Testing Distributional Inequalities and Asymptotic Bias

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

February 11, 2008

Abstract

This note studies the asymptotic bias properties of one-sided tests of distributional inequalities under $\sqrt{n}$-converging Pitman local alternatives. Examples of these tests include tests of stochastic dominance and tests of independence or conditional independence against positive dependence or positive conditional dependence. As regards the distributional inequalities, there are two types of testing problems: testing a distributional equality against an inequality that holds uniformly over the support of the random variable, and testing the presence of such a uniform inequality relation between two random variables. This paper focuses on tests of the first type which, in particular, have a sup norm of a standard Brownian bridge process with a drift as their limit under $\sqrt{n}$-converging Pitman local alternatives. It is shown that there exist a class of $\sqrt{n}$-converging Pitman local alternatives against which the tests are asymptotically biased. This phenomenon of asymptotic bias does not arise for tests of the second type in general, due to Anderson's lemma.

Key words and Phrases: Asymptotic Bias, One-sided Tests, Stochastic Dominance, Conditional Independence, Pitman Local Alternatives, Brownian Bridge Processes

JEL Classifications: C12, C14, C52.

1 Introduction

Many empirical researches in economics have long been centered around investigating distributional relations between random variables. For instance, many literatures have focused on

[1] This note was originally inspired when I was working on a paper on testing stochastic dominance with Oliver Linton and Yoon-Jae Whang. 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.

a stochastic dominance relation between two or several distributions of investment strategies or income distributions. (See Barrett and Donald (2003), Linton, Massoumi, and Whang (2005), and references therein.) Other examples are testing the independence and testing conditional independence between two random variables. This independence property has been widely used as an identifying assumption, and numerous testing procedures have been proposed, making it impossible to do justice to all the literature in this limited space. To name but a few, see Linton and Gozalo (1998), Su and White (2004), Delgado and Gonzalez Manteiga.(2001), and Song (2007).

Many nonparametric tests proposed in this literature are omnibus tests whose rejection probability converges to one for all types of violation of the null hypothesis. It is also well-known that one can often obtain a test that is asymptotically unbiased against $\sqrt{n}$-converging Pitman local alternatives. This asymptotic unbiasedness property is a more refined property than the typical consistency property of tests, and requires an analysis of the local asymptotic power function of the test. In two-sided Cramer-von Mises tests, the analysis of local asymptotic power function has been performed by a principal component decomposition of the tests. (Anderson and Darling (1952), Durbin and Knott (1972), Neuhaus (1976), Eubank and LaRiccia (1992), Stute (1997), and Escanciano (2006), to name but a few.) In the case of Kolmogorov-Smirnov tests, Milbrodt and Strasser (1990) and Janssen (1995) analyzed the curvature of the local asymptotic power function. Global bounds of the asymptotic power function for two-sided Kolmogorov-Smirnov tests were obtained by Strasser (1990). However, less is known for one-sided nonparametric tests. For a limited class of alternatives, the global local asymptotic power function and local efficiency of one-sided Kolmogorov-Smirnov tests has been studied by Andel (1967) and Hajek and Sidak (1967).

This paper draws attention to the omnibus property of asymptotic unbiasedness of a test: a property that a test is asymptotically unbiased against all the directions of $\sqrt{n}$-converging Pitman local alternatives. This "omnibus" property of asymptotic unbiasedness is often established by invoking Anderson's Lemma, and satisfied by many two-sided tests that have a Gaussian process with a drift as a limit under local alternatives. However, as far as the author is concerned, much less is known for the case of one-sided nonparametric tests where Anderson's Lemma does not apply. This paper attempts to analyze the local asymptotic power properties in terms of their global bounds. Unlike the result of Strasser (1990) for two-sided tests, the bounds are not tight, yet this approach still reveals interesting aspects of local asymptotic powers and provides an intuitive link between the shape of the local alternatives and the local asymptotic powers.

More specifically, this paper analyzes Kolmogorov-Smirnov type nonparametric tests whose limit under $\sqrt{n}$-converging Pitman local alternatives is a standard Brownian bridge process with a drift. This paper formulates a useful bound for the tail-probabilities of the sup norm of the Brownian bridge process with a drift. This bound is used to compute the upper bound for the rejection probability of the nonparametric test. The derivation of the upper bound utilizes a result of Ferger (1995) who characterizes the joint distribution of a standard Brownian bridge process and its maximizer. Then, we introduce a class of $\sqrt{n}$-converging Pitman local alternatives against which the test is indeed asymptotically biased. The class of local alternatives that this paper considers cannot be thought of as a pathological case. We provide intuitive explanations behind this asymptotic unbiasedness property. We apply this finding to two classes of nonparametric tests: stochastic dominance tests, and tests of independence or conditional independence.

We may view this paper's result in the light of Janssen (1990)'s finding that any nonparametric test has nearly trivial local asymptotic powers against all the directions except for a finite dimensional subspace. The paper demonstrates that in the case of one-sided nonparametric tests, there exists a set of Pitman local alternatives against which the local asymptotic powers are *strictly lower* than the asymptotic size. Furthermore, this set is not contained by any finite dimensional space. Some intuitive examples are provided to illustrate this phenomenon of asymptotic biasedness.

The paper illustrates the result using small scale Monte Carlo simulations of stochastic dominace tests. The result of Monte Carlo simulations show the biasedness of the test with the finite samples, as predicted by the asymptotic theory. Indeed, some of the rejection probabilities are shown to lie below their empirical sizes as one moves away from the null. The test recovers nontrivial power when the distribution moves farther away from the null hypothesis beyond these values. The result of this paper suggests that the study of asymptotic bias in one-sided nonparametric tests in full shape poses a nontrivial problem.

2 A Preliminary Result and Discussions

In this section, we analyze the local asymptotic power properties of nonparametric tests whose limit under $\sqrt{n}$-converging Pitman local alternatives is a standard Brownian bridge process with a drift. Let us introduce a lemma that serves as a basis for the analysis in this paper. For any random element ν in $L_2([0,1])$, define $\tau(\nu) = \inf\{t \in [0,1] : \nu(t) = \sup_{0\leq s\leq 1}\nu(s)\}$. Suppose that $\tau(\nu)$ is a random variable. For given ν, let us introduce a real function $H_\nu : \mathbf{R} \times [0,1] \to [0,1]$ such that

$$H_\nu(y,z) \triangleq P\{\sup_{0\leq t\leq 1}\nu(t) \leq y \text{ and } \tau(\nu) \leq z\}. \tag{1}$$

For each $y \in \mathbf{R}$, the function $H_\nu(y, z)$ increases in z, and hence it is of bounded variation in z. Let $\{t_m\}_{m=1}^M$ be a partition of $[0, 1]$ such that $\Delta t_m = |t_m - t_{m-1}| \to 0$ as $M \to \infty$. For any Lebesgue measurable set $A \subset [0, 1]$, and for any function $g : \mathbf{R} \to \mathbf{R}$, we write

$$\int_A H_v(g(t), dt) = \lim_{M \to \infty : \Delta t_m \to 0} \sum_{m=1 : t_m \in A}^{M} \{H_v(g(t_{m+1}), t_{m+1}) - H_v(g(t_{m+1}), t_m)\}$$

when the limit exists and does not depend on the choice of partition. We first introduce a general lemma that presents bounds for the local asymptotic powers.

Lemma 1: *Suppose ν is a random element in $L_2([0, 1])$ such that $\tau(\nu)$ is a random variable.* (i) *Suppose that $\delta(t) \leq D(t)$ where $D : \mathbf{R} \to [\delta_L, \delta_U]$ is decreasing in* $[0, 1]$. *Then, for each* $c \geq 0$,

$$P\left\{\sup_{0 \leq t \leq 1} [v(t) + \delta(t)] > c\right\} \leq 1 - \int_0^1 H_v(c - D(t), dt). \tag{2}$$

(ii) *Suppose that $\delta(t) \geq D(t)$ where $D : \mathbf{R} \to [\delta_L, \delta_U]$ is increasing in* $[0, 1]$. *Then, for each* $c \geq 0$,

$$P\left\{\sup_{0 \leq t \leq 1} [v(t) + \delta(t)] > c\right\} \geq 1 - \int_0^1 H_v(c - D(t), dt).$$

The lemma provides bounds of the asymptotic power of any nonparametric test whose limit under the local alternatives with a drift $\delta(t)$ is given by $\sup_{0 \leq t \leq 1} [v(t) + \delta(t)]$. This representation is convenient because in certain cases with $\nu(t)$ as a Gaussian process, the function $H_\nu(y, z)$ is explicitly known. Both the inequalities in (i) and (ii) become equalities only when $\delta(t)$ is a constant. In the context of a stochastic dominance test, it is required that $\delta(t) \to 0$ as $|t| \to \infty$. Therefore, both the bounds do not hold with equality for any type of Pitman local alternatives in this context.

Let us consider the following type of one-sided Kolmogorov-Smirnov tests. Suppose $\nu_n(t)$ is a stochastic process such that $\nu_n(t)$ is constructed by observed random variables and

$$\begin{aligned} \nu_n(\cdot) &\Longrightarrow \nu(\cdot) \text{ under the null hypothesis and} \\ \nu_n(\cdot) &\Longrightarrow \nu(\cdot) + \delta(\cdot) \text{ under local alternatives,} \end{aligned}$$

where $\nu(t)$ is a Gaussian process, and the notation $\Longrightarrow$ indicates the weak convergence in the sense of Hoffman and Jorgensen. Then, the upper bound in (2) serves as an upper bound for the asymptotic rejection probability of the one-sided Kolmogorov-Smirnov type test

$$T_n = \sup_{0 \leq t \leq 1} \nu_n(t).$$

This upper bound can be explicitly computed when the function $H_\nu(y,z)$ is fully known. This is indeed the case when ν is a standard Brownian bridge process, B^0, on $[0,1]$. First, note that the almost sure uniqueness of the maximizer $\tau(B^0)$ is well-known (e.g. Ferger (1995), Kim and Pollard (1990)). Hence $\tau(B^0)$ is a random variable. By Ferger (1995), we have in general for all $y \geq 0$,

$$\begin{aligned} H_{B^0}(y,z) = {} & \Phi\left(\frac{y}{\sqrt{z(1-z)}}\right) - \exp\left(-2y^2\right)\Phi\left(\frac{y(2z-1)}{\sqrt{z(1-z)}}\right) \\ & -(1-z)\left(2\Phi\left(\frac{y}{\sqrt{z(1-z)}}\right)-1\right) \end{aligned}$$

if $z \in (0,1)$, and $H_{B^0}(y,1) = 1 - \exp\left(-2y^2\right)$, and $H_{B^0}(y,0) = 0$. Here Φ denotes the distribution function of a standard normal random variable. Hence we can explicitly compute the upper bound for the asymptotic local power of the test for various classes of local drifts $\delta(t)$.

3 Testing Stochastic Dominance

3.1 Testing Stochastic Dominance

Let $\{X_i\}_{i=1}^n$ be a set of i.i.d. random variables. We consider testing the following null hypothesis

$$H_0^{BD}: F(t) \leq t \text{ for all } t \in [0,1]. \tag{3}$$

This test is a simple version of the stochastic dominance test studied by Barret and Donald (2003). The null hypothesis says that the distribution of X_i is stochastically dominated by a uniform distribution. The paper's framework applies to the situation where we are interested in testing whether the marginal distribution of X_i is stochastically dominated by a distribution that has a strictly increasing distribution function G. Then, the null hypothesis becomes

$$H_0^{BD}: F(t) \leq G(t) \text{ for all } t.$$

Since we know G, we can reformulate the null hypothesis as

$$H_0^{BD}: G^{-1}(F(t)) \leq t \text{ for all } t.$$

By writing $\tilde{F} = G^{-1} \circ F$, we are back to the original null hypothesis in (3).

The natural test statistics are obtained by using the following stochastic process

$$\nu_n(t) = \sqrt{n}\left(\hat{F}_n(t) - t\right),$$

where $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq t\}$, the empirical distribution function of $\{X_i\}_{i=1}^{n}$. When $F(t)$ is the distribution function of Uniform(0,1), we may consider the following test statistics:

$$\begin{aligned} T_n^{BD} &\triangleq \sup_{t\in\mathbf{R}}\sqrt{n}\left(\hat{F}_n(t) - t\right) \underset{D}{\rightarrow} \quad \sup_{t\in[0,1]}B^0(t) \text{ and} \\ T_n^{BD+} &\triangleq \sup_{t\in\mathbf{R}}\sqrt{n}\left(\hat{F}_n(t) - t\right)_+ \underset{D}{\rightarrow} \quad \sup_{t\in[0,1]}(B^0(t))_+ \end{aligned}$$

where $B^0(t)$ is a Brownian bridge process, i.e. a Gaussian process whose covariance kernel is given by $t_1 \wedge t_2 - t_1 t_2$, and $(f)_+ = \max(f, 0)$. The result follows from the well-known weak convergence of the empirical process: $\nu_n(\cdot) \Longrightarrow B^0(\cdot)$ combined with the continuous mapping theorem.

One can show that the test based on T_n^{BD} is consistent against all the violations of the null hypothes is. However, we demonstrate that the test is not asymptotically unbiased against all types of Pitman local alternatives that converge to the null hypothesis at the rate of $\sqrt{n}$. In nonparametric tests, Anderson's Lemma is usually used to establish the asymptotic unbiasedness of a test. Note that here, we cannot apply Anderson's Lemma because the class of functions: $\{f : \sup_{t\in[0,1]} \max\{f(t), 0\} \leq c\}$ is convex, but not symmetric.

Let $F_n(t)$ be the distribution function of X_i under local alternatives $\delta(t)$ and satisfy

$$F_n(t) = F(t) + \frac{\delta(t)}{\sqrt{n}} \tag{4}$$

where $F(t)$ is the distribution function of Unif[0,1]. The following result shows that there exists a test of the form $1\{T_n > c\}$ that is asymptotically biased against a certain class of Pitman local alternatives.

Corollary 1: *There exists $c_\alpha, c_\alpha^+ > 0$ and a subset $\mathcal{A}$ of uniformly bounded drifts $\delta(t)$ such that for each $\delta(t) \in \mathcal{A}$,*

$$\begin{aligned} \lim_{n\to\infty} P\left\{T_n^{BD} > c_\alpha\right\} &= \alpha \text{ and } \lim_{n\to\infty} P\left\{T_n^{BD+} > c_\alpha^+\right\} = \alpha \text{ under the null hypothesis and} \\ \lim_{n\to\infty} P\left\{T_n^{BD} > c_\alpha\right\} &< \alpha \text{ and } \lim_{n\to\infty} P\left\{T_n^{BD+} > c_\alpha^+\right\} < \alpha \text{ under the local alternatives.} \end{aligned}$$

Furthermore, this set $\mathcal{A}$ is not contained by any finite dimensional space of drifts $\delta(t)$.

Corollary 1 demonstrates the existence of a set of local shifts against which the test is asymptotically biased. The proof is based on a construction of local alternatives of a simplified form. This facilitates the computation of the upper bound in Lemma 1, and at the same time, provides an intuitive view of the local asymptotic power of the test against local alternatives. To construct this class, we fix numbers $x \in (0,1)$ and $b \in \mathbf{R}$ and let $D(t)$ be a function such that

$$D(t; x, b_1, b_2) = \begin{cases} 0 \text{ if } t < 0 \\ b_1 \text{ if } 0 \leq t \leq x \\ -b_2 \text{ if } x < t \leq 1 \\ 0 \text{ if } t > 1. \end{cases} \tag{5}$$

Note that $D(t; x, b_1, b_2)$ is decreasing in $t \in [0,1]$. When $x > b_2/\sqrt{n}$ and $b_1 > 0$, there exists a class of local shifts $\mathcal{A}$ such that for all $\delta(t) \in \mathcal{A}$, $\delta(t) \leq D(t; x, b_1, b_2)$ and $t + \delta(t)/\sqrt{n}$ is nondecreasing in t, taking values in $[0,1]$. Using Lemma 1, we can obtain that

$$\lim_{n\to\infty} P\left\{T_n^{BD} > c_\alpha\right\} \leq 1 - \int_0^1 H_{B^0}(c_\alpha - D(t; x, b_1, b_2), dt) \triangleq \pi(x, b_1, b_2), \text{ say,}$$

where $\pi(x, b_1, b_2)$ can be computed as

$$\pi(x, b_1, b_2) = H_{B^0}(c_\alpha + b_2, x) - H_{B^0}(c_\alpha - b_1, x) + \exp(-2(c_\alpha + b_2)^2).$$

An example of $F_n(t) = t + \delta(t)$ with $n = 1$ appears in Figure 1.

INSERT FIGURES 1 AND 2 HERE

If we take $c_\alpha = 1.224$, $P\{\sup_{t\in[0,1]} B^0(t) > c\} = \exp(-2c^2) \approx 0.05$. The graph of an upper bound $\pi(x, b_1, b_2)$ for the asymptotic rejection probability with this critical value c_α and $b_2 = 0.2$ is plotted in Figure 2. For example, when x is equal to 0.3 and b_1 is chosen to be smaller than 0.2, the upper bound for the asymptotic rejection probability lies below the nominal level 0.05.

The main reason for this phenomenon is purely due to the nature of the alternative hypothesis that is both infinite dimensional and one-sided. To see this, we consider $\delta(t)$ defined in (10). (See Figure 4.) In this example, when x is small, the Brownian bridge process is more likely to attain its maximum in the area where $\delta(t)$ takes a negative value than in the area where $\delta(t)$ takes a positive value. Therefore, the supremum of the Brownian bridge process plus the local shift $\delta(t)$ is more likely to lie below the supremum of the Brownian bridge process. Since the maximizer of the Brownian bridge process follows Uniform(0,1), this asymptotic biasedness result purely follows by the fact that the Lebesgue measure of t's in which $\delta(t)$ takes a negative value is greater than that of t's in which $\delta(t)$ takes a positive value and that b_1 is equal to b_2. Intuitively, this appears to imply that $x < 0.5$ if and only if the asymptotic power falls below the size 0.05. The reason that this does not happen is in Figure 2 is that when $x = 0.5$, the monotonicity restriction imposed upon $t + \delta(t)$ still renders the area of t's giving negative $\delta(t)$ larger than the area of t's giving positive $\delta(t)$. This asymmetry disappears as b_1 and b_2 become smaller. Figure 3 shows the graphs of the upper bound $\pi(x, b, b)$ with varying b's. The graph shows that as b becomes smaller, the upper bound of the asymptotic power, $\pi(x, b, b)$, crosses the size 0.05 at a value nearer to 0.5.

Observe that the result of Theorem 1 applies to many other one-sided nonparametric tests whose limiting distribution under the null hypothesis is a supremum of a standard Brownian bridge. Later, we discuss the case of testing independence and testing conditional independence.

3.2 Testing the Absence of Stochastic Dominance Relationship

A different formulation of stochastic dominance is the following :

$$H_0^{LMW} : \text{neither } F(t) \le t \text{ for all } t \in [0,1], \text{ nor } t \le F(t) \text{ for all } t \in [0,1]. \tag{6}$$

This is the test of the absence of stochastic dominance relation of $F(t)$ with a uniform distribution. The test is a special case of the stochastic dominance relation studied by Linton, Massoumi, and Whang (2005). Similarly as in Linton, Massoumi, and Whang (2005), we may consider the following test statistics:

$$T_n^{LMW} = \min\left\{\sup_{t\in\mathbf{R}}\sqrt{n}\left(\hat{F}_n(t) - t\right), \sup_{t\in\mathbf{R}}\sqrt{n}\left(t - \hat{F}_n(t)\right)\right\},$$

where $\hat{F}_n(t)$ is the empirical distribution function of $\{X_i\}_{i=1}^n$ as defined previously. Then, under the null hypothesis of $F(t) = t$,

$$T_n^{LMW} \underset{D}{\to} \min\left\{\sup_{t\in[0,1]} B^0(t), \sup_{t\in[0,1]} - B^0(t)\right\},$$

from the well-known Donsker's theorem. The asymptotic size is obtained by choosing the critical value c such that the following probability

$$\begin{aligned} & P\left\{\min\left\{\sup_{t\in[0,1]} B^0(t), \sup_{t\in[0,1]} - B^0(t)\right\} > c\right\} \\ = \; & 2P\left\{\sup_{t\in[0,1]} B^0(t) > c\right\} - \left[1 - P\left\{\sup_{t\in[0,1]} |B^0(t)| \leq c\right\}\right] \end{aligned} \tag{7}$$

equals 0.05. The last two probabilities can be precisely evaluated; in particular,

$$P\{\sup_{t\in[0,1]} |B^0(t)| \leq c\} = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2c^2}.$$

When we set $c = 0.6781$, the asymptotic size is set to be approximately 0.05. The asymptotic unbiasedness immediately follows from Anderson's Lemma for Pitman local alternatives with any $\delta(t)$.

4 Testing Independence and Conditional Independence

The asymtotic biasedness result applies to other examples of one-sided nonparametric tests. We introduce two examples: testing independence and testing conditional independence, and propose one-sided tests. These tests contain empirical quantile transforms in the indicator functions which requires slightly involved techniques. We resort to Song (2007) to deal with this.

4.1 Testing Independence

Suppose that we are given a random sample $\{X_i, D_i\}_{i=1}^n$ of (X_i, D_i) where X_i is a continuous random variable and D_i is a binary random variable. We say that X_i and D_i are positively dependent if and only if

$$\mathbf{E}[1\{F(X) \leq t\}D] > P\{F(X) \leq t\}\mathbf{E}[D] \text{ for all } t \in [0, 1], \tag{8}$$

where $F(\cdot)$ is the distribution function of X. When the inequality is reversed, we say that they are negatively dependent. Taking the quantile transform of X into $F(X)$ is to normalize the marginal distribution of X to a uniform distribution on $[0, 1]$. When the strict inequality in (8) is replaced by equality, the equality indicates the independence of X_i and D_i.

Consider the following test:

$$\begin{aligned} H_0 &: \quad X_i \text{ and } D_i \text{ are independent} \\ H_1 &: \quad X_i \text{ and } D_i \text{ are positively dependent.} \end{aligned}$$

Then, the null hypothesis can be formulated as

$$H_0 : \mathbf{E}\left[1\{X_i \leq t\}D_i\right] = P\{X_i \leq t\}\mathbf{E}\left[D_i\right] \text{ for almost all } t.$$

In order to obtain an asymptotically pivotal test, let us consider the empirical quantile transform of $\{X_i\}_{i=1}^n$. Define $U_{n,i} = F_{n,i}(X_i)$ where $F_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n 1\{X_j \leq x\}$. In this case, one might consider the following process:

$$\zeta_n(t) = \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^n \left(1\{U_{n,i} \leq t\}D_i - \frac{1}{(n-1)(n-2)}\sum_{j=1,j\neq i}^n \sum_{k=1,j\neq i,j}^n 1\{U_{n,j} \leq t\}D_k\right),$$

where $\hat{\sigma}_D^2 = \bar{D}_n - \bar{D}_n^2$ and $\bar{D}_n = \frac{1}{n}\sum_{i=1}^n D_i$. Then, we can show that this process weakly converges to a standard Brownian bridge:

$$\zeta_n \Longrightarrow B^0.$$

Consider the following Kolmogorov-Smirnov test:

$$T_n^I = \sup_{t\in[0,1]} \zeta_n(t).$$

Then, the result of this paper shows that there exists a class of local alternatives against which this test is asymptotically biased. The local alternatives take a form as

$$\frac{\mathbf{E}\left[1\{U_i \leq t\}D_i\right]}{\mathbf{E}\left[D_i\right]} - P\{U_i \leq t\} = \frac{1}{\sqrt{n}}\delta(t),$$

where $U_i = F(X_i)$ and $\delta(t)$ is given as before.

Theorem 1: *Under the assumption that $(X_i, D_i)_{i=1}^n$ is i.i.d. with finite second moments, and that $\mathbf{E}\left[D_i\right] \in (0,1)$,*

$$\begin{aligned} T_n^I &\underset{D}{\to} \sup_{t\in[0,1]} B^0(t) \text{ under the null hypothesis and} \\ T_n^I &\underset{D}{\to} \sup_{t\in[0,1]} \{B^0(t) + \delta(t)\} \text{ under the local alternatives.} \end{aligned}$$

Hence the asymptotic biasedness result of this paper applies to this case with $\delta(t)$ defined previously. The result of this paper implies that when the Lebesgue measure of the set of t's such that the negative dependence between $1\{X_i \leq t\}$ and D_i arises is much greater than that of t's such that the positive dependence arises, the rejection probability can be smaller than the size.

Following the same manner as in (6), we may consider the following null and alternative hypotheses:

$$\begin{aligned} H_0 &: \quad X_i \text{ and } D_i \text{ are neither positively dependent nor negatively dependent} \\ H_1 &: \quad X_i \text{ and } D_i \text{ are positively dependent or negatively dependent} \end{aligned}$$

The null hypothesis H_0 is weaker than the hypothesis of independence. We can formulate the test statistic as

$$T_n^{I\prime} = \min\left\{ \sup_{t\in[0,1]} \zeta_n(t), \sup_{t\in[0,1]} -\zeta_n(t) \right\}.$$

Then, the limiting behavior of the test statistic $T_n^{I\prime}$ can be derived similarly as before. This test does not suffer from the previous kind of asymptotic biasedness, due to Anderson's lemma.

4.2 Testing Conditional Independence

The result of this paper also applies to one-sided tests of conditional independence. For example, various forms of one sided tests of conditional independence have been used in the literature of contract theory. (e.g. Cawley and Phillipson (1999), Chiappori and Salanié (2000), Chiappori, Jullien, Salanié, and Salanié (2002).) Conditional independence restrictions also have been used in the identification of treatment effects parameters in program evaluations. One-sided test of conditional independence can be used when the conditional positive or negative dependence of the participation in the program and counterfactual outcomes is excluded *a priori*. Following the suit of testing independence, we say that X_i and D_i are conditionally positively dependent (CPD) given Z_i if and only if

$$\mathbf{E}[1\{F(X|Z) \leq t\}D|Z] > P\{F(X|Z) \leq t|Z\}\mathbf{E}[D|Z] \text{ for all } t \in [0,1],$$

where $F(\cdot|Z)$ is the conditional distribution function of X given Z. When the reverse inequality holds, we say that they are conditionally negatively dependent (CND) given Z_i. Again, taking the conditional quantile transform of X into $F(X|Z)$ is to normalize the conditional distribution of X given Z to a uniform distribution on $[0,1]$. When the strict inequality in

(8) is replaced by equality, the equality indicates the conditional independence of X_i and D_i given Z_i.

Let X_i be a continuous variable and D_i a binary variable taking values from $\mathcal{D} = \{0, 1\}$, and Z_i a discrete random variable taking values from a finite set $\mathcal{Z} \subset \mathbf{R}$. We assume that $P\{Z_i = z\} \in (\varepsilon, 1-\varepsilon)$ for some $\varepsilon > 0$. Consider the following test

$$\begin{aligned} H_0 &: \quad X_i \text{ and } D_i \text{ are conditionally independent given } Z_i \\ H_1 &: \quad X_i \text{ and } D_i \text{ are CPD given } Z_i, \end{aligned}$$

The null hypothesis of conditional independence can be written as

$$H_0 : \mathbf{E}\left[1\{X_i \le t\} D_i | Z_i\right] = P\{X_i \le t | Z_i\} \mathbf{E}\left[D_i | Z_i\right] \text{ for almost all } t.$$

In order to obtain an asymptotically pivotal test, we follow the idea of Song (2007) and consider the empirical conditional quantile transform of $\{X_i\}_{i=1}^n$ given Z_i. Define $\tilde{X}_{n,i} = F_{n,i}(X_i|Z_i)$ where $F_{n,i}(x|z) = \frac{1}{n-1}\sum_{j=1, j\neq i}^n 1\{Z_j = z\}1\{X_j \le x\} / \frac{1}{n-1}\sum_{j=1,j\neq i}^n 1\{Z_j = z\}$. Then the test can be constructed as a one-sided Kolmogorov-Smirnov functional of the following process:

$$\zeta_n(t, z) = \frac{1}{\sqrt{n\hat{\sigma}_D^2}} \sum_{i=1}^n 1\{Z_i = z\} \left(D_i - \hat{F}(D_i | Z_i = z)\right) \left(1\{\tilde{X}_{n,i} \le t\} - t\right),$$

where $\hat{\sigma}_D^2 = \frac{1}{n}\sum_{i=1}^n 1\{Z_i = z\}\{D_i - \hat{F}(D_i|Z_i = z)\} \left(1\{\tilde{X}_{n,i} \le t\} - t\right)^2$ and $\hat{F}(D_i|Z_i = z) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n 1\{Z_j = z\}D_j / \frac{1}{n-1}\sum_{j=1,j\neq i}^n 1\{Z_j = z\}$. The local alternatives that we focus on are

$$\mathbf{E}\left[1\{F(X_i|Z_i) \le t\} D_i | Z_i\right] - t\mathbf{E}\left[D_i | Z_i\right] = \frac{1}{\sqrt{n}}\delta(t)\sigma_D^2, \text{ for almost all } t.$$

Define the test statistic

$$T_n^{CI} = \sup_{z \in \mathcal{Z}} \sup_{t \in [0,1]} \zeta_n(t, z).$$

The theorem below presents the asymptotic distribution of T_n^{CI}. Note that a similar result was obtained by the author (Song, 2007) when Z_i is a continuous variable and contains an unknown parameter.

Theorem 2 : *Suppose that $(X_i, D_i, Z_i)_{i=1}^n$ is i.i.d. with finite second moments and that $P(Z_i = z) \in (\varepsilon, 1-\varepsilon)$ for all $z \in \mathcal{Z}$ for some $\varepsilon > 0$. Furthermore assume that the conditional distribution of X_i given $Z_i = z$ is absolutely continuous with respect to the Lebesgue measure* *for all* $z \in \mathcal{Z}$. *Then,*

$$T_n^{CI} \underset{D}{\to} \sup_{z\in\mathcal{Z}} \sup_{t\in[0,1]} \zeta(t,z) \text{ under the null hypothesis and}$$
$$T_n^{CI} \underset{D}{\to} \sup_{z\in\mathcal{Z}} \sup_{t\in[0,1]} \{\zeta(t,z) + \delta(t)\} \text{ under the local alternatives,}$$

where $\zeta(t,z)$ *is a mean-zero Gaussian process such that*

$$\begin{aligned}
\mathbf{E}\left[\zeta(t_1,z)\zeta(t_2,z)\right] &= t_1 \wedge t_2 - t_1 t_2 \text{ and} \\
\mathbf{E}\left[\zeta(t_1,z)\zeta(t_2,z)\right] &= 0 \text{ if } z_1 \neq z_2.
\end{aligned}$$

The test T_n^{CI} is asymptotically pivotal. We can construct asymptotic critical values for the test T_n^{CI} in the following way. Let $c_{1-\alpha}$ be such that

$$P\left\{\sup_{t\in[0,1]} B^0(t) \geq c_{1-\alpha}\right\} = (1-\alpha)^{1/|\mathcal{Z}|},$$

where $|\mathcal{Z}|$ denotes the cardinality of the set $\mathcal{Z}$. Then, it follows that

$$\begin{aligned}
P\left\{T_n^{CI} \geq c_{1-\alpha}\right\} &\to P\left\{\sup_{z\in\mathcal{Z}} \sup_{t\in[0,1]} \zeta(t,z) \geq c_{1-\alpha}\right\} \qquad (9)\\
&= \left[P\left\{\sup_{t\in[0,1]} \zeta(t,z) \geq c_{1-\alpha}\right\}\right]^{|\mathcal{Z}|} = \left[P\left\{\sup_{t\in[0,1]} B^0(t) \geq c_{1-\alpha}\right\}\right]^{|\mathcal{Z}|} = 1-\alpha.
\end{aligned}$$

The first equality above follows because $\sup_{t\in[0,1]} \zeta(t,z)$ is i.i.d across different z's in $\mathcal{Z}$. Although the test statistic T_n^{CI} does converges to a Kolmogorov-Smirnov functional of a standard Brownian bridge process, we can obtain the same result of Theorem 1, namely that the test is asymptotically biased against a class of Pitman local alternatives. If the negative conditional dependence arises for a larger set of t's than the positive dependence, the rejection probability of the test can be smaller than the size of the test.

Again, we may be interested in the following null hypothesis and alternative hypothesis:

$$\begin{aligned}
H_0 &: X_i \text{ and } D_i \text{ are neither CPD nor CNP given } Z_i \\
H_1 &: X_i \text{ and } D_i \text{ are CPD or CNP given } Z_i
\end{aligned}$$

The null hypothesis H_0 is weaker than the hypothesis of conditional independence. We can formulate the test statistic as

$$T_n^{CI\prime} = \min \left\{ \sup_{z \in \mathcal{Z}} \sup_{t \in [0,1]} \zeta_n(t,z), \sup_{z \in \mathcal{Z}} \sup_{t \in [0,1]} -\zeta_n(t,z) \right\}.$$

Then, the limiting behavior of the test statistic $T_n^{I\prime}$ can be derived similarly as before. The computation of asymptotic critical values can be done by combining (7) and (9). More specifically, let α_1 and α_2 be such that $1 - 2(1-\alpha_1)^{|\mathcal{Z}|} + (1-\alpha_2)^{|\mathcal{Z}|} = \alpha$ and let $c_{1-\alpha}$ such that

$$P\left\{ \sup_{t \in [0,1]} B^0(t) \geq c_{1-\alpha} \right\} = \alpha_1 \text{ and } P\left\{ \sup_{t \in [0,1]} |B^0(t)| \geq c_{1-\alpha} \right\} = \alpha_2.$$

Then, similarly as in (7), we write

$$\begin{aligned} & P\left\{ \min\left\{ \sup_{z \in \mathcal{Z}} \sup\nolimits_{t \in [0,1]} \zeta(t,z), \sup_{z \in \mathcal{Z}} \sup\nolimits_{t \in [0,1]} -\zeta(t,z) \right\} > c_{1-\alpha} \right\} \\ = \;& 2P\left\{ \sup_{z \in \mathcal{Z}} \sup_{t \in [0,1]} \zeta(t,z) \geq c_{1-\alpha} \right\} - \left[1 - P\left\{ \sup_{z \in \mathcal{Z}} \sup\nolimits_{t \in [0,1]} |\zeta(t,z)| \leq c_{1-\alpha} \right\} \right] \\ = \;& 1 - 2P^{|\mathcal{Z}|}\left\{ \sup_{t \in [0,1]} B^0(t) \leq c_{1-\alpha} \right\} + P^{|\mathcal{Z}|}\left\{ \sup\nolimits_{t \in [0,1]} \left| B^0(t) \right| \leq c \right\} = \alpha. \end{aligned}$$

Hence the asymptotic critical value $c_{1-\alpha}$ is valid. Again, this test does not suffer from the previous kind of asymptotic biasedness, due to Anderson's Lemma.

5 Simulations

To illustrate the implications of the asymptotic biasedness of the test, we consider the following simulation exercise. Let U_1 be a random variable distributed Uniform(0,1). Then, we define

$$X_i = (U - ab_1)1\{ab_1 \leq U \leq x\} + (U + \alpha b_2)1\{x < U \leq 1 - \alpha b_2\}$$

for $a \in (0,1)$. As a becomes closer to zero, the distribution of X_i becomes closer to the uniform distribution and when $a = 0$, the data generating process corresponds to the null hypothesis. Then the distribution function of X_i becomes

$$P(X_i \leq t\} = t + a\delta(t),$$

where

$$\delta(t) = \begin{cases} 0 \text{ if } t \leq 0 \\ b_1 \text{ if } 0 < t \leq x - b_1 \\ x - t \text{ if } x - b_1 < t \leq x + b_2 \\ -b_2 \text{ if } x + b_2 < t \leq 1 \\ 0 \text{ if } t > 1. \end{cases} \tag{10}$$

Note that this $\delta(t)$ is one example satisfying $\delta(t) \leq D(t)$ for $D(t)$ in (5). The shape of $F(t) + \delta(t)$ is depicted in Figure 4. The Monte Carlo simulation number is set to be 2000 and the sample size is equal to 600. The following table contains the results of the finite sample size and power of the test.

Table 1: The Rejection Probability of the Test T_n^{BD} of H_0^{BD} with $b_1 = 0.1$ and $b_2 = 0.1$ and $n = 600$.

	$x = 0.2$	$x = 0.3$	$x = 0.4$	$x = 0.5$
$a = 0$	0.0525	0.0540	0.0420	0.0425
$a = 0.03$	0.0295	0.0435	0.0520	0.0565
$a = 0.06$	0.0315	0.0435	0.0540	0.0700
$a = 0.09$	0.0295	0.0455	0.0755	0.0895
$a = 0.12$	0.0370	0.0550	0.0860	0.1275
$a = 0.15$	0.0465	0.0880	0.1335	0.1685
$a = 0.2$	0.0840	0.1535	0.2130	0.2435
$a = 0.3$	0.2510	0.3495	0.4105	0.4550

The numbers in the first row with $a = 0$ represent the empirical size of the test. The finite sample distribution of the test does not depend on the choice of x and hence the variation among these numbers show the sampling variations in Monte Carlo simulations. Under the alternatives, the rejection probability tends to increase with x because the area of t's giving $\delta(t)$ a negative value becomes smaller. Under the alternatives with $x = 0.2$ and a's from 0.03 to 0.15, the rejection probabilities lie below the empirical size, as predicted by the asymptotic bias result. However, as a moves farther from zero beyond these values, the empirical power of the test becomes nontrivial. A similiar phenomenon arises when $x = 0.3$ but less conspicuously. As we move x to farther away from the corner of 0, the triviality of the empirical power disappears, as consistent with the theoretical results of this note.

For comparison, we present the results from the two-sided Kolmogorov-Smirnov test in Table 2.

Table 2: The Rejection Probability of the Two-Sided Test of H_0^{BD} with $b_1 = 0.1$ and $b_2 = 0.1$ and $n = 600$.

	$x = 0.2$	$x = 0.3$	$x = 0.4$	$x = 0.5$
$a = 0$	0.0440	0.0450	0.0420	0.0460
$a = 0.03$	0.0465	0.0505	0.0450	0.0490
$a = 0.06$	0.0560	0.0595	0.0605	0.0675
$a = 0.09$	0.0675	0.0785	0.0785	0.0945
$a = 0.12$	0.0940	0.1235	0.1300	0.1330
$a = 0.15$	0.1240	0.1505	0.1685	0.1765
$a = 0.2$	0.2240	0.2335	0.2695	0.2725
$a = 0.3$	0.5070	0.5525	0.5785	0.6120

Table 2 shows that the rejection probabilities are larger than the empirical sizes, as predicted by the asymptotic unbiasedness results.

The following table is the test of the presence of stochastic dominance test that is considered in Section 2.2. Although much more general cases were studied by Linton, Massoumi and Whang (2005), a simulation result for this simple case is presented for comparison.

Table 3: The Rejection Probability of the Test of T_n^{LMW} of H_0^{LMW} with $b_1 = 0.1$ and 0.1 and $n = 600$.

	$x = 0.2$	$x = 0.3$	$x = 0.4$	$x = 0.5$
$a = 0$	0.0300	0.0385	0.0420	0.0315
$a = 0.03$	0.0495	0.0490	0.0400	0.0455
$a = 0.06$	0.0665	0.0720	0.0675	0.0770
$a = 0.09$	0.1040	0.1145	0.1220	0.1325
$a = 0.12$	0.1700	0.2160	0.2090	0.2025
$a = 0.15$	0.2660	0.3110	0.3490	0.3270
$a = 0.2$	0.5070	0.5385	0.5525	0.5530
$a = 0.3$	0.8990	0.9105	0.9080	0.9190

First, observe that the results show size distortions with the sample size 600. The test does not show any "dip" in power as in the previous case. This is expected by the omnibus asymptotic unbiasedness of the test as explained previously. It is also interesting to note that the rejection probabilities are larger than the case of testing H_0^{BD}. This demonstrates that the presence of a stochastic dominance relation is easier to detect from the data than both the existence and the direction of the stochastic dominance relation. This phenomenon is intuitive because the null hypothesis H_0^{BD} is stronger than the null hypothesis H_0^{LMW} and hence requires stronger evidence to be rejected.

6 Closing Remarks

This paper demonstrates that there exist $\sqrt{n}$-converging Pitman local alternatives against which the one-sided Kolmogorov test of distributional inequalities is asymptotically biased. Among the examples are testing stochastic dominance tests, testing independence or conditional independence. The examples are not pathological ones, nor does it apply only to a narrow class of examples. The result of this paper demonstrates that in the case of one-sided nonparametric tests, it is not a trivial problem to characterize the class of Pitman local alternatives against which the test is asymptoticall unbiased.

7 Appendix

Lemma A1 : Let $\{A_m\}_{m=1}^{\infty}$ and $\{B_m\}_{m=1}^{\infty}$ be sequences of sets such that $\cup_{m=1}^{M} A_m \subset \cup_{m=1}^{M} B_m$ for all $M \geq 1$. Then, for any decreasing sequence of sets $C_1 \supset C_2 \supset C_3 \supset \cdots$,

$$\cup_{m=1}^{M}(A_m \cap C_m) \subset \cup_{m=1}^{M}(B_m \cap C_m) \text{ for all } M \geq 1.$$

Proof of Lemma A1: Take $x \in A_{m'} \cap C_{m'}$ for some $m' \leq M$. Then, obviously, $x \in (\cup_{m=1}^{m'} B_m) \cap C_{m'} \subset \cup_{m=1}^{m'}(B_m \cap C_m) \subset \cup_{m=1}^{M}(B_m \cap C_m)$. ■

Proof of Lemma 1: First let us consider (i). The case with (ii) can be dealt with similarly. Observe that

$$P\left\{\sup_{0\leq t\leq 1}(\nu(t)+\delta(t)) \leq c\right\} \geq P\left\{\sup_{0\leq t\leq 1}(\nu(t)+D(t)) \leq c\right\}$$

by the assumption that $\delta(t) \leq D(t)$. Choose a set of grid points $\{b_1, \cdots, b_M\} \subset [\delta_L, \delta_U]$ such that

$$\delta_L = b_1 < b_2 < \cdots < b_{M-1} < b_M = \delta_U.$$

Let $A_m = \{t \in \mathbf{R} : b_m < \delta(t) \leq b_{m+1} \text{ and } \delta(t) \geq 0\}$. Then, $\mathbf{R} = \cup_{m=1}^{M} A_m$. Note that A_m's are disjoint. Define $D_M : \mathbf{R} \to [\delta_L, \delta_U]$ as

$$D_M(t) = \sum_{m=1}^{M} b_{m+1} 1\{t \in A_m\}.$$

By construction, $\delta(t) \leq D_M(t)$, for all $t \in \mathbf{R}$. Hence $\sup_{0\leq t\leq 1}(\nu(t)+\delta(t)) \leq \sup_{0\leq t\leq 1}\{\nu(t)+D_M(t)\}$. Define

$$\tau^D(\nu) = \arg\max_{t\in[0,1]}\{\nu(t)+D_M(t)\}.$$

Since $D_M(t)$ is constant almost everywhere in $\mathbf{R}$, $\tau^D(\nu)$ is a random variable. Now,

$$P\left\{\sup_{0\leq t\leq 1}(\nu(t)+D_M(t)) \leq c\right\} = \sum_{m=1}^{M} P\left\{\sup_{0\leq t\leq 1}(\nu(t)+D_M(t)) \leq c, \tau^D(\nu) \in A_m\right\}. \quad (11)$$

Observe that

$$
\begin{aligned}
& P\left\{\sup_{0\leq t\leq 1}\left(\nu(t)+D_M(t)\right)\leq c, \tau^D(\nu)\in A_m\right\} \\
= \ & P\left\{\sup_{t\in A_m}\left(\nu(t)+D_M(t)\right)\leq c, \tau^D(\nu)\in A_m\right\} \\
= \ & P\left\{\sup_{t\in A_m}\nu(t)\leq c-b_{m+1}, \tau^D(\nu)\in A_m\right\} \\
\geq \ & P\left\{\sup_{0\leq t\leq 1}\nu(t)\leq c-b_{m+1}, \tau^D(\nu)\in A_m\right\}.
\end{aligned}
$$

Because D_M is a decreasing function, $\tau^D(\nu)\leq\tau(\nu)$. We deduce from Lemma A1 above that

$$
\sum_{m=1}^{M} P\left\{\sup_{0\leq t\leq 1}\nu(t)\leq c-b_{m+1}, \tau^D(\nu)\in A_m\right\}\geq\sum_{m=1}^{M} P\left\{\sup_{0\leq t\leq 1}\nu(t)\leq c-b_{m+1}, \tau(\nu)\in A_m\right\}.
$$

Therefore,

$$
\begin{aligned}
P\left\{\sup_{0\leq t\leq 1}\left(\nu(t)+D(t)\right)\leq c\right\} \ &\geq \ \sum_{m=1}^{M} P\left\{\sup_{0\leq t\leq 1}\nu(t)\leq c-b_{m+1}, \tau(\nu)\in A_m\right\} \\
&= \ \sum_{m=1}^{M}\left\{H(c-b_{m+1},a_{m+1})-H(c-b_{m+1},a_m)\right\}.
\end{aligned}
$$

The inequality in the above does not depend on a specific choice of partitions. By choosing the partition $[\delta_L,\delta_U]$ finer, we obtain that as $M\to\infty$,

$$
P\left\{\sup_{0\leq t\leq 1}\left(\nu(t)+D(t)\right)\leq c\right\}\geq\int_{[0,1]} H_v(c-D(z),dz).
$$

Now, as $M\to\infty$, we have $D_M(t)\to D(t)$. The proof is complete. ■

Proof of Theorem 1: We first show that

$$
\hat{V}_{n,i}(r)\triangleq\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\left(1\{U_{n,j}\leq t\}-t\right)=o_P(1)
$$

uniformly over $t\in[0,1]$. Let $\Phi_{\delta,r}(y)=\Phi\left(\frac{r-y}{\delta}\right)$ where Φ is the distribution function of a standard normal variate. Define

$$
\hat{V}_{n,\delta}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\Phi_{\delta,r}(U_{n,i})-r\right) \text{ and } V_{n,\delta}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\Phi_{\delta,r}(U_i)-r\right).
$$

We can show that

$$
\hat{V}_n(r)-V_n(r)=\hat{V}_{n,\delta}(r)-V_{n,\delta}(r)+o(\delta)
$$

as $\delta\to 0$. Now, using an appropriate linearization, we can show that

$$
\hat{V}_{n,\delta}(r)-V_{n,\delta}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi'_{\delta,r}(U_i)\left\{U_{n,i}-U_i\right\}+o_P(1). \tag{12}
$$

The leading term on the right-hand side of (12) becomes

$$
\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\Phi'_{\delta,r}(U_i)\left\{1\{U_j\leq U_i\}-U_i\right\}+o_P(1).
$$

Note that $\mathbf{E}\left[\Phi'_{\delta,r}(U_i)\left\{1\{U_j \leq U_i\} - U_i\right\}|U_i\right] = 0$ and the Hajek projection of the leading sum becomes

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{E}\left[\Phi'_{\delta,r}(U_i)\left\{1\{U_j \leq U_i\} - U_i\right\}|U_j\right] + o_P(1) \\
= & \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_0^1 \Phi'_{\delta,r}(u)\left\{1\{U_j \leq u\} - u\right\}du + o_P(1).
\end{aligned}$$

The leading term is equal to (as $\delta \to 0$)

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\Phi_{\delta,r}(U_j) - \int_0^1 \Phi_{\delta,r}(u)du\right\} + o_P(1) \text{ (by integration by parts).}$$

However, by integration by parts, $\int_0^1 \Phi'_{\delta,r}(u)\left\{1\{U_j \leq u\} - u\right\}du$ is equal to

$$\begin{aligned}
[\Phi_{\delta,r}(u)]_{U_j}^1 - \int_0^1 \Phi'_{\delta,r}(u)udu &= \Phi_{\delta,r}(1) - \Phi_{\delta,r}(U_j) - [\Phi_{\delta,r}(u)u]_0^1 + \int_0^1 \Phi_{\delta,r}(u)du \\
&= -\Phi_{\delta,r}(U_j) + \int_0^1 \Phi_{\delta,r}(u)du = -1\{U_j \leq r\} + \int_0^1 1\{u \leq r\}du + o(\delta) \\
&= -1\{U_j \leq r\} + r + o(\delta).
\end{aligned}$$

Therefore, $\hat{V}_n(r) - V_n(r) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1\{U_i \leq r\} - r\right\} + o_P(1) = -V_n(r) + o_P(1)$, implying that $\hat{V}_n(r) = o_P(1)$.

Hence

$$\frac{1}{\sqrt{\hat{\sigma}_D^2(n-1)}}\sum_{i=1}^{n}\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}(1\{U_{n,j} \leq t\} - t)\left\{\frac{1}{n-2}\sum_{k=1,j\neq i,j}^{n}D_k\right\} = o_P(1).$$

Therefore,

$$\begin{aligned}
\zeta_n(t) &= \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}\left(1\{U_{n,i} \leq t\}D_i - \frac{t}{n-2}\sum_{k=1,j\neq i,j}^{n}D_k\right) + o_P(1) \\
&= \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}(1\{U_{n,i} \leq t\}D_i - tP\{D_i = 1\}) \\
&\quad - \frac{t}{\sqrt{\hat{\sigma}_D^2}}\frac{\sqrt{n}}{n-2}\sum_{k=1,j\neq i,j}^{n}(D_k - P\{D_i = 1\}) + o_P(1).
\end{aligned}$$

Let us deal with the first term:

$$\begin{aligned}
& \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}(1\{U_{n,i} \leq t\}D_i - tP\{D_i = 1\}) \\
= & \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}P\{D_i = 1\}(1\{U_{n,i} \leq t\} - t) + \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}1\{U_{n,i} \leq t\}(D_i - P\{D_i = 1\}) \\
= & \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^{n}1\{U_{n,i} \leq t\}(D_i - P\{D_i = 1\}) + o_P(1)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\zeta_n(t) &= \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^n \left(1\{U_{n,i}\le t\}-t\right)\left(D_i - P\{D_i=1\}\right)+o_P(1)\\
&= \frac{1}{\sqrt{n\hat{\sigma}_D^2}}\sum_{i=1}^n \left(1\{U_{i}\le t\}-t\right)\left(D_i - P\{D_i=1\}\right)+o_P(1)\\
&= \frac{1}{\sqrt{n\sigma_D^2}}\sum_{i=1}^n \left(1\{U_{i}\le t\}-t\right)\left(D_i - P\{D_i=1\}\right)+o_P(1)
\end{aligned}$$

because $\hat{\sigma}_D^2 \to_P \sigma_D^2$. Now, the last term weakly converges to a Gaussian process because the classes indexing the process is P-Donsker. It is easy to check that this Gaussian process has the same covariance kernel as that of a standard Brownian bridge. ■

Proof of Theorem 2 : The proof closely follows that of Theorem 1 in Song (2007). We sketch the steps here. First, write

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i &= z\}\left(D_i-\hat{F}(D_i|Z_i=z)\right)\left(1\{\tilde{X}_{n,i}\le t\}-t\right) \qquad (13)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i=z\}\left(D_i-F(D_i|Z_i=z)\right)\left(1\{\tilde{X}_{n,i}(z)\le t\}-t\right)\\
-\frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i &= z\}\left(\hat{F}(D_i|Z_i=z)-F(D_i|Z_i=z)\right)\left(1\{\tilde{X}_{n,i}(z)\le t\}-t\right).
\end{aligned}$$

We turn to the leading sum on the right-hand side, which we write as

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i &= z\}\left(D_i-F(D_i|Z_i=z)\right)\left(1\{\tilde{X}_{i}(z)\le t\}-t\right)\\
+\frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i &= z\}\left(D_i-F(D_i|Z_i=z)\right)\left(1\{\tilde{X}_{n,i}(z)\le t\}-1\{\tilde{X}_{i}(z)\le t\}\right).
\end{aligned}$$

Similarly in the proof of Claim 2 of Theorem 1 of Song (2007), we can obtain that the last term is $o_P(1)$ uniformly in $(t,z)\in[0,1]\times\mathcal{Z}$.

We turn to the last sum in (13). Recall that $\tilde{X}_{n,i}(z)=F_{n,i}(X_i|z)$ where $F_{n,i}(\cdot|z)$ is uniformly bounded in $[0,1]$ and increasing for each $z\in\mathcal{Z}$. Furthermore, we can establish that

$$\sqrt{n}\sup_{z\in\mathcal{Z}}||F_{n,i}(\cdot|z)-F(\cdot|z)||_\infty = O_P(1).$$

We consider the following sum:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n 1\{Z_i=z\}\left(\hat{F}(D_i|Z_i=z)-F(D_i|Z_i=z)\right)\left(1\{G_z(X_i)\le t\}-t\right),$$

where $G_z\in\mathcal{G}_{z,n}=\{G : G \text{ is uniformly bounded in } [0,1] \text{ and increasing, and } ||G-F(\cdot|z)||_\infty\le Cn^{-1/3}\}$. By noting the bracketing entropy bound for $\mathcal{G}_{z,n}$ (Birman and Solomjak (1967)), we can apply Lemma B1 below to deduce that the sum is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}1\{Z_i=z\}\left(D_i-F(D_i|Z_i=z)\right)\mathbf{E}\left[1\{\tilde{X}_i(z)\leq t\}-t|Z_i=z\right]+o_P(1)=o_P(1),$$

uniformly in $(t,z)\in[0,1]\times\mathcal{Z}$. The last equality follows because $\tilde{X}_i(z)=F(X_i|z)$ and conditional on $Z_i=z$, X_i is absolutely continuous with respect to the Lebesgue measure. By collecting the results, we conclude that the first sum in (13) is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}1\{Z_i=z\}\left(D_i-F(D_i|Z_i=z)\right)\left(1\{\tilde{X}_i(z)\leq t\}-t\right)+o_P(1).$$

The class of functions indexing this process is obviously P-Donsker, and its weak convergence to a Gaussian process follows by the usual procedure. (e.g. van der Vaart and Wellner (1996)). The computation of the covariance kernel for this Gaussian process follows straightforwardly. ■

The following lemma is used to prove Theorem 2 and useful for other purposes. Hence we make the notations for the lemma self-contained here. The lemma is a variant of Lemma B1 of Song (2007), but not a special case, because the conditioning variable here is discrete. The proof is much simpler in this case. Let Ψ_n and Φ_n be classes of functions $\psi:\mathbf{R}^{d_S}\rightarrow\mathbf{R}$ and $\varphi:\mathbf{R}^{d_W}\rightarrow\mathbf{R}$ that satisfy Assumptions B1 and B2 below and let $(W_i,Z_i,X_i)_{i=1}^n$ be i.i.d. from P where Z_i is a discrete random variable. We establish the uniform asymptotic representation of the following process $\hat{\Delta}_n(\varphi,\psi)$, $(\varphi,\psi)\in\Phi_n\times\Psi_n$, defined by

$$\hat{\Delta}_n(\varphi,\psi)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i)\left\{\hat{g}_{\varphi,i}(Z_i)-g_\varphi(Z_i)\right\}$$

where $\hat{g}_{\varphi,i}(\cdot)$ is an estimator of $g_\varphi(z)=\mathbf{E}\left[\varphi(X)|Z=z\right]$:

$$\hat{g}_{\varphi,i}(z)=\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\varphi(X_j)1\{Z_j=z\}}{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}1\{Z_j=z\}}. \tag{14}$$

For Φ_n and Ψ_n, we assume the following.

Assumption B : (i) For classes Φ_n and Ψ_n, there exist $b_\Phi,b_\Psi\in[0,1)$ and $p>2$ such that

$$\log N_{[]}(\varepsilon,\Phi_n,||\cdot||_p)<b_n\varepsilon^{-b_\Phi},\ \log N_{[]}(\varepsilon,\Psi_n,||\cdot||_p)<d_n\varepsilon^{-b_\Psi}$$

and envelopes $\tilde{\varphi}$ and $\tilde{\psi}$ for Φ_n and Ψ_n satisfy that $\mathbf{E}[|\tilde{\varphi}(X)|^p|Z]<\infty$ and $\mathbf{E}[|\tilde{\psi}(W)|^p|Z]<\infty$, a.s.
(ii) $P\{Z_i=z\}\in(\varepsilon,1-\varepsilon)$ for some $\varepsilon>0$.

Lemma B1 : Suppose that Assumption 5 holds for the kernel and the bandwidth and Assumptions B1-B2 hold. Then

$$\sup_{(\varphi,\psi)\in\Phi_n\times\Psi_n}\left|\hat{\Delta}_n(\varphi,\psi)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[\psi(W_i)|Z_i\right]\left\{\varphi(X_i)-g_\varphi(Z_i)\right\}\right|=o_P(1).$$

Proof of Lemma B1 : Define $\hat{\rho}_{\varphi,i}(z)=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}1\{Z_j=z\}\varphi(X_j)$ and $\hat{p}_i(z)=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}1\{Z_j=z\}$, and write $\hat{g}_{\varphi,i}(Z_i)-g_\varphi(Z_i)$ as

$$\begin{aligned}&\left[\hat{\rho}_{\varphi,i}(Z_i)-g_\varphi(Z_i)\hat{p}_i(Z_i)\right]\left\{1/p(Z_i)+1/\hat{p}(Z_i)-1/p(Z_i)\right\}\\=\;&\left[\hat{\rho}_{\varphi,i}(Z_i)-g_\varphi(Z_i)\hat{p}_i(Z_i)\right]/p(Z_i)+o_P(n^{-1/2}),\end{aligned}$$

where $\tilde{\delta}_{n,i} = \delta_{n,i} + \delta_{n,i}^2(1-\delta_{n,i})^{-1}$ and $\delta_{n,i} = p(Z_i) - \hat{p}(Z_i)$ and $p(z) = P\{Z_j = z\}$. By the law of large numbers, $\max_{1\leq i\leq n}\left|\tilde{\delta}_{n,i}\right| = o_P(1)$. Hence it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi(W_i)}{p(Z_i)}\left[\hat{\rho}_{\varphi,i}(Z_i) - g_\varphi(Z_i)\hat{p}_i(Z_i)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[\psi(W_i)|Z_i\right]\left\{\varphi(X_i) - g_\varphi(Z_i)\right\} + o_P(1). \tag{15}$$

Write the right-hand side as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\psi(W_i)}{p(Z_i)}\left[\hat{\rho}_{\varphi,i}(Z_i) - g_\varphi(Z_i)\hat{p}_i(Z_i)\right] = \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} q(S_i, S_j;\varphi,\psi),\ \ S_i = (W_i, Z_i, X_i),$$

where $q(W_i, Z_i, X_j;\varphi,\psi) = \psi(W_i)1\{Z_j = Z_i\}\left\{\varphi(X_j) - g_\varphi(Z_i)\right\}/p_i(Z_i)$. Since $\mathbf{E}[q(S_i, S_j;\varphi,\psi)|S_i] = 0$, the usual Hoeffding's decomposition allows us to write the above sum as

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{E}\left[q(S_i, S_j;\varphi,\psi)|S_j\right] + r_n(\varphi,\psi) + o_P(1)$$

where $r_n(\varphi,\psi)$ is a degenerate U-process. Following similar steps in Song (2007) using the maximal inequality of Turki-Moalla (1998), we can show that it is $o_P(1)$ uniformly in $(\varphi,\psi)\in\Phi_n\times\Psi_n$. Note that

$$\begin{aligned}\mathbf{E}\left[q(S_i, S_j;\varphi,\psi)|S_j\right] &= \mathbf{E}\left[\psi(W_i)1\{Z_j = Z_i\}\left\{\varphi(X_j) - g_\varphi(Z_i)\right\}|S_j\right]/p_i(Z_i)\\ &= \mathbf{E}\left[\mathbf{E}\left[\psi(W_i)|Z_i\right]1\{Z_j = Z_i\}\left\{\varphi(X_j) - g_\varphi(Z_i)\right\}|S_j\right]/p_i(Z_i)\\ &= \mathbf{E}\left[\psi(W_j)|Z_j\right]\left\{\varphi(X_j) - g_\varphi(Z_j)\right\}.\end{aligned}$$

Hence we obtain the result in (15) for each $(\psi,\varphi)\in\Psi_n\times\Phi_n$. Now, the unifomity over these latter space is obtained by the stochastic equicontinuity of the process, which follows from the bracketing entropy conditions in Assumption B. ■

References

[1] Anderson, T. W. (1955), "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proceedings of American Mathematical Society*, 6, 170-176.

[2] Anderson, T. W. and D. A. Darling (1952), "Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes," *Annals of Mathematical Statistics*, 23, 193-212.

[3] Barret, G. F. and S. G. Donald (2003), "Consistent tests for stochastic domiance," *Econometrica*, 71, 71-104.

[4] Bickel, P. J., C. A. J. Klassen, Y. Ritov, J. A. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.

[5] Birman and Solomjak (1967), "Piecewise-polynomial approximation of functions of the classes W_p^α," *Mathematics of the USSR-Sbornik* 2, 295-317.

[6] Cawley, J. and T. Phillipson (1999), "An empirical examination of information barriers to trade insurance," *American Economic Review* 89, 827-846.

[7] Chiappori, P-A, and B. Salanié (2000), "Testing for asymmetric information in insurance markets," *Journal of Political Economy* 108, 56-78.

[8] Chiappori, P-A, B. Jullien, B. Salanié, and F. Salanié (2002), "Asymmetric information in insurance: general testable implications," mimeo.

[9] Durbin, J. and M. Knott (1972), "Components of the Cramer-von Mises statistics, I." *Journal of the Royal Statistic Society Series*, B, 34, 290-307.

[10] Escanciano, J. C. (2006), "A consistent diagnostic test for regression models using projections", *Econometric Theory*, 22, 1030-1051.

[11] Eubank, R. L. and V. M. LaRiccia (1992), "Asymptotic comparison of Cramer-von Mises and nonparametric function estimation techniques for testing goodness-of-fit," *Annals of Statistics* 20, 2071-2086.

[12] Ferger, D. (1995), "The Joint Distribution of the Running Maximum and Its Location of D-Valued Markov Processes," *Journal of Applied Probability*, 32, 842-845.

[13] Janssen, A. (1995), "Principal component decomposition of non-parametric tests," *Probability Theory and Related Fields*, 101, 193-209.

[14] Janssen, A. (2000), "Global power functions of goodness of fit tests," *Annals of Statistics*, 28, 239-253.

[15] Kim, J. and D. Pollard (1990), "Cube-root asymptotics," *Annals of Statistics*, 18, 191-219.

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

[17] Linton, O., E. Maasoumi, and Y-J. Whang (2005), "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies* 72, 735-765.

[18] Milbrodt, H. and H. Strasser (1990), "On the asymptotic power of the two sided Kolmogorov-Smirnov test," *Journal of Statistical Planning and Inference*, 26, 1-23.

[19] Neuhaus, G. (1976), "Asymptotic power properties of the Cramér-von Mises test under contiguous alternatives," *Journal of Multivariate Analysis*, 6, 95-110.

[20] Song, K. (2007), "Testing Conditional Independence using Rosenblatt Transforms," Working Paper.

[21] Strasser, H. (1990), "Global extrapolations of local efficiency," *Statistics and Decisions,* 8, 11-26.

[22] Stute, W. (1997), "Nonparametric model checks for regression," *Annals of Statistics* 25, 613-641.

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

[24] Turki-Moalla, K. (1998), "Rates of convergence and law of the iterated logarithm for U-processes," *Journal of Theoretical Probability* 11, 869-906.

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

Figure 1: An Example of the Distribution Function Under the Alternatives

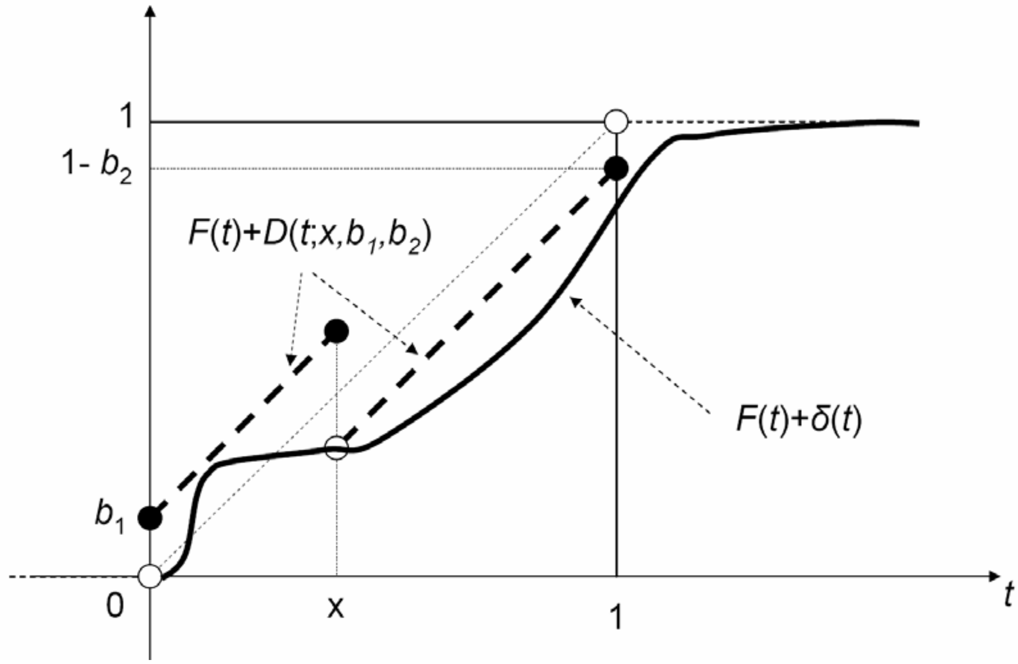

Figure 2: The Upper Bound for the Asymptotic Local Power of the Test with b2=0.2

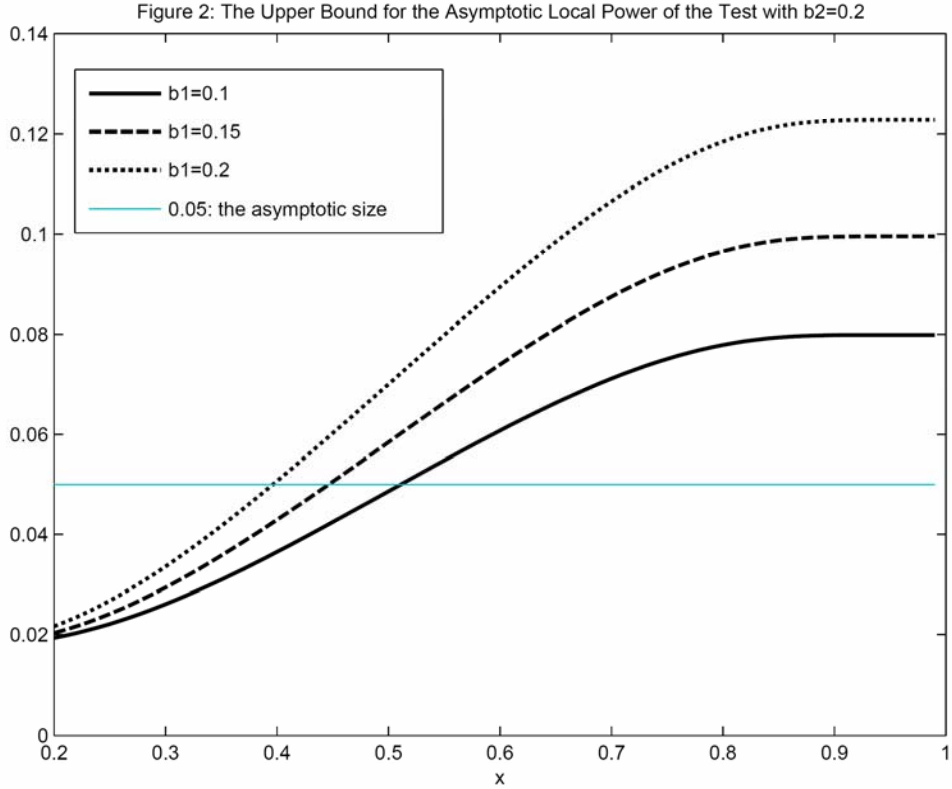

Figure 3: The Upper Bound for the Asymptotic Local Power of the Test with Varying b

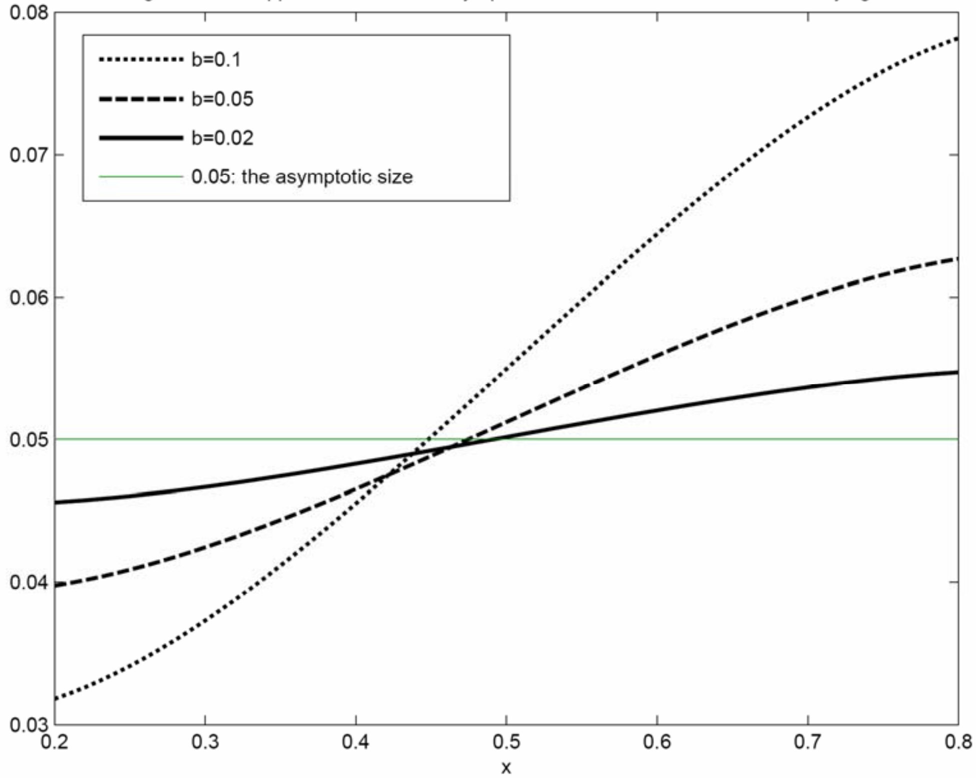

Figure 4: The Distribution Function Under the Alternatives