^K(r), \quad \text{in } D(S), \quad (16)
\]

where \( \nu^K \) is a Gaussian process on \( S \) whose covariance kernel is given by

\[
c_3(r_1; r_2) \triangleq \int_{-\infty}^{u_1 \wedge u_2} C_3(w_1, w_2; u) dF(u), \quad r_j = (u, w_j), \quad j = 1, 2,
\]

and the function \( C_3(w_1, w_2; u) \) is defined by \( C_3(w_1, w_2; u) \triangleq g_{z_1 \wedge z_2}^u(u) g_{y_1 \wedge y_2}^u(u) \).
(b) Under the fixed alternatives,

\[ n^{-1/2} \nu_n^K(r) \sim E[\gamma_u(U)(\gamma^K_u, \gamma^K_Z)_U] \quad \text{and} \quad n^{-1/2} \nu_{n,E}^K(r) \sim E[\gamma_u(U)(\gamma^K_u, \gamma^K_Y)_U], \quad \text{in } D(S), \]

where \( \gamma^K_u \) and \( \gamma^K_Y \) are defined in (13).

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 processes \( \nu_n^K(r) \) and \( \nu_{n,E}^K(r) \).

The limiting Gaussian process \( \nu^K(r) \) is a time-transformed Brownian sheet whose kernel still depends on the unknown nonparametric functions \( g_Z(z) \) and \( g_Y(y) \). Following a similar idea in Khmaladze (1993), we can turn the process into a standard Brownian sheet by replacing \( \gamma^K_Z(u, \bar{z}) \) and \( \gamma^K_Y(u, \bar{y}) \) by

\[
\tilde{\gamma}^K_Z(u, \bar{z}) \triangleq \frac{\gamma^K_Z(u, \bar{z})}{\sqrt{f_{Z|U}(\bar{z}|u)}} \quad \text{and} \quad \tilde{\gamma}^K_Y(u, \bar{y}) \triangleq \frac{\gamma^K_Y(u, \bar{y})}{\sqrt{f_{Y|U}(\bar{y}|u)}},
\]

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

\[ c(r_1, r_2) \triangleq (u_1 \wedge u_2)(z_1 \wedge z_2)(y_1 \wedge y_2), \quad r_1, r_2 \in S, \]

which is that of a standard Brownian sheet. Observe that the test based on this conditional martingale transform is not yet asymptotically pivotal because the range \( S \) of the standard Brownian sheet depends on the distribution of \( (Y, Z) \). This can be dealt with by rescaling the marginals of \( Y \) and \( Z \) appropriately. We will explain this later in a section devoted to the feasible conditional martingale transform.

### 4.3 Feasible Conditional Martingale Transforms

The method of conditional martingale transform proposed in the previous section is not yet feasible for two reasons. First the conditional martingale transform involves unknown components of the data generating process. Second, the range of index \( S \) is not known.

Suppose \( F_n^Y \) and \( F_n^Z \) are the empirical quantile transform of \( (Y_i)_{i=1}^n \) and \( (Z_i)_{i=1}^n \). Let us
define two functions in the following way:

\[ V^Y_i \triangleq F^Y_n(Y_i) \quad \text{and} \quad V^Z_i \triangleq F^Z_n(Z_i). \]  

(17)

We define a feasible transform as follows. Let us define \( \hat{P} \) as in (5) and let

\[
\beta(y, \bar{y}) \triangleq \begin{bmatrix}
1\{c_0 \leq V^Y_1 \leq y \wedge \bar{y}\}/ \left\{ \sqrt{\hat{f}(V^Y_1, \hat{U}_1)[1 - p^K(\hat{U}_1)\hat{\pi}(V^Y_1)]} \right\} \\
\vdots \\
1\{c_0 \leq V^Y_n \leq y \wedge \bar{y}\}/ \left\{ \sqrt{\hat{f}(V^Y_n, \hat{U}_n)[1 - p^K(\hat{U}_n)\hat{\pi}(V^Y_n)]} \right\}
\end{bmatrix}
\]

(18)

where \( \hat{\pi}(y) \triangleq [\hat{P}^{\prime} \hat{P}]^{-1}\hat{P}^{\prime}b(y) \), \( b(y) \) being the column vector of \( i \)-th entry equal to \( 1\{0 < V^Y_i \leq y\} \), and \( \hat{f}(v, u) \) denotes a nonparametric estimator of the joint density of \( V^Y_i \) and \( U_i \).

Thus the feasible transform is obtained by

\[
(\gamma^K_y)(u, \bar{y}) = \frac{1\{\bar{y} \leq y\}}{\sqrt{\hat{f}(\bar{y}, u)}} - p^K(u)\hat{P}^{\prime} [\hat{P}^{\prime} \hat{P}]^{-1}\hat{P}^{\prime} \beta(y, \bar{y}).
\]

(19)

We can similarly obtain \( (\gamma^K_z)(u, \bar{z}) \) by exchanging the roles of \( V^Y_i \) and \( V^Z_i \). An estimated version of the transformed process is now defined by

\[
\hat{\nu}^K_n(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i)\gamma^K_z(\hat{U}_i, V^Z_i)\gamma^K_y(\hat{U}_i, V^Y_i).
\]

Due to the random rescale of \( Y_i \) and \( Z_i \) in (17), the process is indexed by \( r \in S_0 \), where \( S_0 \triangleq [0, 1)^2 \times [0, 1] \). Hence the dependence of the process upon the unknown support \( S \) has disappeared.

We can construct the transformed two-sided test statistics in the following way:

\[
T^K_{KS} = \sup_{r \in S_0} |\hat{\nu}^K_n(r)| \quad \text{and} \quad T^K_{CM} = \int_{S_0} |\hat{\nu}^K_n(r)|^2 dr.
\]

(20)

In the context of testing semiparametric conditional moment restrictions, Song (2007) delineated conditions under which the feasible conditional martingale transforms are asymptotically valid. We can establish the asymptotic validity of the feasible conditional martingale

\[ \text{Note that since } U_i \text{ is uniformly distributed, we have } f(v|u) = f_{V^Y_i, U}(v, u) \text{ where } f_{V^Y_i, U}(v, u) \text{ is a joint-density of } V^Y_i \text{ and } U_i. \]  Hence we can use simply the estimator of \( f_{V^Y_i, U}(v, u) \) instead of \( f(v|u) \).
transforms in a similar manner, so that we obtain, combined with Theorem 2,

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

where \( W \) is a standard Brownian sheet on \( S \). We will sketch the derivation of asymptotic validity in the appendix.

When we are interested in testing the null of conditional independence against alternatives of conditional positive dependence, we can construct the following test statistics

\[ T_{KS,1}^K = \sup_{r \in S_0} \hat{\gamma}_n(r) \to_d \sup_{r \in S_0} W(r) \text{ and } \]
\[ T_{CM,1}^K = \int_{S_0} |\max\{\hat{\gamma}_n(r), 0\}|^2 dr \to_d \int_{S_0} |\max\{W(r), 0\}|^2 dr. \]

### 5 When \( Z_i \) is a Binary Variable

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 \( P\{Z = 1|X\} \). The application of conditional martingale transform in this case becomes even simpler. First, we do not need to transform the indicator function of \( Z \). 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

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

for all \((u, y, z) \in [0, 1]^2 \times \{0, 1\}\). From the binary character of \( Z \), we observe that the null hypothesis is equivalent to\(^6\)

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

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 \([0, 1]^2\) instead of \([0, 1]^2 \times \{0, 1\}\). The corresponding feasible empirical process in the test

\(^6\)Equivalently, we can write the null hypothesis as \( E[\gamma_u(U)(1 - Z)\{\gamma_y(Y) - E[\gamma_y(Y)|U]\}] = 0. \)
statistic before the martingale transform are given by
\[ \hat{\nu}_n(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i) \left(Z_i \{ \gamma_y(Y_i) - \hat{E}[\gamma_y(Y_i)|\hat{U}_i] \} \right), \quad (u, y) \in [0, 1]^2. \]

Using similar arguments in the proof of Theorem 1 and assuming regularity conditions for the estimator \( \hat{P}\{Z = 1|U_i\} \), we obtain that
\[ \hat{\nu}_n(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \left(Z_i - \hat{E}[Z_i|U_i]\right) \{ \gamma_y(Y_i) - \hat{E}[\gamma_y(Y_i)|U_i] \} + o_P(1). \]

Hence, it turns out that in this case of binary \( Z \), we have only to take a conditional martingale transform on the indicator function of \( \gamma_y \). Let us consider the following three martingale transformed processes:

\[ \hat{\nu}_{n,\alpha}(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \gamma_u(\hat{U}_i) \left(\hat{\gamma}_y(\hat{U}_i, Y_i) \right), \]

(21)
\[ \hat{\nu}_{n,\beta}(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - Z_i) \gamma_u(\hat{U}_i) \left(\hat{\gamma}_y(\hat{U}_i, Y_i) \right), \]

and
\[ \hat{\nu}_{n,\gamma}(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \gamma_u(\hat{U}_i) \left\{ \left(\hat{\gamma}_y(\hat{U}_i, Y_i) - \hat{E}[\hat{\gamma}_y(\hat{U}_i, Y_i)|\hat{U}_i] \right) \right\} \left(\hat{\gamma}_y(\hat{U}_i, Y_i) \right) \left(\hat{\gamma}_y(\hat{U}_i, Y_i) \right). \]

Then in view of the previous results, these processes will weakly converge in \( D([0, 1] \times [0, 1]) \) to a standard Brownian sheet under the null hypothesis, i.e., a Gaussian process with covariance function:
\[ c(u_1, y_1; u_2, y_2) \triangleq (u_1 \wedge u_2)(y_1 \wedge y_2). \]

In a next subsection, we compare the asymptotic power properties of tests based on the three processes.

Let us consider the concrete procedure of constructing a test statistic using a feasible conditional martingale transform. For brevity, we consider only the processes \( \hat{\nu}_{n,\alpha}(u, y) \) and \( \hat{\nu}_{n,\beta}(u, y) \). Let \( \hat{P}(Z_i = 1|\hat{U}_i) \) denote a nonparametric estimator of \( P\{Z_i = 1|U_i\} \) and let \( \hat{\gamma}_y \)
be as defined in (19). Then we define the processes
\[
\nu^K_{na}(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i 1\{\hat{U}_i \leq u\} \hat{\gamma}^{K}_{y}(\hat{U}_i, V^Y)}{\sqrt{\hat{P}(Z_i = 1|\hat{U}_i)}}
\]
and
\[
\nu^K_{nb}(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - Z_i) 1\{\hat{U}_i \leq u\} \hat{\gamma}^{K}_{y}(\hat{U}_i, V^Y) \sqrt{\hat{P}(Z_i = 0|\hat{U}_i)}.
\]

We can construct the following three Kolmogorov-Smirnov type test statistics:
\[
T^K_{1a} = \sup_{(u,y) \in [0,1] \times [0,1]} \left| \nu^K_{1na}(u, y) \right|
\]
\[
T^K_{1b} = \sup_{(u,y) \in [0,1] \times [0,1]} \left| \nu^K_{1nb}(u, y) \right|
\]
\[
T^K_{1S} = \sup_{(u,y) \in [0,1] \times [0,1]} \left\{ \left| \nu^K_{1na}(u, y) \right| + \left| \nu^K_{1nb}(u, y) \right| \right\} / \sqrt{2}.
\]

The three tests have the same asymptotic critical values. However, as we shall see, their asymptotic power properties are quite different as we can see in the next section and the section devoted to simulation studies. The asymptotic power of \(T^K_{1a}\) can be large or small depending on whether \(Z\) and \(Y\) conditionally positively or negatively dependent given \(U\). The symmetrized version \(T^K_{1S}\) does not show such asymmetric power properties depending on the direction of alternatives.

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.

<table>
<thead>
<tr>
<th>significance level</th>
<th>0.5</th>
<th>0.25</th>
<th>0.20</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>critical values</td>
<td>1.46</td>
<td>1.81</td>
<td>1.91</td>
<td>2.21</td>
<td>2.46</td>
<td>2.70</td>
<td>3.03</td>
</tr>
</tbody>
</table>

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. Therefore, one reject the null hypothesis of conditional independence at the level of 0.05 if \(T_n > 2.46\).

We can similarly construct one-sided Kolmogorov-Smirnov type test statistics:
\[
T^K_{1a,1} = \sup_{(u,y) \in [0,1] \times [0,1]} \nu^K_{1na}(u, y)
\]
\[
T^K_{1b,1} = \sup_{(u,y) \in [0,1] \times [0,1]} \nu^K_{1nb}(u, y).
\]
The statistic $T_{K_{a1}}$ is for testing conditional independence against the positive conditional dependence given $\lambda_\theta(X)$ and the test statistic is $T_{K_{b1}}$ is for testing conditional independence against the negative conditional dependence.

6 Asymptotic Power Analysis

In this section, we embark on a formal comparison of a variety of tests in which we analyze the effects of the nonparametric estimation error and the conditional martingale transform upon the asymptotic power of the tests. The basic tool of comparison that we employ is the liming Pitman efficiency which we can compute without difficulty due to the well-known results of Bahadur (1960) and Wieand (1976). In particular, Wieand (1976) specified sufficient conditions under which the limiting Pitman efficiency for asymptotically non-normal tests can be computed via the approximate Bahadur efficiency (Bahadur (1960)).

First, we find that the nonparametric estimation error improves the asymptotic power of the test. This finding may strike one as counter-intuitive because it is tantamount to saying that even when perfect knowledge of the nonparametric function is available, it is better to use its estimator. A similar observation was made in estimation by Hahn (1998) and Hirano, Imbens, and Ridder (2003). We provide some discussions regarding this phenomenon later.

Second, we analyze the effect of the conditional martingale transform upon the asymptotic power properties. We find that when both $Y$ and $Z$ are continuous, the improvement of asymptotic power by nonparametric estimation error disappears under the conditional martingale transform, but when either $Y$ or $Z$ is binary, the improvement is partially preserved even under the conditional martingale transform. We also characterize the set of local alternatives in which the conditional martingale transform improves the asymptotic power of the tests.

6.1 Limiting Pitman Efficiency

First, we define the limiting Pitman efficiency as follows (c.f. Nikitin (1995), p.2-3.). Suppose that $\mathcal{P}$ is the space of probabilities $P$ for $(Y, Z, X)$, and we are given a sequence of test statistics $T_n$ for testing the null of conditional independence. We choose a parametric submodel in $\mathcal{P}$ and define the limiting Pitman efficiency under the parametric submodel. And then, we demonstrate that certain bounds of limiting Pitman efficiency do not depend on the specific parametrization.

We first consider a parametric submodel \( \{P_\tau \in \mathcal{P} : \tau \in [0, \tau_0]\} \) for some $\tau_0 > 0$ in which $P_0$ represents a probability under the null hypothesis and $P_\tau$ with $\tau \in (0, \tau_0]$ represents a
probability under the alternative hypothesis. Then, for any $\beta \in (0, 1)$ and $\tau \in [0, \tau_0]$, a real sequence $c_{T_n}(\beta, \tau)$ is chosen to be such that

$$P_\tau\{T_n > c_{T_n}(\beta, \tau)\} \leq \beta \leq P_\tau\{T_n \geq c_{T_n}(\beta, \tau)\}.$$ 

Then, the number $\alpha_{T_n}(\beta, \tau) \triangleq P_0\{T_n \geq c_{T_n}(\beta, \tau)\}$ denotes the minimal size of the test based on $\{T_n\}$ such that the power at $P_\tau$ is not less than $\beta$. For any given two test statistics $T_n$ and $S_n$, the relative efficiency of $T_n$ with respect to $S_n$ is defined as

$$e_{T,S}(\alpha, \beta, \tau) = \min\{n : \alpha_{S_n}(\beta, \tau) \leq \alpha \text{ for all } m \geq n\}.$$ 

When $e_{T,S} > 1$, the test based on $\{T_n\}$ is preferable to that based on $\{S_n\}$ because the required sample size to achieve power $\beta$ at $P_\tau$ with the given level $\alpha$ is smaller for the test $T_n$ than that for the test $S_n$. In general, the exact computation of $e_{T,S}(\alpha, \beta, \tau)$ is extremely difficult in many cases, and there has been proposed a variety of notions of asymptotic relative efficiency. In this paper, we focus on the limiting Pitman efficiency. For other notions, the reader is referred to Nikitin (1995).

Following Wieand (1976), we define the upper and lower Pitman efficiencies as follows:

$$e_{T,S}^+(\alpha, \beta) = \sup\{\limsup_{j \to \infty} e_{T,S}(\alpha, \beta, \tau_j) : \{\tau_j\} \in T_1, \tau_j \downarrow 0\}$$

$$e_{T,S}^- (\alpha, \beta) = \inf\{\liminf_{j \to \infty} e_{T,S}(\alpha, \beta, \tau_j) : \{\tau_j\} \in T_1, \tau_j \downarrow 0\},$$

where $T_1$ denotes the space of sequences $\{\tau_j\}_{j \in \mathbb{N}_+}$, $\tau_j \in (0, \tau_0]$. When $\lim_{\alpha \to 0} e_{T,S}^+(\alpha, \beta)$ and $\lim_{\alpha \to 0} e_{T,S}^- (\alpha, \beta)$ exist and coincide, we call the common number the limiting Pitman efficiency and denote it by $e_{T,S}^*(\beta)$. This limiting Pitman efficiency is the criterion that we employ to compare the tests. When the test statistics are asymptotically normal, Bahadur (1960) showed that the limiting Pitman efficiency coincides with the limit of the approximate Bahadur efficiency as $\tau \downarrow 0$. Wieand (1976) showed that such coincidence is preserved even under a more general setting once a stronger condition on the asymptotic behavior of the test statistic under the local alternatives is satisfied. In this latter case, the limiting Pitman efficiency is independent of $\beta$ and hence we simply denote it by $e_{T,S}^*$.

### 6.2 The Effect of Nonparametric Estimation Error

We consider the situation of Section 3. We analyze how our use of a nonparametric estimator of $g_y(u)$ in the process $\hat{\nu}_n(r)$ affects the power of the test. 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 \( \hat{g}_Y(u) \) even if we know the true conditional distribution function \( g_Y(u) \).

As for the stochastic processes \( \nu_n(r) \) and \( \hat{\nu}_n(r) \) defined in (3) and (6), define

\[
T_n = \sup_{r \in S} |\nu_n(r)| \quad \text{and} \quad S_n = \sup_{r \in S} |\hat{\nu}_n(r)|,
\]

and denote \( e_{T,S}(\beta) \) to be the limiting Pitman efficiency of \( T_n \) with respect to \( S_n \). Let \( g_Z(\cdot;\tau) \) and \( g_Y(\cdot;\tau) \) denote the conditional distribution functions of \( Z \) and \( Y \) given \( U \) under \( P_\tau \). Then, we obtain the following result.

**Corollary 2**: Suppose the conditions of Theorem 1 hold. Then

\[
e_{T,S}^* = \lim_{\tau \to 0} \sup_{(z,y) \in \mathbb{R}^2} \int_0^1 \left( g_Z^2(u;\tau) - g_Z^2(u;\tau) \right) \left( g_Y^2(u;\tau) - g_Y^2(u;\tau) \right) du \leq \frac{1}{4},
\]

if the limit above exists.

The result in Corollary 2 is well expected from the result of Theorem 1. The main reason is because while the nonparametric estimator \( \hat{g}_Y(u) \) in place of \( g_Y(u) \) has an effect of "shrinking" the kernel of the limiting Gaussian process, the local shift of the test statistics under the local alternatives remains the same regardless of whether one uses \( g_Y(u) \) or \( \hat{g}_Y(u) \). This yields the counter-intuitive result that even when we know the conditional distribution function \( g_Y(u) \), it is better to use its estimator \( \hat{g}_Y(u) \) in the test statistic.\(^7\)

In the literature of estimation, 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.

The primary source of the improvement of the asymptotic power in Corollary 2 stems from the following asymptotic representation (for a general result, see Lemma 1U of Escanciano and Song (2007))

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_u(U_i) \gamma_z(Z_i) \{ \gamma_y(Y_i) - \hat{g}_y(U_i) \} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \gamma_u(U_i) \{ \gamma_z(Y_i) - g_z(U_i) \} \{ \gamma_y(Y_i) - g_y(U_i) \} + o_P(1).
\]

\(^7\)Khmaladze and Koul (2004) also discuss the case when the power is increased by using estimators instead of true ones in the context of parametric tests.
As noted in the remarks after Theorem 1, the variance of the leading sum on the right-hand side is smaller than that of \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) g^Z_i(Z_i) \{ \gamma_y(Y_i) - g^Y_y(U_i) \} \) under the null hypothesis. In fact, when we replace \( \gamma_z(Z) \) by \( g^Z_i(U) \) on the left-hand side, we obtain the following result

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) g^Z_i(U_i) \{ \gamma_y(Y_i) - g^Y_y(U_i) \} = o_p(1). \tag{22}
\]

This fact provides a contrast to a result obtained using the true nonparametric function \( g^Y_y \):

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) g^Z_i(U_i) \{ \gamma_y(Y_i) - g^Y_y(U_i) \} = O_p(1). \tag{23}
\]

The rate of convergence for the sample version of \( \mathbb{E} [ \gamma_u(U) g^Z_i(U) \{ \gamma_y(Y) - g^Y_y(U) \} ] \) is faster with the nonparametric estimator \( \hat{g}^Y_y \) than the true one. Actually we can obtain the same contrast between (22) and (23) when we replace \( \gamma_u(U) g^Z_i(U) \) by any function \( \varphi(U) \). In other words, by using the nonparametric estimation in the test statistic, we utilize the conditional moment restriction \( \mathbb{E} [ \gamma_y(Y) - g^Y_y(U) | U] = 0 \) more effectively than by using the true function \( g^Y_y \). 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 which is dominated by the \( \sqrt{n} \)-diverging shift term that remains the same regardless of whether we use \( g^Y_y \) or \( \hat{g}^Y_y \). 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 \( g^Y_y \).

### 6.3 The Effect of Nonparametric Estimation Error After the Martingale Transform

In this section, we analyze the effect of nonparametric estimation error for the test statistics under the conditional martingale transforms. We divide the discussion into the case of \( Y \) and \( Z \) being continuous and the case of either \( Y \) or \( Z \) being binary.

#### 6.3.1 When \( Y \) and \( Z \) Are Continuous

We consider the situation of Section 4.2. Recall the definitions of the processes \( \nu^K_n(r) \) and \( \nu^K_{n,E}(r) \) in (15). Now, we compare the two tests defined by

\[
T_n = \sup_{r \in S} |\nu^K_n(r)| \quad \text{and} \quad S_n = \sup_{r \in S} |\nu^K_{n,E}(r)|,
\]

24
and denote \( e^*_{T,S} \) to be the limiting Pitman efficiency of \( T_n \) with respect to \( S_n \). The test \( T_n \) uses the estimated version \( \hat{g}^K_y(U) = \hat{E}[\gamma^K_y(Y)|U] \) and the test \( S_n \) uses the true one \( g^K_y(U) = E[\gamma^K_y(Y)|U] \) which is zero by Theorem 2(i). Then we obtain the following result.

**Corollary 3:** Suppose the conditions of Theorems 1 and 2 hold. Then we have

\[ e^*_{T,S} = 1. \]

The result of Corollary 3 shows that the effect of nonparametric estimation error disappears under the conditional martingale transform. This was expected from the result of Theorem 2 in which we saw that the asymptotic behavior of the two tests \( T_n \) and \( S_n \) are the same both under the null hypothesis and the local alternatives. The main reason is that the conditional martingale transform completely wipes out the additional term due to the nonparametric estimation error that is responsible for the improvement in asymptotic power. Therefore, under the conditional martingale transform, one may prefer the simpler test statistic \( S_n \) than \( T_n \) which involves the nonparametric estimator \( \hat{g}^K_y(U) \).

### 6.3.2 When Either \( Y \) or \( Z \) are Binary

In this subsection, we consider the situation of Section 5. We noted 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 \( \gamma^K_y \) and thereby, wiping out the effect of nonparametric estimation error only partially.

To substantiate these remarks, let us analyze the behavior of the two processes \( \tilde{\nu}^K_{na}(u, y) \) and \( \tilde{\nu}^K_{n,E}(u, y) \) in (21) under a fixed alternative. Based on Theorems 1 and 2, we can easily show that

\[
\begin{align*}
n^{-1/2} \tilde{\nu}^K_{na}(u, y) & \rightarrow_p \varphi_0(u, y; \tau) \triangleq E_\tau \left[ \frac{Z \gamma_y(U) (\tilde{\gamma}^K_y)(U, Y)}{\sqrt{P_\tau\{Z = 1|U\}}} \right] \quad \text{and} \\
n^{-1/2} \tilde{\nu}^K_{n,E}(u, y) & \rightarrow_p \varphi_1(u, y; \tau) \triangleq E_\tau \left[ \frac{Z \gamma_y(U) (\tilde{\gamma}^K_y)(U, Y)}{\sqrt{P_\tau\{Z = 1|U\} - P_\tau\{Z = 1|U\}^2}} \right]
\end{align*}
\]

where \( E_\tau \) denotes the expectation under \( P_\tau \). Since the denominator of the shift term for \( \tilde{\nu}^K_{na}(u, y) \) dominates that of the shift term for \( \tilde{\nu}^K_{n,E}(u, y) \), the test based on the process

---

*8* The second statement follows by the law of iterated conditional expectations and the orthogonality conditions in Theorem 2(i).
\( \tilde{\nu}_{n,E}(u,y) \) has a better local asymptotic power than that based on \( n^{-1/2} \nu_{na}(u,y) \). This also shows the interesting phenomenon that the test has a better power when we use the non-parametric estimator for \( E [\hat{\gamma}_y(U,Y)|U] \) instead of the true one. As before, we can confirm this phenomenon in terms of the limiting Pitman efficiency. Define now

\[
T_n = \sup_{(u,y) \in [0,1] \times \mathbb{R}} |\tilde{\nu}_{n,E}(u,y)| \quad \text{and} \quad S_n = \sup_{(u,y) \in [0,1] \times \mathbb{R}} |\nu_{na}(u,y)|,
\]

and denote \( e^*_{T,S} \) to be the limiting Pitman efficiency of \( T_n \) with respect to \( S_n \).

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

\[
e^*_{T,S} = \lim_{\tau \downarrow 0} \left( \frac{\sup_{(u,y) \in [0,1] \times \mathbb{R}} |\varphi_1(u,y;\tau)|}{\sup_{(u,y) \in [0,1] \times \mathbb{R}} |\varphi_0(u,y;\tau)|} \right)^2 \leq 1,
\]

if the limit above exists.

The result of the improved power property in Corollary 4 is demonstrated in the simulation studies. Note that the local shift \( E [Z\gamma_y(U)(\hat{\gamma}_y(U,Y))/\sqrt{P\{Z = 1\}}] \) in (24) can be large or small depending on whether \( Z \) and \( Y \) are conditionally positively or negatively dependent. Hence the asymptotic power properties can be asymmetric depending on the alternatives. The symmetrized version \( T_{15}^E \) removes part of this asymmetry.

### 6.4 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 the limiting Pitman efficiency. In this subsection, we compare two tests based on the processes \( \tilde{\nu}_n(r) \) and \( \nu^E_n(r) \) defined in (6) and (15). 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 into account the trade-off between these two changes. To facilitate the analysis, we consider the following specification of \( Y \):

\[
Y = \xi(Z,\varepsilon),
\]

where \( \xi(z,\varepsilon) : \mathbb{R}^2 \to \mathbb{R} \) and \( \varepsilon \) is a random variable taking values in \( \mathbb{R} \) and is conditionally independent of \( Z \) given \( U \). In the following we give the limiting Pitman efficiency of
Kolmogorov-Smirnov tests based on the processes $\nu_{1n}(r)$ and $\nu_{n}^{K}(r)$ 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

$$C_{Z,Y|U}(z,y|U;\tau) \triangleq P_{\tau}\{g_{z}^{Y}(U;\tau) \leq z, g_{Y}^{Y}(U;\tau) \leq y|U\}.$$  

The conditional copula $C_{Z,Y|U}(z,y|U;\tau)$ is the conditional joint distribution function of $Z$ and $Y$ under $P_{\tau}$ 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). Define

$$T_{n} = \sup_{r \in S_{0}}|\hat{\nu}_{n}(r)| \text{ and } S_{n} = \sup_{r \in S_{0}}|\nu_{n}^{K}(r)|,$$

and denote $e_{T,S}^{*}$ to be the limiting Pitman efficiency of $T_{n}$ with respect to $S_{n}$.

**Corollary 5:** 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 $\varepsilon$ or strictly decreasing in $z$ for all $\bar{\varepsilon}$ in the support of $\varepsilon$. Then

$$e_{T,S}^{*} = \lim_{\tau \to 0} \frac{\sup_{(u,z,y) \in S} \left| \int_{0}^{u} \{C_{Z,Y|U}(g_{z}^{z}(u;\tau), g_{y}^{Y}(u;\tau)|u;\tau) - g_{z}^{z}(u;\tau)g_{y}^{Y}(u;\tau)\} du\right|^{2}}{\sup_{(z,y) \in \mathbb{R}^{2}} \int_{0}^{u} (g_{z}^{z}(u;\tau) - g_{z}^{z}(u;\tau)^{2})(g_{y}^{Y}(u;\tau) - g_{y}^{Y}(u;\tau)^{2})du},$$

if the limit above exists. In particular, when $g_{z}^{z}(u;\tau)$ and $g_{y}^{Y}(u;\tau)$ do not depend on $u$ for all $\tau \in [0,\varepsilon)$ for some $\varepsilon > 0$,

$$e_{T,S}^{*} \leq \frac{1}{4}.$$  

Corollary 5 provides a representation of the limiting Pitman efficiency $e_{T,S}^{*}$. When $g_{z}^{z}(u;\tau)$ and $g_{y}^{Y}(u;\tau)$ do not depend on $u$ on a neighborhood of $\tau = 0$, so that the test of conditional independence becomes test of independence, it follows that $e_{T,S}^{*} \leq 1/4$, and hence in this case, the conditional martingale transform weakly improves the asymptotic power as long as $\xi(z,\bar{\varepsilon})$ satisfies the condition of Corollary 5. However, in general, we cannot say whether either of the two tests dominate the other.
7 Simulation Studies

7.1 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 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 \( \beta \) is set to be 0.5. The variable \( X_i \) is drawn from the uniform distribution on \([-1, 1]\) and the errors \( \eta_i \) and \( \varepsilon_i \) are independently drawn from \( N(0, 1) \). Note that \( \kappa = 0 \) corresponds to the null hypothesis of conditional independence between \( Y \) and \( Z \) given \( X \). The magnitude of \( \beta \) 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. For both tests, we take a sample quantile transform of the marginals of \( Y_i \) and \( X_i \) to generate \( V_{Y,i} \) and \( \hat{U}_i \). The bootstrap procedure is as follows:

1. Take \((\hat{\varepsilon}_{y,i})_{i=1}^n \) defined by \( \hat{\varepsilon}_{y,i} = \gamma_y(Y_i) - \hat{E}(\gamma_y(Y_i)|\hat{U}_i) \).
2. 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 \( E(\omega_i) = 0 \), \( E(\omega_i^2) = 1 \).
3. Construct \( \gamma_{y,i}^* = \hat{E}(\gamma_y(Y_i)|\hat{U}_i) + \varepsilon_{y,i}^* \).
4. Using \((\gamma_{y,i}^*, \hat{U}_i)_{i=1}^n \), estimate the bootstrap conditional mean function \( \hat{E}(\gamma_{y,i}^*|\hat{U}_i) \).
5. Obtain the bootstrap empirical process \( \nu_{n}^*(u, y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\hat{U}_i)Z_i \left[ \gamma_{y,i}^* - \hat{E}(\gamma_{y,i}^*|\hat{U}_i) \right] \).

The bootstrap procedure is the same as that considered in Delgado and González-Manteiga (2001), p.1490, except that they did not take the quantile transform of \( X \). Based upon the bootstrap process, we can construct the bootstrap test statistics:

\[
T_{KS}^B = \sup_{(u,y) \in [0,1] \times \mathbb{R}} |\nu_{n}^*(u, y)| \quad \text{and} \quad T_{CM}^B = \int \int \nu_{n}^*(u, y)^2dudy.
\]

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 estimation than the bootstrap approach. The empirical size overall seems reasonable.

[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 $\kappa$. 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 $T^K_{1A}$ is shown to be disproportionate between alternatives of positive dependence (with $\kappa$ positive) and negative dependence (with $\kappa$ negative). This disproportionate behavior disappears in the case of its "symmetrized" version $T^K_{1B}$. Second, the power of tests based on $T^K_2$ is better than that based on $T^K_{1A}$ or $T^K_{1B}$. Recall that the test statistic $T^K_2$ involves the nonparametric estimator for $E[(\hat{\gamma}^K_y)(U_i,Y_i)|U_i] = 0$, whereas $T^K_{1A}$ and $T^K_{1B}$ 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]
7.2 The Effect of Nonparametric Estimation of $g_Y^Y(u)$

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 $\varepsilon$ follows a standard normal distribution and independent of $U$ and $Z$, and $Z$ is generated by

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

where $\eta$ 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 $\kappa$ 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 $g_Y^Y(u)$ as follows

$$g_Y^Y(u) = P\{Y \leq y|U = u, Z = 1\} P\{Z = 1\} + P\{Y \leq y|U = u, Z = 0\} P\{Z = 0\}$$

$$= \frac{1}{2} \left( \Phi \left( \frac{y - \alpha(1 - \kappa)u}{1 - \alpha(1 - \kappa)} \right) + \Phi \left( \frac{y - \alpha u}{1 - \alpha} \right) \right).$$

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

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

$$T_n^{INF} = \sup_{(u, y) \in [0, 1]^2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1\{U_i \leq u\}Z_i}{\sigma_2^2(U_i)} (1\{Y_i \leq y\} - \hat{g}_Y^Y(U_i)) \right|$$

and

$$T_n^{FEAS} = \sup_{(u, y) \in [0, 1]^2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1\{U_i \leq u\}Z_i}{\sigma_2^2(\hat{U}_i) - \sigma_2^2(\hat{U}_i)} (1\{Y_i \leq y\} - \hat{g}_Y^Y(U_i)) \right|.$$
from our analysis of power from the previous section, and reflects our theoretical observation that the use of $\hat{g}_y(Y_i)$ instead of $g_y(Y_i)$ improves the local power properties of the tests. This is also observed in terms of empirical processes inside $T_{n\INF}$ and $T_{n\FEAS}$ as shown in Figure 2.

[INSERT FIGURES 1 AND 2]

8 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 (2007), 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 Nishiyama (2000).) It seems that the results in this paper can be extended to time series context based on those results.

9 Appendix: Mathematical Proofs

9.1 Main Results

In this section, we present the proofs of the main result. Throughout the proofs, the notation $C$ denotes a positive absolute constant that may assume different values in different contexts in which it is used. We first prove Theorem 1. The following lemma is useful for this end.
Define
\[ \gamma_{u,n,\lambda}(x) \triangleq 1 \left\{ \frac{1}{n} \sum_{i=1}^{n} 1 \{ \lambda(X_i) \leq \lambda(x) \} \leq u \right\}. \]

The function \( \gamma_{u,n,\lambda}(\cdot) \) is indexed by \((u, \lambda) \in [0, 1] \times \Lambda_0 \) where \( \Lambda_0 \triangleq \{ \lambda_\theta(\cdot) : \theta \in B(\theta_0, \delta) \} \). 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:** Assume that \( \sup_{\lambda \in \Lambda_0} \sup_{X \in \mathbb{R}} f_\lambda(\bar{\lambda}) < \infty \), where \( f_\lambda(\cdot) \) denotes the density of \( \lambda(X) \). Then there exists a sequence of classes \( G_n \) of measurable functions such that

\[ P \{ \gamma_{u,n,\lambda} \in G_n \text{ for all } (u, \lambda) \in [0, 1] \times \Lambda_0 \} = 1, \]

and for any \( r \geq 1 \),

\[ \log N_{\| \cdot \| \cdot} (C_2 \varepsilon, G_n, \| \cdot \|_r) \leq \log N(\varepsilon^r / 2, \Lambda_0, \| \cdot \|_\infty) + \frac{C_1}{\varepsilon}, \]

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

**Proof of Lemma A1:** For a fixed sequence \( \{x_i\}_{i=1}^n \in \mathbb{R}^{nd_X} \), we define a function:

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

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

\[ G_n' \triangleq \{ g_{n,\lambda,u} : (\lambda, \{x_i\}_{i=1}^n, u) \in \Lambda_0 \times \mathbb{R}^{nd_X} \times [0, 1] \} \quad \text{and} \]
\[ G_n \triangleq \{ g \circ \lambda : (g, \lambda) \in G_n' \times \Lambda_0 \}. \]

Then we have \( \gamma_{u,n,\lambda} \in G_n \). 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 \( r \geq 1 \)
\[
\log N\{\varepsilon, G_n', ||\cdot||_{Q,r}\} \leq \frac{C_1}{\varepsilon},
\]
for any probability measure \( Q \) and for some constant \( C_1 > 0 \).

Without loss of generality, assume that \( \lambda(X) \in [0, 1] \), because otherwise, we can consider \( \Phi \circ \Lambda_0 = \{\Phi \circ \lambda : \lambda \in \Lambda_0\} \) in place of \( \Lambda_0 \), where \( \Phi \) is the distribution function of a standard normal variate. We choose \( \{\lambda_1, \ldots, \lambda_{N_2}\} \) such that for any \( \lambda \in \Lambda_0 \), there exists \( j \) with \( ||\lambda_j - \lambda||_\infty < \varepsilon^r/2 \). For each \( j \), we define \( \tilde{\lambda}_j(x) \) as follows.

\[
\tilde{\lambda}_j(x) = m\varepsilon^r/2 \text{ when } \lambda_j(x) \in [m\varepsilon^r/2, (m + 1)\varepsilon^r/2) \text{ for some } m \in \{0, 1, 2, \ldots, \lfloor 2/\varepsilon^r \rfloor\},
\]
where \( \lfloor z \rfloor \) denotes the greatest integer that does not exceed \( z \). Note that the range of \( \tilde{\lambda}_j \) is finite and \( ||\lambda - \tilde{\lambda}_j||_\infty \) is equal to
\[
\sup_x \sum_{k=0}^{\lfloor 2/\varepsilon^r \rfloor} \left| \lambda(x) - \frac{k\varepsilon^r}{2} \right| 1 \left\{ \lambda_j(x) \in \left[ \frac{k\varepsilon^r}{2}, \frac{(k + 1)\varepsilon^r}{2} \right) \right\} \leq \varepsilon^r.
\]
From (25), we have \( \log N\{\varepsilon, G_n', ||\cdot||_{Q,r}\} \leq C_1/\varepsilon \). We choose \( \{g_1, \ldots, g_{N_1}\} \) such that for any \( g \in G_n' \), there exists \( j \) such that \( \int_0^1 |g(\tilde{\lambda}) - g_j(\tilde{\lambda})|^r d\tilde{\lambda} < \varepsilon^r \). Then we have for any \( \lambda \in \Lambda_0 \),
\[
\mathbf{E}|g(\lambda(X)) - g_j(\lambda(X))|^r = \int_0^1 |g(\tilde{\lambda}) - g_j(\tilde{\lambda})|^r f_\lambda(\tilde{\lambda})d\tilde{\lambda} \leq \sup_{\lambda} f_\lambda(\tilde{\lambda})\varepsilon^r.
\]
Now, take any \( h \in G_n \) such that \( h \triangleright= g \circ \lambda \) and choose \( (g_{j_1}, \lambda_{j_2}) \) from \( \{g_1, \ldots, g_{N_1}\} \) and \( \{\lambda_1, \ldots, \lambda_{N_2}\} \) such that \( \int_0^1 |g(\tilde{\lambda}) - g_{j_1}(\tilde{\lambda})|^r d\tilde{\lambda} < \varepsilon^r \) and \( ||\lambda_{j_2} - \lambda||_\infty < \varepsilon^r/2 \). Then consider
\[
\left\| h - g_{j_1} \circ \tilde{\lambda}_{j_2} \right\|_r \leq \|g \circ \lambda - g_{j_1} \circ \lambda\|_r + \|g_{j_1} \circ \lambda - g_{j_1} \circ \tilde{\lambda}_{j_2}\|_r \leq \left( \sup_{\lambda} \sup_{\tilde{\lambda}} f_\lambda(\tilde{\lambda}) \right)^{1/r} \varepsilon + \left( \mathbf{E}\left|(g_{j_1} \circ \lambda)(X) - (g_{j_1} \circ \tilde{\lambda}_{j_2})(X)\right|^r \right)^{1/r}.
\]
The absolute value in the expectation of the second term is bounded by
\[
1 \left\{ \frac{1}{n} \sum_{i=1}^n 1 \left\{ \lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) - \varepsilon^r \right\} \leq u_{j_1} \right\} - 1 \left\{ \frac{1}{n} \sum_{i=1}^n 1 \left\{ \lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) + \varepsilon^r \right\} \leq u_{j_1} \right\},
\]

\[
33
\]
or by

\[
1 \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) - \varepsilon^r \right\} \leq u_{j_1} < \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \tilde{\lambda}_{j_2}(X) + \varepsilon^r \right\} \right\} \\
= \sum_{m=0}^{2/\varepsilon} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \frac{m\varepsilon^r}{2} - \varepsilon^r \right\} \leq u_{j_1} < \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \frac{m\varepsilon^r}{2} + \varepsilon^r \right\} \right\} \\
\times 1 \left\{ \frac{m\varepsilon^r}{2} \leq \lambda_{j_2}(X) < \frac{(m+1)\varepsilon^r}{2} \right\}.
\]

Since \( P \left\{ \lambda_{j_2}(X) \in \left[ m\varepsilon^r/2, (m+1)\varepsilon^r/2 \right] \right\} \leq \sup_{\lambda} \sup_{\lambda} f_\Lambda(\tilde{\lambda})\varepsilon^r \leq C\varepsilon^r \) by the mean-value theorem, and the sets

\[
A_m = \left\{ u \in [0, 1] : \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \frac{m\varepsilon^r}{2} - \varepsilon^r \right\} \leq u < \frac{1}{n} \sum_{i=1}^{n} \left\{ \lambda_{j_1}(x_i) \leq \frac{m\varepsilon^r}{2} + \varepsilon^r \right\} \right\}
\]

constitutes the partition of \([0, 1] \), we deduce that

\[
E \left\| (g_{j_1} \circ \lambda)(X) - (g_{j_1} \circ \tilde{\lambda}_{j_2})(X) \right\|_r^r \leq C\varepsilon^r.
\]

Therefore, \( \| h \circ g_{j_1} \circ \tilde{\lambda}_{j_2} \|_r \leq C\varepsilon \), yielding the result that

\[
\log N \left( C\varepsilon, \mathcal{G}_n, \| \cdot \|_r \right) \leq \log N \left( C\varepsilon^r, \Lambda_0, \| \cdot \|_{\infty} \right) + \log N \left( \varepsilon, \mathcal{G}_n, \| \cdot \|_r \right) \leq \log N \left( C\varepsilon^r, \Lambda_0, \| \cdot \|_{\infty} \right) + C/\varepsilon.
\]

\[\blacksquare\]

**Assumption A1:** (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 \( \{ \pi_y : y \in \mathbb{R} \} \) and \( \{ \pi_z : z \in \mathbb{R} \} \) such that

\[
\sup_{(y,u) \in \mathbb{R} \times [0,1]} | p^K(u)' \pi_y - g_y^Y(u) | = O(K^{-d_1}) \quad \text{and} \\
\sup_{(z,u) \in \mathbb{R} \times [0,1]} | p^K(u)' \pi_z - g_z^Z(u) | = O(K^{-d_2}),
\]

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

\[
\sup_{(y,\bar{u}) \in \mathbb{R} \times [0,1]} \left( \int_0^1 | p^K(u)' \pi_{y,\bar{u}} - \tilde{g}_y^Y(u) 1 \{ \bar{u} \leq u \} |^p du \right)^{1/p} = O(K^{-d_1}).
\]

34
where \( \hat{g}_y^Y(\cdot) \) denotes the first order derivative of \( g_y^Y(\cdot) \).

(iii) For \( d_1 \) and \( d_2 \) in (ii), \( \sqrt{n} \zeta_{0,K}^2 K^{-d} = o(1) \) and \( \sqrt{n} \zeta_{1,K}^{-d_2} = o(1) \).

(iv) For \( b \) in the definition of \( \Lambda_0 \), we have \( n^{1/2 - 2b} \zeta_{0,K}^3 = o(1) \), \( n^{-1/2 + 1/p} K^{1 - 1/p} \zeta_{0,K}^2 = o(1) \) and \( n^{-b} \zeta_{0,K} \{ \sqrt{\zeta_0 K_{2,K}} + \zeta_1 K \} = o(1) \).

**Proof of Theorem 1:** Choose a sequence \( \tilde{\lambda} \) within the shrinking neighborhood \( B(\lambda_0, \delta_0) \subset \Lambda \) of \( \lambda_0 \) with diameter \( \delta_0 = o(1) \) in terms of \( || \cdot ||_2 \). For \( G_n \) in Lemma A1, we define a process \( \hat{v}_n(g, r, \lambda) \) indexed by \( (g, r, \lambda) \in G_n \times S \times \Lambda_0 \) as

\[
\hat{v}_n(g, y, z, \lambda) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) \gamma_z(Z_i) \left[ \gamma_y(Y_i) - p\hat{K}(\hat{U}_i) \hat{\pi}_{y,\lambda} \right],
\]

where \( \hat{\pi}_{y,\lambda} \) is equal to \( \hat{\pi}_y \) defined below (4) except that in place of \( \lambda_\theta(\cdot) \), \( \lambda_\theta \in \Lambda_0 \), is used. By Lemma A1 and Assumption 3(iii), we obtain

\[
\log N_{||}(\varepsilon, G_n, || \cdot ||_p) \leq \log N(\varepsilon^p/2, \Lambda_0, || \cdot ||_\infty) + C_1/\varepsilon \leq C \log \varepsilon + C_1/\varepsilon.
\]

This implies that the bracketing entropy of \( \{g \gamma_z : (z, \gamma) \in \mathbb{R} \times G_n\} \) is bounded by \( C \log \varepsilon + C_1/\varepsilon \). We apply Lemma 1U of Escanciano and Song (2007) to \( \hat{v}_n(g, y, z, \lambda) \) to deduce that it is equal to

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

where \( o_p(1) \) is uniform over \( (g, y, z, \lambda) \in G_n \times \mathbb{R} \times \mathbb{R} \times \Lambda_0 \). Hence the above holds for any arbitrary sequence \( g_n \in G_n \) such that \( g_n(x) \rightarrow \gamma_u((F_{\theta_0} \circ \lambda_{\theta_0})(x)) \). This yields the wanted result.

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: \( \gamma_y^{\perp}(Y, U) \triangleq \gamma_y(Y) - \mathbb{E} [\gamma_y(Y) | U] \). We also denote for a real-valued function \( g(y, z) \) on \( \mathbb{R} \times \mathbb{R} \), \( \mathbb{P} g \triangleq \mathbb{E} [g(Y, Z)] \) and \( (\mathbb{P}^U g)(u) \triangleq \mathbb{E} [g(Y, Z) | U = u] \).

**Proof of Corollary 1:** (i)(a) and (ii)(b): Assume the null hypothesis. We first consider \( \nu_{1n}(r) \). The convergence of finite dimensional distributions of the process \( \sqrt{n} \mathbb{P}_n [\gamma_u \gamma_z \gamma_y^{\perp}] \) follows by the usual CLT. 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 \( \sqrt{n} \mathbb{P}_n [\gamma_u \gamma_z \gamma_y^{\perp}] \) to verify the covariance function of the limit Gaussian process in the theorem. The weak convergence \( \sqrt{n} \mathbb{P}_n [\gamma_u \gamma_z \gamma_y^{\perp}] \) of follows from the fact 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)). We can likewise analyze the limit behavior of the process \( \nu_n(r) = \sqrt{n} p_n [\gamma_u \gamma_z \gamma_y^\perp] \) to obtain the wanted result.

(i)(b) and (ii)(b): Write

\[
\frac{n^{-1/2}}{\nu_n(r)} = \sqrt{n} p_n [\gamma_u \gamma_z \gamma_y^\perp] = \sqrt{n} p_n [\gamma_u [\gamma_z \gamma_y^\perp] + \gamma_u [\gamma_z \gamma_y^\perp] + \gamma_y [\gamma_z \gamma_y^\perp] + \gamma_y [\gamma_z \gamma_y^\perp] + o_p(1),
\]

where \( \rho_w = \gamma_z (\gamma_y - pU_y). \) Now, assume that the null hypothesis does not hold. Using the preliminary result in (7), we write

\[
\frac{n^{-1/2}}{\hat{\nu}_n(r)} = \sqrt{n} p_n [\gamma_u [\gamma_z \gamma_y^\perp] + \gamma_u [\gamma_z \gamma_y^\perp] + \gamma_y [\gamma_z \gamma_y^\perp] + o_p(1),
\]

by adding and subtracting terms. The first and second processes are 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

\[
\nu_n^K(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \gamma_z(Z_i) \{\gamma_y(Y_i) - (pU_y)(Y_i, U_i)\}
\]

by replacing \( \gamma_z \) and \( \gamma_y \) by \( \gamma_z^K \) and \( \gamma_y^K \). In view of the proof of Theorem 1, the main step is applying Lemma 1U of Escanciano and Song (2007) to the above process. To this end, it suffices to show that the classes \( \{\gamma_y^K : y \in \mathbb{R}\} \) and \( \{\gamma_z^K : z \in \mathbb{R}\} \) satisfy the required bracketing entropy conditions. This is shown in Lemma A3(a) of Song (2007). Hence by following the same steps in the proof of Theorem 1, we obtain that under the null hypothesis,

\[
\nu_n^K(r) = \sqrt{n} p_n [\gamma_u (\gamma_z^K)^\perp (\gamma_y^K)^\perp] + o_p(1).
\]

Furthermore, by observing from (i) of this theorem, \( pU (\gamma_y^K) = 0 \) and \( pU (\gamma_z^K) = 0 \). Note that under the null hypothesis of conditional independence, \( (\gamma_z^K)(U_i, Z_i) \) and \( (\gamma_y^K)(U_i, Y_i) \) are

36
conditionally independent given $U_i$. Therefore

$$\mathbf{P}^U(\gamma^K Y \gamma^K Y) = \mathbf{P}^U(\gamma^K Y) \mathbf{P}^U(\gamma^K Y) = 0.$$ 

Therefore, under the null hypothesis, we have

$$\nu_n^K(r) = \sqrt{\nu_n^K[\gamma_u(\gamma^K Z)(\gamma^K Y)]} + o_P(1).$$

We investigate the process $\sqrt{\nu_n^K[\gamma_u(\gamma^K Z)(\gamma^K Y)]}$ and establish its weak convergence. The convergence of finite dimensional distributions follows by the usual central limit theorem. Note that the functions $\gamma^K Y$ and $\gamma^K Z$ are uniformly bounded. By Lemma A3(a) of Song (2007), the class $\mathcal{A} = \{\gamma_u(\gamma^K Z)(\gamma^K Y) : (u, z, y) \in S\}$ is totally bounded and the process $\sqrt{\nu_n^K[\gamma_u(\gamma^K Z)(\gamma^K Y)]}$ is stochastically equicontinuous in $\mathcal{A}$, and hence the weak convergence of the process $\sqrt{\nu_n^K[\gamma_u(\gamma^K Z)(\gamma^K Y)]}$ follows by Proposition in Andrews (1994), p. 2251.

The covariance function of this process is equal to $\mathbf{P} \left[ \gamma_{u_1} \wedge u_2 (\gamma^K Z_1)(\gamma^K Z_2)(\gamma^K Y_1)(\gamma^K Y_2) \right]$. Under the null hypothesis of conditional independence, the covariance function becomes

$$\mathbf{P} \left[ \gamma_{u_1} \wedge u_2 \mathbf{P}^U \left[ (\gamma^K Z)(\gamma^K Y)(\gamma^K Y_2) \right] \right] = \mathbf{P} \left[ \gamma_{u_1} \wedge u_2 \mathbf{P}^U \left[ \gamma^K Z_1 \gamma^K Z_2 \right] \mathbf{P}^U \left[ \gamma^K Y_1 \gamma^K Y_2 \right] \right] = \mathbf{P} \left[ \gamma_{u_1} \wedge u_2 \gamma^K Z_1 \gamma^K Z_2 \gamma^K Y_1 \gamma^K Y_2 \right].$$

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

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

Now, when we replace $\gamma^K Z$, $\gamma^K Y$, by $\tilde{\gamma}^K Z$, $\tilde{\gamma}^K Y$, the covariance function becomes

$$\int_{u_1 \wedge u_2}^u \left[ \int_{z_1 \wedge z_2} \hat{f}_{Z|U}^{-1}(z | u) \hat{f}_{Z|U}(z | u) dz \right] \left[ \int_{y_1 \wedge y_2} \hat{f}_{Y|U}^{-1}(y | u) \hat{f}_{Y|U}(y | u) dy \right] du = (u_1 \wedge u_2)(z_1 \wedge z_2)(y_1 \wedge y_2).$$

(b) The result is immediate from (26). \(\Box\)

Here, we review the notion of asymptotic Bahadur efficiency. We consider the parametrization that we introduced in the beginning of subsection 6.1. Denote by $h_T(v)$ and $h_S(v)$ the limits of the rejection probability of tests based on $\Gamma_{\nu_n}$ and $\Gamma_{\nu_n}$ under the null
hypothesis. That is, under the null hypothesis

\[ P\{T_n \geq v\} \to h_T(\kappa) \text{ and } P\{S_n \geq \kappa\} \to h_S(\kappa) \]

as \( n \to \infty \). This often boils down to the computation of tail probabilities of a functional of Gaussian processes. Suppose that there exist constants \( b_T \) and \( b_S \) satisfying

\[
\ln h_T(\kappa) = -\frac{1}{2} b_T \kappa^2 + o(1) \quad \text{and} \quad \ln h_S(\kappa) = -\frac{1}{2} b_S \kappa^2 + o(1) \text{ as } \kappa \to \infty.
\]

(27)

On the other hand, suppose that under the fixed alternative \( P_\tau \), we have numbers \( \varphi_T \) and \( \varphi_S \) on \( S \) such that

\[
\frac{1}{\sqrt{n}} T_n \to p \varphi_T, \quad \text{and} \quad \frac{1}{\sqrt{n}} S_n \to p \varphi_S.
\]

(28)

Then the approximate Bahadur efficiency (denoted by \( e_{T,S}^B(\tau) \)) at \( P_\tau \) of a test based on the test statistic \( T_n \) relative to one that is based on \( S_n \) is computed from (Nikitin (1995))

\[
e_{T,S}^B(\tau) = \frac{b_T}{b_S} \times \left\{ \frac{\varphi_T}{\varphi_S} \right\}^2.
\]

(29)

The approximate Bahadur efficiency is composed of two ratios. The first ratio \( b_T/b_S \) measures the relative tail behavior of the limiting distribution of the test statistic under the null hypothesis. The second ratio \( \varphi_T/\varphi_S \) measures how sensitively the test statistics deviate from the null limiting distribution in large samples under the alternatives. The limiting Pitman efficiency \( e_{T,S}^* \) coincides with \( \lim_{\tau \downarrow 0} e_{T,S}^B(\tau) \) when the conditions of Wieand (1976) is satisfied and the limit \( \lim_{\tau \downarrow 0} e_{T,S}^B(\tau) \) exists.

**Proof of Corollary 2 :** It is obvious that the sequences \( \{T_n\} \) and \( \{S_n\} \) are standard sequences as defined in Bahadur (1960). Hence, we prove Condition III* of Wieand (1976). Let \( b_p = \sup_{r \in S} |\sqrt{n} \langle \gamma_u, \gamma_z (\gamma_u - g_y^Y) \rangle| \) and

\[ U_n(r) = \hat{\nu}_n(r) - \sqrt{n} \langle \gamma_u, \gamma_z (\gamma_u - g_y^Y) \rangle. \]

Then, by the result of Corollary 1, \( \lim_{n \to \infty} P\{\sup_{r \in S}|U_n(r)| < v\} = Q(v) \), where \( Q \) is the distribution function \( \sup_{r \in S}|\nu_1(r)| \). Now, observe that for any \( \varepsilon > 0 \) and \( \delta \in (0,1) \), there exists \( C > 0 \) such that for all \( n > C/b_p^2 \),

\[ P\{|n^{-1/2}T_n - b_p| < \varepsilon b_p\} \geq P\{\sup_{r \in S}|U_n(r)| < \varepsilon b_p\} > 1 - \delta \]

if \( P \in \{P_\tau : \tau \in (0, \tau_0]\} \), by Lemma of Wieand (1976), p.1007. Hence the sequence \( T_n \)
satisfies the Condition III* of Wieand (1976). Similarly, we can show that the sequence \( S_n \) also satisfies Condition III* because the local shifts under the fixed alternatives are the same. (See Corollary 1(i)(b) and (ii)(b).) Hence the limiting Pitman efficiency coincides with the approximate Bahadur efficiency.

From now on, we omit the notation that represents dependence on \( \tau \) for notational brevity. The covariance kernels of the limit Gaussian processes of \( \nu(r) \) and \( \nu_1(r) \) in Theorem 1 are respectively given by

\[
c(r_1, r_2) = \int_0^{u_1 \land u_2} g_{Z_1 \land Z_2}(u; \tau)(g_{Y_1 \land Y_2}(u) - g_{Y_1}(u)g_{Y_2}(u))du,
\]

\[
c_1(r_1, r_2) = \int_0^{u_1 \land u_2} (g_{Z_1 \land Z_2}(u) - g_{Z_1}(u)g_{Z_2}(u))(g_{Y_1 \land Y_2}(u) - g_{Y_1}(u)g_{Y_2}(u))du.
\]

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

\[
\lim_{\kappa \to \infty} \kappa^{-2} \log P\{\sup_{r \in S} |\nu(r)| \geq \kappa\} = -\frac{1}{2}b_T,
\]

\[
\lim_{\kappa \to \infty} \kappa^{-2} \log P\{\sup_{r \in S} |\nu_1(r)| \geq \kappa\} = -\frac{1}{2}b_S,
\]

where the constants \( b_T \) and \( b_S \) are given by

\[
b_T^{-1} = \sup_{r \in S} c(r, r) = \sup_{y \in \mathbb{R}} \int_0^1 (g_y^Y(u) - g_y^Y(u)^2)du
\]

and

\[
b_S^{-1} = \sup_{r \in S} c_1(r, r) = \sup_{(z, y) \in \mathbb{R}^2} \int_0^1 (g_z^Z(u) - g_z^Z(u)^2)(g_y^Y(u) - g_y^Y(u)^2)du
\]

Hence

\[
\frac{b_T}{b_S} = \sup_{(z, y) \in \mathbb{R}^2} \frac{\int_0^1 (g_z^Z(u) - g_z^Z(u)^2)(g_y^Y(u) - g_y^Y(u)^2)du}{\sup_{y \in \mathbb{R}} \int_0^1 (g_y^Y(u) - g_y^Y(u)^2)du} \leq \frac{1}{4}.
\]

Now let us consider the case of alternatives. The results of Theorem 1 yields that the numbers \( \varphi_T \) and \( \varphi_S \) in (28) are equal, given by

\[
\varphi_T = \sup_{(u, z, y) \in S} |P_T [\gamma_u(\langle \gamma_z, \gamma_y \rangle u - \langle \gamma_z, 1 \rangle u, \langle \gamma_y, 1 \rangle u)]|.
\]

Therefore, we obtain the wanted result from (29).

**Proof of Corollary 3:** Trivial from the result of Theorem 2.

**Proof of Corollary 4:** We can easily check Condition III* of Wieand(1976) for \( T_n \) and \( S_n \) similarly as in the proof of Corollary 2. Hence, we compute the approximate Bahadur
efficiency. The covariance kernels of the limit Gaussian processes of \( \hat{\nu}_{n,a}^K(r) \) and \( \hat{\nu}_{n,E}^K(r) \) in Theorem 1 are identical. Hence for \( b_T \) and \( b_S \) in (27), \( b_T = b_S \). Now, the wanted result follows by plugging (24) into (29).

**Proof of Corollary 5:** Again, we can easily check Condition III* of Wieand (1976) for \( T_n \) and \( S_n \) similarly as in the proof of Corollary 2. The computation of \( e_{T,S}^* \) is again via the approximate Bahadur efficiency.

The covariance kernels of the limit Gaussian processes of \( \nu^K(r) \) is given by

\[
\varphi_K(r_1, r_2) = \int_{u_1 \wedge u_2}^{u_1 \wedge u_2} g_{u_1 \wedge u_2}(u) g_{u_1 \wedge u_2}(u) du,
\]

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

\[
\lim_{\kappa \to \infty} \kappa^{-2} \log P\{\sup_{r \in S} |\nu^K(r)| \geq \kappa\} = -b_S/2,
\]

where the constant \( b_S \) is given by \( b_S^{-1} = \sup_{r \in S} \varphi_K(r, r) = 1 \). Therefore,

\[
\frac{b_T}{b_S} = 1/ \left\{ \sup_{(z,y) \in \mathbb{R}^2} \int_{0}^{1} (g^z_z(u) - g^y_y(u))^2 (g^y_y(u) - g^y_y(u))^2 du \right\},
\]

as \( b_S \) in (30) is \( b_T \) now. On the other hand, under \( P_T \), we have

\[
\varphi_T = \sup_{(u,z,y) \in S} |E_T \left[ \gamma_u \left\{ \langle \gamma_z, \gamma_y \rangle u - \langle \gamma_z, 1 \rangle u \langle \gamma_y, 1 \rangle u \right\} \right] | = \sup_{(u,z,y) \in S} |E_T \left[ C_{Z,Y \mid u}(g^z_z(U), g^y_y(U)) - g^z_z(U) g^y_y(U) \right] |.
\]

Now, we turn to \( \varphi_S \). Observe that

\[
\varphi_S = \sup_{(u,z,y) \in S} |P_T \left[ \gamma_u \left( \langle \gamma_z^K, \gamma_y^K \rangle u - \langle \gamma_z^K, 1 \rangle u \langle \gamma_y^K, 1 \rangle u \right) \right] | = \sup_{(u,z,y) \in S} |E_T \left[ (K\gamma_z)(U; Z)(K\gamma_y)(U; Y) \right] |.
\]

The second equality follows from the conditional orthogonality property of \( K \).

First, assume that \( \xi(z, \varepsilon) \) is strictly increasing in \( z \). We observe that the expectation in the last term in (31) is equal to

\[
E_T [(K\gamma_z)(U; Z)(K\gamma_{\epsilon^{-1}(y,z)})(U; Z)]
\]

in this case. (When \( \xi(z, \varepsilon) \) is strictly decreasing in \( z \), the expectation in the last term in (31)
is equal to

$$E_{\tau}[\mathcal{K}\gamma_{\varepsilon}(U;Z)(\mathcal{K}(1 - \gamma_{\varepsilon}(y,z)))(U;Z)|U]$$

$$= -E_{\tau}[\mathcal{K}\gamma_{\varepsilon}(U;Z)(\mathcal{K}\gamma_{\varepsilon}(y,z))(U;Z)|U],$$

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 (32) is equal to

$$E_{\tau}[\gamma_{\varepsilon}(Z)\gamma_{\varepsilon-1}(y,z)(Z)|U] = E_{\tau}[F_{\varepsilon}(U;Z)g_{\varepsilon}(U;\tau), g_{\varepsilon}(U;\tau)|U].$$

Therefore, (omitting the notation for dependence on $\tau$)

$$\varphi_T = \sup_{(z,y,u) \in \mathcal{S}} \left| E\left[\gamma_u(U) \left\{ C_{Z,Y|U}(g_{\varepsilon}^{Z}(U),g_{\varepsilon}^{Y}(U)|U) - g_{\varepsilon}^{Z}(U)g_{\varepsilon}^{Y}(U) \right\} \right] \right|$$

and

$$\varphi_S = \sup_{(z,y,u) \in \mathcal{S}} E\left[\gamma_u(U)C_{Z,Y|U}(g_{\varepsilon}^{Z}(U),g_{\varepsilon}^{Y}(U)|U) \right] = 1.$$

Combining the results, we obtain the wanted statement.

Suppose $g_{\varepsilon}^{Z}(u)$ and $g_{\varepsilon}^{Y}(u)$ does not depend on $u$. Then

$$e_{T,S}^* \leq \frac{1}{4} \max \left\{ \sup_{(z,y) \in [0,1]^2} \{ z \wedge y - zy \}^2, \sup_{(z,y) \in [0,1]^2} \{ zy - \max\{ z + y - 1, 0 \} \}^2 \right\} = \frac{1}{4}.$$

Recall that $z \wedge y$ and $\max\{ z + y - 1, 0 \}$ represent the upper and lower Frechet-Hoeffding bounds. (e.g. Nelsen (1998)).

9.2 Asymptotic Validity of Feasible Martingale Transforms

In this section, we briefly sketch how the feasible martingale transform that we introduced in a previous section can be shown to be asymptotically valid. The details can be filled following the proof of Theorem 2 of Song (2007) and hence are omitted.

Let $\tilde{V}_i^{Y} = F_{Y}(Y_i)$ and $\tilde{V}_i^{Z} = F_{Z}(Z_i)$. From Theorem 2(ii), it is not hard to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i)\tilde{z}_{\varepsilon}^{K}(U_i,\tilde{V}_i^{Z})\tilde{z}_{\varepsilon}^{Y}(U_i,\tilde{V}_i^{Y}) \sim W(r),$$

41
where \( \hat{\gamma}_x^K(U_i, \tilde{V}_i^Z) \) and \( \hat{\gamma}_y^K(U_i, \tilde{V}_i^Y) \) are equal to \( \tilde{\gamma}_x^K(U_i, Z_i) \) and \( \tilde{\gamma}_y^K(U_i, Y_i) \) with \( Z_i \) and \( Y_i \) replaced by \( \tilde{V}_i^Z \) and \( \tilde{V}_i^Y \). Hence it suffices to show that

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\bar{U}_i) \hat{\gamma}_x^K(\bar{U}_i, V_i^Z) \hat{\gamma}_y^K(\bar{U}_i, \bar{V}_i^Y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \tilde{\gamma}_x^K(U_i, \tilde{V}_i^Z) \tilde{\gamma}_y^K(U_i, \tilde{V}_i^Y) + o_P(1),
\]

uniformly over \( r \in S_0 \). The difference between the two sums is equal to

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(\bar{U}_i) \hat{\gamma}_x^K(\bar{U}_i, V_i^Z) \{ \hat{\gamma}_y^K(\bar{U}_i, \bar{V}_i^Y) - \tilde{\gamma}_y^K(U_i, \tilde{V}_i^Y) \}
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \gamma_u(U_i) \tilde{\gamma}_x^K(U_i, \tilde{V}_i^Y) \{ \hat{\gamma}_y^K(U_i, \bar{V}_i^Z) - \tilde{\gamma}_y^K(U_i, \tilde{V}_i^Z) \}
\]

\[
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \gamma_u(\bar{U}_i) - \gamma_u(U_i) \} \hat{\gamma}_y^K(U_i, \bar{V}_i^Z) \tilde{\gamma}_x^K(U_i, \bar{V}_i^Y)
\]

\[= A_{1n} + A_{2n} + A_{3n} \text{ say.}
\]

We consider \( A_{1n} \) first. First define

\[\hat{\psi}(z', x; z, u) \triangleq \gamma_u(F_{n, \tilde{\theta}}(\lambda_\tilde{\theta}(x))) \hat{\gamma}_x^K(F_{n, \tilde{\theta}}(\lambda_\tilde{\theta}(x)), F_n^Z(z')) \]

and

\[\psi_0(z', x; z, u) \triangleq \gamma_u(F_{\theta_0}(\lambda_{\theta_0}(x))) \tilde{\gamma}_x^K(F_{\theta_0}(\lambda_{\theta_0}(x)), F_Z(z')) \]

and

\[\varphi(y', x; y, v) \triangleq \frac{1\{0 < F_n^Y(y') \leq y \land v\}}{\sqrt{\hat{f}(F_n^Y(y'), F_{n, \tilde{\theta}}(\lambda_\tilde{\theta}(x)))} \{ 1 - \hat{F}(F_n^Y(y')|F_{n, \tilde{\theta}}(\lambda_\tilde{\theta}(x)))) \}}\]

\[\varphi_0(y', x; y, v) \triangleq \frac{1\{0 < F_Y(y') \leq y \land v\}}{\sqrt{\hat{f}(F_Y(y'), F_{\theta_0}(\lambda_{\theta_0}(x)))} \{ 1 - F(F_Y(y')|F_{\theta_0}(\lambda_{\theta_0}(x)))) \}}.\]

Then, \( A_{1n} \) is equal to

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}(Z_i, X_i; z, u) \left\{ \hat{E} \left[ \varphi(Y_i, X_i; y, \bar{y})|\bar{U}_i \right]_{\bar{y}=Y_i} - \hat{E} \left[ \varphi_0(Y_i, X_i; y, \bar{y})|U_i \right]_{\bar{y}=Y_i} \right\}.
\]

Under the regularity conditions of the kind in Assumption 6 and conditions for basis functions in Assumption 2U of Song (2007) for \( \varphi \) and \( \hat{\psi} \), we can deduce

\[
A_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{E} \left[ \psi_0(Z_i, X_i; z, u) \left\{ \varphi_0(\bar{y}, X_i; y, Y_i) - \hat{E} \left[ \varphi_0(Y_i, X_i; y, \bar{y})|U_i \right]_{\bar{y}=Y_i} \right\} |U_i \right]_{\bar{y}=Y_i} + o_P(1).
\]

(33)
Under the null hypothesis of conditional independence,

\[
\mathbb{E}\left[\psi_0(Z_i, X_i; z, u) \left\{ \varphi_0(\bar{y}, X_i; y, Y_i) - \mathbb{E}[\varphi_0(Y_i, X_i; y, \bar{y})|U_i] \right\} | U_i \right] | y = Y_i = 0
\]

since

\[
\mathbb{E}[\psi_0(Z_i, X_i; z, u)|Y_i, U_i] = \mathbb{E}[\psi_0(Z_i, X_i; z, u)|U_i] = \gamma_u(U_i) \langle \hat{\gamma}_z, 1 \rangle_{U_i} = 0
\]

by conditional independence and by Theorem 2(i). Under the local alternatives such that

\[
|\mathbb{E}[\psi_0(Z_i, X_i; z, u)|Y_i, U_i]| = o_P(1),
\]

\( A_{1n} = o_P(1) \) because the process in (33) is a mean-zero process. (This follows by establishing that the class of functions that index the process in (33) is \( P\)-Donsker.) Hence both under the null hypothesis and under local alternatives, \( A_{1n} = o_P(1) \). The second term \( A_{2n} \) can be dealt with similarly.

Let us turn to the third term \( A_{3n} \). We replace \( F_{n, \hat{\theta}}(\lambda_{\hat{\theta}}(\cdot)) \) in \( \hat{U}_i \) by an arbitrary deterministic function \( G \) that belongs to a class \( \mathcal{F}_n \) such that

\[
P\{F_n \in \mathcal{F}_n\} \to 1 \quad \text{and} \quad \log N(\varepsilon, \mathcal{F}_n, ||\cdot||_p) < C\varepsilon^{-b}
\]

for \( b \in [0, 2) \), and for each \( G \in \mathcal{F}_n \), we have \( ||F_{\theta_0} \circ \lambda_{\theta_0} - G||_\infty = O(n^{-1/2}) \). Then, we consider the process

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \gamma_u(G(X_i)) - \gamma_u(U_i) \} \hat{\gamma}_y(U_i, \hat{V}_i^Y) \hat{\gamma}_z(U_i, \hat{V}_i^Z), \quad (G, u, y, z) \in \mathcal{F}_n \times \mathcal{S},
\]

which can be shown to be equivalent to

\[
\sqrt{n} \{ \gamma_u(G(X_i)) - \gamma_u(U_i) \} \mathbb{E} \left[ \hat{\gamma}_y(U_i, \hat{V}_i^Y) \hat{\gamma}_z(U_i, \hat{V}_i^Z)|G(X_i), U_i \right] + o_P(1).
\]

Then using Lemma A2(ii) of Song (2006) and Theorem 2(i),

\[
\mathbb{E} \left[ \hat{\gamma}_y(U_i, \hat{V}_i^Y) \hat{\gamma}_z(U_i, \hat{V}_i^Z)|G(X_i), U_i \right] = O_P(n^{-1/2}).
\]

Now, from the fact that \( ||F_{\theta_0} \circ \lambda_{\theta_0} - G||_\infty = O(n^{-1/2}) \), it follows \( A_{3n} = o_P(1) \).
References


when the criterion function is not smooth," *Econometrica* 71, 1591-1608.


in insurance: general testable implications," mimeo.


tistics*, 8, 598-617.


parametric and semiparametric functional forms,” *Econometrica*, 64, 865-890.

[22] Hahn, J. (1998), "On the role of the propensity score in efficient semiparametric esti-
mation of average treatment effects," *Econometrica*, 64, 315-331.


evaluations and social experiments: accounting for heterogeneity in programme im-

treatment effects using the estimated propensity score," *Econometrica*, 71, 1161-1189.

[27] Khmaladze, E. V. (1993), “Goodness of fit problem and scanning innovation martin-


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

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

\textsuperscript{9}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\textsuperscript{K}_{1A}) and MG trns (T\textsuperscript{K}_{1B}) indicate tests based on martingale transforms, the former using the test statistic $T\textsuperscript{K}_{1A}$ which is based on the true conditional expectation of $\gamma^K_y(Y)$ given $U$ and the latter using the test statistic $T\textsuperscript{K}_{2}$ which is based on the estimated conditional expectation of $\gamma^K_y(Y)$ given $U$. MG trns (T\textsuperscript{K}_{1B}) is the symmetrized version of MG trns (T\textsuperscript{K}_{1A}).
Table 2: Power of the Tests: Conditional Positive Dependence

<table>
<thead>
<tr>
<th>$n = 300$</th>
<th>Order 1</th>
<th>Order 2</th>
<th>MG trns ($T_{1A}^\kappa$)</th>
<th>MG trns ($T_{1B}^\kappa$)</th>
<th>MG trns ($T_2^\kappa$)</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa = 0.2$</td>
<td>5$^2$</td>
<td>5</td>
<td>0.193</td>
<td>0.191</td>
<td>0.229</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.214</td>
<td>0.190</td>
<td>0.225</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>6$^2$</td>
<td>5</td>
<td>0.187</td>
<td>0.196</td>
<td>0.226</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.245</td>
<td>0.213</td>
<td>0.269</td>
<td>0.243</td>
</tr>
<tr>
<td>$\kappa = 0.4$</td>
<td>5$^2$</td>
<td>5</td>
<td>0.575</td>
<td>0.581</td>
<td>0.678</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.604</td>
<td>0.605</td>
<td>0.667</td>
<td>0.664</td>
</tr>
<tr>
<td></td>
<td>6$^2$</td>
<td>5</td>
<td>0.591</td>
<td>0.585</td>
<td>0.679</td>
<td>0.673</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.626</td>
<td>0.612</td>
<td>0.694</td>
<td>0.701</td>
</tr>
<tr>
<td>$\kappa = 0.6$</td>
<td>5$^2$</td>
<td>5</td>
<td>0.909</td>
<td>0.938</td>
<td>0.953</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>0.930</td>
<td>0.921</td>
<td>0.957</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>6$^2$</td>
<td>5</td>
<td>0.898</td>
<td>0.927</td>
<td>0.949</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>0.919</td>
<td>0.912</td>
<td>0.951</td>
<td>0.948</td>
</tr>
</tbody>
</table>
Table 3: Power of the Tests: Conditional Negative Dependence

<table>
<thead>
<tr>
<th>$n = 300$</th>
<th>Order 1</th>
<th>Order 2</th>
<th>MG trns ($T_{1_A}^k$)</th>
<th>MG trns ($T_{1_B}^k$)</th>
<th>MG trns ($T_{2}^k$)</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa = -0.2$</td>
<td>$5^2$</td>
<td>5</td>
<td>0.024</td>
<td>0.176</td>
<td>0.208</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.039</td>
<td>0.217</td>
<td>0.204</td>
<td>0.208</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$6^2$</td>
<td>5</td>
<td>0.029</td>
<td>0.197</td>
<td>0.191</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.036</td>
<td>0.229</td>
<td>0.201</td>
<td>0.201</td>
<td></td>
</tr>
<tr>
<td>$\kappa = -0.4$</td>
<td>$5^2$</td>
<td>5</td>
<td>0.070</td>
<td>0.614</td>
<td>0.652</td>
<td>0.636</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.068</td>
<td>0.647</td>
<td>0.634</td>
<td>0.640</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$6^2$</td>
<td>5</td>
<td>0.072</td>
<td>0.644</td>
<td>0.644</td>
<td>0.649</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.062</td>
<td>0.656</td>
<td>0.628</td>
<td>0.650</td>
<td></td>
</tr>
<tr>
<td>$\kappa = -0.6$</td>
<td>$5^2$</td>
<td>5</td>
<td>0.367</td>
<td>0.944</td>
<td>0.956</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.319</td>
<td>0.965</td>
<td>0.957</td>
<td>0.928</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$6^2$</td>
<td>5</td>
<td>0.359</td>
<td>0.943</td>
<td>0.941</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.306</td>
<td>0.949</td>
<td>0.950</td>
<td>0.933</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: The Empirical Density of Test Statistics both under the Null and under the Alternative: n=3000

- With true $F(y|x)$ under $H_0$
- With estimated $F(y|x)$ under $H_0$
- With true $F(y|x)$ under $H_1$
- With estimated $F(y|x)$ under $H_1$
Figure 2: Empirical Density of Empirical Processes at \((x,y)=(0.5,0.5)\) under the Null and under the Alternative: \(n=3000\)

with true \(F(y|x)\) under \(H_0\)

with estimated \(F(y|x)\) under \(H_0\)

with true \(F(y|x)\) under \(H_1\)

with estimated \(F(y|x)\) under \(H_1\)